

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
СХІДНОУКРАЇНСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ ІМ. В. ДАЛЯ
ФАКУЛЬТЕТ ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ ТА ЕЛЕКТРОНІКИ
КАФЕДРА КОМП'ЮТЕРНИХ НАУК ТА ІНЖЕНЕРІЇ

До захисту допускається
Т.в.о. завідувача кафедри
_____ Кардашук В.С.
« ____ » _____ 20__ р.

МАГІСТЕРСЬКА РОБОТА

НА ТЕМУ:

Метод нечіткої кластеризації даних з пропусками

Освітній рівень “Магістр”
Спеціальність 122 “Комп’ютерні науки”

Науковий керівник роботи:

(підпис)

О.І.Рязанцев

(ініціали, прізвище)

Консультант з охорони праці:

(підпис)

Я.О.Критська

(ініціали, прізвище)

Студент:

(підпис)

Халілова Айнур Гіяс кизи

(ініціали, прізвище)

Група:

КН-19ім

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
СХІДНОУКРАЇНСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
ІМЕНІ ВОЛОДИМИРА ДАЛЯ

Факультет Інформаційних технологій та електроніки

Кафедра Комп'ютерних наук та інженерії

Освітній рівень магістр

Напрямок підготовки _____

(шифр і назва)

Спеціальність 122 "Комп'ютерні науки"

(шифр і назва)

ЗАТВЕРДЖУЮ:

Т.в.о. завідувача кафедри _____

В.С.Кардашук

« _____ » _____ 20 ____ р.

**З А В Д А Н Н Я
НА МАГІСТЕРСЬКУ РОБОТУ СТУДЕНТУ**

Халілова Айнур Гіяс кизи

(прізвище, ім'я, по батькові)

1. Тема роботи Метод нечіткої кластеризації даних з пропусками

керівник проекту (роботи) Рязанцев Олександр Іванович, д.т.н., проф.

(прізвище, м.я, по батькові, науковий ступінь, вчене звання)

затверджені наказом вищого навчального закладу від «11» 10 2020 р. № 136/15.15

2. Строк подання студентом роботи 10.01.2021

3. Вихідні дані до роботи Матеріали науково-дослідної практики, карти

Кохонена, адаптивний метод ймовірнісної нечіткої кластеризації даних, перелік
використовуваних програмних засобів: програмна платформа MATLAB, UCI Iris
repository, UCI Wine repository

4. Зміст розрахунково-пояснювальної записки (перелік питань, які потрібно розробити) Огляд проблеми та постановка задачі дослідження, адаптивні
нейро-фаззі методи для кластеризації викривлених даних, програмна реалізація,
охорона праці та безпека в надзвичайних ситуаціях, висновки

5. Перелік графічного матеріалу (з точним зазначенням обов'язкових креслень)

Електронні плакати

6. Консультанти розділів проекту (роботи)

Розділ	Прізвище, ініціали та посада консультанта	Підпис, дата	
		завдання видав	завдання прийняв
Охорона праці та безпека в надзвичайних ситуаціях	Критська Я.О. ст. викл. кафедри КНІ		

7. Дата видачі завдання 14.10.2020

Керівник

_____ (підпис)

Завдання прийняв до виконання

_____ (підпис)

КАЛЕНДАРНИЙ ПЛАН

№ з/п	Назва етапів дипломного проекту (роботи)	Строк виконання етапів проекту (роботи)	Примітка
1	Розробка технічного завдання	02.09.2020-15.09.2020	
2	Аналіз літератури з досліджуваної проблеми	16.09.2020-22.09.2020	
3	Аналіз технічних засобів та розробка методу	23.09.2020-25.09.2020	
4	Програмна реалізація	26.09.2020-06.10.2020	
5	Аналіз результатів дослідження	07.10.2020-25.11.2020	
6	Розробка частини проекту "Охорона праці та безпеки в надзвичайних ситуаціях"	26.11.2020-1.12.2020	
7	Оформлення пояснювальної записки, автореферату та презентації	2.12.2020-09.01.2021	

Студент

_____ (підпис)

Халілова Айнур Гіяс кизи

_____ (прізвище та ініціали)

Науковий керівник

_____ (підпис)

Рязанцев О.І.

_____ (прізвище та ініціали)

АНОТАЦІЯ

Халілова Айнур Гіяс кизи. Метод нечіткої кластеризації даних з пропусками.

Метою роботи є розробка ймовірнісного методу нечіткої on-line кластеризації даних що може бути вирішена на основі самоорганізованої нейро-фаззи мережі Кохонена.

Об'єктом дослідження є послідовність вибірки даних UCI репозиторія, а саме такі як Iris та Wine.

Використано методи нечіткої кластеризації даних та само організовані карти Кохонена. Проведено дослідження методів кластеризації даних, а також порівняльний аналіз роботи відомих методів кластеризації та запропонованого методу по критеріям якості кластеризації.

У результаті роботи розглянуто задачу ймовірнісної нечіткої on-line кластеризації даних і показано, що вона може бути вирішена на основі само організованої нейро-фаззи мережі Кохонена та адаптивного алгоритму самонавчання, який було запропоновано, що є гібридом правила «переможець отримує більше» і рекурентних алгоритмів нечіткої кластеризації.

Ключові слова: нечітка кластеризація, самоорганізовна мережа кохонена, правила навчання.

ABSTRACT

Khalilova Ainur Gias kizi. Fuzzy data clustering method with gaps /

The purpose of the paper is to develop a probabilistic method of fuzzy on-line data clustering that can be solved on the basis of the Kohonen network self-organizing neuro-phase.

The object of the study is a sample sequence of UCI repository data, such as Iris and Wine.

Fuzzy data clustering methods and self-organized Kohonen maps are used. The methods of data clustering were investigated, as well as a comparative analysis of the work of known clustering methods and the proposed method on the criteria of clustering quality.

As a result, the problem of probabilistic fuzzy on-line data clustering is considered, and it is shown that it can be solved on the basis of Kohonen's self-organized neuro-phase network and the proposed adaptive self-learning algorithm, which is a hybrid of the winner-receive-more rule and recurrent fuzzy algorithms.

Keywords: fuzzy clustering, self-organizing kohonen network, training rules.

ЗМІСТ

ПЕРЕЛІК СКОРОЧЕНЬ ТА УМОВНИХ ПОЗНАЧЕНЬ.....	6
ВСТУП	7
1 ОГЛЯД ПРОБЛЕМИ ТА ПОСТАНОВКА ЗАДАЧІ ДОСЛІДЖЕННЯ	8
1.1 Тип і характер навчання	8
1.2 Нечітка логіка і системи нечіткого виведення.....	11
1.3 Нейромережеві і нейро-фаззі моделі для класифікації даних і їх навчання	11
1.3.1 Самоорганізовані карти Кохонена	13
1.3.2 Нейро-фаззі системи.....	18
1.4 Постановка задачі дослідження.....	22
2 АДАПТИВНІ НЕЙРО-ФАЗЗІ МЕТОДИ ДЛЯ КЛАСТЕРИЗАЦІЇ ВИКРИВЛЕНИХ ДАНИХ.....	23
2.1 Формальна постановка задачі кластеризації.....	23
2.2 Функції відстані.....	24
2.2.1 Евклідова відстань	24
2.2.2 Відстань міських кварталів (Манхеттенська відстань).....	25
2.2.3 Відстань Чебишева	25
2.2.4 Часткова відстань.....	25
2.2.5 Методи кластеризації даних	26
2.2.6 Адаптивний метод ймовірнісної нечіткої кластеризації даних	26
3 ПРОГРАМНА РЕАЛІЗАЦІЯ	29
3.1 MATLAB	29
3.2 DATA BASE (Виборка)	31
3.2.1 Iris	31
3.2.2 Wine	32
3.3 Експериментальні дослідження.....	33
4 ОХОРОНА ПРАЦІ	37

4.1 Аналіз потенційних небезпечних і шкідливих виробничих чинників проектованого об'єкту, що мають вплив на персонал	37
4.2 Заходи щодо техніки безпеки	38
4.3 Заходи, що забезпечують виробничу санітарію і гігієну праці.....	41
4.4 Рекомендації по пожежній безпеці	44
4.5 Вплив на навколишнє природне середовище.....	47
ВИСНОВКИ.....	48
ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАНЬ.....	49
ДОДАТОК А. Лістинг коду	57
ДОДАТОК Б. Електронні плакати	61

ПЕРЕЛІК СКОРОЧЕНЬ ТА УМОВНИХ ПОЗНАЧЕНЬ

НМ – нейронні мережі

ШНМ - штучні нейронні мережі

РБНМ – радіально-базисні нейронні мережі

НРБНМ – нормовані радіально-базисні нейронні мережі

УРНС – узагальнені регресійні нейронні мережі

ННМ – нейронні нечіткі мережі

РБФ – радіально-базисні функції

NEFCLASS – Neuro Fuzzy CLASSifier

ВСТУП

Для вирішення складних завдань, що часто не мають аналітичного вирішення, на сьогоднішній день нерідко і вдало використовують технології обчислювального інтелекту. Ці технології і, особливо, штучні нейронні мережі (ШНМ) стають все поширенішими для вирішення різних завдань обробки сигналів, оптимізації, оптимального і адаптивного управління, розпізнавання образів, ідентифікації, прогнозування в реальному часі і т.д. Створено реальні системи обробки зображень та комп'ютерного зору, управління аерокосмічними об'єктами, діагностики технічної та медичної, у фінансах та економіці, у військовій справі, управління дорожнім рухом, в енергетичній промисловості, криміналістиці, аналізу сигналів різноманітної природи походження та інше, причому цей список стає довшим кожного дня.

Сьогодні є достатня кількість інформації про підприємницьку діяльність підприємств, лікарень, фірм, яка показує діяльність цих закладів.

В наслідок розгляду інформації стає можливим побачити закономірності при умовах того, що зроблена таблиця показує фактичні дані, що демонструють причинно-наслідкові залежності.

Ця інформація збирається протягом дового часу (темпи інфляції, рівень прибутку населення, витрати народу, ціну послуг житлово-комунального господарства, стан промислового і сільськогосподарського виробництва, своєчасність виплати заробітної плати та пенсій, тощо). Зрозуміло, такі дані складно обробити вручну через великий обсяг інформації і складних нелінійних причинно-наслідкових зв'язків. Тому і виникла потреба у створенні новіших методів аналізу, кластеризації і прогнозування, до яких відносять машинні методи виявлення послідовних зв'язків.

1 ОГЛЯД ПРОБЛЕМИ ТА ПОСТАНОВКА ЗАДАЧІ ДОСЛІДЖЕННЯ

Однією з головних властивостей біологічного мозку є його здібність до навчання, і через те, що штучна нейронна мережа моделює роботу мозку «навчання» є ключовим поняттям в теорії штучних нейронних мереж. Процес навчання має сталий характер. З плином часу мережа робить краще свої характеристики, поступово приходячи до найбільш ефективного рішення поставленої задачі.

Нейронні мережі використовують для вирішення складних задач, які потребують аналітичних обчислень схожих тим, що робить справній мозок. Найчасішими вживання нейронних мереж:

1) класифікація – розподіл даних по параметрам. Наприклад, на вхід подається список людей і потрібно прийняти рішення, кому з них одобрити кредит, а кому ні. Цю роботу може виконати нейронна мережа, обробляючи таку інформацію як: вік, платоспроможність, кредитна історія і тд,

2) передбачення – можливість передбачити наступний крок. Наприклад, зростання або падіння акцій, ґрунтуючись на ситуацію фондового ринку,

3) розпізнавання – на даний момент, саме популярне використання нейронних мереж. Застосовується в Google, коли ви шукаєте фотографію або в камерах телефонів, коли воно визначає положення вашого обличчя і виділяє його і багато іншого.

1.1 Тип і характер навчання

Тип і характер навчання визначаються передусім об'ємом апріорної і поточної інформації про середовище, в якій знаходиться мережа, а також цільовою функцією (критерієм якості), що характеризує ступінь доречності нейронної мережі до розв'язуваної нею задачі. Інформація про зовнішнє середовище задана, як правило, у виді навчальної вибірки образів або прикладів.

На даний момент існує кілька головних парадигм навчання штучних нейронних мереж таких, як навчання «з учителем», «без учителя», або «самонавчання», а так само навчання «з підкріпленням».

Найбільш поширеною і очевидною до тепер є парадигма навчання «з учителем». У даній парадигмі «вчителю» відома інформація про зовнішнє середовище. Вона подана у вигляді деякої послідовності даних або в пакетах вхідних векторів x , а також реакції на ці сигнали, показана у вигляді навчального сигналу d .

Альтернативою цій парадигмі є навчання «без вчителя», або самонавчання, коли правильна реакція на сигнали зовнішнього середовища невідома. В даний час існує декілька основних парадигм навчання штучних нейронних мереж таких, як навчання «з учителем», «без вчителя», або «самонавчання», а так само навчання «з підкріпленням». Найбільш поширеною і очевидною до тепер є парадигма навчання «з учителем». У даній парадигмі «вчителю» відома інформація про зовнішнє середовище. Вона задана у вигляді деякої послідовності даних або в пакетах вхідних векторів x , а також реакція на ці сигнали, показана у вигляді навчального сигналу d . Альтернативою цій парадигмі є навчання «без вчителя», або самонавчання, коли правильна реакція на сигнали зовнішнього середовища невідома.

Можливість самонавчання являється необхідним атрибутом нейронних мереж (НМ) для аналізу і знаходження скритих закономірностей в багатовимірній інформації, що властиві більшості завдань інтелектуального аналізу даних, зокрема, задачам класифікації та кластеризації.

Своєрідним компромісом між парадигмами «самонавчання» і «навчання з учителем» є навчання з підкріпленням, при якому доступна лише непряма інформація про правильні реакції на вхідний сигнал x .

Процес навчання з підкріпленням розділяється на два етапи:

- 1) навчання тому, як вихідний сигнал мережі впливає на спостережувані середовища, що змінюються, тобто відновлення зображення \tilde{F} ;
- 2) навчання мережі на базі мінімізації прийнятого критерію.

У великій кількості додатків, пов'язаних з управлінням технологічними виробничими процесами, функціонально-наближена архітектура мережі достатньо проста, і вона успішно справляється з моделюванням виробничих процесів. Нейронні мережі повинні бути привчені маючи набори даних, отриманих за допомогою найкращого з можливих варіантів параметричного тестування. Щоб зімітувати управління процесом відповідним чином, користувач обирає змінну процесу і робить набір даних для навчання. Цей набір повинен закрити весь можливий діапазон навчання і відповідати його функціональному призначенню. Набір даних містить дані, пов'язані з входами, і корелює їх до цільового призначення. Входи, які використовуються у навчанні, демонструють простір управління нейронної мережевої моделі.

Всі записи набору даних виконуються з тимчасовими мітками, що визначають їх цільовий відлік до часу захоплення і значення параметра входу. Набори даних мають бути прив'язані до конфігурації обладнання, яка мала місце в момент покупки системи. Для покращення надійності моделі, ліквідації перешкод і зниження вартості критично велике значення має правильний вибір моделі входів. При цьому калібрування повинне бути завершено для всіх входів ще до поступання даних. Це потрібно для того, щоб встановити їх базовий опорний рівень.

Після того як набір даних був отриманий, він повинен бути розібраний на три підмножини: навчання, випробування і валідацію. Навчання нейронної мережі передбачає зворотне поширення помилки між виходом моделі і цільовим значенням. Навчання є ітеративним процесом зворотнього поширення, який вимагає стратегії, що робить упор на надійність і підтвердження достовірності. Значення (ваги) і їх зміщення для кожного нейрона надаються в процесі навчання. Звіт про валідацію має бути розглянутий ще до імплементації системи на платформі управління. Цей звіт повинен показати збіг цільової функції управління у фазі з моделлю при мінімальній похибці.

