

УДК 681.518:004
КП
№ держреєстрації 0109U001380
Інв. №

Міністерство освіти і науки України
Сумський державний університет
(СумДУ)
40007, м. Суми, вул. Римського-Корсакова, 2
тел. (0542) 33-54-79 факс (0542) 33-54-79

ЗАТВЕРДЖУЮ
Проректор
з наукової роботи
д.ф-м.н., професор
А.М. Черноус
2010-12-24 . . .

ЗВІТ
ПРО НАУКОВО-ДОСЛІДНУ РОБОТУ
«РОЗРОБЛЕННЯ НАУКОВО-МЕТОДИЧНИХ ОСНОВ ТА
ІНФОРМАЦІЙНИХ ЗАСОБІВ ПРОЕКТУВАННЯ ЗДАТНИХ
САМОНАВЧАТИСЯ АДАПТИВНИХ СИСТЕМ КЕРУВАННЯ
ТЕХНОЛОГІЧНИМИ ПРОЦЕСАМИ»

(проміжний звіт)

Начальник НДЧ
канд. техн. наук, доцент
Керівник НДР
доктор техн. наук, професор

В.А. Осіпов
А.С. Довбиш

2010

Рукопис завершено 23 грудня 2010 р.

Результати даної роботи розглянуто
науковою радою СумДУ, протокол від 2010-12-23 № 6

СПИСОК АВТОРІВ

Керівник НДР, доктор техн. наук, професор	А.С.Довбиш (вступ, підрозділ 1.3)
Молодший наук. співробітник	І.В. Шелехов (підрозділи 2.5, 3.4)
Молодший наук. співробітник	С.О. Петров (підрозділ 3.6)
Молодший наук. співробітник	С.С. Мартиненко (підрозділ 3.1)
Інженер другої категорії	Н.М.Бровко (підрозділ 1.1)
Провідний фахівець	О.О. Дзюба (підрозділ 3.5)
Провідний фахівець	С.М. Котенко (підрозділ 3.1)
Провідний фахівець	О.А. Якушев (підрозділ 3.3)
Аспірант	Саад Джулгам (підрозділ 2.2)
Аспірант	О.В. Ліщинський (підрозділ 3.6)
Аспірант	О.Б. Берест (підрозділ 1.2)
Аспірант	М.О. Руденко (підрозділ 2.1)
Аспірант	К.В. Барило (підрозділ 3.2)
Аспірант	О.Б. Барило (підрозділ 3.2)
Студент	Н.І. Андрієнко (підрозділ 3.5))
Студент	С.П. Занченко (підрозділ 2.4)
Студент	В.В. Москаленко (підрозділ 3.4)

РЕФЕРАТ

Звіт: 125 с., 48 рисунків, 3 таблиці, 74 джерела.

Мета роботи – розроблення методологічних та теоретичних засад створення адаптивних систем керування складним технологічним процесом.

Об’єкт дослідження – слабо формалізований технологічний процес.

Предмет дослідження – моделювання, оцінка функціональної ефективності, алгоритми оптимізації та адаптації СППР, що самонавчається.

Даний звіт є проміжним.

На основі аналітичного огляду та аналізу тенденції розвитку інтелектуальних автоматизованих систем керування сформовано задачу аналізу та синтезу системи підтримки прийняття рішень (СППР), що самонавчається в режимі факторного кластер-аналізу. Розроблено науково-методологічні основи інформаційно-екстремального методу аналізу і синтезу СППР, що самонавчається в режимі факторного кластер-аналізу за умов апріорної невизначеності, інформаційних і ресурсних обмежень. Для побудови безпомилкових за навчальною вибіркою вирішальних правил розроблено алгоритми самонавчання СППР з оптимізацією словника ознак розпізнавання в режимі факторного кластер-аналізу. Розроблено категорійні моделі та досліджено вплив параметрів навчання на функціональну ефективність СППР.

АДАПТИВНА СИСТЕМА КЕРУВАННЯ, ОПТИМІЗАЦІЯ,
ФУНКЦІОНАЛЬНА ЕФЕКТИВНІСТЬ, ІНФОРМАЦІЙНИЙ КРИТЕРІЙ,
СИСТЕМА ПІДТРИМКИ ПРИЙНЯТТЯ РІШЕНЬ, ТЕХНОЛОГІЧНИЙ
ПРОЦЕС, ІНФОРМАЦІЙНО-ЕКСТРЕМАЛЬНА ІНТЕЛЕКТУАЛЬНА
ТЕХНОЛОГІЯ, ПРОГНОЗУВАННЯ

ЗМІСТ

ПЕРЕЛІК УМОВНИХ СКОРОЧЕНЬ	5
ВСТУП	6
РОЗДІЛ 1 АНАЛІЗ ПІДХОДІВ І МЕТОДІВ АВТОМАТИЗАЦІЇ ТЕХНОЛОГІЧНИХ ПРОЦЕСІВ	8
1.1 Сучасний стан і тенденції проектування та розвитку АСК, що самонавчаються.....	8
1.2 Аналіз проблеми оптимізації словника ознак в задачах автоматизації складних технологічних процесів	12
1.3. Постановка задачі аналізу та синтезу СППР, що самонавчається в режимі факторного кластер-аналізу.....	34
РОЗДІЛ 2 АНАЛІЗ І СИНТЕЗ СППР, ЩО САМОНАВЧАЄТЬСЯ	40
2.1 Інформаційна модель слабо формалізованого технологічного процесу ...	40
2.2 Основні принципи аналізу та синтезут СППР, що самонавчається.....	42
2.3 Математичні моделі СППР, що самонавчається.....	46
2.4 Оцінка функціональної ефективності СППР, що самонавчається.....	55
2.5 Вплив потужності словника ознак на структуру бінарного простору ознак.....	58
РОЗДІЛ 3 ОПТИМІЗАЦІЯ ПАРАМЕТРІВ ФУНКЦІОНУВАННЯ СППР, ЩО САМОНАВЧАЄТЬСЯ.....	70
3.1 Інформаційно-екстремальний алгоритмом навчання СППР	71
3.2 Оцінка інформативності ознак	80
3.3 Оптимізація словника ознак з використанням випадкового пошуку	84
3.4 Інформаційно-екстремальна оптимізація словника ознак за послідовно-спадковою селекцією	90
3.5 Гібридний алгоритм оптимізації словника ознак.....	96
3.6 Унімодальний класифікатор.....	104
ВИСНОВКИ.....	115
ПЕРЕЛІК ПОСИЛАНЬ.....	118

ПЕРЕЛІК УМОВНИХ СКОРОЧЕНЬ

АСК – автоматизована система керування;

АСКТП – автоматизована система керування технологічним процесом;

БПОІ – блок первинного оброблення інформації;

БКОІ – блок кінцевого оброблення інформації;

БФВР – блок формування векторів розпізнавання;

БП – блок прогнозування;

ГА – генетичний алгоритм;

ІЕІТ – інформаційно-екстремальна інтелектуальна технологія;

ЗР – зменшення розмірності;

КФЕ – критерій функціональної ефективності;

ОПР – особа, що приймає рішення;

ПСС – послідовно-спадна селекція;

СО – селекція ознак;

СКД – система контрольних допусків;

СНД – система нормованих (експлуатаційних) допусків;

СМД – складні мінеральні добрива;

СППР – система підтримки прийняття рішень.

ВСТУП

Основною тенденцією сучасного розвитку науково-технічного прогресу всіх галузей соціально-економічної сфери суспільства є інтелектуалізація творчої діяльності шляхом моделювання на базі ЕОМ притаманних людині когнітивних процесів при прийнятті рішень. Особливо актуальним для промисловості України є використання інтелектуальних технологій при проектуванні та виготовленні виробів, що повинні задовольняти вимогам світових стандартів. У теперішній час будь-який складний виріб не є конкурентно спроможним, якщо він не містить інтелектуальної складової. І далі ця тенденція буде тільки підсилюватися. Тому основні світові виробники сучасних автоматизованих систем керування (АСК) зрозуміли, що орієнтація на виготовлення високо точного технологічного обладнання без інтелектуальної складової не дозволяє ефективно здійснювати керування слабо формалізованими процесами за умов апріорної невизначеності та ресурсних обмежень. Наприклад, основною умовою використання такого обладнання є вимога жорсткого вхідного контролю сировини та матеріалів. Але для сировини та матеріалів природничого походження, які широко використовуються в хімічній, металургійній, харчовій та інших галузях промисловості, виконання такої умови є ускладненим через невирішені технічні проблеми поточного контролю. Інтелектуалізація АСК складними слабо формалізованими технологічними процесами дозволяє надати їм властивість адаптивності на основі самонавчання та розпізнавання образів. При цьому науково-методологічні питання підвищення функціональної ефективності АСК, що самонавчаються, все ще залишаються недостатньо дослідженими. Так, основними недоліками відомих методів самонавчання АСК, які роблять неефективним їх застосування на практиці, є ігнорування перетину класів розпізнавання у просторі ознак розпізнавання та відсутність оптимізації

словників ознак в процесі навчання системи. Це призводить як до невисокої достовірності розпізнавання функціональних станів технологічних процесів, так і до необхідності застосування надлишкових в інформаційному розумінні словників ознак великих обсягів. Указані вище методологічні та теоретичні ускладнення вирішення проблеми побудови високоефективних інтелектуальних АСК обумовлюють актуальність розроблення математичних моделей класифікаційного керування, методів оцінки функціональної ефективності та оптимізації просторово-часових параметрів функціонування, включаючи параметри словника ознак, а так само відповідних засобів інформаційної технології проектування високоефективних АСК, що навчаються (самонавчаються), та широке їх впровадження в різні галузі соціально-економічної сфери України.

У першому розділі наведено аналітичний огляд методів аналізу та синтезу інтелектуальних СППР для керування складними технологічними процесами і постановку задачі синтезу СППР, що самонавчаються в режимі факторного кластер-аналізу.

Другий розділ присвячено висвітленню теоретичних і методологічних положень інформаційно-екстремального методу аналізу і синтезу СППР, що самонавчаються в режимі факторного кластер-аналізу з оптимізацією словника ознак.

У третьому розділі наведено результати оптимізації параметрів функціонування СППР, що самонавчаються в режимі факторного кластер-аналізу.

Одержанні при виконанні науково-дослідної роботи результати опубліковано у працях [1-19].

РОЗДІЛ 1

АНАЛІЗ ПРОБЛЕМИ ТА ПОСТАНОВКА ЗАДАЧІ ДОСЛІДЖЕННЯ

1.1 Сучасний стан і тенденції проектування та розвитку АСК, що самонавчаються

Підвищення функціональної ефективності автоматизації керованих технологічних процесів, розподілених в просторі і часі, пов'язано з розробленням науково-методологічних основ проектування та впровадження адаптивних АСК на основі самонавчання та автоматичної класифікації [20]. Основними властивостями таких систем є:

- наявність слабо формалізованого керованого процесу, що характеризується нестационарністю, імплікативністю, багато критеріальністю та впливом неконтрольованих факторів;
- наявність нечітких вхідних даних, що обумовлено довільними (не нульовими) початковими умовами функціонування АСК, що навчається, в режимі моніторингу за умов апріорної невизначеності, інформаційних та ресурсних обмежень;
- наявність замкнених зворотних зв'язків, що допускає можливість роботи в режимі відпрацювання внутрішніх та зовнішніх збурень з метою стабілізації керованого процесу;
- наявність локальних систем автоматичного керування шляхом використання аналогових і цифрових ПД-регуляторів;
- наявність системи підтримки прийняття рішень (СППР), яка може функціонувати у двох роздільних (або об'єднаних) у часі режимах: навчання, в процесі якого формуються безпомилкові за навчальною вибіркою вирішальні правила, і екзамену, на якому безпосередньо приймаються рішення про оцінку функціонального стану системи та вироблення керуючих команд;

- багатопотоковість режиму функціонування, що призводить до асинхронності оброблення інформації, що робить ускладненим застосування традиційних математичних методів моделювання СППР, що навчаються;
- спроможність самооцінки функціональної ефективності систем.
- знання-орієнтованість (здатність до засвоювання і виведення знань);
- використання сучасних електронно-обчислювальних комплексів з потужною довготривалою пам'яттю й оперативністю оброблення інформації;
- ергономічність і можливість функціонування в інтерактивному режимі;
- об'єктно-орієнтована методологія проектування систем;

Аналіз сучасного стану та тенденції розвитку основних напрямів синтезу та аналізу АСК, що самонавчаються, свідчить про інтенсивну розробку та впровадження методів проектування таких систем на базі штучних нейронних мереж []. Серед основних прикладних задач, що розв'язуються за допомогою нейронних мереж, можна зазначити фінансове прогнозування, інтелектуальну обробку даних, діагностику систем, контроль за функціонуванням мереж, шифрування даних та інше. Основними нерозв'язаними задачами даного напрямку вважаються:

- задача інтерпретованості вагових коефіцієнтів, пов'язана із складністю інтерпретації змісту інтенсивності вхідних сигналів і вагових коефіцієнтів;
- задача інтерпретованості передатної функції (складність інтерпретації і обґрунтування адитивності аргументу і вигляду активаційної функції нейрона);
- задача багатовимірності, що призводить до "комбінаторного вибуху", який виникає при визначенні структури зв'язків нейронів, підборі вагових коефіцієнтів і передатної функції;
- задача лінійної розділимості, оскільки збудження нейронів приймає виключно булеві значення (0 або 1).

При цьому проблема інтерпретованості призводить до зниження якості одержаних результатів, проблема багатовимірності – до значних обмежень на місткість і складність структури мереж, а проблема лінійної розділимості – до необхідності застосування складних багат шарових мереж навіть для нескладних задач. Логічним кроком вирішення останньої проблеми є тенденція до розробки та впровадження систем керування на базі нечітких штучних нейронних мереж.

Недостатня ефективність і модельність відомих методів статистичної класифікації обумовили інтенсивний розвиток у теперішній час методів нечіткої класифікації, основи якої закладено Л. Заде [21]. Методи нечіткої класифікації знайшли застосування переважно в гібридних АСК, що функціонують за умов невизначеності та дії багатьох неконтрольованих факторів [22]. Але при цьому питання оцінки функціональної ефективності та оптимізації процесу навчання АСК залишаються все ще мало дослідженими.

Одним з важливих напрямків проектування інтелектуальних СППР, які є складовою частиною адаптивних АСК розподіленими в часі і просторі технологічними процесами, є створення високоефективних алгоритмів машинного навчання, які дозволяють будувати безпомилкові за багатовимірною навчальною матрицею вирішальні правила. Вирішення цієї проблеми у рамках детермінованого або статистичного підходів, що є базовими для даної галузі досліджень, ускладнено через модельність алгоритмів навчання, які не придатні для практичного застосування. При цьому ігнорування перетину класів розпізнавання і відсутність алгоритмів оптимізації процесу навчання за прямим КФЕ вказує на незавершеність даного підходу навчання систем керування. Детерміновано-статистичний підхід до аналізу і синтезу здатних навчатися СППР [23] намагається поєднати переваги детермінованих і статистичних методів і обійти їх недоліки. Тенденція до синтезу оптимальних алгоритмів функціонування СППР, що навчаються, шляхом якісного аналізу існуючих класичних

підходів та їх модифікацій з урахуванням поточного рівня розвитку комп'ютерної техніки та додаткових вимог практичного застосування проявляє себе в останніх перспективних розробках даної галузі, однією з яких є інформаційно-екстремальна інтелектуальна технологія (ІЕІ-технологія).

Основна ідея методів аналізу і синтезу здатних навчатися (самонавчатися) СППР у рамках ІЕІ-технології полягає в оптимізації структурованих просторово-часових параметрів функціонування системи шляхом трансформації в процесі навчання відношення схожості на нечіткому розбитті простору ознак на класи у відношення еквівалентності. При цьому оптимізація параметрів функціонування здійснюється за ієрархічною ітераційною процедурою транзитивного замикання на динамічні відображення у відповідних контурах оптимізації параметрів шляхом пошуку глобального максимуму інформаційного критерію функціональної ефективності (КФЕ) навчання в робочій області визначення його функції. Побудова безпомилкового за навчальною матрицею класифікатора згідно з принципами дуальності оптимального керування [24], редукції [25] і максимізації інформації [1] здійснюється у дискретному субпарацептуальному просторі шляхом допустимих перетворень апріорного нечіткого унімодального розподілу реалізацій образу з метою його вписування в оптимальний контейнер класу розпізнавання, що відновлюється в радіальному базисі. При цьому трансформація апріорного розподілу реалізацій образу здійснюється шляхом цілеспрямованої зміни значень ознак. При цьому зростання продуктивності сучасних комп'ютерів в сполученні з підвищенням якості алгоритмів дозволяє застосовувати створені раніше теоретичні наукові розробки на практиці. Аналогічним чином проходить зміна і інтелектуальних СППР для автоматизації складних технологічних процесів, які намагаються використовувати достатньо прості, але ресурсномісткі алгоритми адаптивної поведінки.

Серед основних проблем даної сфери можна зазначити:

- модельний характер переважної більшості відомих методів автоматичної класифікації, що становить певну теоретико-методологічну цінність, але є непридатним для практичного застосування;

- незавершеність теорії ефективного машинного навчання АСК складними технологічними процесами;

- нелінійність, нестационарність і багатфакторність виробничих технологічних процесів, які відбуваються за умов апріорної невизначеності, що й обумовлює їх слабку формалізованість;

- підвищенні вимоги до оперативності етапів навчання і перенавчання АСК, що навчаються (самонавчаються) в режимі реального часу.

