

Міністерство освіти і науки України

Системні технології

System technologies

5 (142) 2022

Регіональний міжвузівський збірник наукових праць

Засновано у січні 1997 року.

У випуску:

- ПРОГРЕСИВНІ ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ
ТА ОРГАНІЗАЦІЯ СУЧАСНОГО ВИРОБНИЦТВА**
- МАТЕМАТИЧНЕ ТА ПРОГРАМНЕ ЗАБЕЗПЕЧЕННЯ
ІНТЕЛЕКТУАЛЬНИХ СИСТЕМ**
- СИСТЕМНІ ТЕХНОЛОГІЇ ОБРОБКИ ІНФОРМАЦІЇ
ТА КІБЕРБЕЗПЕКА**

Системні технології. Регіональний міжвузівський збірник наукових праць. – Випуск 5 (142). - Дніпро, 2022. – 125 с.
ISSN 1562-9945 (Print).
ISSN 2707-7977 (Online).

Редакційна колегія випуску:

Алпатов А.П. - д.т.н., проф. (відп. редактор)
Архипов О.Є. - д.т.н., проф.
Білозьоров В.Є. - д.ф.-м.н., проф.
Бабічев С.А. (Чеська Республіка) - д.т.н., доц.
Єрьомін О.О. - д.т.н., проф.

Прогресивні інформаційні
технології та організація
сучасного виробництва

Гече Ф.Е. - д.т.н., проф., (відп. редактор)
Гуда А.І. - д.т.н., проф.
Гнатушенко Вік.В. - д.т.н., проф.
Скалозуб В.В. - д.т.н., проф.

Математичне
та програмне забезпечення
інтелектуальних систем

Гнатушенко В.В. - д.т.н., проф., (відп. редактор)
Гожий О.П. - д.т.н., проф.
Кіріченко Л.О. - д.т.н., проф.
Светличний Д.С. (Польща) - д.т.н., проф.
Хандецький В.С. - д.т.н., проф.

Системні технології
обробки інформації
та кібербезпека

Збірник друкується за рішенням Вченої Ради
Українського державного університету науки і технологій
від 28.03.2022 р., № 5

Адреса редакції: 49600, Дніпро, пр. Гагаріна, 4
Український державний університет науки і технологій,
ННІ «Інститут промислових та бізнес технологій»
кафедра Інформаційних технологій та систем.
Тел. +38(097)6854525
E-mail: st@nmetau.edu.ua
<https://journals.nmetau.edu.ua/index.php/st>

© Український державний університет науки і технологій,
ННІ «Інститут промислових та бізнес технологій»,
ІВК «Системні технології», 2022

Н.В. Карпенко, В.І. Стаценко, Н.І. Федоряцька, І.І. Коломоєць
**ВПРОВАДЖЕННЯ СТАРТАПІВ ПІД ЧАС ПІДГОТОВКИ ФАХІВЦІВ
У ГАЛУЗІ ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ**

Анотація. У статті наведено погляд на освітній процес з точки зору різних стейкхолдерів – зі сторони освітян, роботодавців та зі сторони здобувачів вищої освіти. Порівняно розвиток стартап екосистеми в Україні та в Європі, показано їхній вплив на економічний зріст країни. Авторами запропоновано комплексний підхід до впровадження стартапів в освітній процес під час підготовки спеціалістів для ІТ-галузі та показано переваги, що отримує кожен зі стейкхолдерів від цього впровадження.

Ключові слова: стартап, освітній процес, ЗВО, стейкхолдери, ІТ-технології, інноваційна діяльність в ЗВО.

Вступ. Статистичні дослідження показують зростання зацікавленості мешканців України в отриманні вищої освіти. Кількість людей відповідного віку, що отримали вищу освіту з 2000 р. по 2014 р. зросла майже на 30% і досі попит на диплом про вищу освіту не знижується. За даними громадської спілки «Освіторія», 82% українців мають вищу освіту станом на 2018 рік. Одночасно з цим роботодавці повідомляють про наявність у працівників, які щойно закінчили заклад вищої освіти (ЗВО), суттєвих прогалин у навичках, що заважає досягненню цілей бізнесу [1]. З іншого боку, опитування показує, що серед тих, хто отримав чи отримує вищу освіту, менше половини студентів задоволені набутими знаннями та навичками. Цікавим є те, що лише третина з них працює за отриманою спеціальністю, ще третині потрібна була перекваліфікація, інші не змогли продати свої знання та навички [2].

Аналіз останніх досліджень і публікацій. На сайті DOU періодично проводять опитування про ЗВО, в яких навчаються чи навчалися ІТ-фахівці. Статистика показує, що за шкалою від 0 до 10 балів актуальність знань фахівці оцінили в 4 бали. Також відмічається слабка технічна база ЗВО, замало практики та те, що знання не досить корисні для роботи в ІТ галузі. Оцінки різних параметрів навчання коливаються від 5 до 6 балів за 10-бальною шкалою, рідко пе-

ревищуючи 6 балів. Максимально високо оцінено престижність факультету – 6,5 у середньому по всіх вишах.

З іншого боку, викладачі ЗВО не задоволені слабкою математичною підготовкою школярів. Так міжнародні дослідження PISA показали, що 36% українських 15-річних учнів не досягають базового рівня знання з математики, 25,9% не здатні належним чином розуміти та використовувати письмові тексти, у 26,4% – немає мінімальних природничих знань [3]. Різниця між успіхами наших школярів та учнів із країн Організації економічного співробітництва та розвитку еквівалентна відставанню в один рік навчання.

На жаль, сьогодні ми у вищій освіті маємо застарілу матеріально-технічну базу, яку майже не оновлювали протягом останніх років та розрив між ЗВО та роботодавцями, які, в свою чергу, мають зараз дефіцит кваліфікованих кадрів. Катастрофічних масштабів досягає демотивація студентів [4], відволікання їх на підробітки під час навчального процесу. І навіть якщо студент паралельно з навчанням працює за фахом, він на роботі отримує практичні навички по вузькій спеціалізації, що не забезпечує набуття всіх компетентностей, які повинна надавати вища освіта. Окрім цього, не менш важливим чинником є втрата популярності у школярів інженерних спеціальностей, за винятком ІТ-галузі, популярність якої забезпечується бурхливим розвитком галузі у цілому, нестачею кадрів та можливістю працювати віддалено у міжнародних командах.

Знаючи таке положення речей, Український інститут майбутнього розробив кроки подолання кризи в освіті аж до 2030 року [5]. Свої поради щодо реформування вищої освіти в Україні надав Світовий Банк [1]. Окрім цього, існує дуже багато статей на сайті Освіторії про можливі шляхи розвитку освіти, роздуми про те, що навчання є трендом чи вимогою часу тощо. Але усі ці матеріали є досить узагальненими і залишають місце для деталізації кожного кроку. Таким чином, в даній статті буде розглянуто переваги та особливості введення стартапів в освітню програму.

Основна частина. У рейтингу Global Startup Ecosystem 2021 Україна посіла 34 місце серед інших країн за стартапами, незважаючи на колосальний потенціал нашої економіки, але війна сильно вплинула на суспільство, завдяки чому Україна в 2022 р. втратила 16 позицій та посіла 50-е місце. В Україну стартапи зайшли не так давно, але серед них вже є проекти Grammarly та GitLab, які набули світової знанності. І хоча в Україні найсприятливішими визначають лише кілька міст – Київ, Одеса, Львів та Харків (Дніпро та Тернопіль випали зі

списку тисячі найкращих міст у 2022 р), стартапами займаються майже всі області України.

Порівнюючи країни, які займають 1 та 2 сходинку в рейтингу найсильніших економік світу, можна побачити, що Китай почав наздоганяти США різко підвищивши кількість державного замовлення на підготовку інженерів (рис. 1), у той час як США залучають багато спеціалістів з інших країн. Індія за ВВП стоїть на 3 місці. Слід відзначити, що всі ці країни надають перевагу STEM освіті. Таким чином, посилюючи STEM освіту можна позитивно вплинути на економічний розвиток країни.

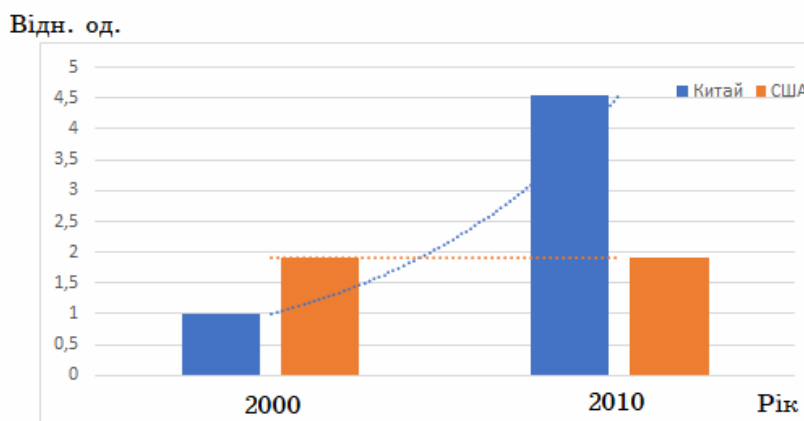


Рисунок 1 – Випуск бакалаврів-інженерів

Підготовку спеціаліста можна розглядати як випуск будь-якої продукції – тобто існують зацікавлені особи, бізнес процеси та ризики, які можуть вплинути на досягнення цілей стейкхолдерів. Відбувається цей процес, звісно, в рамках певної системи. Для того, щоб система працювала належним чином, вона повинна бути гнучкою та сталою [6]. В західній моделі стартап екосистеми всі стейкхолдери (державна, університет, підприємці, корпорації) гармонійно взаємодіють між собою, причому університети грають ключову роль в екосистемі. Не менш важливим чинником є інвестиції, які надходять до університетів від інших стейкхолдерів. В Україні, на відміну від західної моделі стартап екосистеми, між представниками освіти (міністерство освіти, ЗВО, випускаюча кафедра, викладачі) і роботодавцями в більшості випадків не існує співпраці (рис. 2). Інколи ЗВО декларує угоду про співпрацю з ІТ-компанією і, найчастіше, такі наміри залишаються лише на папері. Рушіями співпраці випускаючої кафедри з ІТ-компанією, чи, навіть, з ІТ-community, є викладачі, які займають активну життєву позицію та вболівають за своє діло. Але цього замало, оскільки така співпраця тримається лише на ентузіазмі конкретного викладача і доброзичливості працівника ІТ-компанії, з яким йде співпраця. Оскільки все це

ніким не оплачується та забирає певний час, який потрібно відривати від основної роботи чи від свого відпочинку, то така співпраця не може тривати довгий час. З іншого боку, мотивований студент одночасно навчається в ЗВО і на курсах, які проводять ІТ-компанії. І як тільки студента взяли в лабораторію якоїсь ІТ-компанії, як правило, він закидає навчання. Таким чином, ЗВО «втрачає» мотивованих студентів, оскільки не може їм запропонувати нічого, окрім навчання.

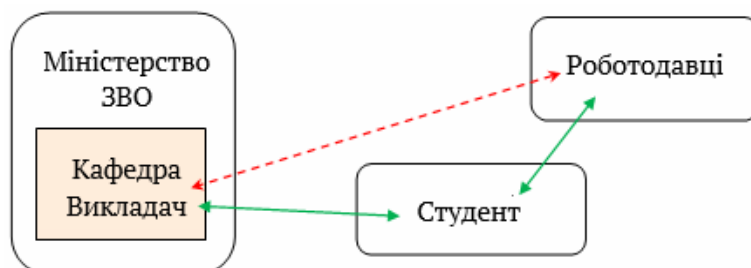


Рисунок 2 – Зв'язок між стейкхолдерами в Україні

Зміна стратегії підготовки спеціаліста із залученням стейкхолдерів із ІТ-галузі дуже важлива, оскільки студентам більш цікаво навчатися на реальних прикладах, приймати посильну участь у вирішенні актуальних задач, які існують у галузі. Таким чином, вони отримують мотивацію, набувають корисні навички. Одночасно зі студентами розвиваються і викладачі: вони набувають актуальні знання і зможуть надавати студентам корисні поради виходячи з власного практичного досвіду, а не лише теоретичні знання.

Перехід на освіту «за компетенціями» вимагає докорінної зміни парадигми мислення та зміни традиційних методів викладання. Як показує досвід роботи університетів Ізраїлю, Нідерландів та інших країн ЄС, дуже добрі результати дає розвиток стартап культури в університетах, що дозволяє розвинути підприємницькі навички у студентів та викладачів. При цьому, інновації та підприємництво стають головними драйверами зростання та виходу університетів на новий рівень якісної освіти та взаємодії з підприємствами. Механізм полягає в створенні серед студентів ЗВО команд, до складу яких можуть входити студенти різних спеціальностей. Команди під час навчання, одночасно з опануванням базових знань за фахом, колективно працюють над власним інноваційним стартап-проектом, генерують гіпотези стосовно створення власного Продукту, його виробництва, бізнес-моделі, каналів просування на ринки та драйверів зростання. Цей процес проходить за допомогою викладачів при менторській підтримці з боку бізнес-інкубаторів та акселераторів, які залучають

представників провідних підприємств та фахівців у якості менторів. Проміжними етапами є публічні захисти (пітч) проектів перед журі, до складу якого включені представники ЗВО, підприємці та інвестори. Команди, які підготували якісні проекти, отримують шанс потрапити до акселераторів та отримати фінансову підтримку для подальшого розвитку власного проекту [7]. Така практика дозволить мотивувати студентів до навчання, сформувати навички командної роботи над проектами, в тому числі і в «дистанційному режимі» за допомогою сучасних інформаційних технологій [8], що є фактично нормою в сучасній інноваційній стартап практиці. Також студенти отримують можливість застосовувати на практиці свої знання та починають свідомо будувати власну кар'єру як фахівця.

Ми вважаємо, що наступні зміни освітнього процесу підвищать якість підготовки спеціалістів:

1) Студенти самостійно обирають цікаву їм тему, або тим, хто зацікавлений, стейкхолдери пропонують проблемні теми, з яких можна розвинути стартап.

2) Формування стартап команд відбувається з числа студентів, викладачів і представників стейкхолдерів.

3) Практичні та лабораторні роботи розробляються кафедрами спільно з майбутніми роботодавцями для відпрацювання навичок, без яких студент, який мріє працювати в ІТ-галузі, не зможе бути залученим для роботи над реальним проектом.

4) ІТ-компанії допомагають викладачам в розробці актуальних навчальних курсів з врахуванням сучасних тенденцій та потреб ІТ-галузі.

5) Вибіркові курси є частиною навчальної програми студента. Для того, щоб зацікавити викладачів у засвоєнні новітніх технологій або підготовці дисципліни, яка надасть більш глибокі знання з одного з напрямків інформаційних технологій, слід зробити так, щоб вибіркові курси могли обирати не лише студенти за навчальним планом, але й будь-хто інший, але на платній основі. Пропорційна матеріальна винагорода викладачу, який зміг зібрати велику кількість слухачів на своєму курсі, також принесе дохід й ЗВО.

6) Для підвищення валідності навчальних посібників, що написані викладачами ЗВО, їх рецензування слід надати представникам ІТ-галузі (з подальшою реалізацією підручників на ринку).

7) Гнучке планування освітнього процесу. Для тих, хто приймає участь в стартапі є можливість дострокового зарахування дисциплін, отже менший термін навчання студента.

8) Формування фонду для фінансування найбільш вдалих стартапів за рахунок учасників стартап екосистеми та інвесторів.

9) Введення в МОН критерію "Інноваційна діяльність" та "Стартап-діяльність" (для рейтингу ЗВО / акредитації / відповідності).

Впровадження цих кроків приведе до того, що всі студенти будуть отримувати більш якісні знання. Робота над стартапами є опціональною, тобто ті студенти, які не мають бажання займатися стартапами після успішного навчання, як і раніше, будуть отримувати диплом, у той час як ті, хто пройшов стартап додатково отримують сертифікат (рис. 3).

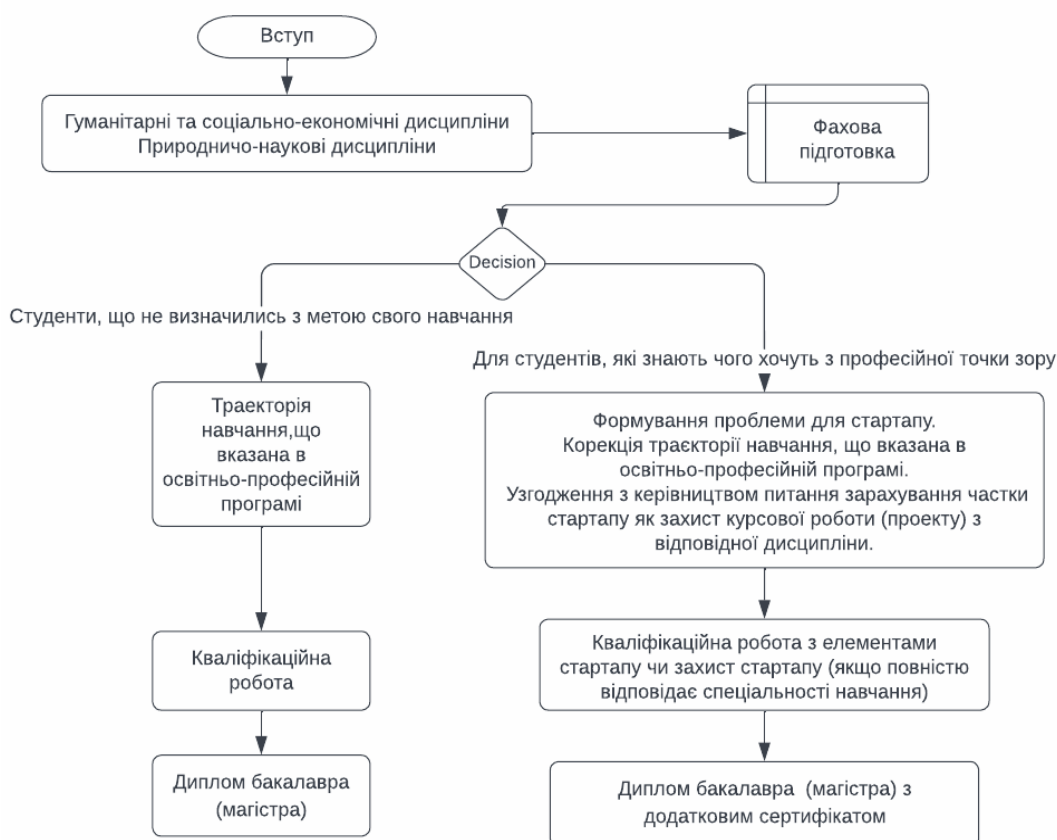


Рисунок 3 – Діаграма запропонованих змін

Висновки. Реорганізований таким чином освітній процес дозволить:

- Виявляти найталановитіших студентів, які будуть позитивно впливати на рівень економічного розвитку нашої країни.

- Завдяки тому, що студенти набули практичні навички, роботодавцям не потрібно буде додатково витратити час і гроші на “донавчання”, а студенти вже зробили перший реальний крок у своїй кар’єрі.

- З-ВО та викладачі мають гроші з реалізації підручників і проведення оплачуваних вибіркового курсів, а також додаткові бали до рейтингу за рахунок критерію інноваційної діяльності.

- Збільшення відсотку студентів, які працюють за спеціальністю буде підвищувати престижність випускаючої кафедри та ЗВО.

ЛІТЕРАТУРА

1. Як реформувати українську освіту — рекомендації Світового Банку. [Електронний ресурс]. Режим доступу: <https://osvitoria.media/experience/yak-reformuvaty-ukrayinsku-osvitu-rekomendatsiyi-svitovogo-banku/> — Загол. з екрана (дата звернення 20.09.2022).

2. Як зміна поведінки українців може зробити країну заможнішою. [Електронний ресурс]. Режим доступу: <https://www.epravda.com.ua/columns/2021/09/16/677896/> — Загол. з екрана (дата звернення 20.09.2022).

3. Нация безработных бакалавров. Шесть проблем образования в Украине. [Електронний ресурс]. Режим доступу: <https://www.liga.net/society/articles/natsiya-bezrobotnyh-bakalavrov-shest-problem-obrazovaniya-v-ukraine> — Загол. з екрана (дата звернення 20.09.2022).

4. Результати дослідження: українським студентам не вистачає мотивації для навчання. [Електронний ресурс]. Режим доступу: <http://vnz.org.ua/statti/9945-rezultaty-doslidzhennja-ukrayinskim-studentam-ne-vystachae-motyvatsiyi-dlja-navchannja> — Загол. з екрана (дата звернення 20.09.2022).

5. Україна – learning nation. [Електронний ресурс]. Режим доступу: <https://strategy.uifuture.org/ukraina-learning-nation.html> — Загол. з екрана (дата звернення 20.09.2022).

6. Стаценко В.І., Федоряцька Н.І. Особливості керування ризиками на ранніх етапах розвитку інноваційних проектів для підвищення енергетичної ефективності технічних пристроїв перетворення енергії відновлюваних джерел. // Відновлювана енергетика. – 2019. – № 3 (58). – с. 6-13.

7. Стаценко В. І. Актуальні питання підготовки спеціалістів в умовах реформи вищої освіти. Розвиток стартап культури в університетах як важливий рушій якості освіти / XI Міжнародна науково-практична конференція «Сучасні енергетичні установки на транспорті, технології та обладнання для їх обслуговування». Херсон ХДМА, 8-10 вересня 2020 р. – с. 324-326.

8. Литвинов А. А., Карпенко Н. В. Особенности применения личностно-деятельного подхода при подготовке специалистов в области информационных технологий. //Системні технології. Збір. наук. праць. — Вип. 1(114) — 2018, с. 136-141.

REFERENCES

1. How to reform Ukrainian education — recommendations of the World Bank. <https://osvitoria.media/experience/yak-reformuvaty-ukrayinsku-osvitu-rekomendatsiyi-svitovogo-banku/>
2. How changing the behavior of Ukrainians can make the country more prosperous. <https://www.epravda.com.ua/columns/2021/09/16/677896/>
3. Nation of unemployed bachelors. Six problems of education in Ukraine. <https://www.liga.net/society/articles/natsiya-bezrobotnyh-bakalavrov-shest-problem-obrazovaniya-v-ukraine>
4. Research results: Ukrainian students lack motivation to study. <http://vnz.org.ua/statti/9945-rezultaty-doslidzhennja-ukrayinskym-studentam-ne-vystachae-motyvatsiyi-dlja-navchannja>
5. Ukraine – learning nation. <https://strategy.uifuture.org/ukraina-learning-nation.html>
6. Statsenko V. I., Fedoryatskaya N.I., Features of risk management in the early stages of the development of innovative projects to enhance the energy efficiency of technical devices for conversing energy.// No. 3(58) (2019): Scientific and Applied Journal Vidnovlavana energetika <https://ve.org.ua/index.php/journal/article/view/210>
7. Statsenko V. I. Actual issues of specialist training in the conditions of higher education reform. Development of start-up culture in universities as an important driver of education quality. https://ksma.ks.ua/wp-content/uploads/2021/02/seutto_2020.pdf – p. 324-326.
8. Litvinov A.A., Karpenko N.V. Peculiarities in applying the personal-active approach to specialists’ training in information technologies// System technologies. – N 1 (114). – Dnipro, 2018. – p. 136-141.

Received 01.06.2022.

Accepted 07.06.2022.

Implementation of startups during the training of specialists in the field of information technology

Now, the training of a specialist is considered as an output. When students enter higher education, the curriculum is drawn up for them, which is strictly followed throughout subsequent years, for example, four years for the bachelor. However, this approach does not suit motivated students who lack creativity along with actual knowledge. Thus, these students, in parallel with higher education, attend special courses

in IT-companies. At the end of the course, students start working for an IT company and lose motivation for further training at the university.

The authors of the article offer a comprehensive approach to the introduction of startups in the educational process of universities during the training of specialists for the IT-field. Attraction to the educational process of stakeholders from IT-companies to solve actual problems will lead to the fact that students will be involved in the creative process, in which they will receive useful knowledge and skills not only in the specialty, but also learn to communicate in a team. The introduction of startups will lead to an increase in the number of motivated students, because students can for creativity to choose not only topics proposed by stakeholders, but also to offer their ideas. Close cooperation with stakeholders from the IT-field will lead to an improvement in the quality of the materials that teachers will offer students for training and will also allow universities to earn on author courses and textbooks written by teachers. The authors of the article propose, together with start-ups to introduce flexible planning of the educational process for students who take part in start-ups, i.e. each motivated student will be able to engage in their educational trajectory and after graduation to obtain not only a diploma, But also certificates that will confirm the skills gained in startups. The article also indicates other benefits that each of the stakeholders will receive from the introduction of startups in the learning process.

Карпенко Надія Валеріївна - к.ф.-м.н., доцент кафедри ЕОМ Дніпровського національного університету імені Олеся Гончара.

Karpenko Nadiia - Ph.D in Physics and Mathematical Sciences, Associate Professor, Computer Engineering Department, Oles Honchar Dnipro National University.

Стаценко Володимир Іванович - старший викладач кафедри РЕА Дніпровського національного університету імені Олеся Гончара.

Statsenko Volodymyr - Senior Lecturer, Radioelectronic Automation Department, Oles Honchar Dnipro National University.

Федоряцька Надія Іванівна - аспірант кафедри РЕА Дніпровського національного університету імені Олеся Гончара.

Fedoriatska Nadiia - graduate student, Department of Radioelectronic Automation, Oles Honchar Dnipro National University.

Коломоець Ілона Іванівна - магістр 2 курсу кафедри ЕОМ Дніпровського національного університету імені Олеся Гончара.

Kolomoiets Ilona - 2nd year master's student, Computer Engineering Department, Oles Honchar Dnipro National University.

ВИМІРЮВАННЯ СКЛАДОВИХ ІНДУКТИВНОГО ІМПЕДАНСУ НА НИЗЬКИХ ЧАСТОТАХ

Анотація. Досліджені можливості методу трьох вольтметрів для виміру індуктивності та її активного опору, а також їх частотних характеристик. Встановлено, що в діапазоні частот до 100 кГц похибка виміру індуктивності складає не більше 0,3%, активного опору - 8,7%, добротності – 5%.

Ключові слова: індуктивність, опір, імпеданс, метод трьох вольтметрів, вимір, частотна характеристика, добротність, похибка, зразкова міра.

Вступ. Серед методів виміру індуктивного імпедансу виділяється метод трьох вольтметрів [1,2], який використовують на промислових частотах (50, 400 Гц) для визначення потужності, споживаної приймачем електричної енергії, його активного опору, реактансу і повного опору. Порівняння методів Q-метра і мостового, використовуваних для виміру складових імпедансу в широкому діапазоні частот, з методом трьох вольтметрів показує, що останній має безперечну перевагу - експлуатаційна простота і зручність, що полягає в тому, що для виміру параметрів приймача досить усього лише визначити напругу на трьох ділянках електричного кола, що складається з приймача електричної енергії і послідовно включеним зразковим активним опором.

Постановка завдання. Оцінити можливості методу трьох вольтметрів для виміру складових індуктивного імпедансу, а також визначення їх частотних характеристик в діапазоні вище за промислові частоти.

Основна частина. На рис.1 представлена еквівалентна схема вимірювального кола (а) для визначення складових індуктивного імпедансу методом трьох вольтметрів і її векторна діаграма (б), де E - напруга генератора гармонійних коливань, R_o - зразковий опір, L, r - індуктивність і активний опір вимірюваного імпедансу z_x , U_o, U_x - спад напруги на зразковому опорі і вимірюваному імпедансі, U_L, U_r - реактивна і активна складові напруги U_x, I - струм у ви-

мірjuвальному колу. Імпеданси вимірюваної індуктивності z_x і вимірювального кола z мають вигляд

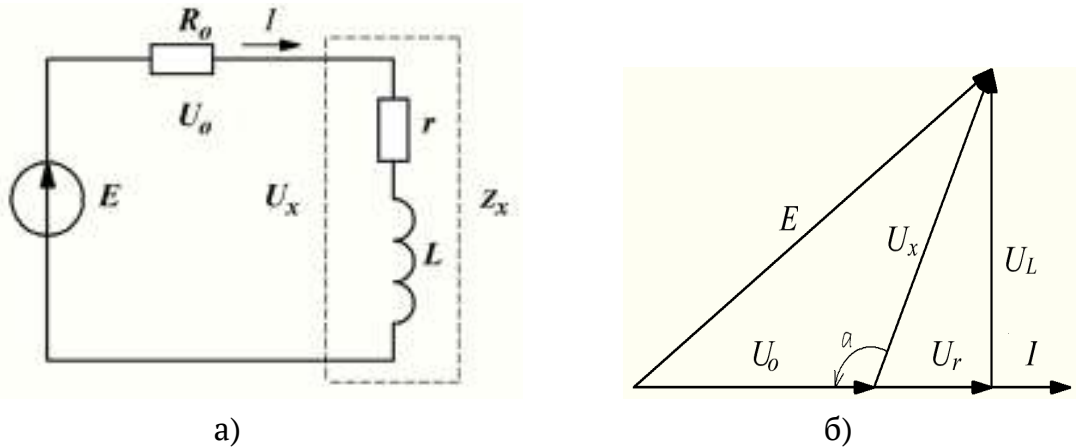


Рисунок 1 – Вимірювальне коло (а) і його векторна діаграма напруг і струму (б)

$$z_x = \sqrt{r^2 + \omega^2 L^2}, \quad (1)$$

$$z = \sqrt{((R_0 + r))^2 + \omega^2 L^2}, \quad (2)$$

де ω - частота коливань гармонійного генератора E . Для послідовного кола значення імпедансів можна представити як

$$z_x =, \quad z = \frac{R_0 E}{U_0}. \quad (3)$$

Із (1) – (3) випливає система рівнянь

$$\sqrt{r^2 + \omega^2 L^2} = \frac{R_0 U_x}{U_0}, \quad (4)$$

$$\sqrt{((R_0 + r))^2 + \omega^2 L^2} = \frac{R_0 E}{U_0}, \quad (5)$$

рішення якої дозволяє отримати вирази для активної складової індуктивного імпедансу

$$r = \frac{R_0}{z} \left(\frac{E^2 - U_x^2}{U_0^2} - 1 \right) \quad (6)$$

і його індуктивності

$$L = \frac{R_0}{\omega} \sqrt{\frac{U_x^2}{U_0^2} - \frac{1}{4} \left(\frac{E^2 - U_x^2}{U_0^2} - 1 \right)^2}. \quad (7)$$

З двох останніх виразів виходить, що величини r і L можна знайти виміром напруг E , U_x і U_0 з подальшим обчисленням по (6) і (7). Слід зазначити, що знайдені складові індуктивного імпедансу r і L будуть дійсними і додатними при виконанні умови

$$U_x^2 + U_0^2 \leq E^2 \leq (U_x + U_0)^2, \quad (8)$$

на векторній діаграмі (рис. 1б) це означає, що кут α може набувати значень в обмеженому діапазоні $\pi/2 \leq \alpha \leq \pi$.

Для визначення можливості виміру складових індуктивного імпедансу була зібрана вимірювальна схема згідно рис.1а. В якості вимірюваного імпедансу використані котушки з набору зразкових мір індуктивності типу L - 0170 III розряду класу 0,2. Виміри проводилися на частоті 1 кГц вольтметром В7-27А/1 з використанням генератора гармонійних коливань ГЗ-118. Зразковим опором R_0 було вибрано резистор типу МЛТ- 2 з активним опором 1475 Ом (на частоті 1 кГц).

У таблиці 1 представлені результати вимірів, де L_0 - номінальні значення зразкових мір індуктивності; L_∂ , r_∂ - дійсні значення індуктивностей і активного опору мір, отримані мостовим вимірювачем Е7-8; L , r - індуктивність і активний опір зразкових мір, визначені методом трьох вольтметрів.

Таблиця 1

Результати вимірів зразкових мір індуктивності

L_0 , мГн	0,2	0,5	1	2	5	10	20	50	100	200
L_∂ , мГн	0,201	0,501	1	1,999	4,997	10	20,01	49,99	100	199,9
r_∂ , Ом	1,612	1,944	2,97	4,73	9,98	20,6	75,2	125,5	294	241
L , мГн	0,225	0,518	0,997	1,984	4,907	9,99	19,9	49,98	99,97	200,5
r , Ом	1,499	1,999	2,987	4,948	11,67	20,7	76,92	125,6	292,7	217,6

Використовуючи метод найменших квадратів отримані рівняння лінійної регресії для індуктивності $L = f(L_\partial)$

$$L = - 0,0641 + 1,0026L_\partial \quad (9)$$

і активного опору $r = \varphi(r_\partial)$

$$r = 1,0182 + 0,9601r_\partial; \quad (10)$$

при рівні значущості 0,001 коефіцієнти кореляції склали для індуктивності $r_L \approx 1$, для активного опору $r_r \approx 0,9983$; середньоквадратичне відхилення окремого виміру індуктивності $\sigma_L = 0,1168$ мГн, активного опору $\sigma_r = 6,5419$ Ом, середня відносна похибка виміру індуктивності склала 0,3%, активного опору - 8,7%.

Результати дослідження частотних характеристик показані на рис. 2.

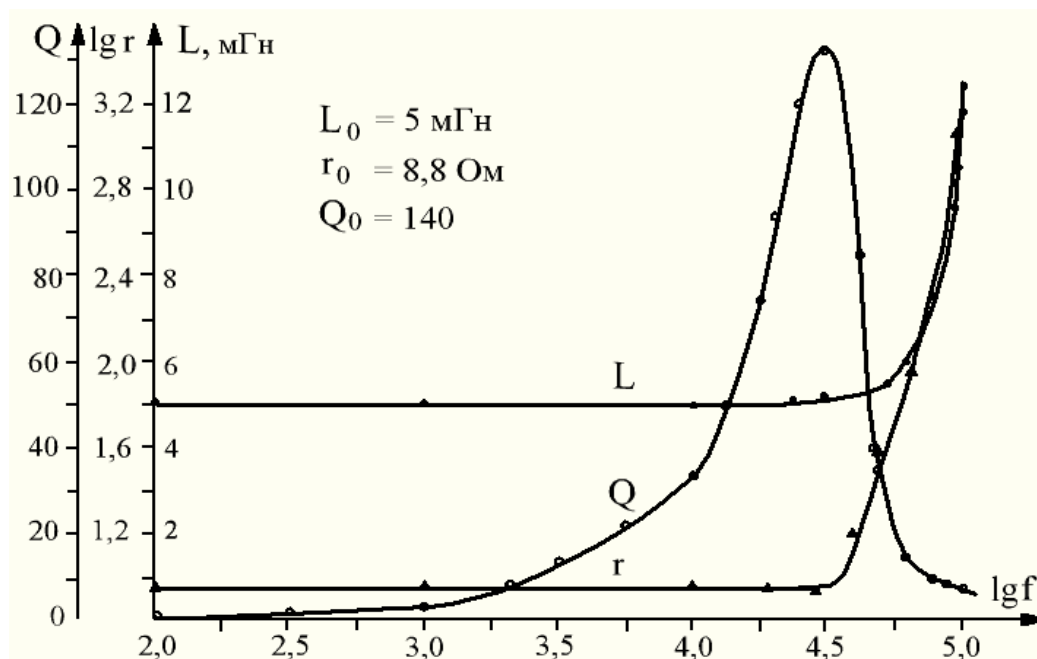


Рисунок 2 - Частотні характеристики індуктивності L , активного опору r і добротності Q зразкової міри індуктивності L_0

Експериментальні частотні залежності індуктивності L , активного опору r і добротності Q визначені в діапазоні 0,1 – 100 кГц. В якості досліджуваного індуктивного імпедансу була використана зразкова міра з індуктивністю $L_0 = 5$ мГн, омичним опором 8,8 Ом і добротністю 140. З отриманих залежностей видно, що в діапазоні частот від 0.1 до 30 кГц активний опір r і індуктивність L залишаються практично незмінними з відхиленнями в межах похибки експерименту. Після 30 кГц активний опір r починає збільшувати своє значення, що можна пояснити зростаючим впливом ефекту близькості і скін-ефекту. Також спостерігається збільшення індуктивності L завдяки шунтуючому впливу розподіленої ємності котушки і вимірювального кола в цілому, які утворюють з індуктивністю паралельний контур з певною частотою власного резонансу f_0 . Із зростанням частоти f при $f < f_0$ імпеданс паралельного контуру збільшується, що викликає уявне збільшення індуктивності L . Це явище вносить похибку у визначення індуктивності на частотах, що наближаються до частоти власного резонансу котушки. Похибку можна зменшити за умови врахування величини розподіленої ємності, для цього знадобиться розгляд дещо іншої еквівалентної схеми вимірювального кола, в якому враховується міжвиткова ємність вимірюваної котушки, входні ємності вимірювальних приладів і з'єднувальних провідників.

Частотна залежність добротності $Q = 2\pi fL/r$ досліджуваної зразкової міри індуктивності має максимум $Q = 133$ на частоті $f = 30$ кГц, величина якого відрізняється на 5% від добротності, вказаної в паспорті зразкової міри.

Висновки. Результати дослідження показують:

- методом трьох вольтметрів досить просто вимірювати складові індуктивного імпедансу на фіксованих частотах вище промислових; у діапазоні до 100 кГц відносна похибка визначення індуктивності складає не більше 0,3%, активного опору - 8,7%, добротності - 5%, що відповідає класу точності використаних вимірювальних приладів;

- метод дозволяє отримувати інформативну характеристику у вигляді частотних залежностей активного опору і індуктивності; це дає можливість вибрати діапазон частот, в якому заважаючий вплив розподіленої ємності незначний, а також знайти оптимальну частоту, при якій добротність індуктивності максимальна;

- подальше підвищення точності визначення дійсних значень складових індуктивного імпедансу можливо з урахуванням шунтуючого впливу розподіленої ємності.

ЛІТЕРАТУРА

1. Ознакомление с методом трех вольтметров: веб-сайт.
URL: <http://www.vunivere.ru/work/36704> (дата звернення: 23.08.2022).
2. Твердоступ Н.И. Особенности измерения индуктивного импеданса на низких частотах. *Системні технології*. Дніпропетровськ, 2016. Вип. 1(102). С. 87 - 92.

