DOI 10.34185/1562-9945-2-127-2020-01
УДК 004

С.М. Вовк

ІНФОРМАЦІЙНА ТЕХНОЛОГІЯ ЗГЛАДЖУВАННЯ ДАНИХ НА ОСНОВІ КРИТЕРІЮ МІНІМУМУ ПРОТЯЖНОСТІ

Анотація. Запропонована інформаційна технологія згладжування даних, спотворених аномальними значеннями й шумом. У рамках даної технології передбачається, що модель даних задана непараметрично за допомогою властивості її гладкості, аномальні значення мають вигляд викидів або коротких за протяжністю аномальних фрагментів даних, а шум проявляється в кожному елементі даних у вигляді випадкової добавки до його значення. Дана технологія реалізується шляхом розв'язання задачі мінімізації цільової функції, побудованої на основі критерію мінімуму протяжності, який застосовується до відхилу розв'язку, і на обмеженні величини похідної розв'язку заданого порядку. Рішення задачі мінімізації досягається чисельно шляхом застосування методу спряжених градієнтів. Керування процесом згладжування даних здійснюється за допомогою параметрів налаштування, значення яких можуть встановлюватися вручну або автоматично. Пропонована технологія випробувана на даних, отриманих за допомогою чисельного моделювання, а також на експериментальних даних, що являють собою спектри фотолюмінесценції.
Ключові слова: інформаційна технологія, згладжування даних.

Вступ. Для багатьох прикладних додатків актуальною є задача згладжування даних, отриманих в умовах наявності шуму й аномальних значень [1]. Складність розв'язання цієї задачі обумовлена тим, що параметрична модель згладжуваних даних звичайно є невідомою, а наявність аномальних значень може стати причиною суттєвих помилок [2]. За цих умов для практичної роботи з обробки даних можна залучити досвідченого дослідника, здатного вручну виконати необхідне згладжування даних. На противагу цьому рутинному й суб'єктивному підходу в даній роботі пропонується інформаційна технологія обробки даних, яка призначена для автоматичного видалення з отриманих даних аномальних значень і шуму. Дана технологія заснована на критерії мінімуму протяжності [3], який застосовується до відхилу розв'язку, та на вимозі гладкості розв'язку. У межах даної технології участь людини є потрібною тільки на етапі налаштування початкових параметрів роботи.

Постановка задачі та мета досліджень. Постановка задачі полягає в розробці інформаційної технології згладжування даних на основі критерію мінімуму протяжності й вимоги гладкості шуканого розв'язку. Метою досліджень є отримання інформаційної технології згладжування даних, спотворених аномальними значеннями й шумом.

Аналіз останніх досліджень і публікацій. Згладжування даних є процесом їх апроксимації гладкою функцією. Поняття гладкої функції означає, що вона має неперервну похідну в усій області її визначення. Якщо для заданої функції існують її неперервні похідні до r-го порядку включно, то говорять про гладку функцію з порядком гладкості r.

Існують різні підходи до розв'язання задачі згладжування даних, які умовно можна поділити на локальні й глобальні [4]. Локальні підходи до згладжування даних використовують розбиття вхідної послідовності даних на набір підпослідовностей, для кожної з яких застосовується та або інша модель апроксимації. Наприклад, відомий метод згладжування Савицького-Голея [5] у якості моделі даних використовує поліном невисокого порядку й вікно невеликих розмірів, яке ковзає вхідною послідовністю даних, виробляючи згладжені значення. Основна перевага локальних методів згладжування полягає в тому, що вони забезпечують високу швидкість обробки даних, а їх недолік полягає в тому, що вони забезпечують тільки локальну гладкість апроксимуючої функції [6].