Не зважаючи на суттєві досягнення в галузі автоматичної класифікації і великий обсяг публікацій, значних зрушень у підвищенні ефективності навчання АСК слабо формалізованими процесами все ще не відбувається. Детальний аналіз даної тенденції вказує на те, що практично всі проблеми в цій галузі пов'язані з нехваткою методологічних ресурсів, що породжує гострий дефіцит робочих алгоритмів оптимізації просторово-часових параметрів функціонування, які прямо або непрямо впливають на функціональну ефективність АСК, що самонавчаються.

Таким чином, актуальною є розробка теоретичних основ для проектування оптимальних в інформаційному розумінні ефективних АСК, що навчаються (самонавчаються), а також відповідних засобів інформаційної технології, які дозволять впровадження таких систем на практиці.

1.2 Аналіз проблеми оптимізації словника ознак в задачах автоматизації складних технологічних процесів

Вирішення проблеми оптимізації словника ознак пов'язано з подоланням суттєвих ускладнень, які обумовлено такими причинами:

- незавершеність теорії оптимізації словника ознак через методологічні ускладнення переходу від формалізованих (конструктивних) методів і алгоритмів оптимізації до неформалізованих (евристичних);
- значні часові та вартісні витрати при зборі, передачі, зберіганні, структуризації даних та складність оцінки їх валідності та релевантності;
- проблема багатовимірності, що пов'язана з функціонуванням інтелектуальних АСК складними технологічними процесами за умов великої потужності алфавіту класів розпізнавання та словника ознак.

На практиці часто недооцінюється важливість релевантності функціонування СППР, що самонавчається. Найбільшою проблемою, що пов'язана з цим, є врахування характеристик, що змінюються з часом. Вимоги до характеристик найбільш бажаного класу стають суворішими, змінюються повністю чи частково, зменшуючи ефективність функціонування СППР в цілому. Для багатьох систем єдиним способом вирішення цієї проблеми є періодичне перенавчання на основі нової інформації, що не є завжди можливим, зважаючи на суттєві витрати, пов'язані з залученням експертів на час перенавчання та зупинкою підприємства або технологічного процесу, яким керувала АСК, до складу якої входить інтелектуальна СППР.

Редукція простору ознак відноситься до широко відомих методологій, що намагаються відобразити вектор багатовимірної гіперпростору в вектор гіперпростору меншої розмірності [25]. Існують дві основні категорії методів, що використовуються для цього:

- 1) зниження розмірності (ЗР) – спрощення гіперпростору шляхом трансформування його осей;
- 2) селекція ознак (СО) – вибір підмножини інформативних ознак з початкового словника без зміни осей.

Головна відмінність між даними підходами полягає у відношенні до структурних зв'язків образів – семантики: ЗР модифікує вісі простору, таким чином безповоротно змінюючи його семантику, а селекція ознак видаляє

певні вісі, не модифікуючи ті, що залишилися; тобто не змінюючи семантики.

Класи, що повністю покривають пов'язаний з ними простір, досить незвичне явище в задачах розпізнавання образів. Не зважаючи на це в більшості СППР використовуються надлишкові словники ознак, що пояснюється значними витратами або надмірною складністю повного аналізу предметної галузі на етапі формування словника.

Редукція простору ознак доцільна з таких причин:

1) досить суттєва частина ознак в реальних системах в певній мірі корелює між собою, що є підтвердженням надлишковості даних. Причиною цього є практична складність або неможливість гарантувати незалежність предметних ознак, тому що їх належність до однієї предметної галузі по суті вказує на наявність певних взаємозв'язків між ними;

2) перешкоди при передачі даних досить часто є причиною виникнення уявної кореляції між незалежними ознаками або навіть включення до словника непередметних та неінформативних ознак;

3) проблема багатовимірності, щодо експоненціальності залежності розміру вибірки, необхідної для визначення виду багатопараметричної функції, від кількості параметрів даної функції. Так, наприклад, вважається, що для побудови точного класифікатора для класу з n характеристиками необхідно 10^n реалізацій, що, враховуючи необхідність створення систем класифікації для декількох класів, призводить до значних складностей при зберіганні та використанні навчальних вибірок та унеможливорює створення СППР, що навчаються в реальному часі.

ЗР застосовується для вирішення проблеми проєкціювання даних у простори меншої розмірності за допомогою трансформування осей початкового простору. Нові осі визначаються таким чином, що проявляється максимальна різноманітність при мінімальному зменшенні кількості корисної інформації. Виділяють чотири основні групи методів ЗР:

1) лінійні, що спрямовані на пошук нового координатного простору, у якому кожна координатна вісь є лінійною комбінацією вихідних ознак;

2) нелінійні, що нехтують аналітичним виразом перетворення вхідного простору ознак у нових координатних осях, намагаючись побудувати таке відображення сукупності даних, що дає найбільше уявлення про особливості їхньої структури;

3) методи багатовимірного шкалування, що дозволяють за заданою інформацією між об'єктами розглянутої сукупності приписувати кожному з них вектор його кількісних характерних показників; при цьому розмірність шуканого координатного простору задається заздалегідь, а занурення у нього аналізованих об'єктів виробляється таким чином, щоб структура взаємних розходжень між ними, обмірюваних за допомогою приписуваних їм допоміжних координат, у середньому найменше відрізнялася б від заданої в змісті того чи іншого функціонала якості;

4) методи розгорнення простору ознак, що намагаються заповнити даний простір гіперкривою таким чином, щоб близькі в просторі об'єкти виявилися по можливості близькими на цій кривій. Після чого проводиться аналіз гістограм розподілів об'єктів на побудованій гіперкривій.

Наведемо найбільш відомі методи ЗР.

Метод головних компонент (МГК) [26], що полягає в максимізації дисперсії комбінації змінних уздовж осей, які визначають нову систему координат та відповідно трансформують навчальну вибірку. Нові осі створюються відповідно до зменшення дисперсії таким чином, що перша ознака в нових трансформованих даних характеризується найбільшою дисперсією. Кореляція між ознаками у сформованому таким чином просторі ознак зменшується або повністю зникає, зменшуючи відповідно і надлишковість словника ознак. Після зниження розмірності за МГК вибирають необхідну кількість так названих головних компонент, починаючи з однієї. Всі інші ознаки не враховуються. До основних недоліків МКГ слід віднести: безповоротне знищення базової семантики простору

ознак, неможливість застосування даного методу як методу попереднього оброблення в процедурах символічного та описового нечіткого моделювання. Крім того МГК обробляє виключно числові (не символічні) дані. Цей метод здатен виконувати лише лінійне проєкціювання, ігноруючи всі нелінійні структурні зв'язки образів. Зрештою, необхідно врахувати, те що кількість ознак, які увійдуть в словник, а отже і співвідношення інформації, втраченої при ЗР, визначається людиною.

Метод регресії на основі цілеспрямованого проєкціювання (ЦП) [27] подібно до МГК створює одно-, дво- або тривимірну проєкцію на основі вихідного простору ОР. ЦП визначає параметри проєкцій на базі максимізації певних критеріїв якості. Основним недоліком даного методу є занадто значні обчислювальні витрати при обробці просторів високої розмірності та його нездатність створити точну функцію для опису образу.

Узагальнені адитивні моделі (УАМ) [28] подібні моделям, що використовуються у методі регресії на основі ЦП. Базове припущення даного методу про відсутність взаємозв'язків між вхідними змінними є одночасно і його головним недоліком. Незважаючи на це УАМ знайшли досить широке застосування у нейромережному інформаційному моделюванні.

Багатовимірні адаптивні регресійні сплайни (БАРС) є модифікацією УАМ, яка враховує взаємозв'язки між вхідними змінними [29]. Даний підхід використовує більш складні моделі та регресійні методи для створення точніших моделей. Замість хребтових функцій (*ridge functions*) метод БАРС використовує сплайни, параметричні криві, що визначаються на основі контрольних точок – вузлів, та базисної функції або матриці. Кожен вузол характеризується асоціативною вагою. Метод БАРС створює багатовимірний сплайн для апроксимації форми простору ознак. Подібно до багатьох апроксимуючих функцій, даний метод використовує неперервну диференційовану апроксимуючу функцію. Застосування методу БАРС в багатьох областях характеризується надзвичайною ефективністю та якістю отриманих результатів. Неперервність результуючої апроксимуючої моделі є

одним з найбільш бажаних результатів при застосуванні статистичного аналізу. Втім проблема багатовимірності має прояв і в методі БАРС. Кожний вимір в просторі ознак збільшує вимірність моделі, тобто обчислювальні витрати на обробку та зберігання сплайнів. Складність має експоненціальну залежність від кількості ознак. Крім того до недоліків БАРС слід віднести його чутливість до перешкод, врахування яких вимагає використання більш складних моделей.

До незаперечних переваг наведених вище методів ЗР слід віднести можливість отримання візуального уявлення про структури досліджуваних даних. Візуалізація припускає одержання тим чи іншим способом графічного відображення сукупності об'єктів на числову вісь, на площину чи в тривимірний простір, що максимально відбиває особливості розподілу цих об'єктів у багатовимірному об'єму. Швидко зростаюча роль візуального аналізу обумовлена розвитком технічних засобів, що забезпечують побудову обробку і модифікацію візуальних образів в інтерактивному режимі за короткий час.

СО вирішує задачу редукції простору ознак через видалення малоінформативних ознак. Головною перевагою СО над ЗР є збереження характерної семантики образів, тобто такої форми даних, яка є придатною для людського сприйняття та обробки відповідними методами.

Виділяють чотири основні групи методів СО [68]:

1) ідеальні, що намагаються сформуванню словник мінімальної потужності, використання якого буде необхідним та достатнім для розв'язання задачі класифікації;

2) обмежені заданою потужністю словника або класичні, що намагаються знайти оптимальну комбінацію заданої кількості ознак (N^*) з початкового словника потужністю N ($N^* \leq N$) серед усіх можливих;

3) орієнтовані на безпосереднє поліпшення точнісних характеристик системи, які намагаються провести процес оптимізації словника ознак таким

чином, щоб значення точності системи на кожному кроці зростало або залишалось без змін;

4) апроксимуючі оригінальний розподіл реалізацій, які намагаються обрати таку підмножину ознак, що розподіл реалізацій класів, отриманий з використанням тільки цих ознак, наближаються до розподілу, отриманого з використанням початкового словника .

Така систематика є досить умовною, що пояснюється існуванням певних характеристик спільних для більш ніж однієї групи і дозволяє створювати інші форми класифікації алгоритмів СО. Найбільш відомі з них подано у працях [31,32]. У праці [31] методи СОР розділені на три групи: методи минулого, теперішнього і майбутнього. Головна увага приділялася методу гілок та границь [33] та його модифікаціям.

Подальший розвиток даного підходу дозволяє створити систему класифікації, що практично повністю включає всі існуючі методи СО. Для її детального опису розглянемо архітектуру типового алгоритму СОР (рис. 1.1).

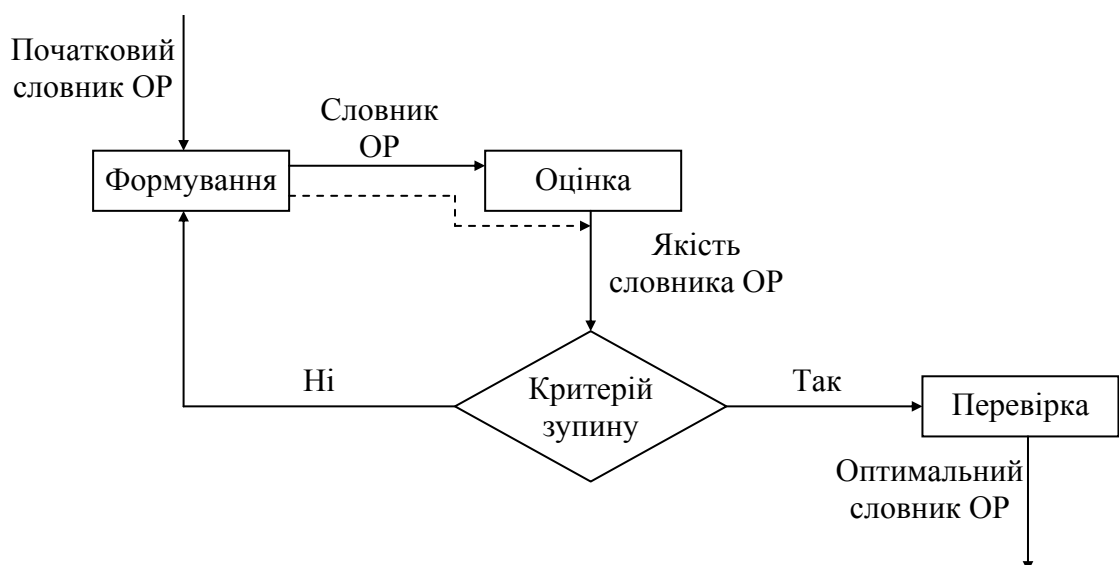


Рисунок 1.1 – Структурна схема типового алгоритму СО

Архітектура типового алгоритму СО може бути подана у вигляді таких чотирьох етапів:

- 1) формування, що створює наступний варіант словника ознак;

- 2) оцінка ефективності використання сформованого словника ознак;
- 3) перевірка на виконання певного критерію зупину (*Stopping criterion*);
- 4) додаткова перевірка валідності застосованого методу СО.

Етап формування базується на процедурі пошуку [31]. Власне кажучи, він створює один з можливих варіантів словника для подальшої його оцінки. Хоча при формуванні використовується початковий словник, даний етап може розпочинатися не тільки з варіанту словника, що включає всі ознаки, але з пустого або випадково сформованого. Наступний словник створюється шляхом додавання/видалення ознак із сформованого на попередній ітерації варіанту або повторного випадкового формування. Існують три групи методів формування словників ознак.

- 1) Оптимальні, що базуються на алгоритмі повного перебору всіх можливих варіантів словника ознак цієї групи включають також методи, які використовують різноманітні евристичні функції для проведення направлено пошуку без ризику втрати якості отриманого результату. Таким чином, хоча порядок простору пошуку дорівнює $O(2^N)$, кількість варіантів словника може бути зменшена [34]. Крім того, використання процедури пошуку з поверненням (*backtracking*) дає додаткові гарантії знаходження оптимального згідно обраної функції оцінки словника. Реалізація такої процедури здійснюється за допомогою різних підходів, наприклад: методу гілок та границь з поверненням (*branch and bound with backtracking*), пошуку ушир (*breadth first search*), пошуку по першому найліпшому збігу з поверненням (*best first search with backtracking*).

- 2) Евристичні, що на кожній ітерації намагаються таким чином змінити склад поточного словника, щоб максимально наблизити його до оптимального. Методи даної групи характеризуються легкістю застосування і досить високою оперативністю.

- 3) Стохастичні, що базуються на процедурі випадкового формування словника. При цьому кількість варіантів словників обмежується

максимальною кількістю ітерацій, що задається виходячи з конкретних характеристик поставленої задачі та наявних ресурсів системи.

Евристичні та стохастичні методи крім того поділяються на такі, що в процесі оптимізації зберігають і модифікують тільки один поточний словник, а також методи, що працюють з популяцією словників

Функція оцінки визначає ефективність використання сформованого словника і порівнює її з найкращим значенням оцінки, отриманої на попередніх ітераціях. У випадку її зростання, поточний варіант словника приймається за найкращий. За функцією оцінки СО поділяють на фільтри (методи попередньої обробки) та вкладені методи. Фільтри використовують елементи теорії інформації, що робить їх найбільш універсальними методами оцінки. Недоліком фільтрів можна вважати те, що вони безпосередньо не використовують інформацію про класифікаційні особливості ознак [35]. Вкладені методи навпаки базуються на оцінці того, наскільки ефективно проводиться класифікація з використанням даного словника ознак, що вказує на наявність певних симбіотичних відносин між ними та системою класифікації. Завдяки цьому вкладені методи характеризуються більш якісними, ніж фільтри, результатами, але втрачають в оперативності та універсальності. Крім того застосування вкладених методів може бути проблематичним для деяких особливо специфічних класифікаторів.

Окрема класифікація функцій оцінки вводиться за мірами якості, що використовуються в них:

1) Дистанційні, що розраховують та максимізують ступінь рознесеності образів, наприклад, Евклідову кодову відстань. Характеризуються універсальністю та незначними обчислювальними витратами.

2) Інформаційні, що визначають приріст кількості інформації (різницю між апіорною та очікуваною апостеріорною невизначеністю) при використанні кожної ознаки, наприклад, ентропійні міри Шеннона або

інформаційна міра Кульбака. Характеризуються універсальністю та незначними обчислювальними витратами.

3) Кореляційні, що визначають малоінформативні ознаки за критеріями пов'язаності. Характеризуються універсальністю та незначними обчислювальними витратами.

4) Сумісні, що визначають сумісність навчальних даних, використовуючи поняття мінімального розміру репрезентативної навчальної вибірки, який визначається для поточного варіанту словника. Характеризуються універсальністю та помірними обчислювальними витратами.

