

ДОСЛІДЖЕННЯ ІНТЕЛЕКТУАЛЬНИХ МОДЕЛЕЙ КЛАСИФІКАЦІЇ НЕВИЗНАЧЕНИХ ДАНИХ З ВИМОГАМИ ДОСТОВІРНОСТІ РЕЗУЛЬТАТІВ

Скалозуб В.В., Горячкін В.М., Терлецький І.А.

Український державний університет науки і технологій, Україна

Анотація. У доповіді приведені результати досліджень та розвитку інтелектуальних моделей управління складними системами за умов невизначеності даних на основі процедур класифікації, які забезпечують достовірне вирішення завдань з урахуванням оцінки граничної розмірності моделей. Досліджені можливості удосконалення нейронних мереж Хеммінга для класифікації даних у форматах нечітких величин і *certainty factor CF(A)*. Визначені особливості математичної моделі завдань класифікації на основі набору шаблонів ознак. Приведено програмний комплекс інформаційної технології управління призначенням/відбором виконавців, а також визначення авторства україномовних творів на основі класифікації наборів шаблонів із певних нечітких ознак. Програмний комплекс використовує запропоновані у доповіді процедури редукції і статистики каппа Коена.

Ключові слова: класифікація наборів шаблонів, невизначені дані, редукція розмірності, статистика каппа Коена, нечіткі величини, *certainty factor CF(A)*, модифікована мережа Хеммінга, інформаційна технологія, програмне забезпечення, завдання визначення авторства та призначення/відбір виконавця.

Для різноманітних складних систем існує необхідність у застосуванні процедур класифікації та діагностування неповних даних, щоб вирішувати завдання вибору керування технологічними процесами, відбору виконавців завдань, визначення авторства творів тощо. З цим пов'язані проблеми формування математичних моделей класифікації та перевірки їхньої коректності та достовірності. Представлена доповідь присвячена дослідженню і розвитку інтелектуальних моделей управління складними системами в умовах невизначеності даних, зокрема за допомогою процедур класифікації, з використанням методів редукції (ПКР) та каппа-статистики Коена. Використання цих методів гарантує достовірні результати класифікації з урахуванням граничних розмірностей моделей. Виконання процедури редукції на основі каппа статистики має такі етапи: - 1) для моделей шаблонів класів з різним числом або складом параметрів необхідно визначити оцінку показника

«каппа», - 2) якщо результати класифікації для різних моделей-шаблонів відповідають вимогам «подібності», скоротити модель, залишити один із шаблонів, - 3) враховуючи граничне значення розмірності моделі класифікації « n_0 » (розрахованої за вихідними експериментальними даними) визначити і видалити найменше значимі чи найбільше «подібні» між собою параметри моделі класифікації. Повторювати процедуру ПКР до формування розмірності модифікованої моделі « n_0 », для якої була задана оцінка ймовірності безпомилкового розділення випадкової і незалежної вибірки довжини « L » при встановленій граничній величині помилкової класифікації « ε ».

Сутність та достовірність результатів ПКР визначається через показник «каппа» статистики Коена. За ним встановлюється «подібність», або належність конкуруючих багатопараметричних математичних моделей класифікації до одного типу, категорії. Для такої перевірки «подібності» необхідно утворити певну множину математичних моделей, шаблонів із різними наборами змінних, для яких в подальшому виконувати порівняльний аналіз їх відповідності між собою у відповідності до процедури ПКР.

У доповіді також розглядаються можливості вдосконалення нейронних мереж Хеммінга для класифікації даних у формі нечітких величин і certainty factor CF(A). У дослідженнях були виявлені особливості запропонованої моделі нечіткої класифікації на основі набору ознак, які визначають класи аналізованих об'єктів.

У роботі сформовані та досліджені нові математичні моделі класифікації для визначення авторства україномовних творів (ЗАТ), детально розглядаються їх властивості. Особливості завдання ЗАТ та його реалізації на основі моделі нечіткої класифікації полягають у відсутності вимог щодо числа етапів процедури прийняття рішень, а також не обов'язковості формування єдиної моделі класифікації. Такі особливості процедур класифікації були враховані при формуванні процедури редукції та каппа-статистики Коена.

Для вирішення завдань класифікації параметрів складних систем за умов невизначеності даних було розроблено відповідне програмне забезпечення. У доповіді приведено структуру програмного комплексу інформаційної

технології управління призначенням/відбором виконавців, а також завданнями із визначення авторства україномовних творів, на основі класифікації наборів шаблонів із певних нечітких ознак. Програмний комплекс використовує процедури редукції та каппа-статистики.

ЛІТЕРАТУРА

1. Скалозуб В.В., Горячкін В.М., Терлецький І.А., Дудник І.П. Формування моделей класифікації невизначених даних процедурами редукції і каппа статистики. Системні технології. – Дніпро, УДУНТ, 2023. – № 5 (148). – С. 141-155. DOI: 10.34185/1562-9945-5-148-2023-13
2. Колесник А. С., Хайрова Н. Ф. Обґрунтування використання статистики каппа Коена в експериментальних дослідженнях NLP Text Mining. Кібернетика та системний аналіз. Т. 58, № 2. 2022. С. 143–153.

RESEARCH OF INTELLECTUAL MODELS FOR CLASSIFICATION OF UNCERTAIN DATA WITH REQUIREMENTS FOR RESULT RELIABILITY

Skalozub V., Horyachkin V., Terletsky I.A.

Abstract. *The report contains the results of research and development of intellectual models for managing complex systems under conditions of data uncertainty based on classification procedures, which ensure reliable solution of tasks considering the assessment of the boundary dimensionality of models. The possibilities of improving Hamming neural networks for data classification in formats of fuzzy variables and certainty factor $CF(A)$ are investigated. The features of the mathematical model of classification tasks based on a set of feature templates are determined. A software complex of information technology for assignment/selection of performers, as well as determination of authorship of Ukrainian-language works based on classification of sets of templates with certain fuzzy features, is presented. The software complex utilizes the reduction and Cohen's kappa statistics procedures proposed in the report.*

Keywords: *template set classification, uncertain data, dimensionality reduction, Cohen's kappa statistics, fuzzy variables, certainty factor $CF(A)$, modified Hamming network, information technology, software, authorship determination task and performer assignment/selection.*

REFERENCES

1. Skalozub V.V., Horyachkin V.M., Terletsky I.A., Dudnik I.P. Formation of classification models of uncertain data using reduction and kappa statistics procedures. System Technologies. – Dnipro, USUST, 2023. – № 5 (148). – P. 141-155. DOI: 10.34185/1562-9945-5-148-2023-13
2. Kolesnyk A.S., Khayrova N.F. Justification of using Cohen's kappa statistics in experimental research of NLP Text Mining. Cybernetics and Systems Analysis. Vol. 58, No. 2. 2022. P. 143–153.