

**МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ ІМЕНІ ІГОРЯ
СІКОРСЬКОГО»**

СЕМЕНЮК РОМАН СЕРГІЙОВИЧ



УДК 681.142

**МЕТОДИ ОПРАЦЮВАННЯ ВИМІРЮВАЛЬНОЇ ТА ЕКСПЕРТНОЇ
ІНФОРМАЦІЇ З ЗАСТОСУВАННЯМ ШКАЛ КЛАСИФІКАЦІЇ**

Спеціальність 05.01.02 – стандартизація,
сертифікація та метрологічне забезпечення

АВТОРЕФЕРАТ
дисертації на здобуття наукового ступеня
кандидата технічних наук

Київ - 2021

Дисертацією є рукопис.

Робота виконана на кафедрі інформаційно-вимірювальних технологій Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського» Міністерства освіти і науки України.

Науковий керівник: кандидат технічних наук, доцент

Яремчук Ніна Антонівна

Національний технічний університет України

«Київський політехнічний інститут імені Ігоря

Сікорського», професор кафедри інформаційно-вимірювальних технологій.

Офіційні опоненти: доктор технічних наук, професор

Кошева Лариса Олександрівна

Національний авіаційний університет,

професор кафедри біокібернетики та

аерокосмічної медицини;

кандидат технічних наук

Борщов Павло Іванович

Інститут електродинаміки НАН України,

старший науковий співробітник відділу

електричних і магнітних вимірювань №5.

Захист відбудеться «27» квітня 2021 р., о 14:30 годині на засіданні спеціалізованої вченої ради Д 26.002.20 при Національному технічному університеті України «Київський політехнічний інститут імені Ігоря Сікорського» за адресою: 03056, м. Київ, пр. Перемоги, 37, корп. №22, ауд. №701.

З дисертацією можна ознайомитися у Науково-технічній бібліотеці ім.Г.І. Денисенка Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського» за адресою: 03056, м. Київ, пр.Перемоги, 37.

Автореферат розісланий «24» березня 2021 р.

Учений секретар
спеціалізованої вченої
ради



А.І. Замулко

ЗАГАЛЬНА ХАРАКТЕРИСТИКА РОБОТИ

Актуальність теми. В сучасних засобах вимірювальної техніки, особливо в інтелектуальних засобах вимірювання (ІЗВ), застосовується процедура класифікації, яку можна представити як додаткову операцію, що дозволяє удосконалити процедуру вимірювання. Тому в представленій роботі проведено дослідження і розробку методів застосування класифікації при опрацюванні експериментальних даних, як числових (кількісних), так і вербальних (якісних). У розроблені та вирішенні питань застосування класифікації внесли вагомий вклад вітчизняні та зарубіжні вчені: Орнатський П.П., Айвазян С.А., Кендал М., Стьюарт А., Розенберг В.Я., Bashkansky E., Franceschini F. В області опрацювання числових даних актуальним питанням є оптимізація вибору алгоритмів функціонування і опрацювання даних за умов активного використання апріорної і поточної вимірювальної інформації. До таких задач відноситься визначення виду розподілу або класу розподілу вибірок малого об'єму, тобто їх попередня класифікація з метою визначення виду розподілу. Наявність такої інформації дозволяє вибрати найбільш ефективну оцінку вимірюваної величини, що має найменшу стандартну невизначеність і відповідно коефіцієнт охоплення при обчисленні розширеної невизначеності. В ІЗВ ідентифікація виду розподілу включається в операцію класифікації і повинна виконуватися без участі оператора. Проте на даний момент залишаються недостатньо розробленими задачі визначення розподілу малих вибірок, тому в дисертації набули подальшого розвитку питання непараметричної класифікації виду розподілу, яка нечутлива до помірних змін центральної тенденції вибірок, що є значною перевагою для її застосування в контрольних картах (КК) технологічних процесів. Про необхідність визначення класу розподілу при побудові КК зазначено в стандарті ISO 8258 – 2001.

В інтелектуальних вимірювальних системах (ІВС) результати вимірювань використовуються при нечітких виведеннях і нечітких (м'яких) обчисленнях. Тому в таких системах має місце поліморфізм, тобто відображення вимірюваних властивостей декількома шкалами, а саме метричною шкалою і шкалою класифікації. За М. Кендалом і А. Стьюартом класифікація, що забезпечує перехід від числових даних, поданих за метричною шкалою, до вербальних (класифікованих) даних, називається метричною. Основною складністю цього переходу, на думку спеціалістів, є необхідність розробки теоретичного обґрунтування побудови шкали класифікації з урахуванням складових невизначеності вимірювальної інформації та неоднозначності або нечіткості експертної інформації. Технологія об'єднання вимірювальної та експертної інформації повинна супроводжуватись характеристиками якості метричної класифікації. Розв'язання цих задач набуває особливого значення для ІВС, що використовуються в промисловості і в народному господарстві, при медичних дослідженнях, тобто там, де невизначеність вимірювальної інформації сумірна з нечіткістю експертної інформації.

Опрацювання вербальних даних, отриманих за метричною класифікацією, може бути проведено з урахуванням відстані між класами еквівалентності шкали класифікації. Це забезпечує можливість розробки технологій опрацювання вербальних даних при нерівномірному розташуванні класів еквівалентності. Існуючі технології орієнтовані на рівномірне розміщення класів еквівалентності в області їх визначення.

Актуальною задачею є розробка методів опрацювання експертної інформації у вигляді вибірок вербальних даних з метою визначення (класифікації) стану об'єкта за центральною тенденцією вибірок експертних даних, що відображають властивості об'єкта, та ранжування вибірок для побудови контрольних карт за вербальними даними.

Зв'язок роботи з науковими програмами, планами, темами. Дослідження, результати яких викладені в дисертації пов'язані з науковою тематикою та планами

кафедри інформаційно-вимірювальних технологій НТУУ «КПІ ім. Ігоря Сікорського», а також НДР держбюджетної теми «Система метрологічного забезпечення реалізацій завдань агромоніторингу» (№ ДР 0115U000004) та НДР «Теоретичні та практичні засади застосування м'яких вимірювань в системах визначення якості продукції» (№ ДР 0118U001478).

Мета і задачі дослідження. *Метою дисертаційної роботи є розробка методів опрацювання вимірювальної та експертної інформації (числової та вербальної) з використанням шкал класифікації.*

Для досягнення поставленої мети вирішено наступні задачі дослідження:

1. Розробити непараметричну класифікацію виду розподілу малих вибірок числових даних, що може бути використана для оптимізації методів опрацювання даних, а також при побудові контрольних карт технологічних процесів.

2. Визначити основні етапи переходу від числових до вербальних даних при метричній класифікації і розробити питання врахування нечіткості експертної інформації і невизначеності числових даних при побудові шкали класифікації, розробити критерії оцінювання якості класифікації.

3. Удосконалити методи опрацювання вербальних (експертних) даних для їх подальшої класифікації за центром малої вибірки для характеристики стану об'єктів та за домінуванням при побудові КК за якісними характеристиками контрольованих властивостей.

4. Розробити технології використання шкали класифікації при вирішенні практичних задач господарської діяльності, а саме для оптимізації норми внесення добрив в ґрунти сільськогосподарського призначення, для перевірки методик отримання проб ґрунту, для визначення рівня заповнення підземних урн сміттям, для ІВС медичного призначення, тощо.

Об'єктом дослідження є процес опрацювання числових та вербальних даних в вибірках малого об'єму.

Предметом дослідження є методи опрацювання вимірювальної та експертної інформації зі встановленням шкал класифікації як для удосконалення процедур опрацювання даних, так і для переходу від числових даних до вербальних.

Методи дослідження. Розв'язання поставлених задач виконано з використанням основних положень системного аналізу та теорії вимірювань; теорії ймовірностей, математичної статистики, теорії нечітких множин та методів експертного оцінювання, теорії шкал.

Наукова новизна одержаних результатів. На основі проведених досліджень отримано наступні результати:

– вперше розроблено метод непараметричної класифікації видів розподілу малих вибірок, який відрізняється можливістю вибрати ефективну оцінку центру розподілу вибірки та визначити клас сумарного розподілу вибірок за нестабільності технологічного процесу при побудові контрольних карт;

– набули подальшого розвитку методи переходу від числових до вербальних даних в інформаційних вимірювальних системах із застосуванням шкали метричної класифікації шляхом об'єднання нечіткої експертної інформації і невизначеності вимірювань, що дозволило отримати характеристику якості встановленої шкали класифікації у вигляді матриці відповідності;

– вперше розроблено метод перевірки узгодженості між вербальними даними в вибірках, отриманих за шкалою метричної класифікації, який відрізняється можливістю врахування нерівномірності розташування класів еквівалентності в області їх визначення;

– вперше розроблено метод класифікації стану об'єкту за експертною інформацією на основі застосування методів математичної статистики і нечіткої логіки для роботи з вибірками вербальних даних, який відрізняється можливістю використання отриманих шкал класифікації для побудови контрольних карт за якісними ознаками.