5) Точнісі, що використовуються у вкладених методах, для яких функція оцінки залежить від точнісних характеристик СППР, і характеризуються високою точністю але значними обчислювальними витратами.

На вибір КЗ впливають особливості попередніх етапів СО. Серед КЗ, які базуються на особливостях етапу формування, виділяють такі, що зупиняють ітеративний процес при досягненні певної потужності словника ознак або при проведенні певної кількості ітерацій. Серед КЗ, які базуються на особливостях етапу оцінки, виділяють ті, що активуються у випадку, коли додавання/видалення ознак не призводить до формування кращого словника.

Процедура перевірки валідності методу СО не входить в процес оптимізації словника безпосередньо, але обов'язково застосовується на практиці. Вона полягає в перевірці словника за допомогою тестів та критеріїв відмінних від тих, що використовувалися на етапі оцінки, та порівняння результатів роботи методу з результатами відомими раніше або отриманими за допомогою інших конкуруючих методів.

Таким чином, при аналізі методів СО достатньо розглянути 25 категорій, що характеризуються одним з трьох типів процедури формування та одним з п'яти типів функції оцінки. У табл. 1.1 подано найбільш відомі методи з кожної категорії (якщо такі існують):

Таблиця 1.1 – Класифікація методів СО за їх архітектурою

		Функція оцінки					
		Дистанційні	Інформаційні	Кореляційні	Сумісні	Точосні	
Процедура формування	Оптимальні	МГГ, МГГ+, ВFF, Бобров- ській88	ММДО*	–	FOCUS,** Шліммер93, MIFES_1	Ічино- Скланські84а, Ічино- Скланські84б, АМВ&В	
	Евристичні	Єдиний розв'язок	** Relief, Relief-E, Relief-F, Сеген84	МДР***, Коллер- Сахамі96 *	ПП+СКК*, PRESET**	RSAR**, QuickReduct *	**ПВС, ПСС, ПВСС, ПССС СПСС, рqПП, ДНП, ПС, ПРК, метод Квейроса- Гелсема
		Сукупність розв'язків	–	–	–	–	Метод променевого пошуку, BFS, FSL
	Стохастичні	Єдиний розв'язок	–	–	–	* Фільтри Лас-Вегаcy	ВЛВ, Випадковий пошук, СФПС, ПЕВМ–OP1
		Сукупність розв'язків	–	–	–	–	Генетичні алгоритми

Найбільш відомий серед оптимальних дистанційних методів СО є метод гілок та границь (МГГ), запропонований в праці [33]. Він використовується для більш швидкого пошуку оптимального словника ознак ніж метод повного перебору. Як функція оцінки в МГГ найчастіше застосовується: кодова відстань Махалобіса [35], критерій Фішера, кодова відстань Бхаттачарія та дивергенція [33]. Недоліком цього методу є те, що функція оцінки повинна бути монотонною, тобто поточний словник не повинен бути більш оптимальним ніж будь-який більший за потужністю словник, що включає поточний. Але через існування проблеми

багатовимірності ця умова може не виконуватися для вибірок малого обсягу. Крім того МГГ недоцільно використовувати для задач з дуже великим словником ознак, тому що обчислювальні витрати для цього алгоритму зростають експоненційно.

Подальший розвиток цього методу призвів до появи великої кількості модифікацій, що намагалися тим чи іншим способом позбавитися вищезазначених недоліків. Так, модифікація BFF [37] використовує стратегію інформованого пошуку по першому найліпшому збігу на зваженому дереві, яка гарантує знаходження оптимального словника без повного перебору для будь-якої монотонної функції оцінки.

У праці [38] подавалась модифікація алгоритму Нарендри і Фукунаги, яка має назву МГГ+ (BAB+), і доводилося, аналітично і експериментально, що вона дає кращі результати ніж оригінальний алгоритм. Їх модифікація істотно розрізняла всі “строкові піддерева” (такі піддерева, що складаються з повного шляху від кореню до кінцевого вузла) і не проводила пошук по відповідним кінцевим вузлам, пропускаючи їх і таким чином зменшуючи обчислювальні витрати.

У праці [39] вказувалося на те, що для певних дистанційних мір МГГ працює навіть тоді, коли цільова функція не є монотонною. У праці [40] доведено можливість створення універсальної монотонної функції оцінки для МГГ, на прикладі коефіцієнта однорідності ϕ_k^* для визначення ступеня лінійної залежності між будь-якими дистанційними мірами, що дозволило додати процедуру повернення до МГГ та його модифікацій.

До наступної групи оптимальних інформаційних методів СО відноситься метод мінімальної довжини опису (ММДО) (Minimum Description Length Method, MDLM) [41]. Він намагається видалити ознаки, що не використовуються (надлишкові або неінформативні) з початкового словника. Згідно з цим методом, якщо ознаки із словника V можуть бути подані у вигляді фіксованої незалежної від класу розпізнавання функції F від ознак іншого словника U , тоді при відомих значеннях ознак із словника U

ознаки із словника V є зайвими. Таким чином, задача СО за ММДО полягає в розділенні словників U і V , що разом складають початковий словник. Для її розв'язання використовують критерій мінімальної дожини опису. Формується вираз, який можна інтерпретувати як кількість біт необхідних для передачі класів, оптимальних параметрів, ознак, що використовуються та не використовуються. Алгоритм повністю перебирає всі можливі варіанти словників (2^N) та обирає такий, що задовольняє вищезазначеному критерію. Даний метод здатен визначити всі ознаки, що використовуються, тільки для Гаусівських спостережень.

До оптимальних сумісних методів можна віднести такі три методи:

1) Метод FOCUS [42], відноситься до вкладених алгоритмів СО та використовує висхідну селекцію (пряму селекцію) для знаходження словника ознак мінімальної потужності, що зберігає сумісність початкових даних. Даний алгоритм базується на припущенні про достатню сумісність початкових даних, що не завжди можна гарантувати на практиці. Крім того, обчислювальні витрати FOCUS у випадку великої потужності оптимального словника досить суттєві. Для їх зменшення використовують процедуру, що полягає в застосуванні двох різних процедур оптимізації словника ознак за FOCUS одночасно, один з яких, наприклад, використовує пряму селекцію, а інший – пошук в глибину для здійснення зворотної селекції (виключення ознак з поточного словника). Інформативність ознак, що визначається в процесі селекції за FOCUS, може приймати два значення (0 – неінформативна ознака, 1 – інформативна ознака).

2) Метод Шліммер93 [34] використовує схему направленої перебору в процедурі формування та критерій несумісності в функції оцінки. Для прискорення пошуку використовується гіпотеза про те, що ймовірність виявлення несумісності пропорційна кількості ознак, що будуть визначені як інформативні відносно поточного словника, та твердження про те, що всі об'єднання ненадійних з точки зору несумісності словників є також ненадійними.

3) Метод MIFES_1 [43] подає навчальну вибірку у вигляді матриці, кожний елемент якої представляє собою неповторювану комбінацію позитивних та негативних реалізацій класу. Ознака f називається покриттям елементу матриці, якщо вона приймає протилежні значення для позитивних і негативних реалізацій відповідного елемента. Оптимізація полягає в пошуку покриття з $N-1$ ознак, починаючи з однієї з усіх N ознак, та повторюючи цей процес доти, доки можлива подальша редукція розміру покриття. Недоліком даного методу є те, що він здатен обробляти лише бінарні навчальні матриці та логічні ознаки.

До групи оптимальних точнісних слід віднести два методи, запропоновані Ічино та Скланські, що базуються на двох різних класифікаторах: лінійному [44] та матричному [45]. Для оптимізації словника ознак вони використовують методи двійкового цілочисельного програмування [46]. До даної групи методів СО слід віднести також і наближений монотонний МГГ (*AMB&B (Approximate Monotonic Branch and Bound)*), що є однією з модифікацій МГГ, направлених на подолання його недоліків, пов'язаних з використанням немонотонних функцій оцінки.

Евристично-дистанційні методи базуються на методі RELIEF [47]. Він є вкладеним методом СО, що використовує статистичні методи для оптимізації словника та вагову методику визначення інформативності ознак. Метод складається з декількох етапів. Спочатку з навчальної матриці RELIEF випадковим чином вибирає вказану кількість реалізацій, для кожної з яких визначається найближча серед всіх реалізацій класу, до якого належить вибрана реалізація, (*nearHit*) та найближча серед всіх реалізацій інших класів (*nearMiss*) з використанням Евклідової кодової відстані. Далі ваги ознак, які спочатку мають нульове значення, корегуються на основі наступного правила: ознака вважається більш інформативною, якщо вона збільшує відстань між вибраною реалізацією і *nearMiss*, та менш інформативною, якщо вона збільшує відстань між вибраною реалізацією і *nearHit*. Після перегляду всіх вибраних реалізацій, проводиться селекція ознак, вага яких перевищила

задану межу. Дана межа може бути подана у вигляді функції, що залежить від кількості відібраних ознак, або визначатися константою (наприклад, відбираються всі ознаки, ваги яких більше нуля). Обчислювальні витрати даного методу лінійно залежать від кількості ознак. Його головним недоліком є те, що RELIEF не вирішує проблеми дублюючих ознак, тобто не оптимізує словник за потужністю. Це вирішується додаванням процедури повного перебору всіх комбінацій з ознак, що були відібрані на основних етапах RELIEF. Крім того даний метод здатен оперувати лише з ознаками логічного типу. Дві модифікації RELIEF-E [48] та RELIEF-F [49] даного методу можна вважати вдалимими спробами подолання вищезазначених недоліків.

Метод Сеген84 [50] використовує функцію оцінки, що базується на сумі статистичних різниць та складності ознак (в бітах). Він визначає ознаку, що відповідає найбільшому значенню даного критерію, та ітеративно додає по одній ознаці, яка разом з вже відібраними ознаками підвищує це значення. Даний процес зупиняють при досягненні мінімального репрезентативного значення критерію.

До групи евристичних інформаційних методів СО можна віднести такі два методи:

1) Метод дерева рішень (МДР) [51] використовує входження ОР в усічене дерево рішень в якості критерію селекції. Тобто, словник формується як об'єднання множин ознак, які були задіяні на шляху до будь-якої кінцевої вершини в усіченому дереві.

2) Метод Коллера та Сахамі [52] базується на припущенні, що ознака, яка містить невелику кількість або зовсім не містить додаткової інформації, повинна бути видалена зі словника. Для реалізації даного механізму було застосовано апроксимацію покриття Маркова. (Словник T є Марковим покриттям ознаки f_i , якщо для даного T f_i є умовно незалежною як за номером класу, так і за всіма ознаками, що не входять в T). Хоча даний метод

є теоретично оптимальним, але характеризується значною обчислювальною трудоемністю.

Згідно з працею [53] ймовірність помилки та середній коефіцієнт кореляції (ПП+СКК) є найбільш вдалим поєднанням серед інших оцінок словника ознак серед евристичних кореляційних методів СО. Метод, що використовує дані оцінки, обирає таку першу ознаку, що має найменше значення ймовірності помилки (P_e). Наступна ознака, яку включають до складу словника, відповідає мінімуму зваженої суми P_e та середнього коефіцієнта кореляції між даною ознакою та ознаками, що вже увійшли до словника раніше і так далі. Цей метод дозволяє ранжувати всі ознаки за значенням зваженої суми та створити словник потрібної потужності.

Метод RSAR (Rough Set-based Attribute Reduction) та QuickReduct [54] належать до групи алгоритмів, розроблених на базі концепції застосування евристичних процедур формування словника та сумісних функцій його оцінки. RSAR крім того підсилювався використанням елементів теорії нечітких множин при визначенні додаткової інформаційної складової функції оцінки.

Методи QuickReduct та QuickReduct II (ReverseReduct) – це фільтри прямої та зворотної СО. На відміну від FOCUS методи даної групи не використовують гіпотезу про достатню сумісність початкових даних, а намагаються знайти словник ознак з максимально можливим значенням функції оцінки. Єдиним недоліком даної групи методів є те, що вони направлені на оптимізацію словника без мінімізації його потужності.

Стохастичні сумісні методи СО в повній мірі характеризуються фільтрами Лас-Вегасу (ФЛВ, LVF) [55], які проводять випадковий пошук у просторі можливих словників із застосуванням для прискорення оптимізації алгоритму Лас-Вегасу та використовують сумісні міри у функції оцінки. Для кожного словника ознак визначається рівень несумісності на базі гіпотези про те, що реалізації різних класів, які співпали для даного словника, визначають той клас, реалізацій якого серед них більше. Якщо

несумісність перевищує задане граничне значення, то такий словник відхиляється. Такий метод дозволяє знайти оптимальний словник навіть для випадку, коли навчальна матриця була отримана з перешкодами, якщо спочатку було встановлено приблизний їх рівень. Ефективність даного методу пояснюється тим, що після кожного кроку оптимізації зменшується простір пошуку (Проводиться перевірка на несумісність лише тих словників, потужність яких не перевищує потужності поточного найкращого по сумісності словника). Таким чином одночасно проводиться оптимізація по сумісності та по потужності словників ознак. Недоліком даного методу можна вважати те, що пошук оптимального словника може зайняти більше часу ніж у алгоритмів СО, які використовують евристичні процедури формування, через неможливість використання додаткових обмежень простору пошуку, отриманих на попередніх ітераціях.

Група евристичних точнісних методів СО з єдиним розв'язком в процесі оптимізації оперує тільки одним поточним словником. На кожній ітерації додаються або видаляються ознаки, доки одержаний словник не буде задовольняти певному критерію зупину. Ці методи, відомі також як “послідовні”, найбільш часто використовується для формування та оптимізації словника ознак. Вони поділяються на дві категорії: такі, що починають з пустого словника і додають ознаки (послідовна “пряма”, “висхідна”(bottom-up) або “направлена вперед” (forward) селекція) (ПВС), і такі, що видаляють ознаки з повного вихідного словника (послідовна “зворотна”, “спадна” (top-down) та “направлена назад” (backward) селекція) (ПСС) [26]. Недоліком цих методів є те, що без перевірки всіх можливих підмножин вихідного словника вони не гарантують оптимального результату через проблему вкладеності, коли ознака, що була внесена до словника не може бути видалена в процесі оптимізації. У праці [56] проведено порівнювальний аналіз цих методів, на базі чого було розроблено оптимальний метод гілок та границь, що використовує штучні навчальні та екзаменаційні вибірки, сформовані на базі Гаусовського розподілу, для двох

класів. У праці [57] розроблено два методи “спрямованої” оптимізації словника ознак: послідовної висхідної спрямованої селекції (ПВСС) (*sequential forward floating selection (SFFS)*) та послідовної спадної спрямованої селекції (ПССС) (*sequential backward floating selection (SBFS)*), що характеризуються складністю та часом обчислень більшими ніж ПВС і ПСС, але вирішують проблему вкладеності завдяки застосуванню пошуку з поверненням.

Крім модифікацій, спрямованих на вирішення проблеми вкладеності існують також такі, що підвищують оперативність: скорочена послідовна спадна селекція (СПСС) [58], (p,q) послідовний пошук (РQПП), двонаправлений пошук (ДНП) [32], схематичний пошук (СП) [59], пошук по релевантності в контексті (ПРК) [60] і метод Квейроса-Гелсема [61]. СПСС використовує твердження, що при досить значних потужностях словника ознак, певні алгоритми класифікації не використовують більшу їх частину. Як і ПСС даний метод розпочинає оптимізацію з повного початкового словника, але після проведення кроку навчання видаляє будь-яку ознаку, що не приймала участь в ньому. ДНП проводить одночасно процедуру як висхідної так спадної селекції, а РQПП певним чином обмежує як процес додавання та і видалення ознак у відповідних процедурах ДНП. СП на кожній ітерації зменшує/збільшує потужність поточного словника на 1. Для визначення оптимального словника на поточній ітерації оцінюються всі варіанти таких словників, кількість яких дорівнює потужності поточного словника. Оптимізація зупиняється, якщо додавання/видалення однієї ознаки з поточного словника не призводить до поліпшення значення функції оцінки. ПРК базується на твердженні, що деякі ОР будуть релевантні тільки в певній частині простору пошуку. Таким чином, визначення інформативності кожної ознаки проводиться не глобально для всієї навчальної матриці, як в ПСС, а локально для певних реалізацій або класів. Метод Квейроса-Гелсема є модифікацією ПВС, що при додаванні ознак враховує різницю між оцінками інформативності ознак, які розраховуються з використанням певних

припущень про взаємозв'язки між даними ознаками і поточним словником. Даний метод використовує припущення про незалежність ознаки, тобто не враховує вплив словника, та навпаки. Як функція оцінки застосовується коефіцієнт помилки байєсівського класифікатора.

Евристичні точнісні методи СО здатні оперувати і сукупністю розв'язків. У статті [31] розглядається метод пошуку по першому найкращому збігу у просторі підмножин ознак, як окремий випадок методу “променевого пошуку”. Обидва ці методи оперують послідовністю можливих розв'язків.

Дані методи мають загальну назву методи “гратки СО” (*feature selection lattice*), де кожний вузол – це окрема підмножина ознак, а ребро – локалізація взаємозв'язків між ними. Таким чином, вони представляють простір підмножин у вигляді графа, а потім застосовують один з багатьох стандартних алгоритмів пошуку на графі. Але ці методи не знайшли широкого застосування і практично не зустрічаються в літературі.

