

Міністерство освіти і науки, молоді та спорту України
Харківський національний університет радіоелектроніки

ШКЛОВЕЦЬ АРТЕМ ВАДИМОВИЧ

УДК 004.272

КУСОЧНО-ГЛАДКІ САМООРГАНІЗУЮЧІ КАРТИ КОХОНЕНА ДЛЯ
ВІЗУАЛІЗАЦІЇ БАГАТОВИМІРНИХ ДАНИХ

05.13.23 – системи та засоби штучного інтелекту

АВТОРЕФЕРАТ

дисертації на здобуття наукового ступеня
кандидата технічних наук

Харків – 2012

Дисертацією є рукопис.

Робота виконана у Харківському національному університеті радіоелектроніки, Міністерство освіти і науки, молоді та спорту України.

Науковий керівник — кандидат технічних наук, доцент
Аксак Наталія Георгіївна,
Харківський національний університет
радіоелектроніки, доцент кафедри електронних
обчислювальних машин.

Офіційні опоненти: доктор технічних наук, професор
Ахметшина Людмила Георгіївна,
Дніпропетровський національний університет
імені О. Гончара, професор кафедри електронних
обчислювальних машин, м. Дніпропетровськ;

доктор технічних наук, професор
Гороховатський Володимир Олексійович,
Харківський інститут банківської справи Університету
банківської справи Національного банку, професор,
завідувач кафедри інформаційних технологій, м. Харків.

Захист відбудеться „_____” _____ 2013 р. о _____ годині на засіданні спеціалізованої вченої ради Д 64.052.01 у Харківському національному університеті радіоелектроніки за адресою: 61166, м. Харків, просп. Леніна, 14.

З дисертацією можна ознайомитися в бібліотеці Харківського національного університету радіоелектроніки за адресою: 61166, м. Харків, просп. Леніна, 14.

Автореферат розіслано „_____” _____ 2012 р.

Вчений секретар
спеціалізованої вченої ради

Є.І. Литвинова

ЗАГАЛЬНА ХАРАКТЕРИСТИКА РОБОТИ

Актуальність теми. У зв'язку зі стрімким накопиченням великої кількості даних збільшується потреба у швидких і точних методах їх аналізу. Сьогодні розроблені численні методи аналізу даних малої вимірності, здатних виявити складні закономірності та моделювати систему або процес, що описується досліджуваними даними. Перевага роботи з даними малої вимірності полягає у простоті методів аналізу та можливості наочного уявлення, як самих даних, так і результатів їхнього аналізу.

Чималий інтерес становлять методи виявлення закономірностей, укладених у великій кількості даних. Але під час реалізації розроблених методів аналізу багатовимірних даних з'являється проблема наочного уявлення як самих даних, так і результатів їхнього аналізу. Відображення багатовимірних даних у простір малої вимірності, яке зберігає їх структуру, дозволяє виявити велику кількість нових закономірностей та визначити в подальшому методи аналізу цих даних. Таке відображення може бути реалізовано шляхом розв'язання задачі візуалізації багатовимірних даних.

Існуючі методи візуалізації багатовимірних даних дозволяють швидко будувати лінійне відображення багатовимірних даних у простір малої вимірності, що вельми ефективно для даних, які створюють просту унімодальну структуру. У випадку багатокластерної структури багатовимірних даних лінійне відображення приводить до значних помилок візуалізації таких даних. У зв'язку з цим була розроблена велика кількість нелінійних методів візуалізації багатовимірних даних, які дозволяють відображати їх у простір малої вимірності з великим ступенем точності. Основним недоліком цих методів є велика кількість скалярних операцій, необхідних для їх реалізації, що робить їх непридатними для візуалізації великої кількості багатовимірних даних. Використання самоорганізуючих карт Кохонена (СОК) дозволяє будувати кусочно-лінійне відображення багатовимірних даних у простір малої вимірності з прийнятною точністю та високою швидкістю. Однак СОК мають недоліки, які призводять до зменшення точності візуалізації багатовимірних даних.

В області візуалізації багатовимірних даних активно ведуться наукові розробки такими вченими як Горбань А.Н., Зінов'єв А.Ю., Kegl B., Wunsch D., Макаренко Н. Г. Значний внесок у розвиток штучних нейронних мереж Кохонена роблять такі українські вчені, як Руденко О.Г., Бодянський Є.В., Куcssуль Н.М. та зарубіжні – Кохонен Т., Хайкін С., Нечаєв Ю.І., Крижановський Б.В., Дунін-Барковський В. Л., Терехов С.А., Яхно В.Г.

Однією з невирішених проблем візуалізації великої кількості багатовимірних даних за допомогою СОК є відображення багатовимірних даних в одну точку, що зводить до їх невиразності у просторі малої вимірності. Цей недолік пов'язаний з кусочно-лінійною структурою самоорганізуючої карти Кохонена.

У зв'язку з цим розробка методів візуалізації багатовимірних даних за допомогою самоорганізуючих карт Кохонена з гладкою структурою є актуальною як з практичної, так і з теоретичної точки зору. Задачі, які при цьому виникають, зумовили напрямок досліджень дисертаційної роботи.

Зв'язок роботи з науковими програмами, темами, планами. Тема дисертаційної роботи та отримані результати відповідають проблематиці держбюджетної теми «Еволюційні гібридні системи обчислювального інтелекту зі змінною структурою для інтелектуального аналізу даних», розділ «Еволюційні гібридні методи і моделі інтелектуальної обробки інформації зі змінною структурою в умовах невизначеності» (№ ДР 0110U000458), у якій автор був одним з виконавців.

Мета і задачі дослідження. Метою дисертаційної роботи є розробка методів побудови кусочно-гладких самоорганізуючих карт Кохонена, які за рахунок апроксимації кусочно-лінійних карт Кохонена кубічними сплайнами дозволяють підвищити точність візуалізації багатовимірних даних.

Для досягнення поставленої мети необхідно вирішити такі задачі:

- розробити методи побудови одновимірних і двовимірних кусочно-гладких самоорганізуючих карт Кохонена;
- розробити методи відображення багатовимірних даних на одновимірні та двовимірні кусочно-гладкі самоорганізуючі карти Кохонена;
- розробити методи обчислення відстаней між багатовимірними даними на кусочно-гладких самоорганізуючих картах Кохонена;
- провести експериментальні дослідження та реалізувати запропоновані методи для розв'язання практичних задач.

Об'єктом дослідження є процес візуалізації багатовимірних даних.

Предметом дослідження є самоорганізуючі карти Кохонена для візуалізації багатовимірних даних.

Методи дослідження. Для розв'язання поставлених задач використані теорія штучних нейронних мереж для кусочно-лінійного відображення багатовимірних даних у простір малої вимірності з невисокою трудомісткістю; сплайн-апроксимація, диференціальна геометрія, теорія графів і теорія оптимізації для побудови кусочно-гладких самоорганізуючих карт Кохонена; чисельні методи для відображення багатовимірних даних на кусочно-гладкі самоорганізуючі карти Кохонена.

Наукова новизна результатів дисертаційної роботи:

1. Вперше запропоновано метод побудови одновимірних кусочно-гладких самоорганізуючих карт Кохонена, який базується на апроксимації одновимірних кусочно-лінійних самоорганізуючих карт Кохонена кубічними параметричними сплайнами, що дозволяє формувати самоорганізуючі карти Кохонена без зломів у нейронах для підвищення точності візуалізації багатовимірних даних в одновимірному просторі.

2. Вперше запропоновано методи побудови двовимірних кусочно-гладких самоорганізуючих карт Кохонена, які характеризуються апроксимацією двовимірних кусочно-лінійних самоорганізуючих карт Кохонена кубічними параметричними сплайн-поверхнями, що дає можливість будувати самоорганізуючі карти Кохонена без зломів у дотику сторін трикутників із триангуляції Делоне для збільшення точності візуалізації багатовимірних даних у двовимірному просторі.

3. Вперше запропоновано методи відображення багатовимірних даних на кусочно-гладкі самоорганізуючі карти Кохонена, які базуються на використанні методу Ньютона з декількома початковими наближеннями, що дозволяє розрізняти

дані на карті та зменшити похибку візуалізації багатовимірних даних під час збереженні лінійної обчислювальної складності.

4. Набув подальшого розвитку метод обчислення відстаней між елементами багатовимірних даних на самоорганізуючих картах Кохонена шляхом введення на них метричного тензора, обчислюються викривлення самоорганізуючих карт Кохонена, що дозволяє зменшити похибку відображення візуалізованих даних.