Глобальні методи згладжування забезпечують гладкість апроксимуючої функції на всьому інтервалі спостереження даних. Одним з них є метод згладжування на основі регуляризації Тихонова, який можна подати задачею мінімізації [7]:

$$\min_f \{ \| f - g \|_2^2 + \gamma \| D_r f \|_2^2 \}, \tag{1}$$

де g позначає вхідні дані, f позначає апроксимуючу функцію, D_r є оператор диференціювання r-го порядку, γ є параметр регуляризації, а $\| \ldots \|_2^2$ позначає квадрат евклідової норми. З (1) видно, що в цій задачі пе-

ISSN 1562-9945 (Print)
ISSN 2707-7977 (Online)

редбачається існування похідних апроксимуючої функції аж до r-го порядку включно. Розв'язок задачі (1) у матричній формі запису є таким:

$$\mathbf{f} = (\mathbf{I} + \gamma \mathbf{D}_r^T \mathbf{D}_r)^{-1} \mathbf{g} \,, \tag{2}$$

де $\mathbf{g}$ - вектор вхідних даних, що складається з N елементів, $\mathbf{f}$ - шуканий вектор згладжених даних, $\mathbf{I}$ - тотожна (одинична) матриця розміром $N \times N$. Оператор D_r для різних r описують такі матриці: для $r = 0$ маємо матрицю $\mathbf{D}_0 = \mathbf{I}$ розміром $N \times N$; для $r = 1$ та $r = 2$ маємо матриці $\mathbf{D}_1 = \begin{vmatrix} -1 & 1 & 0 & 0\ldots \\ 0 & -1 & 1 & 0\ldots \\ 0 & 0 & -1 & 1\ldots \\ \ldots & \ldots & \ldots & \ldots\ldots \end{vmatrix}$ розміром $(N-1) \times N$ та $\mathbf{D}_2 = \begin{vmatrix} 1 & -2 & 1 & 0\ldots \\ 0 & 1 & -2 & 1\ldots \\ 0 & 0 & 1 & -2\ldots \\ \ldots & \ldots & \ldots & \ldots\ldots \end{vmatrix}$ розміром $(N-2) \times N$, відповідно, і т.д. Проте матриця $\mathbf{D}_r^T \mathbf{D}_r$; r=1,2,... завжди має розмір $N \times N$. Перевага (2) полягає в тому, що розв'язання задачі згладжування досягається за один крок шляхом множення оберненої матриці $(\mathbf{I} + \gamma \mathbf{D}_r^T \mathbf{D}_r)^{-1}$ на вектор вхідних даних $\mathbf{g}$. Недоліки (2) полягають у необхідності обертати матрицю великого розміру тоді, коли значення N є великим, у зростанні числа обумовленості матриці зі збільшенням r та γ, а також у неефективній обробці викидів, які внаслідок згладжування розмиваються на сусідні елементи даних, формуючи на згладженій залежності "горби" і "ями".

У [8] наведено метод чисельного диференціювання даних за умови наявності шуму й викидів. Він заснований на постановці задачі:

$$\min_u \{\Psi[Au - g] + \gamma \,\|u'\|_2^2\}, \tag{3}$$

де A є оператор антидиференціювання, u є похідна функції f, що апроксимує вхідні дані g, функція u' є похідна u, а неквадратичний функціонал Ψ реалізує концепцію "квазіпротяжність функції" на основі вартісної функції [9]:

$$\psi^{(\alpha,\beta,q)}(x) = k^{(\alpha,\beta,q)}[(1 + |x|^q / \alpha^q)^{\beta/q} - 1]; \quad |x| < \infty\,, \tag{4}$$

де $\alpha > 0$; $-\infty < \beta \le 1$; $0 < q < \infty$; $\beta < q$; $k^{(\alpha,\beta,q)} = 1/[(1 + |x_0|^q / \alpha^q)^{\beta/q} - 1]$; $\psi^{(\alpha,\beta,q)}(x_0) = 1$; $x_0 = 1$. Цей метод демонструє гарну якість результатів чисельного диференціювання даних за умов наявності шуму й викидів і може бути основою інформаційної технології згладжування даних. Але

ISSN 1562-9945 (Print)
ISSN 2707-7977 (Online)

зважаючи на особливості задачі, що розлядається, його доцільно переформулювати відносно апроксимуючої функції f, видаляючи з розгляду оператор A і спосіб отримання апроксимуючої функції за її похідною [8].