Найбільш відомі точнісні стохастичні методи з єдиним розв'язком базуються на випадковому пошуку та його модифікаціях. За цими методами випадковим образом відбираються і l ознак, і оцінюється інформативність цієї підсистеми. Далі знову незалежно від попереднього набору випадково формується інша система з l ознаками. Так повторюється h раз. Із h наборів ознак відбирається той, який має найбільшу інформативність. Чим h більше, тем вище ймовірність відбору оптимальної підсистеми. При $h \geq 2$ можна стверджувати, що система не є неоптимальною (якщо обрані підсистеми не є однаковими за інформативністю). Модифікацією даного методу є випадковий пошук з адаптацією. Це послідовно направлена процедура, яка базується на випадковому пошуку з врахуванням результатів попередніх відборів. На початку процедури шанси всіх вихідних ознак на входження в підсистему, що складається з $K > 1$ ознак, приймаються рівними. Для випадкового відбору використовується датчик рівномірно розподілених на інтервалі $[0;1]$ випадкових (псевдовипадкових) чисел. Крім того базовий алгоритм

випадкового пошуку досить часто поєднується з методами, як належать до інших груп. Розглянемо деякі вдалі модифікації, отримані таким чином.

ВЛВ [55] представляє собою спробу застосування алгоритму Лас-Вегасу до вкладених методів СО, подібно ФЛВ до фільтрів. Основною відмінністю ВЛВ від свого аналогу є застосування точнісних критеріїв у функції оцінки, що є характерною рисою вкладених методів СО. Нажаль дана модифікація практично не усуває основні недоліки ФЛВ.

СФПС (Стохастичне формування плюс послідовна селекція) [32] підсилює алгоритми ПСС або ПВС процедурою рандомізації. Даний метод випадковим чином формує словник ознак, який використовується як початковий замість повного або порожнього словника відповідно.

Пошук екстремуму функції оцінки з використанням випадкової мутації (алгоритми ПЕВМ–ОР1 і RMHC-PF1 [62]) здійснюється з метою селекції реалізацій та ознак, поданих у вигляді бінарного вектору. Як функції оцінки алгоритм ПЕВМ–ОР1 використовує точнісні характеристики, обчислені за методом найближчого сусіда. На кожній ітерації проводиться випадкова мутація компонентів бінарного вектору для формування наступного варіанту словника ознак.

У праці [63] вказується на можливість застосування генетичних алгоритмів (ГА) для формування словника ознак, а отже виділення групи точнісних стохастичних методів СО з сукупністю розв'язків. У ГА вихідна множина ознак представляється у вигляді бінарного вектору (хромосоми) довжиною n . Значення i -ої компоненти хромосоми приймають значення 1 або 0 в залежності від того входить відповідна ознака до словника чи ні. Оперують популяцією хромосом, в якій і проводиться селекція у відповідності зі значенням функції пристосованості. Нові хромосоми створюються з існуючих за допомогою одноточкового або багатоточкового схрещування (*crossover*) – обміну частинами хромосом між двома або більше хромосомами популяції, або мутації – стохастичної зміни частини хромосоми.

Досить складно порівняти ГА з іншими методами оптимізації словника ознак. Основна відмінність від послідовних методів полягає в тому, що ГА не намагається знайти оптимальний словник визначеного розміру, а використовує всі можливі комбінації ознак. Наприклад, результати порівняння ГА з послідовними методами (“висхідним” та “спадним”) і модифікацією МГГ, для випадку, коли цільова функція не є монотонною, вказують на перевагу ГА як за класифікаційними характеристиками, так і за обчислювальними витратами [63]. Результати аналогічних порівнянь ГА з ПВС і ПВСС у праці [64] вказують на залежність класифікаційних характеристик ГА від розмірності вихідного словника.

Існує окрема група методів СО, які дуже важко віднести до будь-якої з вищезазначених груп. У праці [65] розроблено алгоритм оптимізації словника ознак для навчання багаторівневих мереж зворотного поширення з механізмом прогнозування подій [66]. Вони визначили критерій “виразності вузлу” (*node saliency*) і представили процедуру скорочення найменш виражених вузлів для зменшення складності ШНМ після її навчання. Скорочення вихідних нейронів еквівалентно видаленню відповідних ознак із словника. Метод скорочення кількості вузлів одночасно оптимізує як словник ознак, так і класифікатор.

При навчанні ШНМ застосовується квадратична функція втрат. Виразність вузлу визначається як сума приросту втрат по всій навчальній матриці, що має місце після видалення вузлу. У праці [66] виразність вузлу апроксимується у вигляді поліному другого порядку і потім обчислюється конкретне значення через відповідні похідні у режимі зворотного поширення. Оскільки обчислення виразності для великих ШНМ характеризується істотними обчислювальними витратами, то використовують метод зворотного поширення, що дає змогу проводити оптимізацію за один крок (замість кроку на вузол).

Таким чином алгоритм скорочення кількості вузлів складається з двох етапів: а) навчання ШНМ вперше і видалення вузлу з найменшим значенням

виразності; б) перенавчання ШНМ без видаленого вузла. Цей процес повторюється, доки не буде досягнуто бажаного взаємовідношення між помилкою класифікації та розмірами ШНМ.

Методи ЗР і СО досить рідко використовуються самостійно. Зазвичай, вони разом з відповідними алгоритмами навчання або екзамену входять до складу систем автоматичної класифікації. Виділяють три класи таких систем:

1) Клас методів, що відокремлюють етап оптимізації словника ознак від етапу навчання системи і використовують його як підготовчий. Хоча такий спосіб інтеграції і не є особливо ефективним, але його вважають найбільш доречним для підвищення оперативності та покращення точнісних характеристик систем, що характеризуються значною потужністю словника [30]. Головною перевагою цього класу є гнучкість, а головним недоліком – невелике, порівняно з іншими класами, поліпшення точнісних характеристик. Найбільш вдалим прикладами, що характеризують такі системи, можна вважати практичне застосування алгоритму Relief та його модифікацій [69,98].

2) Клас методів, що використовують методи оптимізації СО безпосередньо в процесі навчання. У цьому випадку замість фільтрів застосовують вкладені методи, для них характерною рисою є те, що процес навчання повторюється, доки не буде досягнуто достатнього або максимально можливого значення точнісних параметрів системи. Це і є основною перевагою методів цього класу. Серед їх недоліків слід зазначити значний час навчання, що унеможлиблює практичне використання в системах, які мають навчатися або перенавчатися в реальному часі. Крім того, гнучкість та універсальність дещо зменшується через необхідність створення певних симбіотичних відносин між алгоритмом оптимізації словника ознак та алгоритмом навчання [30]. Як приклад систем цього класу можна назвати практичне застосування методів FOCUS [38,58].

3) Клас методів, для яких не можливо відділити оптимізацію словника ознак від процесу навчання. Канонічним прикладом таких систем вважається

застосування методу дерева рішень [67], у процесі побудови якого використовуються стільки ознак, скільки необхідно для виявлення понять, що закладені в апріорно надлишкової початковій вибірці. Іншим прикладом можна вважати БАРС [29] або системи лінгвістичної класифікації [68].

1.3 Постановка задачі інформаційного синтезу СППР, що самонавчається

Розглянемо бінарний простір ознак розпізнавання Ω_B , який є підмножиною простору Хеммінга з потужністю $Card \Omega_B = 2^N$, де N – кількість ознак розпізнавання. При класифікаційному аналізі i -та ознака X_i розглядається як випадкова величина, значення якої утворюють на генеральній сукупності повторну вибірку $\{x_i^{(j)} \mid j = \overline{1, n}\}$ обсягу n . Тоді множину відбитих властивостей m -го функціонального стану системи керування і відношень між її елементами, яку можна задати в просторі ознак розпізнавання визначеною певним чином областю, будемо вважати класом розпізнавання (образом) $X_m^o, m = \overline{1, M}$. Множина таких класів розпізнавання $\{X_m^o \mid m = \overline{1, M}\}$ формує алфавіт класів розпізнавання.

Детерміновано-статистичний підхід до прийняття рішень вимагає завдання як системи нормованих (експлуатаційних), так і контрольних допусків на ознаки. Нехай X_1^o – базовий клас, який характеризує максимальну функціональну ефективність здатної навчатися АСК, тобто є найбільш бажаним для особи, що приймає рішення (ОПР). Тоді нормованим є поле допусків $\{\delta_{H,i} \mid i = \overline{1, N}\}$, в якому значення ознаки X_i знаходиться з імовірністю $p_i = 1$ або $p_i = 0$, за умови, що функціональний стан АСК відноситься до базового класу X_1^o , а контрольним – поле допусків

$\{\delta_{K,i} | i = \overline{1, N}\}$, в якому значення ознаки X_i знаходиться з імовірністю $0 < p_i < 1$ за умови, що функціональний стан характеризується класом X_1^o .

В ІЕІ-технології контрольні допуски на ознаки вводяться з метою рандомізації процесу прийняття рішень, оскільки для повноти дослідження необхідно використовувати як детерміновані, так і статистичні характеристики керованого процесу. Областю визначення системи контрольних допусків (СКД) є відповідна система нормованих допусків.

Як вектор-реалізацію образу $x_m^{(j)} \in X_m^o$ будемо розглядати двійковий випадковий структурований вектор, що є j -м рядком бінарної навчальної матриці $\|x_{m,1}^{(j)}\|$, $m = \overline{1, M}$, $i = \overline{1, N}$, $j = \overline{1, n}$:

$$x_m^{(j)} = \langle x_{m,1}^{(j)}, \dots, x_{m,i}^{(j)}, \dots, x_{m,N}^{(j)} \rangle, \quad j = \overline{1, n_{\min}},$$

де $x_{m,i}^{(j)}$ – i -та координата вектора, яка приймає одиничне значення, якщо значення ознаки X_i знаходиться в полі допусків $\delta_{K,i}$, і нульове значення, якщо не знаходиться; n_{\min} – мінімальна кількість випробувань, яка забезпечує репрезентативність навчальної вибірки.

Оскільки при нормальному розподілі реалізацій образу виправданою є гіпотеза компактності (чітка або нечітка) реалізацій образу, то в ІЕІ-технології введено поняття “контейнер” [132], який є апроксимованим наближенням “точної” складної замкненої роздільної гіперповерхні класу розпізнавання, яке відновлюються на кожному кроці навчання в радіальному базисі простору ознак розпізнавання вигляді правильної геометричної фігури. При цьому геометричний центр контейнера вважається визначеним будь-яким способом. Одним із способів визначення геометричного центру контейнеру класу X_m^o є формування еталонного вектора-реалізації образу

x_m , який є математичним сподіванням випадкових векторів-реалізацій $\{x_m^{(j)}\}$ класу X_m^o .

Структура двійкового еталонного вектору класу X_m^o , має вигляд

$$x_m = \langle x_{m,1}, \dots, x_{m,i}, \dots, x_{m,N} \rangle, \quad m = \overline{1, M},$$

де $x_{m,i}$ – i -та координата вектора, яка приймає одиничне значення, якщо значення ознаки X_i знаходиться в нормованому полі допусків $\delta_{H,i}$, і нульове значення, якщо не знаходиться.

Таким чином, для класу X_m^o контейнер $K_m^o \subset X_m^o$ у рамках ІЕІ-технології є його “прозорою” оболонкою. При цьому “прозорість” цієї оболонки полягає у тому, що при побудові нечіткого розбиття простору ознак на класи контейнери можуть перетинатися.

Ефективність функціонування АСК залежить від її просторово-часових параметрів функціонування – характеристик інформаційного забезпечення, які впливають на функціональну ефективність системи. Параметри функціонування, які оптимізуються в процесі навчання, будемо називати параметрами навчання. Як критерій оптимізації процесу навчання в рамках ІЕІ-технології застосовується будь-який статистичний інформаційний КФЕ, який є природною мірою різноманітності класів розпізнавання і одночасно функціоналом від точнісних характеристик розподіленої інтегрованої АСК, що самонавчається. Важливими параметрами функціонування АСК є параметри словника ознак. Оптимізація його параметрів у рамках ІЕІ-технології базується на понятті інформативності як окремої ознаки, так і їх груп. При цьому інформативність розглядається як ступінь впливу ознаки на критерій оптимізації процесу навчання.

У рамках ІЕІ-технології можна виділити чотири основні групи ознак:

- інформативні, наявність в словнику яких збільшує значення КФЕ;

- неінформативні, що не впливають на функціонування СППР, що самонавчається;
- заважаючі, або дезінформуючі, включення яких в словник ознак призводить до втрати ефективності при навчанні СППР;
- латентні, «схованість» яких обумовлена малою частотою їх появи, що не перевершує вибраний поріг прийняття рішень.

Розглянемо постановку задачі інформаційно-екстремального синтезу СППР, здатної самонавчатися в режимі ФКА.

Нехай відомий апріорний алфавіт у загальному випадку нечітких класів розпізнавання $\{X_m^o \mid m = \overline{1, M}\}$, який характеризує M функціональних станів технологічного процесу, навчальна матриця типу «об'єкт-властивість» $\|y_{m,i}^{(j)}\|, i = \overline{1, N}, j = \overline{1, n}$, де N, n – кількість ознак розпізнавання і випробувань відповідно і апріорний у загальному випадку надлишковий в інформаційному розумінні словник ознак $\Sigma^{[N]}$. Нехай базовий клас X_1^o характеризує найбільш бажаний функціональний стан технологічного процесу. Необхідно:

1) Для апріорно класифікованого нечіткого розбиття $\tilde{\mathfrak{R}}^{|M|}$ побудувати шляхом допустимих перетворень в субпарацептуальному дискретному просторі ознак Ω_B оптимальне (тут і далі в роботі в інформаційному розумінні) чітке розбиття еквівалентності класів $\mathfrak{R}^{|M|}$:

$$\left(\forall X_m^o \in \tilde{\mathfrak{R}}^{|M|} \right) \left[X_m^o \neq \emptyset, m = \overline{1, M} \right]; \quad (1.3.1)$$

$$\left(\exists X_m^o \in \tilde{\mathfrak{R}}^{|M|} \right) \left(\exists X_c^o \in \tilde{\mathfrak{R}}^{|M|} \right) \left[X_m^o \neq X_c^o \rightarrow X_m^o \cap X_c^o \neq \emptyset, m, c = \overline{1, M} \right]; \quad (1.3.2)$$

$$\left(\forall X_m^o \in \tilde{\mathfrak{R}}^{|M|} \right) \left(\forall X_c^o \in \tilde{\mathfrak{R}}^{|M|} \right) \left[X_m^o \neq X_c^o \rightarrow \text{Ker} X_m^o \cap \text{Ker} X_c^o = \emptyset \right]; \quad (1.3.3)$$

$$\left(\forall X_m^o \in \mathfrak{R}^{|M|} \right) \left(\Sigma_M^{(i)} \in \Omega_B \right) \left(E_m \in G_E \right) \left[\text{if } E_m^* = \max_{\{k\}} E_m \text{ then } \Sigma_M^* := \Sigma_M^{(i)}, i = \overline{1, N} \right]; \quad (1.3.4)$$

$$\left(\forall X_m^o \in \tilde{\mathfrak{R}}^{|\Lambda|} \right) \left(\forall X_c^o \in \tilde{\mathfrak{R}}^{|\Lambda|} \right) \left[X_m^o \neq X_c^o \rightarrow (d_m^* < d(x_m \oplus x_c)) \& \right. \\ \left. \& (d_c^* < d(x_m \oplus x_c)) \right]; \quad (1.3.5)$$

$$\bigcup_{X_m^o \in \tilde{\mathfrak{R}}} X_m^o \subseteq \Omega_B, \quad (1.3.6)$$

де $KerX_m^o$ – ядро класу X_m^o ; $KerX_l^o$ – ядро класу X_l^o , найближчого сусіднього до класу X_m^o ; $\Sigma_M^{(i)}$ – поточний словник ознак для апріорного алфавіту класів, який містить i ознак, $i = \overline{1, N}$; E_m – інформаційний КФЕ навчання СППР; G_E – область значень КФЕ; Σ_M^* – оптимальний словник ознак для M класів; d_m^* – оптимальний радіус контейнера $K_m^o \subset X_m^o$; $d(x_m \oplus x_c)$ – міжцентрова кодова відстань для класів X_m^o і X_c^o ; d_c^* – оптимальний радіус контейнера $K_l^o \subset X_c^o$.