Більшість підходів навчання нейронної мережі працюють на настільних комп'ютерах, а основні платформи управління вже мають весь необхідний набір команд для налаштування моделі нейронної мережі. Там, де це можливо, кодування нейронів на платформі управління повинно бути виконано в підпрограмі, що дозволяє повторне застосування коду на майбутніх моделях. У працюючих моделях всі входи повинні бути перевірені на відповідні якісні показники і допустимий простір контролю. Якщо входи знаходяться не в колі допустимого простору контролю, то повинна передбачатись стратегія альтернативного управління.

НМІ повинен показувати вихід моделі, входи, сигнали тривоги, технологічні уставки процесу, давати можливість вибору режиму (автоматичний / ручний) і стратегії управління (звичайна / альтернативна). Тенденції і логіка з обслуговування в порядку поступання є достатньо важливим рішенням для оперативного інформування та діагностики. Для введення необхідних констант, сигналізації і обмежень краще використовувати окремий екран налаштування.

1.2 Нечітка логіка і системи нечіткого виведення

Нечітка логіка вперше була запропонована Л. Заде для подання нечітких і неточних знань. Вона надає ефективний спосіб опису поведінки дуже складних, погано визначених або погано формалізованих математичних систем. Маніпуляції з нечіткими змінними утворюють формальну систему, яку називають нечіткою логікою. Нечітка логіка має у своєму складі: фазифікації - переклад чіткого вхідного значення в нечітке, нечіткий висновок вилучення даних з бази знань, а також дефазифікація - переводить нечіткий результат в чіткий.

Головною перевагою нечітких множин можна вважати можливість лінгвістичної інтерпретації. А саме, використовуючи нечіткі множини, вдається легко визначати різні лінгвістичні терміни, які можуть мати нечіткі значення.

1.3 Нейромережеві і нейро-фаззі моделі для класифікації даних і їх навчання

На сьогоднішній день ШНМ користуються попитом для рішення задач ідентифікації, прогнозування та управління нелінійними об'єктами.

Більшість реальних завдань описуються тим, що дані надходять на обробку послідовно. В даному випадку альтернативою може бути радіально-базисна нейронна мережа (РБНМ), тому що класичний багат шаровий персептрон не підходить. Ці мережі ще є універсальними апроксиматорами. Для їх навчання в реальному часі може бути використано рекурентний метод найменших квадратів або його модифікації, які по суті є алгоритмами оптимізації другого порядку, які забезпечують квадратичну збіжність до оптимального рішення. Практичне використання РБНМ обмежується, так званим, прокляттям розмірності, а також появою пропусків в просторі радіально-базисних функцій, що призводять до появи областей, де всі нейрони мережі не являються активними. Уникнути схожих ситуацій можна, скориставшись одиначною розбивкою простору, що реалізовується за допомогою нормалізованих радіально-базисних нейронних мереж (НРБНМ), в яких вихідний сигнал нормується на суму виходів всіх нейронів. Ці мережі навчаються за допомогою рекурентних градієнтних алгоритмів, загальним недоліком яких є не велика швидкість збіжності і можливість попадання в локальні

мінімуми прийнятого критерію навчання. На сьогоднішній день ШНМ користуються попитом для рішення задач ідентифікації, прогнозування та управління нелінійними об'єктами. Більшість реальних завдань описуються тим, що дані надходять на обробку послідовно. В даному випадку альтернативою може бути радіально-базисна нейронна мережа (РБНМ), що показана на рис. 1.1, тому що класичний багатошаровий перцептрон не підходить. Ці мережі ще є універсальними апроксиматорами. Для їх навчання в реальному часі може бути використано рекурентний метод найменших квадратів або його модифікації, які по суті є алгоритмами оптимізації другого порядку, які забезпечують квадратичну збіжність до оптимального рішення. Практичне використання РБНМ обмежується, так званим, прокляттям розмірності, а також появою пропусків в просторі радіально-базисних функцій, що призводять до появи областей, де всі нейрони мережі не являються активними. Уникнути схожих ситуацій можна, скориставшись одиначною розбивкою простору, що реалізовується за допомогою нормалізованих радіально-базисних нейронних мереж (НРБНМ), в яких вихідний сигнал нормується на суму виходів всіх нейронів. Ці мережі навчаються за допомогою рекурентних градієнтних алгоритмів, загальним недоліком яких є не велика швидкість збіжності і можливість попадання в локальні мінімуми прийнятого критерію навчання.

Вхідні шари такої мережі – це сенсори, які пов'язують ІНМ з навколишнім середовищем. Єдиний прихований шар, створений нейронами, робить нелінійне перетворення вхідного простору R^n розкритого простору R^h , як правило, більш високої розмірності ($h \gg n$).

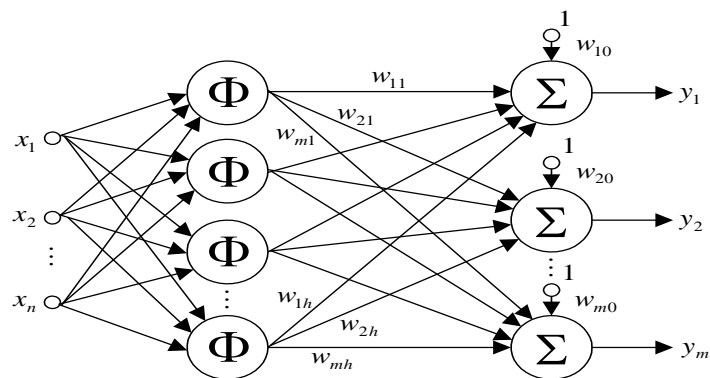


Рисунок 1.1 – Радіально-базисна нейронна мережа

І, нарешті, вихідний шар, утворений адаптивними лінійними асоціатором, формує реакцію мережі $y = (y_1, y_2, \dots, y_m)^T$ на вхідний сигнал $x = (x_1, x_2, \dots, x_n)^T$.

Альтернативами мережам, заснованим на оптимізації, являються мережі засновані на пам'яті, найяскравішим представником яких являється узагальнена регресійна нейронна мережа (УРНС), запропонована Д. Шпехтом [38]. Ці мережі можуть бути віднесені до

миттєво навчених моделей [39], налаштованим завдяки одному проходу алгоритму навчання. Збігаючись з архітектурою з нормалізованими радіально-базисними нейронними мережами (НРБНС), УРНМ вчиться на багато швидше, встановлюючи центри РБФ в точках з координатами, визначеними вхідними сигналами об'єкта за принципом «нейрони в точках даних» [40], і з висотою РБФ, що збігається з відповідними значеннями вихідного сигналу об'єкта. Саме висока швидкість навчання ОРНС забезпечила їх ефективне використання в задачах реального часу [7, 10].

1.3.1 Самоорганізовані карти Кохонена

В основі карти Кохонена лежить нейробіологічна модель Д. Віллшоу - К. фон дер Мальсбурга. Здатність до самоорганізації мережі пов'язані з тим, що налаштування синаптичних ваг відбувається без наявності зовнішнього навчального сигналу, тобто в режимі самонавчання, при цьому кожен наступний сигнал викликає адаптацію тих чи інших параметрів ШНМ. Даний процес може бути безперервним, забезпечуючи можливість вирішення завдань послідовно.

Самоорганізована карта має просту архітектуру з прямою передачею інформації, а також крім нульового рецепторного шару містить єдиний шар нейронів, який іноді називають шаром Кохонена [41]. Кожен нейрон шару Кохонена пов'язаний з кожним рецептором нульового шару прямими зв'язками і з усіма іншими нейронами поперечними всерединішаровими (латеральний) зв'язками. Саме латеральні зв'язки забезпечують збудження одних нейронів і гальмування інших.

Завдяки такій організації мережі, кожен нейрон отримує всю інформацію про аналізований образ і генерує на своєму виході відповідну реакцію після чого в шарі Кохонена виникає режим конкуренції, в результаті якого визначається єдиний нейрон-переможець з максимальним вихідним сигналом. Цей сигнал забезпечує збудження найближчих «сусідів» переможця і придушення реакції вузлів, що далеко знаходяться.

Таким чином, в процесі конкурентного самонавчання формуються групи нейронів, кожен з яких максимальним відгуком реагує на образи з відповідних підобластей вхідного простору, що дозволяє карті Кохонена успішно справлятися з компресією великих обсягів інформації. Самоорганізовані карти можуть мати різну топологію, як це показано на рис. 1.2.

Найпростіша карта Кохонена, наведена на рис. 1.2, має 1D топологію, n рецепторів і $m(m < n)$ нейронів в шарі Кохонена, кожен з яких характеризується власним вектором синаптичних ваг w_q , $q = 1, 2, \dots, m$.

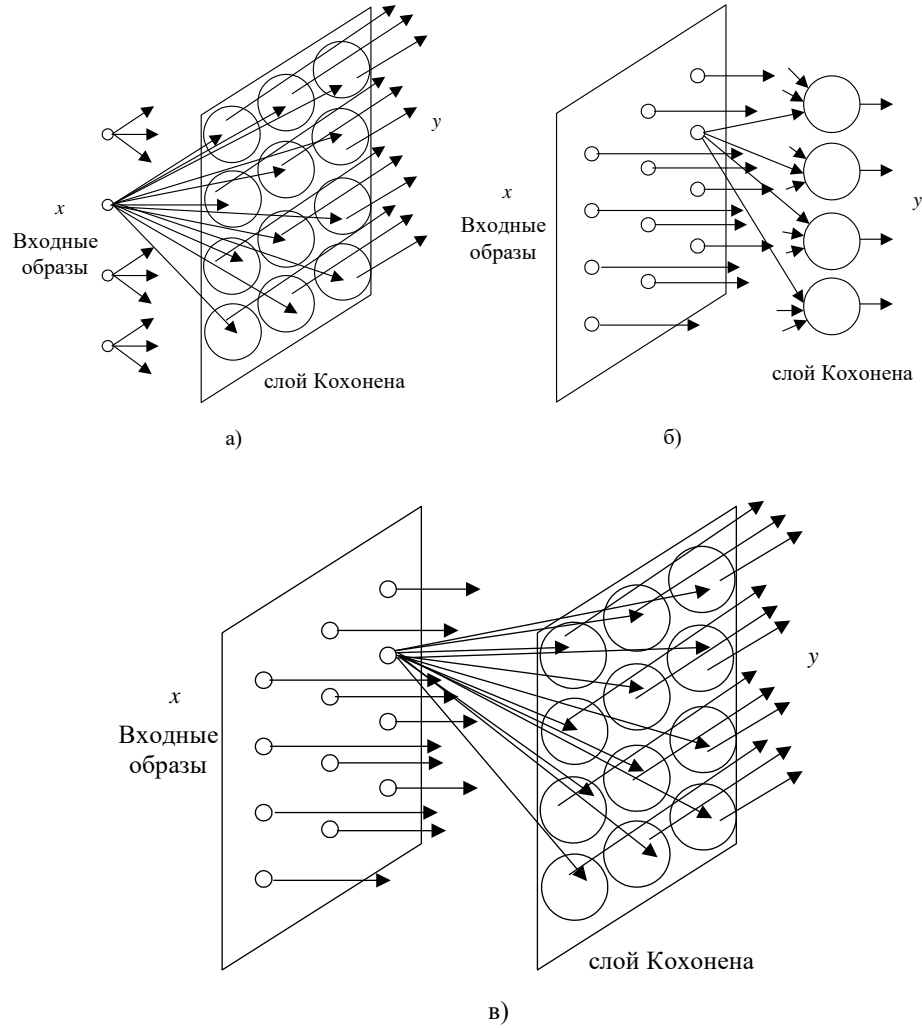


Рисунок 1.2 – Топологія карт Кохонена

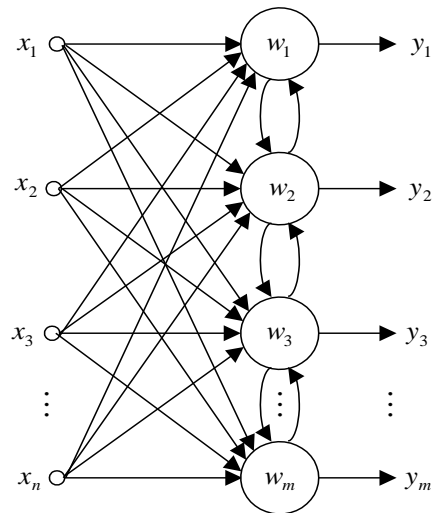


Рисунок 1.3 – 1D-карта Кохонена

В основі алгоритму самоорганізації карти Кохонена було покладено принципи конкурентного самонавчання. Як і будь-яка інша процедура навчання, робота алгоритму починається з ініціалізації синаптичних ваг мережі, які зазвичай обираються за допомогою застосування генератора випадкових чисел, при цьому бажано, щоб для кожного з нейронів виконувалася умова $\|w_q(0)\| = 1$.

Процедура самоорганізації реалізується в три основних етапи: конкуренції, кооперації і синаптичної адаптації. Самоорганізація починається з аналізу образу, що надходить з рецепторного шару на всі нейрони прошарку Кохонена. Для кожного з нейронів обчислюється відстань

$$D(\tilde{x}_k, w_q(k)) = \|\tilde{x}_k - w_q(k)\|, \quad (1.1)$$

причому, входи слід попередньо пронумерувати так, щоб $\|\tilde{x}_k\| = 1$. Як відстань (1.1) використовується евклідова метрика, а мірою близькості між векторами \tilde{x}_k і $w_q(k)$ може служити скалярний добуток

$$D(\tilde{x}_k, w_q(k)) = \tilde{x}_k^T w_q(k) = \cos(\tilde{x}_k, w_q(k)) = \cos \theta_q. \quad (1.2)$$

Далі визначається нейрон-переможець, «найближчий» до вхідного образу такий, що

$$D(\tilde{x}_k, w^*(k)) = \min_q D(\tilde{x}_k, w_q(k)), \quad (1.3)$$

після чого в найпростішому випадку, «перестрибуючи» через етап кооперації, можна підлаштувати синаптичні ваги мережі

$$w_q(k+1) = \begin{cases} w_q(k) + \eta(k)(\tilde{x}_k - w_q(k)), & \text{якщо } q \text{ – тій нейрон виграв,} \\ w_q(k) & \text{в іншому випадку.} \end{cases} \quad (1.4)$$

Процедура (1.4) реалізує принцип «переможець отримує все», при цьому вектор синаптичних ваг нейрона-переможця «підтягується» до вхідного вектору на відстань, що дорівнює кроку пошуку.

Величина кроку має монотонно зменшуватися в процесі навчання, наприклад за правилом

$$\eta(k) = r^{-1}(k), r(k) = \alpha r(k-1) + \|\tilde{x}_k\|^2 = \alpha r(k-1) + 1, 0 < \alpha \leq 1, \quad (1.5)$$

При цьому при $\alpha = 1$ параметр кроку $\eta(k) = \frac{1}{k}$, тобто задовольняє умовам Дворецького. Варіюючи фактором забування α , нескладно забезпечити достатньо широкий інтервал зміни кроку пошуку

$$\frac{1}{k} \leq \eta(k) \leq 1. \quad (1.6)$$

Однією з особливостей карти Кохонена є наявність етапу кооперації в процесі самоорганізації, коли нейрон-переможець визначає так звану локальну область топологічного сусідства, в якій псується не тільки він сам, але і його найближче оточення, при цьому ближчі до переможця нейрони збуджуються сильніше, ніж віддалені. Ця топологічна область визначається функцією сусідства $\phi(q, l)$, яка залежить від відстані $D(w_q(k), w_l(k))$ між переможцем $w_q^*(k)$ і будь-яким з нейронів шару Кохонена $l = 1, 2, \dots, m$ і деякого параметра, який задає її «ширину».

Використання функцій сусідства призводить до модифікованого правила навчання Кохонена

$$w_l(k+1) = w_l(k) + \eta(k)\phi(q, l, k)(\tilde{x}_k - w_l(k)), l = 1, 2, \dots, m, \quad (1.7)$$

реалізує принцип «Переможець отримує більше» (Winner takes most) замість традиційного «Переможець отримує все».