Практичне значення одержаних результатів полягає у тому, що розроблені методи опрацювання даних з застосуванням класифікації можуть бути використані при побудові ІВС для удосконалення алгоритмів опрацювання даних багаторазових вимірювань, при побудові контрольних карт технологічних процесів на основі числових і вербальних даних, при отриманні результатів «м'яких» вимірювань. Результати дослідження метричної класифікації були використані в методиках визначення нерівномірності розподілу добрив при агромоніторингу, при побудові ІВС для визначення стану заповнення сміттям підземних урн. Способи класифікації вибірок вербальних даних можуть бути широко застосовані при візуальному і тактильному контролі якості продукції.

Особистий внесок здобувача. Всі результати теоретичного аналізу та експериментальних досліджень дисертаційної роботи, висновки та рекомендації, які винесені до захисту, одержані автором особисто. В спільних публікаціях автором зроблено наступне: [1, 2] – дослідження використання непараметричної класифікації виду розподілу для вибірок малого об'єму, [6] - застосування методів непараметричної класифікації для оптимізації кількості проб ґрунту в задачах агромоніторингу, [4, 7] – дослідження використання контрольних карт технологічних процесів при опрацюванні даних, [12] – застосування лінгвістичних шкал при вимірюваннях і діагностиці, [3] – дослідження використання вербальної шкали при розробці вимірювача сміття в підземній урні, [13] – дослідження способів урахування інструментальної невизначеності при формуванні результатів м'яких вимірювань, [14] - дослідження способів побудови шкали класифікації, за якої отримують ступені приналежності результатів вимірювання окремих ознак встановленим класам еквівалентності.

Апробація результатів дисертації. Основні положення та результати дослідження доповідалися та обговорювалися на: XI Міжнародній науково-технічній конференції «Гіротехнології, навігація, керування рухом і конструювання авіаційно-космічної техніки» (Київ, 13-14 квітня 2017), VI Міжнародній науково-технічній конференції «Метрологія, інформаційно-вимірювальні технології та системи», (Київ, 24-25 жовтня 2017), III Всеукраїнській науково-практичній конференції «Приладобудування та метрологія: сучасні проблеми, тенденції розвитку» (Київ, 11-12 жовтня 2018), II Міжнародній науково-практичній конференції «Мехатронні системи: інновації та інжиніринг» (Київ, 15 червня 2018), XV Міжнародному науково-технічному семінарі «Неопределенность измерений: научные, нормативные и методические аспекты» UM-2018, (Созополь, 10 сентября 2018), VII Міжнародній науково-технічній конференції «Метрологія, інформаційно-вимірювальні технології та системи», (Харків, 18-19 лютого 2020), XII Міжнародній науково-технічній конференції «Метрологія та вимірювальна техніка» та на семінарі «Невизначеність вимірювань: наукові, прикладні, нормативні та методичні аспекти» UM-2020, (Харків 7 жовтня 2020).

Публікації. Основні положення дисертаційної роботи опубліковано у 14 наукових працях, у тому числі 6 статей у наукових фахових виданнях (з них 1 стаття у періодичних наукових виданнях інших держав, які входять до ОЕСР та/або Європейського Союзу), 8 тез доповідей в збірниках матеріалів конференцій.

Структура та обсяг дисертаційної роботи. Дисертаційна робота складається з переліку умовних скорочень, вступу, 4 розділів, висновків, списку використаних

найменувань і 2 додатків. Загальний обсяг дисертації складає 149 сторінок друкованого тексту, у тому числі 137 основного тексту, робота містить 19 рисунків, 46 таблиць.

ОСНОВНИЙ ЗМІСТ РОБОТИ

У вступі обґрунтовано актуальність теми дисертації, мета і завдання дослідження, наукова новизна і практичне значення отриманих результатів роботи, зв'язок з науковими програмами і планами НДР, показано особистий внесок здобувача, представлено інформацію з апробації і публікації результатів дисертації, реалізація і впровадження основних положень роботи. Розкрито сутність і стан досліджуваної науково-технічної проблеми, наведено обґрунтування доцільності проведення досліджень, подана загальна характеристика роботи.

У першому розділі проведено аналіз наукових праць вчених М. Кендала, А. Стьюарта, С.А. Айвазяна, В.Я. Розенберга, що працювали в області теорії класифікації. За цими працями основними ознаками класифікації є: наявність шкали класифікації, що охоплює всю генеральну сукупність і яка розробляється заздалегідь або в процесі класифікації, виходячи з умови розрізнення; можливість формування навчальних вибірок; наявність інформації про класи еквівалентності. Характеристиками якості класифікації можуть бути ймовірності помилок першого і другого роду та похідні від цих характеристик. При складних випадках класифікації, коли рішення подається за кількома градаціями, це висновок за умовною ймовірністю віднесення до певного класу еквівалентності.

Статистичні феномени, з якими зустрічається дослідник в галузі класифікації, складні і не завжди передбачувані, тому дуже важливим є розробка методів і технологій для різних випадків застосування процедури класифікації.

Розроблено метод непараметричної класифікації розподілів вибірок числових даних. Операція класифікації дозволяє отримати додаткову інформацію про класи розподілу ймовірності вибірових даних, що може бути використано для удосконалення процедури опрацювання даних багаторазових вимірювань і при побудові контрольних карт технологічних процесів.

Для дослідження і удосконалення було обрано метод непараметричної ідентифікації, розроблений Ю. Клікушиним, який дозволяє вирішити задачу ідентифікації ряду розподілів (Коши, Лапласа, арксинусного, експоненціального, Сімпсона) за вибіркою малого об'єму (починаючи з 9). Ідентифікація заснована на визначенні місць нечітких оцінок (НО), що формуються за комбінаціями порядкових статистик з певними номерами. За порядком слідування НО формується код вибірки, який порівнюється з кодами зазначених вище розподілів. При повному співпадінні кодів отримують висновок про певний закон розподілу вибірки, при частковому співпадінні фіксується наближеність до певного розподілу.

Для розширення можливостей було удосконалено метод непараметричної ідентифікації при різному об'ємі вибірок, було сформовано алгоритм знаходження імен нечітких оцінок, тобто комбінацій порядкових статистик, які задіяні в формулах обчислення НО.

При дослідженні метода непараметричної ідентифікації при різних розподілах вибірок з метою визначення класифікаційних ознак були використані «навчальні» (еталонні) вибірки, що були «входом» задачі класифікації. На «виході» задачі повинен бути клас або перелік об'єктів, що створює цей клас. Дослідження було проведено з метою визначення «виходу» класифікації (тобто коду розподілу вибірки) за «входом» - вибіркою $\zeta = \{\zeta_1, \zeta_2, \dots, \zeta_n\}$ об'ємом n , що отримана моделюванням за формулою теоретичної функції

розподілу ймовірності, і складалося з наступних етапів: формування «входу» класифікації, а саме навчальної вибірки ζ об'єму n , що відповідає заданому теоретичному розподілу ймовірностей $F(x)$; опрацювання навчальної вибірки за алгоритмом непараметричної ідентифікації; отримання «виходу» - коду розподілу вибірки, що відповідає теоретичному $F(x)$.

Навчальні вибірки було отримано за зворотним перетворенням від функції розподілу ймовірності $\zeta = F'(\alpha)$, де α є послідовністю дискретних значень $\alpha = \{1/(n+1), 2/(n+1), \dots, n/(n+1)\}$ за умов зростання функції щільності ймовірності. У разі монотонного спадання функції щільності ймовірності $\zeta = F^{-1}(1-\alpha)$. Для прикладу двостороннього розподілу Лапласа з параметром ζ була отримана навчальна вибірка $\zeta = \frac{1}{\lambda} \{-1,61; -0,92; -0,51; -0,22; 0; 0,22; 0,51; 0,92; 1,61\}$. З використанням навчальної вибірки були обчислені нечіткі оцінки HO_i ($i=1..6$) у відповідності з ім'ям, до якого входять номери задіяних порядкових статистик.

Кодування НО відбувалось за правилами: якщо $HO_i < med(NO)$, тоді кодівий номер 1; якщо $HO_i > med(NO)$, тоді кодівий номер 0. «Вихід» класифікації – це код 101010 при збереженні порядку НО. Аналогічне моделювання було проведено для розподілів Коши, Гаусса, Сімпсона, арксинусного, рівномірного, одностороннього Лапласа та експоненціального.

Проведений аналіз на навчальних вибірках показав, що в дійсності непараметрична ідентифікація дозволяє розрізнити розподіли з різних класів еквівалентності. Результати моделювання і остаточний розподіл на класи наведено в табл.1. Вибір ознак, за якими проведено класифікацію, наступний: симетричність розподілу і кількість максимумів кривої щільності ймовірності, що відповідають модам розподілу. З аналізу загальної матриці «ознаки» - «класи» видно, що вектор ознак у кожного з класів різний, тобто класи не перерізаються. Визначено правило класифікації: якщо вибіркового код співпадає з одним з кодів табл.1, то розподіл вибірки належить відповідному класу. Якщо такого співпадіння немає, класифікація неможлива.