Практичне значення результатів дисертаційної роботи полягає в тому, що запропоновані методи побудови кусочно-гладких самоорганізуючих карт Кохонена та методи відображення на них багатовимірних даних було використано для розв'язання широкого кола різноманітних за змістом задач, таких як візуалізація, кластеризація та класифікація багатовимірних даних.

Методи візуалізації багатовимірних даних за допомогою кусочно-гладких самоорганізуючих карт Кохонена доведені до рівня програмної реалізації, що дозволило:

1) здійснити аналіз ринку полімерів компаній із країн СНД у відділі моніторингу та річних оглядів і стратегічного консалтингу ООО «Маркет Репорт» (акт впровадження від 7.02.2012);

2) виконати розв'язання задачі ефективного планування та розробки тарифів у відділі роботи із споживачами ОАО «Укртелеком» (акт впровадження від 15.12.2011).

Проведені експериментальні дослідження для оцінки розроблених методів підтверджують основні положення, що виносяться на захист.

Результати роботи були впроваджені в навчальний процес Харківського національного університету радіоелектроніки (акт впровадження від 13.03.2012).

Особистий внесок здобувача. Всі основні результати, що виносяться на захист, отримано автором самостійно. У роботах, що опубліковані у співавторстві, здобувачеві належать: [1] – новий метод побудови одновимірної кусочно-гладкої СОК; [2] – метод побудови чотирикутних СОК для прискорення побудови двовимірних кусочно-гладких СОК; [3] – новий метод відображення багатовимірних даних на одновимірну кусочно-гладку СОК; [5] – новий метод відображення багатовимірних даних на двовимірну кусочно-гладку СОК; [6] – метод визначення відстаней між точками на кусочно-гладких СОК з використанням Риманової метрики; [7] – новий метод побудови двовимірних кусочно-гладких СОК на триангуляції Делоне; [8] – порівняльний аналіз одновимірних кусочно-гладких та кусочно-лінійних СОК для візуалізації багатовимірних даних; [9] – порівняльний аналіз двовимірних кусочно-гладких та кусочно-лінійних СОК для візуалізації багатовимірних даних; [10] – порівняльний аналіз методів візуалізації великої кількості багатовимірних даних за допомогою СОК; [11] – модифікацію алгоритму навчання нейронної мережі Кохонена, структура якої зростає; [12] – паралельний алгоритм навчання зростаючої нейронної мережі Кохонена; [13] – паралельний алгоритм навчання нейронної мережі Кохонена; [14] – метод побудови двовимірних кусочно-гладких СОК на нерегулярній решітці; [15] – метод відображення багатовимірних даних на одновимірну кусочно-гладку СОК; [16] – аналіз методів візуалізації багатовимірних даних.

Апробація результатів дисертації. Основні результати дисертаційної роботи доповідалися й обговорювалися на наукових конференціях: 9 Міжнародна

конференція-семінар «Высокопроизводительные вычисления на кластерных системах» (м. Владивір, Росія, 2009); 17 Міжнародна конференція з автоматичного управління «Автоматика -2010» (м. Харків, 2010); 13 та 14 - всеросійські науково-технічні конференції «Нейроинформатика - 2011» та «Нейроинформатика - 2012» (м. Москва, Росія, 2011, 2012); 9 Всеросійська наукова конференція "Нейрокомпьютеры и их применение" (м.Москва, Росія, 2011); 2 Міжнародна науково-практична конференція «Информатика, математическое моделирование, экономика» (м. Смоленськ, Росія, 2012); 15 та 16 Міжнародних молодіжних форумах «Радиоэлектроника и молодежь в XXI веке» (м. Харків, 2011, 2012); Міжнародна наукова конференція «Интеллектуальные системы принятия решений и проблемы вычислительного интеллекта (ISDMCI'2010)» (м. Евпатория, 2010); 1 Міжнародна науково-технічна конференція «Сучасні напрями розвитку інформаційно-комунікаційних технологій та засобів управління» (м. Харків-Київ, 2010).

Публікації. За темою дисертаційної роботи опубліковано 18 наукових праць, з них: 6 статей у наукових фахових виданнях України з технічних наук, серед яких 1 стаття без співавторів, 12 публікацій у збірниках матеріалів та тез доповідей на міжнародних наукових конференціях, семінарах, форумах.

Структура та обсяг дисертаційної роботи. Дисертаційна робота складається із вступу, чотирьох розділів, висновків, списку використаних літературних джерел з 116 найменувань на 11 сторінках, 3 додатків на 6 сторінках. Робота містить 40 рисунків, 15 таблиць. Загальний обсяг роботи складає 153 сторінки, з них 136 - основного тексту.

ОСНОВНИЙ ЗМІСТ РОБОТИ

У **вступі** стисло розглянуто стан досліджень у галузі візуалізації багатовимірних даних (ВБД), обґрунтовано актуальність теми та доцільність роботи, зазначено зв'язок роботи з науковими програмами, темами, сформульовано мету та задачі дослідження, визначено об'єкти, предмет і методи дослідження, охарактеризовано наукову новизну і практичне значення отриманих результатів, наведено кількість публікацій за темою роботи, виділено особистий внесок здобувача.

У **першому розділі** наведена змістовна та формальна постановка задачі ВБД і проведено аналіз методів розв'язання цієї задачі.

Задача ВБД полягає в побудові такого відображення простору багатовимірних даних у простір малого розміру, яке зберігає структуру та головні закономірності, властиві багатовимірним даним.

Класичним методом розв'язання задачі ВБД є метод головних компонент, головна ідея якого полягає у побудові лінійного різноманіття, званого картою, на яке проектуються дані. Цей метод ефективно візуалізує дані з унімодальним розподілом. Для більш складних структур багатовимірних даних було розроблено ряд нелінійних методів ВБД, таких як карти Саммона, метод головних різноманіть, самоорганізуючі карти Кохонена (СОК), пружні карти. Найбільш висока точність ВБД досягається за допомогою карт Саммона, яка полягає у мінімізації викривлення структури даних. ВБД за допомогою методу головних різноманіть полягає у

побудові нелінійної карти, на яку проектується дані. Головним недоліком наведених методів є висока трудомісткість їх реалізації, що робить ці методи неприйнятними для візуалізації великої кількості багатовимірних даних.

Головна ідея ВБД за допомогою СОК полягає в навчанні нейронної мережі Кохонена, вихідні нейрони якої адаптивно розташовуються у хмарі даних. В наслідок цього формується кусочно-лінійна самоорганізуюча карт Кохонена (КЛСОК), на яку відображуються багатовимірні дані. Основна перевага СОК полягає у малій кількості скалярних операцій, необхідних для їх побудови, що дозволяє ефективно використовувати їх для візуалізації великої кількості багатовимірних даних. Більшість виявлених недоліків ВБД за допомогою КЛСОК були розв'язані шляхом розробки різноманітних модифікацій алгоритму навчання нейронної мережі Кохонена, таких як алгоритм регуляризації, Batch SOM, Density Tracking SOM, Adaptive SOM, ієрархічні алгоритми, нейронний газ, пружні карти тощо. Але кусочно-лінійна структура СОК приводить до невиразності даних на карті і значно зменшує точність ВБД.

На основі проведеного аналізу визначено сукупність перспективних напрямків збільшення точності ВБД за допомогою використання гладких структур СОК і сформульовано задачі дисертаційної роботи.

У **другому розділі** вперше запропоновано метод побудови одновимірних кусочно-гладких самоорганізуючих карт Кохонена (КГСОК) шляхом апроксимації одновимірних КЛСОК кубічними параметричними сплайнами та метод відображення багатовимірних даних на побудовану одновимірну КГСОК.

Кубічний параметричний сплайн у n -мірному просторі задається вектор-функцією

$$S(t) = (s_1(t), \dots, s_n(t)), \quad (1)$$

$$s_p(t) = \bigoplus_{k=1}^{\ell-1} s_p^k(t) \in C^2[1, \ell], p = \overline{1, n}, \quad (2)$$

$$s_p^k(t) = \sum_{j=0}^3 a_{jp}^k (t-k)^j, t \in [k, k+1], S^k(t) = (s_1^k(t), \dots, s_n^k(t)), \quad (3)$$

де $s_p(t)$ – сплайн 3-го порядку одиничного дефекту з вузлами на сітці Δ , яка є рівномірним розбиттям відрізка $[1, \ell]$ на $(\ell-1)$ інтервал, ℓ – кількість нейронів вихідного шару нейронної мережі Кохонена з матрицею вагових коефіцієнтів $W = (w_i^k)_{i=\overline{1, \ell}}^{k=\overline{1, n}}$, t – параметр сплайна.