REFERENCES

1. Oznakomlenie s metodom threh voltmetrov: web-sait.
<http://www.vunivere.ru/work/36704> (data zvernennja: 23.08.2022).
2. Tverdostup N.I. Osobennosti izmereniya induktivnogo impedansa na nizkih tchastotah. *Systemni tehnologii*. Dnipropetrovsk, 2016. Vyp. 1(102). S. 87 – 92.

Received 03.06.2022.

Accepted 10.06.2022.

Measurement of inductive components impedance at low frequencies

The method of three voltmeters allows you to determine the components of the inductive impedance by measuring the voltages on the tested impedance, on the reference active resistance and at the output of the harmonic voltage source that feeds the measuring circuit. The analysis of the measuring circuit using the method of three voltmeters was carried out, mathematical expressions were found for calculating the active resistance and inductance as components of the investigated impedance based on the

results of the measurement of three voltages. Regarding the values of the voltages on the investigated impedance, on the reference active resistance and on the output of the harmonic voltage source, there is a well-founded condition under which the obtained results will be valid. Using the method of three voltmeters, the frequency characteristics of the active resistance, inductance and Q-factor of the reference coil were experimentally determined in the range up to 100 kHz from a set of exemplary inductance measurements with previously known parameters. Experimental frequency dependences indicate the presence of a region of constant value of the impedance components at low frequencies. At frequencies higher than 30 kHz, there is a region of growth of the active component due to the influence of the proximity effect and skin effect and a region of apparent increase in inductance due to the influence of the distributed capacitance of the coil. The obtained characteristics make it possible to choose a frequency range in which the interfering influence of the distributed capacitance is minimal, as well as to determine the optimal frequency at which the Q-factor of the inductance is maximal. The relative error of inductance determination was no more than 0.3%, active resistance – 8.7%, Q factor – 5%.

Твердоступ Микола Іванович – доцент, к.т.н., доцент кафедри електронних обчислювальних машин Дніпровського національного університету імені Олеся Гончара.

Tverdostup Mykola Ivanovych – Associate Professor of Computer Systems Engineering Department of the Oles Honchar Dnipro National University.

**DETECTION OF ATTACKS OF THE U2R CATEGORY
BY MEANS OF THE SOM ON DATABASE NSL-KDD**

Annotation. Creating an effective system for detecting network attacks requires the use of qualitatively new approaches to information processing, which should be based on adaptive algorithms capable of self-learning. The mathematical apparatus of the Kohonen self-organizing map (SOM) was used as a research method. Python language with a wide range of modern standard tools was used as a software implementation of the Kohonen SOM addition, this section compiles the Python software model «SOM_U2R» using a Kohonen SOM. Created «SOM_U2R» software model on database NSL-KDD an error research was performed for different number of epochs with different map sizes. On the «SOM_U2R» model the research of parameters of quality of detection of attacks is carried out. It is determined that on the «SOM_U2R» created software model the error of the second kind of detection of network classes of attacks Buffer_overflow and Rootkit is 6 %, and for the class Loadmodule reached 16 %. In addition, a survey of the F-measure was conducted for a different number of epochs of learning the Kohonen SOM. It is determined that for all network attack classes (except Buffer_overflow) the F-measure increases, reaching its maximum value at 50 epochs.

Keywords: category; class; NSL-KDD; SOM; Python; error; epoch; F-measure.

Formulation of the problem. Creating an effective system for detecting network attacks requires the use of qualitatively new approaches to information processing, which should be based on adaptive algorithms capable of self-learning. The most promising direction in the creation of such systems for detecting attacks on a computer network is the use of neural network technologies, which confirms the relevance of the topic of this work.

The aim of this article is to identify U2R network classes by means of a Kohonen SOM. The following tasks are set according to the purpose: 1) review neural networks to identify network attacks; 2) create a software model of a Kohonen SOM to identify network classes of U2R category; 3) determine the optimal parameters of the Kohonen SOM; 4) conduct research on network quality detection parameters.

Analysis of recent research. At the present stage, the most promising direction in the creation of attack detection systems is the use of neural networks: Multi Layer Perceptron, MLP [10-11]; Radial Basis Function Network, RBF [4]; Self Organizing Maps, SOM [2, 5-6, 8]; Adaptive-Network-Based Fuzzy Inference System, ANFIS) [4] and based on a combination of computational intelligence methods [1, 4].

On the one hand, neural networks with different topologies can detect different attacks, but erroneous triggers also do not always occur on the same network packets when analyzed using different types of neural networks. In addition, each type of neural network has its advantages and disadvantages that need to be considered or additional research. For example, the RBF learns faster than the MLP, it is necessary to determine the number of radial elements, their location and deviation values, the RBF model requires slightly more elements, namely it will run slower and requires more memory than the MLP model.

On the other hand, attempts are being made to use neural networks at different levels. For example, in [11] the structure of a hypothetical complex is considered, and consists of five neural networks (NNs) of the multilayer perceptron type. In [9] reviewed existing datasets, the most common of which is the NSL-KDD database, initiated by the US DARPA Agency based on the KDD'99 database [7]. It should be noted that today there are a number of scientific papers by various scientists and scholars on the definition of network attacks in the categories of DoS and Probe, but there is little work on the study of network classes for categories R2L and U2R. According to [6], existing intrusion detection systems based on SOM have difficulties due to the long computation time and low detection rate of U2R and R2L attacks. In [3], a research of two approaches to detecting network attacks using a single neural network and a set of neural networks based on the calculation of quality indicators for detecting attacks, among which errors of the first and second kind are important.

Setting task. The rapid development of computer networks and information technology causes a number of problems related to the security of network resources, which require effective approaches. The use of neural network technology is the most rational, because neural networks have the following advantages: solving problems with unknown patterns; resistance to input data noise; adaptation to changes in the environment; potential ultra-high speed. In this paper, it is necessary to identify network attack classes of the U2R category. U2R network attacks are system attacks in which a hacker starts a system with a normal user account and tries to abuse vulnerabilities in the system to gain superuser privileges. This type of attack is

divided into the following classes: Buffer_overflow, Loadmodule, Perl, Rootkit. The NSL-KDD database [7] presents a sufficient number of parameters for the network classes Buffer_overflow, Rootkit, Loadmodule. Because the database does not have enough parameters for the Perl class, it will not be used.

The structure of the Kohonen SOM. Kohonen neural networks – class of neural networks used to solve classification problems. It is divided into many types according to the methods of adjusting the scales. This work used a Kohonen SOM, structure of the SOM is presented in Fig. 1. The initial parameters are 41 parameters of network traffic. As the resulting data: Y1 – there was an attack of the Rootkit class; Y2 – there was an attack of the Loadmodule class; Y3 – there was an attack of the Buffer overflow class; Y4 (normal) – there was no attack.

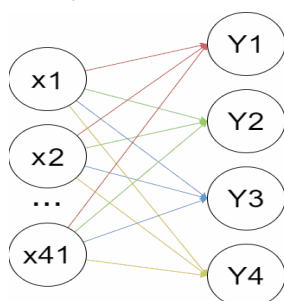


Figure 1 – The structure of the Kohonen SOM

Algorithm of functioning of the Kohonen SOM can be represented:

Step 1. Arrange the weight vectors of the node in random order on the map.

Step 2. Randomly select the input vector.

Step 3. Bypass each node on the map.

Step 4. To find the similarity between the input vector and the weight vector of the map node, you must use the euclidean distance.

Step 5: Remember the node that has the shortest distance as the Best Matching Unit (BMU).

Step 6. Update the weight vectors of the nodes near the BMU inclusive, by approaching the input vector according to the following formula:

$$W_v^{new} = W_v^{old} + \theta(u, v, s) * a(s) * (D(t) - W_v^{old}),$$

where W_v^{new} – new node weight vector; W_v^{old} – previous node weight vector; $\theta(u, v, s)$ – proximity function; $a(s)$ – learning rate; $D(t)$ – vector target input; s – current iteration; u – index of the best matching node on the map; v – node index on the map.

Step 7. Increase s and repeat until $s < \lambda$, where λ – limit of iterations.

Characteristics of the created «SOM_U2R» software model. To detect U2R attacks, the «SOM_U2R» software model was developed, which is based on the algorithm of the Kohonen SOM. The structure of «SOM_U2R» model is shown Fig. 2. 41 parameters of network traffic were fed to the input «SOM_U2R», the result of execution is a two-dimensional map with distribution on it of network classes of attacks. The software model is waiting for the input clearly 41 parameters, otherwise it will not work properly. Connect the libraries needed to implement the required functions, namely: Numpy, Matplotlib, MiniSom.

Neural network training and testing. A training sample of 55 selections is presented at the input of the neural network. The number of epochs of study was 10; the dimension of the card was 20*20. A sample containing 40 vectors (examples) was used to test the neural network. Neural network testing has been conducted for 10 epochs. The result of the «SOM_U2R» software model is presented in Fig. 3; Rootkit class (red circle); Loadmodule class (green square); Buffer_overflow class (blue cross); Normal (yellow cross).

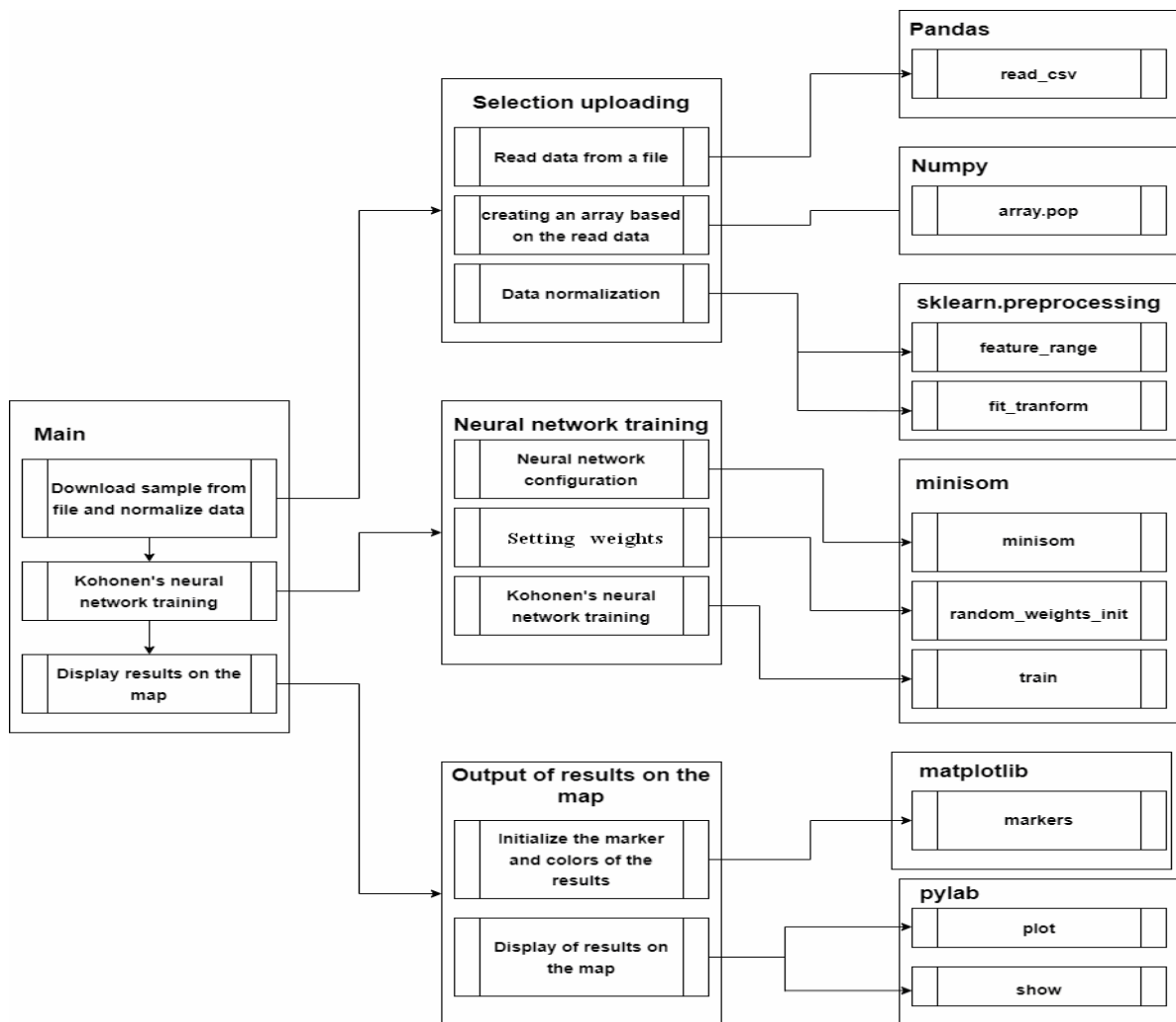


Figure 2 – The structure of the «SOM_U2R» created software model

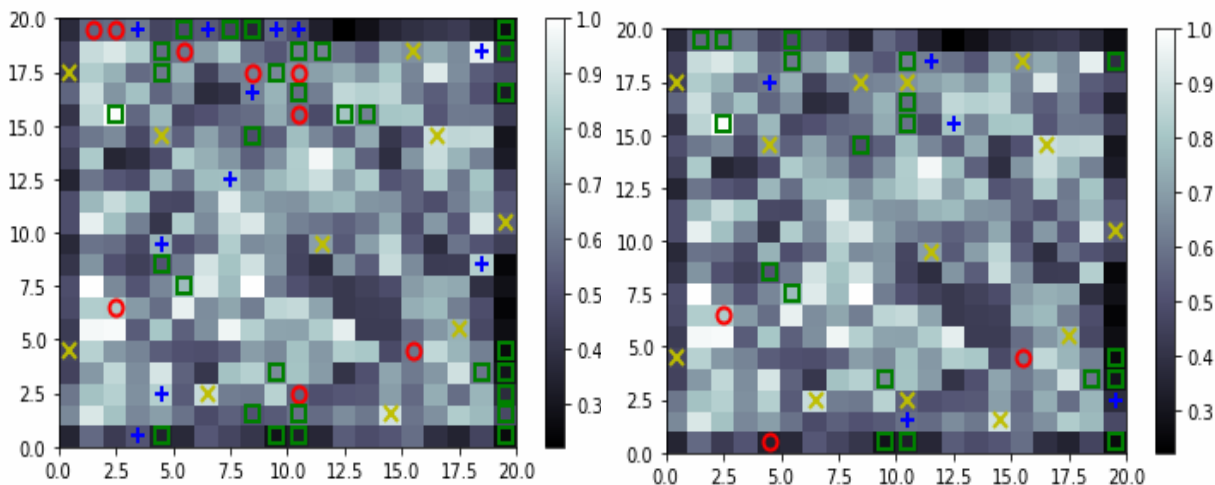


Figure 3– The result of the «SOM_U2R» model for training and testing

Determining the optimal parameters of the Kohonen SOM. Conducted on the created «SOM_U2R» model errors for different number of epochs (10, 20 and 50) for different map sizes: 5*5; 10*10; 20*20. The table shows that the smallest value of error is achieved for Kohonen SOM, the size of which is 20*20, with 10 epochs.

Research of parameters of quality of detection of network attacks. The assessment of the quality of detection of network attacks on the «SOM_U2R» model is performed according to the following parameters: TP (True Positive) – the classifier correctly assigned the object to the class under consideration; TN (True Negative) – the classifier correctly states that the object does not belong to the class under consideration; FP (False Positive) – the classifier incorrectly assigned the object to the class under consideration; FN (False Negative) – the classifier incorrectly states that the object does not belong to the class under consideration. One of the main ones is the second kind of error; the results obtained are summarized in table 1.

Table 1

The results of research of different network classes on «SOM_U2R»

Rootkit				Loadmodule				Buffer_overflow			
<u>TP</u>	<u>FP</u>	<u>TP, %</u>	<u>FP, %</u>	<u>TP</u>	<u>FP</u>	<u>TP, %</u>	<u>FP, %</u>	<u>TP</u>	<u>FP</u>	<u>TP, %</u>	<u>FP, %</u>
6	0	11	0	18	0	33	0	8	0	14	0
<u>FN</u>	<u>TN</u>	<u>FN, %</u>	<u>TN, %</u>	<u>FN</u>	<u>TN</u>	<u>FN, %</u>	<u>TN, %</u>	<u>FN</u>	<u>TN</u>	<u>FN, %</u>	<u>TN, %</u>
2	46	6	83	9	28	16	51	3	44	6	80

The table shows that the largest value of the error of the second kind of 16 % is achieved when detecting the network class Loadmodule. The obtained values of

other parameters are summarized in table 2, where TPR (True Positive Rate) – shows the proportion of found objects in the class; FPR (False Positive Rate) – shows the proportion of incorrect classifier triggers to the total number of objects outside the class; accuracy – shows the share of correct classifications; precision – shows the share of class objects among the objects selected by the classifier; recall – shows the proportion of class objects found in the total number of class objects.

Table 2

Parameters for assessing the quality of attack detection on «SOM_U2R»

<u>Indicator</u>	<u>TP</u>	<u>FP</u>	<u>FN</u>	<u>TN</u>	<u>TPR</u>	<u>FPR</u>	<u>Accu- racy</u>	<u>Preci- sion</u>	<u>Re- call</u>
<u>Buffer_overflow</u>	<u>8</u>	<u>0</u>	<u>3</u>	<u>44</u>	<u>0,73</u>	<u>0</u>	<u>0,95</u>	<u>1</u>	<u>0,73</u>
<u>Loadmodule</u>	<u>18</u>	<u>0</u>	<u>9</u>	<u>28</u>	<u>0,67</u>	<u>0</u>	<u>0,84</u>	<u>1</u>	<u>0,67</u>
<u>Rootkit</u>	<u>6</u>	<u>0</u>	<u>3</u>	<u>46</u>	<u>0,67</u>	<u>0</u>	<u>0,95</u>	<u>1</u>	<u>0,67</u>
<u>Normal</u>	<u>8</u>	<u>0</u>	<u>2</u>	<u>45</u>	<u>0,80</u>	<u>0</u>	<u>0,96</u>	<u>1</u>	<u>0,80</u>

The created «SOM_U2R» software model well detects network attacks such as Buffer_overflow and Rootkit with an accuracy of 0,95; but errors may occur when determining the Loadmodule network class (accuracy was 0,84).

Research of F-measures on different number of epochs. A research of the F-measure of detecting network classes of attacks on the «SOM_U2R» model for a different number of learning epochs (Fig. 4).

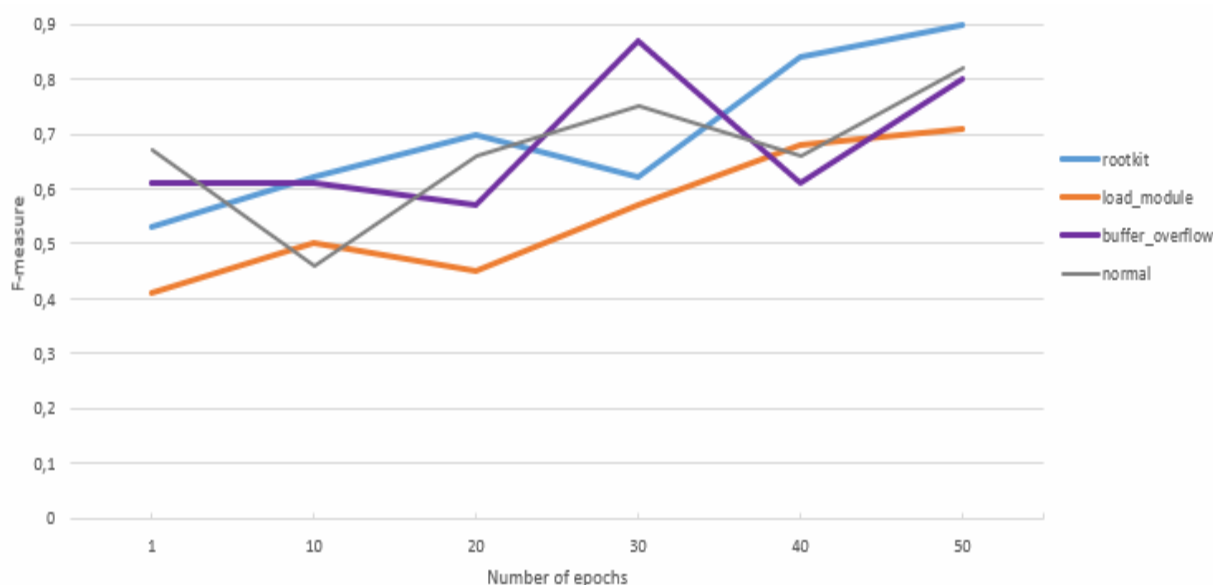


Figure 4 – The value of the F-measure by the number of epochs of learning

The F-measure is the average harmonic value between accuracy and completeness. The figure shows that for all types of attacks, except Buffer_overflow, the F-measure increases, reaching its maximum value

at around 50 epochs. For Buffer_overflow, the maximum value of the F-measure is observed in 30 epochs, but to ensure maximum efficiency in detecting network attacks, the neural network must be stopped at 50 epochs. As the Rootkit and Loadmodule network attack classes show the best results at the 50 epoch mark, thus sacrificing the growth of the Buffer_overflow network class.

Conclusions

- Based on the results of the survey, the following neural networks may be used to determine attacks: MLP; RBF; SOM; fuzzy network. To detect U2R network classes: Buffer_overflow; Load modules; Rootkit using NSL-KDD open database and further research selected Kohonen SOM.

- To identify U2R network classes, the «SOM_U2R» software model was created on Python using a Kohonen SOM, the input of which was supplied with 41 network traffic parameters. On the «SOM_U2R» software model error studies on the number of epochs (10, 20 and 50) with different map sizes: 5*5; 10*10; 20*20. It is determined that the smallest value of error is achieved on the map 20*20 at 10 epochs.

- On the «SOM_U2R» model researches of parameters of quality of detection of network classes of the U2R category are carried out. It is determined that on the «SOM_U2R» model the error of the second kind of detection of network classes of attacks Buffer_overflow and Rootkit made 6 %, and for the Loadmodule class reached 16 %.

- On the «SOM_U2R» created software model the research of F-measure on various quantity of epochs of training of the Kohonen SOM is carried out. It is determined that using the «SOM_U2R» model for all classes of attacks of the U2R category (except Buffer_overflow) the F-measure increases, reaching its maximum value at the level of 50 epochs.

ЛІТЕРАТУРА

1. Браницкий А.А. Обнаружение аномальных сетевых соединений на основе гибридизации методов вычислительного интеллекта : автореф. дис. ... канд. техн. наук. Санкт-Петербург, 2018. 18 с.
2. Пахомова В.М., Павленко І. І. Дослідження параметрів якості визначення мережевих атак категорії PROBE з використанням самоорганізуючої карти. SworldJournal. 2022. Issue 11. Part 1. pp.100-104. DOI: 10.30888/2663-5712.2022-11-01-022

3. Пахомова В.М., Коннов М.С. Дослідження двох підходів до виявлення мережних атак із використанням нейромережної технології. Наука та прогрес транспорту. 2020. № 3 (87). pp. 81-93. DOI: <https://doi.org/10.15802/stp2020/208233>
4. Amini M., Rezaeenour J., Hadavandi E. A Neural Network Ensemble Classifier for Effective Intrusion Detection using Fuzzy Clustering and Radial Basis Function Networks. International Journal on Artificial Intelligence Tools. 2016. Vol. 25. Iss. 02. pp. 1-32. DOI: <https://doi.org/10.1142/s0218213015500335>
5. Gunes K., Zincir-Heywood A., Malcolm I. H. A hierarchical SOM-based intrusion detection system. Engineering Applications of Artificial Intelligence. 2007. pp. 439-451.
6. Kruti C., Bhavin S., Ompriya K. Improving user-to-root and remote-to-local attacks usinggrowing hierarchical self organizing map. International journal of engineering sciences and research technology. 2015. том 4. № 6. URL: <http://paper.researchbib.com/view/paper/45808>
7. NSL-KDD dataset. URL: <https://www.unb.ca/cic/datasets/nsl.html>
8. Ortiz A. Improving Network Intrusion Detection with Growing Hierarchical Self-Organizing Maps. University of De La Plata Argentin. 2011. URL: <https://www.semanticscholar.org/paper/Improving-Network-Intrusion-Detection-with-Growing-Ortiz-Ortega/f3fbcf7dfd84d9f2f2ace73580c32eb7c469b6e7>
9. Ring M., Wunderlich S., Scheuring D., Landes D., Hotho A. A Survey of Network-based Intrusion Detection Data Sets. Компьютер и безопасность. 2019. DOI: 10.1016 / j.cose.2019.06.005
10. Zhukovyts'kyi I. V., Pakhomova V. M. Identifying threats in computer network based on multilayer neural network. Science and Transport Progress. 2018. № 2 (74). pp. 114-123. DOI: <https://doi.org/10.15802/stp2018/130797>
11. Zhukovyts'kyi I. V., Pakhomova V. M., Ostapets D. O., Tsyhanok O. I. Detection of attacks on a computer network based on the use of neural network complex. Science and Transport Progress. 2020. № 5(89). pp. 68-79. URL: <https://doi.org/10.15802/stp2020/218318>

REFERENCES

1. Branitskiy, A.A. (2018). Obnaruzhenie anomalnykh setevykh soedineniy na osnove gibridizatsii metodov vychislitelnogo intellekta (Extended abstract of PhD dissertation). St. Petersburg, Russia. (in Russian)
2. Pakhomova, V.M., & Pavlenko, I.I. (2022). Research of parameters of quality of definition of network attacks of the PROBE category with use of the self organizing map. SworldJournal, 11-1, 100-104. DOI: 10.30888/2663-5712.2022-11-01-022 (in Ukrainian)

3. Pakhomova, V. M., & Konnov, M. S. (2020). Research of two approaches to detect network attacks using neural network technologies. *Science and Transport Progress*, 3(87), 81-93. DOI: <https://doi.org/10.15802/stp2020/208233> (in Ukrainian)
4. Amini, M., Rezaeenour, J., & Hadavandi, E. (2016). A Neural Network Ensemble Classifier for Effective Intrusion Detection Using Fuzzy Clustering and Radial Basis Function Networks. *International Journal on Artificial Intelligence Tools*, 25(02), 1-32. DOI: <https://doi.org/10.1142/s0218213015500335> (in English)
5. Gunes, K., Zincir-Heywood, A., & Malcolm, I. H. (2007). A hierarchical SOM-based intrusion detection system. *Engineering Applications of Artificial Intelligence*, 439-451.
6. Kruti, C., Bhavin, S., & Ompriya, K. (2015). Improving user-to-root and remote-to-local attacks using growing hierarchical self organizing map. *International journal of engineering sciences and research technology*, 4(6),
URL: <http://paper.researchbib.com/view/paper/45808> (in English)
7. NSL-KDD dataset. URL: <https://www.unb.ca/cic/datasets/nsl.html> (in English)
8. Ortiz, A. (2011). Improving Network Intrusion Detection with Growing Hierarchical Self-Organizing Maps. University of De La Plata Argentina. URL: <https://www.semanticscholar.org/paper/Improving-Network-Intrusion-Detection-with-Growing-Ortiz-Ortega/f3fbcf7dfd84d9f2f2ace73580c32eb7c469b6e7> (in English)
9. Ring, M., Wunderlich, S., Scheuring, D., Landes, D., & Hotho, A. (2019). A Survey of Network-based Intrusion Detection Data Sets. *Komp'yuter and bezopasnost*. DOI: 10.1016 / j.cose.2019.06.005 (in English)
10. Zhukovyts'kyu, I. V., & Pakhomova, V. M. (2018). Identifying threats in computer network based on multilayer neural network. *Science and Transport Progress*, 2(74), 114-123. DOI: <https://doi.org/10.15802/stp2018/130797> (in English)
11. Zhukovyts'kyu, I.V., Pakhomova, V.M., Ostapets, D.O., & Tsyhanok, O. I. (2020). Detection of attacks on a computer network based on the use of neural network complex. *Science and Transport Progress*, 5(89), 68-79. URL: <https://doi.org/10.15802/stp2020/218318> (in English)

Received 13.06.2022.

Accepted 17.06.2022.

***Визначення атак категорії U2R засобами SOM
на основі бази даних NSL-KDD***

Створення ефективної системи виявлення мережевих атак вимагає застосування якісно нових підходів до обробки інформації, які повинні ґрунтуватися на адаптивних алгоритмах здатних до самонавчання. Найбільш

перспективним напрямком у створенні подібних систем виявлення атак на комп'ютерну мережу є застосування нейромережних технологій, що підтверджує актуальність теми даної роботи. У якості методу дослідження використаний математичний апарат самоорганізуючої карти Кохонена 41-2-4, де 41 - кількість вхідних нейронів (параметри мережевого трафіку); 2 - кількість шарів; 4 - кількість результуючих нейронів (Rootkit, Loadmodule, Buffer_overflow та відсутність атаки). У якості програмної реалізації самоорганізуючої карти Кохонена використана мова Python з широким спектром сучасних стандартних засобів. На створеній програмній моделі «SOM_U2R» з використанням відкритої бази даних NSL-KDD проведено дослідження помилки за різною кількістю epoch при різних розмірах карти: 5*5; 10*10; 20*20. Визначено, що найменше значення помилки досягається на карті 20*20. На створеній програмній моделі «SOM_U2R» проведено дослідження параметрів якості виявлення атак: True Positive; True Negative; False Positive; False Negative та інші. Визначено, що на програмній моделі «SOM_U2R» помилка другого роду склала 6 % для Buffer_overflow і Rootkit, 16 % для класу Loadmodule. Крім того, проведено дослідження F-мірки (середнегармонічного значення між точністю та повнотою) за різною кількістю epoch навчання самоорганізуючої карти Кохонена. Визначено, що для всіх атак (крім Buffer_overflow) F-мірка зростає, досягаючи свого максимального значення (50 epoch).

Пахомова Вікторія Миколаївна – к.т.н., доц. кафедри електронних обчислювальних машин Українського державного університету науки та технологій (Дніпро); ORCID 0000-0002-0022-099X

Pakhomova Victoria – PhD, Assoc. Department of Electronic Computers of the Ukrainian State University of Science and Technology (Dnipro); ORCID 0000-0002-0022-099X

Мегельбей Егор Олександрович – бакалавр спеціальності «Кібербезпека» кафедри електронних обчислювальних машин Українського державного університету науки та технологій (Дніпро)

Mehelbei Yehor – Bachelor of Cybersecurity in Department of Electronic Computers of the Ukrainian State University of Science and Technology (Dnipro)

TWO-STAGE METHOD FOR PARAMETER ESTIMATING IN UNKNOWN NOISE ENVIRONMENT

Abstract. This paper addresses the parameter estimation problem in the case of an unknown noise environment. A two-stage method consisting of the tuning and estimating stages is proposed for solving this problem. At the tuning stage, the estimator is tuned to the noise environment by minimizing the estimation root-mean-square error for a known data fragment in the space of the three free parameters; these free parameters allow modifying the behavior of the minimization problem objective function. At the estimating stage, estimation is carried out by solving the corresponding minimization problem using already tuned free parameters. The features of the proposed method, including algorithms for the tuning and estimating stages, are demonstrated for estimating a Gaussian pulse that slowly moves in the unknown noise environment and locates on a known constant background. The numerical simulations confirm the high performance of the proposed method.

Keywords: parameter estimation, cost function, Gaussian pulse.

Introduction. Classical estimation methods use an assumption about the completeness of data statistics. These methods solve the estimation problem in a noise environment formed by additive noise with a known distribution. A theoretical basis of these methods is the maximum-likelihood criterion, where the likelihood function is a common probability density function to be maximized [1] – [2]. The well-known results of this theory include the least-squares method and the least-absolute deviation method. These methods obey the assumption that data samples are independent and identically Gaussian or Laplacian distributed, respectively. The maximum-likelihood criterion has remarkable achievements related to the myriad and meridian filterings, which are generalized in [3]. These methods assume that the data samples are independent and identically distributed according to the Cauchy and Meridian statistics. They produce the so-called myriad (or Cauchy [4]) estimator and meridian estimator, respectively. Because of the impulsive nature of the Cauchy and Meridian noise, these estimators belong to the set of robust estimators.

Robust estimation methods use the incompleteness of data statistics [5] – [9]. They solve the estimation problem in the noise environment formed by additive noise and anomalous values. A theoretical basis of these methods is the generalized maximum-likelihood criterion that does not have a strictly probabilistic sense [7]. For that reason, the corresponding estimators usually are referred to as the “maximum-likelihood type” estimators or the M-estimators [6]. In contrast to the classical approach, the robust one requires choosing a cost (loss) function that should keep the data (inliers) and reject the anomalous values (outliers). The appropriate cost functions used for these purposes have been proposed by Tukey [6], Hampel [7], Andrews [8], Meshalkin [9], Geman and McClure [10], Demidenko [11], as well as by other authors [12] – [13]. Assembling various cost functions into a common set, which we call the “superset of cost functions,” eliminates the redundancy of well-known cost functions and provides a wide range of possible solutions, including the mean, median, myriad, and meridian [14]. The modified version of this superset is given in [15]. An advantage of using the superset of cost functions is the possibility of tuning the M-estimator to the current noise environment.

The purpose of this paper is to present the two-stage method for parameter estimating in the unknown noise environment under the general assumption about the symmetry of noise distribution. This method uses the generalized maximum-likelihood criterion and the superset of cost functions. Through this paper, we show the features and performance of the proposed method by solving the problem of estimating a slowly moving Gaussian pulse located on the known constant background, where a mixture of Cauchy noise and positive outliers simulates the unknown noise environment.

Problem formulation. Let for a given observation interval X the observed data $g(x)$ are the sum of a given data model $s(x; \theta)$, which specified up to a vector $\theta = (\theta_1, \dots, \theta_M)$ of the unknown parameters $\theta_1, \dots, \theta_M$, and a random distortion $\xi(x)$, which describes the unknown noise environment, i.e.,

$$g(x) = s(x; \theta) + \xi(x), \quad (1)$$

where $x \in X$, and M is given. The estimation problem is to obtain the estimates of $\theta_1, \dots, \theta_M$. Using (1), this problem can be formulated as the optimization problem:

$$\hat{\theta} = \arg \min_{\theta} F[g(x) - s(x; \theta)]; \quad x \in X, \quad (2)$$

where F denotes an error functional that defines the estimation quality, and $\hat{\theta} = (\hat{\theta}_1, \dots, \hat{\theta}_M)$ is a vector of the estimates $\hat{\theta}_1, \dots, \hat{\theta}_M$ for $\theta_1, \dots, \theta_M$. For the discrete case, the problem (2) has the form:

$$\hat{\theta} = \arg \min_{\theta} \sum_{n=1}^N \rho[g_n - s_n(\theta)]; \quad x \in X, \quad (3)$$

where $g_n = g(x_n)$, $s_n(\theta) = s(x_n; \theta)$, $x_n = n \cdot \Delta x$, Δx is a sampling step size, and N is a number of data samples on the observation interval X . From (3), it follows that the value of F is the sum of the values of cost function $\rho(\dots)$ for each n -th data sample, and these values depend on θ . Since the cost function forms an objective function to be minimized, the problem of parameter estimation in the unknown noise environment consists of the problem of choosing the best cost function for the noise environment and the problem of parameter estimation after the choice of the best cost function.

Analysis of recent research and publications. A classical approach to the parameter estimation problem uses two main assumptions. Firstly, it assumes that the statistical nature of random distortions is known completely. Secondly, it assumes that each parameter to be estimated takes a single value. Within the framework of estimation theory, these assumptions dictate the search for a global minimum (or a global maximum) of an estimation quality indicator. Usually, the unconditional average risk is taken as this indicator. For a given loss (cost) function, the minimum of this risk corresponds to a Bayesian estimate. Using the “0-1” loss function and assuming that the probability density of the unknown parameter is a constant, its Bayesian estimate transfers to the maximum likelihood estimate [16]. The latter is widely used to obtain optimal or suboptimal solutions. For example, under the assumption of Gaussian data distribution, the corresponding maximum likelihood criterion becomes a quadratic one, leading to the least-squares problem [17]. An advantage of this problem is that, for the case of the data model described by a constant function, the least-squares problem leads to the analytical solution, yielding the arithmetic mean of data samples as the estimate. The assumption about the Laplace distribution of data samples leads to the least-absolute deviations problem, which gives a median value of data samples as the estimate for the same data model described by the constant function [18]. However, the quadratic or other convex functions are unsatisfactory for processing the data containing anomalous values [19]. In this case, robust approach based on the generalized maximum likelihood criterion is used. For the vector θ of unknown parameters of data model, it has the form of the following minimization problem:

$$\min_{\theta} \sum_{n=1}^N \rho[r_n; \theta], \quad (4)$$

where r_n is a residual between the observed data and data model for a given n , and ρ is the cost function [6]-[7]. Usually, the robust approach (4) is implemented by the cost function with horizontal asymptotes [9].

To obtain a wide range of cost functions, a “superset of cost functions” has been proposed [14]. This “superset” is defined by the cost function:

$$\rho_s(x) = k_s [(1 + |x/\alpha|^q)^{\beta/q} - 1], \quad (5)$$

where $\alpha > 0$ is a smoothing parameter with the same unit as x ; q is a smoothing degree parameter ($0 < q < \infty$); β is a shape parameter ($-\infty < \beta \leq 1$; $\beta < q$); $k_s(x) = 1/[(1 + |x_0/\alpha|^q)^{\beta/q} - 1]$ is a constant, which is used to transform the cost function within the “superset” framework; x_0 is a normalization point, where $\rho_s(x_0) = 1$. Let $x_0 = 1$. The following limits are true:

$$\begin{aligned} \lim_{\alpha \rightarrow \infty} \rho_s(x) &= |x|^q; & \lim_{\alpha \rightarrow 0} \rho_s(x) &= \begin{cases} |x|^\beta; & 0 < \beta \leq 1 \\ \varphi(x); & -\infty < \beta \leq 0 \end{cases}; \\ \lim_{\beta \rightarrow \pm 0} \rho_s(x) &= k_1 \cdot \ln(1 + |x|^q / \alpha^q); & \lim_{\beta \rightarrow -\infty} \rho_s(x) &= \varphi(x); \end{aligned} \quad (6)$$

$$\lim_{\beta \rightarrow 1} \rho_s(x) = k_2 \cdot [(1 + |x|^q / \alpha^q)^{1/q} - 1]; \quad \lim_{\substack{\beta \rightarrow -2 \\ q \rightarrow 2}} \rho_s(x) = k_3 \cdot \frac{x^2}{x^2 + \alpha^2},$$

where $\varphi(x) = \text{sgn}^2(x)$; $k_1 = 1/\ln(1 + 1/\alpha^q)$; $k_2 = 1/[(1 + 1/\alpha^q)^{1/q} - 1]$; $k_3 = (1 + \alpha^2)$.