2) На етапі навчання СППР в режимі ФКА побудувати шляхом допустимих перетворень в субпарацептуальному дискретному просторі ознак Ω_B оптимальне відкрите чітке розбиття еквівалентності класів $\mathfrak{R}^{|\Lambda|}$, $\Lambda > M$, за умов:

$$\left(\forall X_m^o \in \tilde{\mathfrak{R}}^{|\Lambda|} \right) \left[X_m^o \neq \emptyset, m = \overline{1, \Lambda} \right]; \quad (1.3.7)$$

$$\left(\exists X_m^o \in \tilde{\mathfrak{R}}^{|\Lambda|} \right) \left(\exists X_c^o \in \tilde{\mathfrak{R}}^{|\Lambda|} \right) \left[X_m^o \neq X_c^o \rightarrow X_m^o \cap X_c^o \neq \emptyset, c = \overline{1, \Lambda} \right]; \quad (1.3.8)$$

$$\left(\forall X_m^o \in \tilde{\mathfrak{R}}^{|\Lambda|} \right) \left(\forall X_c^o \in \tilde{\mathfrak{R}}^{|\Lambda|} \right) \left[X_m^o \neq X_c^o \rightarrow KerX_m^o \cap KerX_c^o = \emptyset \right]; \quad (1.3.9)$$

$$\left(\forall X_m^o \in \mathfrak{R}^{|\Lambda|} \right) \left(\Sigma_\Lambda^{(i)} \in \Omega_B \right) \left(E_m \in G_E \right) \left[\text{if } \bar{E}^* = \max_{\{k\}} \bar{E} \text{ then } \Sigma_\Lambda^* := \Sigma_\Lambda^{(i)}, i = \overline{1, N} \right], \quad (1.3.10)$$

де $\Sigma_\Lambda^{(i)}$ – поточний словник ознак для нового алфавіту класів $\{X_m^o\}^{|\Lambda|}$,

який містить i ознак, $i = \overline{1, N_{\Lambda-1}^*}$; $\bar{E} = \frac{1}{\Lambda} \sum_{m=1}^{\Lambda} \sum_{i=1}^N E_{m,i}$ – усереднене значення

КФЕ навчання СППР; Σ_Λ^* – оптимальний словник ознак для Λ класів;

$$\left(\forall X_m^o \in \tilde{\mathfrak{R}}^{|\Lambda|} \right) \left(\forall X_c^o \in \tilde{\mathfrak{R}}^{|\Lambda|} \right) \left[X_m^o \neq X_c^o \rightarrow (d_m^* < d(x_m \oplus x_c)) \& \right. \\ \left. \& (d_c^* < d(x_m \oplus x_c)) \right], \quad (1.3.11)$$

$$\bigcup_{X_m^o \in \tilde{\mathfrak{R}}} X_m^o \subseteq \Omega_B. \quad (1.3.12)$$

У виразі (1.3.9) $E_{m,i}$ – значення КФЕ навчання системи розпізнавати реалізації класу X_m^o , $m = \overline{1, \Lambda}$, для поточного словника ознак $\Sigma_{\Lambda}^{(i)}$.

3) на етапі екзамену, тобто безпосередньо у робочому режимі СППР, оцінити поточний функціональний стан керованого технологічного процесу і при зміні потужності алфавіту класів сформувати репрезентативну навчальну матрицю $\|x_{m,i}^{(j)}\|_{m=1, \Lambda; i=1, N_{\Lambda-1}^*; j=\overline{1, n}}$, де $N_{\Lambda-1}^*$ – потужність попереднього оптимального словника $\Sigma_{\Lambda-1}^*$.

4) перенавчити СППР для алфавіту класів $\{X_m^o\}^{|\Lambda|}$ з оптимізацією словника ознак.

4) за умови невідповідності поточного функціонального стану технологічного процесу класу X_1^o реалізувати оператор корекції для стабілізації АСКТП.

Таким чином, специфіка задачі самонавчання СППР у рамках ІЕІ-технології полягає в поєднанні задачі ФКА з оцінкою інформативності ознак та оптимізації параметрів словника ознак шляхом багатоциклічної структурованої за параметрами функціонування ітераційної процедури пошуку глобального максимуму інформаційного КФЕ самонавчання в робочій (допустимій) області визначення його функції.

РОЗДІЛ 2

АНАЛІЗ І СИНТЕЗ ІНФОРМАЦІЙНО-ЕКСТРЕМАЛЬНОЇ СППР, ЩО САМОНАВЧАЄТЬСЯ

2.1 Інформаційна модель слабо формалізованого технологічного процесу

Інформаційну модель складного технологічного процесу у загальному випадку показано у вигляді субблоку на рис.2.1

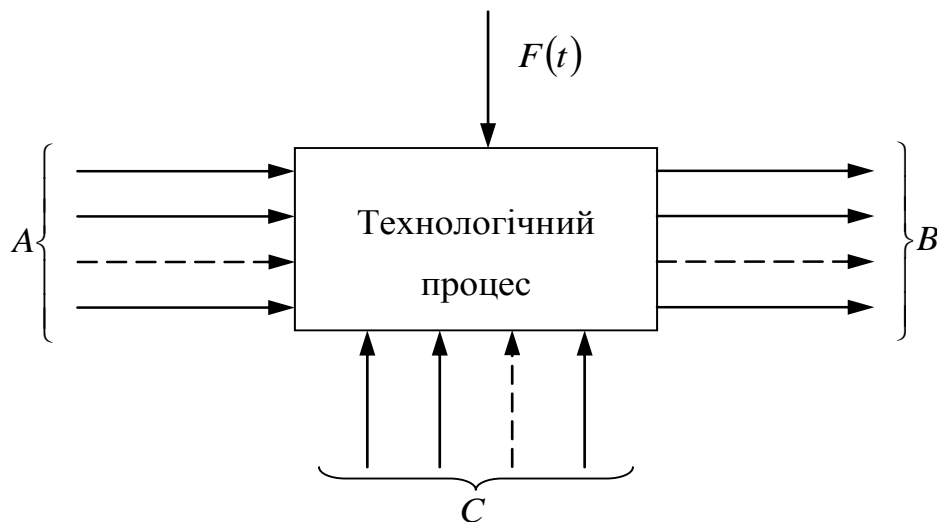


Рисунок 2.1 – Інформаційна модель слабо формалізованого технологічного процесу

Функціональний стан слабо формалізованого технологічного процесу характеризується вхідними векторами параметрів, що керуються, $A = \{a_i\}$, параметрів, що не керуються, $C = \{c_p\}$, вектором вихідних змінних $B = \{b_j\}$ і вектором випадкових збурень $F(t)$. Параметри векторів A, B, C є ознаками, що контролюються. При цьому вектор C може суттєво впливати на функціональну ефективність АСКТП, знижуючи достовірність керуючих рішень, особливо за умов відсутності через технічні ускладнення неперервного вхідного контролю, наприклад для сировини та матеріалів, що мають природне походження.

Основною характеристикою інформаційної моделі об'єкту автоматизації є кількість середньої умовної інформації про вектор A (або B), що міститься у векторі B (або A), яка визначається за відомою симетричною формулою [43,126]

$$I_B(A) = H(A) - H(A/B) = H(B) - H(B/A) = I_A(B),$$

де $H(A)$ – середня безумовна ентропія входу АСКТП; $H(B)$ – середня безумовна ентропія виходу АСКТП; $H(A/B)$ – апостеріорна умовна ентропія входу АСКТП; $H(B/A)$ – апіорна середня умовна ентропія виходу АСКТП.

Розглянемо інформаційний опис керованого технологічного процесу для загального випадку, тобто із урахуванням вектора $C = \{C_p\}$. Тоді кількість середньої умовної інформації про вектори A і B відносно один одного при наявності на вході вектора C буде дорівнювати

$$I_C(A, B) = H(A/C) - H(A/BC) = H(B/C) - H(B/A, C) = I_C(B, A). \quad (2.1.1)$$

При цьому кількість інформації $I_A(B)$ можна виразити через імовірності подій на вході і виході об'єкту автоматизації:

$$I_C(A, B) = \sum_i \sum_j p(a_i, b_j / c_p) * \log_2 \frac{p(a_i, b_j / c_p)}{p(a_i / c_p) p(b_j / c_p)}.$$

Оскільки

$$I_{A,C}(B) = H(B) - H(B/A, C), \quad (2.1.2)$$

то відняв з (2.1.2) $I_C(B) = H(B) - H(B/C)$ із урахуванням (2.1.1) одержимо

$$I_{A,C}(B) = I_C(B) + I_C(B) + I_C(A, B). \quad (2.1.3)$$

Із виразу (2.1.3) витікає, що

$$I_{A,C}(B) \geq I_C(B), \quad (2.1.4)$$

тобто кількість інформації, що міститься у векторі B про залежні вектори A і C не менше кількості інформації, що міститься у векторі B про вектор C . При цьому у виразі (2.1.4) має місце рівність за умови, що вектори A і B незалежні, оскільки $I_C(A,B)=0$ і тоді $H(A/C)=H(A/B,C)$.

На практиці може мати інтерес інформаційна модель керованого процесу за умови, що вектори A і C є незалежними. Тоді аналогічно (2.1.2) одержимо вираз

$$I_{B,C}(A) = I_B(A) + I_B(A,C). \quad (2.1.5)$$

Із (2.1.5) при незалежних B і C витікає $I_{B,C}(A) = I_B(A)$ і $I_B(A,C) = 0$.

Таким чином, функціональна ефективність АСК як складова загальної ефективності, визначає згідно з критерієм мети ступінь відповідності функціонування системи її робочому алгоритму досягнення поставленої задачі. При цьому важливою складовою критерію мети є інформаційний критерій функціональної ефективності (КФЕ) навчання системи.

2.2. Основні принципи аналізу та синтезу СППР, що самонавчається в режимі факторного кластер-аналізу

Методи ІЕІ-технології аналізу та синтезу здатної навчатися (самонавчатися) АСК, в яку інтегровано інтелектуальну СППР, ґрунтуються на максимізації інформаційної спроможності системи шляхом введення

додаткових інформаційних обмежень за умов апріорної невизначеності, нечітких даних і ресурсних обмежень. Подальший розвиток ІЕІ- технології дозволив створити низку методів, що доповнюють та розширюють можливості базового методу – методу функціонально-статистичних випробувань (МФСВ) і дозволяють ефективно його використовувати для розв’язання практичних задач автоматизації розподілених в просторі та часі технологічних процесів шляхом автоматичної класифікації їх функціональних станів за умови апріорної невизначеності та формування для користувача-оператора оптимальних керуючих повідомлень.

Основні концептуальні положення інформаційно-екстремального методу аналізу та синтезу СППР, що самонавчається, такі:

- критерій функціональної ефективності (КФЕ) АСК, що самонавчається, безпосередньо пов’язаний з прямою оцінкою інформаційної спроможності системи, яка однозначно визначає функціональну ефективність СППР;

- процес самонавчання проводиться в рамках детерміновано-статистичного підходу і полягає в побудові відносно простих безпомилкових за навчальною багатовимірною навчальною матрицею вирішальних правил, що дозволяють в режимі екзамену, тобто безпосередньо в робочому режимі, одержати повну ймовірність оцінки поточного функціонального стану технологічного процесу, наближену до граничного значення;

- процес самонавчання здійснюється за умови нечіткої компактності реалізацій образу, тобто припускає перетин класів розпізнавання, що має місце в практичних задачах автоматизації розподілених технологічних процесів, і полягає в цілеспрямованій ітераційній багатоциклічній оптимізації просторово-часових параметрів функціонування з метою послідовного наближення глобального максимуму інформаційного КФЕ, що обчислюється в робочій (допустимій) області визначення його функції, до максимально граничного, яке визначає побудову безпомилкового класифікатора;

- через застосування логарифмічних статистичних інформаційних мір, які мають за А.О. Колмогоровим [69] властивість стиснення обсягу вибіркової випадкової послідовності без втрати статистичних закономірностей, допускається можливість використання репрезентативних навчальних вибірок обсяг яких на порядок менше обсягу вибірок для обчислення статистик у багатовимірному статистичному аналізі [70];

- закладені в ІЕІ-технології принципи об'єктно-структурованого проектування дозволяють здійснювати успадкування і довизначення її методів, що активізує їх розвиток в рамках вирішення проблеми інформаційного синтезу широкого класу АСК, що навчаються.

Таким чином, методи ІЕІ-технології в певній мірі проявляють властивість універсальності при проектуванні здатної навчатися АСК, що дозволяє вирішувати як загальну, так і частинні задачі її інформаційного синтезу.

На рис. 2.2 подано основні принципи, на яких ґрунтується інформаційно-екстремальний метод аналізу та синтезу СППР, що самонавчається. Крім того метод ґрунтуються на загально відомих принципах системного підходу і розпізнавання образів [71], а також об'єктно-орієнтованого проектування [72].

Принципи оцінки функціональної ефективності системи керування обґрунтовують доцільність використання інформаційних КФЕ як з логіко-гносеологічної, так і кібернетичної точок зору, визначають основні властивості інформації як міри різноманітності функціональних станів і режимів системи при її використанні за призначенням.

Принципи організації процесу навчання та екзамену деталізують способи подання та накопичення знань в режимі навчання та їх використання в режимі екзамену. Крім того вони обґрунтовують доцільність спрощення структури узагальнюючих дескрипторів (концептів) реалізацій класів та способів їх побудови. У цих принципах закладено основу оптимізації надлишкового словника ознак, сформовано поняття інформативність та

визначено його зв'язок з достовірністю прийняття рішень СППР, що самонавчається.



Рисунок 2.2 – Структура основних принципів інформаційно-екстремального методу аналізу та синтезу СППР, що самонавчається

Принцип апіорної нечіткої компактності векторів-реалізацій класу забезпечує перехід від традиційного апіорного чіткого розбиття класів, що носить модельний характер, до його нечіткого варіанту, притаманного практичним задачам керування складними організаційно-технічними об'єктами та технологічними процесами. Принципи формування навчальної

вибірки дозволяють визначити репрезентативний обсяг таких вибірок, їх статистичну стійкість та однорідність.

За ідеологією ІЕІ-технології допустимі перетворення апріорного нечіткого розбиття простору ознак здійснюються в дискретному субпарацептуальному просторі, що дозволяє:

- підвищити заводо захищеність методів аналізу і синтезу здатних самонавчатися АСК;
- здійснювати цілеспрямовану трансформацію апріорного нечіткого розбиття простору ознак розпізнавання в чітке розбиття класів, що дозволяє побудувати безпомилковий за навчальною вибіркою класифікатор.

Таким чином, ІЕІ-технологія ґрунтується як на відомих принципах автоматичної класифікації та теорії прийняття рішень, так і на специфічних принципах, що відбивають інформаційну природу оптимального функціонування інтелектуальної СППР, що самонавчається.

2.3 Математичні моделі СППР, що самонавчається

При обґрунтуванні гіпотези нечіткої компактності, яка має місце на практиці, розглянемо апріорне нечітке розбиття $\tilde{\mathfrak{R}}^{|M|} \subset \Omega$. Застосуємо оператор θ допустимих перетворень вхідного математичного опису СППР в бінарному просторі ознак Ω_B з метою нечіткої факторизації простору ознак: $\theta: Y \rightarrow \tilde{\mathfrak{R}}^{|M|}$. Нехай оператор класифікації $\psi: \tilde{\mathfrak{R}}^{|M|} \rightarrow I^{|I|}$ перевіряє основну статистичну гіпотезу про належність реалізацій $\{x_m^{(j)} | j = \overline{1, n}\}$ класу X_m^o , де $I^{|I|} = \{\gamma_1, \gamma_2, \dots, \gamma_l\}$ – множина гіпотез. Оператор $\gamma: I^{|I|} \rightarrow \mathfrak{Z}^{|q|}$ шляхом оцінки статистичних гіпотез формує множину точнісних характеристик $\mathfrak{Z}^{|q|}$, де $q = l^2$ – кількість точнісних характеристик. Оператор $\phi: \mathfrak{Z}^{|q|} \rightarrow E$ обчислює множину значень інформаційного КФЕ, який є функціоналом точнісних

характеристик. Контур оптимізації геометричних параметрів розбиття $\tilde{\mathfrak{R}}^{|M|}$ шляхом пошуку максимуму КФЕ навчання розпізнаванню реалізацій класу X_m^o замикається оператором $r: E \rightarrow \tilde{\mathfrak{R}}^{|M|}$. Тоді категорійна модель у вигляді діаграми відображення множин, задіяних в процесі навчання за базовим алгоритмом у рамках ІЕІ-технології, для апріорно нечіткого розбиття має вигляд, показаний на рис. 2.3 [1].

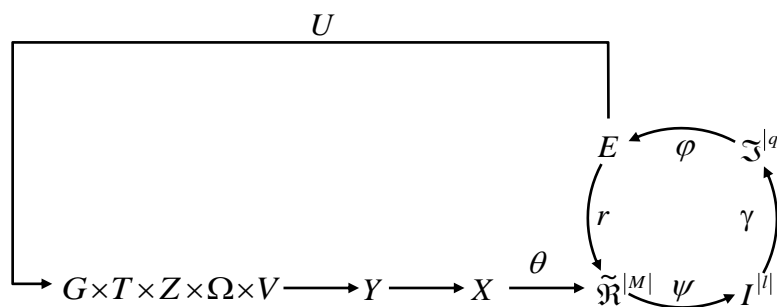


Рисунок 2.3 – Діаграма відображення множин за базовим алгоритмом навчання

На рис.2.3 прийнято такі позначення:

G – простір вхідних сигналів (факторів), які діють на СППР;

T – множина моментів часу зняття інформації;

Ω – простір ознак розпізнавання;

Z – простір можливих станів АСК;

V – множина типів вирішальних правил;

$\Phi: G \times T \times \Omega \times Z \rightarrow Y$ – оператор формування вибіркової множини Y на вході СППР.