Метод побудови одновимірної кусочно-гладкої самоорганізуючої карти Кохонена полягає у визначенні $N_A = 4n(\ell-1)$ коефіцієнтів $A = (a_{jp}^k)_{j=\overline{0,3}; p=\overline{1, n}}^{k=\overline{1, \ell-1}}$ параметричного сплайна $S(t)$ шляхом складання та розв'язку системи лінійних алгебраїчних рівнянь (СЛАР), що складається з $4n(\ell-1) - 2n$ рівнянь, еквівалентних умовам проходження параметричного сплайна через нейрони (4) та його гладкості другого порядку (5), (6)

$$\begin{cases} S^k(t)|_{t=k} = W_k, \\ S^k(t)|_{t=k+1} = W_{k+1}, \end{cases} \Rightarrow \begin{cases} A_0^k = W_k, \\ \sum_{j=0}^3 A_j^k = W_{k+1}, \end{cases} \quad k = \overline{1, \ell-1} \quad (4)$$

$$\frac{dS^k(t)}{dt}|_{t=k+1} = \frac{dS^{k+1}(t)}{dt}|_{t=k+1} \Rightarrow \sum_{j=1}^3 jA_j^k = A_1^{k+1}, \quad k = \overline{1, \ell-2} \quad (5)$$

$$\frac{d^2S^k(t)}{dt^2}|_{t=k+1} = \frac{d^2S^{k+1}(t)}{dt^2}|_{t=k+1} \Rightarrow \sum_{j=2}^3 j(j-1)A_j^k = 2A_2^{k+1}, \quad k = \overline{1, \ell-2}. \quad (6)$$

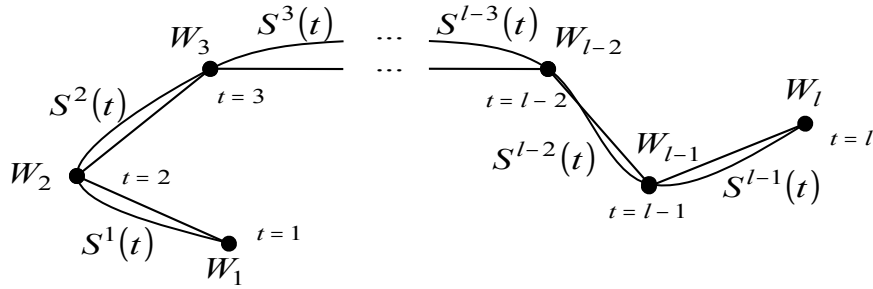


Рис. 1. Побудова одновимірної КГСОК

СЛАР (4)-(6) була розв'язана за лінійний час методом прогонки відносно коефіцієнтів $A_1^1 = (a_{11}^1, \dots, a_{1n}^1)$ та $A_3^{\ell-1} = (a_{31}^{\ell-1}, \dots, a_{3n}^{\ell-1})$. Решта коефіцієнтів була знайдена трьома шляхами:

1) шляхом визначення

$$\begin{cases} \frac{dS^1(t)}{dt}|_{t=1} = 0, \\ \frac{dS^{\ell-1}(t)}{dt}|_{t=\ell} = 0, \end{cases} \Rightarrow \begin{cases} A_1^1 = 0, \\ \sum_{j=2}^3 jA_j^{\ell-1} = 0; \end{cases} \quad (7)$$

2) шляхом мінімізації довжини одновимірної КГСОК

$$\sum_{k=1}^{\ell-1} \int_k^{k+1} \sqrt{\sum_{p=1}^n \left(\frac{ds_p^k(t)}{dt} \right)^2} dt \rightarrow \min; \quad (8)$$

3) мінімізації відхилення одновимірної КГСОК від відповідної їй одновимірної КЛСОК

$$\sum_{k=1}^{\ell-1} \frac{1}{\rho_k} \int_k^{k+1} \left(\sum_{p=1}^n [s_p^k(t) - w_p^k + (t-k)(w_p^{k+1} - w_p^k)]^2 \right) dt \rightarrow \min \quad (9)$$

де ρ_k – довжина відрізка $[W_{k+1}, W_k]$, $W^k = (w_1^k, \dots, w_n^k)$. Усі способи дозволяють побудувати одновимірну КЛСОК аналітично за лінійний час.

Запропоновано *метод відображення багатовимірних даних на одновимірну КГСОК*. Для відображення багатовимірних даних $X = \{X^1, \dots, X^N\}$, $X^c = (x_1^c, \dots, x_m^c)^T$, $c = \overline{1, N}$ на одновимірну КГСОК була розв’язана система

$$\frac{dS(t)}{dt} \cdot (S(t) - X^c) = 0, \quad c = \overline{1, N}, \quad (10)$$

коренями якої є проекції елемента X^c на одновимірну КГСОК (рис. 2.). Розв’язок системи (5) було зведено до розв’язку системи алгебраїчних рівнянь 5-го порядку на відрізках, отриманих розбиттям Δ . Система (5) була розв’язана за лінійний час шляхом локалізації та пошуком коренів методом Ньютона на відповідних відрізках.

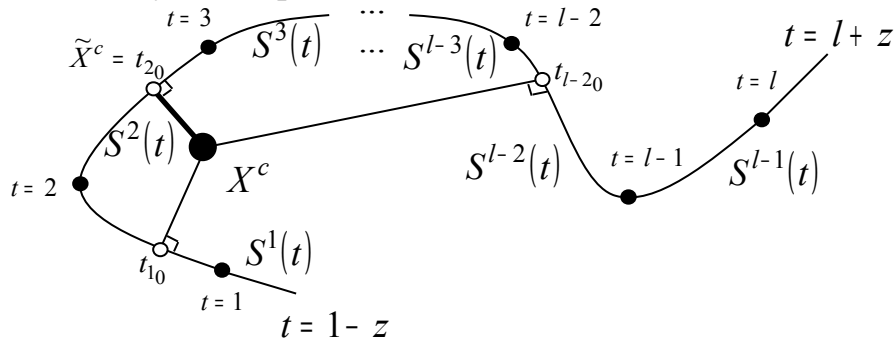


Рис. 2. Відображення елемента X^c на одновимірну КГСОК $S(t)$

Елемент $\tilde{X}^c \in S(t)$ із множини коренів системи (5) називається візуалізацією елемента X^c на X^c на одновимірну КГСОК, якщо вона розташована найближче до елемента X^c відносно інших коренів системи (5). Результатом ВБД є одновимірна множина $\tilde{X} = \{\tilde{X}^1, \dots, \tilde{X}^N\}$.

У третьому розділі вперше запропоновано методи побудови двовимірних КГСОК шляхом апроксимації двовимірних КЛСОК кубічними параметричними сплайн-поверхнями та метод відображення багатовимірних даних на побудовану двовимірну КГСОК.

Двовимірна кубічна сплайн-поверхня задається вектор-функцією

$$S(t_1, t_2) = (s_1(t_1, t_2), \dots, s_n(t_1, t_2)), \quad (11)$$

якщо на деякому двовимірному різноманітті $\Omega \subset \mathbb{R}^n$ задано розбиття Δ на під-області Ω^k , $k = \overline{1, N_\Omega}$, N_Ω - кількість під областей розбиття Δ , причому кожній під-області Ω^k із розбиття Δ відповідає вектор-функція

$$S^k(t_1, t_2)|_{\Omega^k} = (s_1^k(t_1, t_2), \dots, s_n^k(t_1, t_2)), \quad (12)$$

параметри якої змінюються в деякій області $\tilde{\Omega}^k \in \mathbb{R}^2$, функція $s_p^k(t_1, t_2)$ є поліномом третього порядку відносно параметрів t_1 і t_2

$$s_p^k(t_1, t_2) = \sum_{i=0}^3 \sum_{j=0}^3 a_{ijp}^k t_1^i t_2^j, \quad k = \overline{1, N_\Omega}, \quad p = \overline{1, n} \quad (13)$$

$$S(t_1, t_2) = \bigoplus_{k=1}^{N_\Omega} S^k(t_1, t_2) \Big|_{\Omega} \quad (14)$$

та $S(t_1, t_2) \in W_2^{2,2}[\Omega]$, де $W_2^{2,2}[\Omega]$ – простір Соболева.

Запропоновано метод побудови двовимірних кусочно-гладких самоорганізуючих карт Кохонена на триангуляції Делоне.