Таблиця 1 - Результати моделювання з визначення класів еквівалентності

Найменування розподілу	Коди розподілів за ідентифікацією	Клас еквівалентності за класифікацією
Коши	101010	Одноmodalний симетричний
Нормальний	101010	
Двосторонній Лапласа	101010	
Сімпсона	101010	
Арксинусний	010101	Антимодальний
Рівномірний	110001	Розподіл типу Шапо (безmodalний)
Односторонній Лапласа	110010	Правосторонній
Експоненціальний	001110	Лівосторонній

Аналіз кодів, що відповідають певним класам розподілів, показав, що неправильна класифікація (відкидання гіпотези про певний розподіл вибірки) можлива за умов зміни 0 в 1 (чи навпаки) третьої і четвертої позиції в коді вибірки. Для характеристики достовірності класифікації обрано ймовірність правильного віднесення до певного класу еквівалентності P_{np} . Була проведена теоретична оцінка достовірності, заснована на статистичних змінах

порядкових статистик, що підлягають β -розподілу з дисперсією, яка визначається за об'ємом вибірки і номером порядкової статистики, була обчислена дисперсія НО за лінійною комбінацією порядкових статистик. Для оцінки ймовірності правильної класифікації визначена критична відстань, тобто відстань між НО, найближчими до медіани, і $med(NO)$. Тоді ймовірність зміни лінгвістичного коду, або ймовірність неправильної класифікації визначається як: $P = P_{11} + P_{12} + P_{13}$, де P_{11} – ймовірність одночасного знаходження обох НО в критичній області, P_{12} та P_{13} – добуток ймовірності знаходження однієї оцінки в критичній області на умовну ймовірність, що одна оцінка менша за другу.

За результатами обчислення ймовірності правильної класифікації за однією вибіркою наступні: нормальний розподіл – 0,8; Коши – 0,78; рівномірний – 0,76; Лапласа – 0,82; арксинусний – 0,84.

Експериментальна оцінка достовірності була отримана за моделюванням процедури класифікації з використанням 200 вибірок з нормальним розподілом. Ймовірність правильної класифікації за моделюванням становила 0,79 (158+, 42-), що збігається з результатами, наведеними вище (для $n = 200$ довірчий інтервал для ймовірності $\pm 0,054$).

Для зменшення дисперсії НО можна проводити непараметричну класифікацію за декількома вибірками малого об'єму. Тому в роботі проведено визначення необхідної кількості вибірок за заданою ймовірністю правильної класифікації. Наведено рекомендації для необхідної кількості вибірок для різних розподілів за заданою достовірністю класифікації ($P = 0,95; 0,99$). Розроблено алгоритм проведення класифікації за декількома вибірками.

В розділі 2 розглянуто можливості удосконалення методів опрацювання багаторазових вимірювань за допомогою попередньої непараметричної класифікації розподілів вибірок. В першу чергу це стосується формування результату багаторазових вимірювань, який за VIM в загальному випадку включає інформацію, що має відношення до нього (relevant information), що може бути як інформацією про певний розподіл, так і про клас розподілу. Крім того бажано, щоб результат багаторазового вимірювання був ефективною статистичною оцінкою. Тому в роботі запропоновано удосконалити алгоритм опрацювання багаторазових вимірювань шляхом введення попередньої непараметричної класифікації розподілу вибірки. Це продемонстровано за допомогою моделювання процедури отримання результату вимірювання за декількома вибірками з генеральної сукупності випадкових чисел з рівномірним законом розподілу, де за результатами попередньої класифікації отримано висновок – безмодальний. Ефективна оцінка центру розподілу – це середина розмаху, а оцінки стандартної і розширеної невизначеності зменшуються у порівнянні з ситуацією, коли розподіл невідомий і його приймають нормальним. Для вибірок більшого об'єму (наприклад 25) для класифікації розподілу можна обмежитись однією вибіркою.

Таким чином, застосування метода непараметричної класифікації дозволяє забезпечити результат вимірювання релевантною інформацією про клас розподілу; обрати ефективну оцінку результату вимірювання і тим самим зменшити і стандартну і розширену невизначеність. Попередня класифікація розподілів вибірки може бути використана для удосконалення алгоритму опрацювання результатів вимірювання при калібруванні при наявності декількох вимірювань в одній точці калібрувальної характеристики і наявності декількох точок. Класифікація емпіричних розподілів може бути першим етапом при використанні узагальнених розподілів (наприклад, gh – розподілу Тьюки, в якому при $g = 0$ мають сім'ю симетричних розподілів).

Розроблено технології застосування метода непараметричної класифікації при побудові КК. Якщо для контролю якості процесу визначено кількісні характеристики, то для удосконалення вибору статистичних характеристик може бути застосована непараметрична класифікація сумарного розподілу вибірок, отриманих для окремих точок, за якими будується КК. Зазвичай цей розподіл вважають нормальним з вибором центральної тенденції за середнім арифметичним. Так як на карті статистичного контролю позначають час або номер вибірки, а на вісі ординат значення контрольованого параметра, центральну тенденцію (за однією чи декількома вибірками) або за середнім арифметичним, або за медіаною вибірки, або за серединою розмаху, то нечітка непараметрична класифікація дозволяє обрати найбільш ефективну.

Алгоритм побудови КК було розроблено для визначення стабільності технологічних процесів. Проведено порівняння двох алгоритмів опрацювання даних для побудови КК: перший орієнтований на нормальний розподіл вибірових даних; другий включає процедуру попередньої класифікації розподілів і вибір ефективної оцінки для центральної тенденції. Порівняння проводилось за умов генерації декількох вибірок рівномірно розподілених чисел з центральним значенням 10,0. В першому алгоритмі центральна тенденція визначалась за середнім арифметичним, в другому – за серединою розмаху (табл.2).

Таблиця 2 - Результати опрацювання даних

Номер алгоритму	Центральна тенденція	СКВ центральної тенденції	Попереджувальні границі	Контрольні границі
1	10,02	0,04	9,94; 10,10	9,90; 10,14
2	10,01	0,023	9,96; 10,06	9,94; 10,08

За результатами моделювання видно, що попередня класифікація сумарного розподілу дозволяє обрати ефективну для даного класу розподілу оцінку центру вибірки, завдяки чому зменшується дисперсія і СКВ за даною вибіркою. Тому попереджувальні і контрольні межі стають вузькими, що забезпечує більшу чутливість саме до змін параметрів технологічного процесу.

Якщо ж зміни відбуваються в контрольних межах, то неможливо вирішити чим вони обумовлені: нестабільністю технологічного процесу чи обмеженістю статистичних даних. Рішення про клас розподілу декількох вибірок завдяки непараметричному оцінюванню приймається за певним порядком слідування нечітких оцінок і не залежить від зміни центральної тенденції. При зміні центральної тенденції всі порядкові статистики зміщуються в один бік і їх порядок слідування не змінюється. Завдяки цьому непараметричну класифікацію можна застосовувати при нестабільності технологічного процесу.

Розділ 3 дисертаційної роботи присвячений технології застосування класифікації при переході від числових даних до вербальних в ІВС. Результати вимірювання, які отримують на виході вимірювальних каналів системи, перетворюють в форму, що може бути застосована надалі при нечітких обчисленнях, прийнятті рішень, отриманні діагнозу про стан об'єктів, а також при необхідності регулювання параметрів об'єктів. Для отримання кінцевого результату використовується метрична класифікація, за якої на метричному носіїв будується шкала з нечіткою лінгвістичною змінною, що дозволяє визначити клас еквівалентності якому відповідає результат вимірювання. Автором запропонована наступна послідовність етапів встановлення або відтворення шкали класифікації з нечіткою лінгвістичною змінною: орієнтовне визначення кількості термів

множини ЛЗ і їх границь; аналіз нечіткості семантичного правила і складових невизначеності вимірювання; вибір загального критерію, що дозволяє об'єднати складові нечіткості і невизначеності; вибір алгоритму об'єднання складових нечіткості і невизначеності; перевірка максимальної кількості термів множини ЛЗ за умовою розрізненості; вибір функцій приналежності (ФП) шкали з нечіткою ЛЗ; апробація (при наявності еталонних елементів) або моделювання нечіткого виведення з застосуванням встановленої лінгвістичної шкали.

В роботі проведено аналіз наведених вище етапів. Особливу увагу приділено аналізу чинників, які впливають на розмитість ФП окремих термів лінгвістичної змінної і які повинні враховуватись при побудові лінгвістичної шкали (рис.1).