Аналіз збіжності процесів конкурентного самонавчання, показав, що в процесі налаштування синаптичних ваг, повинен ставати менше не тільки крок пошуку, але і параметр ширини функції сусідства, яка таким чином стає залежною від моменту у часі.

Для гауссовської функції

$$\phi(q, l) = \exp\left(-\frac{\|w_q^*(k) - w_l(k)\|^2}{\sigma^2}\right). \quad (1.8)$$

Г. Ріттер і К. Шультен запропонували для налаштування параметра ширини застосовувати процедуру

$$\sigma(k) = \sigma(0) \exp\left(-\frac{k}{\beta}\right), \quad (1.9)$$

де $\beta > 0$ - скалярний параметр, що визначає швидкість зменшення сили впливу нейрона переможця на своє оточення. Звичайно, що при цьому змінюється і форма області топологічного сусідства, що набуває вид

$$\phi(q, l, k) = \exp\left(-\frac{\|w_q^*(k) - w_l(k)\|^2}{\sigma^2(k)}\right). \quad (1.10)$$

Треба звернути увагу на те, що експоненціальне убавання параметра ширини може бути забезпечено і за допомогою більш простого, ніж (1.9) вираження

$$\sigma(k) = \beta\sigma(k-1), 0 < \beta < 1. \quad (1.11)$$

Для навчання самоорганізованої карти пропонується взагалі не обирати переможця як такого, а в якості найбільш універсальної функції сусідства використовувати вихідний сигнал кожного нейрона $y_l(k)$ так, що

$$\begin{aligned} w_l(k+1) &= w_l(k) + \eta(k)\phi(q, l, k)(\tilde{x}_k - w_l(k)) = \\ &= w_l(k) + \eta(k)y_l(k)(\tilde{x}_k - w_l(k)), l = 1, 2, \dots, m. \end{aligned} \quad (1.12)$$

Якщо в якості нейронів шару Кохонена використовується лінійний асоціатор, то (1.12) можна переписати у формі

$$w_l(k+1) = w_l(k) + \eta(k)w_l^T(k)\tilde{x}_k(\tilde{x}_k - w_l(k)), \quad (1.13)$$

а якщо крім того $\|w_l(k)\| = \|\tilde{x}_k\| = 1$, то і в більш простому вигляді

$$\begin{aligned} w_l(k+1) &= w_l(k) + \eta(k) \cos(\tilde{x}_k, w_l(k))(\tilde{x}_k - w_l(k)) = \\ &= w_l(k) + \eta(k) \cos \theta_l(k)(\tilde{x}_k - w_l(k)). \end{aligned} \quad (1.21)$$

Увесь процес самоорганізації має дві тимчасові фази: початкова фаза упорядкування, в якій відбувається топологічне розбиття вхідного простору, і подальша фаза збіжності, в якій здійснюється точне налаштування синаптичних ваг. Після закінчення цього процесу нейронна мережа може вирішувати поставлені задачі без уточнення ваг. У разі появи вхідного образу, який не може бути віднесеним до жодного з сформованих кластерів, картою повинен бути утворений додатковий нейрон в прошарку Кохонена, що несе інформацію про цей образ. Ці ШНМ найбільш ефективні для роботи в реальному часі, коли налаштування синаптичних ваг і обробка вхідних сигналів виконуються паралельно

1.3.2 Нейро-фазі системи

Завдання комплексного забезпечення адекватного уявлення взаємодіючих динамічних нечітких процесів, а також оптимізації ресурсів і вибору альтернатив розвитку нечітких процесів на безлічі обмежень є важливими і на даний момент не мають рішень, які б знайшли ефективне використання більшості практичних реалізацій і часто носять емпіричний, вузькоспеціалізований характер.

Підходи на основі штучних нейронних мережах (ШНМ), нейронних нечітких мережах (ННМ) є універсальним засобом моделювання складних процесів великої розмірності, але в меншій мірі ефективні при моделюванні процесів і процедур малої і середньої розмірності. Вони також потребують додаткових ресурсів на машинне навчання моделям.

Динамічні процеси в реальних системах характеризуються складними багатофункціональними залежностями і істотною нелінійністю. Моделювання таких процесів в задачах прогнозування, діагностики, ідентифікації, класифікації на основі існуючих класичних підходів спричиняє певні труднощі. Теоретико-системний аналіз на

основі подання процесів математичними рівняннями дозволяє створювати досить точні моделі, але досволі трудомісткий і складний в реальних розробках. Експериментальний системний аналіз або ідентифікація базуються на моделях, параметри яких ґрунтуються на даних вимірювань. Перевагами таких моделей є більш короткі терміни розробки, але їх якість істотно залежить від коректного вибору структури та інтерпретації результатів вимірювань.

Особливістю ННМ є те, що вони відносяться до не параметричних моделей, а це викликає складності у визначенні відповідності (інтерпретації) її параметрів в термінах реальних процесів. В задачах ідентифікації на основі апроксимуючих властивостей ШНМ найбільш ефективними представляються моделі на основі багат шарового перцептрона і мереж з радіально-базисними функціями (РБФ). Архітектури моделей на основі мереж Кохонена або Хопфілда і деяких інших загалом використовуються в задачах угруповання і класифікації.

Нечіткі системи як і нейронні мережі мають універсальні апроксимуючі властивості, за допомогою яких можна моделювати довільні функціональні залежності.

Нейронна мережа може вчитися на вхідних і вихідних даних для визначення поведінки системи. Ці знання можуть бути використані для створення нечітких правил і функцій приналежності, що істотно зменшує час, необхідний на розробку. Таке об'єднання також допомагає вирішити проблему неінтерпретованості результатів, отриманих за допомогою нейронних мереж. Подання ваг нейронної мережі за допомогою нечітких правил забезпечує розуміння роботи нейронної мережі, що, у свою чергу, допомагає створювати більш ефективні програми.

Нейро-фаззі системи можуть генерувати нечіткі логічні правила і функції приналежності для складних систем, яким недостатньо стандартного нечіткого підходу. Для таких систем стандартна нечітка логіка використовує складні ієрархічні правила, для зменшення кількості самих правил, що відповідно знижує ефективність і точність рішення.

Дані ШНМ використовують нелінійні функції приналежності. Перевагою таких функцій є рівномірний розподіл знань між базою правил і базою функцій приналежності, що демонструється в скороченні обсягу бази правил. Важливим є той факт, що властивості узагальнення і навченості нейро-фаззі систем дозволяють генерувати правила і функції приналежності, що забезпечують більш достовірне і точне рішення в порівнянні з альтернативними методами. За допомогою коректної комбінації нечіткої логіки і нейронних мереж представляється можливим повне відображення знань нейронної мережі нечіткою логікою, у випадку, якщо алгоритми нечіткої логіки повністю засновані на

архітектурі нейронної мережі. Це дозволяє генерувати нечіткі логічні рішення з заздалегідь заданою точністю виходу.

Алгоритми, що дозволяють на основі даних спостережень будувати в кінцевому підсумку нечітку систему, на початкових етапах використовують або нечітку кластеризацію, яке покриття простору багатовимірними гіперкубами, або оптимізацію багатовимірних ґратчастих структур, що виникають при розбитті координатних осей n -мірного простору на нечіткі підмножини. Основними проблемами в цьому випадку є забезпечення єдності подання одного і того ж лінгвістичного значення і інтерпретується індукованого набору правил приведення одержаних дискретних функцій приналежності до однієї зі стандартних форм, а також уникнути серйозних інформаційних втрат після проектування одержаних в результаті кластеризації областей.

Один з напрямків розвитку систем нечіткого виводу було запропоновано в [42], де розглядаються адаптивні мережеві системи нечіткого виведення (Adaptive-Network-Based Fuzzy Inference Systems, ANFIS), побудовані на базі архітектури Такаґі-Сугено, а також гібридний метод настройки параметрів антецедента (тобто параметрів функції належності). В [43], доведена еквівалентність ANFIS і радіально-базисних нейронних мереж. Головною властивістю ANFIS є те, що апріорно задані правила налаштовуються в процесі навчання для отримання більш точних результатів.

Підходи на основі штучних нейронних мережах (ШНМ), нейронних нечітких мереж (ННМ) є універсальним засобом моделювання складних процесів великої розмірності, але в меншій мірі ефективні при моделюванні процесів і процедур малої і середньої розмірності. Вони також потребують додаткових ресурсів на машинне навчання моделей.

Динамічні процеси в реальних системах характеризуються складними багатофункціональними залежностями і істотною нелінійністю. Моделювання таких процесів в задачах прогнозування, діагностики, ідентифікації, класифікації на основі існуючих класичних підходів викликає певні труднощі. Теоретико – системний аналіз на основі представлення процесів математичними рівняннями дозволяє створювати досить точні моделі, але досить трудомісткий і складний в реальних розробках. Експериментальний системний аналіз або ідентифікація базуються на моделях, параметри яких ґрунтуються на даних вимірювань. Перевагами таких моделей є більш короткі терміни розробки, але їх якість істотно залежить від коректного вибору структури та інтерпретації результатів вимірювань.

Особливістю ННМ є те, що вони відносяться до непараметрическим моделям, а це викликає складності у визначенні відповідності (інтерпретації) її параметрів в термінах

реальних процесів. В задачах ідентифікації на основі апроксимуючих властивостей ШНМ найбільш ефективними представляються моделі на основі багат шарового перцептрона і мереж з радіально - базисними функціями (РБФ). Архітектури моделей на основі мереж Кохонена або Хопфілда і деяких інших більшою мірою використовуються в задачах угруповання і класифікації.

Нечіткі системи як і нейронні мережі мають універсальні апроксимуючими властивостями, з їх допомогою можна моделювати довільні функціональні залежності.

Нейронна мережа може вчитися на вхідних і вихідних даних для визначення поведінки системи. Ці знання можуть бути використані для створення нечітких правил і функцій приналежності, що істотно зменшує час, необхідний на розробку. Таке об'єднання також допомагає вирішити проблему неінтерпритуємості результатів, одержуваних за допомогою нейронних мереж. Вираз ваг нейронної мережі за допомогою нечітких правил забезпечує розуміння роботи нейронної мережі, що, у свою чергу, допомагає створювати більш ефективні програми.

Нейро-фаззі системи можуть генерувати нечіткі логічні правила і функції приналежності для складних систем, яким не досить стандартного нечіткого підходу. Для таких систем стандартна нечітка логіка використовує складні ієрархічні правила, для зменшення кількості самих правил, що відповідно знижує ефективність і точність рішення.

Дані ШНМ використовують нелінійні функції приналежності. Перевагою таких функцій є рівномірний розподіл знань між базою правил і базою функцій приналежності, що демонструється в скороченні обсягу бази правил. Важливим є той факт, що властивості узагальнення і навченості нейро-фаззі систем дозволяють генерувати правила і функції приналежності, що забезпечують більш достовірне і точне рішення в порівнянні з альтернативними методами. За допомогою коректної комбінації нечіткої логіки і нейронних мереж представляється можливим повне відображення знань нейронної мережі нечіткою логікою, у випадку, якщо алгоритми нечіткої логіки повністю засновані на архітектурі нейронної мережі. Це дозволяє генерувати нечіткі логічні рішення з заздалегідь заданою точністю виходу.

В іншій архітектурі нейро-фаззі систем NEFPROX (Neuro-Fuzzy function approximator) основна увага приділяється отриманню добре інтерпретуються правил. Платою за це є зниження точності апроксимації.

Метою моделі NEFCLASS (NEuro Fuzzy CLASSifier) є отримання нечітких правил з безлічі даних, які можна розділити на різні класи. Нечіткі множини та лінгвістичні правила представляють апроксимацію і визначають результат системи NEFCLASS. Вони

виходять з безлічі вибірок шляхом навчання. Обов'язково повинно виконуватися правило, що для кожного лінгвістичного значення може існувати тільки одне подання нечіткої множини.

1.4 Постановка задачі дослідження

У багатьох задачах інтелектуального аналізу даних, пов'язаних з обробкою емпіричних кількісних спостережень, дані можуть бути спотворені пропусками. Задачі відновлення таких спостережень приділялася достатня увага, при цьому вельми ефективними в даній ситуації опинилися підходи, засновані на математичному апараті обчислювального інтелекту і, перш за все, штучних нейронних мереж, які вирішують задачу відновлення загублених спостережень.

Разом з тим, описані підходи до відновлення даних працездатні лише у випадках, коли вихідні дані задані апріорно, а сама таблиця «об'єкт-властивість» або тимчасової ряд мають фіксовану кількість спостережень, тобто не змінюються в процесі обробки. У той же час існує широкий клас задач, коли дані надходять на обробку послідовно так, як це відбувається при навчанні самоорганізованих карт Кохонена [9] або їх модифікацій [10].

Для роботи з такими даними найбільш ефективними представляються штучні нейронні мережі, нейро-фаззи системи, гібридні системи, відомі своїми універсальними апроксимуючими властивості і здатністю до навчання, під яким зазвичай розуміється можливість налаштування їх параметрів шляхом оптимізації деякого критерію якості (цільової функції, критерію навчання) . У більш широкому сенсі налаштовувати можна не тільки параметри, але і архітектуру системи.

В даний час існує ряд підходів, хоча найбільшого поширення набув, так званий, конструктивний підхід, при якому система обчислювального інтелекту, стартуючи з найпростішої архітектури, поступово нарощує свою складність, одночасно налаштовуючи і свої параметри, до досягнення необхідної якості виконання завдання. У той же час більшість таких систем обробляє інформацію в пакетному режимі, що ускладнює їх використання в випадках, коли дані на обробку надходять послідовно в online режимі.

2 АДАПТИВНІ НЕЙРО-ФАЗЗИ МЕТОДИ ДЛЯ КЛАСТЕРИЗАЦІЇ ВИКРИВЛЕНИХ ДАНИХ

Завдання кластеризації масивів даних, що описуються наборами векторів-образів, досить часто зустрічається в багатьох додатках, пов'язаних з інтелектуальним аналізом даних, при цьому останнім часом особлива увага приділяється, так званій, нечіткій кластеризації [70-71], коли обробляється вектор-образ ознак з різними рівнями ймовірностей, можливостей або приладдя може належати одночасно до кількох класів.

Разом з тим, у багатьох задачах DataMining, включаючи, звичайно, кластеризації, вихідні масиви даних можуть містити пропуски, інформація в яких з тих чи інших причин відсутня. Для роботи в подібних умовах досить ефективним може виявитися використання математичного апарату обчислювального інтелекту [73] і, перш за все, штучних нейронних мереж [44], що вирішують завдання відновлення втрачених спостережень, і модифікацій популярного методу нечітких k -середніх [5], що вирішують завдання кластеризації без такого відновлення.

Крім зазначених, існує ще цілий ряд підходів до обробки даних, що містять пропуски, однак, всі вони працездатні тільки в випадках, коли масив вихідних спостережень заданий заздалегідь в повному обсязі і не змінюється в процесі обробки. У той же час існує широкий клас задач, коли дані надходять на обробку послідовно в режимі реального часу так, як це відбувається при навчанні самоорганізованих карт Кохонена [1] або їх модифікацій [6].

2.1 Формальна постановка задачі кластеризації

Вихідною інформацією для стандартної задачі кластеризації є вибірка даних, сформована з N n -мірних векторів (образів) ознак $X = \{x_1, x_2, \dots, x_N\} \subset R^n$, $x_k \in X$, $k = 1, 2, \dots, N$ - або номер спостереження в таблиці «об'єкт-властивість», або поточний дискретний час в задачах on-line обробки.

Результатом кластеризації є розбиття X на m -пересічних класів з деяким рівнем приналежності $U_q(k)$ -го вектора x_k ознак до q -му кластеру.