У випадку двовимірної КГСОК розбиття Δ являє собою триангуляцію Делоне $T = \{T^1, \dots, T^M\}$, побудовану на нейронах Кохонена W , де $T^k = \{T_1^k, T_2^k, T_3^k\}$, $k = \overline{1, M}$, M - кількість трикутників, T_p^k – номер нейрона, $p = \overline{1, 3}$.

Для побудови двовимірної КГСОК визначені $N_A = 16nM$ коефіцієнтів $A = (A_{ij}^k)_{i=\overline{0,3}; j=\overline{0,3}}^{k=\overline{1,M}} = (a_{ijp}^k)_{i=\overline{0,3}; j=\overline{0,3}; p=\overline{1,n}}^{k=\overline{1,M}}$ кубічної параметричної сплайн-поверхні шляхом складання та визначення рангу r СЛАР, складеної з умов (10) проходження двовимірної КГСОК через нейрони W та умов (11)-(12) гладкого дотику всіх пар трикутників із триангуляції Делоне T , що мають загальну сторону.

$$\begin{cases} S^k(t_1, t_2) \Big|_{t_1=0, t_2=0} = W^{T_1^k}, \\ S^k(t_1, t_2) \Big|_{t_1=1, t_2=0} = W^{T_2^k}, \\ S^k(t_1, t_2) \Big|_{t_1=0, t_2=1} = W^{T_3^k}, \end{cases} \Rightarrow \begin{cases} A_{00}^k = W^{T_1^k}, \\ \sum_{i=0}^3 A_{i0}^k = W^{T_2^k}, \quad k = \overline{1, M} \\ \sum_{j=0}^3 A_{0j}^k = W^{T_3^k}, \end{cases} \quad (15)$$

$$S^{k_1}(t_1, t_2) \Big|_{\partial\Omega} = S^{k_2}(t_1, t_2) \Big|_{\partial\Omega}, \quad (16)$$

$$\frac{\partial S^{k_1}(t_1, t_2)}{\partial t_2} \Big|_{\partial\Omega} = - \frac{\partial S^{k_2}(t_1, t_2)}{\partial t_2} \Big|_{\partial\Omega}, \quad \frac{\partial S^{k_1}(t_1, t_2)}{\partial t_2} \Big|_{\partial\Omega} = - \frac{\partial S^{k_2}(t_1, t_2)}{\partial t_2} \Big|_{\partial\Omega}, \quad (17)$$

де $\partial\Omega$ - крива дотику поверхні $S^{k_1}(t_1, t_2)$ і $S^{k_2}(t_1, t_2)$.

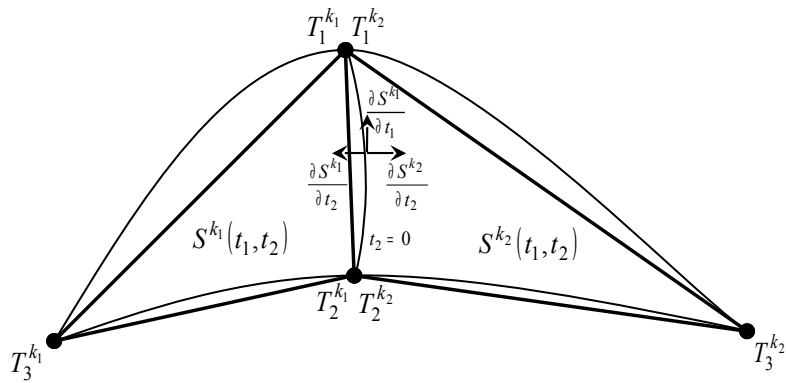
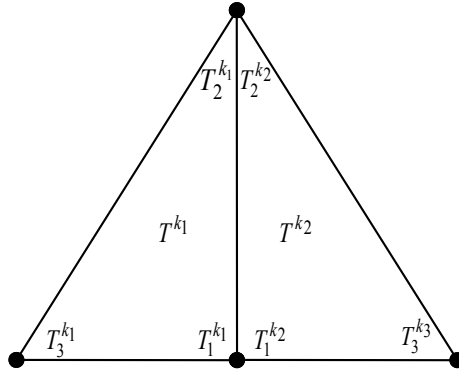
Розглянуто 21 випадок дотику сторін трикутників T^{k_1} і T^{k_2} .

У випадку $(T_1^{k_1} = T_1^{k_2}) \vee (T_2^{k_1} = T_2^{k_2})$, тобто у трикутників T^{k_1} і T^{k_2} загальною є сторона $(1;2) \Leftrightarrow (1;2)$, $\partial\Omega$ для трикутників T^{k_1} і T^{k_2} виражається співвідношенням $(t_2 = 0)$ (рис. 3.). Умови (16) і (17) набувають вигляд відповідно:

$$A_{i0}^{k_1} = A_{i0}^{k_2}, i = \overline{0,3}, \quad (18)$$

$$A_{i1}^{k_1} = -A_{i1}^{k_2}, i = \overline{0,3}, \quad (19)$$

$$A_{i2}^{k_1} = A_{i2}^{k_2}, i = \overline{0,3}. \quad (20)$$



(a)

(б)

Рис. 3. Випадок $(1;2) \leftrightarrow (1;2)$. (а) – вигляд на площині, (б) – вигляд у просторі

Для випадку $r \geq N_A$ видалено $(r - N_A)$ нееквівалентних рівнянь із СЛАР, що дозволило розв'язати її, але звело до зменшення дефекту кубічної параметричної сплайн-поверхні.

Для $r < N_A$ виражено r коефіцієнтів СЛАР через $(N_A - r)$. Коефіцієнти, які залишилися, були визначені шляхом розв'язку задачі оптимізації. Розглянуті три види цільової функції:

1) сума відстаней між точками $S^k(t_1, t_2) \Big|_{t_1=\frac{1}{2}, t_2=0}, S^k(t_1, t_2) \Big|_{t_1=0, t_2=\frac{1}{2}},$

$$S^k(t_1, t_2) \Big|_{t_1=\frac{1}{2}, t_2=\frac{1}{2}} \text{ та відповідними точками } \frac{W^{T_1^k} + W^{T_2^k}}{2}, \frac{W^{T_1^k} + W^{T_3^k}}{2}, \frac{W^{T_2^k} + W^{T_3^k}}{2},$$

$k = \overline{1, M};$

2) сума довжин кривих, відповідних сторонам топологічних трикутників T

$$\sum_{k=1}^M \left(\int_0^1 \sqrt{\sum_{p=1}^n \left(\frac{\partial s_p^k(t_1, 0)}{\partial t_1} \right)^2} dt_1 + \int_0^1 \sqrt{\sum_{p=1}^n \left(\frac{\partial s_p^k(0, t_2)}{\partial t_2} \right)^2} dt_2 + \int_0^1 \sqrt{\sum_{p=1}^n \left(\frac{\partial s_p^k(t_1, 1-t_1)}{\partial t_1} \right)^2} dt_1 \right); \quad (21)$$

3) сума відстаней від кожного трикутника T^k до відповідної йому поверхні $S^k(t_1, t_2)$

$$\sum_{k=1}^M \sum_{p=1}^n \int_0^1 dt_1 \int_0^{1-t_1} \left(s_p^k(t_1, t_2) - w_p^{T_1^k} - (w_p^{T_3^k} - w_p^{T_1^k})t_1 - (w_p^{T_2^k} - w_p^{T_1^k})t_2 \right)^2 dt_2. \quad (22)$$

Для прискорення та спрощення процесу побудови двовимірної КГСОК запропоновано *метод побудови двовимірних кусочно-гладких самоорганізуючих карт Кохонена на основі чотирикутних СОК*. Побудова чотирикутних СОК полягає у побудові графа G , ізоморфного вихідної триангуляції T (рис. 4).

Для графа G побудовано остовий ліс O_1 , наведений на рис. 5.