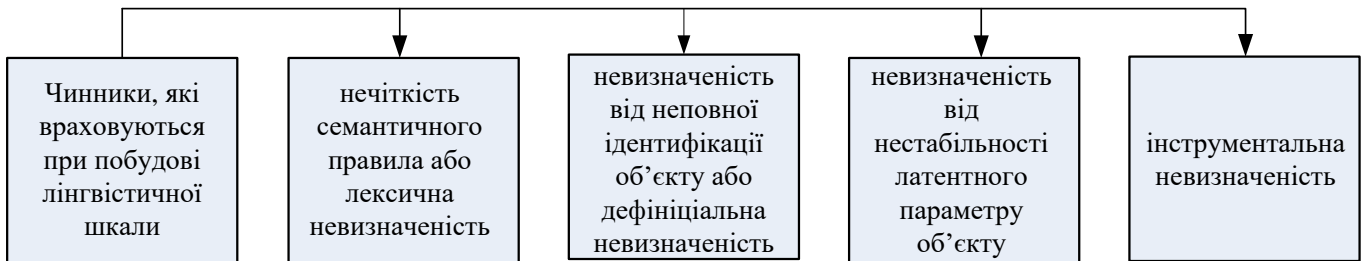


Рисунок 1 – Перелік чинників, що впливають на побудову лінгвістичної шкали

Визначено, що спосіб об'єднання чинників залежить від алгоритму роботи нечіткого класифікатора, за яким результат вимірювання подають як нечітке число (fuzzy number), або чітке число (singleton). Останній вид алгоритму використовується в нечіткому додатку до пакету MatLab, за яким визначають саме активовані ФП класів еквівалентності. Ці ФП модифікують у відповідності з максимумом і мінімумом перерізу і за центром ваги визначають клас еквівалентності, якому відповідає результат вимірювання. Тому автором роботи був обраний спосіб об'єднання чинників, що характеризують нечіткість і невизначеність.

Для розробки підходів і рекомендацій до способів об'єднання чинників необхідно було визначити критерій з дотриманням вимог обумовленості, практичності і реалістичності. Цим вимогам відповідає індекс нечіткості, що є мірою нечіткості множини і визначається як відстань цієї множини до найближчої до неї звичайної множини. Ця метрика може бути застосована і для дискретної і для неперервної множини, тобто при всіх способах подання нечіткості семантичного правила.

Лінійний індекс нечіткості неперервної множини A з $\mu_A^{(1)}(x)$ визначається як

$$I_A^L = \frac{2}{b-a} \int_a^b |\mu_A^{(1)}(x) - \mu_{\underline{A}}^{(2)}(x)| dx, \quad (1)$$

де $\mu_{\underline{A}}^{(2)}(x)$ – звичайна множина, найближча до нечіткої, а a і b – нижня і верхня границі носія нечіткої множини A відповідно.

Дослідження зв'язку індексу нечіткості з частиною носія, що характеризує нечіткість семантичного правила було проведено для різних форм ФП. Для прикладу обрано клас еквівалентності «малий», тобто початковий терм лінгвістичної шкали з ФП

$$\mu_{T_1}(x) = \begin{cases} 1, & \text{якщо } 0 \leq x \leq a_1; \\ \frac{a_2 - x}{a_2 - a_1}, & \text{якщо } a_1 \leq x \leq a_2; \\ 0, & \text{якщо } a_2 < x, \end{cases}$$

де $a=0$; $b=a_2$; $0 \leq x \leq a_1$ – область толерантності, де $\mu_T(x)=1$; $a_1 \leq x \leq a_2$ – область, яка характеризує нечіткість семантичного правила; значення $(a_1+a_2)/2$ відповідає чіткій верхній границі класу еквівалентності «малий»; $(a_2-a_1)/2$ – область розмитості границі.

Якщо ввести позначення $\Delta=(a_2-a_1)/2$ і $\delta=2\Delta/(a_1+a_2)$, δ – відносна розмитість границі, тоді $I_A^L = \delta/(1+\delta)$.

Для індексу нечіткості, заснованого на евклідовій відстані

$$I_A^E = \frac{2}{b-a} \sqrt{\int_a^b (\mu_A^{(1)}(x) - \mu_A^{(2)}(x))^2 dx} = \sqrt{\frac{2\delta}{3 \cdot (1+\delta)}} \quad (2)$$

Виходячи з (1) і (2) при об'єднанні чинників, що впливають на індекс нечіткості, треба визначити загальне значення області розмитості Δ , як носія нечіткого числа, симетричного відносно границі терму або класу еквівалентності $(a_1+a_2)/2$. Подібна задача вирішується з застосуванням теорії можливостей і теорії нечітких множин при об'єднанні результатів вимірювання, отриманих за допомогою «фізичних» систем і результатів оцінювання з «суб'єктивною» неточністю. В цих роботах визначають вкладений інтервал (nested interval) за заданим рівнем довіри.

Аналіз складових нечіткості і невизначеності був застосований у процедурі встановлення лінгвістичної шкали артеріального тиску при визначенні з допомогою ІВС допустимого фізичного навантаження людини, що характеризує стан її здоров'я.

При встановленні шкали артеріального тиску терм-множина лінгвістичної змінної визначається за експертними даними. Розбіжність даних, отриманих з різних джерел для певного віку і статі особи, ідентифікується як нечіткість семантичного правила. Наприклад, для терму «нормальний» вона складала ± 3 мм рт. ст.. Для повної ідентифікації об'єкта треба було розробити шкали з урахуванням статі і віку (по 3 роки в кожній групі). Це збільшує об'єм бази правил і, відповідно, кількість антецедентів в ній. Тому необхідно знайти компроміс між критеріями точності при побудові шкали і критеріями компактності бази правил. Виходячи з чинників, що впливають на індекс нечіткості, а саме інструментальної невизначеності ± 3 мм рт. ст. і нестабільності латентного параметру ± 5 мм рт. ст. (від зміни тиску протягом доби), залишкову складову невизначеності від неповної ідентифікації об'єкту було встановлено на рівні $\pm 2,5$ мм рт. ст. з організацією шкал для трьох вікових груп: «молодий», «середній», «похилий». У відповідності з рекомендаціями було визначено загальний покривний інтервал ± 10 мм рт. ст. для довірчої ймовірності $P=0,95$.

Після визначення загального СКВ чинників, що впливають на розмитість терм-множини ЛЗ, необхідно перевірити правильність вибору кількості термів за умов їх розрізнення. Визначення максимальної кількості термів або градацій лінгвістичної шкали N_{\max} запропоновано проводити за інформаційним критерієм, як $N_{\max} = (x_2 - x_1)/d$, де x_2 , x_1 – максимальне і мінімальне значення за областю визначення лінгвістичної шкали; d – ентропійний інтервал невизначеності. Якщо кількість термів множини ЛЗ $N \leq N_{\max}$, можна переходити до вибору форми ФП окремих термів.

Залежності між індексом нечіткості і інтервалом невизначеності (1) та (2) були отримані для лінійних форм ФП. Подальші дослідження показали, що використання нелінійних форм ФП мало впливає на значення індексу нечіткості, але при цьому порушується правило розділення одиниці, що справедливо при лінійних ФП. Для аналізу впливу форми ФП на рішення, що приймається при класифікації було проведено моделювання, результати якого наведені в табл.3.

З таблиці видно, що тільки для трапецієподібної ФП результати класифікації відповідають умові розділення одиниці. Тому другий рядок результатів класифікації для увігнутої та опуклої ФП отримано з перетворенням для виконання умови розділення одиниці.

Таблиця 3 - Результати класифікації за ступенем приналежності класам «малий» і «середній»

Форма ФП	Результати вимірювань		
	$x_1 = a_1 + 0,25 \cdot (a_2 - a_1)$	$x_2 = a_1 + 0,5 \cdot (a_2 - a_1)$	$x_3 = a_1 + 0,75 \cdot (a_2 - a_1)$
Трапецієподібна	$T_1 0,75; T_2 0,25$	$T_1 0,5; T_2 0,5$	$T_1 0,25; T_2 0,75$
Увігнута	$T_1 0,5; T_2 0,06$	$T_1 0,3; T_2 0,25$	$T_1 0,13; T_2 0,56$
	$T_1 0,89; T_2 0,11$	$T_1 0,55; T_2 0,45$	$T_1 0,19; T_2 0,81$
Випукла	$T_1 0,94; T_2 0,5$	$T_1 0,75; T_2 0,7$	$T_1 0,44; T_2 0,87$
	$T_1 0,65; T_2 0,35$	$T_1 0,52; T_2 0,48$	$T_1 0,34; T_2 0,66$

За всіма формами ФП результат віднесення до класу еквівалентності за максимумом перерізу був однаковим. Якщо в області розмежування двох класів еквівалентності застосувати дві форми ФП – випуклу та увігнуту, тоді пріоритет буде у класу з випуклою ФП (в цьому випадку можна заздалегідь передбачити пріоритетне віднесення до вищого або нижчого за порядком класу еквівалентності).