Вихідні дані в процесі обробки центруються і стандартизовані за всіма компонентами так, щоб всі спостереження належали гіперкубу $[-1,1]^n$. Таким чином вихідні дані набувають вигляду $\tilde{X} = \{\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_N\} \subset R^n$, $\tilde{x}_k = \{\tilde{x}_{k1}, \tilde{x}_{k2}, \dots, \tilde{x}_{kn}\}^T$, $-1 \leq \tilde{x}_{ki} \leq 1$, $1 \leq m \leq N$, $1 < q < m$, $1 \leq i \leq n$, $1 \leq k \leq N$, при цьому N в загальному випадку не обмежена і може збільшуватися з плином часу. Відзначимо, що традиційно прийняте в картах Кохонена (SOM) перетворення даних до виду $\|\tilde{x}_k\| = 1$ в даному випадку не має сенсу, оскільки якщо x_k містить пропуск - обчислення норми такого вектора неможливо, а якщо x_k містить викидень в одній з компонент - \tilde{x}_k буде практично збігатися з відповідним ортом простору ознак. Перетворення $-1 \leq \tilde{x}_{ki} \leq 1$ веде до того, що неспотворені дані концентруються в околиці нуля, а дані з викидами - близько -1 і +1.

2.2 Функції відстані

Вибір відстані між об'єктами є вирішуючим моментом дослідження, від нього багато в чому залежить остаточний варіант розбиття об'єктів на класи при даному алгоритмі розбиття.

2.2.1 Евклідова відстань

Найбільш простий шлях обчислення відстаней між об'єктами в багатовимірному просторі полягає в обчисленні евклідових відстаней. Якщо є дво- або тривимірний простір, то цей захід є реальною геометричною відстанню між об'єктами в просторі.

$$D^2(\tilde{x}_k, w_q) = \|\tilde{x}_k - w_q\|_2^2 \quad (2.1)$$

Зауважимо, що евклідова відстань (і його квадрат) обчислюється по вихідним, а не за стандартизованим даним. Це звичайний спосіб його обчислення, який має певні переваги. Проте, на відстані можуть сильно впливати відмінності між осями, за координатами яких обчислюються ці відстані.

2.2.2 Відстань міських кварталів (Манхеттенська відстань)

Це відстань є просто сумою модулів різниць по координатах. У більшості випадків ця міра відстані призводить до таких же результатів, як і для звичайна відстань Евкліда. Однак зазначимо, що для цього заходу вплив окремих великих різниць (викидів) зменшується (так як вони не зводяться в квадрат). Манхеттенська відстань обчислюється за формулою:

$$D(\tilde{x}_k, w_q) = \|\tilde{x}_k - w_q\|_1 = \sum_{i=1}^n |\tilde{x}_{ki} - w_{qi}|. \quad (2.2)$$

2.2.3 Відстань Чебишева

Це відстань може виявитися корисною при визначенні двох об'єктів як "різні", якщо вони розрізняються за якоюсь однією координатаю (будь-яким одним виміром). Відстань Чебишева обчислюється за формулою:

$$D(\tilde{x}_k, w_q) = \max |\tilde{x}_{ki} - w_{qi}| \quad (2.3)$$

2.2.4 Часткова відстань

Дана міра відстані обчислюється за формулою (2.4)

$$D_P^2(\tilde{x}_k, w_q) = \frac{n}{\delta_{k\Sigma}} \sum_{i=1}^n (\tilde{x}_{ki} - w_{qi})^2 \delta_{ki}, \quad (2.4)$$

де w_{qi} - а компонента q -го прототипу (центроїда) відповідного кластера ($q = 1, 2, \dots, m$),

$$\delta_{ki} = \begin{cases} 0 & | \tilde{x}_{ki} \in X_G, \\ 1 & | \tilde{x}_{ki} \in X_F, \end{cases}$$

$$\delta_{k\Sigma} = \sum_{i=1}^n \delta_{ki}.$$

Часткова відстань (PD) дозволяє врахувати пропущені значення, що дозволяє кластеризувати зіпсовані дані. Жоден з наведених вище методів розрахунку відстаней не надає такої можливості.

2.2.5 Методи кластеризації даних

При великій кількості спостережень ієрархічні методи кластерного аналізу не придатні. У таких випадках використовують неієрархічні методи, засновані на поділі, які представляють собою ітеративні методи подрібнення вихідної сукупності. У процесі поділу нові кластери формуються до тих пір, поки не буде виконано правило зупинки.

Така неієрархічна кластеризація полягає в поділі набору даних на певну кількість окремих кластерів. Існує два підходи. Перший полягає у визначенні меж кластерів як найбільш щільних ділянок в багатовимірному просторі вихідних даних, тобто визначення кластера там, де є велике "зкупчення точок". Другий підхід полягає в мінімізації заходів відмінностей об'єктів.

2.2.6 Адаптивний метод ймовірнісної нечіткої кластеризації даних

Вводячи цільову функцію кластеризації

$$E(U_q(k), w_q) = \sum_{k=1}^N \sum_{q=1}^m U_q^\beta(k) D^2(\tilde{x}_k, w_q)$$

при обмеженнях $\sum_{q=1}^m U_q(k) = 1, 0 < \sum_{k=1}^N U_q(k) < N$ і вирішуючи завдання нелінійного програмування, приходимо до алгоритму ймовірнісної нечіткої кластеризації [73,97]

$$\begin{cases} U_q^{(\tau+1)}(k) = \frac{(D^2(\tilde{x}_k, w_q^{(\tau)}))^{\frac{1}{1-\beta}}}{\sum_{l=1}^m (D^2(\tilde{x}_k, w_l^{(\tau)}))^{\frac{1}{1-\beta}}}, \\ w_q^{(\tau+1)} = \frac{\sum_{k=1}^N (U_q^{(\tau+1)})^\beta \tilde{x}_k}{\sum_{k=1}^N (U_q^{(\tau+1)}(k))^\beta}, \end{cases} \quad (2.5)$$

де w_q - прототип (центр ваги) -го кластера,

$\beta > 1$ - параметр, іменований фаззифікатором і визначає «розмитість» меж між класами,

$D^2(\tilde{x}_k, w_q)$ - відстань між \tilde{x}_k і w_q в прийнятій метриці,

$\tau = 0, 1, 2, \dots$ - індекс епохи обробки інформації, яка організовується у вигляді послідовності $w_q^{(0)} \rightarrow U_q^{(1)} \rightarrow w_q^{(1)} \rightarrow U_q^{(2)} \rightarrow \dots$

Процес обчислень триває до виконання умови

$$\|w_q^{(\tau+1)} - w_q^{(\tau)}\| \leq \varepsilon \quad \forall 1 \leq q \leq m,$$

де ε - наперед задається поріг точності обчислень.

Вибираючи $\beta = 2$ і приймаючи евклідову відстань $D^2(\tilde{x}_k, w_q) = \|\tilde{x}_k - w_q\|^2$, отримуємо популярний алгоритм нечітких с-середніх (FCM) Дж. Бездека:

$$\begin{cases} U_q^{(\tau+1)}(k) = \frac{\|\tilde{x}_k - w_q^{(\tau)}\|^{-2}}{\sum_{l=1}^m \|\tilde{x}_k - w_l^{(\tau)}\|^{-2}}, \\ w_q^{(\tau+1)} = \frac{\sum_{k=1}^N (U_q^{(\tau+1)}(k))^2 \tilde{x}_k}{\sum_{k=1}^N (U_q^{(\tau+1)}(k))^2}. \end{cases}$$

Процес нечіткої кластеризації може бути організований в on-line режимі по мірі послідовного надходження даних на обробку. При цьому пакетний алгоритм може бути переписаний в рекуррентній формі

$$\begin{cases} U_q(k+1) = \frac{(D^2(\tilde{x}_{k+1}, w_q^{(k)}))^{\frac{1}{1-\beta}}}{\sum_{l=1}^m (D^2(\tilde{x}_{k+1}, w_l^{(k)}))^{\frac{1}{1-\beta}}}, \\ w_q(k+1) = w_q(k) + \eta(k+1) U_q^\beta(k+1) (\tilde{x}_{k+1} - w_q(k)), \end{cases} \quad (2.6)$$

(тут $\eta(k+1)$ - параметр кроку навчання), що є узагальненням градієнтної процедури кластеризації Парка-Деггера [79] і алгоритму навчання Чанга-Лі.

Якщо дані надходять на обробку з високою частотою, перерахунок по епохах не проводиться, якщо ж ця частота невелика, то між моментами часу k і $k+1$ можна організувати кілька епох в прискореному часі, при цьому моменту k відповідає $\tau = 0$, а в $(k+1)$ -й момент цей перерахунок закінчується.

Цікаво також зауважити, що перший вираз може бути переписано в формі

$$\begin{aligned} U_q(k+1) &= \frac{(D^2(\tilde{x}_k, w_q(k)))^{\frac{1}{1-\beta}}}{\sum_{l=1}^m (D^2(\tilde{x}_k, w_l(k)))^{\frac{1}{1-\beta}}} = \\ &= \frac{(D^2(\tilde{x}_k, w_q(k)))^{\frac{1}{1-\beta}}}{(D^2(\tilde{x}_k, w_q(k)))^{\frac{1}{1-\beta}} + \sum_{\substack{l=1 \\ l \neq q}}^m (D^2(\tilde{x}_k, w_l(k)))^{\frac{1}{1-\beta}}} = \\ &= \frac{1}{1 + (D^2(\tilde{x}_k, w_q(k)))^{\frac{1}{\beta-1}} \sum_{\substack{l=1 \\ l \neq q}}^m (D^2(\tilde{x}_k, w_l(k)))^{\frac{1}{1-\beta}}}, \end{aligned}$$

для евклідової метрики і $\beta = 2$ набуває вигляду функції Коші з параметром ширини σ^2 :

$$U_q(k+1) = \frac{1}{1 + \frac{\|\tilde{x}_k - w_q(k)\|^2}{\sigma^2}},$$

$$\sigma^2 = \left(\sum_{\substack{l=1 \\ l \neq q}}^m \|\tilde{x}_k - w_l(k)\|^{-2} \right)^{-1}.$$

Дана обставина дозволяє переписати другий вираз (2.6) при $\beta = 2$ у формі

$$\begin{aligned} w_q(k+1) &= w_q(k) + \eta(k+1)U_q^2(k+1)(\tilde{x}_{k+1} - w_q(k)) = \\ &= w_q(k) + \eta(k+1)\varphi_q(k+1)(\tilde{x}_{k+1} - w_q(k)), \end{aligned}$$

де $U_q^2(k+1) = \varphi_q(k+1)$ дзвінокоподібна функція сусідства нейро-фаззі мережі Кохонена, призначеної для вирішення завдань нечіткої кластеризації [80] на основі принципу «переможець отримує більше» (WTM).

3 ПРОГРАМНА РЕАЛІЗАЦІЯ

Для перевірки коректної роботи програми та запропонованого алгоритму було проведено імітаційне моделювання з використанням UCI репозиторію Iris і Wine. Імітаційне моделювання проводилося в середовищі MATLAB.

3.1 MATLAB

MATLAB - це програмна платформа, розроблена спеціально для інженерів і вчених. Серцем MATLAB є мова MATLAB, матричний мову, що дозволяє найбільш природне вираження обчислювальної математики.

MATLAB поєднує в собі настільну середу, налаштовану для ітеративного аналізу і процесів проектування з мовою програмування, який безпосередньо виражає математику матриць і масивів.

Використовуючи MATLAB, ви можете:

- 1) Аналізувати дані;
- 2) Розроблений алгоритм;
- 3) Створення моделей і додатків.

Мова, Додатки і вбудовані математичні функції дозволяють швидко вивчити кілька підходів, щоб прийти до вирішення. MATLAB дозволяє вам переходити від досліджень до виробництва, розгортаючи корпоративні додатки і вбудовані пристрої, а також інтегруючись з моделями і додатками Simulink і Model-Based Design.

Мова, інструментарій та вбудовані математичні функції якого дозволяють вам досліджувати різні підходи і отримувати рішення швидше, ніж з використанням електронних таблиць або традиційних мов програмування, таких як C / C ++ або Java.

MATLAB широко використовується в таких областях, як:

- 1) обробка сигналів і зв'язок
- 2) обробка зображень і відео
- 3) системи управління
- 4) автоматизація тестування і вимірювань
- 5) фінансовий інжиніринг
- 6) обчислювальна біологія і т.п.

Понад мільйон інженерів і вчених по всьому світу використовують MATLAB в якості мови технічних обчислень.

MATLAB в порівнянні з традиційними мовами програмування (C / C ++, Java, Pascal, FORTRAN) дозволяє на порядок скоротити час вирішення типових завдань і значно спрощує розробку нових алгоритмів. MATLAB являє собою основу всього сімейства продуктів MathWorks і є головним інструментом для вирішення широкого спектра наукових і прикладних задач, в таких областях як: моделювання об'єктів і розробка систем управління, проектування комунікаційних систем, обробка сигналів і зображень, вимірювання сигналів і тестування, фінансове моделювання, обчислювальна біологія та ін.

Ядро MATLAB дозволяє максимально просто працювати з матрицями реальних, комплексних і аналітичних типів даних і зі структурами даних і таблицями пошуку. MATLAB містять задану вбудовані функції лінійної алгебри (LAPACK, BLAS), швидкого перетворення Фур'є (FFTW), функції для роботи з поліномами, функції базової статистики та чисельного рішення диференціальних рівнянь; розширені математичні бібліотеки для Intel MKL. Всі вбудовані функції ядра MATLAB розроблені й оптимізовані фахівцями і працюють швидше або так само, як їх еквівалент на C / C ++.

MATLAB надає безліч методів для аналізу даних, розробки алгоритмів і створення моделей. Мова MATLAB включає в себе математичні функції для інженерних і наукових операцій. Вбудовані математичні функції використовують процесор-оптимізовані бібліотеки, призначені для прискорення векторних і матричних обчислень. Доступні наступні операції:

- 1) Інтерполяція і регресія
- 2) Диференціювання та інтегрування
- 3) Системи лінійних рівнянь
- 4) Фур'є аналіз
- 5) Власні значення і сингулярні числа матриць
- 6) Звичайні диференціальні рівняння
- 7) Розріжені матриці

Розширення MATLAB надають спеціалізований функціонал в таких областях як статистика, оптимізація, обробка сигналів, машинне навчання.

MATLAB надає інструменти для отримання, аналізу та візуалізації даних, що дозволяють досліджувати проблему швидше, ніж це можливо за допомогою електронних таблиць або традиційних мов програмування. Також ви можете документувати результати у вигляді графіків, звітів або публікації коду MATLAB.

MATLAB дозволяє вам отримувати доступ до даних з файлів, інших додатків, баз даних, зовнішніх пристроїв. Ви можете читати дані з файлів таких популярних форматів як Microsoft Excel, текстових або довічних файлів, зображень, аудіо та відео файлів, наукових форматів (netCDF і HDF). Функції вводу-виводу дозволяють працювати з файлами даних будь-яких форматів. Використовуючи розширення MATLAB можна отримувати дані з різних пристроїв, таких як послідовний порт комп'ютера або звукова карта, а також потокові дані в реальному часі з вимірювальних пристроїв безпосередньо в MATLAB для аналізу і візуалізації. Крім того, ви можете управляти такими приладами, як осцилографи, аналізатори сигналів і генератори коливань спеціальної форми.

MATLAB дозволяє управляти, фільтрувати і здійснювати попередню обробку даних. Ви можете досліджувати дані для знаходження трендів, перевірки гіпотез, побудови описових моделей. В MATLAB включені функції для фільтрації, згладжування, згортки і швидкого перетворення Фур'є (FFT). Продукти-розширення включають можливості підбору кривих і поверхонь, багатовимірної статистики, спектрального аналізу, аналізу зображень, ідентифікації систем та інші інструменти аналізу.