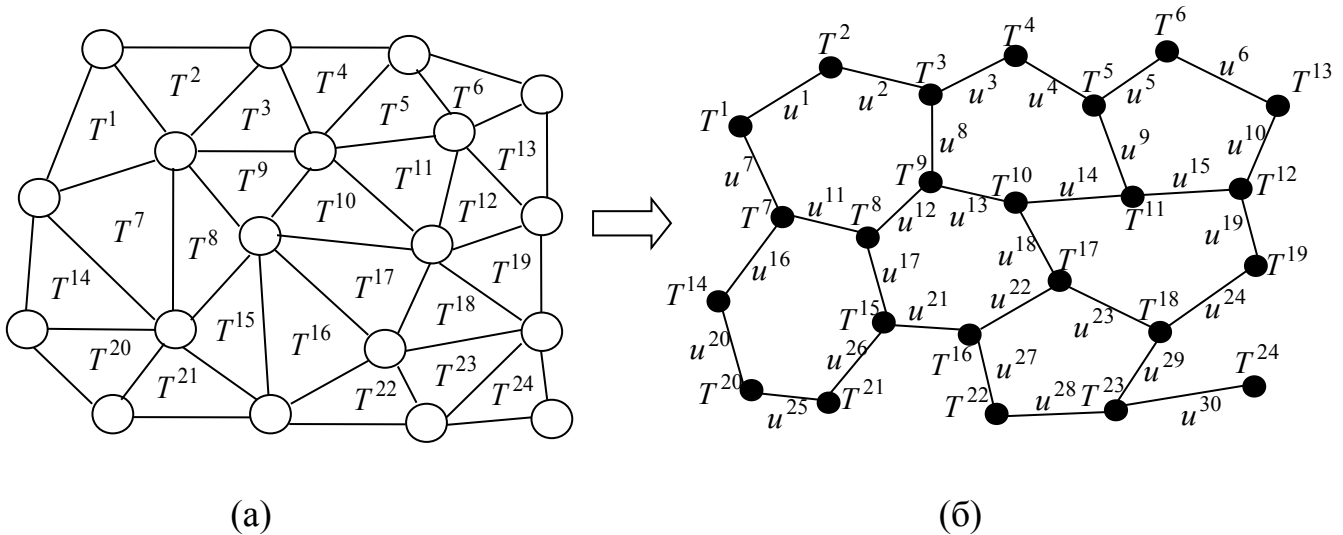


Рис. 4. – (а) – триангуляція T , (б) – відповідний граф G

На основі остовного лісу O_1 графа G побудована чотирикутна самоорганізуюча карта Кохонена Q , як показано на рис. 6.

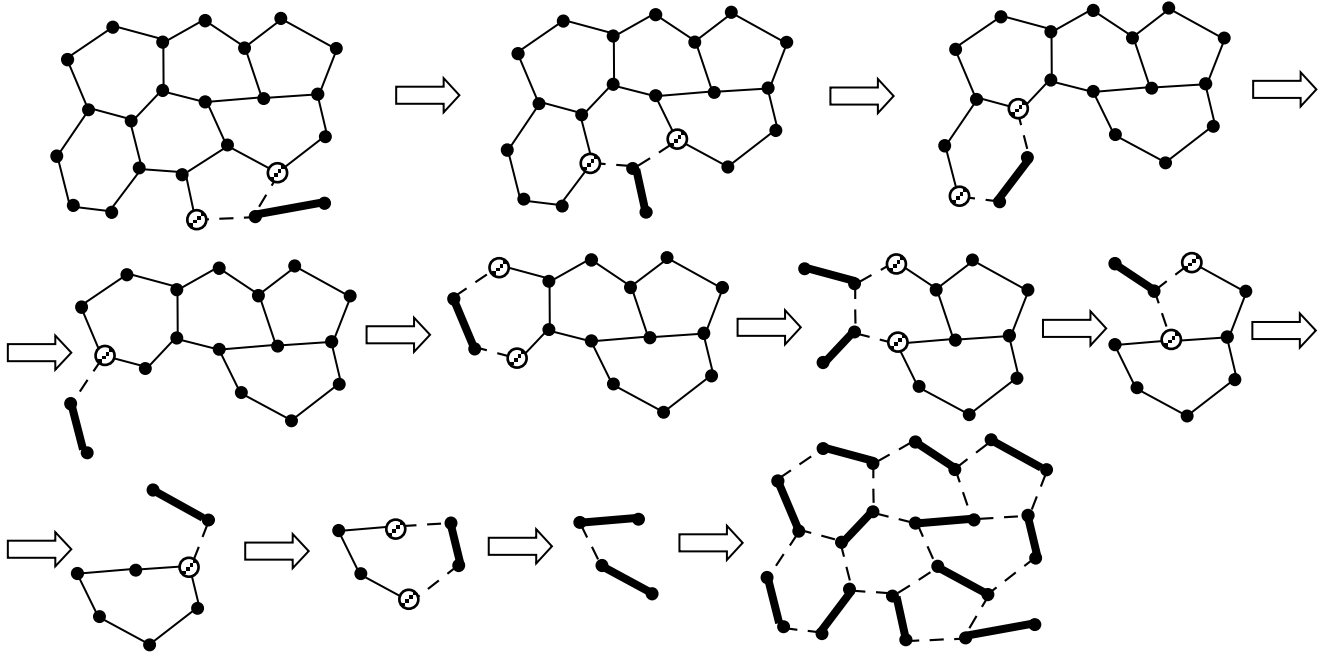
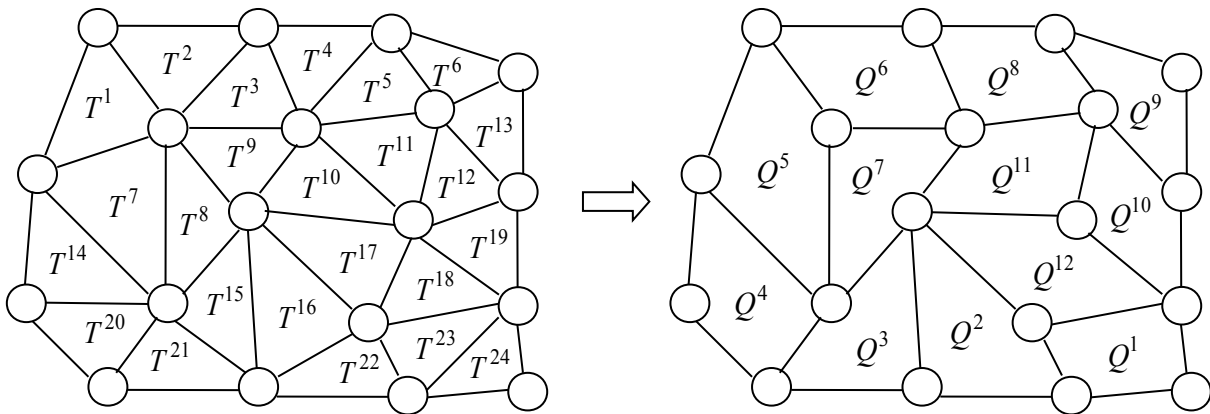
Рис. 5. Процес побудови остовного лісу O_1 графа G 

Рис. 6. Процес побудови чотирикутної СОК

Побудова двовимірної КГСОК на чотирикутних СОК здійснюється аналогічно методу побудови двовимірної КГСОК у вигляді триангуляції Делоне. Задача оптимізації (22) має вигляд

$$\begin{aligned}
& \sum_{k=1}^{\frac{M}{2}} \sum_{p=1}^n \left[\int_0^1 dt_1 \int_0^{1-t_1} \left(s_p^k(t_1, t_2) - x_p^{Q_1^k} - (x_p^{Q_2^k} - x_p^{Q_1^k})t_1 - (x_p^{Q_2^k} - x_p^{Q_1^k})t_2 \right)^2 dt_2 + \right. \\
& \left. + \int_0^1 dt_1 \int_0^{1-t_1} \left(s_p^k(1-t_1, 1-t_2) - x_p^{Q_4^k} - (x_p^{Q_2^k} - x_p^{Q_4^k})t_1 - (x_p^{Q_3^k} - x_p^{Q_4^k})t_2 \right)^2 dt_2 \right] \rightarrow \min.
\end{aligned} \quad (23)$$

Побудовано двовимірні кусочно-гладкі прямокутні та гексагональні СОК розміром $\ell_1 \times \ell_2$, які є окремим випадком чотирикутних СОК. Кількість коефіцієнтів СЛАР складає $N_A = 16n(\ell_1 - 1)(\ell_2 - 1)$, а її ранг $r = 16n(\ell_1 - 1)(\ell_2 - 1) - 2n(\ell_1 - \ell_2 - 2) - 8n$. СЛАР розв'язана за лінійний час методом прогонки відносно коефіцієнтів

$$\begin{aligned}
& \{A_{01}^1, A_{02}^1, A_{10}^1, A_{11}^1, A_{12}^1, A_{13}^1, A_{20}^1, A_{21}^1, A_{22}^1, A_{23}^1, A_{31}^1, A_{32}^1, A_{31}^{k_1}, A_{32}^{k_1}, A_{13}^{k_2(l_1-1)+1}, A_{23}^{k_2(l_1-1)+1}\} \\
& k_1 = \overline{2, \ell_1 - 1}, k_2 = \overline{1, \ell_2 - 2}.
\end{aligned} \quad (24)$$

Коефіцієнти (24) знайдені за лінійний час шляхом аналітичного розв'язання задачі оптимізації (23).