Основна частина. Пропонована постановка задачі згладжування даних на основі критерію мінімуму протяжності й обмеження величини похідної заданого порядку $r > 0$ має вигляд:

$$\min_{f}\{\Psi[f-g]+\gamma\,\|D_r f\|_2^2\}. \tag{5}$$

Відмінність (5) від (3) полягає в тому, що для (5) шуканою є апроксимуюча функції f, а не її похідна. Підставляючи (4) в (5), одержуємо розгорнутий запис постановки задачі згладжування у вигляді:

$$\min_{f_1,\ldots,f_N}\left\{k^{(\alpha,\beta,q)}\sum_{j=1}^{N}[(1+|f_j-g_j|^q/\alpha^q)^{\beta/q}-1]+\gamma\sum_{j=1}^{N}|\sum_{n=1}^{N}d_{jn}f_n|^2\right\}, \tag{6}$$

де d_{jn} є елемент матриці $\mathbf{D}_r$, яка відповідає оператору диференціювання r-го порядку D_r. Оскільки $k^{(\alpha,\beta,q)} > 0$ для $0 < \beta \le 1$ і $k^{(\alpha,\beta,q)} < 0$ для $-\infty < \beta < 0$, то для випадку $\beta \ne 0$ після відкидання несуттєвих коефіцієнтів задачу (6) можна записати у вигляді:

$$\min_{f_1,\ldots,f_N}\left\{\operatorname{sgn}(\beta)\sum_{j=1}^{N}[(1+|f_j-g_j|^q/\alpha^q)^{\beta/q}]+\gamma_1\sum_{j=1}^{N}|\sum_{n=1}^{N}d_{jn}f_n|^2\right\}, \tag{7}$$

де $\gamma_1 = \gamma\cdot|(1+|x_0|^q/\alpha^q)^{\beta/q}-1|$ та $\operatorname{sgn}(\beta)=1$ для $0<\beta\le 1$ і $\operatorname{sgn}(\beta)=-1$ для $-\infty<\beta<0$. Для випадку $\beta=0$ з (6) шляхом граничного переходу за $\beta\to 0$ одержуємо задачу

$$\min_{f_1,\ldots,f_N}\left\{\sum_{j=1}^{N}\ln[1+|f_j-g_j|^q/\alpha^q]+\gamma_2\sum_{j=1}^{N}|\sum_{n=1}^{N}d_{jn}f_n|^2\right\}, \tag{8}$$

де $\gamma_2 = \gamma\cdot\ln(1+|x_0|^q/\alpha^q)$. Згідно (6), процесом згладжування можна керувати за допомогою параметрів α, β, q, γ та r. При цьому α, β й q слід використовувати для налаштування на поточне шумове оточення, γ – для встановлення компромісу між квазіпротяжністю відхилу й величиною похідної заданого порядку гладкості, а r – для бажаного порядку гладкості розв'язку. Налаштування цих параметрів може бути виконане або вручну, або автоматично шляхом встановлення рекомендованих значень. Одним з рекомендованих наборів значень є такий: $\alpha=100$, $\beta=0$,

ISSN 1562-9945 (Print)
ISSN 2707-7977 (Online)

$q=2$, $\gamma=10000$, $k=2$. Цей набір був отриманий в [8] для випадку, коли значення елементів даних були подані 16-а двійковими розрядами, з яких корисні значення займали діапазон від 1 до 13 розрядів, значення адитивного шуму – від 1 до 4 розрядів, значення адитивного шуму, що залежить від сигналу, – від 1 до 6 розрядів, а значення викидів – увесь діапазон від 1 до 16 розрядів.

Для чисельного розв'язання задачі (6) доцільно використати метод спряжених градієнтів, який подається стандартною схемою: $\mathbf{f}^{(t+1)}=\mathbf{f}^{(t)}+h^{(t)}\mathbf{p}^{(t)}$, $t\geq 0$; $\mathbf{p}^{(0)}=-\mathbf{grad}^{(0)}$; $\mathbf{p}^{(t)}=-\mathbf{grad}^{(t)}+b^{(t-1)}\mathbf{p}^{(t-1)}$, $t\geq 1$; $b^{(t-1)}=\|\mathbf{grad}^{(t)}\|^2/\|\mathbf{grad}^{(t-1)}\|^2$; $h^{(t)}=\arg\min\limits_{h\geq 0}\Phi(\mathbf{f}^{(t)}+h\mathbf{p}^{(t)})$, де $\mathbf{f}^{(0)}=\mathbf{g}$, а ціле число t задає номер ітерації. У цій схемі одновимірна цільова функція $\Phi(\mathbf{f}^{(t)}+h\mathbf{p}^{(t)})$ залежить від h і отримується підстановкою елементів вектора $\mathbf{f}^{(t)}+h\mathbf{p}^{(t)}$ у вираз, записаний у фігурних дужках задачі (6). Далі представлено випадок побудови вектора градієнта для $\beta=0$ (задача (8)) і $q=2$, коли забезпечується оптимальна (за критерієм максимальної правдоподібності) обробка даних, спотворених шумом Коші, а також застосовуваний метод мінімізації одновимірної цільової функції.