Як характеристику якості встановленої шкали класифікації запропоновано використовувати матрицю відповідності шкали, рядки якої відповідають носіям окремих термів або градаціям шкали, а стовпчики – класам еквівалентності, до яких віднесено розмір величини за класифікацією. Якщо позначити терм-множину T_i ($i=1, \dots, N$), де N - кількість класів еквівалентності, що відповідає вимірюваній величині, а T_j^i ($j=1, \dots, N$) - множину результатів класифікації, то на перерізі рядків і стовпчиків отримаємо ймовірність P_{ij} віднесення розміру величини з i -того терму до j -го класу еквівалентності. При цьому, якщо $i = j$, це відповідає ймовірності правильної класифікації, значення якої розташовані в діагональних елементах матриці відповідності.

За матрицею відповідності може бути визначений функціонал, що відповідає інтуїтивно поняттю «якості класифікації». Таким функціоналом або загальним показником якості встановленої шкали обрана норма Фробеніуса: $G = \left(\sum_{i=1}^N \sum_{j=1}^N (P_{ij} - I_{ij})^2 \right) / 2N$, де I_{ij} - ідеальна матриця відповідності. Показник якості знаходиться в межах $0 \leq G \leq 1$, при цьому 0 відповідає правильній класифікації.

Матриця відповідності характеризує ситуацію, коли апріорної інформації немає і ймовірність знаходження вимірюваної величини однакова за усіма класами еквівалентності. При наявності результату вимірювання з інформацією про його класифікацією за певними класами еквівалентності результат класифікації може бути скоригований за формулою Байеса. Необхідність корекції визначається за нормою Фробеніуса.

Для аналізу вибірок класифікованих даних важлива оцінка їх розсіювання відносно центральної тенденції, тому що прийняття рішення за вибіркою з меншим розсіюванням більш інформативне ніж тоді, коли розсіювання велике.

Недоліком відомих інформаційних критеріїв як міри узгодженості між ординальними даними є відсутність врахування розподілу вибірки або порядку групування

членів вибірки. Тому за ранговою арифметизацією було запропоновано наступний критерій:

$$Cns(x) = 1 + \sum_{i=1}^N p_i \log_2 \left(1 - \frac{|x_i - \mu_x|}{d_x} \right)$$

де x_i - ранг класу еквівалентності, p_i - ймовірність або частота, поєднана з x_i , d_x - область визначення, що дорівнює $x_{\max} - x_{\min}$, μ_x - середнє значення $\mu_x = \sum_{i=1}^N p_i x_i$.

Недоліком наведеного способу обчислення міри узгодженості є застосування операцій з рангами, які проявляються тільки у відношенні порядку та визначення центральної тенденції, як середнього арифметичного рангів. Запропоновано визначати центральну тенденцію за операторами *med* і *OWA*. Якщо визначено клас еквівалентності T_u , що відповідає центру вибірки, тоді з застосуванням встановленої шкали класифікації може бути отримане значення $x(T_u)$, що відповідає центру (тобто середині) класу еквівалентності T_u . Запропоновано наступне визначення міри узгодження:

$$Cns(x) = 1 + \sum_{i=1}^N p_i \log_2 \left(1 - \frac{|x(T_i) - x(T_u)|}{x_{\max} - x_{\min}} \right) \quad (3)$$

де $x(T_i)$ - середина класу еквівалентності T_i за областю визначення шкали класифікації.

Такий метод оцінювання міри узгодження дозволяє її застосовувати при нерівномірному розміщенні класів еквівалентності. При рівномірному розташуванні класів еквівалентності можна отримати наслідок з формули (3), як

$$Cns(x) = 1 + \sum_{i=1}^N p_i \log_2 \left(1 - (n(T_i | T_u)) / N \right) \quad (4)$$

де $n(T_i | T_u)$ - кількість класів еквівалентності між T_i і T_u .

Результати обчислення $Dnt(x) = 1 - Cns(x)$ за формулою (4) приблизно збігаються з ординальною дисперсією, отриманою для рівномірного розташування класів еквівалентності

$$D = \left(\sum_{k=1}^N F_k (1 - F_k) \right) / ((N - 1) / 4) \quad (5)$$

де F_k - кумулятивна відносна частота для k -го класу еквівалентності.

Практичне застосування отриманих результатів теоретичних досліджень при вирішенні таких задач як тривалий моніторинг земельних ресурсів, поліпшення якості ґрунтів при внесенні добрив, визначення придатності земель при оптимізації різних способів використання, тощо, отримане з визначення терм-множини ЛЗ в системі класифікації ґрунтів за вмістом рухомого фосфору для оптимізації норми внесення добрив. Виділено наступні етапи: отримання проб ґрунту за детально визначеною програмою; визначення показників родючості ґрунтів за стандартизованими методиками; прийняття рішень щодо стану якості ґрунту або способів поліпшення цього стану за внесенням необхідної кількості добрив. Шкала класифікації дозволяє представити дані, що є певними рівнями ЛЗ «показник якості ґрунту» (табл.4). Групування ґрунтів за показником якості «вміст рухомого фосфору в пробах ґрунту» (ВРФ) згідно ДСТУ 4114-2002 засноване на 5 рівнях або класах еквівалентності.

Виходячи з точності вимірювання рухомого фосфору за методом Мачигіна (довірчий інтервал при $P = 0,95$ складає 30% при вмісті P_2O_5 до 15 мг/кг і 20% при вмісті 15 мг/кг) за інформаційним критерієм було отримано $N_{\max} = 5$. Терм-множина ЛЗ «рівень

вмісту рухомого фосфору в пробах ґрунту» побудована за лінійними ФП (рис.3). Форма ФП окремих термів змінюється від трапецієподібної до трикутної в залежності від обчисленої невизначеності вимірювання. Якщо ентропійний інтервал невизначеності займає практично весь носій ФП терма, форма ФП відповідає трикутній.

Таблиця 4 - Матриця відповідності шкали класифікації показника якості ґрунту

Результат класифікації \ Вміст рухомого фосфору (мг/кг)	Низький	Середній	Підвищений	Високий	Дуже високий
0 – 15,5	0,952	0,048	0	0	0
15,5 – 30,5	0,050	0,850	0,1	0	0
30,5 – 45,5	0	0,100	0,775	0,125	0
45,5 – 60,5	0	0	0,125	0,75	0,125
60,5 – 75	0	0	0	0,129	0,871

Враховуючи переріз ФП сусідніх термів побудовано матрицю відповідності встановленої шкали, де в комірках наведено значення, що є ймовірностями віднесення до певних класів еквівалентності P_{ij} . Наприклад, для першого класу еквівалентності

$$P_{ij} = \int_0^{x_{zp}} \mu_{T_1}(x) dx / \left(\int_0^{x_{max}} \mu_{T_1}(x) dx \right),$$
 де $x_{zp} = x_{kp}$, x_{zp} - встановлена чітка границя класу еквівалентності, що є одночасно критичною точкою x_{kp} ; x_{max} - максимальна границя з урахуванням невизначеності вимірювання. Матрицю відповідності шкали класифікації рис.2 наведено в табл.7.

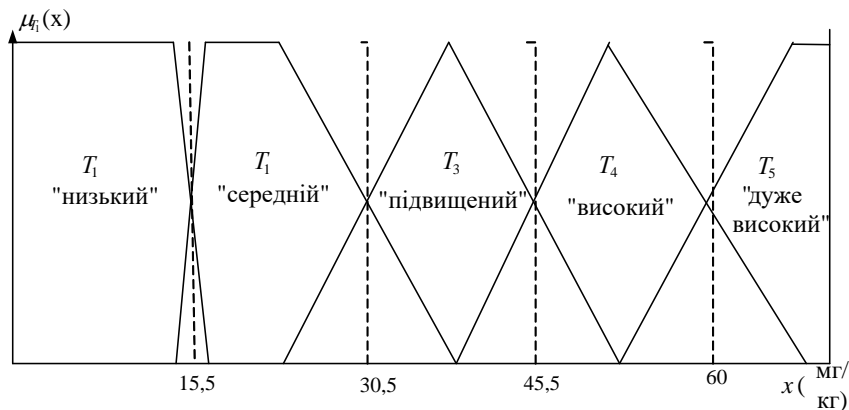


Рисунок 2 – Шкала класифікації показника якості ґрунту «рівень вмісту рухомого фосфору в пробах ґрунту»

За даними табл.7 було отримано загальний показник якості встановленої шкали класифікації – норму Фробеніуса, що дорівнює $G = 0,024$.