The modified version of the “superset” is given in [15]. It has been obtained by equalizing behavior of cost function at zero. This modification expands the “superset” by a generalized Meshalkin’s cost function [11]:

$$\rho_M(x) = [1 - \exp(-|x/\alpha|^q)]/[1 - \exp(-|x_0/\alpha|^q)], \quad (10)$$

where the value of α must be recalculated concerning the value of α for the original “superset.” For the modified superset, it holds: $\lim_{\beta \rightarrow -\infty} \rho_{Sm}(x) = \rho_M(x)$, where $\rho_{Sm}(x)$ denotes the cost function of the modified superset. From (10), it follows that $\lim_{\alpha \rightarrow 0} \rho_M(x) = \varphi(x)$, $\lim_{\alpha \rightarrow \infty} \rho_M(x) = |x|^q$, and $\lim_{q \rightarrow 0} \rho_M(x) = \varphi(x)$, where $x_0 = 1$. Besides this, if $0 < \alpha < x_0$, then $\lim_{q \rightarrow \infty} \rho_M(x) = \rho_{pit}(x)$, where $\rho_{pit}(x)$ has the shape of a “rectangular pit” of 2α wide, i.e.,

$$\lim_{q \rightarrow \infty} \rho_M(x) \Big|_{|x| < \alpha_M} = 0; \quad \lim_{q \rightarrow \infty} \rho_M(x) \Big|_{|x| = \alpha_M} = 1 - 1/e \approx 0,63; \quad \lim_{q \rightarrow \infty} \rho_M(x) \Big|_{|x| > \alpha_M} = 1, \quad (11)$$

where e is Euler’s number. The use of (11) changes the maximum generalized likelihood method into the maximum histogram method, where the histogram bin width

is equal to 2α . Since the “superset” covers many symmetric cost functions, it can be used to tune the estimation to the noise environment with the symmetric noise distribution and with the outliers to be rejected.

Theory of the two-stage estimation method. The proposed two-stage estimation method uses the generalized maximum likelihood criterion and the superset of cost functions. There are two ways to apply this method. For the first one, it needs to set the three free parameters α , β , and q to a priori known values. For the second one, it needs to tune these three free parameters to the current noise environment. The latter we discuss for a Gaussian pulse that slowly moves in the unknown noise environment and locates on the known constant background. Here, we assume that the observed data contain a background fragment with unknown location in the measuring space, and a sufficient number of data realizations is available. Then, by estimating the background at the tuning stage, we choose the best values of free parameters, thereby realizing some training procedure. At the estimating stage, we estimate the location, amplitude, and half-width of the Gaussian pulse by the iterative algorithm proposed.

Let the data model is:

$$s(x; A_0, A, m, \sigma_p) = A_0 + A \cdot \exp[-(x - m)^2 / (2\sigma_p^2)], \quad (12)$$

where A_0 is a known constant background, and A, m, σ_p are the Gaussian pulse parameters to be estimated. Let m is slowly changed in time comparing to the data sampling rate, and the length of the observation interval X is larger than the half-width σ_p of the Gaussian pulse in several times. Also suppose that $\sigma_p \gg \Delta x$, $m \in X$, and A is within a known range. In such a case, we have a Gaussian pulse, which has a new location for each data measuring; we call this pulse a “moving Gaussian pulse.” By using (12) for the discrete case, we state the following two estimation problems. The first one is:

$$\hat{A}_0 = \arg \min_{A_0} \sum_{n=1}^N \rho_s \{g_n - A_0 - A \exp[-(x_n - m)^2 / (2\sigma_p^2)]\}, \quad (13)$$

where ρ_s denotes the superset cost function, which depends on the three free parameters α, β, q [14]. Assuming $A = 0$ and excluding the Gaussian pulse model from (13), we have the first problem in the following form:

$$\hat{A}_0 = \arg \min_{A_0} \sum_{n=1}^N \rho_s [g_n - A_0]. \quad (14)$$

The problem (14) describes the case of a “pure” noise environment when observed data do not contain a Gaussian pulse to be estimated. However, below we ap-

ply (14) both in this case and in the case of the presence of the Gaussian pulse. Solving (14) gives the estimate \hat{A}_0 of the background parameter A_0 for the given values of free parameters α , β , and q . Since A_0 has a known value, the best values of free parameters should minimize the estimation error. Note that subtracting the known constant background from observed data gives a similar problem for zero background estimating.

The second problem is the following:

$$(\hat{A}, \hat{m}, \hat{\sigma}_p) = \arg \min_{A, m, \sigma_p} \sum_{n=1}^N \rho_S \{g_n - A_0 - A \exp[-(x_n - m)^2 / (2\sigma_p^2)]\}, \quad (15)$$

This is the Gaussian pulse estimation problem under the assumption that values are known for A_0 , α , β and q .

Let us consider the tuning stage. Substituting (5) into (14) yields:

$$\hat{A}_0 = \arg \min_{A_0} \left\{ k_S \cdot \sum_{n=1}^N [(1 + |g_n - A_0|^q / \alpha^q)^{\beta/q} - 1] \right\}, \quad (16)$$

where \hat{A}_0 is the estimate of A_0 . For most cases, the objective function in (16) is a non-convex. Therefore, its minimization needs using the zero-order optimization methods. Naturally assuming $\hat{A}_0 \in [g_{\min}, g_{\max}]$, where $g_{\min} = \min_{1 \leq n \leq N} (g_n)$ and $g_{\max} = \max_{1 \leq n \leq N} (g_n)$, we offer to solve the minimization problem (16) by searching for the best value of A_0 either among data samples or given values. The former is more attractive for the following reason. Because the value of each data sample turns into zero at least one term of the objective function in (16), such a value can be considered a “quasi-optimal” one. There are no more than N quasi-optimal values for the problem (16). Besides this, if $q < 1$, the quasi-optimal value coincides with the optimal value that minimizes the objective function in (16) (although, for the given values of α, β, q this optimal value of A_0 may not coincide with the true value of A_0).

The tuning stage is implemented by training performed over a given number of random data realizations and aims to obtain the best values for α, β, q . For training, we use the Mean-Square-Error (MSE) criterion. Let J be a number of data realizations used for training. The MSE, which we denote by the ε symbol, is the three-dimensional function for a fixed value of J , i.e.,

$$\varepsilon(\alpha, \beta, q) = \left[\frac{1}{J} \sum_{j=1}^J |\hat{A}_0(j; \alpha, \beta, q) - A_0|^2 \right]^{1/2}, \quad (20)$$

where $\hat{A}_0(j; \alpha, \beta, q)$ is the estimate of A_0 for the j -th realization and given α, β, q .

The MSE minimum points to the best values $\hat{\alpha}, \hat{\beta}, \hat{q}$, i.e.,

$$(\hat{\alpha}, \hat{\beta}, \hat{q}) = \arg \min_{\alpha, \beta, q} \varepsilon(\alpha, \beta, q), \quad (21)$$

For tuning, it needs to choose the sequences of free parameters values. Assuming that the scale of inlier distortions is less than 1, we propose the following ones. For α , the sequence is $(10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 1, 10)$. For β , the sequence is $(-\infty, -16, -4, -2, -1, -1/2, -1/4, -1/8, 0, 1/8, 1/4, 1/2, 1)$. For q , the sequence is $(1/2, 1, 3/2, 2, 5/2, 3, 4, 5, 10)$.

Let us consider the estimating stage. Substituting (5) into (15) yields:

$$(\hat{A}, \hat{m}, \hat{\sigma}_p) = \arg \min_{A, m, \sigma_p} k_s \sum_{n=1}^N (\{1 + |g_n - A_0 - A \exp[-(x_n - m)^2 / (2\sigma_p^2)]\}^q / \alpha^q)^{\beta/q} - 1), \quad (22)$$

where α, β, q are already tuned parameters and A_0 is known. To solve (22), we propose to use a zero-order minimization method that consists in generating the initial estimates and their refining.

Generating the initial estimates is to obtain the approximate values of A, m, σ_p . To generate the initial estimates, we assume that each of the data samples $g_n; n = 1, \dots, N$ can specify the location and the amplitude of the Gaussian pulse. In that case, the problem reduces to estimate the half-width σ_p . To solve this problem, we construct a set of trial values for σ_p and select such a trial value that minimizes the objective function presented in (22). We include in this set the following values. The first value is $\sigma_p = \infty$, indicating the transformation of the Gaussian pulse into a constant background. Although it contradicts the assumption about the presence of the Gaussian pulse, it can be used to calculate a maximum value of the objective function for the given data sample. Other trial values are generated by the formula:

$$\hat{\sigma}_p = \frac{|j - i| \cdot \Delta x}{\sqrt{2 \ln[(g_i - A_0)/(g_j - A_0)]}} \quad (23)$$

under the condition that $(g_i - A_0)/(g_j - A_0) > 1$, where i is the index of the data sample g_i considered as a center of the Gaussian pulse, and j is the index of some other data sample g_j , for which the inequality mentioned above holds. If, for the given g_i , it is impossible to find $g_j; j \neq i$ such that $(g_i - A_0)/(g_j - A_0) > 1$, then g_i is excluded from contenders to be a center of the Gaussian pulse, and the trial value is not generated for σ_p . However, if for the given data sample g_i there is $g_j; j \neq i$ such

that the inequality: $(g_i - A_0)/(g_j - A_0) > 1$ holds, the corresponding trial value is generated for σ_p . After obtaining the set of trial values, we select among them such a value that minimizes the objective function in (22). Thus, for each data sample, we obtain the initial estimates of A, m, σ_p and the corresponding value of the objective function. Then, passing through all the data samples, we select the best initial estimates that give a minimum value of the objective function.

Refining of the estimates consists in iteratively refining the estimates of A, m, σ_p . To implement refining, we use the trial values calculated as:

$$\hat{m} = i \cdot \Delta x - \hat{\sigma}_p \sqrt{2 \ln[\hat{A}/(g_i - A_0)]}; \text{ if } \hat{A}/(g_i - A_0) > 1, \quad (24)$$

$$\hat{A} = (g_i - A_0) \cdot \exp[(i \cdot \Delta x - \hat{m})^2 / (2\hat{\sigma}_p^2)], \quad (25)$$

$$\hat{\sigma}_p = \frac{|i \cdot \Delta x - \hat{m}|}{\sqrt{2 \ln[\hat{A}/(g_i - A_0)]}}; \text{ if } \hat{A}/(g_i - A_0) > 1, \quad (26)$$

where the current estimate is at the left side of the equal sign, and the previous estimates are at the right side of the equal sign. We refine the Gaussian pulse parameters estimates by the following algorithm. (1) Store the parameter estimate and the objective function value. (2) Create an array of trial values of the estimated parameter. (3) Check the performance of each trial value by calculating the objective function value and comparing it with the stored value of the objective function. If some trial value of the parameter minimizes the objective function, it is the estimate of this parameter at the current iteration. We use this algorithm iteratively through the following sequence of Gaussian pulse parameters: the location, the amplitude, and the half-width, and finish it when the estimates achieve the given accuracy.

Simulations. For simulations, we used (16) - (26), as well as the models of positive outliers and Cauchy noise. We set $N=256$, $A_0=1$, $A=1$, $m=l \cdot \Delta x$, $\sigma_p=8 \cdot \Delta x$, $\Delta x=1$, where l has an integer value from 60 to 160. We used the Cauchy noise distribution $p(\xi) = (v/\pi) \{1/[(\xi - \vartheta)^2 + v^2]\}$ with $\vartheta=0$ and $v=0.1$. For positive outliers, we set the probability of their occurrence to 0.1, and the uniform distribution of their amplitudes in the range from 0 to 5. For each new data realization containing a Gaussian pulse, changing l created the effect of Gaussian pulse moving. Therefore, averaging the data realizations will not be effective for estimating. Moreover, averaging the Cauchy noise is ineffective.

At the tuning stage, we simulated 1000 random realizations of the noise environment, formed by the Cauchy noise and positive outliers in both cases of the absence and the presence of the Gaussian pulse. For the latter case, we used different

values of l obtained first by its decrementing by 1 from 160 to 60, and then by its incrementing by 1 from 60 to 160, etc. By estimating the constant background for various values of α , β , and q , we calculated the three-dimensional function (20) for each of the indicated cases. Fig. 1 presents these simulations, where Fig. 1a, Fig. 1b, and Fig. 1c show cross-sections of the decimal logarithm of the three-dimensional function (20) by the planes $\beta = \hat{\beta}$ and $q = \hat{q}$, by the planes $\alpha = \hat{\alpha}$ and $q = \hat{q}$, as well as by the planes $\alpha = \hat{\alpha}$ and $\beta = \hat{\beta}$, respectively. In these figures, the symbol $@$ denotes the ordinal number of the element in the given sequences for α , β , q , which starts from one and changes by one. Thus, if $@\alpha = 1$, then $\alpha = 10^{-6}$; if $@\alpha = 2$, then $\alpha = 10^{-5}$, etc. In a like manner, if $@\beta = 1$, then $\beta = -\infty$, etc., as well as if $@q = 1$, then $q = 0.5$, etc. Fig. 1d shows the cost functions for values $\hat{\alpha} = 0.1$, $\hat{\beta} = -4$, $\hat{q} = 2$ and $\hat{\alpha} = 0.1$, $\hat{\beta} = -\infty$, $\hat{q} = 1.5$, which minimized the MSE for the indicated cases, when $J = 1000$. Here, we see that these cost functions visually almost coincide. The obtained dependences of $\hat{\alpha}, \hat{\beta}, \hat{q}$ and MSE minimum on the number J of random realizations show that tuning of α was the fastest, while tunings of β and q were slower in both cases.

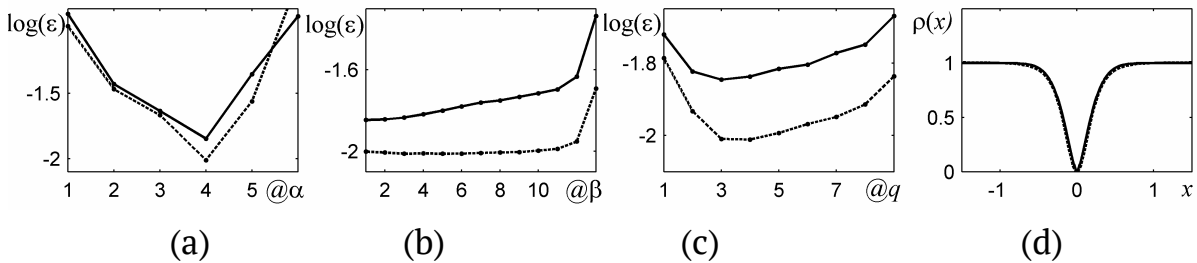


Figure 1 – Tuning to the noise environment formed by the Cauchy noise and positive outliers: the log of MSE along α -axis (a), β -axis (b), q -axis (c), and the best cost functions (d); dotted and solid lines correspond to the absence and the presence of Gaussian pulse, respectively

The second stage of the proposed method is the estimating stage. Fig. 2 shows two examples of this stage for estimating the Gaussian pulse. Here, Fig. 2a and Fig. 2e show two different data realizations. Fig. 2b and Fig. 2f show the estimates obtained for "best" tuning (when $\alpha = \nu = 0.1$, $\beta = -4$, $q = 2$), where plots of the true Gaussian pulse and its estimates visually coincide. Fig. 2c and Fig. 2g present the estimates obtained for "good" tuning (when $\alpha = \nu = 0.1$, $\beta = -\infty$, $q = 1.5$). Fig. 2d and

Fig. 2h show the estimates obtained for "bad" tuning (when $\alpha = \nu/100 = 0.001$, $\beta = -\infty$, $q = 1.5$). A full demonstration of this stage is available by the reference <https://drive.google.com/file/d/1RVz9AKShqhWAUDpzGPffJC5INeCMtiY1>.

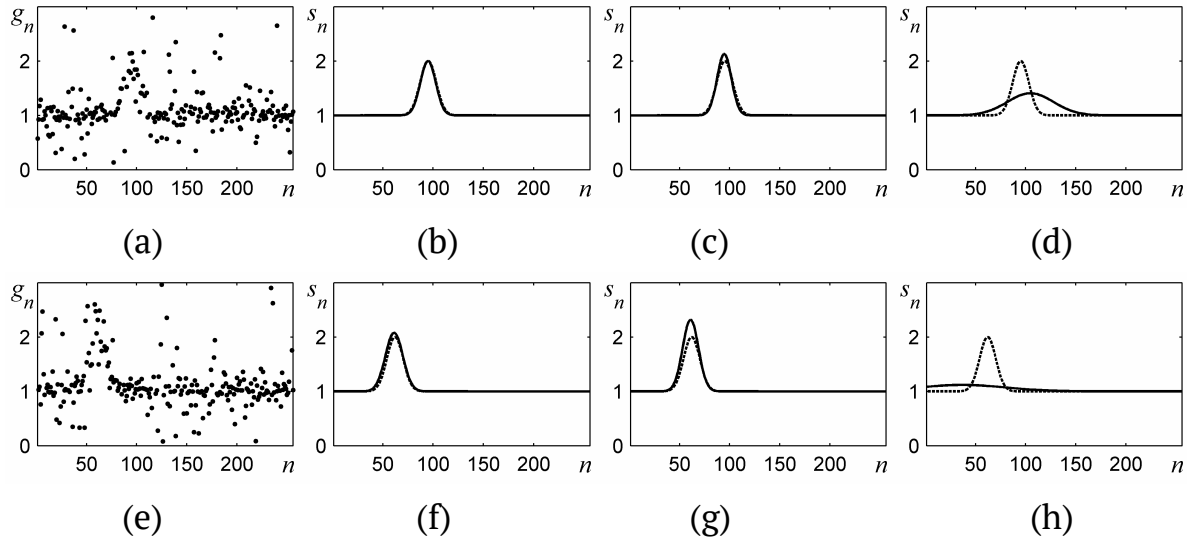


Figure 2 – Gaussian pulse estimation with known constant background: input data (a,e); the estimate (solid line) for “best” tuning (b, f), "good" tuning (c, g), and "bad" tuning (d, h); dotted line shows the true Gaussian pulse

Table 1 presents the average and maximal MSE values of the Gaussian pulse estimation obtained over 1000 realizations by using the mean, median, myriad, and meridian estimators, as well as by using the proposed estimator with the best, good, and two bad tunings, respectively. The plots in the table cells show the cost functions used for estimation. A feature of these cost functions is their normalization at $x_0 = 1$, i.e. $\rho_s(x_0) = 1$. Here, we see that the shape of the cost function and the scale of its argument are important at once.

Discussion. For considered example, the Gaussian pulse estimation should be performed for each data realization obtained after measuring. Unfortunately, the presence of the Gaussian pulse complicates tuning to the noise environment since the Gaussian pulse also contaminates it. However, Fig. 1 shows if the ratio of the observation interval length to the half-width of the Gaussian pulse is large enough, then the tuning gives the same results both in the absence and in the presence of the Gaussian pulse. Simulations also show that the tuning time can be reduced essentially by using the following algorithm. (1) Solve the three-dimensional tuning problem for α , β , q with a small number (e.g., 100) of realizations and take the best value of α . (2) Fixing this value of α , solve the two-dimensional tuning problem for

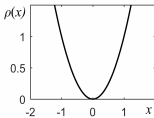
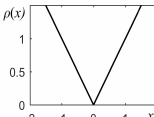
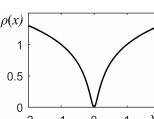
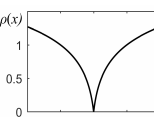
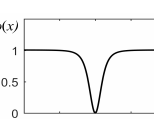
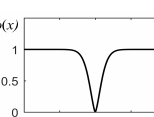
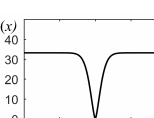
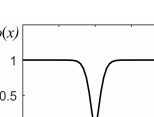
q and β with a moderate number (e.g., 200-500) of realizations and take the best value of q . (3) Fixing this value of q , solve the one-dimensional tuning problem for β with a large number (e.g., 500-1000) of realizations and take the best value of β .

Fig. 2 shows that the quality of estimation depends on the tuning quality. In this context, Table 1 shows that the best estimate corresponds to the "best" tuning (both in terms of the average value and maximal value of MSE). But for the "good" tuning, the estimate is slightly worse. Table 1 also shows that a small average value does not guarantee a small maximal value. In total, Fig. 2 and Table 1 confirm the effectiveness of the proposed approach for Gaussian pulse estimation for the case when the free parameters are properly tuned to the noise environment.

Problem (2) was formulated for a parametric data model without using the constraints on the values of data model parameters. Using such information and including it in the formulation of the estimation problem, we can improve the estimation quality. Besides this, if the unknown parameter takes several values, it is advisable to look for several local minima of the objective function.

Conclusion. The proposed two-stage method is an efficient method for parameter estimating of a given data model in an unknown noise environment. It can be used under the condition that it is possible to tune the parameter estimation procedure to the noise environment. It is shown that tuning can be performed on a fragment of known data. Further, subtracting the estimated data from the obtained data after the estimating stage allows refining tuning. The proposed method can be used to develop an intelligent system that can tune itself to the unknown noise environment and perform the best estimation of the parameters of a given data model.

Table 1

MSE of Gaussian pulse estimation			
Estimator	Cost function	Average value of MSE	Maximal value of MSE
Mean		59.76	1502.44
Median		0.36	2.24
Myriad ($\alpha=v, \beta=0, q=2$)		0.50	3.00
Meridian ($\alpha=v, \beta=0, q=1$)		0.40	1.52
proposed: “best” tuning ($\alpha=v, \beta=-4, q=2$)		0.34	1.47
proposed: “good” tuning ($\alpha=v, \beta=-\infty, q=1.5$)		0.37	2.39
proposed: “bad 1” tuning ($\alpha=v*100, \beta=-\infty, q=1.5$)		1.19	22.08
proposed: “bad 2” tuning ($\alpha=v/100, \beta=-\infty, q=1.5$)		3.39	20.17

REFERENCES

1. Smith B.W. Parameter estimation in the presence of measurement noise. *International Journal of Control*. 1966. Vol. 3, N. 4. P. 297-312.
2. Tong X., Gao F., Chen K., Cai D., Sun J. Maximum likelihood estimation in transformed linear regression with nonnormal errors. *The Annals of Statistics*. 2019. Vol. 47, N. 4. P. 1864-1892.
3. Carrillo R.E., Aysal T.C., Barner K. E. Generalized Cauchy distribution based robust estimation. 2008 IEEE International Conference on Acoustics, Speech and Signal Processing, Las Vegas, NV, 31 March – 4 April 2008. P. 3389–3392.
4. Balmer D.W., Boulton M., Sack R. A. Optimal solutions in parameter estimation problems for the Cauchy distribution. *Journal of the American Statistical Association*. 1974. Vol. 69, N. 345. P. 238-242.
5. Rousseeuw P.J., Hubert M. Anomaly detection by robust statistics. *WIREs Data Mining Knowledge Discovery*. 2018. Vol. 8, N. 2. P. 1-14.
6. Huber P. *Robust statistics*. New York: Wiley. 1981. 312 p.
7. Hampel F.R., Ronchetti E.M., Rousseeuw P. J., Stahel W. A. *Robust statistics: the approach based on influence functions*. Hoboken: Wiley, 2011. 502 p.
8. Andrews D. F. Bickel P. J., Hampel F. R., Huber P. J., Rogers W. H., Tukey J. W. *Robust estimates of location. Survey and advances*. Princeton : Princeton University Press, 1972. 372 p.
9. Shevlyakov G. L., Vil'chevski N. O. *Robustness in data analysis: criteria and methods*. Utrecht : VSP, 2002. 310 p.
10. Geman S., McClure D. Bayesian image analysis. An application to single photon emission tomography. *Proceedings of the American Statistical Association, Statistical Computing Section*. 1985. P. 12-18.
11. Vovk S., Borulko V. Family of generalized Demidenko functionals for robust estimation. *Proceedings of the 17th Int. Seminar/Workshop on Direct and Inverse Problems of Electromagnetic and Acoustic Wave Theory, Tbilisi, Georgia, September 2012*. P. 151–154.
12. Guo W., Lou Y., Qin J. et al. A Novel Regularization Based on the Error Function for Sparse Recovery. *Journal of Scientific Computing*. 2021. Vol. 87. 31.
13. Zhao Z., Wang S., Wong D., Sun C., Yan R., Chen X. Robust enhanced trend filtering with unknown noise. *Signal Processing*. 2021. Vol. 180, 107889.
14. Borulko V. F., Vovk S. M. Minimum-duration filtering. *Radio Electronics, Computer Science, Control*. 2016. N. 1. P. 7–14.

15. Vovk S.M. General approach to building the methods of filtering based on the minimum duration principle. Radioelectronics and Communications Systems. 2016. Vol. 59, N. 7. P. 281–292.
16. Millar R.B. Maximum Likelihood Estimation and Inference: With Examples in R, SAS and ADMB. New York : Wiley, 2011. 376 p.
17. Wolberg J. Data Analysis Using the Method of Least Squares: Extracting the Most Information from Experiments. Berlin : Springer-Verlag, 2005. 250 p.
18. Elgmati E.A., Gredni N.B. Quartile Method Estimation of Two-Parameter Exponential Distribution Data with Outliers. International Journal of Statistics and Probability. 2016. Vol. 5, N. 5. P. 12-15.
19. Chandola V. Banerjee A., Kumar V. Anomaly detection: A survey. ACM Computing Surveys. 2009. Vol. 41, N. 3. P. 15-58.

Received 20.06.2022.

Accepted 24.06.2022.

Двоетапний метод оцінювання параметрів у невідомому шумовому середовищі

Запропоновано двоетапний метод вирішення задачі оцінювання параметрів заданої моделі даних для випадку невідомого шумового середовища, який складається з етапів налаштування та оцінювання. На етапі налаштування оцінювач налаштовується на шумове середовище шляхом мінімізації середньоквадратичної помилки оцінювання для відомого фрагмента даних у просторі трьох вільних параметрів. На етапі оцінювання оцінювач розв'язує відповідну задачу мінімізації, використовуючи вже налаштовані вільні параметри. Наведено приклад оцінювання гауссівського імпульсу, який повільно рухається у невідомому шумовому середовищі.

Вовк Сергій Михайлович - к.ф.-м.н., доцент, Дніпровський національний університет, доцент кафедри комп'ютерних наук та інформаційних технологій.

Vovk Serhii Mikhailovich - Ph.D, Associate Professor, Oles Honchar Dnipro National University, Associate Professor of Department of Computer Science and Information Technology.

В.У. Ігнаткін, Ю.С. Хомош, В.С. Дудніков, О.П. Юшкевич, В.А. Тіхонов

**ОЦІНКА ЕФЕКТИВНОСТІ СИСТЕМИ АВТОМАТИЗОВАНОГО
УПРАВЛІННЯ МЕТРОЛОГІЧНИМ ОБСЛУГОВУВАННЯМ ЗАСОБІВ
ВИМІРЮВАЛЬНОЇ ТЕХНІКИ – ЯК ІНФОРМАЦІЙНО-ДОВІДКОВОЇ СИСТЕМИ
КОЛЕКТИВНОГО КОРИСТУВАННЯ**

Анотація. В якості оцінки ефективності діалогової системи колективного користування, на прикладі системи автоматизованого управління метрологічним обслуговуванням засобів виміральної техніки (САУМОЗВТ), використовується відношення середньої реактивності системи до середніх витрат на один запит. Досліджено відношення середньої реактивності цільової системи до середніх витрат на один запит (С). Під реактивністю прийнята величина, зворотна часу проходження запиту у системі. Потік запитів у системі підпорядковується закону розподілення Пуассона. Для дослідження вибраний метод імітаційного моделювання. Одержані: залежності (С) від часу поміж двома сусідніми запитами; залежність витрат на один запит від інтенсивності надходження запитів; залежність витрат на один запит від пропускної спроможності системи та ін. Збільшення ефективності діалогових систем колективного користування можливим шляхом більш повного навантаження більшої потужності інформаційної системи. Вибраний критерій оцінки показує, який запас в пропускній здатності системи є надмірним. Наведено результати розрахунків економічної ефективності від запровадження САУМОЗВТ на кількох підприємствах та НДІ. У загальному випадку методи визначення економічної ефективності САУ різних призначень розрізняються між собою складом елементів невпорядкованості (витрат) і ступенем впливу використаної інформації на стан керованої системи. Основним критерієм оцінки економічної ефективності розробки САУ є зростання прибутку промислового підприємства з урахуванням зростання продуктивності суспільної праці, тобто. переважно з допомогою зниження витрат за виробництво продукції. Як кількісного вираження економічної ефективності створення САУ застосовується річна економія від запровадження САУ і термін окупності додаткових капітальних вкладень. Ключові слова: реактивність системи; імітаційне моделювання; закон Пуассона; пропускна спроможність системи; економічна ефективність; прибуток; термін окупності

Постановка задачі. При побудові автоматизованих інформаційних систем однією з важливих є оцінка ефективності їх функціонування. Це важливо

для визначення раціональної продуктивності системи, коли поряд з вимогами до функціональних можливостей ставиться завдання високоефективного її функціонування [1-4].

У діалогових системах одиницею робочого навантаження є запит. Під реактивністю системи будемо розуміти величину, зворотну часу проходження запиту T_x , тобто $\rho = 1/T_x$.

В якості оцінки ефективності (С) діалогової системи колективного користування будемо розглядати відношення середньої реактивності системи ρ до середніх витрат Z на один запит:

$$C = \frac{\bar{\rho}}{Z} = (T_x \times \bar{Z}) \quad (1)$$

Викладення основного матеріалу. У діалогових системах колективного користування, в яких колективним ресурсом є не тільки обчислювальні засоби, а й дані, витрати розподіляються між користувачами пропорційно часу, необхідного для реалізації запитів користувачів, тобто

$$z = k \cdot T_z, \quad (2)$$

де T_z — середній час реалізації однієї заявки, k — коефіцієнт пропорційності.

Таким чином, з точністю до коефіцієнта пропорційності маємо:

$$C = (T_x \cdot T_z)^{-1} \quad (3)$$

Досліджено величина (С) як критерій ефективності діалогової системи колективного користування. У якості методу дослідження обрано метод імітаційного моделювання [5].

Приймемо в якості робочого навантаження потік заявок з розподілом Пуассона, тобто ймовірність надходження наступної заявки складає $P = 1 - e^{-\lambda t}$, де λ — інтенсивність надходження заявок.

Припустимо, що час обслуговування заявки має такий же закон розподілу $P = 1 - e^{-\mu t}$, де μ — середня кількість вимог, які може бути обслужено за одиницю часу. Будемо розглядати випадок системи з одним каналом обслуговування.

Звісно, що в разі, коли $\mu \ll \lambda$ система набуває нестійкий характер, час очікування заявок на обслуговування істотно збільшується, що веде до зниження ефективності функціонування системи за критерієм (С). Випадок, коли $\mu \gg \lambda$ призводить до стану системи, при якому знижується коефіцієнт завантаження обладнання, збільшується час простою, що також веде до зниження критерію (С), тому що простої системи віднесені за рахунок користувачів. Найбільш ці-

кавою видається область параметрів системи, при яких $\mu \approx \lambda$. У цій області і буде вироблено моделювання роботи інформаційної системи.

Зробимо зауваження з приводу фізичного сенсу обраного нами критерію. Величина (С) представляє числове значення реактивності ρ , поділене на величину витрат, тобто оптимізація діалогової системи в сенсі критерію (С), зводиться до пошуку максимальної реактивності, що віддається системою за одиницю витрат.

В процесі моделювання необхідно з'ясувати чи є обрана нами величина (С) об'єктивним критерієм оцінки ефективності функціонування системи. Крім того, необхідно визначити поведінку системи за різними параметрами в перехідному процесі, коли $\lambda \approx \mu$.

Робота моделюючого алгоритму являє імітацію процесу надходження заявок в деякі моменти часу t_i , розділені інтервалами випадкової тривалості τ_i , і процесу обслуговування заявок в моменти часу t_j , розділені інтервалами випадкової тривалості Θ_j . В процесі моделювання були виявлені наступні закономірності. Залежність величини (С) від величини інтервалу між двома сусідніми запитами τ для різних значень величини Θ , яка показана на рисунку1 дзвоноподібними кривими. Максимум значень величини (С) для різних Θ як видно з рисунка досягається при $\lambda \approx \mu$, що дає підставу вважати величину (С) об'єктивним критерієм ефективності. Зауважимо, що при зменшенні пропускної здатності системи (збільшення Θ) оптимальне значення величини (С) зменшується.

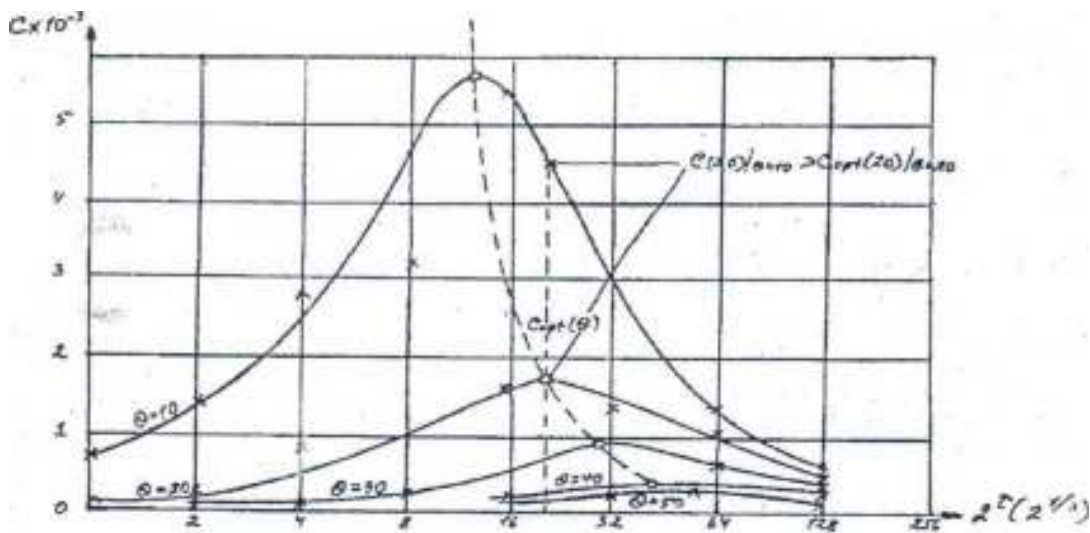


Рисунок 1 – Залежність величини (С) від величини інтервалу між двома сусідніми запитами τ_i для різних значень величини Θ

На рисунку 2 показана залежність величини T_z від величини τ при різних Θ . Як видно, для $\tau \in [1; 16]$ величина T_z практично не змінюється, і тільки при $\tau = 32$ і вище відбувається істотне збільшення T_z , Таким чином витрати на інформаційне обслуговування однієї заявки зменшуються при збільшенні інтенсивності надходження заявок (що може бути досягнуто підключенням нових користувачів).

Рисунок 3 показує залежність витрат на одну заявку від пропускної здатності (або часу обслуговування однієї заявки Θ). Як видно з рисунка при збільшенні пропускної здатності (зменшення величини Θ) витрати віднесені на одну заявку зменшуються.

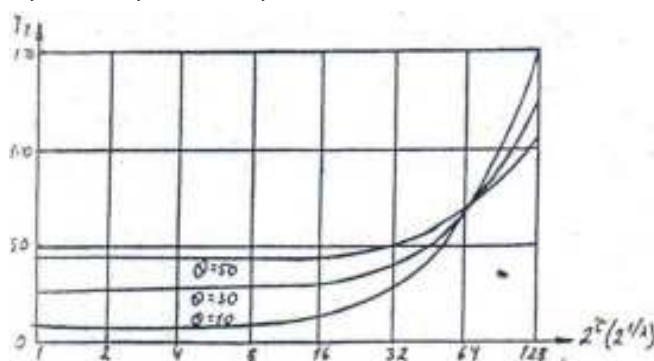


Рисунок 2 -Залежність величини витрат T_z від величини τ

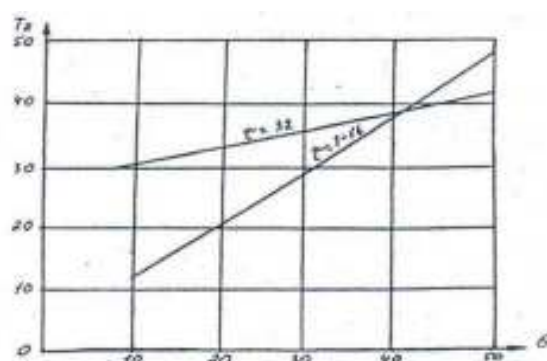


Рисунок 3–Залежність витрат T_z на одну заявку від пропускної здатності

Аналізуючи отримані результати моделювання можна зробити наступні висновки:

1. Величина (С) при виборі відповідної одиниці виміру шляхом підбору коефіцієнта пропорційності k може бути прийнята в якості одного з критеріїв оцінки ефективності функціонування діалогової інформаційної системи. Фізичний сенс критерію визначається збільшенням реактивності системи при збільшенні витрат на обслуговування однієї заявки. Величина, зворотна (С) може бути розглянута як ціна одиниці реактивності системи.

2. При збільшенні інтенсивності надходження заявок, S_{\max} збільшується. Таким чином, збільшення ефективності діалогової системи в сенсі величини (С) можливо шляхом більш повного завантаження більш потужної інформаційної системи. Так система з $\Theta = 10$ навіть наполовину завантажена вважається більш ефективною в сенсі (С), ніж повністю завантажена система з $\Theta = 20$.