Y – вибіркова множина (вхідна навчальна матриця $\| y_{m,i}^{(j)} \|$);

X – вибіркова множина, яка утворює бінарну навчальну матрицю $\| x_{m,i}^{(j)} | i = \overline{1, N}, j = \overline{1, n} \|$, аналогічну за структурою вхідній навчальній

матриці $\| y_{m,i}^{(j)} \|$;

$U: E \rightarrow G \times T \times \Omega \times Z \times V$ – оператор, що регламентує процес навчання і

дозволяє оптимізувати параметри його плану, які визначають, наприклад, обсяг і структуру випробовувань, черговість розгляду класів розпізнавання та інше.

На рис. 2.3 композиція оператора допустимих перетворень $\theta = \theta_1 \circ \theta_2$ складається з оператора θ_1 , який формує вибіркочну бінарну множину X – вхідну у загальному випадку дійсну бінарну навчальну матрицю $\|x_{m,i}^{(j)}\|$ типу “процес-властивість”, і оператора θ_2 , який відновлює в процесі навчання СППР оптимальне розбиття простору ознак на класи еквівалентності.

Серед параметрів навчання, які суттєво впливають на достовірність класифікатора, можуть розглядатися поля контрольних допусків $\{\delta_{k,i} \mid i = \overline{1, N}\}$ на значення ознак, рівні селекції $\{\rho_m\}$ координат еталонних двійкових векторів, крок квантування в часі τ реалізацій образу, параметри словника ознак $\Sigma^{\{N\}}$, параметри впливу середовища та інше.

Таким чином, аналіз розподіленого в просторі та часі технологічного процесу на прикладі виробництва складного мінерального добрива НРК у ВАТ «Сумихімпром» показав, що використання сучасних апаратно-програмних засобів автоматизації (цифрові ПД-регулятори, частотні регулятори та інше) не забезпечує високої функціональної ефективності існуючої АСК за умови впливу випадкових зовнішніх неконтрольованих факторів. Тому застосування адаптивних АСК на основі машинного навчання (самонавчання) для автоматизації складних керованих технологічних процесів особливо актуально для виробництв, де використовуються сировина та матеріали, для яких через технічні ускладнення відсутній неперервний вхідний контроль. Як найбільш перспективним методом аналізу та синтезу адаптивної АСК слід вважати ІЕІ-технологію, що ґрунтується на максимізації інформаційної спроможності СППР шляхом введення в процесі навчання додаткових інформаційних обмежень [1,12].

Розглянемо математичну модель інформаційно-екстремальної СППР, що навчається при незмінній потужності словника ознак. Математична

модель повинна включати як обов'язкову складову частину вхідний математичний опис, який подамо у вигляді теоретико-множинної структури

$$\Delta_B = \langle G, T, \Omega, Z, Y, V; F, \Phi, \Pi, H \rangle,$$

де $F: G \times T \times Z \rightarrow \Omega$ – оператор формування простору ознак;
 $\Pi: G \times T \times \Omega \rightarrow Z$ – оператор переходів, що відбиває механізм зміни станів під дією внутрішніх і зовнішніх збурень; $H: G \times T \times \Omega \times Z \rightarrow V$ – оператор переходу до нового типу вирішальних правил.

У рамках ІЕІ-технології ітераційну оптимізацію процесу самонавчання СППР будемо здійснювати за інформаційним КФЕ, який є функціоналом від точнісних характеристик. Ітераційний процес оптимізації геометричних параметрів розбиття $\tilde{\mathfrak{R}}^{|M|}$ відповідно діаграми, наведеної на рис. 1.2, реалізується оператором $r: E \rightarrow \tilde{\mathfrak{R}}^{|M|}$ шляхом пошуку максимуму КФЕ

$$E_m^* = \max_{\{d\}} E_m(d), \quad (2.3.1)$$

де $\{d\}$ – множина кроків навчання СППР.

Як відомо, основна ідея ІЕІ-технології полягає в зміні значень ознак в субпарацептуальному просторі шляхом допустимих перетворень. Одним із таких перетворень є оптимізація СКД на ознаки, яка у рамках ІЕІ-технології полягає у виборі такої СКД з терм-множини D , що ітераційно наближує значення глобального максимуму інформаційного критерію оптимізації E в робочій (допустимій) області визначення його функції до найбільшого (граничного) його значення. На рис. 2.4 показано категорійну модель оптимізації контрольних допусків в процесі навчання СППР [1,14].

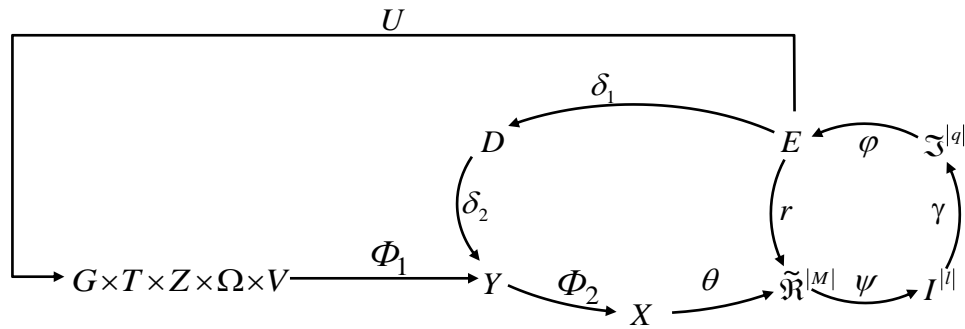


Рисунок 2.4 – Діаграма відображення множин при оптимізації СКД за ІЕІ-технологією

На рис. 2.4 оператори δ_1 та δ_2 відповідно оцінюють вплив параметра, що оптимізується, на функціональну ефективність СППР і регламентують ітераційний процес оптимізації.

На рис. 2.5 показано контур операторів, що безпосередньо оптимізують СКД і який включає контур оптимізації геометричних параметрів розбиття $\tilde{\mathfrak{R}}^{|M|}$.

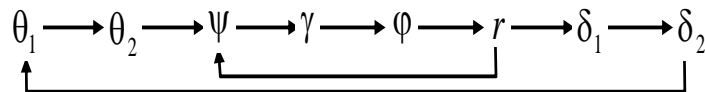


Рис. 2.5. Контур оптимізації контрольних допусків на ознаки

Таким чином, оптимізація геометричних параметрів контейнерів класів проводиться на кожному кроці оптимізації СКД і є внутрішнім циклом інформаційно-екстремального алгоритму навчання СППР.

Математичну модель оптимізації словника ознак (селекції ознак) в рамках ІЕІ-технології можна описати як додатковий контур оптимізації в алгоритмі навчання. Модифіковану математичну модель подамо у вигляді відповідної діаграми відображень множин (рис. 2.6):

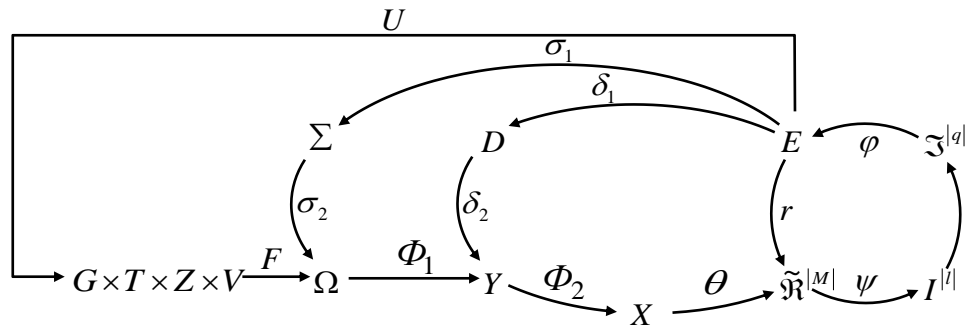


Рисунок 2.6 – Діаграма відображення множин в процесі оптимізації словника ознак розпізнавання в рамках ІЕІ-технології

Контур оптимізації словника ознак показано на рис. 2.7.

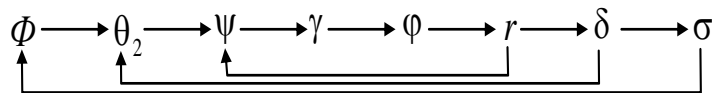


Рисунок 2.7 – Контур оптимізації словника ознак

На рис. 2.7 оператор $\sigma = \sigma_1 \circ \sigma_2 : E \rightarrow \Omega$ змінює простір ознак Ω згідно з відповідним алгоритмом оптимізації словника ознак. Для одержаного поточного варіанту словника Σ в процесі навчання здійснюється оптимізація його параметрів з використанням або тільки базового алгоритму, або алгоритму оптимізації СКД, структурні діаграми яких подано відповідно на рис. 2.3 і рис. 2.5. При цьому оптимізація словника ознак здійснюється шляхом ітераційної процедури пошуку максимуму цільової функції за алгоритмом

$$\Sigma^* = \arg \max_{\Sigma \in \Omega} \{ \max_{G_\delta} \{ \max_{\{k\}} EK_k \} \}, \quad (2.3.2)$$

де EK_k – деяка узагальнена цільова функція, що обчислюється на k -му кроці навчання СППР і включає як інформаційний КФЕ, обчислення

якого є особливістю ІЕІ-технології, так і додаткові умови (наприклад, мінімальна розмірність простору ознак та інше), які є характерними для відповідного алгоритму селекції ознак; G_δ – область допустимих значень поля контрольних допусків; $\{k\}$ – множина кроків навчання.

Наявність додаткових умов обчислення цільової функції вказує на існування допоміжних контурів оптимізації словника ознак, які пов'язані з іншими параметрами функціонування СППР, що оптимізуються. З урахування цього попередня діаграма відображень (рис. 2.6) набуває вигляду, показаному на рис. 2.8, де штрих-пунктирними стрілками позначені можливі додаткові оператори оптимізації словника, що використовують особливості оптимізації інших параметрів функціонування СППР, що навчається. При цьому оператор h_D виділяє групу ознак, які не змінюють КФЕ в процесі оптимізації СКД за умови, що така оптимізація проводилася для кожної ознаки послідовно.

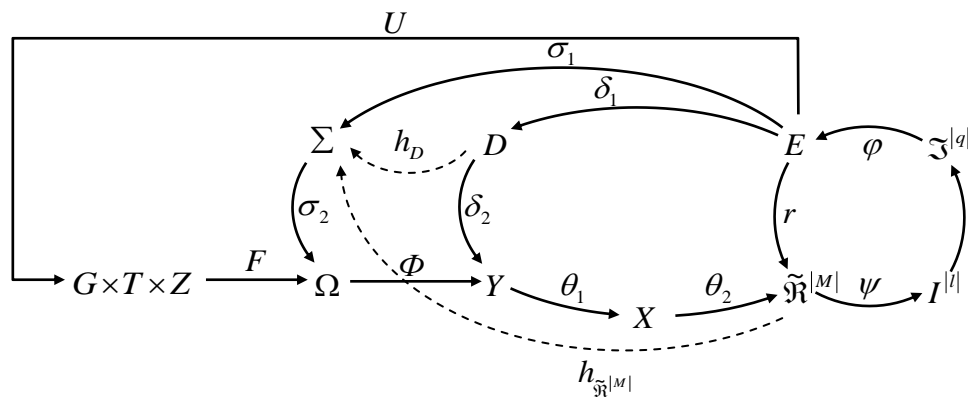


Рисунок 2.8 – Діаграма відображень множин при оптимізації словника ознак з використанням додаткових умов

Показаний на рис. 2.8 оператор $h_{\tilde{R}^{|M|}}$ перевіряє які з ознак було використано при реалізації максимально-дистанційного або мінімально-дистанційного принципів оптимізації геометричних параметрів розбиття простору ознак. Слід зазначити, що ці контури впливають лише на стратегію оптимізації словника, оскільки здатні об'єднувати окремі ознаки в групи відповідно до їх впливу на СППР, що навчається.

На рис. 2.9 показано категорійну модель функціонування СППР в режимі ФКА з самонавчанням.

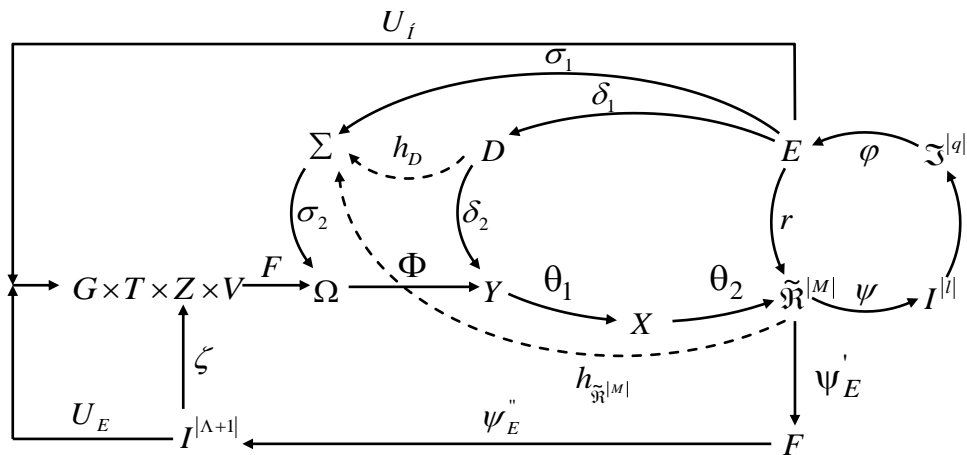


Рисунок 2.9 – Діаграма відображення множин при оптимізації словника ознак в режимі ФКА

Відмінність діаграми відображення множин (рис.2.9) в режимі ФКА від вище наведених полягає в наявності паралельних контурів навчання і екзамену. При цьому оператори U_H і U_E регламентують процес навчання й екзамену відповідно, а оператор класифікації реалізацій образу в режимі екзамену утворює композицію $\Psi_E = \Psi_E' \circ \Psi_E''$, де Ψ_E' – оператор обчислення функції належності реалізації образу відповідному контейнеру; Ψ_E'' – оператор реалізації вирішальних правил.

Перевагою категорійних моделей у вигляді вище наведених діаграм відображення множин є те, що вони дозволяють на етапі системного аналізу СППР, що навчаються (самонавчаються) в режимі ФКА, не тільки встановлювати відношення між елементами інформаційного забезпечення та інформаційними потоками оброблення даних, але і суттєво полегшують розроблення алгоритмів функціонування системи.

Необхідною та достатньою умовою реалізації ФКА у рамках ІЕІ-технології є виконання нерівності

$$\bar{\mu}_m = \frac{1}{n} \sum_{j=1}^n \mu_{m,j} \leq c, \quad (2.3.3)$$

де $\bar{\mu}_m$ – усереднена функція належності вектора-реалізації класу, що розпізнається, контейнеру $K_m^0 \subset X_m^0$; c – порогове значення, що обумовлює прийняття гіпотези відмови від класифікації $\gamma_{\Lambda+1} \in I^{|\Lambda+2|}$. Тут $I^{|\Lambda+1|}$ – множина гіпотез для відкритої абетки, де $\gamma_{\Lambda+1}$ – гіпотеза, що дозволяє формування навчальної матриці нового класу X_{Λ}^0 і відповідно донавчання системи. При цьому для побудованого в радіальному базисі простору ознак гіперсферичного контейнера, який є прийнятним при унімодальному розподілі реалізацій класів, геометрична функція належності, наприклад, для класу X_m^o може мати вигляд [12]

$$\mu_m = 1 - \frac{d(x_m^* \oplus x^{(j)})}{d_m^*}, \quad (2.3.4)$$

де $d(x_m^* \oplus x^{(j)})$ – кодова відстань вектора-реалізації $x^{(j)}$, що розпізнається, від вершини двійкового оптимального еталонного вектора $x_m^* \in \mathfrak{R}^{|\Lambda|^*}$, визначеного в процесі навчання для оптимального чіткого розбиття $\mathfrak{R}^{|\Lambda|^*}$; d_m^* – оптимальний радіус контейнера класу X_m^o , обчислений в процесі навчання СППР.

У рамках ФКА за ІЕІ-технологією алгоритм агрегування нового класу при незмінному словнику ознак розпізнавання полягає у формуванні оператором Ξ додаткової навчальної матриці $\|x_{\Lambda}^{(j)}\|$, де $\Lambda = M + 1$, яка складається з реалізацій екзаменаційної матриці, що дали від'ємні для всіх класів значення функції належності (2.3.4). При досягненні необхідної

репрезентативності матриці $\|x_{\Lambda}^{(j)}\|$ оператор ζ запускає процес донавчання СППР з метою побудови нового розбиття простору ознак.

Таким чином, необхідність реалізації алгоритмів ФКА в сучасних АСК обумовлено невисокою достовірністю оцінки функціональних станів керованих слабо формалізованих технологічних процесів, що відбуваються за умов апіорної невизначеності.