Якщо результат класифікації отримують за одноразовими вимірюваннями, він може бути скоригований з використанням матриці відповідності і формули Байеса. Так, якщо результат класифікації за перерізом результату вимірювання зі шкалою класифікації (рис.2) становить $T_1 | 0,7; T_2 | 0,3$ тоді з розрахунком апостеріорної ймовірності для класу еквівалентності «низький» отримуємо

$$P(T_1 | A) = \frac{P_{11} \cdot P(A | T_1) + P_{12} \cdot P(A | T_2)}{P_{11} \cdot P(A | T_1) + P_{12} \cdot P(A | T_2) + P_{22} \cdot P(A | T_2) + P_{21} \cdot P(A | T_1)} = 0,701$$

звідки $P(T_2 | A) = 0,299$. Таким чином, якщо значення норми Фробеніуса має результати класифікації не потребують корекції. Але якщо розсіяння терм-множини збільшується, це спричиняє зменшення діагональних ймовірностей, що як наслідок приводить до збільшення норми Фробеніуса. Тоді результати класифікації потребують корекції. Так, якщо $P_{11} = 0,8$ з подальшими змінами $G = 0,032$, тоді за шкалою класифікації при попередньому значенні вимірюваної величини отримуємо $T_1 | 0,8; T_2 | 0,2$. При проведенні корекції за формулою Байеса отримуємо $T_1 | 0,74; T_2 | 0,26$.

Отримана матриця відповідності може бути використана при заданих границях термів множини ЛЗ як показник якості класифікації, а також для оптимізації вибору границь, тому що, як видно з табл.4, в центральній частині області визначення шкали класифікації ймовірність правильного віднесення до певного класу еквівалентності зменшується.

Результати нечіткої класифікації використовуються для прийняття рішення про внесення добрив за правилами: якщо вміст рухомого фосфору x , то норма внесення амофосу y . Для класифікації вмісту фосфору нечіткий класифікатор визначає переріз ординати, що відповідає результату вимірювання, з терм-множиною лінгвістичної змінної. Якщо переріз відповідає такому значенню $\mu_{T_1}(x)$, яке дорівнює одиниці, використовується відповідне правило з системи правил. Якщо переріз включає два терма множини, наприклад $T_1 | 0,6; T_2 | 0,4$, тоді норма внесення добрив H розраховується за двома правилами з ваговими коефіцієнтами, а саме $H = 90 \text{ кг} / \text{га} \cdot 0,6 + 71 \text{ кг} / \text{га} \cdot 0,4 = 82,4 \text{ кг} / \text{га}$.

При опрацюванні багаторазових вимірювань і прийнятті рішення за вибіркою виникає необхідність перевірки даних на узгодженість. Для цього може бути використана запропонована в роботі міра узгодженості класифікованих даних (формула (3)). Ця міра може бути використана для перевірки методики рандомізованого відбору проб ґрунту в різних точках контрольованої ділянки. В роботі проведено моделювання процедури перевірки узгодженості вибірок з різним розсіюванням. Так як застосування безводного аміаку можливе лише шляхом його локального внесення у стрічки на певну глибину, зміни, що відбуваються у стрічках його внесення можуть бути доволі контрастними порівняно з показниками ґрунту у міжрядді внаслідок формування осередків з високою концентрацією солей. Тому розробляються методики спеціального відбору проб, наприклад за схемою зигзагоподібного перетинання, що забезпечують однакову кількість проб у стрічці і у міжрядді. Як критерій вибору оптимальної програми відбору проб може бути використана запропонована міра узгодженості, яка заснована на використанні шкали класифікації. Так як невизначеність вимірювання може бути причиною розсіювання класифікованих даних, в роботі визначено граничне значення міри узгодженості.

В дисертації наведено результати розробки інтелектуальної системи для визначення стану заповнення сміттям підземних урн, в якій було використано метричну класифікацію. Наведено структурну схему системи, за якої безпосередньому вимірюванню підлягає відстань від верхньої точки урни до рівня насипу сміття. Максимальна висота урни становила 2400 мм. При побудові шкали класифікації на основі інформаційного критерію була встановлена кількість класів еквівалентності і встановлено 5 термів ЛЗ «стан заповнення урни»: T_1 - урна практично порожня ($0 \div 300 \text{ мм}$); T_2 - урна заповнена менше ніж наполовину ($300 \div 900 \text{ мм}$); T_3 - урна заповнена наполовину ($900 \div 1500 \text{ мм}$); T_4 - урна заповнена більше ніж наполовину ($1500 \div 2100 \text{ мм}$); T_5 - урна практично заповнена ($2100 \div 2400 \text{ мм}$).

Формування ФП окремих термів відбувалось з урахуванням сумарної невизначеності вимірювання відстані, складовими якої були похибка далекоміра, складові від впливу на результати вимірювання щільності та різнорідності сміття, кольору сміття (сірий, білий), форми насипу (наявність «горбика») та інших впливних величин. Загальна невизначеність вимірювання становила $\pm 14\%$. З урахуванням невизначеності встановлено шкалу класифікації і отримано характеристику якості шкали – матрицю відповідності. Розраховано норму Фробеніуса $G = 0,023$. Розглянуто процедуру одноразового і багаторазового вимірювання за різними алгоритмами роботи нечіткого класифікатора.

Четвертий розділ дисертації присвячений розробці технологій застосування класифікації при опрацюванні вербальних даних. Було визначено оператори, які можуть бути використані саме при роботі з вербальними даними (без призначення рангів або арифметизації). Якщо об'єктом дослідження є вибірка вербальних даних, то для визначення центру вибірки або її центральної тенденції можуть бути використані оператори, що побудовані на непараметричних статистиках або на нечіткій логіці. При непарній кількості членів вибірки може бути використана медіана вибірки med , і незалежно від кількості членів вибірки - оператор OWA , що може виступати як емулятор середнього арифметичного.

$OWA = \max_{k=1}^n [\min\{Q(k), b_k\}]$, де $Q(k) = S(f_k)$, $k = 1, 2, \dots, n$, n - об'єм вибірки; $S(f_k)$ - тий рівень лінгвістичної шкали;

$f_k = Int \left\{ 1 + \left[k \frac{t-1}{n} \right] \right\}$, f_k - найближче ціле число в формулі; t - кількість градацій вербальної шкали; b_k - k -тий елемент вибірки, попередньо ранжованої за порядком спадання градацій вербальної шкали.

Відношення домінування може бути розповсюджене на вибірки даних, отриманих з застосуванням вербальних шкал квазіпорядку, для яких встановлені відношення порядку між класами еквівалентності.

На основі використання перелічених вище операторів в роботі запропоновано наступні технології застосування процедури класифікації за центральною тенденцією вербальної вибірки при визначенні стану об'єкту і при побудові контрольних карт технологічних процесів. При перевірці збіжності оцінок центральної тенденції вибірки за операторами med і OWA (табл.5), показано, що результати визначення центральної тенденції за медіаною і оператором OWA не співпадають. Це пояснюється тим, що медіана нечутлива до змін крайніх членів впорядкованої вибірки, а оцінка OWA навпаки – чутлива. Крім того, внаслідок ступінчастості оператора OWA збільшується кількість вибірок, класифікованих за середнім класом еквівалентності (табл.5). Так як обидві оцінки не є універсальними, то для підвищення точності запропоновано використати дві оцінки: за медіаною і за оператором OWA .

В тих випадках, коли результати визначення центральної тенденції не співпадають, а знаходяться в сусідніх класах еквівалентності, запропоновано використовувати проміжні класи еквівалентності (для табл.5 це напівгрубий НГ, напівточний НТ). Таким чином розроблено метод опрацювання вербальних вибірок малого об'єму з метою класифікації стану об'єкту з обчисленням двох оцінок центральної тенденції, за якими при їх нееквівалентності визначають додаткові (проміжні) класи еквівалентності стану об'єкту.

Таблиця 5 - Способи оцінювання центральної тенденції вербальних вибірок, отриманих при візуальному контролі рівня шорсткості оброблених поверхонь, де Г – грубий, С – середній, Т - точний

Досліджувані вибірки	<i>med</i>	<i>OWA</i>	<i>med i OWA</i>	Досліджувані вибірки	<i>med</i>	<i>OWA</i>	<i>med i OWA</i>
ГГГГГ	Г	Г	Г	ТСССГ	С	С	С
СГГГГ	Г	Г	Г	ТСССС	С	С	С
ТГГГГ	Г	Г	Г	ТТССГ	С	С	С
ССГГГ	Г	С	НГ	ТТССС	С	С	С
ТСГГГ	Г	С	НГ	ТТТГГ	Т	С	НТ
ТТГГГ	Г	С	НГ	ТТТСГ	Т	С	НТ
СССГГ	С	С	С	ТТТСС	Т	С	НТ
ССССГ	С	С	С	ТТТТГ	Т	Т	Т
ССССС	С	С	С	ТТТТС	Т	Т	Т
ТССГГ	С	С	С	ТТТТТ	Т	Т	Т
ТТСГГ	С	С	С				

При більших об'ємах вербальних вибірок ($n \geq 10$) збіжність оцінок центральної тенденції може бути перевірена за непараметричними критеріями.

В табл. 6 наведені результати класифікації стану окремих урн з позначеннями: МІД – заповнена; МПД – заповнена більше половини; ПД – наполовину заповнена; БПД – заповнена менше половини; МАД – пуста.