3.2 DATA BASE (Виборка)

3.2.1 Iris

Iris – об'єктно-орієнтована система управління базами даних, що розробляється в лабораторіях Hewlett-Packard. Iris призначена для задоволення потреб нових і нових додатків баз даних, таких як офісні та інженерні інформаційні системи, системи, засновані на знаннях, виробничі програми та проектування апаратного та програмного забезпечення. Ці програми потребують багатого набору можливостей, які не підтримуються поточним поколінням (тобто реляційними) СУБД. Модель даних Iris-це модель об'єкта та функції. Він надає три основні конструкції об'єктів, типів і функцій. Як і в інших об'єктних системах, об'єкти Iris мають унікальний ідентифікатор й можуть бути доступні і управлятися тільки за допомогою функцій. Об'єкти класифікуються за типом. Об'єкти, що належать до одного типу, спільно використовують загальні функції. Типи організовані в ієрархію з успадкованими функціями.

3.2.2 Wine

Інформація про набір даних:

Ці дані є результатами хімічного аналізу вин, вирощених в одному регіоні Італії, але отриманих з трьох різних сортів. Аналіз визначається кількістю 13 виборців у кожному з трьох типів вин.

Атрибути :

- 1) Alcohol;
- 2) Malic acid;
- 3) Ash;
- 4) Alcalinity of ash;
- 5) Magnesium;
- 6) Total phenols;
- 7) Flavanoids;
- 8) Nonflavanoid phenols;
- 9) Proanthocyanins;
- 10) Color intensity;
- 11) Hue;
- 12) OD280/OD315 of diluted wines;
- 13) Proline.

У контексті класифікації це добре поставлена проблема з "добре себе ведуть" структурами класів. Гарний набір даних для першого тестування нового класифікатора, але не дуже складний.

Всі атрибути є безперервними.

Статистики немає, але пропонується стандартизувати змінні для певних цілей (наприклад, для нас з класифікаторами, які не є масштабно-інваріантними)

Примітка: 1-й атрибут-ідентифікатор класу (1-3)

3.3 Експериментальні дослідження

В експериментальних дослідженнях порівнювалися з FCM, за основними параметрами кластеризації: коефіцієнту розбиття (PC), класифікаційної ентропії (CE), індексом розділів (SC), індексу розподілу (S), індексу CE і Бені (XB), індексу даних (DI).

Коефіцієнт розбиття (PC): вимірює кількість "перекриття" між кластерами.

Класифікаційна ентропія (CE): вона вимірює нечіткість тільки розділу кластера, що схоже на коефіцієнт розподілу.

Індекс розділів (SC): це відношення суми компактності та розділеності кластерів. Це сума окремих дій щодо дій кластеру, нормалізованих за допомогою поділу нечіткою кардинальністю кожного кластера. SC корисний при порівнянні різних розділів, що мають однакову кількість кластерів. Менше значення SC вказує на кращий розділ.

Індекс поділу (S): навпаки індексу розділів (SC), поділ in-dex використовує поділ мінімальної відстані для дійсності розділу.

Індекс Сі і Бені (XB): він спрямований на кількісне визначення співвідношення загальної зміни в кластерах та поділу кластерів. Оптимальна кількість кластерів повинна мінімізувати значення індексу.

Індекс даних (DI): цей індекс спочатку пропонується використовувати для ідентифікації "компактних і добре відокремлених кластерів". Таким чином, результат кластеризації потрібно переглядати, оскільки це був алгоритм жорсткого розподілу.

Результати експериментів представлені в Таблиці 1 та Таблиці 2. Результати запропонованого нами алгоритму ми також порівняли з іншими більш відомими, такими як алгоритм кластеризації нечітких С-засобів (FCM).

Порівняльний результат приведено в Таблицях 3.1 та 3.2.

Таблиця 3.1 – Результати експериментальних досліджень, порівняльний аналіз кластеризації даних вибірки Iris UCI repository

Алгоритми	Iris UCI repository					
	PC	CE	SC	S	XB	DI
Адаптивний ймовірнісний нечіткий метод кластеризації даних	0,0199	0,0122	-0,2439	0,0022	0,0015	1
FCM	0,8011	0,3410	0,2567	0,0030	7,1965	0,0080

Таблиця 3.2 – Результати експериментальних досліджень, порівняльний аналіз кластеризації даних вибірки Wine UCI repository

Алгоритми	Wine UCI repository					
	PC	CE	SC	S	XB	DI
Адаптивний ймовірнісний нечіткий метод кластеризації даних	0,0230	0,0219	-0,3007	0,0032	0,0065	0,9999
FCM	0,7411	0,2389	0,3112	0,0040	6,9945	0,0078

	1	2	3	4	5	6	7	8	9	10	11	12
1	1	14.2300	1.7100	2.4300	15.6000	127	2.8000	3.0600	0.2800	2.2900	5.6400	1.0400
2	1	13.2000	1.7800	2.1400	11.2000	100	2.6500	2.7600	0.2600	1.2800	4.3800	1.0500
3	1	13.1600	2.3600	2.6700	18.6000	101	2.8000	3.2400	0.3000	2.8100	5.6800	1.0300
4	1	14.3700	1.9500	2.5000	16.8000	113	3.8500	3.4900	0.2400	2.1800	7.8000	0.8600
5	1	13.2400	2.5900	2.8700	21	118	2.8000	2.6900	0.3900	1.8200	4.3200	1.0400
6	1	14.2000	1.7600	2.4500	15.2000	112	3.2700	3.3900	0.3400	1.9700	6.7500	1.0500
7	1	14.3900	1.8700	2.4500	14.6000	96	2.5000	2.5200	0.3000	1.9800	5.2500	1.0200
8	1	14.0600	2.1500	2.6100	17.6000	121	2.6000	2.5100	0.3100	1.2500	5.0500	1.0600
9	1	14.8300	1.6400	2.1700	14	97	2.8000	2.9800	0.2900	1.9800	5.2000	1.0800
10	1	13.8600	1.3500	2.2700	16	98	2.9800	3.1500	0.2200	1.8500	7.2200	1.0100
11	1	14.1000	2.1600	2.3000	18	105	2.9500	3.3200	0.2200	2.3800	5.7500	1.2500
12	1	14.1200	1.4800	2.3200	16.8000	95	2.2000	2.4300	0.2600	1.5700	5	1.1700
13	1	13.7500	1.7300	2.4100	16	89	2.6000	2.7600	0.2900	1.8100	5.6000	1.1500
14	1	14.7500	1.7300	2.3900	11.4000	91	3.1000	3.6900	0.4300	2.8100	5.4000	1.2500
15	1	14.3800	1.8700	2.3800	12	102	3.3000	3.6400	0.2900	2.9600	7.5000	1.2000
16	1	13.6300	1.8100	2.7000	17.2000	112	2.8500	2.9100	0.3000	1.4600	7.3000	1.2800
17	1	14.3000	1.9200	2.7200	20	120	2.8000	3.1400	0.3300	1.9700	6.2000	1.0700
18	1	13.8300	1.5700	2.6200	20	115	2.9500	3.4000	0.4000	1.7200	6.6000	1.1300
19	1	14.1900	1.5900	2.4800	16.5000	108	3.3000	3.9300	0.3200	1.8600	8.7000	1.2300
20	1	13.6400	3.1000	2.5600	15.2000	116	2.7000	3.0300	0.1700	1.6600	5.1000	0.9600
21	1	14.0600	1.6300	2.2800	16	126	3	3.1700	0.2400	2.1000	5.6500	1.0900
22	1	12.9300	3.8000	2.6500	18.6000	102	2.4100	2.4100	0.2500	1.9800	4.5000	1.0300

Рисунок 3.1 Вихідні дані

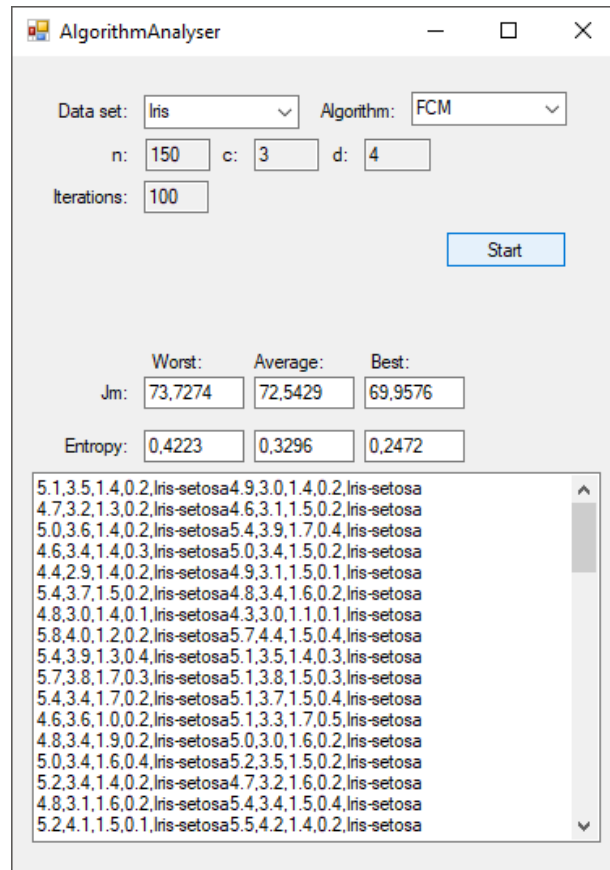


Рисунок 3.2 - Результати роботи програми

4 ОХОРОНА ПРАЦІ

4.1 Аналіз потенційних небезпечних і шкідливих виробничих чинників проєктованого об'єкту, що мають вплив на персонал

У даному дипломному проєкті розробляється програмне забезпечення.

Розроблене програмне забезпечення орієнтоване на роботу з персональним комп'ютером. Експлуатовані для вирішення внутрішньовиробничих завдань ПЕОМ типу IBM PC мають наступні характеристики:

споживана потужність	220 Вт;
робоча напруга	220 В;
напруга джерел живлення	+12 В; - 12 В; +5 В;
робоча частота	50 Гц.

Виходячи з приведених характеристик, вочевидь, що для людини існує небезпека поразки електричним струмом, унаслідок недбалого поводження з комп'ютером і порушення правил експлуатації, залишення частин ПЕОМ, що знаходяться під напругою, відкритими або знятих для ремонту вузлів.

Відповідно до ДСН 3.3.6.042-99 [101] до легкої фізичної роботи відносяться всі види діяльності, виконувані сидячи і ті, що не потребують фізичної напруги. Робота користувача ПК відноситься до категорії 1а.

При роботі на ПЕОМ користувач піддається ряду потенційних небезпек. Унаслідок недотримання правил техніки безпеки при роботі з машиною (невиконання огляду відкритих частин ПЕОМ, що знаходяться під напругою або знятих для ремонту вузлів) для користувача існує небезпека поразки електричним струмом.

Джерелами підвищеної небезпеки можуть служити наступні елементи:

- розподільний щит;
- джерела живлення;
- блоки ПЕОМ і друку, що знаходяться в ремонті.

Ще одна проблема полягає у тому, що спектр випромінювання комп'ютерного монітора включає рентгенівську, ультрафіолетову і інфрачервону області, а також широкий діапазон хвиль інших частот. Небезпека рентгенівського проміння мала, оскільки цей вид випромінювання поглинається речовиною екрану. Проте велику увагу

слід приділяти біологічним ефектам низькочастотних електромагнітних полів (аж до порушення ДНК).

Відповідно до НПАОП 0.00-7.15-18 [102], при обслуговуванні ПЕОМ мають місце фізичні і психофізичні небезпечні, а також шкідливі виробничі чинники:

- підвищене значення напруги в електричному ланцюзі, замикання якої може відбутися через тіло людини;
- підвищений рівень статичної електрики;
- підвищений рівень електромагнітних випромінювань;
- підвищена або знижена температура повітря робочої зони;
- підвищений або знижений рух повітря;
- підвищена або знижена вологість повітря;
- відсутність або недостатність природного світла;
- підвищена пульсація світлового потоку;
- недостатня освітленість робочого місця;
- підвищений рівень шуму на робочому місці;
- розумове перенапруження;
- емоційні навантаження;
- монотонність праці.

4.2 Заходи щодо техніки безпеки

Основним небезпечним чинником при роботі з ЕОМ є небезпека поразки людини електричним струмом, яка посилюється тим, що органи чуття людини не можуть на відстані знайти наявності електричної напруги на устаткуванні.

Проходячи через тіло людини, електричний струм чинить на нього складну дію, що є сукупністю термічної (нагрів тканин і біологічних середовищ), електролітичної (розкладання крові і плазми) і біологічної (роздратування і збудження нервових волокон і інших органів тканин організму) дій.

Тяжкість поразки людини електричним струмом залежить від цілого ряду чинників:

- значення сили струму;
- електричного опору тіла людини і тривалості протікання через нього струму;

- роду і частоти струму;
- індивідуальних властивостей людини і навколишнього середовища.

Розроблений дипломний проект передбачає наступні технічні способи і засоби, що застерігають людину від ураження електричним струмом:

- заземлення електроустановок;
- занулення;
- захисне відключення;
- електричне розділення мережі;
- використання малої напруги;
- ізоляція частин, що проводять струм;
- огорожа електроустановок.

Занулення зменшує напругу дотику і обмежує години, протягом яких людина, ткнувшись до корпусу, може потрапити під дію напруги.

Струм однофазного короткого замикання визначається по наближеній формулі:

$$I_k = \frac{U_f}{Z_{\Pi} + \frac{Z_T}{3}}, \quad (4.1)$$

де U_f - номінальна фазна напруга мережі, В;

Z_{Π} - повний опір петлі, створене фазними і нульовими дротами, Ом;

Z_T - повний опір струму короткого замикання на корпус, Ом.

Згідно таблиці 4 [103]: $Z_T / 3 = 0,1$ Ом.

Для провідників і жил кабелю для розрахунку повного опору петлі використовуємо формулу (4.2.) :

$$Z_{\Pi} = \sqrt{R_{\Pi}^2 + X_{\Pi}^2}, \quad (4.2)$$

де $R_{\Pi} = R_f + R_o$ - сумарний активний опір фазного R_f і нульового R_o дротів, Ом;

X_{Π} - індуктивний опір паяння дротів, Ом.

Перетин 1 км мідного дроту $S = 2.5$ мм, тоді згідно таблицям 5 і 6 [103], має такий опір:

$$X_{\Pi} = 0,11 \text{ Ом};$$

$$R_{\Phi} = 7,55 \text{ Ом};$$

$$R_0 = 7,55 \text{ Ом}.$$

Отже, $R_{\Pi} = 7,55 + 7,55 = 15,1 \text{ Ом}.$

Тоді по формулі (4.2) знаходимо повний опір петлі :

$$Z_{\Pi} = \sqrt{15,1^2 + 0,11^2} \approx 15,1 \text{ (Ом)}.$$

Струм однофазного короткого замикання рівний:

$$I_k = \frac{220}{15,1 + 0,1} = 14,47 \text{ (А)}.$$

Дія плавкої вставки на ПЕОМ забезпечується, якщо виконується співвідношення:

$$I_k \geq k * I_n, \quad (4.3)$$

де I_n - номінальний струм спрацьовування плавкої вставки, А;

k - коефіцієнт кратності нелінійного струму I_n , А.

Коефіцієнт кратності нелінійного струму I_n розраховується по формулі (4.4.) :

$$I_n = P / U, \quad (4.4)$$

де $P = 220 \text{ Вт}$ - споживана потужність;

$U = 220 \text{ В}$ - робоча напруга;

$k = 3 \text{ А}$ - для плавких вставок.

Отже, $I_n = 220 / 220 = 1 \text{ А}.$

Підставивши значення у вираз (4.3), одержимо:

$$14,47 > 3 * 1.$$

Таким чином, доведено, що апарат забезпечить спрацьовування(і захист) при підвищенні номінального струму.

4.3 Заходи, що забезпечують виробничу санітарію і гігієну праці

Вимоги до виробничих приміщень встановлюються ДСН 3.3.6.042-99[101], ДБН, відповідними ГОСТами і ОСТами з урахуванням небезпечних і шкідливих чинників, що утворюються в процесі експлуатації електроустаткування.

Підвищення працездатності людини і збереження її здоров'я забезпечується стабільними метеорологічними умовами.