З необхідної умови мінімуму (8) для випадку $q=2$ одержуємо систему з N нелінійних рівнянь

$$\frac{(f_m^{(t)}-g_m)}{(f_m^{(t)}-g_m)^2+\alpha^2}+\gamma_2\sum_{n=1}^{N}\hat{d}_{mn}f_n^{(t)}=0;\quad m=1,\dots,N, \tag{9}$$

де $\hat{d}_{mn}=\sum\limits_{i=1}^{N}d_{mi}^{T}d_{in}$ є елементом матриці $\mathbf{D}_r^T\mathbf{D}_r$. Ліва частина (9) є m-м елементом вектора градієнта $\mathbf{grad}^{(t)}$ для задачі (8).

Рішення одновимірної задачі: $h^{(t)}=\arg\min\limits_{h}\Phi(\mathbf{f}^{(t)}+h\mathbf{p}^{(t)})$ полягає у виборі кроку h уздовж напрямку спуску $\mathbf{p}^{(t)}$. Оскільки функція Φ не є унімодальною, то доцільно використати метод пасивного пошуку її мінімуму на множині "пробних кроків". Ця множина складається з кроків, які скидають в нуль компоненти відхилу розв'язку й компоненти стабілізуючої частини функціонала, а також з кроку за методом Ньютона.

Кроки, які скидають в нуль компоненти відхилу розв'язку, для випадку $\beta=0$ й $q=2$ визначаються формулою:

ISSN 1562-9945 (Print)
ISSN 2707-7977 (Online)

$$h_j^{[1]} = (g_j - f_j^{(t)}) \big/ p_j^{(t)}; \quad j = 1,\ldots,N. \tag{10}$$

Кроки, які скидають в нуль компоненти стабілізуючої частини функціонала, отримуються з умови рівності нулю цих компонентів. Розглядаючи стабілізуючу частину функціонала в (8) як скалярний добуток, з умови рівності нулю кожного з компонентів цього скалярного добутку $(\mathbf{D}_k^T \mathbf{D}_k (f^{(t)} + h\mathbf{p}^{(t)}), (f^{(t)} + h\mathbf{p}^{(t)}))$ одержуємо два набори із N кроків:

$$h_n^{[2]} = -f_n^{(t)} \big/ p_n^{(t)}; \quad n = 1,\ldots,N, \tag{11}$$

$$h_n^{[3]} = -\sum_{j=1}^{N} \hat{d}_{nj} f_j^{(t)} \Bigg/ \sum_{j=1}^{N} \hat{d}_{nj} p_j^{(t)}; \quad n = 1,\ldots,N, \tag{12}$$

Крок, який відповідає методу Ньютона, використовується за умови, що розв'язок перебуває в околиці локального мінімуму. Цей крок для $\beta = 0$ й $q = 2$ задається формулою:

$$h^{[4]} = -(\mathbf{grad}^{(t)}, \mathbf{p}^{(t)}) \big/ ([\mathbf{Q} + \gamma_2 \mathbf{D}_k^T \mathbf{D}_k]\mathbf{p}^{(t)}, \mathbf{p}^{(t)}), \tag{13}$$