Центральна тенденція за *med* – це рівень МПД, за оператором *OWA* – це рівень ПД. Для визначення збіжності оцінок центральних тенденцій був використаний довірчий інтервал медіани з границями $\Delta_{rl} = x_u - med$, $\Delta_{rh} = x_v - med$, де x_u , x_v - порядкові статистики вибірки з номерами u , v ; u - ближнє ціле число, що менше від $(n+1 - z_p \sqrt{n})/2$; v - ближнє ціле число, що більше від $(n+1 + z_p \sqrt{n})/2$; z_p - квантиль нормального розподілу; для $P = 0,95$, $x_u = x_4 = \text{МПД}$; $x_v = x_{12} = \text{БПД}$.

Таблиця 6 - Стан заповнення окремих урн

№ урни	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Стан заповнення	МІД	МІД	МІД	МПД	МПД	МПД	МПД	МПД	МПД	ПД	БПД	БПД	БПД	МАД	МАД

Так як X_{OWA} знаходиться в межах МПД і БПД, то результати X_{OWA} і X_{med} збігаються. Рішення про виклик сміттєзбиральної техніки приймаємо, якщо X_{med} або X_{OWA} дорівнює або перевищує контрольний рівень, наприклад МПД.

Класифікацію поточних станів технологічних процесів при побудові контрольних карт проводять за вибірками невеликого об'єму, розділених певними проміжками часу. Як показали дослідження, для класифікації вибірок недостатньо одного критерію. Це наочно продемонстровано в табл. 7, де кількість рівнів $t=3$ з терм-множиною $T = \{\text{«низький»}, \text{«середній»}, \text{«високий»}\} = \{H, C, B\}$, об'єм вибірок $n=3$. Важливою характеристикою сукупності вибірок є загальна кількість N можливих сполучень в вибірках, де $N = C_t^n = (t+n-1)! / (n!(t-1)!)$ і кількість розрізнявальних класів еквівалентності $N_{екв}$ після класифікації. Значення N характеризує потенційну розрізняльну здатність при побудові КК, що збільшується при збільшенні t і n . Значення $N_{екв}$ характеризує реальну розрізняльну здатність, що дозволяє визначити на КК центральну тенденцію, попереджувальні і контрольні границі. В таблиці 7 $N = N_{екв}$.

Таблиця 7 - Класифікація поточних станів за вербальними вибірками

Ранжовані вибірки, що відповідають різним станам	ННН	СНН	ВНН	ССН	ВСН	ССС	ВСС	ВВН	ВВС	ВВВ
Номер класу за відношенням домінування вибірок	1	2	3		4		5		6	7
Номер класу за відношенням домінування, медіаною і розсіюванням	1	2	3	4	5	6	7	8	9	10

Але при збільшенні t і n навіть при використанні всіх операторів частина вибірок залишається нерозрізненою і кількість класів еквівалентності стає меншою за кількість можливих сполучень вербальних категорій, тобто $N_{екв} < N$. Велика кількість класів еквівалентності приводить до недоцільності побудови загальної шкали класифікації. Тому вербальні вибірки класифікуються в процесі побудови контрольної карти. При цьому послідовність застосування операторів зворотна по відношенню до табл.7, а саме:

- вибірки впорядковуються за спаданням властивості;
- вибірки розподіляються на класи еквівалентності за центральною тенденцією (використовується або медіана, або *OWA*, або обидві оцінки з визначенням проміжних класів еквівалентності);

вибірки в кожному класі еквівалентності ранжуються за відношенням домінування і розсіювання. Вибірki, що залишились нерозрізненими, належать до одного класу еквівалентності.

За отриманими класами еквівалентності може бути побудована контрольна карта. Результати класифікації 31 вибірки при $t = 3$; $n = 5$ наведено в табл.8.

Таблиця 8 - Класифікація вибірок для побудови контрольної карти

Класифікація за центром вибірки – медіаною	Класифікація за відношенням домінування і розсіюванням	Номер класу еквівалентності	Номери впорядкованих вибірок для застосування непараметричних оцінок
Клас Н	ССННН	1	1
Клас С	СССНН	2	2,3,4,5
	ССССН	3	6,7,8,9
	ССССС	4	10
	ВСССН		11,12,13,14,15
	ВВССН	5	16,17,18,19,20
Клас В	ВВВНН	6	21
	ВВВСН		22,23,24
	ВВВСС	7	25,26,27,28
	ВВВВС	8	29,30,31

Результати розрахунків за даними таблиці 8 для побудови контрольної карти:
 $med = x'_{16} = (ВВССН)$; ; $НПГ = x'_{10} = (ССССС)$; ; $ВПГ = x'_{22} = (ВВВСН)$; ; $НКГ = x'_{7} = (ССССН)$;
 $ВКГ = x'_{25} = (ВВВСС)$.

Вибірki на рис.3 розташовують за лініями, які відповідають 8-ми класам еквівалентності. Якщо класів еквівалентності недостатньо для розрізнення центральної тенденції, попереджувальних та контрольних границь необхідно збільшити або кількість розрізнювальних градацій t або об'єм вибірки n .

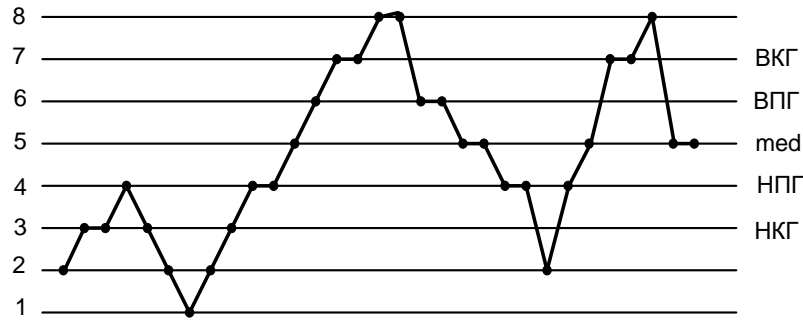


Рисунок 3 – Контрольна карта за даними таблиці 8

Таким чином, розроблено технології встановлення відношення порядку між вербальними вибірками малого об'єму для побудови шкали і метод класифікації вербальних вибірок за рівнем якості технологічного процесу на основі використання критерію домінування, оцінок центральної тенденції вибірки і розсіювання, який може бути використаний при побудові контрольних карт технологічних процесів за якісними характеристиками.

ВИСНОВКИ

Дисертаційна робота містить нові науково обґрунтовані результати застосування шкал класифікації при опрацюванні вимірювальної та експертної інформації. Висновки, що узагальнюють отримані наукові та практичні результати, полягають у наступному:

1. Розроблено метод непараметричної класифікації видів розподілу вибірок малого об'єму, що дозволяє провести попередню класифікацію вимірювальної інформації та вибрати ефективну оцінку центру розподілу і тим самим підвищити точність результату вимірювання.

2. Вперше застосовано метод непараметричної класифікації виду розподілу при побудові контрольних карт технологічного процесу, що дозволяє збільшити чутливість до контрольованих змін центральної тенденції і визначити клас сумарного розподілу вибірок при нестабільності технологічного процесу.

3. Набули подальшого розвитку технології побудови лінгвістичних шкал в інтелектуальних вимірювальних системах для переходу від числових даних до вербальних в напрямку об'єднання вимірювальної та експертної інформації із застосуванням нечіткої метричної класифікації, що дозволило провести адекватне врахування невизначеності при формуванні функцій приналежності терм-множини шкали і отримати характеристику якості встановленої шкали класифікації у вигляді матриці відповідності.

4. Вперше розроблено метод визначення узгодженості вибірок класифікованих даних, заснований на непараметричних оцінках центру вибірки і шкалі метричної класифікації, що дозволяє провести оцінку узгодженості при нерівномірному розташуванні класів еквівалентності в області їх визначення.

5. З метою класифікації стану об'єкту за центральною тенденцією експертних оцінок його властивостей розроблено метод опрацювання вербальних вибірок з обчисленням двох оцінок центральної тенденції, за яким при їх нееквівалентності визначається додаткові (проміжні) класи еквівалентності стану об'єкта.

6. Розроблено метод класифікації стану об'єкта або технологічного процесу за вербальними даними, що ґрунтується на технології використання декількох критеріїв і оцінок і дозволяє вирішити задачу класифікації стану з використанням або встановлених класів еквівалентності або з додатковими (проміжними), це підвищує чутливість контрольних карт за якісними ознаками.

СПИСОК ОПУБЛІКОВАНИХ ПРАЦЬ ЗА ТЕМОЮ ДИСЕРТАЦІЇ

Статті у періодичних наукових виданнях інших держав, які входять до ОЕСР:

1. Semeniuk R.S., Yaremchuk N.A. The methods of obtaining and elaborating classified data in intellectual measurement systems : *Central European Researchers Journal (CERES Journal)*, Volume 6, Issue 1, 2020. С.27-36 (видання держави, що входить до ЄС (Словаччина); посилання на збірник: <http://ceres-journal.eu/iss200601>).