3. При збільшенні інтенсивності надходження заявок витрати на обслуговування однієї заявки зменшуються. Однак при зменшенні інтенсивності є деяка нижня позитивна межа, нижче якій витрати не можуть бути. Так, в нашій

моделі при однаковій інтенсивності надходження заявок для всіх $\tau < 16$ витрати практично однакові.

4. При збільшенні пропускної спроможності системи витрати на обслуговування однієї заявки зменшуються. Залежність між величинами Θ і T_z лінійна. Лінія регресії тим крутіше, чим менше величина t . Це ще раз підтверджує тезу про те, що більш ефективною діалоговою системою є та, яка має більш високу пропускну здатність і є повністю завантажена. Якщо в інформаційну систему ввести документальний і фактографічний контури, то результати дослідження залежностей наведені на рис. 4 і рис. 5, де заштриховані області становлять предмет окремого викладу [2].

На рис.5 обрані залежності (С) від τ для Θ мають однакове відношення затрат між затратами на пошук і затратами на підготовку довідки, яку готує довідкова система (наприклад 1:4), відповідно: 2+8; 4+16; 8+32; 16+64.

На рисунку 5 стрілками вказані T_z при оптимальному (в сенсі С) τ_{opt} . З графіків видно, що, починаючи з τ_{opt} T_z починає збільшуватися істотно і стійко. Тобто критерій (С) показує, який запас в пропускній здатності є надмірною.

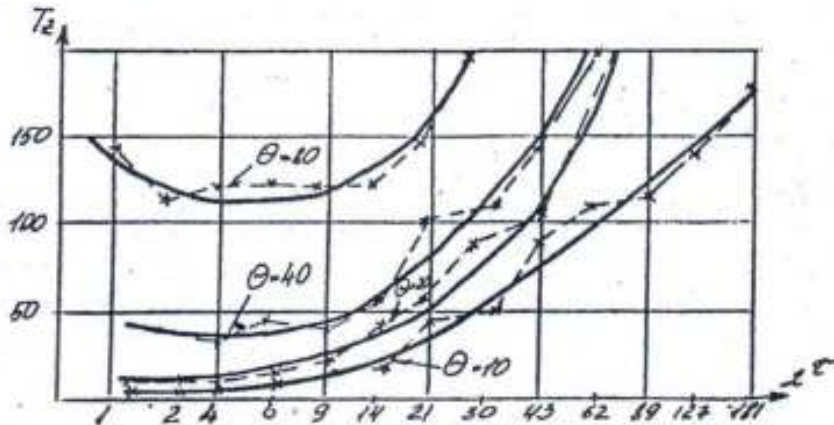


Рисунок 4 – Результати дослідження величини T_z від величини τ

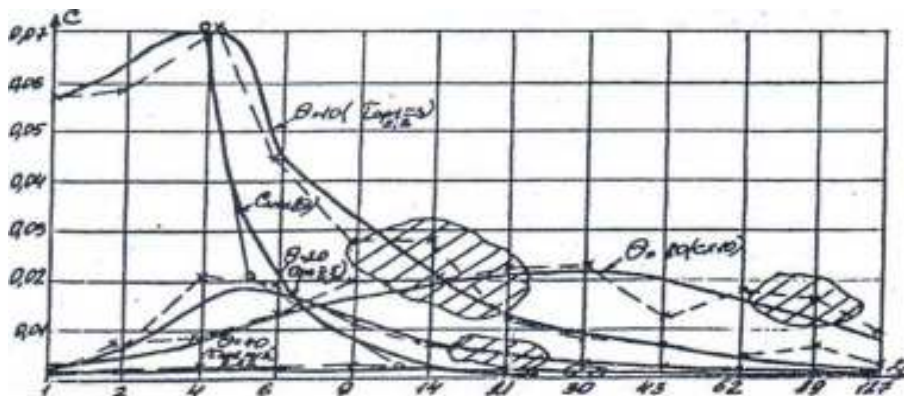


Рисунок 5–Оцінка ефективності функціонування системи (С) від змінної τ

Оцінка ефекту від впровадження системи автоматизованого управління метрологічним обслуговуванням засобів вимірювання на підприємстві. Результати функціонування даної системи як і будь-якої іншої САУ виявляються в джерелах її ефективності при впровадженні. У загальному випадку методи визначення економії ефективності САУ різних призначення значно відрізняються між собою головним чином складом елементів невпорядкованості (втрат) і ступенем впливу використаної на формації стан керованої системи.

Основним критерієм економічної ефективності розробки та впровадження САУ є зростання прибутку промислового підприємства на основі зростання продуктивності громадського праці, тобто в основному за рахунок зниження витрат на виробництво продукції. Як кількісного вираження економічної ефективності створення САУ приймається річна економія від запровадження САУ і термін окупності додаткових капітальних вкладень. Джерелами економічної ефективності у сфері виробництва при впровадженні САУ є: збільшення випуску продукції; зростання продуктивності праці завдяки кращій організації виробництва; скорочення простоїв та понад урочних робіт, підвищення коефіцієнта завантаження обладнання; зниження браку завдяки підвищенню ритмічності виробництва; зменшення нормованих оборотних засобів; скорочення витрат товарно-матеріальних цінностей; зниження непродуктивних витрат, пов'язаних із зменшенням штрафів, пені, неустойок; скорочення умовно незмінних витрат.

Чинники економії і власне економія за кожним чинником часом можуть бути визначені прямим рахунком. Це, зазвичай, можна зробити для САУ, дають велику економію, де здійснюються розрахунки з оптимізації плану виробництва чи використання САУ супроводжується значним скороченням витрат у виробництві. Економія, одержувана в результаті оптимізації плану виробництва, прямо пропорційна витратам і умовно постійним витратам і цілком входить у загальну економію від САУ. Прибуток, одержуваний від збільшення випуску продукції в результаті оптимізації плану виробництва, також додається до загальної економії у повному обсязі.

Найчастіше на етапі застосування САУ їх ефективність доводиться визначати з допомогою низки прийомів, які у сумі і становлять методіку.

Економічна ефективність САУ визначається річним приростом прибутку у зв'язку з функціонуванням САУ, річним економічним ефектом та ефективністю витрат на її створення.

Річний приріст прибутку (або економію) розраховують за формулою [6,7]

$$\Delta_{\text{пр}} = \left(\frac{A_2 - A_1}{A_1} \right) \Pi_1 + \frac{(C_1 - C_2)}{100} \cdot A_2, \quad (4)$$

де A_1, A_2 – річні обсяги продукції до і після впровадження САУ, грошових одиниць; C_1, C_2 – витрати на виробництво реалізованої продукції до і після впровадження САУ; Π_1 – прибуток від продукції до застосування САУ, грошових одиниць.

Річна економія від функціонування САУ використовується для розрахунку річного економічного ефекту і відображається у планах за собівартістю продукції та фінансових планах підприємств та галузей народного господарства.

Річний економічний ефект показує ступінь економічної ефективності САУ з урахуванням витрат на її створення та визначається за формулою:

$$\Delta = \left[\left(\frac{A_2 - A_1}{A_1} \right) \Pi_1 + \frac{C_1 - C_2}{100} \cdot A_2 \right] \cdot E_n \cdot K^a, \quad (5)$$

де E_n – нормативний коефіцієнт економічної ефективності капітальних вкладень у даній галузі; K^a – капітальні вкладення, пов'язані зі створенням та використанням САУ, грошових одиниць.

Ефективність витрат визначається розрахунковим коефіцієнтом ефективності витрат

$$E_p = (\Delta_{\text{пр}} / K^a) \cdot E_n \quad (6)$$

або терміном окупності витрат

$$T = K^a / \Delta_{\text{пр}}. \quad (7)$$

Розрахунковий коефіцієнт ефективності E_p зіставляється з нормативним значенням E_n і у разі, коли розрахунковий коефіцієнт більший за нормативний, САУ вважається ефективною. Як правило, розрахунковий коефіцієнт ефективності САУ має суттєво перевищувати нормативний для цієї галузі.

Термін окупності витрат T на САУ є показником, що характеризує період часу, протягом якого загальні витрати відшкодовуються за рахунок економії поточних витрат та додаткового прибутку від реалізації продукції.

При розрахунку фактичної економічної ефективності САУ використовуються фактичні дані про зміну показників – виробничо-господарської діяльності у зв'язку з її впровадженням (див.табл.1-4). Основними складовими економічного ефекту від функціонування системи САУМОЗВТ є: скорочення докуме-

нтів управління (обліку, контролю, планування), що складаються вручну, та їх уніфікація; покращення організації та підвищення якості планування МО парку ЗВТ; ефективне використання ЗВТ зі зниженою класністю; зниження відсотка браку, викликаного використанням ЗВТ, яке відмовило (метрологічно) при контролі виробленої промислової продукції; зниження трудомісткості управлінських процесів. Отже, річна економічна ефективність системи може бути розрахована за формулою:

$$\mathcal{E} = \mathcal{E}_{к.б.} + \mathcal{E}_и + \mathcal{E}_{з.п.} + \mathcal{E}_{сс} + \mathcal{E}_{уд} - 3 \cdot K_2 \quad (8)$$

$$\mathcal{E}_{к.б.} = \left\{ \left[\frac{\Phi}{\Phi - B \cdot \frac{b_n}{100} \cdot T_{з.б.}} - 1 \right] \left[N + \left(E_n + \frac{a_0}{100} \right) K \right] + B \cdot \frac{b_n}{100} \cdot T_{з.б.} \cdot C_p \right\} \quad (9)$$

де — економічна ефективність підвищення якості перевірочних робіт і зниження відсотка браку продукції Φ річний фонд часу роботи обладнання; B — річний випуск виробів, b_n — зменшення випадків шлюбу продукції % ($b_n = \frac{НБн}{Нге} \cdot 100$); $T_{з.б.}$ — час виявлення та заміну браку на наступних стадіях технологічного процесу виготовлення продукції чи споживача, N — річна сума умовно-постійних витрат при випуску продукції, a_0 — середня амортизація, % (~13%), K — капітальні вкладення; C_p — вартість одиниці часу працівника; $\mathcal{E}_и = ([\Sigma C_{p2} - \Sigma C_{p1}] \cdot N)$ — економічна ефективність від застосування ЗВТ зниженої класності, ΣC_{p2} , ΣC_{p1} — суми вартості ЗВТ відповідно підвищеної та зниженої класності; N — кількість ЗВТ; — економія від уніфікації документів; $\mathcal{E}_{з.п.} = ([Ч - \Delta Ч] \cdot 3п \cdot 12)$ — економічна ефективність від умовного вивільнення чисельності штату у зв'язку зі зниженням трудомісткості управлінських процесів, $Ч$ — чисельність персоналу; $\Delta Ч$ — чисельність персоналу після вивільнення, $3п$ — середня зарплата одного працівника (~185 руб); $\mathcal{E}_{сс} = (\mathcal{E}_{з.п.} \cdot K_1)$ — економія відрахування на соцстрах, K_1 — відсоток відрахувань на соцстрах (~14%), 3 — одноразові витрати на проектування, підготовку та впровадження САУ, включаючи НДР та ДКР, K_2 — нормативний коефіцієнт (~ 15%).

Термін окупності

$$T = \frac{3}{(\mathcal{E}_{к.б.} + \mathcal{E}_и + \mathcal{E}_{з.п.} + \mathcal{E}_{сс} + \mathcal{E}_{уд})} \quad (10)$$

Викладена вище методика розрахунку економічного ефекту враховує всі його складові (на цьому етапі створення системи).

Нині через відсутність низки вихідних даних розрахунок ефективності (фактичної) виробляється дещо спрощеним способом. Так, згідно з цим способом, економічний ефект визначають за формулами:

$$\mathcal{E} = Z_1 - Z_2, \quad (11)$$

де Z_1 — річні експлуатаційні витрати під час ручної обробки, Z_2 — загальні річні експлуатаційні витрати під час запровадження САУ.

$$Z_1 = T \cdot C_{н.ч}, \quad (12)$$

де T — трудомісткість, нормо-год / рік; $C_{н.ч}$ — вартість однієї нормо-години, грошових одиниць.

$$C_{н.ч} = \frac{Z_{п.} (1 + НР + ОС)}{T_{г}}, \quad (13)$$

де $Z_{п.}$ — середня заробітна плата на місяць; $НР$ — накладні витрати, %; $T_{г}$ — середня кількість годин на місяць; $ОС$ — відрахування на соцстрах, %.

$$T = N \cdot T_{уд}, \quad (14)$$

де N — кількість одиниць ЗВТ; $T_{уд}$ — трудомісткість, нормо-год, на один прилад, по всіх розв'язуваних задачах; $T_{уд} = T_1 + T_2 + \dots + T_n$, n — кількість розв'язуваних задач.

$$Z_2 = Z'_2 + K \cdot E_n, \quad (15)$$

де Z'_2 — річні експлуатаційні витрати під час застосування САУ; K — витрати використання САУ; E_n — нормативний коефіцієнт.

$$Z'_2 = Z_{еом} + Z_{п.д}, \quad (16)$$

де $Z_{еом}$ — загальна вартість вирішення завдань на ЕОМ; $Z_{п.д}$ — витрати на підготовку даних на рік

$$Z_{еом} = \left(\sum_{i=1}^n T_{i3} \right) \cdot \Pi_{ч.еом}, \quad (17)$$

де n — кількість розв'язуваних завдань; T_{i3} — час на одне (i -е) завдання; $\Pi_{ч.еом}$ — вартість однієї години роботи ЕОМ.

$$Z_{п.д} = T_{год} \cdot C_{н.ч}, \quad (18)$$

де $T_{год}$ — час, необхідний на підготовку даних на рік;

$C_{н.ч}$ — вартість однієї нормо-год.

Отже,

$$K = Z_{п.} \cdot (1 + НР + ОС) \cdot 12. \quad (19)$$

$$\begin{aligned}
 Z_2 &= T_{\text{зад}} \cdot C_{\text{ч.еом}} + T_{\text{год}} \frac{Z_{\text{п}}}{T_{\text{ч}}} (1 + \text{НР} + \text{ОС}) + Z_{\text{п}} (1 + \text{НР} + \text{ОС}) \cdot 12 \cdot E_{\text{н}} = \\
 &= \left(\sum_{i=1}^n T_{\text{из}} \right) \cdot C_{\text{ч.еом}} + [Z_{\text{п}} \cdot (1 + \text{НР} + \text{ОС})] \left[(T_{\text{год}} / T_{\text{ч}}) + 12 \cdot E_{\text{н}} \right],
 \end{aligned}
 \tag{20}$$

де $T_{\text{зад}}$ – час, необхідний на виконання задач на ЕОМ.

Тоді

$$\begin{aligned}
 \Delta &= Z_1 - Z_2 = N \cdot T_{\text{уд}} \frac{Z_{\text{п}}}{T_{\text{ч}}} (1 + \text{НР} + \text{ОС}) - n \cdot T_{\text{з}} \cdot C_{\text{ч.еом}} + \\
 &+ [Z_{\text{п}} \cdot (1 + \text{НР} + \text{ОС})] \left[(T_{\text{год}} / T_{\text{ч}}) + 12 \cdot E_{\text{н}} \right].
 \end{aligned}
 \tag{21}$$

В економічну ефективність, що розраховується за формулою, слід включити економію від підвищення якості сукупності вимірювань у виробництві промислової продукції і ефект від поліпшення умов експлуатації ЗВТ.

Термін окупності може бути розрахований за формулою

$$T = (Z_2 + K) / \Delta. \tag{22}$$

Наведемо приклади розрахунку фактичного річного економічного ефекту від застосування системи на конкретних підприємствах.

Приклад 1. Розрахунок економічної ефективності від застосування системи на Ступинському металургійному комбінаті. Вихідні дані для розрахунку становлять: $Z_{\text{п}} = 161$ грошових одиниць., $T_{\text{г}} = 174$ год, $\text{НР} = 123\%$, $\text{ОС} = 14\%$, $N = 3000$ шт., $T_{\text{уд}} = 0,21$ нормо-год, $E = 0,15$, $n = 6$, $T_{\text{из}} = 0,05$ год. $C_{\text{ч.еом}} = 60$ грошових одиниць. $T_{\text{год}} = 9$ год.

Підстановка цих даних у вихідну формулу дала такі результати: $\Delta = 13091,4$ грошових одиниць., а термін окупності $T = 0,33$ року (розрахунок зроблено по цінах 80-х років минулого століття).

Приклад 2. Розрахунок економічної ефективності від застосування системи НДІРФ ім. Л. А. Расплетіна. Вихідні дані для розрахунку становлять: $N=2500$ шт., $Z_{\text{п}} = 185$ грошових одиниць., $T = 124$ год., $\text{НР} = 190\%$, $\text{ОС} = 14\%$, $E_{\text{н}} = 0,15$, $n = 6$, $T_{\text{из}} = 0,05$ год, $C_{\text{ч.еом}} = 60$ грошових одиниць., $T_{\text{год}} = 9$ год.

Підстановка цих даних у вихідну формулу дала наступний результат: $\Delta = 18\,575,5$ грошових одиниць. (розрахунок зроблено по цінах 80-х років минулого століття).

Якщо говорити про непрямі складові ефективності застосування системи, то слід зазначити таке. Перше позитивне зрушення після впровадження системи спостерігається в частині систематизації вимірювальної техніки, наведення

порядку в обліку ЗВТ. Оскільки для системи потрібна інформація про наявність різних ЗВТ, їх стан, технічні характеристики, то її впровадженню передують етапи паспортизації вимірювальної техніки. Це призводить до наведення ладу у вимірювальному господарстві. Наступне, що приносить використання системи автоматизації

МОЗВТ — це наведення порядку в організації перевірочних робіт, що можна зробити і вручну, але з величезними витратами; ЕОМ видає графік перевірки приладів за місяцями та підрозділами. Перший графік передається в метрологічну службу для планування перевірочних робіт, інший — підрозділам для планування пред'явлення наявних у них ЗВТ на перевірку. Графік складається з урахуванням рівномірного завантаження метрологічної служби. Це сприяє виробництву перевірочних робіт та найкоротші терміни. З іншого боку, регулярні перевірки ЗВТ сприяють підтримці їх високих технічних характеристик, що, у свою чергу, позитивно позначається на якості продукції, що випускається. У разі відхилень у перевірочних роботах (прилад не пред'явлений вчасно на перевірку або підлягає списанню за результатами перевірки) система видає відповідну інформацію як заявок придбання приладів замість списаних чи попереджень підрозділам про зрив термінів пред'явлення ЗВТ на перевірку. Це сприяє своєчасному оновленню парку ЗВТ, підтримці його у задовільному технічному стані, виконавчій дисципліні підрозділів, що експлуатують ЗВТ тощо.

Далі в процесі функціонування системи накопичується статистика відмов ЗВТ, на її підставі провадиться перерахунок міжперіодичного інтервалу (МПІ) з урахуванням різних критеріїв (якості продукції, що випускається, економічних показників тощо). Подальше планування процесу перевірочних робіт проводиться на підставі нового МПІ, що призводить до зниження втрат на перевірочні роботи, покращення комплексних показників роботи метрологічної служби підприємства.

Система дозволяє розраховувати виробничі площі під метрологічні служби, визначати трудомісткість виконання робіт з метрологічного обслуговування. Причому визначається верхня та нижня межі трудомісткості виконання таких робіт. Т. е. можна оцінити, до якої величини можна скоротити трудомісткість виконання перевірочних робіт за рахунок раціональної організації та суміщення робіт, освоєння суміжних професій і т.п.

Оптимізація вироблення керуючих рішень в САУМОЗВТ заснована на базі моделей (або однієї комплексної багатопараметрової моделі) процесів обслуговування ЗВТ за конкретними приватними критеріями (або за декількома

критеріями одночасно — багатокритеріальна оптимізація) згідно з методом декомпозиції загальної мети функціонування системи (рисунк 6). Для схеми на рис. 6 характерні три етапи; 1) процес налаштування та коригування моделей під поточний стан об'єкта (СМОЗВТ) або на вимогу ззовні; 2) процес коригування моделей за недосягнення цілей оптимізації; 3) програвання (перебір) на моделях варіантів керуючих рішень, порівнювання їх між собою за певним критерієм (критеріями), запам'ятовування найкращого та видача керуючого рішення.

Таким чином, раціональна САУМОЗВТ промисловими підприємствами має будуватися як інформаційно-довідкова система, як правило, на базі ЕОМ.

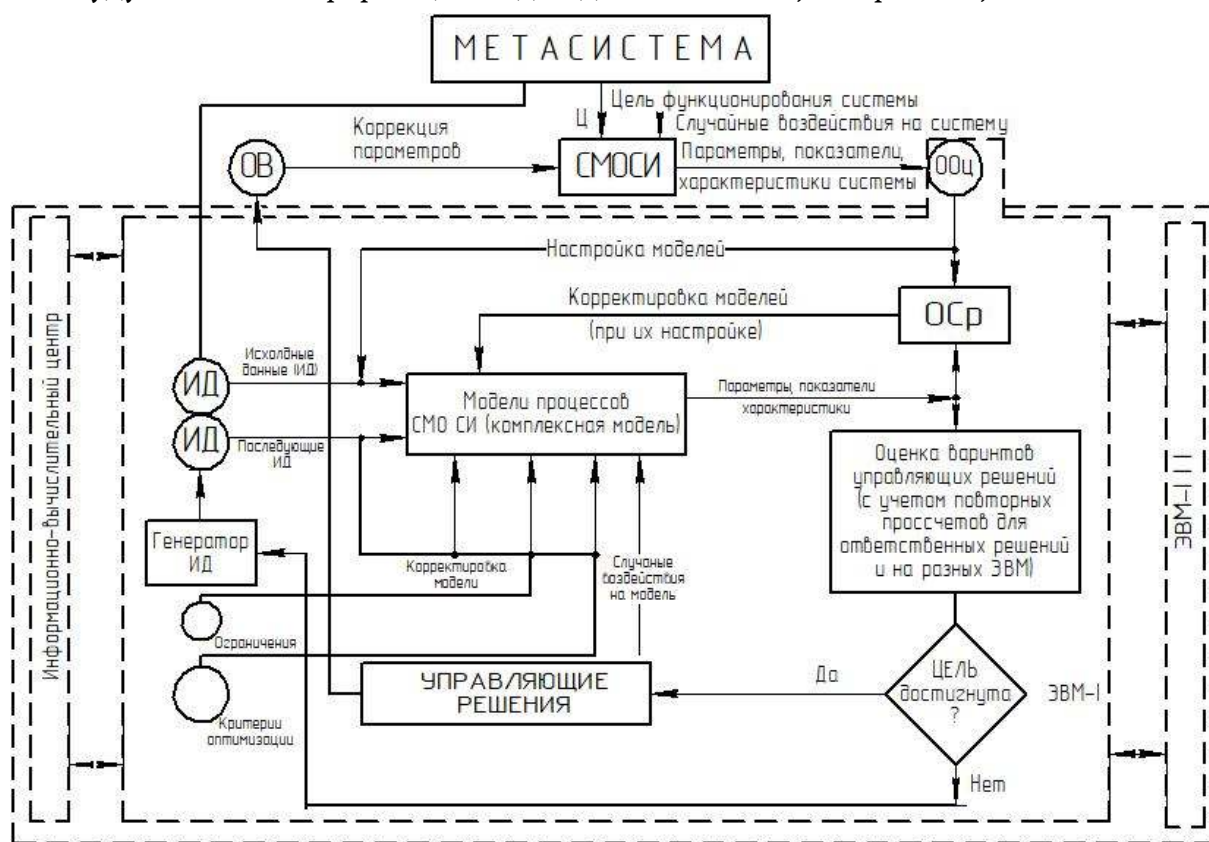


Рисунок 6 — схема оптимального (у певному сенсі) вироблення керуючих рішень в САУМОЗВТ: ОВ — орган впливу; ИД- вхідні дані; ООц — орган оцінки стану; ОСр— орган порівняння [8]

Розробка програмного забезпечення може проводитись із застосуванням мови програмування PL/1. Інформаційне забезпечення може будуватися як у вигляді системи лінійних наборів даних, так і із застосуванням СУБД. Таким вимогам задовольняє система автоматизації метрологічного обслуговування САУ МО ЗВТ , опис якої наведено у [9].

Розрахунок економічної ефективності, одержаної в результаті втілення першої черги автоматизованої системи ЗВТ на Дніпровському машинобудівному заводі. (розрахунок зроблено по цінах 80-х років минулого століття). Результати розрахунків економічної ефективності наведено у таблицях 1-4.

Таблиця 1

Основні техніко-економічні показники

Найменування показників	До впровадження заходів	Після впровадження заходів
Трудомісткість річної програми нормо-год.	22016,2	7477
Трудомісткість річної програми чол-год.	22016,2	7477
Кількість виробничих робочих, необхідні виконання плану, чол.	10,6	3,6
Виробіток на одного робітника, нормо-год.		
Собівартість за статтями витрат, що змінюються, грошових одиниць.	2077	2077
Умовно-річна економія від зниження собівартості, грошових одиниць.	16048	6293
Річний економічний ефект, грошових одиниць.	—	9661
Додаткові капітальні вкладення, грошових одиниць.	—	626
Окупність додаткових капітальних вкладень, рік.	—	0,06

Таблиця 2

Собівартість за статтями витрат, що змінюються

Стаття затрат	До впровадження заходів	Після впровадження заходів	Результат (економія, перевитрата)
Основна заробітна плата, грошових одиниць.	14004	5820	-8724
Додаткова заробітна плата (8%), грошових одиниць.	1120	422	-698
Витрати, пов'язані з роботою обладнання, грошових одиниць.	924	376	-548
Амортизаційні відрахування	—	64	+64
Поточний ремонт та утримання обладнання	—	151	+151
ВСЬОГО	16048	6293	9755

З табл. 2 видно, що економія від зниження собівартості становить 9755 грошових одиниць.

Уточнений розрахунок економічної ефективності та розрахунок терміну окупності капітальних вкладень від запровадження САУМОЗВТ проводиться за формулою

$$\Delta = (C_1 - C_2) \cdot E \cdot K$$

де C_1 - собівартість за змінними статтями витрат до впровадження, що дорівнює 16048 грошових одиниць.; C_2 - собівартість за змінними статтями витрат після застосування, що дорівнює 6293 грошових одиниць.; E - нормативний коефіцієнт порівняльної ефективності капітальних витрат, що дорівнює 0,15; K - капітальні витрати на придбання технічних засобів, рівні 626 грошових одиниць.

Річна економічна ефективність $S = 9661$ грошових одиниць.

Окупність капітальних витрат визначається за формулою $T = K / (C_1 - C_2)$ і дорівнює 0,06 року.

Таблиця 3

Розрахунок кількості виробничих робітників та фонду заробітної плати

Найменування показників	До впровадження заходів	Після впровадження заходів
Трудомісткість річної програми, чол-ч.	22016,2	7477
Виконання норм, %		
Трудомісткість річної програми з урахуванням виконання норм, чол-год.	104,6 21048	104,6 7148
Річний фонд часу одного робітника, Кількість виробничих робітників, чол.	1986	1986
Основна заробітна плата, грошових одиниць.	10,6	3,6
Додаткова заробітна плата, грошових одиниць.	14004	5280
Відрахування на соціальне страхування, грошових одиниць.	1120 924	422 376
ВСЬОГО	16048	6078

Витрати на поточний ремонт та утримання обладнання становлять 151 грошова одиниця.

Вартість обладнання та амортизаційні відрахування

Найменування обладнання	Кількість, шт.	Балансова вартість, грошових одиниць.		Амортизаційні відрахування.\	
		Одиниці	Всього	%	Сума, грошових одиниць.
Устаткування, яке застосовується після впровадження технічних засобів	1	626	626	10,2	63,8

Висновки:

1. В якості оцінки ефективності діалогової системи колективного користування САУМОЗВТ використовується відношення середньої реактивності системи до середніх витрат на один запит.

2. Збільшення ефективності діалогових систем колективного користування можливо шляхом більш повного навантаження більшої потужності інформаційної системи.

3. Вибраний критерій оцінки показує який запас пропускної здатності системи є надмірним.

4. Наведено результати розрахунків економічного ефекту від запровадження САУМОЗВТ на кількох промислових підприємствах та НДІ.

5. Запропоновано схему оптимальної (у певному сенсі) вироблення керуючих рішень в САУМОЗВТ.

Наукова новизна: Запропоновано метод вироблення оптимальних (у певному сенсі) керуючих рішень в САУМОЗВТ — як приклад реалізації конкретної системи підтримки прийняття рішень (СППР) та оптимізації процесів метрологічного обслуговування засобів вимірювальної техніки на промислових підприємствах.

ЛІТЕРАТУРА

1. Стогний А.А. Об основных принципах построения автоматизированных информационных систем // Управляющие системы и машины, 1972, №2. -- С.5-.10.
2. Мордань В.И. О реализации документального контура многоцелевой информационно-справочной системы// Управляющие системы и машины ,1984, №1. -- С.93-98.

3. Дж.Мартин Организация баз данных в вычислительных системах. – М.: Мир, 1980. –662 с.
4. Ігнаткін В.У. та інші. Математичне забезпечення технічних засобів вимірювання і контролю. Монографія за науковою редакцією професора В.У.Ігнаткіна,.- Кам'янське, ДДТУ, 2019. – 340с.
5. Шеннон Р. Иммитационное моделирование систем.– Искусство и наука.- М.: Мир, 1978. – 418 с.
6. Методика оценки экономической эффективности отраслевых автоматизированных систем управления (ОСАУ) в промышленных министерствах, всесоюзных и республиканских промышленных объединениях (утв. 31 июля 14974 г.ОН№485/109/40).
7. Отраслевой справочник нормативов времени для нормирования труда работников отдела метрологии промпредприятий и объединений/ Минприбор СССР. – М., 1984 г.
8. Ігнаткін В.У. Особливості аналізу динамічної погрішності при оцінці метрологічної надійності засобів вимірювання. //Метрологія і прилади, 2019, №2. -- С. 20–30.
9. Ігнаткін В.У. Автоматизована система метрологічного обслуговування на промислових підприємствах. Міжнародна науково-практична конференція «Прикладні науково-практичні дослідження», м. Івано-Франківськ, «Симфонія», 5-7 квітня 2017. – С. 86.

REFERENCES

1. Stogny. A.A. On the basic principles of building automated information systems // Control systems and machines, 1972, no. 2. – P. 5-.10.
2. Mordan V.I. On the implementation of the documentary contour of a multi-purpose information and reference system / / Control systems and machines, 1984, No. 1. – P. 93-98.
3. J. Martin. Organization of databases in computing systems. - М.: Mir, 1980. -662 p.
4. Ignatkin V.U. and others. Mathematical security of technical methods of vimiryuvannya and control. Monograph for the scientific editorial staff of Professor V.U.
5. Shannon R. Simulation modeling of systems. - Art and science. - М.: Mir, 1978. - 418 p.
6. Methodology for assessing the economic efficiency of sectoral automated control systems (OASU) in industrial ministries, all-Union and republican industrial associations (approved July 31 1974, No. 485/109/40).

7. Industry directory of time standards for labor rationing of employees of the department of metrology of industrial enterprises and associations / USSR Ministry of Instruments. - M., 1984.
8. Ignatkin V.U. Peculiarities of the analysis of the dynamic error in the assessment of the metrological reliability of the measurements. //Metrology and fittings, 2019, No. 2. – P. 20–30.
9. Ignatkin V.U. The system of metrological service at industrial enterprises has been automated. International Scientific and Practical Conference "Applied Scientific and Practical Research", Ivano-Frankivsk, Symphony, April 5-7, 2017. - P. 86.

Received 27.06.2022.

Accepted 01.07.2022.

Evaluation of efficiency of the automated control system of metrological maintenance of measuring instruments - as informatic information

As an assessment of the effectiveness of the dialog system of collective use, on the example of the system of automated control of metrological maintenance of measuring equipment (SACMMME), the ratio of the average reactivity of the system to the average cost per request is used. The ratio of the average reactivity of the target system to the average cost per request (C) is studied. Under the reactivity is the value inverse of the time of the request in the system. The flow of queries in the system obeys the Poisson distribution law. The method of simulation modeling is chosen for research. Obtained: time dependences (C) between two adjacent queries; the dependence of the cost of one request on the intensity of requests; the dependence of the cost of one request on the bandwidth of the system, etc. Increasing the efficiency of shared dialogue systems is possible through a fuller load of more information system power. The selected evaluation criterion shows which margin in the system bandwidth is excessive.

The results of calculations of economic efficiency from the introduction of SACMMME at several enterprises and research institutes are given. In the general case, the methods of determining the economic efficiency of ACS for different purposes differ in the composition of the elements of disorder (loss) and the degree of influence of the information used on the state of the managed system. The main criterion for assessing the economic efficiency of ACS development is the growth of profits of industrial enterprises, taking into account the growth of productivity of social labor, ie. mainly by reducing production costs. The annual savings from the introduction of ACS and the payback period of additional capital investments are used as a quantitative expression of the economic efficiency of ACS creation.

Ігнаткін Валерій Устинович – доктор технічних наук, професор, професор кафедри автоматизації та комп’ютерно- інтегрованих технологій Івано-Франківського національного технічного університету нафти і газу, Україна.

Ignatkin Valeri Ustinovich – Doctor of Engineering Sciences, Professor of the Department of Automation and Integrated Computer Technologies at Ivano-Frankivsk National Technical University of Oil and Gas, Ukraine.

Хомош Юрій Степанович – кандидат економічних наук, доцент, директор Дрогобицького фахового коледжу нафти та газу.

Khomosh Yury Stepanovich – Candidate of Economic Sciences, Associate Professor, Director of the Drogobychev Professional College of Oil and Gas.

Дудніков Володимир Степанович – кандидат технічних наук, доцент кафедри мехатроніки Дніпровського національного університету імені Олеся Гончара.

Dudnikov Volodymyr Stepanovich – Candidate of Engineering Sciences (PhD in engineering sciences), associate professor of the department of mechatronics at Oles Honchar Dnipro National University.

Юшкевич Олег Павлович – кандидат технічних наук, доцент кафедри мехатроніки Дніпровського національного університету імені Олеся Гончара

Yushkevich Oleh Pavlovych – Candidate of Engineering Sciences (PhD in engineering sciences), associate professor of the department of mechatronics at Oles Honchar Dnipro National University.

Тіхонов Василь Андрійович – директор Дніпровського фахового коледжу радіоелектроніки.

Tikhonov Vasyl Andriyovich - is the director of the Dnipro Vocational College of Radio Electronics.

К.Ю. Островська, Т.М. Фененко, О.О. Глущенко

СТАТИСТИЧНИЙ АНАЛІЗ ТЕКСТУ

ТА ДОСЛІДЖЕННЯ ДИНАМІКИ ТОЧНОСТІ КЛАСИФІКАЦІЇ

Анотація. Робота присвячена статистичному аналізу тексту та дослідженню динаміки точності класифікації. У роботі проводиться відбір статистичних ознак тексту, класифікація текстів, що належать різним авторам, та дослідження динаміки точності класифікації в залежності від довжини текстових фрагментів. Для вирішення поставленого завдання використовувалися: методи обробки природної мови; статистичні характеристики текстів; методи машинного навчання; методи зниження розмірності для можливості візуалізації. На основі отриманої динаміки зміни точності класифікації в залежності від довжин текстових фрагментів було зроблено відповідні висновки щодо оптимальної довжини текстів, що використовуються для навчання та тестування моделей. Завдання вирішувалося у програмному середовищі Jupyter Notebook дистрибутива Anaconda, який дозволяє одразу встановити Python та необхідні бібліотеки.

Ключові слова: машинне навчання, статистичний аналіз текста, визначення авторства, аналіз даних, обробка природного мови.

Огляд предметної області. Ідентифікація авторства тексту широко використовується у різних напрямках, зокрема, для експертизи у криміналістиці, для боротьби з плагіатом, встановлення авторства анонімного тесту тощо.

Завдання ідентифікації дуже складне. По-перше, виникає проблема формування набору ознак, за якими визначається ймовірність належності певного тексту окремому автору. По-друге, до останнього часу для якісної і точної роботи нині розроблених моделей визначення авторства необхідні великі об'єми авторських текстів у навчальній виборці. По-третє, якісне обмеження розроблених моделей на кількість авторів.

1. Статистичний аналіз тексту.

1. Вилучення ознак із передопрацьованого тексту. Для глибшого вивчення вихідного тексту необхідно звернутися до методів, які дозволять виявити деякі ознаки характерні для наявних даних.

1.1 Мішок слів "bag-of-words". Ця модель тексту є сумативне єдність слів, що становлять вихідний текст.

Одиниці «мішка слів» – слова, кожне з яких має атрибут, а саме кількість зустрічей даного слова в тексті.

Важливими особливостями даної моделі є врахування порядку слів у документі та морфологічних форм подання слів.

Робота «мішка слів»:

1. Дана вибірка, що складається з n різних слів: w_1, \dots, w_n .
2. Текст кодується за допомогою n ознак, де i -та ознака позначає частку входження слова w_i серед усіх входження слів у текст.

Зазвичай «мішок слів» застосовується вже до попередньо обробленого тексту, з якого були виключені всі стоп-слова. Дані слова не несуть у собі смислове навантаження, тому часто частка їх входження до тексту не враховується.

1.2 Мішок термів "bag-of-terms". На відміну від мішка слів, його елементом є терм, який характеризується частотою народження в тексті.

Його елементом є терм, який характеризується частотою народження в тексті.

Під термом розуміється символічне вираження об'єкта формальної моделі (системи, мови).

1.3 TF-IDF метод векторизації текстових даних. Метод TF-IDF Vectorizer використовується для призначення ваг словами вихідних текстів.

1.4 N-грами. Перевага n -грам полягає в тому, що вони дозволяють враховувати порядок слів, на відміну від «мішка слів». Крім цього, використання n -грамів розширює ознаковий простір завдяки обліку словосполучень. Тому прості моделі можуть бути пошуку більш складних закономірностей проти «мішком слів».

Використання n -грамів розширює ознаковий простір завдяки обліку словосполучень. Серед n -грам можна виділити:

- Уніграми: набори, що складаються з одного токена;
- Біграми: набори, що складаються з двох токенів, що йдуть поспіль;
- Триграми: набори, що складаються з трьох токенів, що йдуть поспіль.