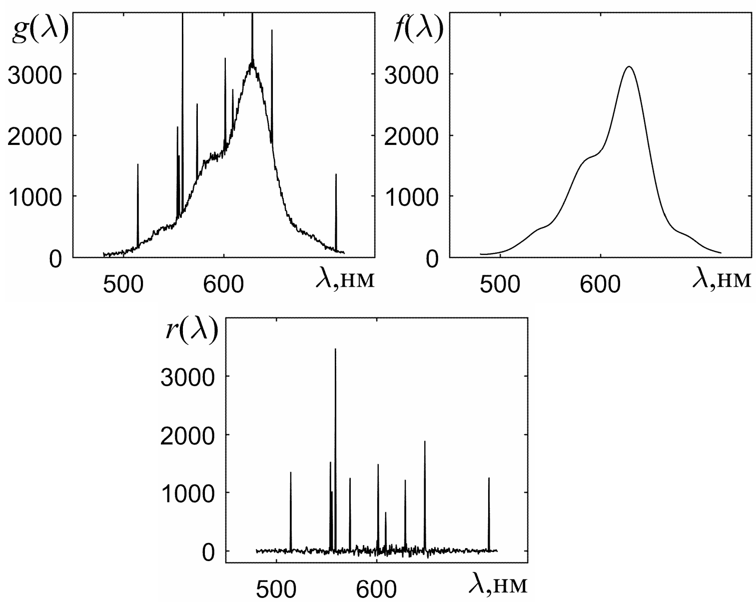
де $\mathbf{Q} = diag\left(\ldots, \dfrac{(\alpha^2 - [f_j^{(t)} - g_j]^2)}{(\alpha^2 + [f_j^{(t)} - g_j]^2)^2}, \ldots\right)$; $j = 1,\ldots,N$ є $N \times N$ діагональна матриця, що залежить від α^2 і від вектора відхилу на t-й ітерації, а круглі дужки з комою позначають операцію скалярного добутку векторів. Усього існує один такий крок на t-й ітерації.

Пробні кроки не є оптимальними, але вони є квазіоптимальними, бо зменшують значення цільової функції Φ. Алгоритм пошуку на множині пробних кроків є наступним. Спочатку на основі (10)-(13) генеруються значення пробних кроків. Потім у якості $h^{(t)}$ обирається крок з найменшим значенням Φ. Але якщо крок $h^{[4]}$ виявляється найкращим N разів поспіль, тоді виконується відновлення напрямку спуску за формулою: $\mathbf{p}^{(t)} = -\mathbf{grad}^{(t)}$. Відновлення напрямку спуску виконується також у випадку, коли число виконаних ітерацій дорівнює N. Якщо жоден із пробних кроків не призводить до зменшення значення Φ, то $h^{(t)}$ покладається рівним нулю й виконується відновлення напрямку спуску. Ітераційний процес методу спряжених градієнтів завершується у випадку, якщо $h^{(t)}$ дорівнює нулю двічі поспіль.

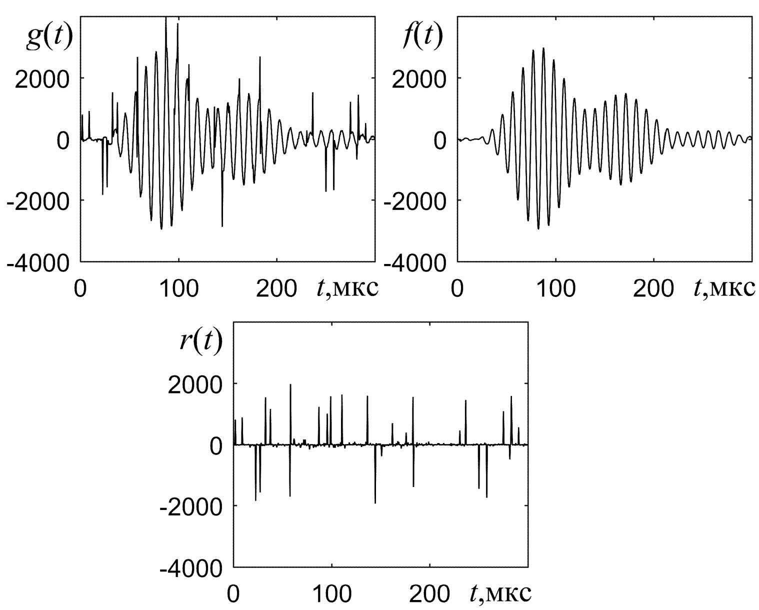
На рис. 1 наведені результати застосування розробленої інформаційної технології згладжування до синтезованих даних, які моделювали

ISSN 1562-9945 (Print)
ISSN 2707-7977 (Online)

спектр фотолюмінесценції (рис. 1а) і ультразвуковий сигнал (рис. 1г) за умов наявності шуму і аномальних значень. Результати згладжування, які наведені на рис. 1б та рис. 1д, візуально збіглися з очікуваними та були отримані для таких наборів значень: $\alpha = 100$, $\beta = 0$, $q = 2$, $\gamma = 1000$, $r = 2$ й $\alpha = 10$, $\beta = 0$, $q = 2$, $\gamma = 0{,}1$ $r = 3$, відповідно.