2.4 Оцінка функціональної ефективності СППР, що самонавчається

Функціональна ефективність, як складова загальної ефективності, визначає ступінь відповідності функціонування системи за її робочим алгоритмом виконанню поставленої перед нею задачі згідно з критерієм мети. Важливою складовою критерію мети є інформаційний КФЕ навчання системи, що є функціоналом від точнісних характеристик рішень, що приймаються системою. Задача вибору та обчислення КФЕ є центральною проблемою оцінки функціональної ефективності інтелектуальної СППР, для вирішення якої пріоритетним в задачах прийняття рішень є інформаційний підхід. У рамках ІЕІ-технології знайшли широке використання дві інформаційні міри [1,12]: ентропійна міра Шеннона, яка є інтегральною мірою

$$E_m^{(k)} = 1 + 0,5 \left(\frac{\alpha^{(k)}}{\alpha^{(k)} + D_2^{(k)}} \log_2 \frac{\alpha^{(k)}}{\alpha^{(k)} + D_2^{(k)}} + \frac{D_1^{(k)}}{\beta^{(k)} + D_1^{(k)}} \log_2 \frac{D_1^{(k)}}{\beta^{(k)} + D_1^{(k)}} + \frac{\beta^{(k)}}{\beta^{(k)} + D_1^{(k)}} \log_2 \frac{\beta^{(k)}}{\beta^{(k)} + D_1^{(k)}} + \frac{D_2^{(k)}}{\alpha^{(k)} + D_2^{(k)}} \log_2 \frac{D_2^{(k)}}{\alpha^{(k)} + D_2^{(k)}} \right), \quad (2.4.1)$$

та міра Кульбака:

$$E_m^{(k)} = 0,5 \log_2 \left(\frac{D_1^{(k)} + D_2^{(k)} + 10^{-r}}{\alpha^{(k)} + \beta^{(k)} + 10^{-r}} \right) * [(D_1^{(k)} + D_2^{(k)}) - (\alpha^{(k)} + \beta^{(k)})] \quad (2.4.2)$$

де $D_1^{(k)}$ – перша достовірність, обчислена на k -му кроці навчання;
 $D_2^{(k)}$ – друга достовірність; $\alpha^{(k)}$ – помилка першого роду; $\beta^{(k)}$ – помилка
 другого роду; 10^{-r} – достатньо мале число для уникнення поділу на нуль.

У загальному випадку побудований за (2.4.1) графік функції $E = f(D_1, D_2)$ є поверхнею в тривимірному просторі (рис. 2.10) і є симетричним відносно бісектриси кута D_1OD_2 , тобто при однакових значеннях першої та другої достовірностей. На рис 2.9 для надання більшої наочності друга частина на графіку не відображена.

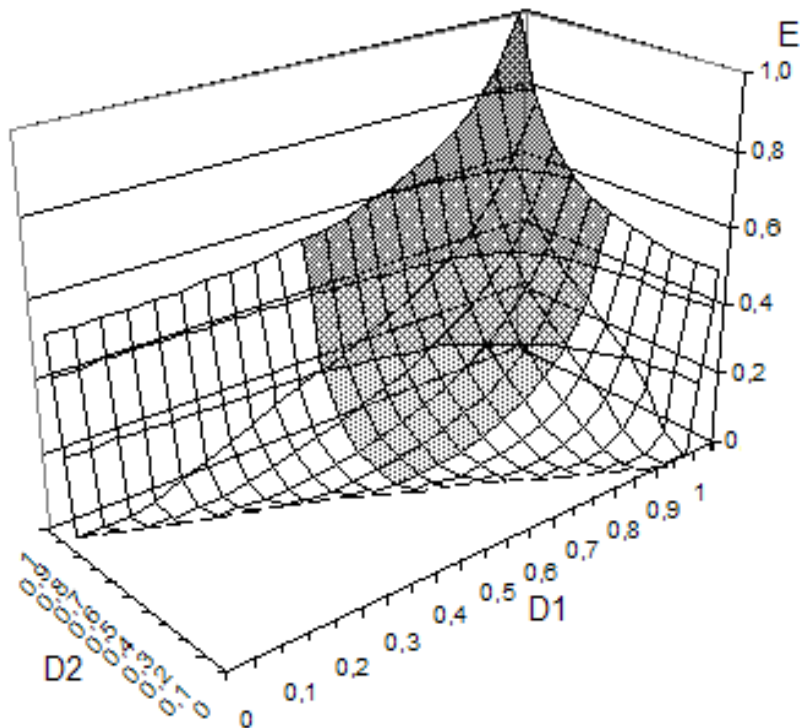


Рисунок 2.10– Графік залежності критерію (2.4.1) від точнісних характеристик при двохальтернативному рішенні

Тривимірну поверхню модифікованого критерію $E = f(D_1, D_2)$, побудовану за формулою (2.4.2), показано на рис. 2.11.

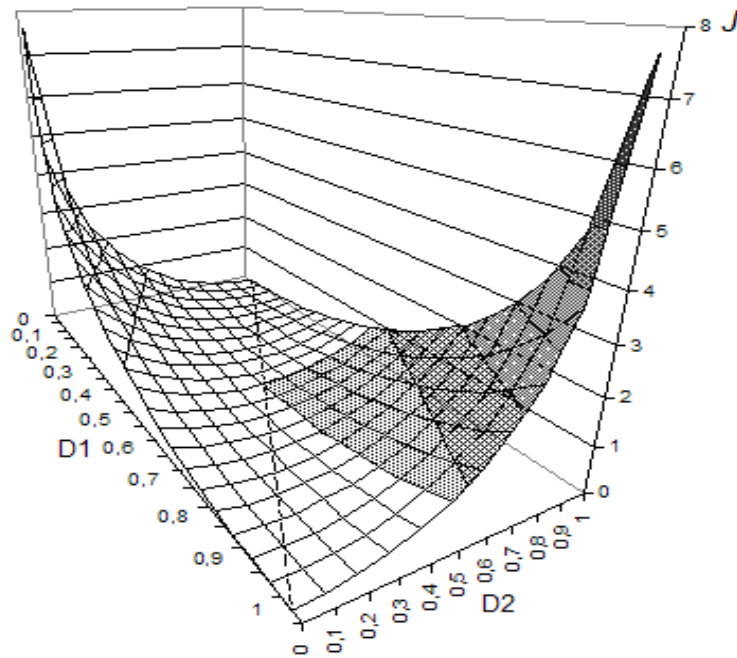


Рисунок 2.11 – Графік залежності критерію (2.4.2) від точнісних характеристик: першої та другої достовірностей

Як видно з рис. 2.10 та рис. 2.11, функції (2.4.1) та (2.4.2) є взаємно-неоднозначні. На практиці цей недолік усувається шляхом введення робочої (допустимої) області визначення функцій, в якій значеннями точнісних характеристик– першої та другої достовірностей, повинні бути більшими відповідних помилок першого та другого роду, тобто $D_1 \geq 0,5$ та $D_2 \geq 0,5$. Робочі області на рис. 2.10 і рис. 2.11 показано на відповідних графіках темним кольором. Аналіз цих графіків показує, що при збільшенні у робочій області як першої, так і другої достовірностей кількість інформації теж збільшується, що знаходиться у відповідності з другим принципом адитивності інформації [12].

Таким чином, аналіз функцій, які використовуються як КФЕ навчання СППР у рамках ІЕІ-технології показує їх відповідність основним вимогам до таких критеріїв:

- вони є прямими і об'єктивними критеріями;
- математично обчислюються і мають геометричний сенс;

- характеризують ступінь відповідності системи своєму призначенню та економічну придатність її використання;
- носять конструктивний характер, тобто дозволяти розробляти методи аналізу та синтезу системи керування;
- є універсальними, тобто здатним для оцінки функціональної ефективності системи керування широкого призначення;
- є чутливими до зміни параметрів функціонування і характеристик системи керування, що навчається;
- дозволяють оптимізувати процес самонавчання СППР з метою максимізації її асимптотичної повної достовірності розпізнавання;
- мають функціональний зв'язок з точносними характеристиками навчання системи керування;
- оцінюють надійність системи керування, що навчається;
- дозволяють прогнозувати зміну функціональної ефективності та надійності адаптивної АСК, що самонавчається.

2.5 Вплив потужності словника ознак розпізнавання на структуру бінарного простору ознак

Побудова безпомилкового класифікатора у рамках ІЕІ-технології за принципом “найближчого сусіда” можлива у частинному випадку за умови невходження всіх реалізацій образу у відповідний контейнер класу розпізнавання, що не гарантує необхідну перформантність машинного навчання, яка може розглядатися як відношення оптимального покриття класів розпізнавання до всього простору ознак. Тому у загальному випадку важливого наукового і практичного значення набуває дослідження впливу потужності як словника ознак, так і алфавіту класів розпізнавання на ефективну та спроможну оцінку асимптотичної повної імовірності

правильного прийняття рішень $P_t^* = 0,5D_1^* + 0,5D_2^*$, де D_1^*, D_2^* – асимптотичні (екстремальні) перша та друга достовірності розпізнавання реалізацій класу X_m^o , обчисленні за результатами оптимізаційного навчання.

Відомо, що у бінарному просторі наближенням гіперсферичного контейнера є гіперкуб. З метою узагальнення та зручності побудови такого контейнера допустимо існування псевдогіперсфери, яка описує гіперкуб, тобто містить всі його вершини. Це дозволяє далі розглядати далі такі параметри оптимізації контейнера в радіальному базисі простору Хеммінга як еталонний вектор, наприклад $x_m \in X_m^o$, вершина якого визначає геометричний центр контейнера $K_m^o \in X_m^o$, і радіус псевдосферичного контейнера, який визначається за формулою

$$d_m = \sum_{i=1}^N (x_{m,i} \oplus \lambda_i), \quad (2.5.1)$$

де $x_{m,i}$ – i -та координата двійкового еталонного вектора x_m ; λ_i – i -та координата деякого вектора λ , вершини якого належать поверхні контейнера $K_m^o \in X_m^o$.

Надалі, з метою спрощення, кодова відстань (2.5.1) між векторами x_m і λ буде позначатися у вигляді $d_m = d(x_m \oplus \lambda)$, а замість терміна “псевдосферичний” будемо використовувати термін “гіперсферичний” контейнер.

Розглянемо вплив на перфомантність навчання потужності словника ознак. Нехай d_0^*, d_1^* – оптимальні радіуси контейнерів класів X_0^o та X_1^o відповідно, а $d_c = d(x_0^* \oplus x_1^*)$ – кодова відстань між їх центрами – еталонними векторами $x_0^* \in X_0^o$ і $x_1^* \in X_1^o$ відповідно. Враховуючи особливості бінарного простору Хеммінга, можна визначити такі допущення:

1) потужність бінарного простору Хеммінга для словника ознак розпізнавання $\Sigma^{|M|}$ дорівнює 2^N ;

2) кількість двійкових реалізацій на кодовій відстані d ($0 \leq d \leq N$) від двійкового вектора x дорівнює

$$B_d^N(x) = C_N^d = \frac{N!}{d!(N-d)!};$$

3) кількість бінарних реалізацій, що належить довільному контейнеру класу X_K^o з радіусом d_K ($0 \leq d_K \leq N$) дорівнює

$$B_{d,K}^N = \sum_{i=0}^{d_K} B_N^i = \sum_{i=0}^{d_K} \frac{N!}{i!(N-i)!}.$$

При цьому

$$B_{d,K}^N = \sum_{i=0}^{d_K-1} \frac{N!}{i!(N-i)!} + \frac{N!}{d_K!(N-d_K)!} = B_{d-1,K}^N + B_d^N(x_K),$$

де $B_{d-1,K}^N$ – кількість реалізацій образу в контейнері з радіусом $d-1$.

Таким чином, у випадку чіткого розбиття $\mathfrak{R}^{|M|}$ для M класів, тобто при $d_0^* + d_1^* < d_c$, кількість відповідних реалізацій, що належить контейнерам класів X_0^o і X_1^o дорівнює

$$B_{\mathfrak{R}^{|M|}}^N = \sum_{k=0}^m B_{d,k}^N = \sum_{k=0}^m \sum_{i=0}^{d_k^*} \frac{N!}{i!(N-i)!},$$

а кількість реалізацій, що знаходяться поза межами цих контейнерів дорівнює

$$B_{\mathfrak{R}}^N = 2^N - B_{\mathfrak{R}^{|M|}}^N.$$

Розглянемо випадок нечіткого розбиття $\tilde{\mathfrak{R}}^{[2]}$ для двох класів X_0^o і X_1^o , що перетинаються, тобто при $d_0^* + d_1^* \geq d_c$ (рис. 2.12).

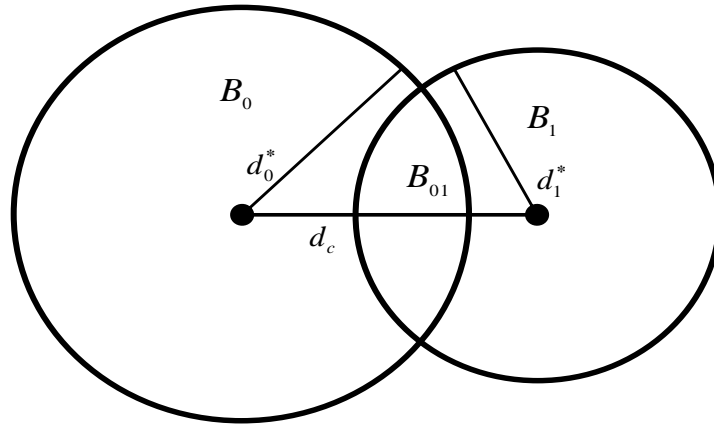


Рисунок 2.12– Геометричні характеристики контейнерів класів X_0^o і X_1^o

Обчислення значення $B_{\tilde{\mathfrak{R}}^{[2]}}^N$ ускладнюється через наявність області бінарного простору $K_0^o \cap K_1^o$, де контейнери класів X_0^o і X_1^o перетинаються:

$$B_{\tilde{\mathfrak{R}}^{[2]}}^N = B_{\mathfrak{R}^{[2]}}^N - B_{\tilde{\mathfrak{R}}^{[2]}}^N(X_0^o, X_1^o).$$

Л е м а 2.5.1. Кількість бінарних реалізацій, що знаходяться на кодovій відстані d_0 від двійкового вектора x_0 та d_1 від двійкового вектора x_1 дорівнює нулю, якщо $d_0 + d_1 < d(x_0 \oplus x_1)$.

Доведення. Нехай двійковий вектор x_0 – нульовий, тоді у векторі x_1 міститься $d(x_0 \oplus x_1)$ одиничних компонент і $N - d(x_0 \oplus x_1)$ нульових. Бінарна реалізація в просторі Хеммінга, що знаходиться на кодovій відстані d_0 від двійкового нульового вектора x_0 , містить d_0 одиничних компонент. При цьому відстань від цієї реалізації до двійкового вектора x_1 залежить від

кількості d_0 компонент, що збігаються з $d(x_0 \oplus x_1)$ одиничними компонентами вектора x_1 . Тоді найменшою кодова відстань буде у випадку, коли кількість таких збігів буде максимально можливою, тобто буде дорівнювати $|d(x_0 \oplus x_1) - d_0|$. Таким чином, для $d_1 < |d(x_0 \oplus x_1) - d_0|$, що включає випадок $d_0 + d_1 < d(x_0 \oplus x_1)$, кількість відповідних бінарних реалізацій дорівнює нулю, що і доводиться.

Аналогічно, найбільша кодова відстань буде у випадку, коли кількість таких збігів буде мінімально можливою, тобто буде дорівнювати $N - d(x_0 \oplus x_1) + d_0$, якщо $d(x_0 \oplus x_1) + d_0 \leq N$, і $-|d(x_0 \oplus x_1) - d_0|$, якщо $d(x_0 \oplus x_1) + d_0 > N$. Таким чином, кількість відповідних бінарних реалізацій дорівнює нулю так само і для випадку, коли

$$d_1 > \begin{cases} N - d(x_0 \oplus x_1) + d_0, & \text{якщо } d(x_0 \oplus x_1) + d_0 \leq N; \\ |d(x_0 \oplus x_1) - d_0|, & \text{якщо } d(x_0 \oplus x_1) + d_0 > N. \end{cases}$$

Таким чином, кодова відстань d_1 , для якої кількість відповідних бінарних реалізацій не нульова, приймає значення в інтервалі $[d(x_0 \oplus x_1) - d_0; d(x_0 \oplus x_1) + d_0]$, якщо $d_0 \leq N - d(x_0 \oplus x_1)$ і $d_0 \leq d(x_0 \oplus x_1)$, або $[d(x_0 \oplus x_1) - d_0; d(x_0 \oplus x_1) + (N - d(x_0 \oplus x_1)) - (d_0 - (N - d(x_0 \oplus x_1)))]$, тобто $[d(x_0 \oplus x_1) - d_0; 2 \cdot N - d(x_0 \oplus x_1) - d_0]$, якщо $d_0 > N - d(x_0 \oplus x_1)$ і $d_0 \leq d(x_0 \oplus x_1)$, або $[d_0 - d(x_0 \oplus x_1); d_0 + d(x_0 \oplus x_1)]$, якщо $d_0 \leq N - d(x_0 \oplus x_1)$ і $d_0 > d(x_0 \oplus x_1)$, або

$$[d_0 - d(x_0 \oplus x_1); (N - d(x_0 \oplus x_1)) - (d_0 - (N - d(x_0 \oplus x_1))) + d(x_0 \oplus x_1)],$$

тобто $[d_0 - d(x_0 \oplus x_1); 2 \cdot N - d(x_0 \oplus x_1) - d_0]$, якщо $d_0 > N - d(x_0 \oplus x_1)$ і $d_0 > d(x_0 \oplus x_1)$.