1.5 Хешування. $h(x)$ – хеш-функція, яка набуває 2^n можливих значень. Даний метод передбачає заміну всіх слів x на їх хеші $h(x)$ та використання цих хешей як токенів. Після цього можна застосовувати описані раніше методи як «мішок слів», TF-IDF векторизатор і т.д.

Застосування хешування має низку переваг. Насамперед, це спрощення зберігання моделі. У разі значення хеша можна ототожнити з індексом конкретного слова (токена).

1.6 Стоп-слова. Стоп-слова – це слова, які зустрічаються у великому обсязі текстів і не несуть особливого змістового навантаження. Зазвичай до них відносять вигуки, прийменники, частинки, деякі займенники, прикметники та інші частини мови. Тому найчастіше передопрацьований текст очищають від стоп-слів.

2. Виділення показників. Очевидно, що з тексту можна виявити нескінченну кількість характеристик.

Ознаки, що витягуються з тексту:

1. Ознаки, засновані на розділових знаках;
2. Ознаки, що базуються на словах як одиницях тексту;
3. Ознаки, де одиницею є окремий символ – літера.

Перша група може досліджувати розподіл розділових знаків по тексту, взаємозв'язок між розділовими знаками і загальним числом символів, слів у тексті.

Друга група дає можливість отримати ознаки, засновані на довжинах слів, частотах вживання різних частин мови, довжинах речень, важливості слів, «стоп-словах» чи унікальних словах.

Третя група, наприклад, дозволить досліджувати відношення між великими і малими літерами, будувати біграми та ін.

Постановка задачі визначення авторства. Завдання визначення авторства полягає в наступному: перед нами є набір текстів (фрагменти творів авторів) української чи іншою мовою. Є неідентифікований текст, але відомо, що він належить комусь із перелічених вище авторів. Необхідно визначити автора цього тексту.

Досліджуватимемо динаміку точності класифікації текстів при зміні довжини тексту, що класифікується. Для вирішення цього завдання будуть використані методи, описані вище, а також будуть відбиратися ознаки та використовуватися у поєднанні з іншими підходами для виявлення поєднання, яке дозволить отримати найкращий результат.

3. Побудова тренувального та тестового множин. Етапи збору та побудови даних:

1. Збір текстів авторів;
2. Завантаження текстів;
3. Розбиття текстів на рівні ділянки, попередня обробка;
4. Конструювання тренувального та тестового наборів.

Спочатку проводився збір творів для відібраних авторів та завантаження їх із електронної бібліотеки. Потім із текстів видалялися службові символи, що перешкоджають подальшому аналізу.

Всі твори для окремого автора об'єднувалися в єдиний рядок, з якого нарізалися потім рядки певної довжини. Динаміка точності

Класифікація текстів досліджувалась залежно від зміни довжини даних рядків. Щоб уникнути перетину тестового та тренувального наборів, відрізки для тренувального набору відбиралися з першої половини рядка, об'єднує твори конкретного авторів, а тестового – з другої відповідно.

4. Розв'язання задачі. Для вирішення завдання визначення авторства необхідно з величезної кількості параметрів (ознак) вибрати ті, які будуть використовуватися для аналізу текстів. На певних етапах відбору ознак текст приводився до нижнього регістру, здійснювалася токенізація та лематизація.

Як основні були відібрані такі статистичні характеристики тексту:

- Відношення кількості великих літер до кількості малих літер;
- Розподіл різних розділових знаків за текстом.
- Розподіл довжин речень.
- Розподіл довжин слів.
- Визначалася водність тексту.
- Різноманітність мови.
- Розподіл частин мови у тексті.

Крім цих ознак, розраховувалися літерні біграми для корпусів текстів. На основі тренувальної вибірки визначалися біграми, що найчастіше зустрічаються, їх частоти використовувалися в якості ознак. При тестуванні підраховувалися частоти відібраних на етапі навчання біграм, що найчастіше зустрічаються.

5. Класифікація текстів. Матриця помилок. Для оцінки отриманих класифікаторів будувалася матриця помилок, що ґрунтується на таблиці сполученості. По одній осі розташовуються оригінальні мітки класів, з іншого – передбачені. Головна діагональ матриці вказує кількість чітко передбачених значень для кожного з класів.

Невірно передбачені елементи будуть розміщені поза головної діагоналі.

Сума діагональних елементів є все правильно класифіковані об'єкти. Загальну точність класифікації можна виразити як відношення суми діагональних елементів (d_i) до загальної кількості елементів (N) формулою (1).

$$Overall Accuracy = \frac{\sum d_i}{N} \quad (1)$$

Можна вважати точність визначення реального (розрахованого) класу: розділивши число правильно класифікованих об'єктів на загальну кількість об'єктів цього класу. Формула для першого класу матиме вигляд (2):

$$Producer's Accuracy = \frac{d_{11}}{a_{1i}} \quad (2)$$

Матриця помилок будувалася більш глибокого аналізу результатів класифікації. На рисунках 1 – 6 наведено дані матриці, побудовані з урахуванням класифікаторів, навчених текстах різної довжини.

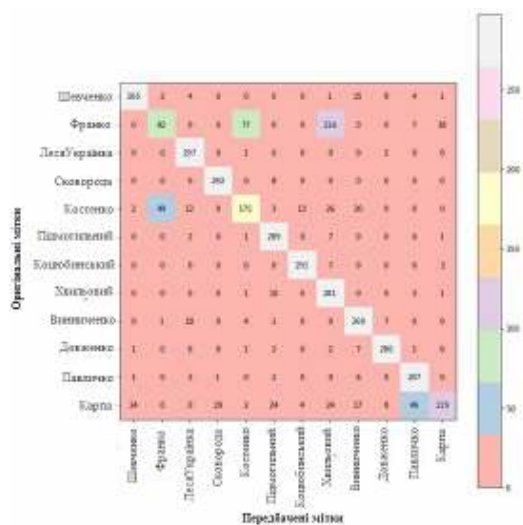


Рисунок 1 - Матриця помилок на текстах довжиною 20000 символів

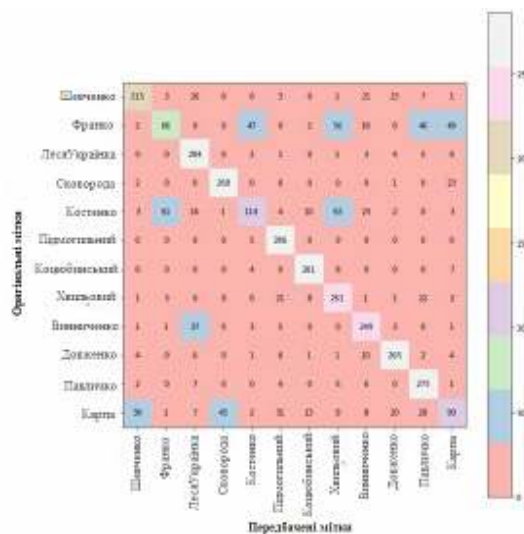


Рисунок 2 – Матриця помилок на текстах довжиною 10000 символів

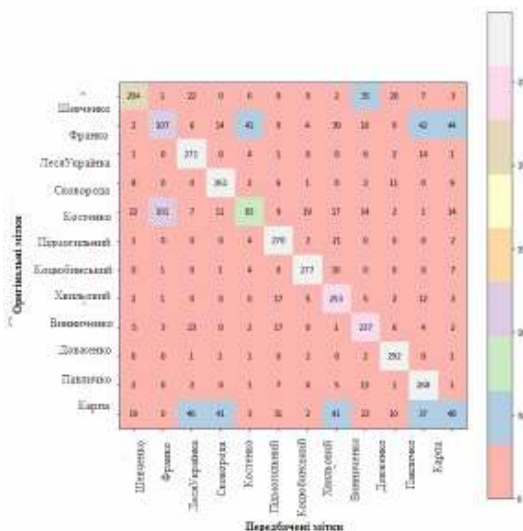


Рисунок 3 - Матриця помилок на текстах довжиною 5000 символів

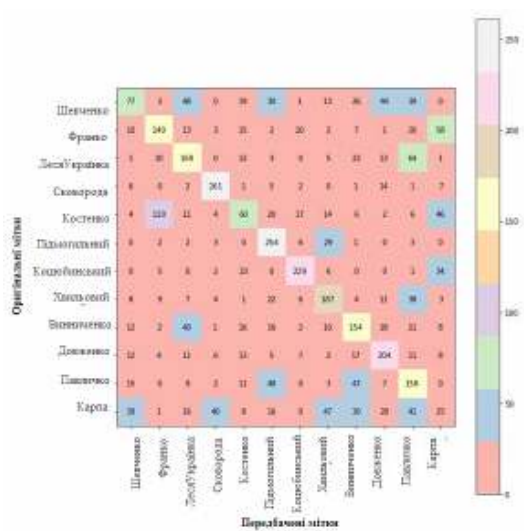


Рисунок 4 - Матриця помилок на текстах довжиною 1000 символів

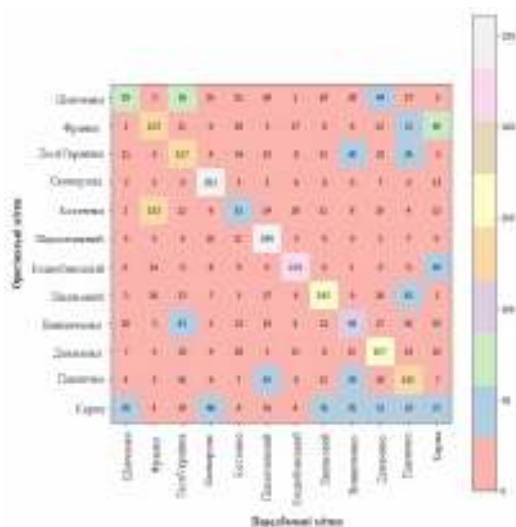


Рисунок 5 – Матриця помилок на текстах довжиною 500 символів

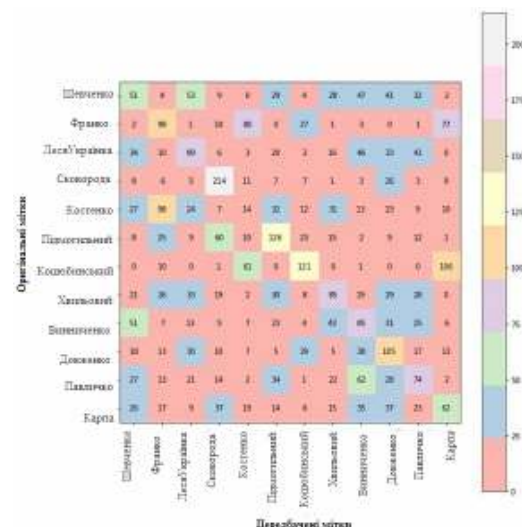


Рисунок 6 - Матриця помилок на текстах довжиною 200 символів

6. Динаміка точності класифікації. Класифікація текстів здійснювалася на основі векторів, побудованих на статистичних ознаках, вилучених із текстів, та порахованих частотах біграм. Для класифікації використовувалися різні алгоритми: логістична регресія, метод k-найближчих сусідів, метод опорних векторів з різними ядрами, градієнтний стохастичний спуск, стохастичний градієнтний спуск на векторизованих текстах методом TF-IDF. Як основна метрика використовувалася «акуратність» (ассурасу). Найкращі результати показали логістична регресія та метод опорних векторів. Для досягнення найкращої точності відбувався підбір параметрів оптимальних параметрів із використанням GridSearchCV. Результати роботи алгоритмів представлені у таблиці 1.

Таблиця 1

Точність класифікації (метрика ассурасу)

	L = 20000	L = 10000	L = 5000	L = 1000	L = 500	L = 200
Logistic Regression	0.813	0.751	0.725	0.536	0.467	0.317
KNN	0.774	0.718	0.665	0.332	0.236	0.133
SVM	0.776	0.743	0.702	0.503	0.475	0.280
Tfidf + SGDClassifier	0.709	0.721	0.733	0.606	0.534	0.332
SGDClassifier	0.753	0.721	0.664	0.492	0.395	0.255

Якість класифікації значно знижується під час навчання текстів довжини менше 5000 символів. Якість класифікації при збільшенні довжина текстів з 5000 до 20000 символів підвищується незначно. Векторизація вихідних текстів – досить трудомістке завдання. Тому за відсутності достатніх ресурсів, можна використовувати тексти найбільшої довжини, не відчуваючи у своїй особливих втрат якості.

7. Візуалізація. В результаті було отримано довгі вектори, що містять безліч ознак, обчислених на створеній вибірці. Було б цікаво подивитися на ці дані та спробувати знайти деякі закономірності між ними. Якби візуально отримані вектори поділялися на кластери, відповідні різним авторам, це означало б, що ознаки було обрано досить успішно.

У межах роботи методи основних компонентів і t-SNE застосовувалися щоб одержати наочного уявлення отриманих ознак.

Вектори, які складаються зі статистичних ознак і частот біграм, відображались на площині. Кольори крапкам присвоювались виходячи з оригінальних міток (оригінальної приналежності будь-якому автору), кожна точка позначає конкретний текст, а саме вектор, отриманий після відображення довгий вектор (ознаки тексту) в простір меншої розмірності. У зв'язку з цим можна було спостерігати одержані розподіли текстів.

Для текстів довжини 20000 було отримано графіки при зниженні розмірності до простору R^2 , які зображені на рисунках 7 - 8. Кожна точка має колір (номер), що означає її належність конкретному автору. Видно, що зниження розмірності за допомогою методу t-SNE для текстів довжини 20000 допомагає виділити різні класи авторів.

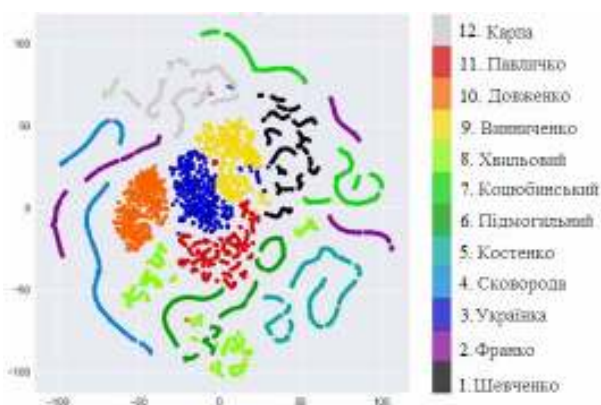


Рисунок 7 - Алгоритм t-SNE для текстів довжини 20 000 символів тренувального набору. Бібліотека Matplotlib

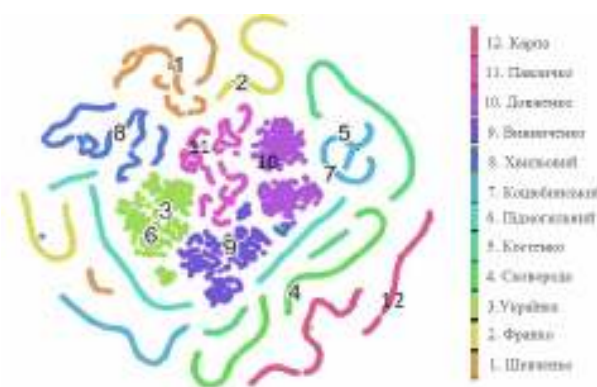


Рисунок 8 - Алгоритм t-SNE для текстів довжини 20 000 символів тренувального набору. Бібліотека Seaborn

Висновки. Вирішувалось завдання класифікації текстів відомих авторів. Крім цього досліджувалась залежність точності класифікації від довжини тексту, яка не вивчалася у роботах, що описують проблематику та перші підходи до вирішення вихідного завдання.

Для вирішення даної проблеми були поставлені і вирішені завдання, що включають:

- Підготовка тексту до аналізу;
- Дослідження тексту як послідовності символів;
- Пошук та виділення закономірностей, здатних охарактеризувати текст;
- Отримання векторного подання тексту;
- Візуальне представлення одержаних ознак;
- Побудова та навчання моделей, які здійснюють класифікацію текстів;
- Порівняння побудованих моделей для різних корпусів текстів та визначення порогових довжин текстів, при яких точність класифікації змінюється незначно чи, навпаки, зазнає спад чи підйом точності класифікації.

ЛІТЕРАТУРА

1. Полинська Г.А. Інформаційні системи маркетингу. Київ : ЮРАЙТ, 2016. 324 с.
2. Мильніков К. Статистичні методи інтелектуального аналізу даних. Україна, 2021. 240 с.
3. Шитиков В.К., Мاستицкий С.Э. Статистичний аналіз та візуалізація даних за допомогою R. Издательство «ДМК Пресс», 2015. 496с.
4. T.Hastie, R.Tibshirani, J.Friedman. The Elements of Statistical Learning. Data Mining, Inference, and Prediction. 2nd Edition. - Springer, 2013.
5. Шитиков В. К., Мастицкий С. Э. Классификация, регрессия и другие алгоритмы Data Mining с использованием R. 2017.
6. Python для анализа данных:
https://mipt-stats.gitlab.io/courses/python/09_seaborn.html

REFERENCES

1. Polynska H.A. Informatsiini systemy marketynhu. Kyiv : YuRAIT, 2016. 324 s.
2. Mylnikov K. Statystychni metody intelektualnoho analizu danykh. Ukraina, 2021. 240 s.
3. Shytykov V.K., Mastytskyi S.Э. Statystychnyi analiz ta vizualizatsiia danykh za dopomohoiu R. Yzdatelstvo «DMK Press», 2015. 496s.
4. T.Hastie, R.Tibshirani, J.Friedman. The Elements of Statistical Learning. Data Mining, Inference, and Prediction. 2nd Edition. - Springer, 2013.

5. Shytykov V.K., Mastytskyi S.Э. Klassyfykatsyia, rehressyia y druhye alhorytmi Data Mining s yspolzovanyem R. 2017.

6. Python dlia analyza danniaxh:

https://mipt-stats.gitlab.io/courses/python/09_seaborn.html

Received 11.07.2022.

Accepted 15.07.2022.

***Statistical text analysis
and study of the dynamics of classification accuracy***

The work is devoted to the statistical analysis of the text and the study of the dynamics of classification. In the work, the selection of statistical features of the text, the classification of texts belonging to different authors, and the study of the dynamics of classification accuracy depending on the length of text fragments are carried out. To solve the problem, the following methods were used: natural language processing methods; statistical characteristics of texts; machine learning methods; dimensionality reduction methods for visualization capability. On the basis of the obtained dynamics of changes in classification accuracy depending on the lengths of text fragments, appropriate conclusions were drawn regarding the optimal length of texts used for training and testing models. The task was solved in the Jupyter Notebook software environment of the Anaconda distribution, which allows you to immediately install Python and the necessary libraries.

Keywords: machine learning, statistical text analysis, authorship determination, data analysis, natural language processing.

Островська Катерина Юрїївна – к.т.н., доцент, доцент кафедри Інформаційних технологій і систем УДУНТ.

Фененко Тетяна Михайлівна- старший викладач кафедри Інформаційних технологій і систем УДУНТ.

Глущенко Олександр- магістр кафедри Інформаційних технологій і систем УДУНТ.

Ostrowska Kateryna - Ph.D., Associate Professor, Associate Professor of the Department of Information Technologies and Systems of USUST.

Fenenko Tetiana - senior teacher of the Department of Information Technologies and Systems of USUST.

Hlushchenko Oleksandr – master of the Department of Information Technologies and Systems of USUST.

A.S. Tonkoshkur, L.V. Nakashydzhe, S.F. Lyagushyn, V.R. Kolbunov

FUNCTIONING FEATURE ANALYSIS OF VOLTAGE LIMITERS BASED ON A COMBINED VARISTOR-POSISTOR STRUCTURE

Abstract. Modeling and analysis results concerning the influence of the electrical and thermal conditions on the functioning of series-connected and thermally contacted layers of varistor ceramics and a posistor composite are presented. Such structures may be useful for the development of a promising limiter of constant and slowly varying voltages. The main attention is paid to the behavior of their functional electrical and thermal characteristics with changes in the internal resistance of the input voltage source and in heat flow intensity to the environment as well as in its temperature. It is shown that a change in the internal resistance of the input voltage source can only lead to a shift towards higher input voltage values ensuring the posistor element actuation. At a low intensity of heat exchange between such a voltage limiter and the environment, it may operate at voltages lower than the classification voltage of the varistor layer, and thus the device under consideration is not capable of fixing a constant voltage at the output and can function as a self-resetting high voltage fuse. In the case of an increase in the ambient temperature (up to the temperature of the phase transition in the posistor layer material), the temperature of the structure and the output voltage of the device change insignificantly around the temperature limitation point, and the current value and dissipation power of the structure are much reduced. For structures based on varistor ceramics with a strong temperature dependence of low-voltage resistance (high activation energies of its temperature sensitivity), the output voltage decrease (to values of one volt order) takes place.

Key words: varistor ceramics; polymeric positive temperature coefficient (PPTC) nanocomposite; varistor-posistor structure, voltage limiter; voltage source internal resistance, heat transfer intensity.

1. Problem statement

Inhomogeneous dielectrics, in particular posistor nanocomposite materials and varistors, have found wide and effective use in modern technology as elements of protection devices against current and impulse overloads of various electrical systems [1-6]. A promising application of inhomogeneous dielectrics of this type is the implementation of protection of solar cells from overvoltage in their series connections (rows) in photovoltaic modules [7].

Therefore, it is very important to model and analyze the influence of the electrical and thermal conditions on the operation of voltage limiters – combined structures based on metal oxide ceramics and a polymer posistor composite used in well-known self-healing fuses of the PolySwitch technology. Hence, relevant is the study of behavior patterns of their functional electrical and thermal characteristics with a change in the internal resistance of the input voltage source, the heat transfer to the environment and its temperature.

2. Analysis of recent research and publications

The analysis of literature data showed that the use of the construction described in [8, 9] in the form of a combined two-layer structure of varistor and posistor layers in thermal contact to limit overvoltage is a promising directional protection against current and impulse overloads. The input voltage is applied to the entire structure (sequential connection of specified layers).

The varistor layer determines the output voltage of such a quadripole and is connected in parallel to the object of protection. When an input overvoltage occurs, the heat generated by the varistor layer initializes the heating of the series-connected posistor layer and causes an increase in its resistance. As a result, there is a redistribution of the input overvoltage between the layers, which can ensure the limitation of the output voltage on the varistor layer and, accordingly, on the load connected in parallel to it.

A number of works are devoted to the study of the electrical characteristics of the considered two-layer varistor-posistor structures [9, 10, 11]. At the same time, the main attention was paid to the research on the substantiation of the very principle of operation of such an overvoltage protection element and the possibility of its application in solar photovoltaic systems [7].

The tasks of developing the devices under consideration as voltage limiters (VL) both for general multi-purpose use and for use as self-resetting surge protectors in solar photovoltaic systems [7, 12] require detailed knowledge of their electrical and thermal characteristics and their behavior under various conditions.

3. The purpose of study

The purpose of the work is to determine the patterns of behavior of the functional electrical and thermal characteristics for such structures with a change in the internal resistance of the input voltage source, in the heat transfer intensity to the environment and its temperature.

4. Description of the main material of research.

4.1. Model equations

The analyzed two-layer structure, consisting of layers of varistor ceramics and a nanocomposite with a positive temperature coefficient of resistance connected in series and in direct thermal contact, can be represented by the scheme in Fig. 1.

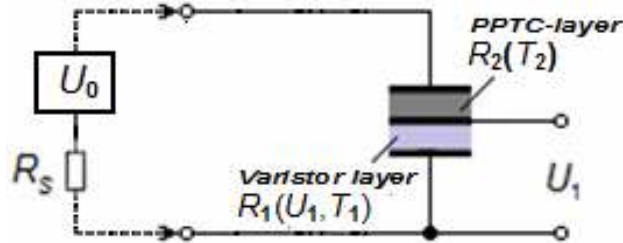


Figure 1 - The structure under study and the circuit for supplying electric voltage to it: U_0 and U_1 are the input and output (to be limited) voltage; $R_1(U_1, T_1)$, $R_2(T_2)$ and R_s are the electrical resistances of the varistor and posistor layers and the internal resistance of the input voltage source

When voltage U_0 is applied to the structure under consideration, it is initially applied to the varistor layer, since the posistor layer at temperatures below the actuation temperature $T_{tr} \approx 398$ K [13, 14] has a low resistance (close to metal). Because of this, the varistor layer begins to heat up due to the increase in the flowing current and its temperature increases. The posistor layer also heats up due to contact with the varistor layer, as well as due to the current flowing through it (self-heating). Its resistance increases sharply as it approaches the actuation temperature T_{tr} (this designation is used because of the traditional term “tripping”). The current flowing through the structure and, accordingly, through the layer of varistor ceramics, slightly decreases, and the structure temperature drops. As a result of such a transient process, the output voltage on the varistor layer and the structure temperature are stabilized.

The set of equations for determining and analyzing the steady-state stationary values of the main parameters of the considered electrothermal voltage limiter can be written in the following form:

$$P_1(T_1, T_2) = \frac{T_1 - T_0}{RT_{dis}} - \frac{T_1 - T_2}{RT_{con}}, \quad (1)$$

$$P_2(T_1, T_2) = \frac{T_2 - T_0}{RT_{dis}} - \frac{T_2 - T_1}{RT_{con}} \quad (2)$$

where $P_1(T_1, T_2) = I(T_1, T_2) \cdot U_1(T_1, T_2)$ and $P_2(T_1, T_2) = I(T_1, T_2)^2 \cdot R_2(T_2)$ – electrical (converted into thermal) power dissipation of the varistor and posistor layers;

$T_1(t)$ и $T_2(t)$ – temperatures; T_0 – ambient temperature (assumed to be 300 K);
 RT_{dis} – temperature resistances of heat dissipation into the environment, which are related to the heat transfer coefficient $RT_{dis} \approx 1 / (h_z S)$, h_z – coefficient of heat transfer per unit area of the plate surface with the environment, S – plate area [15] (taken the same for both layers of the structure); RT_{con} – temperature resistance of heat transfer of the contact between these layers.

Electrical resistance of the varistor layer, taking into account the temperature dependence of their conductivity in weak electric fields (in the ohmic part of the I–V characteristic) $G(T)$, was represented on the basis of the analytical equation given in [5] as

$$R_1(U, T) = \frac{I}{G(T) + \frac{I_c}{U} \cdot \left(\frac{U}{U_c} \right)^\beta}, \quad (3)$$

where U_c – voltage (sometimes called breakdown voltage) across the varistor layer at current $I_c = 10^{-3}$ A [1, 5]; β – non-linearity coefficient;

$G(T) = G_0 \cdot \exp(-\Delta E_g / (kTn))$; k and ΔE_g are the Boltzmann constant and the activation energy of the considered temperature dependence (which for zinc oxide varistor materials ranges from 0.4 to 0.8 eV [16]; n is the correction factor for determining the activation energy of the considered temperature dependence (determined experimentally).

To approximate the temperature dependence of the posistor electrical resistance $R_2(T)$, the equation is used [9], which is known for describing phase transitions and suitable for nanocomposite materials based on PolySwitch technology

$$R_2(T) = \begin{cases} \frac{A_1}{(T_{trip} - T)^m}, & T < \theta \\ A_2, & T \geq \theta \end{cases}, \quad (4)$$

where $A_1 = R_2(T_0) \cdot (T_{trip} - T_0)^m$; $R_2(T_0)$, A and m are values determined by the technical parameters of samples of PPTC composite of PolySwitch technology; $T_{trip} \approx 398$ K is the temperature of its operation (phase transition);

$R_2(T_0) = R_2(300 \text{ K})$ (equals to R_{min} or R_{1max} [7]) is the resistance of the fuse layer in the conducting state; R_{min} is the minimum resistance of the posistor nanocomposite layer at 23 °C; R_{1max} is the maximum layer resistance after switching off and restoration within one hour at 23 °C;

A_2 – resistance of the fuse in the insulating state (from $10^4 \Omega$ to $10^{10} \Omega$) [5]; m is the approximation parameter (from 1 to 4); $\theta \approx T_{trip}$.

Determining the magnitude of the electrical voltage $U_1(T_1, T_2)$ on the varistor layer in a circuit from a series connection of the varistor and posistor layers, to which the voltage U_0 of a power supply with an internal electrical resistance R_s is applied, is carried out by solving the following non-linear equation, which can be obtained from the well-known Kirchhoff laws,

$$U_1(T_1, T_2) = \frac{U_0 \cdot R_1[U_1(T_1, T_2), T_1]}{R_1[U_1(T_1, T_2), T_1] + R_2(T_2) + R_s}. \quad (5)$$

The solution of a non-linear equation (5) can only be numerical (in the form of a two-dimensional array). To find it, the standard built-in root function of the Mathcad package [17] can be used. To represent the function $U_1(T_1, T_2)$ in the analytical form required for solving the set (1)-(3), two-dimensional spline interpolation was used (the interp and cpline functions of the Mathcad package).

The expression for the current passing through the structure under consideration has the form:

$$I(T_1, T_2) = \frac{U_0}{R_1[U_1(T_1, T_2), T_1] + R_2(T_2) + R_s}. \quad (6)$$

The temperatures T_1 and T_2 were found by solving the set of nonlinear equations (1)-(2), which was performed using the already mentioned standard built-in root function of the Mathcad package.

4.2. Modeling the influence of the features of the electrical and thermal conditions on the operation of a varistor-posistor voltage limiter

4.2.1. Influence of the internal resistance of the input circuit

The influence of the internal resistance of the input voltage source R_s on the steady state values of output voltage U_1 , current I , and power dissipation of the varistor layer P_1 after the actuation of the analyzed VL is shown in Fig. 2.

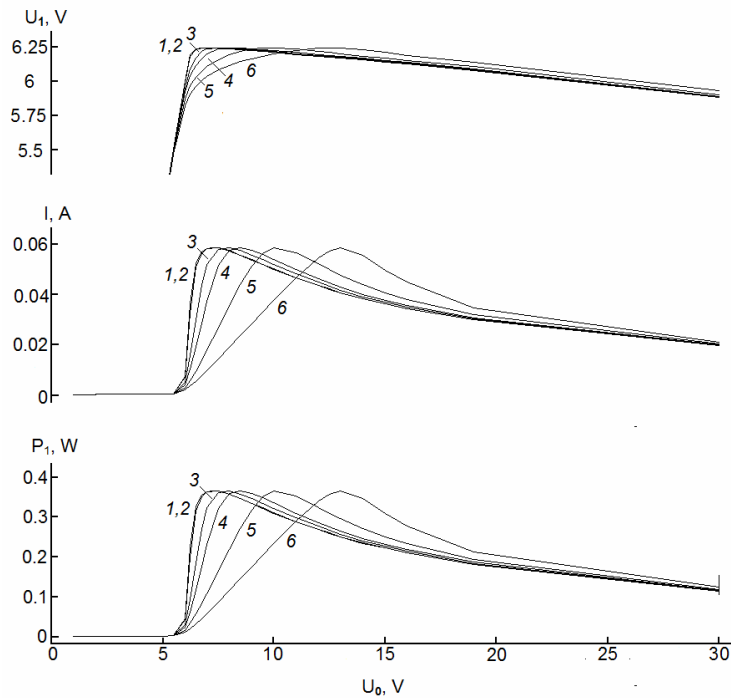


Figure 2 - Anisient $U_1(U_0)$ and input $I(U_0)$ characteristics, as well as dependences of the power dissipation of the varistor layer $P_1(U_0)$ of the varistor-positistor structure on the input voltage when using voltage sources with internal resistances R_s (Ω):

$$1 - 0.1; 2 - 1; 3 - 10; 4 - 20; 5 - 50; 6 - 100$$

In numerical experiments, the typical values of the quantities $R_s=1 \Omega$; $RT_{con}=1.15 \text{ K}\cdot\text{W}^{-1}$; $h_z=0.001 \text{ W}\cdot\text{K}^{-1}\cdot\text{cm}^{-2}$ ($RT_{dis}=318 \text{ K}\cdot\text{W}^{-1}$) and of the physical parameters of the component materials of the structure under study, oriented for use in protection circuits of solar arrays, are accepted. For the varistor layer, the parameters of its working element – a layer assembled from several dozen parallel-connected commercial frameless low-voltage varistors V3.5MLA0603NH ($U_C = 3.7-7 \text{ V}$, $\beta = 30-50$) [18] and for the positistor layer, that of a self-restoring fuse of the FRX110 60F type (element resistance in conductive state $R_2(T_0) \sim 0,3 \Omega$; $T_{trip}=125 \text{ }^\circ\text{C}$) [19] were used.

As can be seen, the dependences $U_1(U_0)$ and $I(U_0)$ presented in Fig. 2 correspond to the known ones [8, 9]: with an increase in the input voltage above the classification voltage of the varistor layer U_C , the output voltage is almost constant and the current and the dissipation power values, respectively, decrease.

These results are in full agreement with the existing ideas about the mechanism of electrothermal voltage limitation by the structure under consideration, which can be described as follows. When an input overvoltage is applied to the structure, the heat generated by the varistor layer heats the series-

connected PTC layer and leads to an increase in its resistance. As a result, there is a redistribution of the input overvoltage between the layers, which ultimately can provide a limitation at a given level of the output voltage on the varistor layer and, naturally, on the load connected in parallel to it.

A change in the internal resistance of the input voltage source can only lead to a shift towards higher values of the input voltage that ensures the operation of the posistor element. This does not contradict the ideas about the allocation of some part of the electric power by the internal resistance of the source, that is, the achievement of the actuation temperature T_{trip} takes place at a higher input voltage (Fig. 3).

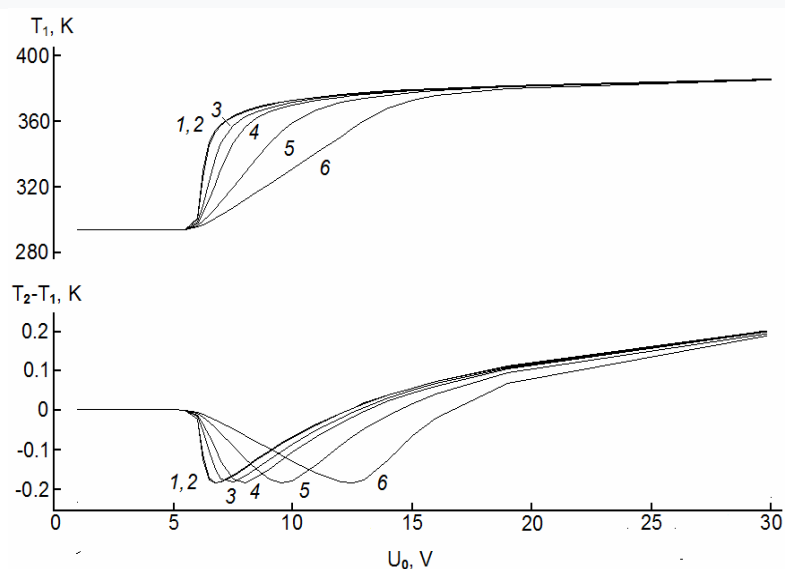


Figure 3 -Dependence of the temperature of the varistor layer T_1 and the temperature difference between the varistor and posistor layers ($T_1 - T_2$) after the VL operation on the applied input voltage for voltage sources with different internal resistances R_s (Ω): 1 - 0.1; 2 - 1; 3 - 10; 4 - 20; 5 - 50; 6 - 100

It should also be noted that at small U_0 the heating temperature of the varistor layer is higher than the temperature of the posistor one ($T_1 - T_2 > 0$), and for large U_0 , the opposite pattern takes place ($T_1 - T_2 < 0$), which is the result of self-heating of the posistor layer by the passing current.

4.2.2 Influence of the intensity of heat exchange with the environment

The intensity of heat exchange with the environment is one of the essential factors that determine the temporal changes in the temperature of the considered two-layer varistor-posistor structure and, accordingly, its functional electrical characteristics. Such process can be regulated within certain limits, for example, by using special metal radiators; this allows to consider it as a regime process.

To simulate the effect of this heat transfer, the coefficient of heat exchange per unit surface of the outer electrode layers of the structure with the environment (h_z) can be used, which determines the thermal resistance of dissipation to the environment (RT_{dis}) in a specific mode of its functioning.

Presented in Fig. 4 data indicate a significant effect of this parameter on the transient response and the dependences of the dissipated electrical power and the temperature of the varistor layer in the active state (after actuation) for the device under consideration.

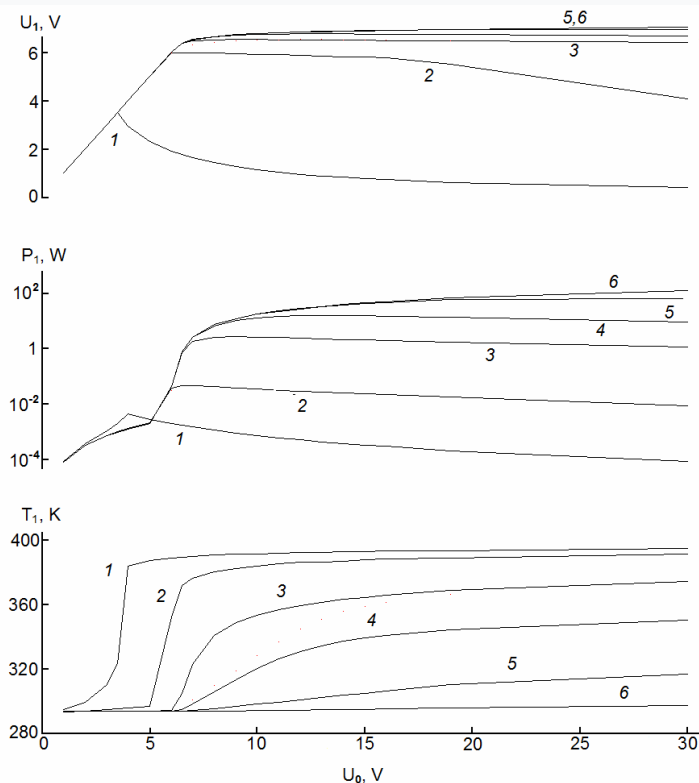


Figure 4 - Dependence of output voltage U_1 , power dissipation P_1 and temperature of varistor layer T_1 after VL actuation on input U_0 at values of specific heat transfer coefficient h_z ($W \cdot K^{-1} \cdot cm^{-2}$): 1 – 0.00001; 2 – 0.0001; 3 – 0.001; 4 – 0.01; 5 – 1; 6 – 10

At low values of the heat transfer coefficient h_z , the VL actuation (at $T = T_{tr}$) takes place at relatively low ($U_0 < U_c$) input voltages U_0 (curves 1 and 2 in Fig. 4). In this case, the output voltage U_1 and the electrical power dissipated by the varistor layer are significantly reduced. It follows that the device under consideration in this mode is not capable of fixing a constant voltage at the output and can only function as a self-resetting high voltage fuse.