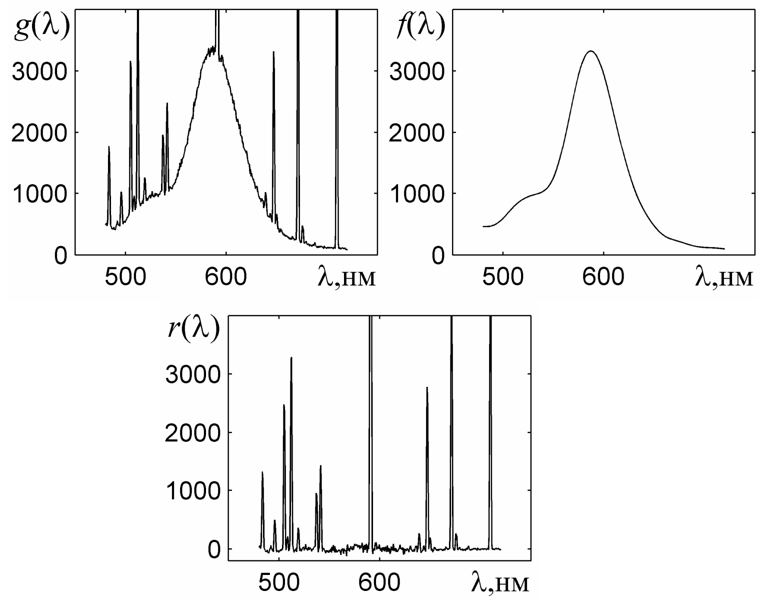
а) б) в)



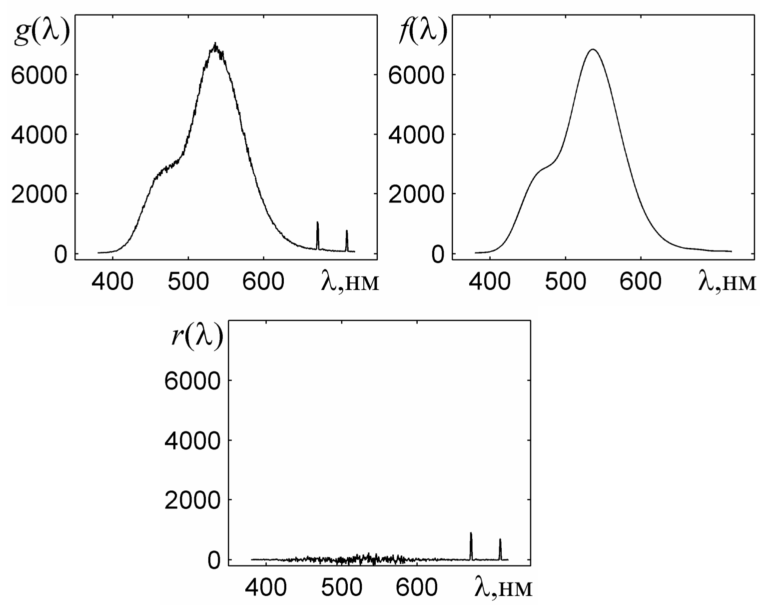
г) д) е)

Рисунок 1 - Чисельне моделювання: а), г) – вхідні дані; б), д) – згладжені дані; в), е) – вилучена частина даних

На рис. 2 наведені результати згладжування експериментальних спектрів фотолюмінесценції для значень: $\alpha = 100$, $\beta = 0$, $q = 2$, $\gamma = 1000$, $r = 2$. Тут залежність $g(\lambda)$ подає вхідні експериментальні дані довжиною в 481 (рис. 2а) й 681 елемент (рис. 2г), відповідно, залежність $f(\lambda)$ подає згладжені дані, а залежність $r(\lambda) = g(\lambda) - f(\lambda)$ подає відхил розв'язку.



а) б) в)



г) д) е)

Рисунок 2 Згладжування спектрів фотолюмінесценції: а), г) – виміряні спектри; б), д) – згладжені спектри; в), е) – вилучена частина даних

ISSN 1562-9945 (Print)
ISSN 2707-7977 (Online)

З розгляду рис.1 та рис.2 видно, що в результаті згладжування відбувається ефективне відокремлення шуму й аномальних значень від гладкої залежності.