Статті у фахових виданнях:

2. Бугреєв С.С., Семенюк Р.С., Яремчук Н.А. Вимірювання рівня сміття в підземних урнах лазерними рівнемірами : *Український метрологічний журнал*. № 3/2017. С. 36-39.

3. Яремчук Н.А., Семенюк Р.С. Опрацювання вербальних даних за побудови контрольних карт : науково-виробничий журнал *Метрологія і прилади*, № 5-1 (67), 2017. С. 58-61.

4. Семенюк Р.С., Павлишин М.М. Використання методів нечіткої класифікації для оптимізації кількості проб ґрунту в задачах агромоніторингу. *Техніко-технологічні аспекти розвитку та випробування нової техніки і технологій для сільського господарства України* : зб. наук. праць, 2018 р., випуск 22(36). С. 264-267.

5. Семенюк Р., Яремчук Н., Гусар І. Спосіб перевірки узгодженості класифікованих даних під час оцінювання показників якості ґрунту. *Техніко-технологічні аспекти розвитку та випробування нової техніки і технологій для сільського господарства України* : зб. наук. праць, 2019 р., випуск 25(39). С.151-159.

6. Семенюк Р.С. Яремчук Н.А. Використання нечіткої класифікації виду розподілу для вибірок малого об'єму. *Інформаційні системи, механіка та керування* : зб. наук. праць, 2017 р., випуск 17. С. 40-50.

Матеріали коференцій:

7. Яремчук Н.А., Семенюк Р.С. Використання нечіткої класифікації при оцінюванні точності і стабільності технологічних процесів. *Гіротехнології, навігація, керування рухом і конструювання авіаційно-космічної техніки* : зб. тез XI міжнар. наук.-техн. конф., м. Київ, 13-14 квітня 2017 р. С. 91-95.

8. Семенюк Р.С. Опрацювання результатів вимірювання з використанням попередньої нечіткої класифікації. *Метрологія, інформаційно-вимірювальні технології та системи* : зб. тез VI міжнар. наук.-техн. конф., м.Київ, 24-25 жовтня 2017 р. С. 129-131.

9. Семенюк Р.С., Яремчук Н.А. Застосування класифікації розподілів вибірки при побудові контрольних карт. *Приладобудування та метрологія: сучасні проблеми, тенденції розвитку* : зб. тез III всеукр. наук.-практ. конф., м.Луцьк, 11-12 жовтня 2018 р. С. 55-57.

10. Семенюк Р.С. Удосконалення побудови контрольних карт технологічних процесів з застосуванням попередньої класифікації. *Мехатронні системи: інновації та інжиніринг* : зб. тез II міжнар. наук.-практ. конф., м.Київ, 15 червня 2018 р. С. 46-47.

11. Семенюк Р.С. Способы определения терм-множества лингвистической переменной с неопределенностью измерений. *Неопределенность измерений: научные, нормативные и методические аспекты УМ-2018* : зб. тез 15-ї міжнар. наук.-техн. сем., г.Созополь, 10 сентября 2018 р. С. 206-207.

12. Семенюк Р.С., Яремчук Н.А. Основні етапи встановлення лінгвістичних шкал при вимірюваннях і діагностиці. *Метрологія, інформаційно-вимірювальні технології та системи* : зб. тез VII міжнар. наук.-техн. конф., 18-19 лютого 2020 р. С. 131-133.

13. Яремчук Н.А., Семенюк Р.С. Способи урахування невизначеності при побудові лінгвістичних шкал. *Метрологія та вимірювальна техніка (Метрологія–2020)* :

зб. доп. XII міжнар. наук.-техн. конф. 6–8 жовтня 2020 р. – Х.: ННЦ Інститут метрології, 2020. – С. 77-81.

14. Ступак С.С., Яремчук Н.А., Семенюк Р.С. Способи урахування впливу невизначеності на результати «м'яких» вимірювань. *Актуальні проблеми автоматики та приладобудування* : матеріали 2-ї міжнар. наук.-техн. конф., м.Харків, 06-07 грудня 2018 р. С. 210-211.

АНОТАЦІЇ

Семенюк Р.С. Методи опрацювання вимірювальної та експертної інформації з застосуванням шкал класифікації – Рукопис.

Дисертаційну роботу присвячено вирішенню важливої науково-практичної проблеми – розробці методів опрацювання експериментальних даних (числових і вербальних) із застосуванням класифікації.

В роботі набув подальшого розвитку метод непараметричної ідентифікації форми розподілу, на основі чого був розроблений метод непараметричної класифікації форми розподілу вибірок малого об'єму, який може бути застосований для удосконалення методів опрацювання результатів багаторазових вимірювань, а також при побудові контрольних карт технологічних процесів.

Набули подальшого розвитку технології побудови лінгвістичних шкал в інтелектуальних вимірювальних системах при застосуванні метричної класифікації для переходу від числових даних до вербальних. Запропонована послідовність етапів встановлення або відтворення лінгвістичної шкали класифікації, причому основна увага приділена аналізу чинників, що характеризують нечіткість правил і невизначеність вимірювання, а також способам їх урахування при побудові терм-множини шкали класифікації, визначена характеристика якості шкали класифікації у вигляді матриці відповідності.

Розроблено метод визначення узгодженості вибірок класифікованих даних, заснований на непараметричних оцінках центру вибірки і шкалі метричної класифікації, що дозволяє провести оцінку узгодженості при нерівномірному розташуванні класів еквівалентності.

Розроблено метод класифікації стану об'єкту або технологічного процесу за вербальними даними, що ґрунтується на технології використання декількох критеріїв і оцінок і дозволяє вирішити задачу класифікації стану з використанням або встановлених класів еквівалентності або з додатковими (проміжними).

Основні наукові положення дисертаційної роботи підтверджені експериментальними дослідженнями і впровадженням в конкретні інформаційно-вимірювальні системи.

Ключові слова: непараметрична класифікація розподілів, шкала метричної класифікації, нечіткість експертної інформації, невизначеність вимірювання, опрацювання вербальних даних, контрольні карти.

Semeniuk R.S. Methods of processing measuring and expert information using classification scales - Manuscript.

The dissertation is devoted to solving an important scientific and practical problem - development of methods for processing experimental data (numerical and verbal) using classification.

The paper reviews the scientific works of scientists who worked in the field of classification theory. The main features of classification are defined, among which one of the

main ones is the presence of a classification scale developed in advance or in the classification process, based on the condition of distinguishing equivalence classes. In complex classification cases, when the solution is presented in several gradations, the classification quality characteristic is the conditional probability of being assigned to a certain equivalence class. Scientists' experience shows that the statistical phenomena a researcher encounters in the field of classification are complex and not always predictable, therefore methods and techniques for different applications of the classification procedure are very important. This issue is the focus of the author's dissertation, which presents techniques for using classification to work through numerical data, transitioning from numerical data to verbal data using measurement and expert information, and working through verbal data.

Among the main points of scientific novelty and practical importance it is worth noting the following:

- methods for the non-parametric identification of the shape of the distribution of small samples have been further developed, based on which a method for the non-parametric classification of distribution types has been developed that makes it possible to select an effective estimate of the sample distribution center and thereby reduce the uncertainty of the measurement result;

- the method of non-parametric classification of the distribution type was applied for the first time in the construction of process control charts, thus increasing the sensitivity to controllable changes in central tendency and determining the class of the total sampling distribution for process instability;

- methods for transition from numerical to verbal data in information and measurement system using a metric classification scale have been further developed, and a theoretical rationale for combining uncertainty in expert information and measurement uncertainty has been developed, that allowed to adequately take into account the uncertainty of the measurement information in the formation of the membership functions of the term set of the scale and to obtain a characteristic of the quality of the established classification scale in the form of a correspondence matrix;

- a method for checking the consistency of samples of verbal data obtained behind a metric classification scale has been developed for the first time, allowing consistency checking when equivalence classes are not evenly spaced in their domain of definition;

- based on the application of several operators for working with samples of verbal data, the technology for classifying the state of objects based on expert information has been developed, followed by the use of the resulting classification scale to build process control charts for qualitative attributes.

Theoretical and practical results of the thesis are connected with the scientific themes and plans of the department of information and measuring technologies of Igor Sikorsky Kyiv Polytechnic Institute, as well as research paper of the state budget / fiscal theme «Research and development of metrological and technical support for the implementation of environmental monitoring tasks in Ukraine» (№ RP 0113U002159) and research paper «Theoretical and practical principles of the application of soft measurements in product quality systems» (№ RP 0118U001478).

Keywords: non-parametric classification of distributions, metric classification scale, expert information uncertainty, measurement uncertainty, verbal data elaboration, control maps.

Підписано до друку 23.03.2021 р. Формат 60/90 1/16.
Папір офсетний. Друк лазерний
Гарнітура *Times*. Умовн. др. арк. 0,9.
Наклад 100 прим. Зам. № 2303/01

Надруковано ФОП Гузік О.М.
Реєстраційний номер №2705814113
м. Київ, вул. Б. Гаврилишина, 16
Тел.: 338-16-61.