At large values of h_z , a significant electric power is required for the corresponding heating of the entire structure (curves 5 and 6 in Fig. 4) and,

therefore, in the stationary state, large currents must flow through the structure, which will eventually lead to the degradation of the varistor layer [20] and the destruction of the voltage limiter. Regarding the accepted physical and technical parameters of the structure, the optimal value of the parameter h_z is within $0.001 - 0.01 \text{ W}\cdot\text{K}^{-1}\cdot\text{cm}^{-2}$.

4.2.3. Influence of ambient temperature

In the technical application of the studied combined electrothermal devices, detailed information is needed on their behavior under various conditions, in particular, in various temperature regimes. One of the urgent tasks is to study the influence of environmental factors on their electrical characteristics, first of all, such factor as ambient temperature.

The calculated dependences of the output voltage, current, and temperature of the structure under study on the applied input voltage in the open circuit mode are shown in Fig. 5.

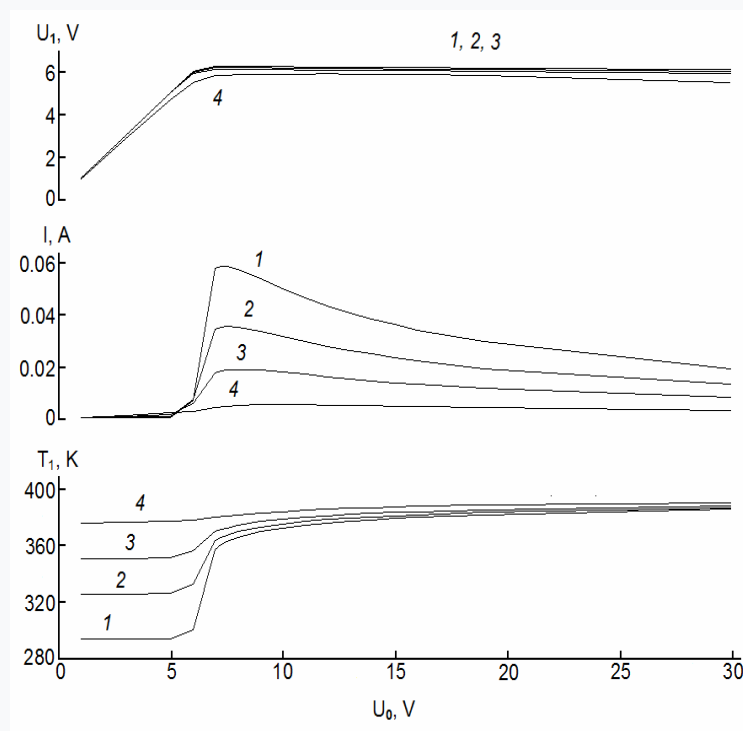


Figure 5 - Dependence of the output voltage U_1 , current and temperature of the varistor layer T_1 after VL actuation on the input U_0 at the ambient temperature values T_0 (K): 1 – 293; 2 – 325; 3 – 350; 4 – 375

The presented patterns are in satisfactory agreement with the known data of the experimental studies [11].

The effect of ambient temperature on the absolute value of the limited output voltage is determined by a negligible thermal resistance factor.

An increase in the ambient temperature T_0 leads to a slight increase in the temperature of the posistor layer of the structure in the output voltage range where the process of its limitation is implemented.

Input characteristics correspond to typical volt-ampere characteristics of posistors [21, 22]. It should be noted that the main function of the varistor – the limitation of the overvoltage pulse – takes only a short period of time at the beginning of its action when the varistor element heats the structure to the temperature of the phase transition of the posistor element, and then it performs only the function of a heater. After switching the posistor element to the low-conducting state, upon reaching T_{tr} , the current of the structure decreases to a value sufficient to keep it in this state [23]. According to the data in Fig. 4, an increase in the ambient temperature T_0 causes a significant decrease in the magnitude of the current I since such increase provides additional heating of the posistor element.

Accordingly, as the voltage U_0 increases, the current I flow through the combined structure under study increases and passes through its maximum (when the temperature of the structure approaches T_{trip} corresponding to the previously indicated phase transition). Further, this current decreases due to an increase in the resistance of the nanocomposite posistor layer.

Figure 6 shows the corresponding dependences of electrical power dissipation by layers of varistor ceramics and posistor polymer nanocomposite/

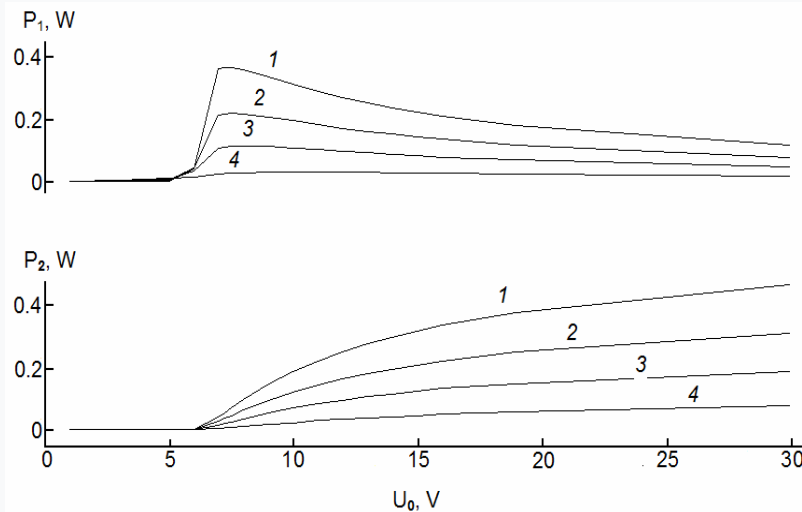


Figure 6 - Dependence of the electrical powers dissipated by the varistor P_1 and posistor P_2 layers of the structure under study after the VL actuation on the input voltage U_0 at the ambient temperature values T_0 (K):

1 – 293; 2 – 325; 3 – 350; 4 – 375

After the transition of the posistor layer to a high-resistance state at T_{trip} , there is a redistribution of the input voltage and dissipated electrical power towards its posistor component increasing.

An increase in the ambient temperature leads to a decrease in the additional electric power required for the posistor layer to reach the phase transition temperature.

In conclusion, it should be noted that the current-voltage characteristics (CVC) of varistor, in particular, metal oxide ceramic materials, have sections that differ somewhat in electrical and thermal properties [24]. At low voltage values, there is a temperature-sensitive linear (ohmic) dependence of the current on voltage in the form of Ohm's law, and the electrical resistance of the varistor has a temperature activation dependence typical of a semiconductor. With a further increase in voltage, the current increases nonlinearly with increasing voltage, and there are sections of weak and varistor nonlinearity of the I–V characteristic, where there is practically no dependence on temperature. At very high voltages, the volt-ampere characteristic is also linear and has a temperature dependence of resistance with low activation energies (<0.05 eV). As a rule, this section of the CVC of varistors is not technically used.

Depending on the type of varistor material, as the temperature rises, its ohmic section and the section of intermediate fields can significantly move to the region of high currents. The activation energy of electrical conductivity in weak fields for different types of ceramics is 0.4–1 eV [16]. This feature of the I–V characteristics of the varistor layer, which is largely dependent on the activation energy of the low-voltage electrical conductivity of metal oxide ceramics, can lead to the effect of transformation of the transient characteristics of the considered LV, like those described in the previous subsection.

The results of the analysis of the effect of the temperature sensitivity of the varistor layer on the characteristics of the considered voltage limiting device within the framework of the adopted model of its CVC are presented below (Fig. 7).

As can be seen from this figure, when using a varistor layer with a significant temperature dependence of ohmic conductivity, a significant decrease in the output (limited voltage) can take place.

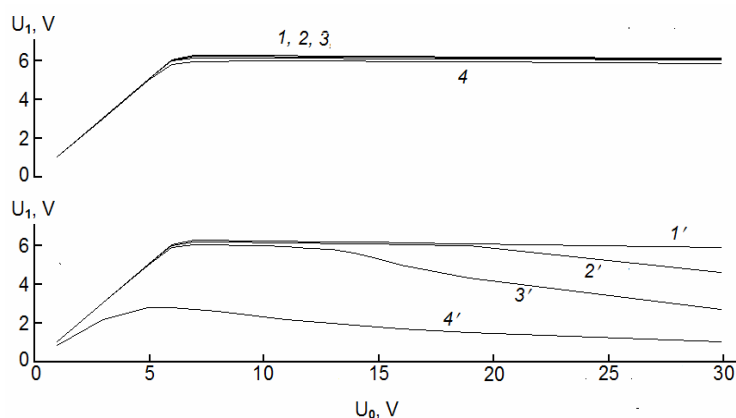


Figure 7 - Dependences of the output voltage U_1 on the input U_0 after actuation for VL based on varistor layers with conduction activation energies in the low-voltage (ohmic) section of their CVC: $\Delta E = 0$ eV (1, 2, 3, 4) and $\Delta E = 0.4$ eV (1', 2', 3', 4').

Other designations correspond to those given in Fig. 6

5. Conclusions

The presented results of modeling and analysis of the influence of the features of the electrical and thermal conditions on the operation of a voltage limiter based on a combined structure of layers of metal oxide varistor ceramics and a posistor polymer nanocomposite made it possible to establish the following regularities.

A change in the internal resistance of the input voltage source may result in a shift towards higher values of the input voltage ensuring the posistor element actuation that corresponds to the loss of some part of the electric power spent for heating the structure on the internal resistance of the source.

Low levels of heat exchange intensity between the considered VL and environment cause it to actuate at voltages lower than the classification voltage of the varistor layer, and thus such a device is not capable of fixing a constant voltage at the output but can only function as a self-resetting high voltage fuse. At high intensity levels, significant currents are required for the appropriate heating of the entire structure, which can lead to the destruction of the VL itself.

If the ambient temperature rises, the structure temperature and the output voltage of the device change insignificantly in the section of the temperature limitation, and the current magnitude and the dissipation power of the structure are substantially reduced. For structures based on varistor ceramics with a strong dependence of low-voltage resistance on temperature (high activation energies of its temperature sensitivity), the output voltage decreases (to values of the order of one volt).

REFERENCES

1. T.K. Gupta, Application of zinc oxide varistors, *J. Am. Ceram. Soc.* 73, 1817 (1990).
2. Li, S., Li, J., Liu, W., Lin, J., He, J., & Cheng, P. (2015). Advances in ZnO varistors in China during the past 30 years—fundamentals, processing, and applications. *IEEE Electrical Insulation Magazine*, 31(4), 35-44.
3. R. B. Standler, *Protection of Electronic Circuits from Overvoltages* (Dover Publications, INC. Mineola, New York, 2002).
4. W. Gretzke, Use of Polyswitch PPTC Protection in Automotive Applications, eds. H. Wallentowitz and C. Amsel, *42 V-Power-Nets* (Springer, Berlin, Heidelberg, 2003), pp. 75–87.
5. S. Cheng, K. Tom and M. Pecht, Failure precursors for polymer resettable fuses, *IEEE Trans. Dev. Mater. Reliab.* 10, 374 (2010)
6. Toth, J., & Protection, R.C. (2020). PolySwitch PPTC device principals of operation
7. Tonkoshkur A.S., Ivanchenko A.V., Nakashydzhe L.V., Lyashkov A.Yu., Gomilko I.V. Application of polymer posistor nanocomposites in systems for protecting photovoltaic components of solar arrays from electrical overloads. Monograph. – Primedia eLaunch, Boston, USA, 2021. – 172 p. DOI: 10.46299/978-1-63972-054-5
8. B. Golubovic, P. N. Becker and R. P. Moore, Circuit protection device having thermally coupled MOV overvoltage element and PPTC overcurrent element, U.S. Patent 7,660,096 B2 (2010).
9. Tonkoshkur A. S. and Ivanchenko A. V. Electrical properties of structures based on varistor ceramics and polymer nanocomposites with carbon filler. *Journal of Advanced Dielectrics*. Vol. 09, No. 03, 1950023 (2019)
10. Hu, H., Shi, G., Wang, Q., Dai, T., & Xia, H. SPICE Model of Polyswitch Device. http://hanbinhu.github.io/data/report/2014_PPTC_TE_Connectivity.pdf
11. Іванченко О.В., Тонкошкур О.С. Вплив температури навколишнього середовища на електричні властивості варисторно-позисторної структури. *Технологія та конструювання в електронній апаратурі*, 2022, № 1–3, с. 3–7. <http://dx.doi.org/10.15222/ТКЕА2022.1-3.03>
12. Ivanchenko A.V., Tonkoshkur A.S., Mazurik S.V. Application of varistor-posistor structure for protection from overvoltages of photovoltaic cells of solar arrays. *Journal of Physics and Electronics*. -2019. - No 27(1). - P. 79-88.
13. Tyco Electronics, Application note, PolySwitch strap devices help protect rechargeable battery packs (2008), <http://www.digikey.jp/Web Export/Supplier Content/Tyco 8004/PDF/TE Strap Device. pdf>.
14. Littelfuse, PolySwitchr PTC devices, Overcurrent protection device, Product: RXEF050 (2016), https://www.littelfuse.com/_/media/electronics/product specifications/resettable ptcs/littelfuse ptc rxef050 product specification.pdf.pdf.

15. Kutateladze S.S. Osnovyi teorii teploobmena, 5-e izd. Moscow, Atomizdat, 1979, 416p.
16. Глот О.Б. Окисні керамічні варистори. Навчальний посібник / О.Б. Глот, О.І. Ивон, О. С. Тонкошкур, І. М. Черненко. Дніпропетровськ: ДДУ, 1998. - 140 с
17. Maxfield B. Engineering with Mathcad: using Mathcad to create and organize your engineering calculations. – Elsevier, 2006.- 521 p.
18. Metal-Oxide Varistors (MOVs) Surface Mount Multilayer Varistors (MLVs) > MLA Series: [Electronic resource]: Littelfuse, Inc. - Mode access: https://m.littelfuse.com/~/media/electronics/datasheets/varistors/littelfuse_varistor_mla_datasheet.pdf.pdf
19. Fuzetec: Radial Leaded PTC Resettable Fuse: FRX Series: [Electronic resource]: Fuzetec Technology Co., LTD. –Mode access: <http://www.fuzetec.com/upload/download/App%20FRX%20Series%20tVer.61.pdf>
20. Тонкошкур А. С. Моделирование деградации металлоксидных варисторных элементов защиты электрических цепей: [монографія] / А. С. Тонкошкур, А.В. Иванченко – Днепр: ТОВ „АКЦЕНТ ПП”, 2019. – 157 с.
21. Шефтель И. Т. Терморезисторы. Москва, Наука, 1973, 415 с.
22. Thermistor types – their workings and applications. <https://www.elprocus.com/introduction-to-thermistor-types-with-its-workings-and-applications/> (Дата звернення: 20 февраля 2022)
23. FRX Series - Radial Leaded PTC. https://www.fuzetec.com/products_2.php?bgid=1&gid=31 (Дата звернення: 20 февраля 2022)
24. Clarke, D. R. (1999). Varistor ceramics. Journal of the American Ceramic Society, 82(3), 485-502

Received 25.07.2022.

Accepted 29.07.2022.

Аналіз особливостей функціонування обмежувачів напруги на основі комбінованої варисторно-позисторної структури

У роботі наведено результати моделювання та аналізу впливу особливостей електричного та теплового режимів на функціонування структури з послідовно з'єднаних та термічно контактованих шарів варисторної кераміки та позисторного композиту, що становлять інтерес для розробки перспективного обмежувача постійних і повільно змінних напруг. Основна увага приділена дослідженню закономірностей поведінки функціональних електричних і теплових характеристик таких структур при зміні внутрішнього опору джерела вхідної напруги та інтенсивності теплообміну з навколишнім середовищем, а також його температури.

Показано, що зміна внутрішнього опору джерела вхідної напруги може призвести лише до зсуву в бік вищих значень вказаної напруги, що забезпечує спрацювання позисторного елемента.

При низькій інтенсивності теплообміну між таким обмежувачем напруги та навколишнім середовищем він може працювати при напругах, нижчих за класифікаційну напругу варисторного шару, і, таким чином, розглянутий пристрій не здатний фіксувати постійну напругу на виході і може функціонувати тільки як самовідновлюваний запобіжник від перенапруги.. При підвищенні температури навколишнього середовища (до температури фазового переходу в матеріалі шару позистора) температура структури та вихідна напруга приладу змінюються несуттєво навколо точки обмеження температури, а значення струму і потужність розсіювання структури значно зменшуються. Для структур на основі варисторної кераміки з сильною залежністю низьковольтного опору від температури (високі енергії активації його температурної залежності) має місце зниження вихідної напруги (до значень порядку одного вольт).

Тонкошкур Олександр Сергійович – д.фіз.-мат.н., професор, професор кафедри електронних обчислювальних машин, Дніпровський національний університет ім.Олеся Гончара.

Накашидзе Лілія Валентинівна – д.т.н., ст.науковий співробітник, провідний науковий співробітник НДІ енергоефективних технологій і матеріалознавства, Дніпровський національний університет ім.Олеся Гончара.

Лягушин Сергій Федорович –к.фіз.-мат.н., доцент, доцент кафедри теоретичної фізики, Дніпровський національний університет ім.Олеся Гончара.

Колбунов Вадим Радиславович – к.фіз.-мат.н., доцент, доцент кафедри прикладної радіофізики, електроніки та наноматеріалів, Дніпровський національний університет ім.Олеся Гончара.

Tonkoshkur Alexander - doctor of physical and mathematical sciences, professor, Professor of the Department of Electronic Computing, Oles HoncharDnipro National University.

Nakashydzhe Liliya - Doctor of Technical Sciences, senior researcher, Leading researcher of the Research Institute of Energy-Efficient Technologies and Materials Science, Oles HoncharDnipro National University.

Lyagushyn Serhiy - candidate of physical and mathematical sciences, associate professor, Associate Professor of the Department of Theoretical Physics, Oles HoncharDnipro National University.

Kolbunov Vadim - candidate of physical and mathematical sciences, associate professor, Associate Professor of the Department of Applied Radiophysics, Electronics and Nanomaterials, Oles HoncharDnipro National University.

ІНФОРМАЦІЙНА СИСТЕМА ПРОВЕДЕННЯ ПРОЦЕСУ LIVENESS DETECTION ЗА ДОПОМОГОЮ AWS

Анотація. У роботі досліджено процес проведення відеоідентифікації особи за допомогою процесу Liveness Detection. Розроблено алгоритм проведення процесу Liveness Detection на основі виконання випадкових завдань особою, що ідентифікується. Проведено дослідження API Amazon DetectFaces, у процесі якого, емпіричним шляхом, були встановлені порогові значення характеристик обличчя на зображенні для проведення процесу Liveness Detection. На основі проведених досліджень розроблена програмна система біометричної відеоідентифікації людини, яка дозволяє автоматично встановлювати присутність особи перед відеокамерою і ідентифікувати її, запобігаючи використанню фальшивих артефактів з метою підробки особистості людини.

Ключові слова: Liveness Detection, відеоідентифікація, Amazon Web Services, розпізнавання обличчя.

Постановка проблеми. Дистанційне проведення процесу відеоідентифікації стає все більш актуальним у нашій країні. З одного боку, поширення у світі вірусу COVID-19 викликало різку зміну патернів споживчої поведінки. Зростає частка тих споживачів, які свідомо чи вимушено відмовляються від фізичного контакту з оточуючими. З іншого боку, велика кількість людей стикається з різними проблемами у зв'язку з переселенням та релокацією підприємств із територій, де ведуться активні бойові дії. Наприклад, людина, яка бажає відкрити особистий рахунок чи взяти кредит, не завжди може знайти відділення банку, де операціоніст зможе перевірити схожість людини з її документами. Або громадянин, який звертається для надання деяких електронних державних послуг, наразі, на певних етапах виконання його запиту, повинен бути особисто присутнім у відповідних державних органах, щоб підтвердити свою особистість.

Одним з шляхів вирішення проблеми дистанційної відеоідентифікації особи є процес Liveness detection [1], який дозволяє відрізнити реальну людину від штучної репродукції її зовнішності. При цьому підходить система, яка вико-

ристовує Liveness detection, повинна взаємодіяти із особою таким чином, щоб значно утруднити підробку її особистості хакером або шахраєм.

Робота має високу актуальність і практичну значимість, через що було прийнято рішення про розробку власного алгоритму проведення процесу Liveness detection, який дозволить дистанційно проаналізувати і оцінити присутність особи, що ідентифікується, та побудувати на його основі програмну систему відеоідентифікації співробітників підприємства, відкритість коду якої дозволить вільно, за потребою, повторно використовувати або змінювати її компоненти при реалізації процесу Liveness detection у будь-яких системах відеоідентифікації людини.

Аналіз останніх досліджень і публікацій. Загальний розвиток комп'ютерних технологій і вдосконалення алгоритмів роботи штучного інтелекту, протягом останніх кількох десятиріч, дозволили розробити безліч алгоритмів розпізнавання обличчя, які умовно можна поділити на три типи, залежно від того, як вони ідентифікують обличчя [2]: на основі зовнішнього вигляду (функції), який використовує повні текстурні функції; розпізнавання обличчя на основі підпростору; прийоми використання нейронних мереж. Конкретні рішення для проведення процесу Liveness detection є закритими для вільного використання та представлені каталогом комерційного API, які базуються на машинному навчанні та використовують нейронні мережі для виконання завдань, такі, як платформа ідентифікації Acuant [3], Kairos Face Recognition [4], API Anometrics Face Recognition [5], API Lambda Labs [6]. Проведений аналіз сучасного ринку програмних і алгоритмічних рішень для виконання процесу Liveness detection показав, що існуючі наразі підходи є цілком комерційними рішеннями із закритими алгоритмами своєї роботи. Окрім цього, алгоритм Liveness detection ще не є стандартизованим і має безліч варіантів реалізації, які іноді можуть приводити до неоднозначних результатів [7]. Саме тому, було прийнято рішення про розробку власного алгоритму та системи Liveness detection на основі отримання характеристик обличчя за допомогою AWS API [8], оскільки цей сервіс пропонує високу точність розпізнавання обличчя, яка складає 99.99% і щомісячно надає безкоштовні 10000 запитів для використання, чого достатньо для виконання мети нашої роботи.

Мета дослідження: розробка алгоритму проведення процесу Liveness Detection за допомогою AWS та побудова системи відеоідентифікації на його основі.

Основна частина: Найпершим кроком нашої роботи, що має бути виконаний для розробки системи відеоідентифікації особи, є розробка алгоритму процесу Liveness detection. Його головною метою є конкретизація шагів комунікації програмної системи із користувачем, з метою отримання висновку щодо фізичної присутності особи, яка проходить відеоідентифікацію, перед камерою. Розглянемо створений алгоритм, зображений на рис. 1. більш детально.

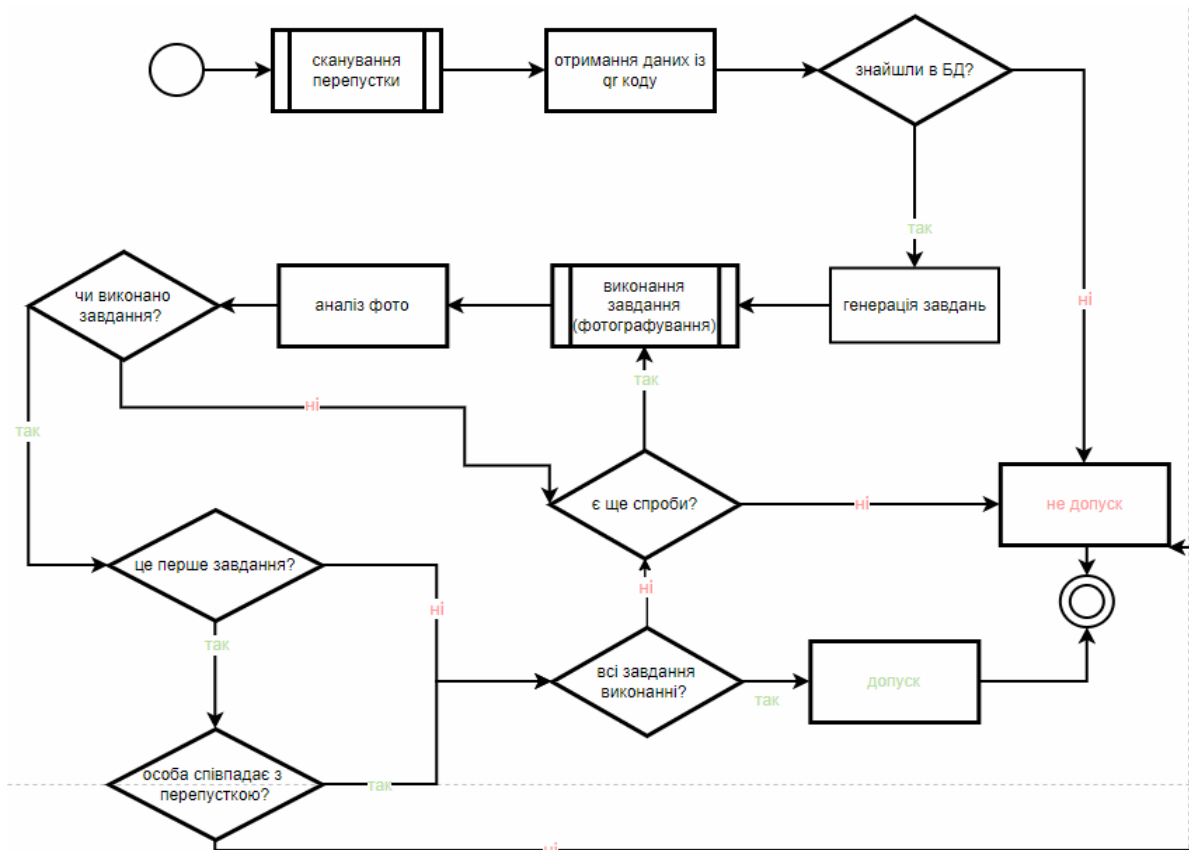


Рисунок 1 - Алгоритм виконання процесу Liveness Detection

Найпершим кроком розробленого алгоритму є отримання еталонного зображення для ідентифікації особи. Таким зображенням може бути фото з документів користувача, або, як зроблено у нашій системі - ідентифікатор перепустки, зашифрований у вигляді QR-коду, який генерується заздалегідь та містить в собі логін співробітника, до якого прив'язана його еталона центральна фотографія.

Відповідно до цього, другим та третім кроком є розшифровка QR-коду з перепустки та пошук розшифрованої інформації у базі даних системи. Якщо працівника за отриманим логіном не знайдено, це значить, що перепустка ви-

явилася підробленою та алгоритм не допускає такого співробітника, завершаючи таким чином процес відеоідентифікації.

Якщо дані перепустки були знайдені, на четвертому кроці алгоритму виконується генерація трьох завдань, націлених на отримання положень голови особи, яка ідентифікується. Першим завданням є отримання центральної фотографії особи, два інших обираються випадково з масиву заздалегідь доступних варіантів завдань.

П'ятим та шостим кроком є безпосереднє фотографування особи, з подальшим аналізом фотографії на положення голови та його відповідності поточному завданню.

На сьомому кроці відбувається перевірка, чи виконане поточне завдання. Якщо ні, то кількість доступних спроб зменшується на одну. Якщо у працівника ще є доступні спроби, то процес переходить до п'ятого кроку, тобто на процес фотографування. Якщо завдання було виконано, система перевіряє, чи співпадає особа на фотографії та на перепустці. Якщо відсоток співпадіння занадто малий, то процес буде завершено з не допуском до підприємства. Якщо особа пройшла перевірку, то надається змога виконати останні два завдання на положення голови, після чого система повертається на п'ятий крок.

Наступним кроком нашої роботи є дослідження можливості використання просторових характеристик обличчя для аналізу положення голови від API AWS, оскільки сам по собі API AWS не надає функціоналу для проведення Liveness detection. З безлічі можливих характеристик положення голови та обличчя на фотографії були обрані наступні: *isHead*, який відповідає за наявність та кількість осіб на фотографії, *brightness*, який відповідає за яскравість фотографії; *distanceHead*, який відповідає за відстань особи до веб-камери; *sunglasses*, який відповідає за наявність сонцезахисних окулярів на обличчі; *posX*, відповідальний за розташування особи щодо осі координат *OX*; *posY*, відповідальний за розташування особи щодо осі координат *OY*; *sharpness*, відповідальний за якість зображення, а саме за його чіткість; *posYaw*, відповідальний за поворот голови по сторонах; *posPitch*, який відповідає за нахил голови вгору і вниз; *posRoll*, відповідальний за нахил верхівки по сторонах; *eyesOpen*, який відповідає за стан очей, а саме, чи є вони відкритими; *mouthClosed*, який відповідає за стан рота, а саме, чи закритий він.

Для того, щоб зрозуміти, чи є дане положення голови користувача вірним, було виконане дослідження підібраних лімітів для кожного параметра окремо, емпіричним шляхом. Після опрацювання кожного з параметрів було виділено

два типи граничних значень: значення, які підходять для кожного з п'яти завдань і значення, які унікальні для певного типу завдання. У першу категорію відносяться параметри, що відповідають за кількість осіб на фотографії, освітлення і чіткість зображення. Значення параметра *isHead* розраховується за розміром масиву, що містить аналіз кожного розпізнаного обличчя сервісом AWS. Якщо розмір масиву дорівнює одиниці, це означає, що у кадрі виявлено лише одне обличчя, що саме і задовольняє систему. Параметр, що відповідає за яскравість, може набувати значення від 0 до 100 (включно), де більш високе значення вказує на яскравіше зображення обличчя. Проведені експерименти показали допустимий інтервал значень від 50 до 97. Для параметра, що відповідає за чіткість зображення, був підібраний інтервал від 40 до 100 включно. Інші параметри та їх допустимі значення є унікальними для кожного завдання.

Параметр, що відповідає за нахил верхівки може приймати негативне значення, що означає, що нахил виконаний в ліву сторону. І навпаки, при нахилі праворуч значення буде позитивним. Таким чином, для виконання завдання на центральне положення голови допустимими вважаються значення, які знаходяться в діапазоні від -5 до 5. Для решти завдань діапазон був збільшений, оскільки, при повороті всієї голови, наприклад, вліво, сервіс AWS дає похибку у значенні нахилу верхівки. Таким чином, в діапазон включені числа від -20 до 20.

Для того, щоб визначити нахил голови вгору або вниз, необхідно проаналізувати значення *possPitch*. Рух голови для даного параметра показано на рис. 2 вісь *pitch*. Дані значення також можуть бути негативними, що означає, що голова опущена вниз, а якщо значення більше за нуль, то голова піднята вгору. Усі підібрані діапазони прийнятних значень для завдань вказані в таблиці 1.

Таблиця 1

Діапазон підібраних значень параметра *possPitch* для виконання завдань

Тип завдання	Мінімальне значення	Максимальне значення
центральне фото	0	18
подивіться праворуч	-5	18
подивіться ліворуч	-5	18
подивіться вгору	20	50
подивіться вниз	-50	-20

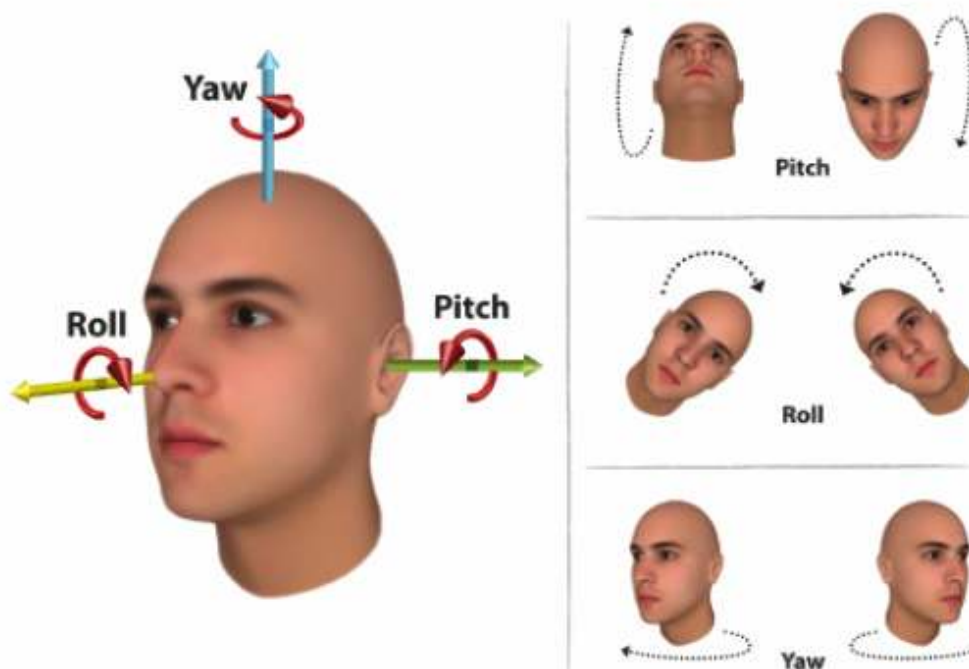


Рисунок 2 - Рухи голови по осях (pose) AWS API

Параметр `poseYaw` відповідає за поворот голови по сторонах. Рух голови для даного параметра показано на рис. 2, вісь `Yaw`. При повороті голови в ліворуч, значення будуть негативними, при повороті праворуч - позитивними. Усі підібрані діапазони прийнятних значень для завдань вказані в таблиці 2.

Таблиця 2

Діапазон значень параметра `poseYaw` для виконання завдань

Тип завдання	Мінімальне значення	Максимальне значення
центральне фото	-5	5
подивіться праворуч	10	50
подивіться ліворуч	-50	-10
подивіться вгору	-5	5
подивіться вниз	-5	5

Параметри, що відповідають за стан очей, рота та наявність сонцезахисних окулярів містять у собі, крім бульового значення, ще й коефіцієнт упевненості даної відповіді. Наприклад, при аналізі фотографії, де на обличчі знаходяться сонцезахисні окуляри, ми отримали значення параметра `sunglasses`,

який дорівнює *true*, але відсоток точності дорівнює 50%. У такому разі ми не можемо стверджувати, що окуляри дійсно присутні на фотографії. Однак, що стосується стану очей і рота, то нам не важливо в якому вони стані, якщо виконується будь-яке завдання на положення голови, крім центрального. Розглянемо варіанти значень кожного параметра докладніше. Якщо очі відкриті або рот закритий, що задовольняє умові виконання центрального завдання, то у відповіді від сервісу AWS ми отримуємо *true*, проте допустимий відсоток впевненості повинен бути не менше 60. Для інших типів завдань неважливо в якому стані очі і рот на фотографіях, тому що вони все одно не будуть брати участь у порівнянні з даними з бази даних і не зможуть вплинути на результат. Для параметра, що відповідає за наявність сонцезахисних окулярів, прийнятним значенням є *false* із впевненістю у 70 відсотків.

Для того, щоб перевірити правильність розташування обличчя в кадрі і відстань обличчя до веб-камери необхідно більш детально проаналізувати параметр *boundingBox*. Даний параметр відповідає за рамку, що обмежує обличчя навколо, де лівий верхній кут зображення - це початок координат (0,0), а верхнє (top) і ліве (left) значення є відношеннями до загального розміру зображення. Наприклад, якщо вхідне зображення має розмір 700x200 пікселів, а розмір рамки становить 350x50 пікселів, то AWS API повертає ліве значення (left) 0,5 (350/700) і верхнє значення (top) 0,25 (50/200). На рис. 3 зображено розташування рамки по відношенню до всього зображення.

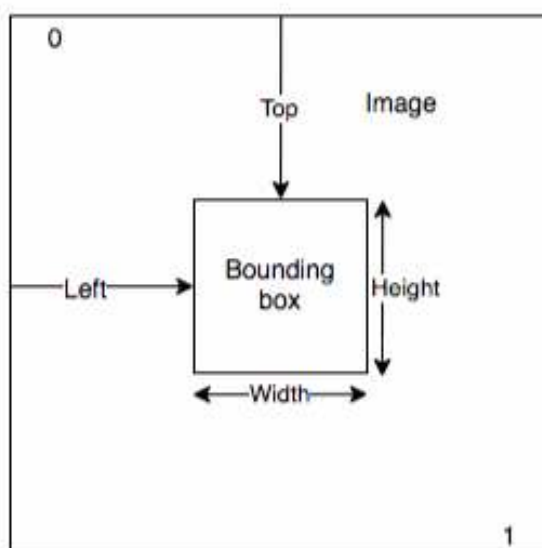


Рисунок 3 - Розташування рамки по відношенню до зображення

Таким чином, для розрахунку відстані від обличчя до веб-камери використовуються значення ширини та висоти рамки, що обмежує обличчя. Допустимі значення ширини лежать у діапазоні від 0.15 до 0.75, а висоти – від 0.2 до 0.65. Значення, які перевищують максимум, означають, що користувач підійшов надто близько до камери, а значення які є менше мінімуму, відповідно означають, що користувач знаходиться занадто далеко від камери. Для визначення розташування особи щодо центру кадру використовуються параметри left і top. Однак при виконанні завдань на повороти голови по сторонах вважається допустимим невеликий зсув від центру кадру убік. Для завдань, де потрібно відтворити нахил голови вгору або вниз, обличчя може бути зміщене щодо центру. У таблиці 3 представлені отримані діапазони значень для кожного із завдань.