Отже, отримані результати підтверджують ефективність застосування запропонованої інформаційної технології.

Висновки. У даній роботі представлена інформаційна технологія згладжування даних, заснована на критерії мінімуму протяжності й вимозі гладкості розв'язку. Результати тестування даної технології на задачах згладжування синтезованих даних і даних експериментальних спектрів фотолюмінесценції дозволяють зробити висновок про її практичну застосовність до обробки послідовностей даних помірної протяжності, які спотворені шумом і аномальними значеннями.

ЛИТЕРАТУРА / ЛІТЕРАТУРА

1. Chandola V. Anomaly detection: A survey / V. Chandola, A. Banerjee, V. Kumar // ACM Computing Surveys. – 2009. – V. 41. – N. 3. – P. 15-58.

2. Лемешко Б.Ю. Расширение области применения критериев типа Граббса, используемых при отбраковке аномальных измерений / Б. Ю. Лемешко, С. Б. Лемешко // Измерительная техника.–2005.–N.6.–С.13-19.

3. Вовк С. М. Критерій мінімуму протяжності / С. М. Вовк // Системні технології. Регіональний міжвузівський збірник наукових праць. Випуск 1 (120). – Дніпро, 2019. – С. 19 – 25.

4. Ермолаев В. А. Методы локального анализа и сглаживание временных рядов и дискретных сигналов / В. А. Ермолаев, Ю. А. Кропотов // Математическое моделирование. – 2017. – Т. 29. – № 2. – С. 119–132.

5. Savitzky A. Smoothing and differentiation of data by simplified least squares procedures / A. Savitzky, M.J. Golay // Analytical Chemistry. – 1964. – N. 36. – P. 1627-1639.

6. Форсайт Дж. Машинные методы математических вычислений / Дж. Форсайт, М. Малькольм, К. Моулер. – М.: Мир, 1980. – 280 с.

7. Любин П. Г. Об одном методе сглаживания двумерной поверхности / П. Г. Любин // Вестник РУДН. Серия Математика. Информатика. Физика. – 2016. – №2. – С. 37–43.

8. Вовк С.М. Метод численного дифференцирования зашумленных данных с выбросами // Радіоелектроніка, інформатика, управління. – 2017. – N. 3 – С. 44-52.
9. Vovk S. M. General approach to building the methods of filtering based on the minimum duration principle / S. M. Vovk // Radioelectronics and Communications Systems. – 2016. – V. 59. – N. 7. – P. 281-292.

REFERENCES

1. Chandola V. Anomaly detection: A survey / V. Chandola, A. Banerjee, V. Kumar // ACM Computing Surveys. – 2009. – V. 41. – N. 3. – P. 15-58.
2. Lemeshko B.Yu. Rasshirenie oblasti primeneniya kriteriev tipa Grabbsa, ispolzuemyih pri otbrakovke anomalnyih izmereniy / B. Yu. Lemeshko, S. B. Lemeshko // Izmeritelnaya tehnika. – 2005. – N 6. – S.13-19.
3. Vovk S. M. Kryterii minimumu protiazhnosti / S. M. Vovk // Systemni tekhnolohii. Rehionalnyi mizhvuzivskyi zbirnyk naukovykh prats. Vypusk 1 (120). – Dnipro, 2019. – C. 19 – 25.
4. Ermolaev V. A. Metodyi lokalnogo analiza i sglazhivanie vremennyih ryadov i diskretnyih signalov / V. A. Ermolaev, Yu. A. Kropotov. – Matematicheskoe modelirovanie. – 2017. – V. 29. – N. 2. – S. 119–132.
5. Savitzky A. Smoothing and differentiation of data by simplified least squares procedures / A. Savitzky, M.J. Golay // Analytical Chemistry. – 1964. – N. 36. – P. 1627-1639.
6. Forsayt Dzh. Mashinnyie metodyi matematicheskih vyichisleniy / Dzh. Forsayt, M. Malkolm, K. Mouler. – M.: Mir, 1980. – 280 s.
7. Lyubin P.G. Ob odnom metode sglazhivaniya dvumernoy poverhnosti / P. G. Lyubin // Vestnik RUDN. Seriya Matematika. Informatika. Fizika. – 2016. – N. 2. – S. 37–43.
8. Vovk S.M. Metod chislennogo differentsirovaniya zashumlennyih dannyih s vyibrosami // RadIoelektronIka, Informatika, upravlInnya. – 2017. – N. 3 – S. 44-52.
9. Vovk S. M. General approach to building the methods of filtering based on the minimum duration principle / S.M. Vovk // Radioelectronics and Communications Systems. – 2016. – V. 59. – N. 7. – P. 281-292.