Мікроклімат виробничих приміщень визначається діючими на організм людини поєднаннями температури, вологості і швидкості руху повітря, а також температури навколишніх поверхонь. Значне коливання параметрів мікроклімату приводить до порушення систем кровообігу, нервової і потовидільної, що може викликати підвищення або пониження температури тіла, слабкість, запаморочення і навіть непритомність.

Відповідно до ДСН 3.3.6.042-99 [101] встановлюють оптимальну і допустиму температуру, відносну вологість і швидкість руху повітря в робочій зоні . За відсутності надмірного тепла, вологи, шкідливих речовин в приміщенні досить природної вентиляції.

У приміщенні для виконання робіт операторського типу(категорія 1а), пов'язаних з нервово-емоційною напругою, проектом передбачається дотримання наступних нормованих величин параметрів мікроклімату (табл. 4.1).

Таблиця 4.1 - Санітарні норми мікроклімату робочої зони приміщень для робіт категорії 1а.

Пора року	Температура, С	Відносна вологість, %	Швидкість руху повітря, м/с
Холодна	22...24	40...60	0,1
Тепло	23...25	40...60	0,1

У приміщенні, де знаходиться ПЕОМ, повітрообмін реалізується за допомогою природної організованої вентиляції (з пристроєм вентиляційних каналів в перекриттях будівлі і вертикальних шахт) й встановленого промислового кондиціонера фірми

Mitsubishi, який дозволяє вирішити переважну більшість завдань по створінню та підтримці необхідних параметрів повітряного середовища. Цей метод забезпечує приток потрібної кількості свіжого повітря, визначеного в ДБН (30 м³ в годину на одного працівника).

Шум на виробництві має шкідливу дію на організм людини. Стомлення операторів через шум збільшує число помилок при роботі, призводить до виникнення травм. Для оператора ПЕОМ джерелом шуму є робота принтера. Щоб усунути це джерело шуму, використовують наступні методи. При покупці принтера слід вибирати найбільш шумозахисні матричні принтери або з великою швидкістю роботи(струменеві, лазерні). Рекомендується принтер поміщати в найбільш віддалене місце від персоналу, або застосувати звукоізоляцію та звукопоглинання(під принтер підкладають демпфуючі підкладки з пористих звукопоглинальних матеріалів з листів тонкої повсті, поролону, пеноплєну).

При роботі на ПЕОМ, проектом передбачені наступні методи захисту від електромагнітного випромінювання : обмеження часом, відстанню, властивостями екрану.

Обмеження годині роботи на ПЕОМ складає 3,5-4,5 години. Захист відстанню передбачає розміщення монітора на відстані 0,4-0,5 м від оператора. Передбачений монітор 20" TFT, Samsung 2043BW відповідає вимогам стандарту ТСО'03.

ТСО'03 пред'являє жорсткі вимоги в таких областях: ергономіка (фізична, візуальна і зручність користування), енергія, випромінювання (електричних і магнітних полів), навколишнє середовище і екологія, а також пожежна та електрична безпека, які відповідають всім вимогам [104].

Для зниження стомлюваності та підвищення продуктивності праці обслуговуючого персоналу в колірній композиції інтер'єру приміщень для ПЕОМ дипломним проектом пропонується використовувати спокійні колірні поєднання і покриття, що не дають відблисків.

У проекті передбачається використання сумісного освітлення. У світлий час доби приміщення освітлюватиметься через віконні отвори, в решту часу використовуватиметься штучне освітлення.

Як штучне освітлення необхідно використовувати штучне робоче загальне освітлення. Для загального освітлення необхідно використовувати люмінесцентні лампи. Вони володіють наступними перевагами: високою світловою віддачею, тривалим терміном служби, хоча мають і недоліки: високу пульсацію світлового потоку.

При експлуатації ПЕОМ виробляється зорова робота. Відповідно до ДБН В.2.5-28-2006 [107] ця робота відноситься до розряду 5а. При цьому нормоване освітлення на робочому місці(E_n) при загальному освітленні рівна 200 лк.

Приміщення завдовжки 12 м, шириною 10 м, заввишки 4 м обладнується світильниками типу ЛПО2П, оснащеними лампами типу ЛБ зі світловим потоком 3120 лм кожна.

Виконаємо розрахунок кількості світильників в робочому приміщенні завдовжки $a=12$ м, шириною $b=10$ м, заввишки $z=4$ м, використовуючи формулу (4.5) розрахунку штучного освітлення при горизонтальній робочій поверхні методом світлового потоку:

$$n = (E \cdot S \cdot Z \cdot k) / (F \cdot U \cdot M), \quad (4.5)$$

де F - світловий потік = 3120 лм;

E - максимально допустима освітленість робочих поверхонь = 200 лк;

S - площа підлоги = 120 м²;

Z - поправочний коефіцієнт світильника = 1,2;

k - коефіцієнт запасу, що враховує зниження освітленості в процесі експлуатації світильників = 1,5;

n - кількість світильників;

U - коефіцієнт використання освітлювальної установки = 0,6;

M - кількість ламп у світильнику = 2.

З формули (4.5) виразимо n (4.6) і визначимо кількість світильників для даного приміщення:

$$n = (E \cdot S \cdot Z \cdot k) / (F \cdot U \cdot M), \quad (4.6)$$

$$\text{Отже, } n = (200 \cdot 120 \cdot 1,2 \cdot 1,5) / (3120 \cdot 0,6 \cdot 2) = 12$$

Виходячи з цього, рекомендується використовувати 12 світильників. Світильники слід розміщувати рядами, бажано паралельно стіні з вікнами. Схема розташування світильників зображена на рис. 4.1.

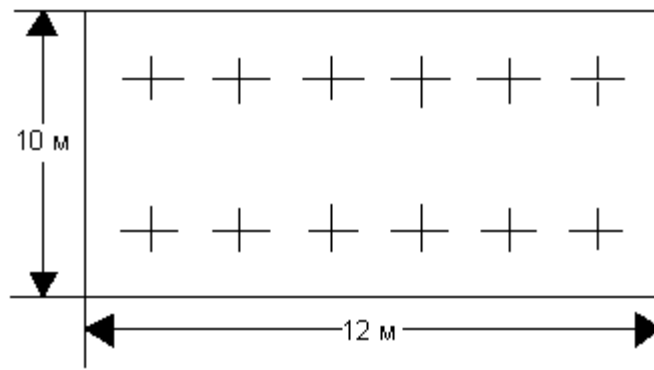


Рисунок 4.1 - Схема розташування світильників

4.4 Рекомендації по пожежній безпеці

Пожежі в приміщеннях, де встановлена обчислювальна техніка, представляють небезпеку для життя людини. Пожежі також пов'язані як з матеріальними втратами, так і з відмовою засобів обчислювальної техніки, що у свою чергу спричиняє за собою порушення ходу технологічного процесу.

Пожежа може виникнути при наявності горючої речовини та внесення джерела запалювання в горюче середовище. Пальними матеріалами в приміщеннях, де розташовані ПЕОМ, є:

- поліамід - матеріал корпусу мікросхеми, горюча речовина, температура самозаймання аерогелю 420 °С ;
- полівінілхлорид - ізоляційний матеріал, горюча речовина, температура запалювання 335 °С, температура самозаймання 530 °С, кількість енергії, що виділяється при згоранні - 18000 - 20700 кДж/кг;
- стеклотекстоліт ДЦ - матеріал друкарських плат, важкозаймистий матеріал, показник горючості 1.74, не схильний до температурного самозаймання;
- пластика кабельний №489 - матеріал ізоляції кабелю, горючий матеріал, показник горючості більш 2.1;
- деревина - будівельний і обробний матеріал, матеріал з якого виготовлені меблі, горючий матеріал, показник горючості більше 2.1, теплота згорання 18731 - 20853 кДж/кг, температура запалювання 399 °С, схильна до самозаймання.

Згідно ДСТУ Б В.1.1-36-2016 [108] приміщення відносяться до категорії В (пожежовибухонебезпечним) і згідно правилам побудови електроустановок простір

усередині приміщення відноситься до вогнебезпечної зони класу П - Па (зони, розташовані в приміщеннях, в яких зберігаються тверді горючі речовини).

Потенційними джерелами запалення при роботі ПЕОМ є:

- іскри при замиканні і розмиканні ланцюгів;
- іскри і дуги коротких замикань;
- перегриви від тривалого перевантаження і наявності перехідного опору.

Продуктами згорання, що виділяються при пожежі, є : оксид вуглецю, сірчистий газ, оксид азоту, синильна кислота, акролеїн, фосген, хлор та ін. При горінні пластмас, окрім звичайних продуктів згорання, виділяються різні продукти термічного розкладання: хлорангідридні кислоти, формальдегіди, хлористий водень, фосген, синильна кислота, аміак, фенол, ацетон, стирол та ін., що шкідливо впливають на організм людини.

Для захисту персоналу від дії небезпечних і шкідливих чинників пожежі проектом передбачається застосування промислового протигазу з коробкою марки В(жовта).

Пожежна безпека об'єктів народного господарства регламентується і забезпечується системами запобігання пожежам і протипожежному захисту[108]. Для успішного гасіння пожеж вирішальне значення має швидке виявлення пожежі і своєчасний виклик пожежних підрозділів до місця пожежі.

Зменшити горюче навантаження не представляється можливим, тому проектом передбачається застосувати наступні способи і їх комбінації для запобігання утворенню(внесення) джерел запалення :

- застосування устаткування, що задовольняє вимогам електростатичної безпеки;
- застосування в конструкції швидкодіючих засобів захисного відключення можливих джерел запалення;
- виключення можливості появи іскрового заряду статичної електрики в горючому середовищі з енергією, рівної і вище мінімальної енергії запалення;
- підтримка температури нагріву поверхні машин, механізмів, устаткування, пристроїв, речовин і матеріалів, які можуть увійти до контакту з палим середовищем, нижче гранично допустимої, становить 80% як найменшої температури самозаймання пального.
- заміна небезпечних технологічних операцій більш безпечними;
- ізольоване розташування небезпечних технологічних установок і устаткування;
- зменшення кількості палих і вибухонебезпечних речовин, що знаходяться у виробничих приміщеннях;
- запобігання можливості утворення палих сумішей на лінії, вентиляційних системах і ін.;

- механізація, автоматизація та справність(потокова) виробництва;
- суворе дотримання стандартів і точне виконання встановленого технологічного режиму;
- запобігання можливості появи в небезпечних місцях джерел запалення;
- запобігання розповсюдженню пожеж і вибухів;
- використання устаткування і пристроїв, при роботі яких не виникає джерел запалення;
- виконання вимог сумісного зберігання речовин і матеріалів;
- наявність громовідводу;
- ліквідація можливості самозаймання речовин і матеріалів .

Для запобігання пожежі в обчислювальних центрах проектом пропонується виконання наступних вимог :

- електроживлення ЕОМ повинно мати автоматичне блокування відключення електроенергії на випадок зупинки системи охолодження і кондиціонування;
- система вентиляції обчислювальних центрів повинна бути обладнана блокуючими пристроями, що забезпечують її відключення на випадок пожежі;
- робочі місця повинні бути оснащені пожежними щитами, сигналізацією, засобами для сповіщення про пожежну небезпеку (телефонами), медичними аптечками для надання першої медичної допомоги, розробленим планом евакуації.

Для зниження пожежної небезпеки в приміщеннях використовуються первинні засоби гасіння пожеж, а також система автоматичної пожежної сигналізації, яка дозволяє знайти початкову стадію загоряння, швидко і точно оповістити службу пожежної охорони про час і місце виникнення пожежі.

Відповідно до правил пожежної безпеки для промислових підприємств приміщення категорії В підлягають устаткуванню системами автоматичної пожежної сигналізації. Проектом передбачається застосування датчика типу ІДФ - 1(димовий фотоелектричний датчик), оскільки специфікою пожеж обчислювальної техніки і радіоапаратури є, в першу чергу, виділення диму, а потім - підвищення температури.

При виникненні пожежі в робочому приміщенні обслуговуючий персонал зобов'язаний негайно вжити заходи по ліквідації пожежі. Для ліквідації пожежі використовують вогнегасники (пінні для повітря ОП-5, ОП-6, ОП-9, вуглекислотні ОУ-5), пісок, пожежний інвентар (сокири, ломи, багри, шерстяну або азбестову ковдри). Як засіб індивідуального захисту проектом передбачається використання промислового протигаза з маскою, фільтруючої коробки В.

В якості організаційно-технічних заходів рекомендується проводити навчання робочого персоналу правилам пожежної безпеки.

4.5 Вплив на навколишнє природне середовище

Діяльність за темою магістерської роботи, а саме: дослідження ефективності використання засобів векторної обробки ядер сучасних процесорів в процесі її виконання впливає на навколишнє природне середовище і регламентується нормами діючого законодавства: Законом України «Про охорону навколишнього природного середовища»[111], Законом України «Про забезпечення санітарного та епідемічного благополуччя населення»[113], Законом України «Про відходи»[112].

В процесі діяльності виконанням дипломного проектування виникають процеси поводження з відходами ІТ галузі. Нижче надано перелік відходів, що утворюються в процесі роботи:

- відпрацьовані люмінесцентні лампи - I клас небезпеки.
- змінні носії інформації - IV клас небезпеки.
- відпрацьовані вогнегасники - IV клас небезпеки.
- макулатура - IV клас небезпеки

ВИСНОВКИ

Задача кластеризації масивів даних часто зустрічається в багатьох додатках, пов'язаних з інтелектуальним аналізом даних. Традиційний підхід до вирішення цих завдань передбачає, що кожне спостереження може відноситися тільки до одного кластеру, хоча більш природною видається ситуація, коли обробляється вектор ознак з різними рівнями ймовірностей або можливостей може належати відразу декільком класів. Дана ситуація є предметом розгляду нечіткого (фаззі) кластерного аналізу, що інтенсивно розвивається в даний час.

Відомі підходи і рішення працездатні лише у випадках, коли вихідний масив даних заданий заздалегідь в повному обсязі і не змінюється в процесі аналізу. У той же час існує досить широкий клас задач, коли дані надходять на обробку послідовно в on-line режимі так, як це відбувається при навчанні самоорганізованих карт Кохонена.

Розглянуто задачу ймовірнісної нечіткої on-line кластеризації даних з пропусками і показано, що вона може бути вирішена на основі самоорганізованої нейро-фаззі мережі Кохонена та запропонованого адаптивного алгоритму самонавчання, що є гібридом правила «переможець отримує більше» і рекурентних алгоритмів нечіткої кластеризації.

У розділі «Охорона праці» виконано аналіз потенційних небезпек при роботі із засобами обчислювальної техніки і механізмами, розроблені заходи щодо техніки безпеки, заходи, які забезпечують виробничу санітарію і гігієну праці, розраховане штучне освітлення, виконані рекомендації по пожежній безпеці. А також визначені основні екологічні аспекти впливу на навколишнє природне середовище та зазначені заходи щодо поводження з ними.

Результати атестаційної роботи були апробовані на міжнародній конференції «Майбутній науковець 2020».