Запропоновано *метод відображення багатовимірних даних на одновимірну КГСОК*. Для відображення елемента X^c на двовимірну КГСОК (рис. 7) розв'язана система

$$\begin{cases} \frac{\partial S(t_1, t_2)}{\partial t_1} \cdot (S(t_1, t_2) - X^c) = 0, \\ \frac{\partial S(t_1, t_2)}{\partial t_2} \cdot (S(t_1, t_2) - X^c) = 0. \end{cases} \quad c = \overline{1, N} \quad (25)$$

Система (25) зведена до систем алгебричних рівнянь 5-го, 6-го та 6-го, 5-го порядків відносно параметрів t_1 і t_2 , які змінюються в області $\tilde{\Omega}^k$, $k = \overline{1, M}$, та розв'язана за лінійний час методом Ньютона.

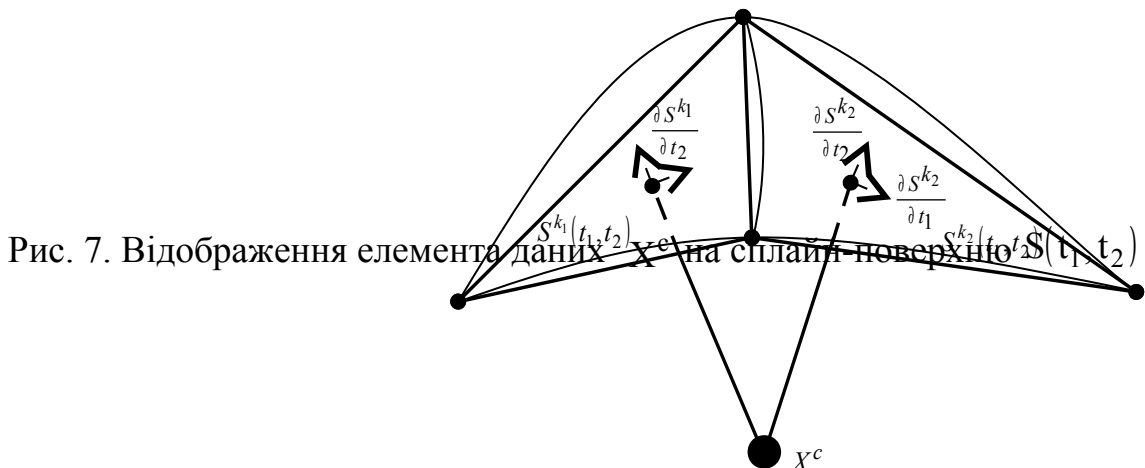


Рис. 7. Відображення елемента даних X^c на сплайн-поверхню $S(t_1, t_2)$

Для пошуку усіх коренів системи алгебраїчних рівнянь системи (25) в області $\tilde{\Omega}^k$, $k = \overline{1, M}$ запропоновано починати пошук коренів з декількох початкових приближень, рівномірно розташованих у $\tilde{\Omega}^k$, $k = \overline{1, M}$. Якщо на деякій ітерації метода Ньютона наближене значення кореня вийшло за область $\tilde{\Omega}^k$, $k = \overline{1, M}$, то цей корінь не розглядається. Якщо відстань між деякими двома коренями менше деякого ε , то корені об'єднуються.

Точка $\tilde{X}^c \in S(t_1, t_2)$ із множини коренів системи алгебраїчних рівнянь (25) є візуалізацією елемента X^c на двовимірну КГСОК, якщо вона розміщена ближче до елемента X^c відносно інших корів системи алгебраїчних рівнянь (25). Результатом ВБД є двовимірна множина $\tilde{X} = \{\tilde{X}^1, \dots, \tilde{X}^N\}$.

У четвертому розділі набув подальшого розвитку метод побудови діаграм Хінтона, проведено експериментальне дослідження та вирішені практичні задачі.

Набув подальшого розвитку *метод побудови діаграм Хінтона*, який дозволяє визначати відстань між багатовимірними даними на КГСОК з урахуванням її викривлення у багатовимірному просторі. Для обчислення відстаней між одновимірними даними, які отримані в наслідок ВБД за допомогою одновимірної КГСОК, з урахуванням її викривлення, використовується Риманова метрика та розраховується відстань між точками по поверхні одновимірної КГСОК

$$\rho_{RS}(\tilde{X}^i, \tilde{X}^j) = \begin{cases} \int_{\tilde{X}^i}^{\tilde{X}^j} \sqrt{\sum_{p=1}^n \left(\frac{ds_p^{[\tilde{X}^i]}(t)}{dt} \right)^2} dt + \int_{\tilde{X}^j}^{\tilde{X}^j} \sqrt{\sum_{p=1}^n \left(\frac{ds_p^{[\tilde{X}^j]}(t)}{dt} \right)^2} dt + \\ \quad + \int_{[\tilde{X}^j]}^{\tilde{X}^j} \sqrt{\sum_{p=1}^n \left(\frac{ds_p^{[\tilde{X}^j]}(t)}{dt} \right)^2} dt, & \text{при } [\tilde{X}^i] \neq [\tilde{X}^j], \\ \int_{\tilde{X}^i}^{\tilde{X}^j} \sqrt{\sum_{p=1}^n \left(\frac{ds_p^{[\tilde{X}^i]}(t)}{dt} \right)^2} dt, & \text{при } [\tilde{X}^i] = [\tilde{X}^j]. \end{cases} \quad (26)$$

Для наочного зображення викривлення одновимірної КГСОК на діаграмі Хінтона відображається її викривлення

$$|S(t)| = \sqrt{\sum_{p=1}^n s_p^k(t)}, \text{ при } t \in [k, k+1], k = \overline{1, \ell-1} \quad (27)$$

у відтінках сірого кольору.

Для визначення відстаней між даними на двовимірній КГСОК з урахуванням її викривлення введено метричний тензор $U\left(\tau, t_1(\tau), t_2(\tau), \frac{dt_1}{d\tau}, \frac{dt_2}{d\tau}\right)$, який є першою квадратичною формою двовимірної КГСОК $S(t_1, t_2)$ та описує її викривлення у точці (t_1, t_2) по кривій $L(\tau) = (t_1(\tau), t_2(\tau)) \in S(t_1, t_2)$. Пошук точної відстані між

елементами $\tilde{X}^i$ та $\tilde{X}^j \in S(t_1, t_2)$ по геодезичній кривій $L(\tau)$, що з'єднує ці елементи $L(\tau^0) \in \tilde{X}^i$, $L(\tau^1) \in \tilde{X}^j$, важкий у зв'язку з необхідністю розв'язувати варіаційну задачу

$$\int_{\tau^0}^{\tau^1} U\left(\tau, t_1(\tau), t_2(\tau), \frac{dt_1}{d\tau}, \frac{dt_2}{d\tau}\right) d\tau \rightarrow \min. \quad (28)$$

Для наближеного будування геодезичної кривої сформована найкоротша ламана лінія між елементами $\tilde{X}^i$ та $\tilde{X}^j$ на двовимірній КГСОК.

На двовимірну діаграму Хінтона нанесена довжина $|S(t_1, t_2)|$ тензора U по напрямку $(t_1(t), t_2(t)) = (t_1, t_2)$.

Для проведення експериментальної ВБД за допомогою КГСОК вибрана база даних компаній із країн СНД. Для візуалізації багатовимірних даних вибрана прямокутна СОК розміром $\ell_1 = 9$, $\ell_2 = 9$, яка побудована шляхом навчання нейронної мережі Кохонена з використанням модифікованого алгоритму регуляризації з параметрами $\alpha_0 = 0,1$, $\tilde{\alpha}_0 = 0,05$. Навчання тривало $T = 500$ епох.

Візуалізація даних дозволила розбити їх на кластери, виявити стабільні, інтенсивно розвиваючі компанії, компанії-новачки тощо. Проведений аналіз визначив якісну відміну українського та російського ринку полімерів. Порівняльний аналіз методів візуалізації багатовимірних даних довів, що КГСОК візуалізують багатовимірні дані порівняно з методом головних компонент у 4,2 рази точніше, але в 2,5 разів довше, порівняно з КЛСОК; в 1,2 рази точніше, але в 1,7 разів довше, у порівнянні з картами Саммона в 2,1 разів точніше, але у 22,2 рази швидше. Таким чином, була показана ефективність використання КГСОК для візуалізації багатовимірних даних великого об'єму.