Таблиця 3

Діапазон значень параметрів рамки для виконання завдань

Тип завдання	Poss_x min	Poss_x max	Poss_y min	Poss_y max
центральне фото	0.25	0.35	0.1	0.25
подивіться праворуч	0.35	0.65	0.1	0.25
подивіться ліворуч	0.05	0.2	0.1	0.25
подивіться вгору	0.25	0.35	0.2	0.6
подивіться вниз	0.25	0.35	0.02	0.25

Проведене тестування розробленої прикладної системи Liveness Detection на основі створеного алгоритму проходження Liveness та підібраних порогових значень AWS API складалося із 100 тестів різної спрямованості, від прямого проходження user-flow до тестувань із медичною маскою, окулярами і бородою, фотокарткою обличчя і відеозображенням людини з метою підробки її особистості. Завдяки початковій точності розпізнавання характеристик обличчя Amazon API у 99.99% та випадковому підбору завдань у розробленому алгоритмі Liveness Detection всі тести дали очікуваний результат – особа або успішно відеоідентифікувалася на підприємстві, або було встановлено підробку її особистості у ході виконання процесу Liveness Detection.

Висновки. Проведене дослідження порогових значень відповідей AWS API та розроблений у роботі алгоритм Liveness Detection дозволяє автоматично оцінювати, що у ході виконання відеоідентифікації система дійсно взаємодіє із фізично присутньою людиною, що дозволяє боротися із випадками шахрайства у багатьох сферах роботи інформаційних систем – від банального врахування часу роботи на підприємстві до запобігання злому систем безпеки банкінгу та інших фінтех рішень. Для розробки програмної імплементації створеного алгоритму була проведена інтеграція з сервісами отримання даних фотозображень AWS, під час якої, емпіричним шляхом були підібрані коефіцієнти якості вхідного зображення для проведення процесу Liveness Detection.

Врешті, на основі проведених досліджень і створених алгоритмів, була розроблена система відеоідентифікації співробітників підприємства та проведено її тестування, яке показало високі результати роботи створеного програмного продукту. Відкритість алгоритмів і коду системи, а також використання передових технологій при її розробці, дозволяють використовувати наробки, отримані у роботі, для інших програмних систем, де необхідно впроваджувати процес відеоідентифікації особи.

ЛИТЕРАТУРА

1. Apgar D. Survey of Face Liveness Detection for Unsupervised Locations / D. Apgar, M. R. Abid // 2021 IEEE 12th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON) – 2021 – P. 162-168.
2. Datta A. Face detection and recognition: Theory and practice / A. Datta, M. Datta – London: CRC Press, 2016 – 323 p.
3. Use-cases: Acuant [Електронний ресурс] – Режим доступу: <https://www.acuant.com/use-case/>.
4. Get to Know Kairos [Електронний ресурс] – Режим доступу: <https://www.kairos.com/docs/>.
5. Animetrics | Face Recognition [Електронний ресурс] – Режим доступу: <http://animetrics.com/>
6. API Documentation - Lambda Labs [Електронний ресурс] – Режим доступу: <https://lambdal.com/api-documentation.html>
7. Fan Y. Research on Liveness Detection Algorithms Based on Deep Learning / Y. Fan, Y. Shi, X. Wang, H. Yi // 2019 IEEE 10th International Conference on Software Engineering and Service Science (ICSESS) – 2019 – P. 1-6.
8. Amazon CloudFront Documentation [Електронний ресурс] – Режим доступу: https://docs.aws.amazon.com/cloudfront/index.html?nc2=h_ql_doc_cf.

REFERENCES

1. Apgar D. Survey of Face Liveness Detection for Unsupervised Locations / D. Apgar, M. R. Abid // 2021 IEEE 12th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON) – 2021 – P. 162-168.
2. Datta A. Face detection and recognition: Theory and practice / A. Datta, M. Datta – London: CRC Press, 2016 – 323 p.
3. Use-cases: Acuant [Electronic resource] – Access mode: <https://www.acuant.com/use-case/>.
4. Get to Know Kairos [Electronic resource] – Access mode: <https://www.kairos.com/docs/>.
5. Animetrics | Face [Electronic resource] – Access mode: <http://animetrics.com/>
6. API Documentation - Lambda Labs [Електронний ресурс] – [Electronic resource] – Access mode: <https://lambdal.com/api-documentation.html>
7. Fan Y. Research on Liveness Detection Algorithms Based on Deep Learning / Y. Fan, Y. Shi, X. Wang, H. Yi // 2019 IEEE 10th International Conference on Software Engineering and Service Science (ICSESS) – 2019 – P. 1-6.
8. Amazon CloudFront Documentation [Electronic resource] – Access mode: https://docs.aws.amazon.com/cloudfront/index.html?nc2=h_ql_doc_cf.

Received 29.08.2022.

Accepted 02.09.2022.

The information system for the liveness detection process using aws

Analysis of recent studies and publications. The analysis of the modern market of software and algorithmic solutions for performing the Liveness detection process showed that the currently existing approaches are completely commercial solutions with closed algorithms of their work. In addition, the Liveness detection algorithm is not yet standardized, and has many implementation options that can sometimes lead to ambiguous results [7]. That is why, it was decided to develop our own algorithm and liveness detection system based on obtaining face characteristics using the AWS API [8], because this service offers a high accuracy of face recognition, which is 99.99%, and provides 10,000 free requests to use every month, which enough to fulfill the purpose of our work.

Purpose of the study. Development of the algorithm for the Liveness Detection process using AWS and the construction of a video identification system based on it.

Main research material. This paper investigates the video identification of a person using the Liveness Detection process. The basis of the question, complexity of the Liveness Detection process and the implementation of biometric human video identification

have been studied. An algorithm for conducting the Liveness Detection process based on the execution of random tasks by an identified person has been developed.

Integration with the Amazon DetectFaces API was carried out in order to obtain the character of a digital image of the head, as a result of which it becomes possible to analyze a photo of a person for a wide range of key facial features. A manual test of the Amazon DetectFaces API was conducted, in the process of which, empirically, thresholds of facial characteristics in the image for the Liveness Detection process were set. Integration with the Amazon CompareFaces API was implemented to execute the biometric video identification.

The testing of the developed Liveness Detection application system based on the created Liveness passing algorithm and selected AWS API thresholds consisted of 100 tests of different orientations, from direct user-flow testing to tests with a medical mask, glasses and beard, a photo of a face and a video image of a person for the purpose of forger his personality. Thanks to the initial accuracy of Amazon API facial recognition of 99.99% and the random selection of tasks in the developed Liveness Detection algorithm, all tests gave the expected result - the person was either successfully video-identified at the enterprise, or his identity was faked during the Liveness Detection process.

Conclusions. *Based on the research, the software system of biometric human video identification has been developed, which allows to automatically establish the presence of a person in front of a video camera and identify it, preventing the use of fake artifacts to falsify a person's identity.*

Яковлева Марія Антонівна - студентка спеціальності «Ком'ютерні науки», кафедра інформаційних технологій. Університет імені Альфреда Нобеля.

Ковилін Єгор Романович - к.т.н., старший викладач кафедри інформаційних технологій Університет імені Альфреда Нобеля.

Yakovlieva Mariia Antonivna - student of the "Computer Science" specialty, Department of Information Technologies. Alfred Nobel University.

Kovylin Yehor Romanovych - Ph.D., Senior lecturer at the Department of Information Technologies, Alfred Nobel University.

ЛЕКСИЧНИЙ АНАЛІЗ ПРОГРАМНОГО КОДУ

Анотація. Будь-який розбір програмного коду починається з лексичного аналізу. Попри те, що лексичний аналіз вважається відносно простим етапом, він грає ключову роль у всій системі аналізу та перетворення програмного коду, а також має велику кількість як теоретичних так і практичних особливостей, що потребують ретельного дослідження. В даній статті розглядаються визначення лексичного аналізатора, принципи його побудови, теоретичні особливості та особливості практичної реалізації.

Ключові слова: лексичний аналізатор, лексема, токен, регулярні вирази, теорія формальних мов, недетермінований кінцевий автомат, детермінований кінцевий автомат.

Етапи перекладу програмного коду, які наведені на рисунку 1, можна розділити на дві глобальні фази – фаза аналізу та фаза синтезу. Етап лексичного аналізу поряд з синтаксичним та семантичним аналізом належить до фази аналізу [1].



Рисунок 1 – Етапи перекладу програмного коду

Лексичний аналіз – це процес перетворення послідовності символів вихідного програмного коду в послідовність токенів (груп символів, що відповідають певним шаблонам) та визначення їх типів. Програма чи функція, що виконує лексичний аналіз, називається лексичним аналізатором, токенізатором чи сканером [2].

Лексичний аналізатор сканує вихідний код рядок за рядком, приймаючи на вхід лексеми він генерує потік токенів. Наприклад, в компіляторах, часто, лексичний аналізатор є функцією, яку використовує синтаксичний аналізатор для отримання чергового токена.

Лексеми аналогічні словам з природніх мов з однією невеликою різницею: слова окремо мають власні значення, тоді як лексеми частіше всього передають зміст групами. Наприклад, визначимо наступний рядок коду:

$$x = a + b * c; \quad (1)$$

У виразі (1) «x» окремо не передає жодного значення, доки ми не розглянемо весь арифметичний вираз. Токен є значенням лексеми. Тоді з точки зору значення у даному виразі «x» є ідентифікатором, символ «=» є оператором (оператором присвоєння), «a» - ідентифікатор і т.д. Як бачимо лексеми «x», «a», «b», «c» мають різний вигляд, але значення передають одне і те саме. А отже, роботою лексичного аналізатора є знаходження значення кожної лексеми. Загалом в мовах програмування виділяють наступні типи токенів [3]:

- ідентифікатор;
- оператор;
- константа;
- ключове слово;
- літерал;
- знак пунктуації;
- спеціальний символ.

На рисунку 2 наведений функціональний склад лексичного аналізатора. Лексеми передаються до фази сканування, яка в свою чергу усуває елементи, що не є токенами, такі як коментарі, послідовні пробіли та інше. В результаті, під час фази аналізу формуються токени. Звичайно, якщо лексичному аналізатору не вдалося знайти відповідний токен до лексеми, він повинен сформулювати повідомлення про помилку з необхідною інформацією.



Рисунок 2 – Функціональний склад лексичного аналізатора

Процес формування токенів потребує певний набір шаблонів для знаходження значень лексем. Стандартним методом опису шаблонів пошуку є регулярні вирази.

Регулярні вирази (англ. regular expressions) – формальна мова, яка використовується в комп'ютерних програмах, що працюють з текстом, для пошуку та здійснення маніпуляцій з підрядками у тексті, заснована на використанні метасимволів. Для пошуку використовується рядок-зразок, який складається з символів та метасимволів та задає правило пошуку [4].

Регулярні вирази мають множину діалектів, особливу популярність набули наступні діалекти: POSIX, Perl, PCRE, ECMAScript, Python. В теорії формальних мов регулярні вирази складаються з констант та операторів, які визначають множину рядків та множину операцій над ними відповідно. На цьому кінцевому алфавіті Σ визначені наступні константи [5]:

- порожня множина \emptyset позначає \emptyset ;
- порожній рядок ϵ позначає множину $\{\epsilon\}$;
- рядок a в Σ позначає множину $\{a\}$;

та наступні операції [5]:

- конкатенація (concatenation) RS позначає набір рядків, який можна отримати конкатенацією множин визначених R та S (у цьому порядку), наприклад: нехай R позначає $\{“ab”, “c”\}$ та S позначає $\{“d”, “ef”\}$, тоді RS позначає $\{“abd”, “abef”, “cd”, “cef”\}$;

- чергування (alternation) $R|S$ позначає об'єднання множин, описаних R та S , наприклад: якщо R описує $\{“ab”, “c”\}$ та S описує $\{“ab”, “d”, “ef”\}$, тоді вираз $R|S$ описує $\{“ab”, “c”, “d”, “ef”\}$;

- замикання Кліні (Kleene closure) R^* позначає найменшу надмножину множин, описаної R , що містить ϵ і є замкнутою відносно конкатенації рядків, це набір всіх рядків, який може бути створений завдяки конкатенації будь-

якого кінцевого числа (включаючи нуль) рядків із набору описаного R , наприклад: якщо R позначає {"0", "1"}, тоді R^* позначає набір всіх кінцевих двійкових рядків (включаючи порожній рядок).

Для прикладу створимо набір шаблонів на основі регулярних виразів для наступної лексики мови програмування:

1. Мова програмування має одне ключове слово "while".
2. Ідентифікатор може складатися з літер латинського алфавіту та цифр, але повинен починатися з літери.
3. Ціле число може бути 0 або складатися з цифр, за умови того, що перша цифра не буде 0. Ціле число може містити знак на початку.
4. Дійсне число з фіксованою крапкою (decimal), наприклад 134.0093, -76.11.
5. Дійсне число з плаваючою крапкою (real), наприклад 1.3400E+02, -98.345E-5, 156E8.

Для зручності можна сформулювати наступні позначення, використовуючи описані операції над рядками:

1. $P^+ = PP^*$. Нехай P це рядок "a", тоді P^+ буде визначати наступну множину: {"a", "aa", "aaa", "aaaa", ...}.
2. $P? = P| \epsilon$. Нехай P це рядок "a", тоді $P?$ буде визначати наступну множину: {"a", "ε"}.
3. $[a-z] = a|b|c|...|z$. Позначення буде визначати множину {"a", "b", "c", "d", ..., "z"}.

Також, для зручності можна давати імена регулярним виразам, щоб потім посилатися на них за іменем. В результаті, набір регулярних виразів для наведеної лексики буде наступним:

```
while-keyword = while
letter = [a-zA-Z]
digit = [0-9]
identifier = letter (letter | digit)*
sign = + | -
integer = sign? (0 | [1-9] digit*)
decimal = integer . digit*
real = (integer | decimal) E sign? digit+
```

Існує два основних правила під час визначення токени [6]:

1. Якщо є неоднозначність, коли у вхідних даних є збіг з декількома шаблонами різних розмірів, то обирається шаблон з найбільшою довжиною – правило найдовшого збігу. Наприклад, якщо в коді зустрінеться послідовність си-

мволів “while”, то згідно сформованих регулярних виразів ми отримуємо токен ідентифікатора, а не токен ключового слова while, тому що шаблон ключового слова має менше символів.

2. Якщо є шаблони, які відповідають однакової максимальній кількості символів, то обирається шаблон, який вказаний першим – правило пріоритету. Наприклад, якщо в кодї зустрінеться послідовність символів “while”, то дана лексема відповідає одночасно токену ключового слова while і токену ідентифікатора, проте, так як шаблон ключового слова описаний раніше в наборі шаблонів, то буде сформований саме токен ключового слова.

На практиці лексичні аналізатори використовують дещо інший механізм пошуку значень лексем. А саме кінцеві автомати (finite-state machine) – особливий різновид автомату - абстракції, що використовується для опису шляху зміни стану об'єкта в залежності від поточного стану та інформації отриманої ззовні. Його особливістю є скінченність множини станів автомату [7].

В реалізаціях лексичних аналізаторів застосовують різновид кінцевого автомату, а саме детермінований кінцевий автомат. Це обумовлено практичною ефективністю використання детермінованого алгоритму на реальних обчислювальних машинах. Існують інструменти, які допомагають переводити відносно легкі у сприйнятті людиною регулярні вирази в кінцеві автомати. Наприклад, таким інструментом є GNU Flex. GNU Flex дозволяє описати лексичний аналізатор простою мовою Flex, перерахувавши регулярні вирази для окремих токенів і вказавши код, який буде виконуватись під час зіставлення кожного токену. В результаті Flex генерує код мовою програмування C або C++. Отриманий код використовує детермінований кінцевий автомат для аналізу тексту [8].

Розглянемо побудову кінцевих автоматів на прикладі декількох токенів з мови програмування C, а саме ключове слово if, ідентифікатор та ціле число. На рисунку 3 наведені індивідуальні кінцеві автомати для цих токенів з власними початковими станами. Для спрощення сприйняття в рисунках кінцевих автоматів спеціально не будуть зображуватись невизначені стани, які приводять до помилки визначення токену. Адже зрозуміло, що за відсутності переходу, який відповідає вхідним даним ми будемо формувати помилку.

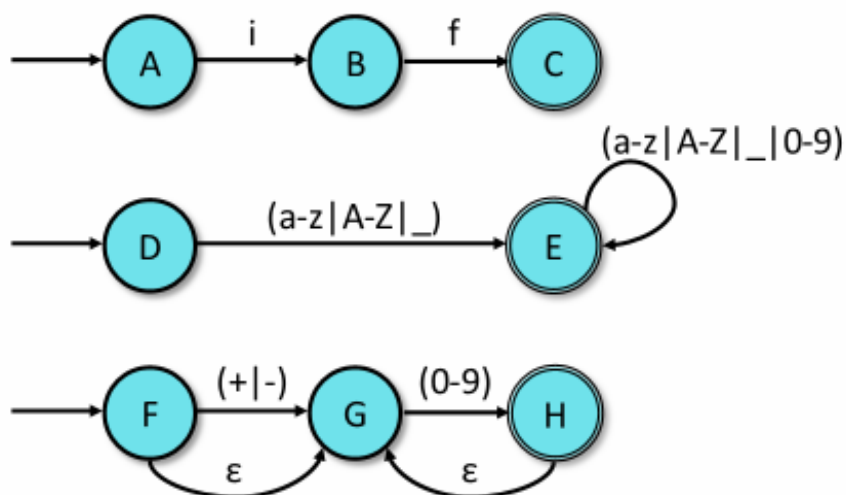


Рисунок 3 – Індивідуальні кінцеві автомати для токенів

Отже, розглянемо автомати на рисунку 3:

1. Для ключового слова *if*. З початкового стану *A*, якщо ми отримуємо літеру “*i*”, то ми переходимо до наступного стану *B*. Потім з *B* отримавши літеру “*f*” ми закінчуємо у кінцевому стані *C*.

2. Для ідентифікатора. З початкового стану *D*, отриманий символ повинен бути будь-якою маленькою літерою з діапазону латинського алфавіту від “*a*” до “*z*” або будь-якою великою літерою з того ж діапазону, або нижнім підкресленням, тоді ми переходимо до фінального стану *E*, тому що одна літера також може бути ідентифікатором. Додатково ідентифікатор може мати будь-яку кількість літер, нижніх підкреслень або цифр, але вони повинні слідувати за будь-якою літерою чи нижнім підкресленням, а тому ми робимо перехід з кінцевого стану *E* в нього ж, доки не завершиться процес переходу.

3. Для цілого числа. З початкового стану *F*, якщо ми отримуємо будь-який знак (позитивний чи негативний), то ми переходимо до стану *G*, і звідси, якщо наступний символ будь-який з діапазону цифр від 0 до 9 ми потрапляємо до кінцевого стану *H*. Цього достатньо для цілого числа з однією цифрою. Для двох або більше цифр в числі ми маємо ϵ -перехід від кінцевого стану *H* в *G*. ϵ -перехід – це перехід без використання вхідного символу, зручний спосіб моделювання. Тепер, якщо ми не бачимо іншого переходу, ми можемо переміститися в стан *G*, і тоді можемо мати будь-яку кількість цифр. Між іншим, ціле число може не мати жодного знаку на початку, тому що для позитивного числа позитивний знак ставити не обов’язково. Тому необхідно сформулювати ϵ -перехід від початкового стану *F* в стан *G*.

Під час фази аналізу в лексичному аналізаторі не застосовуються відокремленні кінцеві автомати, використовується один кінцевий автомат, що об'єднує усі відокремленні автомати. Отже, нам необхідно виконати об'єднання кінцевих автоматів. В нашому випадку можна ввести новий початковий стан S і побудувати ϵ -переходи до існуючих автоматів (рисунок 4). В результаті сформувався недетермінований кінцевий автомат (NFA). Точніше це ϵ -NFA – різновид недетермінованого кінцевого автомата з ϵ -переходами [3].

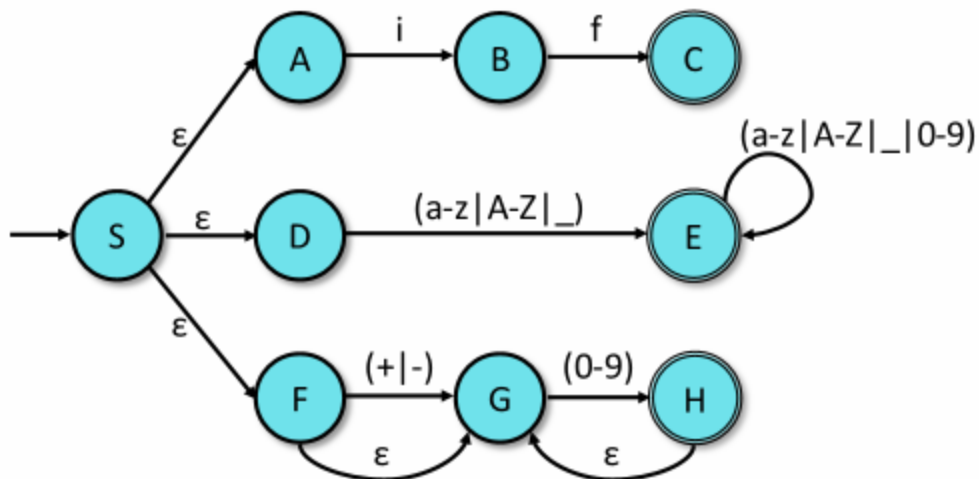


Рисунок 4 – Недетермінований кінцевий автомат з ϵ -переходами

В недетермінованих автоматах вхід може привести до одного, більше ніж одного або зовсім без переходу для даного стану. Якщо в теорії дана особливість не має критичного значення, то на практиці недетермінований автомат не підходить для впровадження в машинних алгоритмах. Лексичні аналізатори використовують детерміновані кінцеві автомати (DFA). На відміну від недетермінованих, в детермінованих автоматах кожен стан має лише один перехід для кожного входу. При цьому, існує алгоритм трансформації будь-якого NFA в складніший DFA з однаковою функціональністю [7]. Отже, на рисунку 5 наведений еквівалентний до автомата з рисунку 4 детермінований кінцевий автомат.

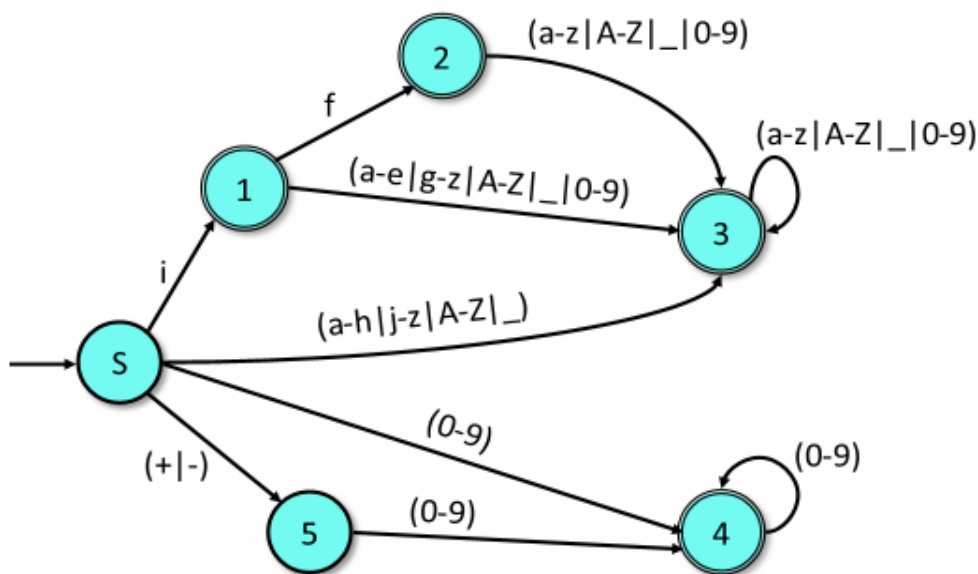


Рисунок 5 – Детермінований кінцевий автомат

Він розпізнає ключове слово *if*, ідентифікатор та ціле число. Розглянемо детальніше:

Починаючи зі стартового стану *S*, якщо автомат отримує маленьку літеру “*i*”, ми переходимо до стану *1*. Звідси, якщо ми отримуємо маленьку літеру “*f*”, то переходимо до стану *2*. Так як стан *2* є кінцевим станом, то якщо ми зупиняємося тут, це означає, що автомат визначив вхідну лексему як токен ключового слова *if*.

Якщо вхідна лексема має більше символів, ніж “*if*”, наприклад, “*iffy*”, то ми перейдемо до кінцевого стану *3*, і далі будемо використовувати перехід з *3* стану в нього ж. Якщо ми зупинимось на стані *3*, це означає, що автомат визначив вхідну лексему як токен ідентифікатора. На рисунку 5 видно, що стан *1* також є кінцевим станом, адже, якщо ми отримаємо літеру “*i*” від початкового стану *S* і за нею не буде слідувати літера “*f*”, тоді це не ключове слово *if*, а ідентифікатор. Якщо ми зупинимось в стані *1*, то автомат визначить лексему як ідентифікатор. Тепер розглянемо перехід від стану *1* до стану *3*. Тут спеціально перевіряється діапазон маленьких літер від “*a*” до “*e*” і від “*g*” до “*z*”, тобто ми виключили з діапазону літеру “*f*”, адже літера “*f*” є умовою переходу від стану *1* до стану *2*. За такого діапазону ми просуваємося до стану *3* і можемо визначити лексему як ідентифікатор. Так само працює перехід від стану *S* до стану *3*, якщо ми отримаємо будь-яку літеру (або нижнє підкреслення) окрім маленької літери “*i*”, то ми перейдемо до стану *3*.

Очевидно, що перехід від стану S до стану 4 дозволяє визначити вхідну лексему як токен цілого числа. Перехід від S до стану 5 враховує цілі числа зі знаком на початку.

В результаті кінцевий стан 2 визначає вхідну лексему як токен ключового слова `if`, кінцеві стани 1 та 3 – як токен ідентифікатора, кінцевий стан 4 – як токен цілого числа.

Детермінований кінцевий автомат як такий, що наведений на рисунку 5 реалізується у фазі аналізу (рисунок 2) в лексичному аналізаторі. Йому на вхід передаються лексеми, за виключенням елементів, що не є токенами, для визначення значень цих лексем, тобто для формування токенів.

Отже, ми розглянули загальне визначення лексичного аналізатора та принципи його побудови, ретельно дослідили теоретичне підґрунтя лексичного аналізу, а також особливості практичної реалізації на обчислювальних машинах.

ЛІТЕРАТУРА

1. В.Г. Павлов, К.О. Шапран Основні принципи побудови лексичних аналізаторів. Матеріали IV Міжнародної науково-технічної конференції молодих учених та студентів. Актуальні задачі сучасних технологій – Тернопіль 25-26 листопада 2015. С. 68-69.
2. Lexical analysis [Електронний ресурс] – Режим доступу: https://en.wikipedia.org/wiki/Lexical_analysis
3. Introduction to compiler design (PPT) [Електронний ресурс] – Режим доступу: <https://www.nesoacademy.org/cs/12-compiler-design/ppts/01-introduction-to-compiler-design>
4. Регулярные выражения [Електронний ресурс] – Режим доступу: <https://bit.ly/3UrPRvp>
5. Regular expression [Електронний ресурс] – Режим доступу: https://en.wikipedia.org/wiki/Regular_expression
6. Regular Expressions (REs) [Електронний ресурс] – Режим доступу: <https://lambda.uta.edu/cse5317/notes/node7.html>
7. Скінченний автомат [Електронний ресурс] – Режим доступу: <https://bit.ly/3ukLJT4>
8. Изучаем генератор кода GNU Flex [Електронний ресурс] – Режим доступу: https://ps-group.github.io/compilers/gnu_flex_doc

REFERENCES

1. V.G. Pavlov, K.O. Shapran Basic principles of building lexical analyzers. Materials of the IV International Scientific and Technical Conference of Young Scientists and Students. Actual tasks of modern technologies - Ternopil, November 25-26, 2015. P. 68-69.
2. Lexical analysis [Electronic resource] – Resource access mode:
https://en.wikipedia.org/wiki/Lexical_analysis
3. Introduction to compiler design (PPT) [Electronic resource] – Resource access mode: <https://www.nesoacademy.org/cs/12-compiler-design/ppts/01-introduction-to-compiler-design>
4. Regular expressions [Electronic resource] – Resource access mode:
<https://bit.ly/3UrPRvp>
5. Regular expression [Electronic resource] – Resource access mode:
https://en.wikipedia.org/wiki/Regular_expression
6. Regular Expressions (REs) [Electronic resource] – Resource access mode:
<https://lambda.uta.edu/cse5317/notes/node7.html>
7. Finite state machine [Electronic resource] – Resource access mode:
<https://bit.ly/3ukLJT4>
8. Learning the GNU Flex Code Generator [Electronic resource] – Resource access mode: https://ps-group.github.io/compilers/gnu_flex_doc

Received 12.09.2022.

Accepted 16.09.2022.

Lexical analysis of program code

The growing volume of technologies, the end of actively used development tools support, outdated API etc., entails the need of program codes conversion. In IT companies and not only, often begged the question of deprecated software support, which customers continue to use, or translation of current software to actual technologies. It is more rational for programmers to use the conversion and save most of code base, than rewriting all software by hand, even if manual adjustment is needed. At this moment, there are few high-quality code conversion systems. Largely, conversion systems work well only with similar programming languages. The task of program codes conversion is a deep and complex topic. To convert the software code, you must first analyze, select components and form a structural representation. Any analysis of program code begins with lexical analysis. Although lexical analysis is considered a relatively simple step, it plays a key role in the entire system of analysis and transformation of software code, and also has a large number of both theoretical and practical features that require careful study.

This article considers the definition of the lexical analyzer, its functional composition and principles of construction, provides key differences between the lexeme and the token. Two approaches have been proposed and considered to solve the search for tokens in the program code: regular expression search and finite state machine search. For these approaches, examples of the formation of search templates under certain rules of vocabulary were given. As a result, the optimality of the use of deterministic finite state machines during the practical implementation of the lexical analyzer on real computing machines was substantiated.

Сокол Іван Олександрович – аспірант, кафедра комп’ютерних наук та інформаційних технологій, Дніпровський національний університет ім.О. Гончара.

Волковський Олег Степанович – к.т.н., доцент кафедри КНІТ, Дніпровський національний університет ім.О. Гончара

Sokol Ivan – graduate student, Department of computer science and information technologies, Oles Honchar Dnipro National University.

Volkovskiyi Oleg – PhD in Technical Science, associate professor of CSIT Department, Oles Honchar Dnipro National University.

**RESEARCH OF POLYNOMIAL APPROXIMATION OF FORCES
IN ROD ELEMENTS OF CORRODING STRUCTURES**

Abstract. The paper considers the problem of polynomial approximation of the "force - time" dependencies in the elements of corroding hinged-rod structures from the point of view of the influence of the degree of the polynomial on the error in calculating their durability. A method for determining the coefficients of approximating polynomials is proposed, which is based on the use of a numerical-analytical algorithm for solving a system of differential equations describing the corrosion process. The results of numerical experiments are presented, illustrating for various constructions the dependence of the error in solving the problem on the degree of approximating polynomials.

Keywords: corrosion process, correction functions method, systems of differential equations, polynomial approximation.

Problem statement. Many structural elements of machines and units, primarily those used in the chemical and mechanical industries, are exposed during operations to the process environment that are corrosive to metal. This leads to its corrosive wear – the destruction of the subsurface layer of the metal, a change in the initial geometric characteristics of the structure, a decrease in the bearing capacity and, as a result, premature, often emergency, failure of the structural element. In the general case of corrosion interaction mechanical stresses accelerate the corrosion process.

The system of differential equations (SDE), which describes the corrosion process taking into account the influence of mechanical stresses, has the form:

$$\frac{d\delta_i}{dt} = v_0 \cdot \Phi(\sigma_i(\bar{\delta})); \quad \delta_i|_{t=0} = 0; \quad i = \overline{1, K}. \quad (1)$$

Here δ_i – is the depth of corrosion damage (damage parameter); v_0 – is the corrosion rate in the absence of stress; Φ – is the known stress function; t – is time; K – the number of damage parameters that uniquely determine the shape and dimensions of the structural element (SDE dimension).

To calculate the stresses that are included in the right parts of the SDE (1), a system of equations of mechanics is used: equations of equilibrium and compatibility of deformations, Cauchy relations and physical relations (for elastic bodies – Hooke’s law). Since the stress functions on the right side of (1) have the form of computational algorithm, only a numerical solution of system is possible (1).

When solving the problem of optimal design of corroding structures, the calculation of constraint functions (CF) involves the calculation of the stress state of the structure at a given point in time, taking into account the corrosion process occurring in it. The scheme for solving the optimization problem is a two-loop scheme (Fig. 1), where OF – is the module for calculating the objective function, CF – is the module for calculating the constraint function, SDE – is the module for solving the system of differential equations, SSS – is the module for solving the problem of mechanics, NLP – is the module for solving the problem of optimal design.

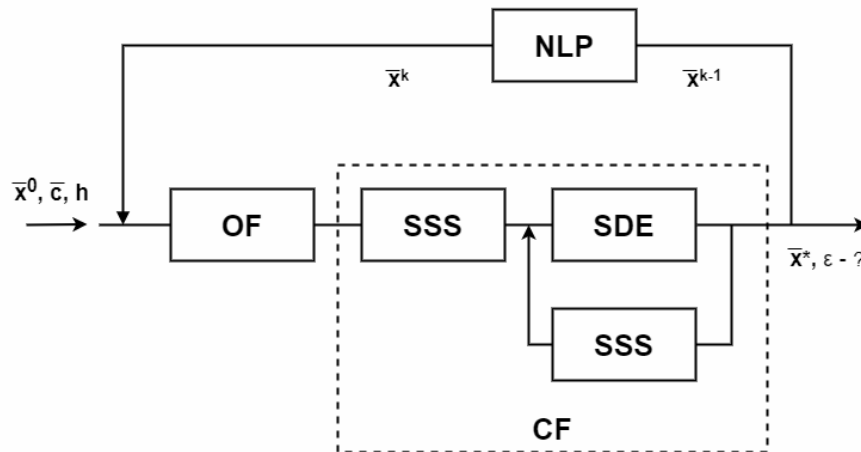


Figure 1 – Scheme for solving the optimization problem

The presence of feedback significantly increases the computational cost of solving the optimization problem and increases its sensitivity to errors that arise when calculating the constraint function.

Consider the error problem, which always arises when the parameter is the numerical solution of system (1) does not change in the process of solving the optimization problem. Let us assume that for some vector of variable parameters (VVP) \bar{x}^s a numerical solution was obtained for the parameter h with an error ε^s , which is in the neighborhood of its maximum allowable value ε^* , set by customer: $|\varepsilon^s - \varepsilon^*| \leq \delta$. For the new VVP \bar{x}^{s+1} the error of the numerical solution ε^{s+1} , obtained with the same parameter h , will differ from ε^s .

Thus, the entire feasible set of solutions can be represented as the union of three subsets:

$$D = D_1 \cup D_2 \cup D_3. \quad (2)$$

Here D_1 – is a subset of solution whose error is higher than the maximum allowable, so any solution from this subset cannot be considered satisfactory; D_2 – a subset of solutions, the error of which is less than the maximum allowable; obtaining any of these solutions is inefficient in terms of computational costs; D_3 – a subset of solutions, the error of which lies in a certain neighborhood of the maximum allowable; any solution from this subset is optimal in terms of the efficiency of the numerical method.

Obviously, when using the traditional approach to solving the problem, the cardinality of the subset D_3 is incomparably less, than that of the other two. The ratio of the cardinalities of the subsets D_1 и D_2 depends on the choice of the parameter of the numerical solution of the SDE (1).

Increasing the accuracy and efficiency of numerical methods for solving the problems of calculating durability and optimal design of corroding structures involves increasing the cardinality of a subset D_3 in the ideal case – to the cardinality of the entire feasible set of solutions. To solve this problem, new approaches to modeling the process of corrosion deformation are needed, using methods of computational intelligence [1]. When building such models, the problem of polynomial approximation becomes very relevant.

Analysis of recent research and publications. In [2], apparently for the first time, an algorithm was proposed for controlling the accuracy of a numerical solution of an SDE of the form (1) using neural networks (NNs). It assumed the presence of several NNs, each of which made it possible to determine the parameter of the numerical solution for a given vector of variable parameters for a specific value of the error margin. Later, this algorithm was modified – is used a single NN, one of the input values of which was the value of the error margin [3]. The disadvantage of these algorithms was ignoring the change in internal forces in the elements of corroding structures when obtaining a training sample, as a result of which the real value of the error in solving the problem did not always coincide with the predicted one.

In [4], a correction functions method was proposed, in which the correction function (the error of the approximate solution of the SDE) was approximated by a neural network. Its input values were, among other things, the coefficients of the

polynomial describing the change in time of internal forces in the elements of the hinged rod structure. The polynomial coefficients were determined at the stage of obtaining an approximate solution. Thus, the accuracy of solving the problem as a whole depended on the accuracy of the polynomial approximation.

Research objective. This paper proposes a research of the influence of the degree of polynomials approximating the dependence «force – time» in the elements of corroding hinged-rod structures, and the method of constructing them on the accuracy of calculating durability.

Presentation of the research main material. The solution of the problem of durability of a corroding hinged-rod structure (HRS) will be sought in the form:

$$t^*(\bar{x}, \bar{y}, \bar{c}, h) = \tilde{t}(\bar{x}, \bar{y}, \bar{c}) \cdot \varphi_k(\bar{y}, \bar{c}, \bar{\alpha}); \quad k = \overline{1, K}, \quad (3)$$

where $\tilde{t}(\bar{x}, \bar{y}, \bar{c})$ – approximate solution obtained with minimal computational costs, $\varphi_k(\bar{y}, \bar{c}, \bar{\alpha})$ – correction function, which is the error of the approximate solution; \bar{x} – vector of variable construction parameters; \bar{y} – vector of constant design parameters; \bar{c} – vector of process environment parameters; $\bar{\alpha}$ – vector of parameters determined in the process of obtaining an approximate solution; h – parameter of the numerical solution of the SDE; K – the number of correction functions required to solve the problem.

In our case $\bar{\alpha}$ is a vector of polynomial coefficients approximating the dependence of internal forces in the rod elements of the structure on time. It is obvious that the degree of the approximating polynomial and the way it is constructed will significantly affect the value of the correction function and the accuracy of solving the problem as a whole. In this case, the values of the coefficients must be determined when implementing the algorithm for obtaining an approximate solution to the durability problem.