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАНЬ

- 1) Ritter, H., & Schulten, K. (1988, July). Kohonen's self-organizing maps: exploring their computational capabilities. In Proceedings of IEEE international conference on neural networks (Vol. 1, pp. 109-116).
- 2) Kasabov, N. K. (2007). Evolving connectionist systems: the knowledge engineering approach. Springer Science & Business Media.
- 3) Lughofer, E. (2011). Evolving fuzzy systems-methodologies, advanced concepts and applications (Vol. 53). Berlin: Springer.
- 4) Gan, G., Ma, C., & Wu, J. (2007). Data clustering: theory, algorithms, and applications (Vol. 20). Siam.
- 5) Загоруйко, Н. Г. (1979). Ёмпирическое предсказание. Наука, Сиб. отд-ние.
- 6) Gorban, A. N., Kégl, B., Wunsch, D. C., & Zinovyev, A. Y. (Eds.). (2008). Principal manifolds for data visualization and dimension reduction (Vol. 58, pp. 96-130). Berlin: Springer.
- 7) Marwala, T. (Ed.). (2009). Computational Intelligence for Missing Data Imputation, Estimation, and Management: Knowledge Optimization Techniques: Knowledge Optimization Techniques. IGI Global.
- 8) Bishop, C. M. (1995). Neural networks for pattern recognition. Oxford university press.
- 9) Gorban, A. N., Rossiev, A. A., & Wunsch, D. C. (2000). Neural network modeling of data with gaps. Радіоелектроніка, інформатика, управління, (1 (3)).
- 10) Tkacz, M. (2005). Artificial neural networks in incomplete data sets processing. In Intelligent Information Processing and Web Mining (pp. 577-583). Springer, Berlin, Heidelberg.
- 11) Golden, R. M., & Golden, R. (1996). Mathematical methods for neural network analysis and design. MIT Press.
- 12) Braun, H. (2013). Neuronale Netze: Optimierung durch Lernen und Evolution. Springer-Verlag.
- 13) Dracopoulos, D. C. (2013). Evolutionary learning algorithms for neural adaptive control. Springer.
- 14) Shepherd, A. J. (2012). Second-order methods for neural networks: Fast and reliable training methods for multi-layer perceptrons. Springer Science & Business Media.
- 15) Haykin, S. (1994). Neural networks: a comprehensive foundation. Prentice Hall PTR.

- 16) Плісс, І. П., Шевякова, А. Ю., & Шевякова, Ю. Ю. (2011). Нейромережеве відновлення пропусків у таблицях даних. Наукові праці [Чорноморського державного університету імені Петра Могили]. Сер.: Комп'ютерні технології, (160, Вип. 148), 59-61.
- 17) Shafronenko, A., Pliss, I., & Ye, B. (2013). The evolving adaptive neural network for data processing with missing observations. *Радіоелектроніка, інформатика, управління*, (2 (29)).
- 18) Волкова, В. В., & Шафроненко, А. Ю. (2011). Нечітка кластеризація масивів даних з пропущеними значеннями. Індуктивне моделювання складних систем.
- 19) Плісс, І. П., Шевякова, А. Ю., & Шевякова, Ю. Ю. (2011). Нейромережеве відновлення пропусків у таблицях даних. Наукові праці [Чорноморського державного університету імені Петра Могили]. Сер.: Комп'ютерні технології, (160, Вип. 148), 59-61.
- 20) McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 5(4), 115-133.
- 21) Braun, H., Feulner, J., & Malaka, R. (2013). *Praktikum Neuronale Netze*. Springer-Verlag.
- 22) Beltratti, A., Margarita, S., & Terna, P. (1996). *Neural networks for economic and financial modelling*. London, UK: International Thomson Computer Press.
- 23) Hunt, K. J., Sbarbaro, D., Żbikowski, R., & Gawthrop, P. J. (1992). Neural networks for control systems—a survey. *Automatica*, 28(6), 1083-1112.
- 24) Leondes, C. T. (1998). *Neural network systems techniques and applications: Advances in theory and applications (Vol. 7)*. Academic Press.
- 25) Nelles, O., Ernst, S., & Isermann, R. (1997). Neuronale Netze zur Identifikation nichtlinearer, dynamischer Systeme: Ein Überblick. *at-Automatisierungstechnik*, 45(6), 251-262.
- 26) Ротштейн, А. П. (1999). Интеллектуальные технологии идентификации: нечеткие множества, генетические алгоритмы, нейронные сети. Винница: УНІВЕРСУМ-Вінниця, 320.
- 27) Tsoukalas, L. H., & Uhrig, R. E. (1996). *Fuzzy and neural approaches in engineering*. John Wiley & Sons, Inc.
- 28) Jang, J. S. (1993). ANFIS: adaptive-network-based fuzzy inference system. *IEEE transactions on systems, man, and cybernetics*, 23(3), 665-685.
- 29) Jang, J. S., & Sun, C. T. (1993). Functional equivalence between radial basis function networks and fuzzy inference systems. *IEEE transactions on Neural Networks*, 4(1), 156-159.
- 30) Han, J., Pei, J., & Kamber, M. (2011). *Data mining: concepts and techniques*. Elsevier.

- 31) Specht D.F. Probobalistic neural networks // *Neural Networks*. – 1990.-3.-p.109-118
- 32) Yamakawa, T. (1992). A neo fuzzy neuron and its applications to system identification and prediction of the system behavior. In *Proc. of the 2nd Int. Conf. on Fuzzy Logic & Neural Networks* (pp. 477-483).
- 33) Uchino, E., & Yamakawa, T. (1997). Soft computing based signal prediction, restoration, and filtering. In *Intelligent Hybrid Systems* (pp. 331-351). Springer, Boston, MA.
- 34) Miki, T. S. U. T. O. M. U., & Yamakawa, T. A. K. E. S. H. I. (1999). Analog implementation of neo-fuzzy neuron and its on-board learning. *Computational Intelligence and Applications*, 144-149.
- 35) Hathaway, R. J., & Bezdek, J. C. (2001). Fuzzy c-means clustering of incomplete data. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 31(5), 735-744.
- 36) Bezdek, J. C. (2013). *Pattern recognition with fuzzy objective function algorithms*. Springer Science & Business Media.
- 37) Суетин, П. (2018). *Классические ортогональные многочлены*. Litres.
- 38) Karlin, S., & Studden, W. J. (1966). *Tchebyeff systems: with applications in analysis and statistics*.
- 39) Семесенко, М. П. (1983). *Методы обработки и анализа измерений в научных исследованиях*. К.: Донецк: Вища шк.
- 40) Yang, S. S., & Tseng, C. S. (1996). An orthogonal neural network for function approximation. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 26(5), 779-785.
- 41) Lee, T. T., & Jeng, J. T. (1998). The Chebyshev-polynomials-based unified model neural networks for function approximation. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 28(6), 925-935.
- 42) András, P. (1999). Orthogonal RBF neural network approximation. *Neural Processing Letters*, 9(2), 141-151.
- 43) Sher, C. F., Tseng, C. S., & Chen, C. S. (2001). Properties and performance of orthogonal neural network in function approximation. *International Journal of intelligent systems*, 16(12), 1377-1392.
- 44) Patra, J. C., & Kot, A. C. (2002). Nonlinear dynamic system identification using Chebyshev functional link artificial neural networks. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 32(4), 505-511.

- 45) Ye, B., Kolodyazhniy, V., & Slipchenko, O. (2004). Artificial neural network with orthogonal activation functions for dynamic system identification. *Synergies between Information Processing and Automation*. Ed. O. Sawodny and P. Scharff–Aachen: Shaker Verlag, 24-30.
- 46) Stasiak, B., & Yatsymirskyy, M. (2006, June). Fast orthogonal neural networks. In *International Conference on Artificial Intelligence and Soft Computing* (pp. 142-149). Springer, Berlin, Heidelberg.
- 47) Rodriguez, N., & Cubillos, C. (2007, September). Orthogonal neural network based predistortion for OFDM systems. In *Electronics, Robotics and Automotive Mechanics Conference (CERMA 2007)* (pp. 225-228). IEEE.
- 48) Wang, H., & Yu, S. (2010). Tracking control of robot manipulators based on orthogonal neural network. *International Journal of Modelling, Identification and Control*, 11(1-2), 130-135.
- 49) Ye, B., Kolodyazhniy, V., & Slipchenko, O. (2004). Structural and synaptic adaptation in the artificial neural networks with orthogonal activation functions. *Sci. Proc. of Riga Technical University. Comp. Sci., Inf. Technology and Management Sci*, 20, 69-76.
- 50) Jain, A. K., Murty, M. N., & Flynn, P. J. (1999). Data clustering: a review. *ACM computing surveys (CSUR)*, 31(3), 264-323.
- 51) Дюк, В., & Самойленко, А. (2001). *Data Mining: учебный курс*. СПб.: Питер, 368, 16.
- 52) Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: data mining, inference, and prediction*. Springer Science & Business Media.
- 53) Xu, R., & Wunsch, D. C. (2005). *Survey of clustering algorithms*.
- 54) Abonyi, J., & Feil, B. (2007). *Cluster analysis for data mining and system identification*. Springer Science & Business Media.
- 55) Olson, D. L., & Delen, D. (2008). *Advanced data mining techniques*. Springer Science & Business Media.
- 56) Höppner, F., Klawonn, F., Kruse, R., & Runkler, T. (1999). *Fuzzy cluster analysis: methods for classification, data analysis and image recognition*. John Wiley & Sons.
- 57) Chung, F. L., & Lee, T. (1994). Fuzzy competitive learning. *Neural Networks*, 7(3), 539-551.
- 58) Park, D. C., & Dagher, I. (1994, June). Gradient based fuzzy c-means (GBFCM) algorithm. In *Proceedings of 1994 IEEE International Conference on Neural Networks (ICNN'94)* (Vol. 3, pp. 1626-1631). IEEE.

- 59) Krishnapuram, R., & Keller, J. M. (1993). A possibilistic approach to clustering. *IEEE transactions on fuzzy systems*, 1(2), 98-110.
- 60) Bodyanskiy, Y. (2005). Computational Intelligence Techniques for Data Analysis. In *Leipziger Informatik-Tage* (pp. 15-36).
- 61) Davé, R. N., & Krishnapuram, R. (1997). Robust clustering methods: a unified view. *IEEE Transactions on fuzzy systems*, 5(2), 270-293.
- 62) Bezdek, J. C., Keller, J., Krishnapuram, R., & Pal, N. (1999). *Fuzzy models and algorithms for pattern recognition and image processing* (Vol. 4). Springer Science & Business Media.
- 63) Klawonn, F. (2013, November). What can Fuzzy cluster analysis contribute to clustering of high-dimensional data?. In *International Workshop on Fuzzy Logic and Applications* (pp. 1-14). Springer, Cham.
- 64) Keller, A., & Klawonn, F. (2000). Fuzzy clustering with weighting of data variables. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 8(06), 735-746.
- 65) Maimon, O., & Rokach, L. (Eds.). (2007). *Soft computing for knowledge discovery and data mining*. Springer Science & Business Media.
- 66) Rutkowski, L. (2008). *Computational intelligence: methods and techniques*. Springer Science & Business Media.
- 67) Perfilieva, I. (2006). Fuzzy transforms: Theory and applications. *Fuzzy sets and systems*, 157(8), 993-1023.
- 68) Perfilieva, I. (2004). Fuzzy transforms. In *Transactions on rough sets II* (pp. 63-81). Springer, Berlin, Heidelberg.
- 69) Cochocki, A., & Unbehauen, R. (1993). *Neural networks for optimization and signal processing*. John Wiley & Sons, Inc..
- 70) Little, R. J., & Rubin, D. B. (2019). *Statistical analysis with missing data* (Vol. 793). John Wiley & Sons.
- 71) Злоба, Е., & Яцкив, И. (2002). Статистические методы восстановления пропущенных данных. *Computer Modelling & New Technologies*, 6(1), 51-61.
- 72) Dave, R. N., & Sen, S. (1997, June). On generalizing the noise clustering algorithms. In *Proc. 7th Int. Fuzzy Systems Association World Congress (IFSA '97)* (Vol. 3, pp. 205-210).
- 73) Höppner, F., Klawonn, F., Kruse, R., & Runkler, T. (1999). *Fuzzy cluster analysis: methods for classification, data analysis and image recognition*. John Wiley & Sons.

- 74) Zhang, J. S., & Leung, Y. W. (2003). Robust clustering by pruning outliers. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 33(6), 983-998.
- 75) Wu, C. W., Chen, J. L., & Wang, J. H. (2001, October). Self-organizing mountain method for clustering. In *2001 IEEE International Conference on Systems, Man and Cybernetics. e-Systems and e-Man for Cybernetics in Cyberspace (Cat. No. 01CH37236) (Vol. 4, pp. 2434-2438)*. IEEE.
- 76) Yao, Y., Chen, Y. Q., & Chen, L. (2000). A novel similarity measure for data clustering. *Intelligent Data Analysis*, 4(5), 421-431.
- 77) Veenman, C. J., & Reinders, M. J. (2005). The nearest subclass classifier: A compromise between the nearest mean and nearest neighbor classifier. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(9), 1417-1429.
- 78) Zhang, C., Qin, Y., Zhu, X., Zhang, J., & Zhang, S. (2006, August). Clustering-based missing value imputation for data preprocessing. In *2006 4th IEEE International Conference on Industrial Informatics (pp. 1081-1086)*. IEEE.
- 79) Yong, Z., Yue'e, L., & Shixiong, X. (2008, December). Robust fuzzy-possibilistic c-means algorithm. In *2008 Second International Symposium on Intelligent Information Technology Application (Vol. 1, pp. 669-673)*. IEEE.
- 80) Han, J., Pei, J., & Kamber, M. (2011). *Data mining: concepts and techniques*. Elsevier.
- 81) Hawkins, S., He, H., Williams, G., & Baxter, R. (2002, September). Outlier detection using replicator neural networks. In *International Conference on Data Warehousing and Knowledge Discovery (pp. 170-180)*. Springer, Berlin, Heidelberg.
- 82) Babuka, R., Van der Veen, P. J., & Kaymak, U. (2002, May). Improved covariance estimation for Gustafson-Kessel clustering. In *2002 IEEE World Congress on Computational Intelligence. 2002 IEEE International Conference on Fuzzy Systems. FUZZ-IEEE'02. Proceedings (Cat. No. 02CH37291) (Vol. 2, pp. 1081-1085)*. IEEE.
- 83) Tipping, M. E., & Bishop, C. M. (1999). Mixtures of probabilistic principal component analyzers. *Neural computation*, 11(2), 443-482.
- 84) Cowell, R. G., Dawid, P., Lauritzen, S. L., & Spiegelhalter, D. J. (2006). *Probabilistic networks and expert systems: Exact computational methods for Bayesian networks*. Springer Science & Business Media.
- 85) Зайченко, Ю. П. (2008). Нечеткие модели и методы в интеллектуальных системах. К.: Слово, 344.
- 86) Круглов, В. В., & Борисов, В. В. (2002). *Искусственные нейронные сети. Теория и практика*. М.: Горячая линия-Телеком.