Розроблені в роботі методи використані для розв'язання практичних задач. Запропоновано розв'язання задачі зменшення ризиків кредитування банками компаній на ринку полімерів із країн СНД. Також результати роботи були використані для розв'язку задачі ефективного планування та розробки тарифів оператора мобільного зв'язку ПАТ «Укртелеком».

ВИСНОВКИ

У дисертаційній роботі наведені результати, які, відповідно до мети дослідження, в сукупності є вирішенням актуальної науково-практичної задачі – повного розрізнення великої кількості багатовимірних даних зі складною топологією на карті під час їх візуалізації за допомогою самоорганізуючих карт Кохонена, що має велике значення для збільшення точності візуалізації багатовимірних даних. Отримані наукові результати:

1. Новий метод побудови одновимірних кусочно-гладких самоорганізуючих карт Кохонена шляхом апроксимації одновимірних кусочно-лінійних карт Кохонена кубічними параметричними сплайнами для анулювання зломів у нейронах, що дозволило збільшити точність візуалізації багатовимірних даних.

2. Новий метод побудови двовимірних кусочно-гладких самоорганізуючих карт Кохонена на триангуляції Делоне без зломів у нейронах та на лініях дотику трикутників із триангуляції Делоне шляхом апроксимації двовимірних кусочно-гладких карт Кохонена кубічними параметричними сплайн-поверхнями, що призвело до збільшення точності візуалізації багатовимірних даних.

3. Новий метод побудови двовимірних кусочно-гладких самоорганізуючих карт Кохонена на чотирикутних СОК шляхом побудови остового лісу ізоморфної триангуляції Делоне, що дозволяє зменшити кількість скалярних операцій.

4. Новий метод відображення багатовимірних даних на одновимірну кусочно-гладку самоорганізуючу карту Кохонена заснований на синтезі методів локалізації та пошуку коренів рівняння. Це дозволяє розрізняти дані на карті. Обчислювальна складність запропонованого методу лінійно залежить від початкових даних.

5. Новий метод відображення багатовимірних даних на двовимірну кусочно-гладку самоорганізуючу карту Кохонена заснований на використанні методу Ньютона з додаванням початкових наближень, обчислювальна складність якого лінійно залежить від початкових даних. Це дозволило розрізняти більшу кількість багатовимірних даних на карті та зменшити похибку візуалізації даних.

6. Набув подальшого розвитку метод визначення відстаней між елементами даних на кусочно-гладких самоорганізуючих картах Кохонена з урахуванням її викривлення у просторі даних, що дозволило збільшити точність візуалізації багатовимірних даних.

7. Проведено експериментальні дослідження запропонованих методів, які дозволили візуалізувати багатовимірні дані за більш короткий проміжок часу, ніж карти Саммона та метод головних різноманіт і з більш високою точністю, ніж метод головних компонент та самоорганізуючі карти Кохонена. Результати роботи були впроваджені і довели свою ефективність під час аналізу бази даних компаній із країн СНД на ринку полімерів та бази даних абонентів ПІБ «Укртелеком».

СПИСОК ОПУБЛІКОВАНИХ ПРАЦЬ ЗА ТЕМОЮ ДИСЕРТАЦІЇ

1. Аксак Н.Г., Метод аппроксимации линейной карты Кохонена кубическим параметрическим сплайном/ Н. Г. Аксак, А. В. Шкловец // Системи управління, навігації та зв'язку - Випуск 2 (14). – Київ, 2010. – С. 70-74.

2. Шкловец А. В. Построение четырёхугольных карт Кохонена на основе триангуляции Делоне для визуализации многомерных данных/ А. В. Шкловец, Н. Г. Аксак. // Бионика интеллекта. науч. техн. журнал. – 2011. – 3(77). – С. 94 – 97.

3. Шкловец А. В. Метод отображения многомерных данных на одномерную кусочно-гладкую самоорганизующуюся карту Кохонена/ А. В. Шкловец, Н. Г. Аксак // Системи управління, навігації та зв'язку - Випуск 1 (21). – Київ, 2012. – С. 70-74.

4. Шкловец А. В. Метод построения кусочно-гладких карт на основе четырёхугольных кусочно-плоских карт Кохонена / А. В. Шкловец. // Системи обробки інформації. – Випуск 2 (100). – Харків, 2012. – С. 168-175.

5. Шкловец А. В. Проецирование большого количества многомерных данных на двумерную кусочно-гладкую самоорганизующуюся карту Кохонена / [Текст] А. В. Шкловец, Н. Г. Аксак. // Системи обробки інформації. – Випуск 3 (101). – Харків, 2012. – С. 127-132.

6. Шкловец А. В. Определение расстояний между точками на кусочно-гладких картах Кохонена / А. В. Шкловец, Н. Г. Аксак. // Бионика интеллекта. науч. техн. журнал. – 2012. – 1(78). – С. 63 – 67.

7. Шкловец А.В. Аппроксимация двумерных карт Кохонена кубическими сплайн поверхностями / А. В. Шкловец, Н. Г. Аксак // Международная научная конференция «Интеллектуальные системы принятия решений и проблемы вычислительного интеллекта (ISDMCI'2010)»: сб. науч. тр. – Евпатория: МОИУ, 2010. – С. 225-228.

8. Шкловец А. В. Метод аппроксимации сплайнами минимальной длины линейных карт Кохонена для визуализации многомерных данных / [Текст] А. В. Шкловец, Н. Г. Аксак // XIII всероссийская научно-техническая конференция «Нейроинформатика-2011»: Сб. науч. трудов. В 3 частях. Ч.1. М.: НИЯУ МИФИ, 2011. – С. 69-78.

9. Шкловец А. В. Метод аппроксимации сплайн поверхностями кусочно-плоских карт Кохонена для визуализации многомерных данных / А.В. Шкловец, Н.Г. Аксак // XIV Всероссийская научно-техническая конференция "Нейроинформатика-2012": Сб. науч. трудов. В 3-х частях. Ч.1. М.: НИЯУ МИФИ, 2012.-С.63-73.

10. Аксак Н. Г. Метод отображения многомерных данных на двумерную кусочно-гладкую самоорганизующуюся карту Кохонена / Н. Г. Аксак, А. В. Шкловец // Информатика, математическое моделирование, экономика: Сборник научных статей по итогам второй Международной научно-практической конференции, г. Смоленск, 20 апреля 2012 г. В 3-х томах. Том 1 – Смоленск: Смоленский филиал АНО ВПО ЦС РФ "Российский университет кооперации", 2012. – с. 109-114с.

11. Шкловец А. В. Решение задачи коммивояжера на многоядерных архитектурах с помощью растущей сети Кохонена / А. В. Шкловец, Н. Г. Аксак // Технологии Microsoft в теории и практике программирования. Материалы конференции / Под ред. проф. В.П. Гергеля. – Нижний Новгород: Изд-во Нижегородского госуниверситета, 2009. С.505-511.

12. Аксак Н.Г. Задача коммивояжера на основе растущей сети Кохонена. Многоядерное решение/ Н. Г. Аксак, А. В. Шкловец //Вестник ХНТУ №1(34)-2009. – С.342-346.

13. Аксак Н.Г. Параллельный алгоритм формирования самоорганизующихся карт Кохонена / Н. Г. Аксак, А. В. Шкловец //9-ая Международная Конференция-семинар. Высокопроизводительные параллельные вычисления на кластерных системах (НРС-2009). Владимир, 2009. Труды конференции- Владимир: Изд. ВГУ2009. – С.27-30.

14. Аксак Н. Г. Метод аппроксимации сплайн поверхностью двух переменных на нерегулярной решетке / Н. Г. Аксак, А. В. Шкловец // 17- Міжнародна конференція з автоматичного управління "Автоматика -2010". Тези доповідей. Том 1. - Харків:ХНУРЕ, 2010. –С.26-27.

15. Шкловец А. В. Визуализация многомерных данных сложной топологии с помощью кусочно-гладких линейных карт Кохонена / А. В. Шкловец // Матеріали першої науково-технічної конференції «Сучасні напрями розвитку інформаційно-

комунікаційних технологій та засобів управління». Тези доповідей. – Харьков-Киев: ХНДІТМ. 2010. – С. 81.