To obtain an approximate solution, it is proposed to use a numerical-analytical method for solving SDEs of the form (1), which is quite fully described in [5].

For the corroding element of the HRC, an analytical formula is known that allows you to determine the time during which the stress in it increases from σ_0 to σ at a constant value of axial force Q :

$$t^*_{an} = t_0 - \frac{2kQ}{v_0 \cdot |d|} \left\{ \operatorname{arctg} \frac{2a\delta - P_0}{|d|} + \operatorname{arctg} \frac{P_0}{|d|} \right\}, \quad (4)$$

Here A_0, P_0 – area and perimeter of the section at the initial moment of time;

k – coefficient of influence of stress on corrosion rate; $t_0 = \frac{\delta^*}{v_0}$; a – section shape

factor; $c = A_0 + kQ$; $d = \sqrt{|P_0^2 - 4ac|}$, ($d \neq 0$); δ^* – depth of corrosion wear corresponding to the limit state of the element.

In well-known numerical-analytical algorithms, it is proposed to use a uniform step by stress $\Delta\sigma = \frac{[\sigma] - \sigma_0}{n} = const$ (Fig. 2), and the corresponding value Δt determine by formula (4).

It is obvious that the accuracy of the numerical-analytical algorithm will depend on the number of splitting points of the interval $(\sigma_0; [\sigma])$.

A 5-rod statically indeterminate HRC was taken as a model design (Fig. 3). The parameters of the HRC and the process environment were assumed to be known: $L = 200$ cm; $E = 2,1 \times 10^5$ MPa; $[\sigma] = 240$ MPa; $v_0 = 0,1$ cm/year, $k = 0,003$ MPa⁻¹; external load value $P = 250$ kN.

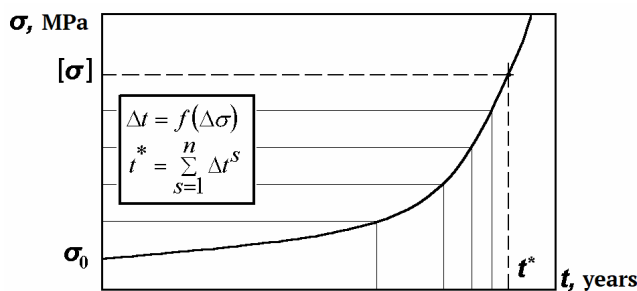


Figure 2 – Graphic illustration of the method

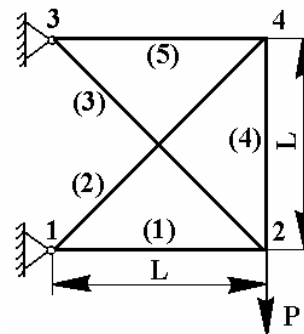


Figure 3 – Design scheme HRC

All rods had sections corresponding to standard shaped profiles (channel, I-beam, equal and unequal angles) of various standard sizes. Numerical experiments were carried out with various parameters of the rods, which were chosen randomly. For each option, a reference solution of the durability problem was obtained. In this case, system (1) was solved numerically using the Runge-Kutta method of the 2nd order of accuracy. The numerical solution parameter was taken equal to

$$\Delta t = 0,005 \cdot t^*,$$

where $t^* = \min \{t^*_1, t^*_2, t^*_3, t^*_4, t^*_5\}$ – element durability value calculated from the formula (4) and determining the durability of the structure as a whole. Internal

forces and stresses in HRC elements were determined by the finite element method (FEM).

Two HRCs were selected for further research, differing only in the parameters of the rod elements. Their durability was determined by the durability of the elements: element (2), working in compression (construction (A), $t_A^* = 2,688$ year), and элементом element (3), working in tension (construction (B), $t_B^* = 2,126$ year). Changes in internal forces over time in these elements are shown in Fig. 4 (solid and dashed lines, respectively). For clarity, the values of time and force are given in dimensionless quantities:

$$\tilde{t} = \frac{t}{t^*}; \quad \tilde{Q} = \frac{Q - Q_0}{Q_t - Q_0}. \quad (5)$$

Here $Q_0 = Q|_{t=0}$; $Q_t = Q|_{t=t^*}$.

The choice of structures was determined by the nature of the changes in the forces. For construction (A) function graph $Q_A(t)$ has points of maximum and inflection; for construction (B) it is a smooth curve. Obviously, the polynomial approximation $Q_A(t)$ is a more difficult problem than the approximation $Q_B(t)$.

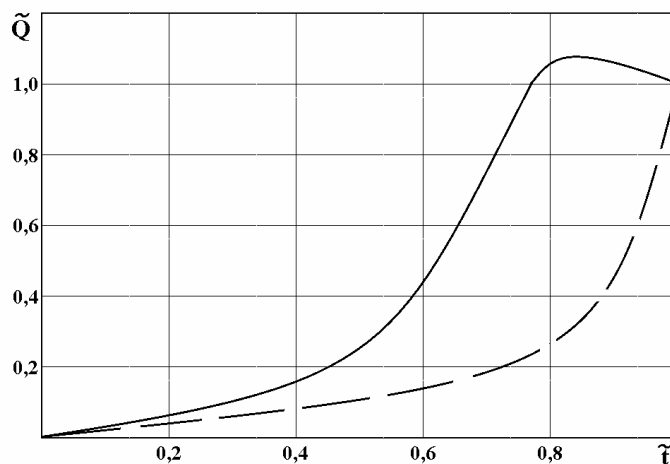


Figure 4 – Change in internal forces over time

Formula (4) makes it possible to obtain an exact solution of the durability problem only in the case when the internal forces in the rod elements of the structure remain constant throughout the entire period of its operation, i.e. for a statically defined construct. On the other hand, if the element number is known m , the durability of which determines the durability of the structure as a whole, and the law of change in the force in it, then the solution of a single differential equation of the form:

$$\frac{d\delta_m}{dt} = v_0 \cdot \Phi \left(\frac{Q_m(t)}{A_m^0 - P_m^0 \delta + a\delta^2} \right) \quad (6)$$

will coincide with the solution of the system (1).

Let approximate the dependence $Q_m(t)$ by a polynomial of degree n :

$$Q_m(t) = P_n(t) = \sum_{k=0}^n \alpha_k \cdot t^k. \quad (7)$$

To determine the coefficients of the polynomial, information is needed on the stress-strain state of the HRC at $(n+1)$ nodal points, including the initial time $t = 0$. The application of the numerical-analytical method will make it possible to obtain an approximate solution $\tilde{t}(\bar{x}, \bar{y}, \bar{c})$ with minimal computational costs (the number of calls to the FEM procedure is determined by the order of the polynomial (7)) and, at the same time, determine the number of the element that determines the durability of the structure and the values of the polynomial coefficients.

Thus, the method for calculating the coefficients of polynomial (7) is determined by the method of obtaining an approximate solution. Next, it is necessary to determine the degree of the polynomial, which provides the required accuracy of the calculation.

When obtaining reference solutions to the durability problem, tabular values of the functions $Q_A(t)$ and $Q_B(t)$ were obtained. We will use these functions in the numerical solution of the differential equation (6). In this case, the results will coincide with the corresponding reference solutions of the durability problem for structures (A) and (B). If approximating polynomials (7) are used in the numerical solution of (6), then the divergence between the results will depend only on their degree. As a criterion for the quality of the polynomial approximation, we take the relative error of the numerical solution (6) using (7).

Some research results are shown in the Table. 1.

Table 1

$P_n(t)$	Construction (A) $t_A^* = 2,688$ year		Construction (B) $t_B^* = 2,126$ year	
	t_n , years	ε , %	t_n , years	ε , %
2	2.871	6,809	2.215	4,191
3	2.775	3,244	2.148	1,032
4	2.727	1,434	2.239	0,533
5	2.713	0,922	2.209	0,391

As follows from the above results, an increase in the degree of approximating polynomials does not lead to a significant increase in the accuracy of calculations.

Conclusions. The study made it possible to substantiate the parameters of polynomial approximation when using the method of correction functions for solving problems of durability and optimal design of corroding hinge-rod systems. Since the correction functions are neural networks, the presence of several matrices of synaptic weights obtained for polynomials of different degrees will allow you to control the accuracy of calculating the constraint functions in the process of solving the optimization problem. This approach is consistent with the strategy of the sliding tolerance method and can significantly reduce computational costs.

ЛІТЕРАТУРА

1. Зеленцов Д.Г., Короткая Л.И. Технологии вычислительного интеллекта в задачах моделирования динамических систем. Днепр: Баланс-Клуб, 2018. – 178 с.
2. Зеленцов Д.Г., Короткая Л.И. Использование нейронных сетей при решении задач долговечности корродирующих конструкций. // Вісник Кременчуцького національного університету ім. М.Остроградського. – Кременчук: КрНУ, 2011. – Вип. 3 (68), част. 1. – С. 24 – 27.
3. Денисюк, О. Р. Определение рациональных параметров численного решения систем дифференциальных уравнений // Вестник Херсонского национального технического университета. – 2016. – №3(58). – С. 208 – 212.
4. Zelentsov D.G., Denysiuk O.R., Korotka L.I. The Method of Correction Functions in Problems of Optimization of Corroding Structures. // Advances in Computer Science for Engineering and Education III (ICCSEEA 2020), 2020. – pp. 132 – 142.
5. Зеленцов Д.Г., Науменко Н.Ю., Ляшенко О.А. Информационное обеспечение расчётов корродирующих объектов. Конечно-элементное моделирование. Днепр: Баланс-Клуб, 2018. – 174 с

REFERENCES

1. Zelentsov D.G., Korotkaya L.I. Tehnologii vyichislitel'nogo intellekta v zadachah modelirovaniya dinamicheskikh sistem. Dnepr: Balans-Klub, 2018. – 178.
2. Zelentsov D.G., Korotkaya L.I. Ispolzovanie neyronnykh setey pri reshenii zadach dolgovechnosti korrodiryuyuschih konstruktsiy. // Visnyk Kremenchutskoho natsionalnoho universytetu im. M.Ostrohradskoho. – Kremenчук: KrNU, 2011. – Vyp. 3 (68), chast. 1. – S. 24 – 27.
3. Denisyuk, O. R. Opredelenie ratsionalnykh parametrov chislennoho resheniya sistem differentsialnykh uravneniy // Vestnik Hersonskogo natsionalnoho tehnikeskogo universiteta. – 2016. – №3(58). – С. 208 – 212.

4. Zelentsov D.G., Denysiuk O.R., Korotka L.I. The Method of Correction Functions in Problems of Optimization of Corroding Structures. // Advances in Computer Science for Engineering and Education III (ICCSEEA 2020), 2020. – pp. 132 – 142.

5. Zelentsov D.G., Naumenko N.Yu., Lyashenko O.A. Informatsionnoe obespechenie raschYotov korrodირuyuschih ob'ektov. Konechno-elementnoe modelirovanie. Dnepr: Balans-Klub, 2018. – 174 s.

Received 26.09.2022.

Accepted 30.09.2022.

Дослідження поліноміальної апроксимації зусиль в стержневих елементах кородуючих конструкцій

У роботі розглядається проблема поліноміальної апроксимації залежностей «зусилля – час» в елементах шарнірно-стержневих конструкцій з точки зору впливу ступеня поліному на похибку обчислення їх довговічності. Процес корозії в конструкціях описується системою диференціальних рівнянь із врахуванням механічних напружень, яка може бути розв'язана тільки з використанням чисельних методів. Алгоритм розв'язання задачі оптимального проектування кородуючих конструкцій вимагає обчислення функцій обмежень, які передбачають обчислення напруженого стану конструкції в певний момент часу. Наявність зворотного зв'язку в даному алгоритмі збільшує обчислювальні витрати та чутливість до похибок. Останні дослідження пропонують використання нейронних мереж для контролю точності чисельних розв'язків систем диференціальних рівнянь, однак ці алгоритми ігнорують зміну внутрішніх зусиль у кородуючих конструкціях. Пізніше був запропонований метод коригувальних функцій, де коригувальна функція була апроксимована нейронною мережею з використанням коефіцієнтів полінома, який описує зміни внутрішніх зусиль з часом. Коефіцієнти поліному визначалися на етапі отримання наближеного розв'язку, таким чином точність розв'язку задачі в цілому залежала від точності поліноміальної апроксимації. Метою роботи є дослідження впливу ступеня поліномів, які апроксимують залежність «зусилля – час» в елементах кородуючих конструкцій, та способу їх отримання на точність обчислення довговічності. В роботі запропоновано спосіб визначення коефіцієнтів апроксимуючих поліномів, який ґрунтується на використанні чисельно-аналітичного алгоритму розв'язання системи диференціальних рівнянь, що описують корозійний процес. Наведено результати чисельних експериментів, що ілюструють для різних конструкцій залежність похибки розв'язку задачі від ступеня апроксимуючих поліномів. Наведений підхід узгоджується зі стратегією нежорсткого допуску і дозволяє суттєво знизити обчислювальні витрати.

Зеленцов Дмитро Гегемонович – доктор технічних наук, професор, завідувач кафедри інформаційних систем, Український державний хіміко-технологічний університет.

Бричковський Олексій Дмитрович – аспірант кафедри інформаційних систем, Український державний хіміко-технологічний університет.

Dmytro Zelentsov – Doctor of Technical Sciences, Professor, Head of Department of Information Systems, Ukrainian State University of Chemical Technologies.

Brychkovskyi Oleksii – Postgraduate Student, Department of Information Systems, Ukrainian State University of Chemical Technologies.

ЗМІСТ

CONTENTS

Карпенко Н.В., Стаценко В.І., Федоряцька Н.І., Коломоєць І. І. Впровадження стартапів під час підготовки фахівців у галузі інформаційних технологій	3	Karpenko N.V., Statsenko V.I., Fedoriatska N.I., Kolomoiets I.I. Implementation of startups during the training of specialists in the field of information technology	3
Твердоступ М.І. Вимірювання складових індуктивного імпедансу на низьких частотах	12	Tverdostup N.I. Measurement of inductive components impedance at low frequencies	12
Пахомова В.М., Мегельбей Е.О. Визначення атак категорії U2R засобами SOM на основі бази даних NSL-KDD	18	Pakhomova V., Mehelbei Y. Detection of attacks of the u2r category by means of the SOM on database NSL-KDD "	18
Вовк С.М. Двоетапний метод оцінювання параметрів у невідомому шумовому середовищі	28	Vovk S.M. Two-stage method for parameter estimating in unknown noise environment	28
Ігнаткін В.У., Хомош Ю.С., Дудніков В.С., Юшкевич О.П., Тіхонов В.А. Оцінка ефективності системи автоматизованого управління метрологічним обслуговуванням засобів вимірювальної техніки – як інформаційно-довідкової системи колективного користування	42	Ignatkin V.U., Khomosh Y.S., Dudnikov V.S., Yushkevich O.P., Tikhonov V.A. Evaluation of efficiency of the automated management system of metrological maintenance of measuring equipment - as infor- matic information	42
Островська К.Ю., Фененко Т.М., Глущенко О.О. Статистичний аналіз тексту та дослідження динаміки точності класифікації	60	Ostrovska K., Fenenko T., Hlushchenko O. Statistical text analysis and study of the dynamics of classification accuracy	60

Тонкошкур А.С., Накашидзе Л.В., Лягшущин С.Ф., Колбунов В.Р. Аналіз особливостей функціонування обмежувачів напруги на основі комбінованої варисторно-позисторної структури	69	Tonkoshkur A.S., Nakashidze L.V., Lyagshushin S.F., Kolbunov V.R. Analysis of the features of the functioning of voltage limiters based on the combined varistor- posistor structure	69
Яковлєва М.А., Ковилін Є.Р. Інформаційна система проведення процесу liveness detection за допомогою AWS	84	Yakovlieva M.A., Kovylin Y.R. The information system for the liveness detection process using AWS	84
Сокол І.О., Волковський О.С. Лексичний аналіз програмного коду	95	Sokol I., Volkovskiy O. Lexical analysis of program code	95
Зеленцов Д.Г., Бричковський О.Д. Дослідження поліноміальної апроксимації зусиль в стержневих елементах кородуючих конструкцій	104	Zelentsov D.G., Brychkovskiy O.D. Research of polynomial approxi- mation of forces in rod elements of corroding structures	104

РЕФЕРАТИ

УДК 378.14

Карпенко Н.В., Стаценко В.І., Федоряцька Н.І., Коломоєць І.І. **Впровадження стартапів під час підготовки фахівців у галузі інформаційних технологій** // Системні технології. Регіональний міжвузівський збірник наукових праць. - Випуск 5(142). - Дніпро, 2022. - С.3 – 11.

У статті наведено погляд на освітній процес з точки зору різних стейкхолдерів – зі сторони освітян, роботодавців та зі сторони здобувачів вищої освіти. Порівняно розвиток стартап екосистеми в Україні та в Європі, показано їхній вплив на економічний зріст країни. Авторами запропоновано комплексний підхід до впровадження стартапів в освітній процес під час підготовки спеціалістів для ІТ-галузі та показано переваги, що отримує кожен зі стейкхолдерів від цього впровадження.

Бібл. 8, іл. 2.

УДК 621.382.08(035.5)

Твердоступ М.І. **Вимірювання складових індуктивного імпедансу на низьких частотах** // Системні технології. Регіональний міжвузівський збірник наукових праць. - Випуск 5(142). - Дніпро, 2022. - С.12 – 17.

Визначено можливості методу трьох вольтметрів для вимірювання індуктивності, активного опору, а також їх частотних характеристик. Встановлено, що в діапазоні частот до 100 кГц похибка вимірювання індуктивності становить не більше 0,3%, активного опору - 8,65%.

Бібл. 2, іл. 2, табл. 1

УДК 004.056.53:[004.7:004.032.26]

Пахомова В.М., Мегельбей Е.О. **Визначення атак категорії U2R засобами SOM на основі бази даних NSL-KDD** // Системні технології. Регіональний міжвузівський збірник наукових праць. - Випуск 5(142). - Дніпро, 2022. - С.18 – 27.

Створення ефективної системи виявлення мережевих атак вимагає застосування якісно нових підходів до обробки інформації, які повинні ґрунтуватися на адаптивних алгоритмах здатних до самонавчання. У якості методу дослідження використаний математичний апарат самоорганізуючої карти Кохонена. У якості програмної реалізації самоорганізуючої карти Кохонена використана мова Python з широким спектром сучасних стандартних засобів. На створеній програмній моделі «SOM_U2R» з використанням бази NSL-KDD проведено дослідження помилки за різною кількістю епох при різних розмірах карти. На моделі «SOM_U2R» проведено дослідження параметрів якості виявлення атак. Визначено, що на створеній програмній моделі «SOM_U2R» помилка другого роду склала 6 % для Buffer_overflow і Rootkit, 16 % для класу Loadmodule. Крім того, проведено дослідження F-мірки за різною кількістю епох навчання самоорганізуючої карти Кохонена. Визначено, що для всіх атак (крім Buffer_overflow) F-мірка зростає, досягаючи свого максимального значення (50 епох).

Бібл. 11, іл. 3.

УДК 004.02:519.254

Вовк С.М. **Двоетапний метод оцінювання параметрів у невідомому шумовому середовищі** // Системні технології. Регіональний міжвузівський збірник наукових праць. - Випуск 5(142). – Дніпро, 2022. – С.28 – 41.

Запропоновано двоетапний метод вирішення задачі оцінювання параметрів заданої моделі даних для випадку невідомого шумового середовища, який складається з етапів налаштування та оцінювання. На етапі налаштування оцінювач налаштовується на шумове середовище шляхом мінімізації середньоквадратичної помилки оцінювання для відомого фрагмента даних у просторі трьох вільних параметрів. На етапі оцінювання оцінювач розв'язує відповідну задачу мінімізації, використовуючи вже налаштовані вільні параметри. Наведено приклад оцінювання гауссівського імпульсу, який повільно рухається у невідомому шумовому середовищі.

Бібл. 19, іл. 2, табл. 1.

УДК 658.52.011.56:389.1/.14

Ігнаткін В.У., Хомош Ю.С., Дудніков В.С., Юшкевич О.П., Тіхонов В.А. **Оцінка ефективності системи автоматизованого управління метрологічним обслуговуванням засобів вимірювальної техніки – як інформаційно-довідкової системи колективного користування** // Системні технології. Регіональний міжвузівський збірник наукових праць. - Випуск 5(142). – Дніпро, 2022. – С.42 – 59.

У статті на прикладі системи автоматизованого управління метрологічним обслуговуванням засобів вимірювальної техніки (САУМОЗВТ) розглядається оцінка ефективності діалогової системи колективного користування. В якості критерія ефективності використовується відношення середньої реактивності системи до середніх витрат на один запит (С). Потік запитів у системі підпорядковується закону розподілення Пуассона. Для дослідження вибраний метод імітаційного моделювання.

Одержані: залежності (С) від часу поміж двома сусідніми запитами; залежність витрат на один запит від інтенсивності надходження запитів; залежність витрат на один запит від пропускнуої спроможності системи та інше.

Збільшення ефективності діалогових систем колективного користування можливо шляхом більш повного навантаження інформаційної системи з більшою потужністю. Наведено результати розрахунків економічної ефективності від запровадження САУМОЗВТ на кількох підприємствах та НДІ.

Бібл. 9, іл. 6, табл. 4.

УДК 004.896

Островська К.Ю., Фененко Т.М., Глущенко О.О. **Статистичний аналіз тексту та дослідження динаміки точності класифікації** // Системні технології. Регіональний міжвузівський збірник наукових праць. - Випуск 5(142). – Дніпро, 2022. – С.60 – 68.

Робота присвячена статистичному аналізу тексту та дослідженню динаміки точності класифікації. У роботі проводиться відбір статистичних ознак тексту, класифікація текстів, що належать різним авторам, та дослідження динаміки точності класифікації в залежності від

довжини текстових фрагментів. Для вирішення поставленого завдання використовувалися: методи обробки природної мови; статистичні характеристики текстів; методи машинного навчання; методи зниження розмірності для можливості візуалізації. На основі отриманої динаміки зміни точності класифікації в залежності від довжин текстових фрагментів було зроблено відповідні висновки щодо оптимальної довжини текстів, що використовуються для навчання та тестування моделей. Завдання вирішувалося у програмному середовищі Jupyter Notebook дистрибутива Anaconda, який дозволяє одразу встановити Python та необхідні бібліотеки.

Бібл. 6, іл. 8, табл. 1.

УДК 621.3

Тонкошкур А.С., Накашидзе Л.В., Лягшушин С.Ф., Колбунов В.Р. **Аналіз особливостей функціонування обмежувачів напруги на основі комбінованої варисторно-позисторної структури** // Системні технології. Регіональний міжвузівський збірник наукових праць. - Випуск 5(142). – Дніпро, 2022. – С.69 – 83.

В статті представлені результати моделювання та аналізу впливу особливостей електричного та теплового режиму на функціонування структури із послідовно з'єднаних та перебувають у тепловому контакті шарів варисторної кераміки та позисторного композиту, що становлять інтерес для розробки перспективного обмежувача постійних і повільно змінюваних напруг. Основну увагу приділено вивченню закономірностей поведінки функціональних електричних та теплових характеристик таких структур при зміні внутрішнього опору джерела вхідної напруги, інтенсивності теплообміну з довкіллям та її температури.

Бібл.24, іл. 7.

УДК 004.93

Яковлєва М.А., Ковилін Є.Р. **Інформаційна система проведення процесу liveness detection за допомогою AWS** // Системні технології. Регіональний міжвузівський збірник наукових праць. - Випуск 5(142). – Дніпро, 2022. – С.84 – 94.

Досліджено проведення відеоідентифікації особи за допомогою процесу Liveness Detection. Розроблено алгоритм проведення процесу Liveness Detection на основі виконання випадкових завдань. Проведено дослідження API Amazon DetectFaces, у процесі якого, емпіричним шляхом, були встановлені порогові значення характеристик обличчя на зображенні для проведення процесу Liveness Detection.

Бібл. 8, табл.3, рис. 3.

УДК 004.4'412

Сокол І.О., Волковський О.С. **Лексичний аналіз програмного коду** // Системні технології. Регіональний міжвузівський збірник наукових праць. - Випуск 5(142). – Дніпро, 2022. – С.95 – 103.

Розглядається визначення лексичного аналізатора, його функціональний склад та принципи побудови, наведено ключові відмінності між лексею та токеном. Для рішення задачі пошуку токенів у програмному коді були запропоновані та розглянуті два підходи: пошук на основі регулярних виразів та пошук за допомогою кінцевих автоматів. Для даних підходів бу-

ли приведені приклади формування шаблонів пошуку за визначених правил лексики. В результаті була обґрунтована оптимальність використання детермінованих кінцевих автоматів під час практичної реалізації лексичного аналізатора на реальних обчислювальних машинах.

Бібл. 8.

УДК 004.94

Зеленцов Д.Г., Бричковський О.Д. **Дослідження поліноміальної апроксимації зусиль в стержневих елементах кородуючих конструкцій** // Системні технології. Регіональний міжвузівський збірник наукових праць. - Випуск 5(142). – Дніпро, 2022. – С.104 – 115.

У роботі розглядається проблема поліноміальної апроксимації залежностей «зусилля – час» в елементах конструкцій шарнірно-стержневих конструкцій з точки зору впливу ступеня полінома на похибку обчислення їх довговічності. Запропоновано спосіб визначення коефіцієнтів апроксимуючих поліномів, який ґрунтується на використанні чисельно-аналітичного алгоритму розв'язання системи диференціальних рівнянь, що описують корозійний процес. Наведено результати чисельних експериментів, що ілюструють для різних конструкцій залежність похибки розв'язку задачі від ступеня апроксимуючих поліномів.

Бібл. 5, іл. 4, табл. 1.

UDC 378.14

Karpenko N.V., Statsenko V.I., Fedoriatska N.I., Kolomoiets I.I. **Implementation of startups during the training of specialists in the field of information technology** // System technologies. N 5(142) - Dnipro, 2022.- P.3 – 11.

Now, the training of a specialist is considered as an output. When students enter higher education, the curriculum is drawn up for them, which is strictly followed throughout subsequent years, for example, four years for the bachelor. However, this approach does not suit motivated students who lack creativity along with actual knowledge. Thus, these students, in parallel with higher education, attend special courses in IT-companies. At the end of the course, students start working for an IT company and lose motivation for further training at the university.

The authors of the article offer a comprehensive approach to the introduction of startups in the educational process of universities during the training of specialists for the IT-field. Attraction to the educational process of stakeholders from IT-companies to solve actual problems will lead to the fact that students will be involved in the creative process, in which they will receive useful knowledge and skills not only in the specialty, but also learn to communicate in a team. The introduction of startups will lead to an increase in the number of motivated students, because students can for creativity to choose not only topics proposed by stakeholders, but also to offer their ideas. Close cooperation with stakeholders from the IT-field will lead to an improvement in the quality of the materials that teachers will offer students for training and will also allow universities to earn on author courses and textbooks written by teachers. The authors of the article propose, together with start-ups to introduce flexible planning of the educational process for students who take part in start-ups, i.e. each motivated student will be able to engage in their educational trajectory and after graduation to obtain not only a diploma, But also certificates that will confirm the skills gained in start-ups. The article also indicates other benefits that each of the stakeholders will receive from the introduction of startups in the learning process.

Bibl. 8, ill. 3.

UDC 621.382.08(035.5)

Tverdostup N.I. **Measurement of inductive components impedance at low frequencies** // System technologies. N 5(142) - Dnipro, 2022.- P.12 – 17.

The possibilities of the method three voltmeters to measure inductance, a resistance, and their frequency characteristics. It is found that in the frequency range up to 100 kHz, the inductance of the measurement error is not more than 0.3%, resistance - 8.65%.

Bibl. 2, ill. 2, tabl. 1

UDC 004.056.53:[004.7:004.032.26]

Pakhomova V., Mehelbei Y. **Detection of attacks of the U2R category by means of the SOM on database NSL-KDD** // System technologies. N 5(142) - Dnipro, 2022.- P.18 – 27.

Creating an effective system for detecting network attacks requires the use of qualitatively new approaches to information processing, which should be based on adaptive algo-

rithms capable of self-learning. The mathematical apparatus of the Kohonen self-organizing map (SOM) was used as a research method. Python language with a wide range of modern standard tools was used as a software implementation of the Kohonen SOM addition, this section compiles the Python software model «SOM_U2R» using a Kohonen SOM. Created «SOM_U2R» software model on database NSL-KDD an error research was performed for different number of epochs with different map sizes. On the «SOM_U2R» model the research of parameters of quality of detection of attacks is carried out. It is determined that on the «SOM_U2R» created software model the error of the second kind of detection of network classes of attacks Buffer_overflow and Rootkit is 6 %, and for the class Loadmodule reached 16 %. In addition, a survey of the F-measure was conducted for a different number of epochs of learning the Kohonen SOM. It is determined that for all network attack classes (except Buffer_overflow) the F-measure increases, reaching its maximum value at 50 epochs.

Bibl. 11, ill. 4.

UDC 004.02:519.254

Vovk S.M. **Two-stage method for parameter estimating in unknown noise environment** // System technologies. N 5(142) - Dnipro, 2022.- P.28 – 41.

A two-stage method consisting of the tuning and estimating stages is proposed for solving the parameter estimation problem in the case of an unknown noise environment. At the tuning stage, the estimator is tuned to the noise environment by minimizing the estimation root-mean-square error for a known data fragment in the space of three free parameters. At the estimating stage, the estimator solves the corresponding minimization problem using already tuned free parameters. An example of estimating a Gaussian pulse, which slowly moves in the unknown noise environment, is presented.

Bibl. 19, fig. 2, tabl. 1.

UDC 658.52.011.56:389.1/.14

Ignatkin V.U., Khomosh Y.S., Dudnikov V.S., Yushkevich O.P., Tikhonov V.A. **Evaluation of efficiency of the automated management system of metrological maintenance of measuring equipment - as informatic information** // System technologies. N 5(142) - Dnipro, 2022.- P.42 – 59.

In the article on the example of the system of automated control of metrological service of measuring equipment (SACMMME) the estimation of efficiency of dialogue system of collective use is considered. The efficiency criterion is the ratio of the average reactivity of the system to the average cost per request (C). The flow of queries in the system obeys the Poisson distribution law. The method of simulation modeling is chosen for research.

Obtained: time dependences (C) between two adjacent queries; the dependence of the cost of one request on the intensity of requests; the dependence of the cost of one request on the bandwidth of the system and more.

Increasing the efficiency of shared dialog systems is possible by more fully loading the information system with more capacity. The results of calculations of economic efficiency from the introduction of SACMMME at several enterprises and research institutes are given.

Bibl. 9, ill. 6, table. 4.

UDC 004.896

Ostrovska K., Fenenko T., Hlushchenko O. **Statistical text analysis and study of the dynamics of classification accuracy** // System technologies. N 5(142) - Dnipro, 2022.- P.60 – 68.

The work is devoted to the statistical analysis of the text and the study of the dynamics of classification. In the work, the selection of statistical features of the text, the classification of texts belonging to different authors, and the study of the dynamics of classification accuracy depending on the length of text fragments are carried out. To solve the problem, the following methods were used: natural language processing methods; statistical characteristics of texts; machine learning methods; dimensionality reduction methods for visualization capability. On the basis of the obtained dynamics of changes in classification accuracy depending on the lengths of text fragments, appropriate conclusions were drawn regarding the optimal length of texts used for training and testing models. The task was solved in the Jupyter Notebook software environment of the Anaconda distribution, which allows you to immediately install Python and the necessary libraries.

Bibl. 6, ill. 8, table. 1.

UDC 621.3

Tonkoshkur A.S., Nakashidze L.V., Lyagshushin S.F., Kolbunov V.R. **Analysis of the features of the functioning of voltage limiters based on the combined varistor-posistor structure** // System technologies. N 5(142) - Dnipro, 2022.- P.69 – 83.

The article presents the results of modeling and analysis of the influence of the features of the electrical and thermal regime on the functioning of the structure of sequentially connected and thermally contacting layers of varistor ceramics and posistor composite, which are of interest for the development of a promising limiter of constant and slowly changing voltages. The main attention is paid to the study of the patterns of behavior of the functional electrical and thermal characteristics of such structures when the internal resistance of the input voltage source changes, the intensity of heat exchange with the environment and its temperature.

Bibl. 24, ill. 7

UDC 004.93

Yakovlieva M.A., Kovylin Y.R. **The information system for the liveness detection process using AWS** // System technologies. N 5(142) - Dnipro, 2022.- P.84 – 94.

Video identification of a person using the Liveness Detection process was investigated. An algorithm for carrying out the Liveness Detection process based on the performance of the random tasks has been developed. Research of the Amazon DetectFaces API was conducted, in the process of which, empirically, the threshold values of the characteristics of the face in the image were established for the Liveness Detection process.

Bibl. 8, tab. 3, fig. 3.

UDC 004.4'412

Sokol I., Volkovskiy O. **Lexical analysis of program code** // System technologies. N 5(142) - Dnipro, 2022.- P.95 – 103.

The growing volume of technologies, the end of actively used development tools support, outdated API etc., entails the need of program codes conversion. In IT companies and not only, often begged the question of deprecated software support, which customers continue to use, or translation of current software to actual technologies. It is more rational for programmers to use the conversion and save most of code base, than rewriting all software by hand, even if manual adjustment is needed. At this moment, there are few high-quality code conversion systems. Largely, conversion systems work well only with similar programming languages. The task of program codes conversion is a deep and complex topic. To convert the software code, you must first analyze, select components and form a structural representation. Any analysis of program code begins with lexical analysis. Although lexical analysis is considered a relatively simple step, it plays a key role in the entire system of analysis and transformation of software code, and also has a large number of both theoretical and practical features that require careful study. This article considers the definition of the lexical analyzer, its functional composition and principles of construction, provides key differences between the lexeme and the token. Two approaches have been proposed and considered to solve the search for tokens in the program code: regular expression search and finite state machine search. For these approaches, examples of the formation of search templates under certain rules of vocabulary were given. As a result, the optimality of the use of deterministic finite state machines during the practical implementation of the lexical analyzer on real computing machines was substantiated.

Bibl. 8.

UDC 004.94

Zelentsov D.G., Brychkovskiy O.D. **Research of polynomial approximation of forces in rod elements of corroding structures** // System technologies. N 5(142) - Dnipro, 2022.- P.104 – 115.

The paper considers the problem of polynomial approximation of the "force - time" dependencies in the elements of corroding hinged-rod structures from the point of view of the influence of the degree of the polynomial on the error in calculating their durability. A method for determining the coefficients of approximating polynomials is proposed, which is based on the use of a numerical-analytical algorithm for solving a system of differential equations describing the corrosion process. The results of numerical experiments are presented, illustrating for various constructions the dependence of the error in solving the problem on the degree of approximating polynomials.

Bibl. 5, fig. 4, tabl. 1.

Системні технології
ЗБІРНИК НАУКОВИХ ПРАЦЬ

Випуск 5 (142)

Головний редактор: к.т.н., доц. Т.В. Селівьорстова

Технічний редактор та секретар збірки: к.т.н., доц. К.Ю. Островська

Здано до набору 25.10.2022. Підписано до друку 28.10.2022.

Формат 60x84 1/16. Друк - різнограф. Папір типограф.

Умов. друк арк. – 8,929. Обл.-видавн. арк. – 7,813.

Тираж 300 прим. Замовл. – 05/22

Український державний університет науки і технологій,
ННІ «Інститут промислових та бізнес технологій»,
кафедра Інформаційних технологій та систем: ІВК «Системні технології»
49600, Дніпро, а/с 493

<http://journals.nmetau.edu.ua/index.php/st>

Свідоцтво про державну реєстрацію друкованого засобу масової інформації:

Серія КВ № 8684 від 23 квітня 2004 рік

Редакційна колегія

Селівьорстова Тетяна Віталіївна
(*головний редактор*)

доцент, кандидат технічних наук

Алпатов Анатолій Петрович

Член-кореспондент НАН України,
професор, доктор технічних наук

Архипов Олександр Євгенійович

професор, доктор технічних наук

Бабічев Сергій Анатолійович

доцент, доктор технічних наук

Білозьоров Василь Євгенович

професор,

доктор фізико-математичних наук

Гече Федір Елемирович

професор, доктор технічних наук

Гуда Антон Ігорович

(*заст. головного редактора*)

професор, доктор технічних наук

Гнатюшенко Вікторія Володимирівна

(*вчений секретар*)

професор, доктор технічних наук

Гнатюшенко Володимир Володимирович

професор, доктор технічних наук

Гожий Олександр Петрович

професор, доктор технічних наук

Єрьомін Олександр Олегович

професор, доктор технічних наук

Кіріченко Людмила Олегівна

професор, доктор технічних наук

Світличний Дмитро Святозарович

професор, доктор технічних наук

Скалозуб Владислав Васильович

професор, доктор технічних наук

Хандецький Володимир Сергійович

професор, доктор технічних наук

Український державний університет науки і технологій, ННІ «Інститут промислових та бізнес технологій», Україна

Інститут технічної механіки

НАНУ і ДКАУ, Україна

Національний технічний університет

України «Київський політехнічний інститут» імені Ігоря Сікорського», Україна

Jan Evangelista Purkyně University
in Ústí nad Labem

Університет імені Яна Євангеліста Пуркіне,
Усті над Лабем, Чеська Республіка

Дніпровський національний університет
імені Олеся Гончара, Україна

Ужгородський національний університет,
Україна

Український державний університет науки і технологій, ННІ «Інститут промислових та бізнес технологій», Україна

Український державний університет науки і технологій, ННІ «Інститут промислових та бізнес технологій», Україна

Національний технічний університет
«Дніпровська політехніка», Україна

Чорноморський національний університет
імені П.Могили, Україна

Український державний університет науки і технологій, ННІ «Інститут промислових та бізнес технологій», Україна

Харківський національний університет
радіоелектроніки, Україна

Akademia Górniczo-Hutnicza

Краківська гірничо-металургійна академія
ім. С. Сташціа, Польща

Український державний університет науки і технологій, ННІ «Дніпровський інститут інфраструктури і транспорту»Україна

Дніпровський національний університет
імені Олеся Гончара, Україна