- 87) Згуровский, М. З., & Зайченко, Ю. П. (2013). Основы вычислительного интеллекта. К.: Наук. думка.
- 88) Yin, Y., Kaku, I., Tang, J., & Zhu, J. (2011). Data mining: Concepts, methods and applications in management and engineering design. Springer Science & Business Media.
- 89) Everitt, B. S., Landau, S., Leese, M., & Stahl, D. (2011). Cluster Analysis. —John Wiley & Sons. Ltd., New York, 330.
- 90) Miamoto, S., Ichihashi, H., & Honda, K. (2008). Algorithms for Fuzzy Clustering: Methods in c-Means Clustering with Applications.
- 91) De Oliveira, J. V., & Pedrycz, W. (Eds.). (2007). Advances in fuzzy clustering and its applications. John Wiley & Sons.
- 92) Mueller, E. T. (2011). Neurocomputing: Learning, Architectures and Modeling. Nova Science Publishers, Inc..
- 93) Suresh, S., Sundararajan, N., & Savitha, R. (2013). Supervised learning with complex-valued neural networks (pp. 125-132). Berlin: Springer.
- 94) Babinec, Š., & Pospíchal, J. (2008, November). Gating echo state neural networks for time series forecasting. In International Conference on Neural Information Processing (pp. 200-207). Springer, Berlin, Heidelberg.
- 95) Crone, S. F., Nikolopoulos, K., & Hibon, M. (2005). Automatic modelling and forecasting with artificial neural networks—A forecasting competition evaluation. Final report for the IIF/SAS Grant, 6, 2008.
- 96) Baczynski, M., & Jayaram, B. (2008). Fuzzy implications. Heidelberg: Springer.
- 97) Zhao, F., Fan, J., & Liu, H. (2014). Optimal-selection-based suppressed fuzzy c-means clustering algorithm with self-tuning non local spatial information for image segmentation. Expert systems with applications, 41(9), 4083-4093.
- 98) Lin, P. L., Huang, P. W., Kuo, C. H., & Lai, Y. H. (2014). A size-insensitive integrity-based fuzzy c-means method for data clustering. Pattern Recognition, 47(5), 2042-2056.
- 99) De Almeida, C. W., De Souza, R. M., & Candeias, A. L. (2013). Fuzzy Kohonen clustering networks for interval data. Neurocomputing, 99, 65-75.
- 100) Mantula, E. V., & Mashtalir, S. V. (2013). Method of Adaptive Forecasting Based On Multidimensional Linear Extrapolation. International Journal of Research in Engineering and Science, 1(4), 31-37.
- 101) ДСН 3.3.6.042-99 Державні санітарні норми мікроклімату виробничих приміщень. Режим доступу: www. URL: <https://zakon.rada.gov.ua/rada/show/va042282-99>

102) НПАОП 0.00-7.15-18 Вимоги щодо безпеки та захисту здоров'я працівників під час роботи з екранними пристроями. Режим доступу: [www. URL: https://zakon.rada.gov.ua/laws/show/z0508-18](http://www.zakon.rada.gov.ua/laws/show/z0508-18)

103) ДСТУ 7237:2011 Національний стандарт України. Система стандартів безпеки праці. Електробезпека. Загальні вимоги та

104) Номенклатура видів захисту. Режим доступу: [www. URL: https://zakon.rada.gov.ua/rada/show/ru/v0037831-11](http://www.zakon.rada.gov.ua/rada/show/ru/v0037831-11)

105) ДСанПіН 3.3.2.007-98. Державні санітарні правила і норми. Гігієнічні вимоги до організації роботи з візуальними дисплейними терміналами електронно-обчислювальних машин. Режим доступу: [www. URL: https://zakon.rada.gov.ua/rada/show/v0007282-98](http://www.zakon.rada.gov.ua/rada/show/v0007282-98)

106) ДБН В.2.5-67:2013. Опалення вентиляція та кондиціонування. Режим доступу: [www. URL: https://zakon.rada.gov.ua/rada/show/v0024858-13](http://www.zakon.rada.gov.ua/rada/show/v0024858-13)

107) ДБН В.2.5-28-2006. Природне і штучне освітлення. Режим доступу: [www. URL: https://zakon.rada.gov.ua/rada/show/v0168667-06](http://www.zakon.rada.gov.ua/rada/show/v0168667-06)

108) ДСТУ Б В.1.1-36-2016. Визначення категорії приміщень, будинків та зовнішніх установок за вибухопожежною та пожежною безпекою. Режим доступу: [www. URL: http://online.budstandart.com/ua/catalog/doc-page.html?id_doc=65419](http://online.budstandart.com/ua/catalog/doc-page.html?id_doc=65419)

109) ДСП 173-96. Державні санітарні правила планування та забудови населених пунктів Режим доступу: [www. URL: https://zakon.rada.gov.ua/laws/show/z0379-96](http://www.zakon.rada.gov.ua/laws/show/z0379-96)

110) Симметрон. Электронные компоненты. Каталог 2002, 2002г. – 192с.

111) Закон України «Про охорону навколишнього природного середовища» . Вводиться в дію Постановою ВР № 1268-ХІІ від 26.06.91, ВВР, 1991, № 41, ст.547. Режим доступу: [www. URL: https://zakon.rada.gov.ua/laws/show/1264-12](http://www.zakon.rada.gov.ua/laws/show/1264-12)

112) Закон України «Про відходи». Відомості Верховної Ради України (ВВР), 1998, № 36-37, ст.242. Режим доступу: [www. URL: https://zakon.rada.gov.ua/laws/show/187/98-вр](http://www.zakon.rada.gov.ua/laws/show/187/98-вр)

113) Закон України «Про забезпечення санітарного та епідемічного благополуччя населення». Відомості Верховної Ради України (ВВР), 1994, № 27, ст.218. Режим доступу: [www. URL: https://zakon.rada.gov.ua/laws/show/4004-12](http://www.zakon.rada.gov.ua/laws/show/4004-12)

ДОДАТОК А.

Лістинг коду

```

close all;
clear all;
load iris.dat;
Exp2=iris(:,1:4);

Exp1=Exp2';
[n,m]=size(Exp1);

%-----normalization-----
%-----

for k=1:n
    for q=1:m
        ExpP(k,:)=Exp1(k,:)-min(Exp1(k,:));
        ExpPP(k,:)=(max(Exp1(k,:)))-(min(Exp1(k,:)));
        ExpPPP(k,:)=ExpP(k,:)/ExpPP(k,:);
        Exper(k,:)=ExpPPP(k,:)*2;
        Experiment(k,:)=Exper(k,:)-1;
    end;
end;

%-----
Exp1=Experiment';
clear ExpP;
clear ExpPP;
clear ExpPPP;
clear Exper;
clear Experiment;

%-----
-----
%-----adaptive probabilistic fcm with gaps-----
-----
%-----on the optimum expansion procedure -----
-----
%-----
-----

[n,m]=size(Exp1);
Exp=Exp1;

c=3;%the number of clusters
eta=0.95;
beta=2;

W1=rand(1,m);

```

```

W2=rand(1,m);
W3=rand(1,m);

x1=rand(1,m);
x2=rand(1,m);
x3=rand(1,m);

e1=ones(n-1,c);
e2=ones(n-1,c);
e3=ones(n-1,c);

err1=1;
err2=1;
err3=1;

P=[err1,err2,err3];

count=0;

m1=0;
n1=2;
for ii=1:n-9
    m1=9+ii;
    for i=1:n
        for j=1:n1
            Exp(i,j)=Exp1(i,j);
        end;
    end;
end;

Exp(2,1)=NaN;
Exp(18,2)=NaN;

for k=1:n
    for q=1:m
        if isnan(Exp(k,q))
            Exp2(k,q)=x1(1,q);
            Exp3(k,q)=x2(1,q);
            Exp4(k,q)=x3(1,q);
        else
            Exp2(k,q)=Exp(k,q);
            Exp3(k,q)=Exp(k,q);
            Exp4(k,q)=Exp(k,q);
        end;
    end;
end;

for k=1:n
    for q=1:m
        D1(k,:)=(norm(Exp2(k,:)-W1)).^2;
        D2(k,:)=(norm(Exp3(k,:)-W2)).^2;
        D3(k,:)=(norm(Exp4(k,:)-W1)).^2;
    end;
end;

```

```

end;
end;

while (P>0.01)

for k=1:n
    for q=1:m
        A(k,:)=(1/(D1(k,:)));
        A1(k,:)=(1/(D2(k,:)));
        A2(k,:)=(1/(D3(k,:)));
        B(k,:)=A(k, :)+A1(k, :)+A2(k, :);

        U1(k,:)=A(k, :)/B(k, :);
        U2(k,:)=A1(k, :)/B(k, :);
        U3(k,:)=A2(k, :)/B(k, :);
        U(k,:)=U1(k, :)+U2(k, :)+U3(k, :);

        E(k,:)=(U1(k, :)).^2;
        E1(k,:)=(U2(k, :)).^2;
        E2(k,:)=(U3(k, :)).^2;

        H1=sum(E);
        H2=sum(E1);
        H3=sum(E2);

        L1(k,:)=Exp2(k, :)*E(k, :);
        N1=sum(L1);
        W1(k,:)=N1/H1;

        L2(k,:)=Exp3(k, :)*E1(k, :);
        N2=sum(L2);
        W2(k,:)=N2/H2;

        L3(k,:)=Exp4(k, :)*E2(k, :);
        N3=sum(L3);
        W3(k,:)=N3/H3;

        S1(k,:)=E(k, :)*W1(k, :);
        Sm1=sum(S1);
        x1=Sm1/H1;

        S2(k,:)=E1(k, :)*W2(k, :);
        Sm2=sum(S2);
        x2=Sm2/H2;

        S3(k,:)=E2(k, :)*W3(k, :);
        Sm3=sum(S3);
        x3=Sm3/H3;

        if isnan(Exp(k, q))
            Exp12(k, q)=x1(1, q);
            Exp13(k, q)=x2(1, q);

```

```
        Exp14(k,q)=x3(1,q);
    else
        Exp12(k,q)=Exp(k,q);
        Exp13(k,q)=Exp(k,q);
        Exp14(k,q)=Exp(k,q);
    end;

    end;
end;

for kk=2:n

    e1=W1(kk,:)-W1(kk-1,:);
    e2=W2(kk,:)-W2(kk-1,:);
    e3=W3(kk,:)-W3(kk-1,:);
end;

err1=mse(e1);
err2=mse(e2);
err3=mse(e3);

count=count+1;

end;
```

ДОДАТОК Б.
Електронні плакати

*Міністерство освіти і науки України
Східноукраїнський національний університет
ім. В.Даля
Кафедра комп'ютерних наук та інженерії*

*Магістерська робота
«Метод нечіткої кластеризації
даних з пропусками»*

*Студентка: Халілова Айнур
Гіяс кизи
Керівник: Рязанцев О.І.*

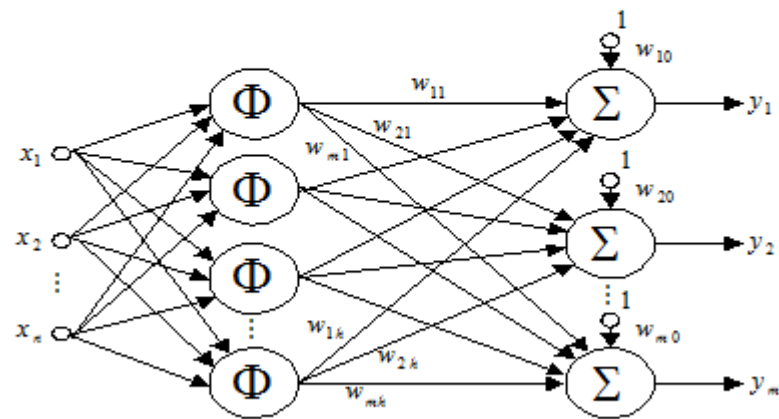
МЕТА ДИПЛОМНОЇ РОБОТИ

Сьогодні є достатня кількість інформації про підприємницьку діяльність підприємств, лікарень, фірм, яка показує діяльність цих закладів. Ця інформація збирається протягом довгого часу (темпи інфляції, рівень прибутку населення, витрати народу, ціну послуг житлово-комунального господарства, стан промислового і сільськогосподарського виробництва, своєчасність виплати заробітної плати та пенсій, тощо). Зрозуміло, такі дані складно обробити вручну через великий обсяг інформації і складних нелінійних причинно-наслідкових зв'язків. Тому і виникла потреба у створенні новіших методів аналізу, кластеризації і прогнозування, до яких відносять машинні методи виявлення послідовних зв'язків.

Метою роботи є розробка ймовірнісного методу нечіткої on-line кластеризації даних що може бути вирішена на основі самоорганізовної нейро-фаззі мережі Кохонена.

Об'єктом дослідження є послідовність вибірки даних UCI репозиторія, а саме такі як Iris та Wine.

РАДІАЛЬНО-БАЗИСНА НЕЙРОННА МЕРЕЖА



ФУНКЦІЇ ВІДСТАНІ

- 1) Евклідова відстань
- 2) Відстань міських кварталів (Манхеттенська відстань)
- 3) Відстань Чебишева
- 4) Часткова відстань
- 5) Методи кластеризації даних
- 6) Адаптивний метод ймовірнісної нечіткої кластеризації даних

ПРОГРАМНА РЕАЛІЗАЦІЯ

Базові операції MATLAB:

- Інтерполяція і регресія
- Диференціювання та інтегрування
- Системи лінійних рівнянь
- Фур'є аналіз
- Власні значення і сингулярні числа матриць
- Звичайні диференціальні рівняння
- Розріжені матриці

ІНФОРМАЦІЯ ПРО НАБІР ДАНИХ

Атрибути:

1. Alcohol;
2. Malic acid;
3. Ash;
4. Alcalinity of ash;
5. Magnesium;
6. Total phenols;
7. Flavanoids;
8. Nonflavanoid phenols;
9. Proanthocyanins;
10. Color intensity;
11. Hue;
12. OD280/OD315 of diluted wines;
13. Proline.

ЕКСПЕРИМЕНТАЛЬНІ ДОСЛІДЖЕННЯ

Таблиця 1 – Результати експериментальних досліджень, порівняльний аналіз кластеризації даних вибірки Iris UCI repository

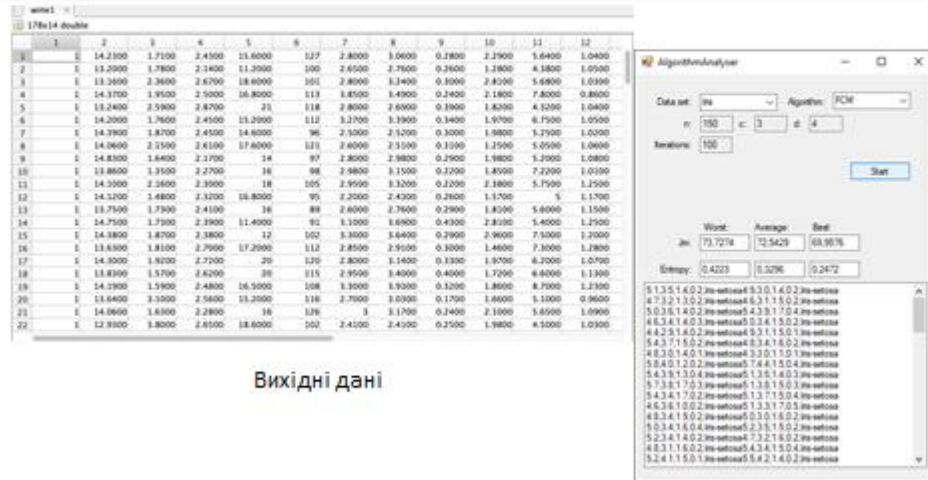
Алгоритми	Iris UCI repository					
	PC	CE	SC	S	XB	DI
Адаптивний ймовірнісний нечіткий метод кластеризації даних	0,0199	0,0122	-0,2439	0,0022	0,0015	1
FCM	0,8011	0,3410	0,2567	0,0030	7,1965	0,0080

ЕКСПЕРИМЕНТАЛЬНІ ДОСЛІДЖЕННЯ

Таблиця 1 – Результати експериментальних

Алгоритми	Wine UCI repository					
	PC	CE	SC	S	XB	DI
Адаптивний ймовірнісний нечіткий метод кластеризації даних	0,0230	0,0219	-0,3007	0,0032	0,0065	0,9999
FCM	0,7411	0,2389	0,3112	0,0040	6,9945	0,0078

ЕКСПЕРИМЕНТАЛЬНІ ДОСЛІДЖЕННЯ



Результати роботи програми

ВИСНОВКИ

Задача кластеризації масивів даних часто зустрічається в багатьох додатках, пов'язаних з інтелектуальним аналізом даних. Традиційний підхід до вирішення цих завдань передбачає, що кожне спостереження може відноситися тільки до одного кластеру, хоча більш природною видається ситуація, коли обробляється вектор ознак з різними рівнями ймовірностей або можливостей може належати відразу декільком класів. Дана ситуація є предметом розгляду нечіткого (фаззі) кластерного аналізу, що інтенсивно розвивається в даний час.

Відомі підходи і рішення працездатні лише у випадках, коли вихідний масив даних заданий заздалегідь в повному обсязі і не змінюється в процесі аналізу. У той же час існує досить широкий клас задач, коли дані надходять на обробку послідовно в on-line режимі так, як це відбувається при навчанні самоорганізованих карт Кохонена.

Розглянуто задачу ймовірнісної нечіткої on-line кластеризації даних з пропусками і показано, що вона може бути вирішена на основі самоорганізованої нейро-фаззі мережі Кохонена та запропонованого адаптивного алгоритму самонавчання, що є гібридом правила «переможець отримує більше» і рекурентних алгоритмів нечіткої кластеризації.