16. Шкловец А. В. Анализ методов визуализации многомерных данных / А.В. Шкловец, Н.Г. Аксак // IX Всероссийская научная конференция "Нейрокомпьютеры и их применение". Тезисы докладов. – М.: МГППУ, 2011. – С.25.

17. Шкловец А. В. Методы решения задачи визуализации многомерных данных / А. В. Шкловец // 15-ый Юбилейный Международный молодежный форум «Радиоэлектроника и молодежь в XXI веке». Сб. материалов форума. Т.9. – Харьков: ХНУРЭ. 2011 –с.660-661.

18. Шкловец А. В. Визуализация базы данных компаний из стран СНГ на рынке полимеров / А. В. Шкловец // 16-ый Международный молодежный форум «Радиоэлектроника и молодежь в XXI веке». Сб. материалов форума. Сб. материалов форума Т.10. – Харьков: ХНУРЭ. 2012 –с. 63-64.

АНОТАЦІЯ

Шкловец А. В. Кусочно-гладкі самоорганізуючі карти Кохонена для візуалізації багатовимірних даних. – На правах рукопису.

Дисертація на здобуття наукового ступеня кандидата технічних наук за спеціальністю 05.13.23. – системи та засоби штучного інтелекту. – Харківський національний університет радіоелектроніки Міністерства освіти і науки, молоді та спорту України, Харків, 2012.

Дисертаційна робота присвячена розв’язку задачі розрізнення багатовимірних даних зі складною топологією на карті під час розв’язку задачі візуалізації великої кількості багатовимірних даних за допомогою кус очно-лінійних самоорганізуючих карт Кохонена, який полягає у побудові кусочно-гладких самоорганізуючих карт Кохонена та проектування даних на них.

Запропоновано метод побудови кусочно-гладких самоорганізуючих карт Кохонена шляхом апроксимації кусочно-лінійних самоорганізуючих карт Кохонена кубічними параметричними сплайнами. Запропоновано метод відображення багатовимірних даних на одновимірну кусочно-гладку самоорганізуючу карту Кохонена. Отримані співвідношення для побудови прямокутних та гексагональних кусочно-гладких самоорганізуючих карт Кохонена.

Проведені обчислювальні експерименти підтвердили ефективність використання кус очно-гладких самоорганізуючих карт Кохонена для візуалізації великої кількості багатовимірних даних.

Результати роботи використані для розв’язання задачі оцінки ризиків кредитування банками компаній на ринку полімерів із країн СНД, а також при розв’язку задачі ефективного планування та розробки тарифів оператора мобільного зв’язку ПІБ «Укртелеком».

Ключові слова: візуалізація багатовимірних даних, самоорганізуюча карта Кохонена, кубічний параметричний сплайн, кубічна параметрична сплайн-поверхня, тріангуляція Делоне.

АННОТАЦИЯ

Шкловец А. В. Метод построения кусочно-гладких самоорганизующихся карт Кохонена для визуализации многомерных данных. – На правах рукописи.

Диссертация на соискание ученой степени кандидата технических наук по специальности 05.13.23. – системы и средства искусственного интеллекта. – Харьковский национальный университет радиоэлектроники Министерства образования и науки, молодежи и спорта, Харьков, 2012.

В диссертационной работе осуществлено решение научной задачи различимости многомерных данных со сложной топологией на карте при решении задачи визуализации большого числа многомерных данных с помощью кусочно-линейных самоорганизующихся карт Кохонена, заключающееся в построении кусочно-гладких самоорганизующихся карт Кохонена и проецировании данных на них. Использование кусочно-гладких самоорганизующихся карт Кохонена для визуализации многомерных данных позволяет повысить точность визуализации по сравнению с кусочно-линейными самоорганизующимися картами Кохонена при незначительном увеличении количества скалярных операций.

Рассмотрены существующие методы визуализации многомерных данных, указаны их области применимости, достоинства и недостатки, заключающиеся в линейной или кусочно-линейной структуре карты, приводящей к низкой точности визуализации многомерных данных, и значительными вычислительными трудностями, возникающими при реализации методов визуализации большого числа многомерных данных. Сделан вывод о целесообразности использования кусочно-гладких самоорганизующихся карт Кохонена для визуализации многомерных данных.

Предложен метод построения одномерных кусочно-гладких самоорганизующихся карт Кохонена путем аппроксимации одномерных кусочно-линейных самоорганизующихся карт Кохонена кубическими параметрическими сплайнами. Предложен метод отображения многомерных данных на одномерную кусочно-гладкую самоорганизующуюся карту Кохонена путем решения соответствующих уравнений 5-го порядка. Предложенные методы позволили повысить точность визуализации многомерных данных при небольшом увеличении количества скалярных операций.

Предложен метод построения двумерных кусочно-гладких самоорганизующихся карт Кохонена в виде триангуляции Делоне путем аппроксимации двумерных кусочно-линейных самоорганизующихся карт Кохонена кубическими параметрическими сплайн-поверхностями. Для ускорения построения двумерных кусочно-гладких самоорганизующихся карт Кохонена предложен метод построения четырехугольных самоорганизующихся карт Кохонена путем построения остового леса графа, изоморфного триангуляции Делоне двумерной кусочно-линейной самоорганизующейся карты Кохонена. Получены соотношения для построения прямоугольных и гексагональных кусочно-гладких самоорганизующихся карт Кохонена. Разработан метод отображения многомерных данных на двумерную кусочно-гладкую самоорганизующуюся карту Кохонена путем решения соответствующих систем нелинейных уравнений. Предложенные методы повысили точность визуализации многомерных данных при незначительном увеличении количества скалярных операций по сравнению с двумерными кусочно-линейными самоорганизующимися картами Кохонена.

Для увеличения точности отображения многомерных данных в пространство малой размерности предложен метод вычисления расстояний между элементами данных с учетом искажения кусочно-гладкой самоорганизующейся карты Кохонена.

Проведенные вычислительные эксперименты подтвердили эффективность использования кусочно-гладких самоорганизующихся карт Кохонена для визуализации большого количества многомерных данных.

Результаты работы использованы при решении задачи оценки рисков кредитования банками компаний на рынке полимеров из стран СНГ, а также при решении задачи эффективного планирования и разработки тарифов оператора мобильной связи ОАО «Укртелеком».

Ключевые слова: визуализация многомерных данных, самоорганизующиеся карты Кохонена, кубический параметрический сплайн, кубическая параметрическая сплайн поверхность, триангуляция Делоне.

ABSTRACT

Shklovets A. V. Piecewise-smooth self organization Kohonen maps for multidimensional data visualization. – A manuscript.

Thesis for candidate's degree in engineering science by specialty 05.13.23 – system and methods of artificial intelligence. – Kharkiv National University of Radio Electronics of the Ministry of Education and Science, Youth and Sport of Ukraine, Kharkiv, 2011.

Dissertation work is devoted to solving the problem of distinguish ability of multi-dimensional data with complex topology map for solving the problem of rendering a large number of multidimensional data using a piecewise-linear self-organizing Kohonen maps, consisting in the construction of piecewise smooth self-organizing Kohonen maps and projected data on them.

The existing methods for visualizing multivariate data, based on the shortcomings of which concluded that the feasibility of using a piecewise-smooth self-organizing Kohonen maps was considered.

A method for constructing piecewise-smooth Kohonen self-organizing maps by fitting one-dimensional piecewise-linear self-organizing Kohonen maps cubic parametric splines was proposed. A method for displaying multidimensional data to one-dimensional piecewise-smooth self-organizing Kohonen map was proposed. Relations are obtained for the construction of rectangular and hexagonal piecewise smooth self-organizing Kohonen maps.

The results are used to solve the problem of risk assessment by banks lending to companies in the polymer from the CIS countries, as well as in solving the problem of effective planning and development rates of mobile operator JSC "Ukrtelecom".

Key words: multidimensional data visualization, self organization Kohonen maps, cubic parametric splines, cubic parametric spline surfaces, Delaunay triangulation.

Відповідальний випусковий В.П. Машталір

Підп. до друку 27.05.2011. Формат 60х84 1/16. Спосіб друку – ризографія.

Умов.друк.арк. 1,2. Облік. вид.арк. 1,1. Тираж 100 прим.

Ціна договірна Зам №2-453

ХНУРЕ. Україна. 61166, Харків, просп. Леніна, 14

Віддруковано в навчально-науковому
видавничо-поліграфічному центрі ХНУРЕ

61166, Харків, просп. Леніна, 14