Received 03.02.2020.
Accepted 05.02.2020.

ISSN 1562-9945 (Print)
ISSN 2707-7977 (Online)

Информационная технология сглаживания данных на основе критерия минимума протяженности

Для многих практических приложений актуальной является задача сглаживания данных, полученных при наличии шума и аномальных значений. Сложность решения этой задачи обусловлена тем, что параметрическая модель данных обычно является неизвестной, а наличие аномальных значений может стать причиной существенных ошибок. Данная работа посвящена разработке информационной технологии сглаживания данных на основе критерия минимума протяженности и требования гладкости искомого решения. Целью данной работы является получение информационной технологии сглаживания данных, искаженных аномальными значениями и шумом.

Сглаживание данных есть процесс их аппроксимации гладкой функцией. Один из эффективных методов сглаживания данных, искаженных аддитивным шумом, основан на методе регуляризации Тихонова. Однако этот метод не является эффективным тогда, когда данные содержат аномальные значения. Для устранения этого недостатка предложена другая постановка задачи сглаживания, которая отличается заменой основного квадратичного члена задачи сглаживания с регуляризацией Тихонова на неквадратичный член, формируемый на основе критерия минимума протяженности. Приведена краткая и развернутая формы предложенной постановки задачи сглаживания. Подчеркнуто, что эта постановка приводит к необходимости решения задачи минимизации с невыпуклой и неунимодальной целевой функцией. В рамках предложенной информационной технологии решение этой задачи минимизации достигается численно методом сопряженных градиентов, а управление процессом сглаживания данных осуществляется с помощью настроечных параметров, значения которых устанавливаются вручную или автоматически. Предложенная информационная технология опробована на данных, полученных численным моделированием, и на экспериментальных данных, представляющих собой спектры фотолюминесценции. Полученные результаты позволяют сделать вывод об эффективности применения предложенной информационной технологии.

Data smoothing information technology based on criterion of minimum-extent

For many practical applications, the smoothing problem of data obtained in the presence of noise and anomalous values is relevant. The complexity of solving this problem is due to the fact that the parametric data model is usually unknown, and the presence of anomalous values can cause significant errors. This work is devoted to the development of smoothing data information technology, which is based on the minimum-extent criterion and the smoothness of desired solution. The goal of this paper is to obtain the information technology for smoothing of data distorted by anomalous values and noise.

The data smoothing is the process of data approximation by a smooth function. One of the effective methods for smoothing of data distorted by additive noise is based on the Tikhonov regularization. However, this method is not effective when the data contains the anomal values. To eliminate this drawback, another formulation of the smoothing problem is proposed, which differs by replacing the main quadratic term of the smoothing problem based on Tikhonov regularization by a non-quadratic term formed on the basis of the minimum-extent criterion. A brief form and detailed form of the proposed statement of the smoothing problem are given. It is em-

phasized that this problem statement leads to the necessity of solving the minimization problem with a non-convex and non-unimodal objective function. Within the proposed information technology framework, the solution of this minimization problem is achieved numerically by the conjugate gradient method, and the data smoothing process is controlled by using tuning parameters, the values of which are set manually or automatically. The proposed information technology has been tested both on data obtained by numerical simulation and on experimental data representing the photoluminescence spectra. The obtained results confirmed the performance of the proposed information technology.

Вовк Сергей Михайлович - к.ф.-м.н., доцент кафедры компьютерных наук и информационных технологий Днепровского национального университета им. Олеся Гончара.

Вовк Сергій Михайлович - к.ф.-м.н., доцент кафедри комп’ютерних наук та інформаційних технологій Дніпровського національного університету ім. Олеся Гончара.

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

ISSN 1562-9945 (Print)
ISSN 2707-7977 (Online)