Л е м а 2.5.2. Кількість бінарних реалізацій, що знаходяться на кодовій відстані d_0 від двійкового вектору x_0 та d_1 від двійкового вектору x_1 дорівнює нулю, якщо $d_1 = |d(x_0 \oplus x_1) - d_0| + 2p + 1$, де $p = 0, 1, 2, \dots$

Доведення. Нехай двійковий вектор x_0 – нульовий, тоді у векторі x_1 міститься $d(x_0 \oplus x_1)$ одиничних компонент та $N - d(x_0 \oplus x_1)$ нульових. Розглянемо двійкову реалізацію x_{\min} , що характеризується мінімально можливою кодовою відстанню від двійкового вектора x_1 $d_1 = |d(x_0 \oplus x_1) - d_0|$. Зрозуміло, що найближча двійкова реалізація $x_{\min \pm 1}$, що відрізняється від даної, характеризується кодовою відстанню $d_1 = |d(x_0 \oplus x_1) - d_0| \pm 1$. Це досягається шляхом зміни значення однієї з компонент даної реалізації на протилежне, що одночасно збільшує або зменшує на одиницю кодову відстань d_0 від $x_{\min \pm 1}$ до двійкового вектору x_0 .

Таким чином, кількість двійкову реалізацій, що знаходяться на кодовій відстані d_0 від двійкового вектора x_0 та d_1 від двійкового вектора x_1 дорівнює нулю, якщо $d_1 = |d(x_0 \oplus x_1) - d_0| \pm 1$. Бінарні реалізації $\{x_{\min \pm 2}\}$, що характеризуються кодовою відстанню $d_1 = |d(x_0 \oplus x_1) - d_0| \pm 2$, можна сформуванати шляхом зміни значення однієї з одиничних компонент та однієї з нульових компонент реалізації x_{\min} на протилежні, не змінюючи кодовій відстані d_0 . Застосування аналогічних міркувань для будь-якого елемента $\{x_{\min \pm 2}\}$ дозволяє встановити, що на їх основі також неможливо сформуванати реалізацій з кодовою відстанню d_0 та $d_1 = |d(x_0 \oplus x_1) - d_0| \pm 3$. Аналізуючи в такий спосіб інтервал кодової відстані d_1 , визначений раніше, отримуємо додаткову множину значень $d_1 = |d(x_0 \oplus x_1) - d_0| + 2p + 1$ ($p = 0, 1, \dots$), для яких кількість відповідних бінарних реалізацій нульова.

Подамо структуру бінарного простору Хеммінга графічно (рис. 2.12).

На рис. 2.13 відображена структура десятивимірного бінарного простору, в якому відтворено контейнери двох класів X_0^o та X_1^o , що перетинаються, радіусами $d_0^* = 5$, $d_1^* = 4$ відповідно та міжцентровою відстанню $d(x_0 \oplus x_1) = 7$. При цьому контейнери гомоморфно подано у вигляді кіл відповідного радіуса, з центрів яких розходяться зображені пунктирною лінією кола.

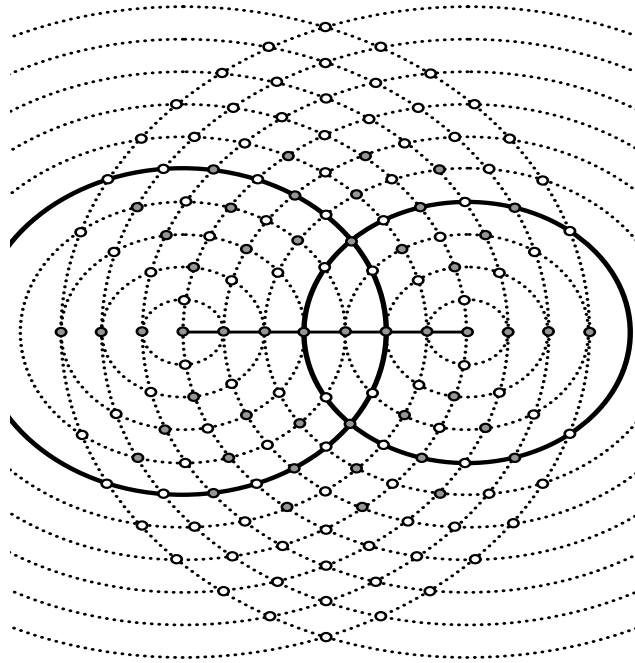


Рисунок 2.13– Структура бінарного простору Хеммінга

На рис.2.13 точки перетину цих окружностей позначено заповненими або незаповненими колами в залежності від кількості двійкових реалізацій в них.

Л е м а 2.5.3. Кількість двійкових реалізацій, що знаходяться на кодовій відстані d_0 від двійкового вектору x_0 та d_1 від двійкового вектору x_1 дорівнює

$$n = \frac{d(x_0 \oplus x_1)!}{d_0!(d(x_0 \oplus x_1) - d_0)!},$$

якщо $d_1 = d(x_0 \oplus x_1) - d_0$ та $d_0 \leq d(x_0 \oplus x_1)$.

Доведення. Нехай двійковий вектор x_0 – нульовий, тоді в x_1 міститься $d(x_0 \oplus x_1)$ одиничних компонент та $N - d(x_0 \oplus x_1)$ нульових. Двійкова реалізація, що характеризується кодовими відстанями з умови леми, містить d_0 одиничних компонентів. При цьому виконання умови $d_1 = d(x_0 \oplus x_1) - d_0$ можливо тоді і тільки тоді, коли координати цих компонентів збігаються з координатами одиничних компонент двійкового

вектора x_1 . Кількість таких двійкових реалізацій буде дорівнювати кількості комбінацій, що можна сформувати з d_0 одиничних та d_1 нульових компонентів, тобто $\frac{(d_0 + d_1)!}{d_0!d_1!}$ або

$$\frac{d(x_0 \oplus x_1)!}{d_0!(d(x_0 \oplus x_1) - d_0)!} = C(d(x_0 \oplus x_1), d_0).$$

Аналогічно, за умов $d_1 = d_0 - d(x_0 \oplus x_1)$ та $d_0 > d(x_0 \oplus x_1)$ двійкові реалізації містять $N - d_0$ нульових компонентів, координати яких збігаються з координатами нульових компонент двійкового вектора x_1 . Кількість таких двійкових реалізацій буде дорівнювати кількості комбінацій, що можна сформувати з $N - d_0$ нульових та d_1 одиничних компонентів, тобто

$$\frac{(N - d_0 + d_1)!}{(N - d_0)!d_1!},$$

або

$$\frac{(N - d(x_0 \oplus x_1))!}{(N - d(x_0 \oplus x_1) - d_1)!d_1!} = C(N - d(x_0 \oplus x_1), d_1).$$

Т е о р е м а 2.5.1. Кількість двійкових реалізацій, що знаходяться на кодовій відстані d_0 від двійкового вектору x_0 та d_1 від двійкового вектору x_1 дорівнює

$$n = \frac{d(x_0 \oplus x_1)!}{(d_0 - p)!(d(x_0 \oplus x_1) - d_0 + p)!} \cdot \frac{(N - d(x_0 \oplus x_1))!}{(N - d(x_0 \oplus x_1) - p)!p!},$$

де $p = \frac{d_1 + d_0 - d(x_0 \oplus x_1)}{2}$.

Доведення. Нехай двійковий вектор x_0 – нульовий, тоді в x_1 міститься $d(x_0 \oplus x_1)$ одиничних компонент та $N - d(x_0 \oplus x_1)$ нульових. Розглянемо двійкові реалізації, механізм формування яких було подано в лемі 2.5.2. При цьому для реалізацій, кількість яких не нульова, виконання умови $d_1 = |d(x_0 \oplus x_1) - d_0| + 2p$, де $p = 0, 1, 2, \dots$, можливо тоді і тільки тоді, коли одночасно координати $d_0 - p$ одиничних компонентів збігаються з координатами одиничних компонент двійкового вектора x_1 та координати p нульових компонентів збігаються з координатами нульових компонент двійкового вектора x_1 , якщо $d_0 \leq d(x_0 \oplus x_1)$; або коли одночасно координати p нульових компонентів збігаються з координатами одиничних компонент двійкового вектора x_1 та координати $d_0 - d(x_0 \oplus x_1) + p$ одиничних компонентів збігаються з координатами одиничних компонент двійкового вектора x_1 , якщо $d_0 > d(x_0 \oplus x_1)$. Кількість таких двійкових реалізацій буде дорівнювати

$$\begin{cases} \frac{d(x_0 \oplus x_1)!}{(d_0 - p)!(d(x_0 \oplus x_1) - d_0 + p)!} \cdot \frac{(N - d(x_0 \oplus x_1))!}{(N - d(x_0 \oplus x_1) - p)!p!}, & \text{якщо } d_0 \leq d(x_0 \oplus x_1), \\ \frac{d(x_0 \oplus x_1)!}{(d(x_0 \oplus x_1) - p)!p!} \cdot \frac{(N - d(x_0 \oplus x_1))!}{(N - d_0 - p)!(d_0 - d(x_0 \oplus x_1) + p)!}, & \text{якщо } d_0 > d(x_0 \oplus x_1), \end{cases}$$

де

$$p = \frac{d_1 - |d(x_0 \oplus x_1) - d_0|}{2} = \begin{cases} \frac{d_1 - d(x_0 \oplus x_1) + d_0}{2}, & \text{якщо } d_0 \leq d(x_0 \oplus x_1), \\ \frac{d_1 + d(x_0 \oplus x_1) - d_0}{2}, & \text{якщо } d_0 > d(x_0 \oplus x_1). \end{cases}$$

Таким чином, кількість двійкових реалізацій, що характеризується кодovими відстанями відповідно з умовами теореми, дорівнює

$$\frac{d(x_0 \oplus x_1)!}{(d_0 - p)!(d(x_0 \oplus x_1) - d_0 + p)!} \cdot \frac{(N - d(x_0 \oplus x_1))!}{(N - d(x_0 \oplus x_1) - p)!p!},$$

$$\text{де } p = \frac{d_1 + d_0 - d(x_0 \oplus x_1)}{2}.$$

У табл. 2.1 наведено розподіл реалізацій у псевдосфері простору Хеммінга з радіусом, що дорівнює 10 кодовим одиницям.

Таблиця 2.1–Розподіл реалізацій образу у псевдосфері простору Хеммінга

$d_0 \backslash d_1$	0	1	2	3	4	5	6	7	8	9	10
0	0	0	0	0	0	0	0	C_7^0	0	0	0
1	0	0	0	0	0	0	C_7^1	0	C_3^1	0	0
2	0	0	0	0	0	C_7^2	0	$C_7^1 C_3^1$	0	C_3^2	0
3	0	0	0	0	C_7^3	0	$C_7^2 C_3^1$	0	$C_7^1 C_3^2$	0	C_3^3
4	0	0	0	C_7^4	0	$C_7^3 C_3^1$	0	$C_7^2 C_3^2$	0	$C_7^1 C_3^3$	0
5	0	0	C_7^5	0	$C_7^4 C_3^1$	0	$C_7^3 C_3^2$	0	$C_7^2 C_3^3$	0	0
6	0	C_7^6	0	$C_7^5 C_3^1$	0	$C_7^4 C_3^2$	0	$C_7^3 C_3^3$	0	0	0
7	C_7^7	0	$C_7^6 C_3^1$	0	$C_7^5 C_3^2$	0	$C_7^4 C_3^3$	0	0	0	0
8	0	$C_7^7 C_3^1$	0	$C_7^6 C_3^2$	0	$C_7^5 C_3^3$	0	0	0	0	0
9	0	0	$C_7^7 C_3^2$	0	$C_7^6 C_3^3$	0	0	0	0	0	0
10	0	0	0	$C_7^7 C_3^3$	0	0	0	0	0	0	0

Аналіз табл. 2.1 показує, що простір Хеммінга не є рівномірним. Крім того, враховуючи таку властивість комбінацій як $C_N^{N-i} = C_N^i$, де $0 \leq i \leq N$, можна стверджувати, що структура бінарного простору є симетричною відносно головної та допоміжної діагоналей таблиці.

Деталізацію структури простору Хеммінга, поданого на рис. 2.13, наведено на рис. 2.14.

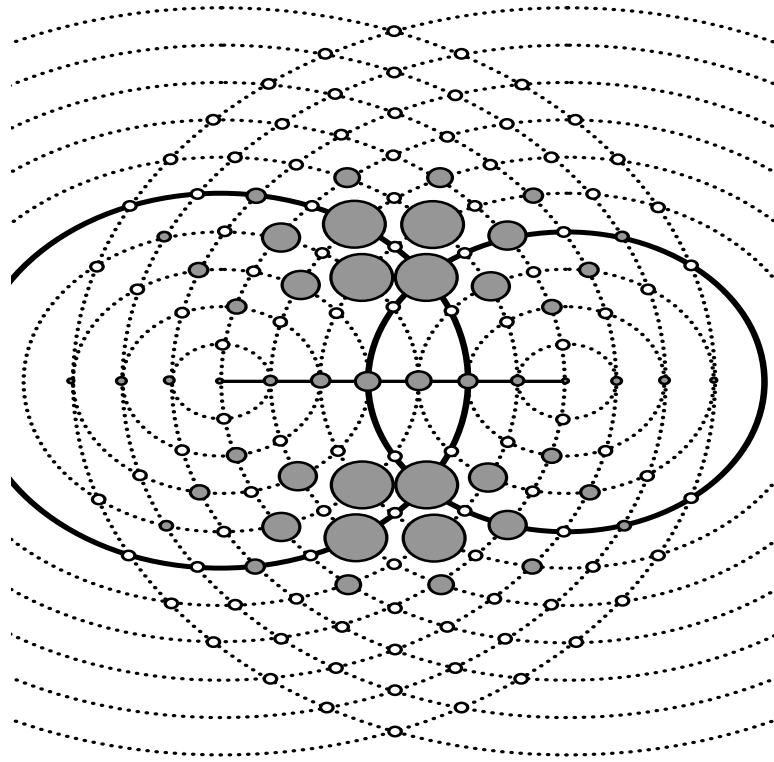


Рисунок 2.13– Деталізована структура простору Хеммінга

На рис.2.14 показано структуру десятивимірного бінарного простору, в якому відтворено контейнери двох класів X_0^o та X_1^o , що перетинаються. При цьому діаметр заповнених кіл відповідає кількості двійкових реалізацій, що характеризуються відповідними відстанями від центрів зображених контейнерів.

Позначимо кількість двійкових реалізацій, що знаходяться на кодовій відстані d_0 від двійкового вектору x_0 та d_1 від двійкового вектору x_1 , як $B_{x_0(d_0), x_1(d_1)}^N$. Тоді кількість двійкових реалізацій, що одночасно належить контейнеру класу X_0^o радіусом d_0^* ($0 \leq d_0^* \leq N$) та контейнеру класу X_1^o радіусом d_1^* ($0 \leq d_1^* \leq N$) дорівнює

$$B_{\mathbb{R}^2}(x_0^o, x_1^o) = \sum_{i=0}^{d_0^*} \sum_{j=0}^{d_1^*} B_{x_0(i), x_1(j)}^N.$$

Таким чином, деталізований аналіз структури розбиття класів в просторі Хеммінга свідчить про імплікативність і симетричність розподілу векторів-реалізацій класів у контейнерах, побудованих в радіальному базисі простору ознак.

РОЗДІЛ 3

ОПТИМІЗАЦІЯ ПАРАМЕТРІВ ФУНКЦІОНУВАННЯ
СППР, ЩО САМОНАВЧАЄТЬСЯ

3.1 Інформаційно-екстремальний алгоритм навчання СППР

Ідея оптимізації вирішальних правил на етапі навчання у рамках ІЕІ-технології полягає в цілеспрямованому наближенні глобального максимуму інформаційного критерію оптимізації, що обчислюється в робочій (допустимій) області визначення функції критерію, до його граничного значення. Параметри функціонування здатної самонавчатися СППР, що оптимізуються, можна поділити на фенотипні, які прямо впливають на геометричну форму контейнерів класів, на генотипні, які впливають на розподіл векторів-реалізацій класу в просторі ознак і на інформаційні, які є параметрами словника ознак і так само прямо впливають на величину інформаційного КФЕ навчання СППР.

Оптимізація геометричних параметрів контейнерів класів, побудованих в радіальному базисі простору ознак, для гіперсферичного класифікатора здійснюється у загальному випадку ($M > 2$) за алгоритмом, який реалізує оператори контуру базової діаграми відображення множин (рис.2.3) за ітераційною процедурою

$$d_m^* = \arg \max_{G_d} \bar{E}^{|M|}, \quad (3.1.1)$$

де d_m^* – оптимальний радіус контейнера класу X_m^o ; $\bar{E}^{|M|}$ – усереднений КФЕ навчання СППР; G_d – область допустимих значень геометричних параметрів.

У виразі (3.1.1) усереднений інформаційний КФЕ навчання СППР обчислюється за формулою:

$$\bar{E}^{|M|} = \frac{1}{M} \sum_{m=1}^M E_m^*, \quad (3.1.2)$$

де E_m^* – максимальне (глобальне) значення інформаційного КФЕ навчання СППР розпізнавати реалізації класу X_m^o , обчислене в робочій області визначення його функції.

Основними задачами базового алгоритму навчання СППР є:

- обчислення на кожному кроці навчання інформаційного КФЕ;
- пошук глобального максимуму КФЕ навчання СППР в робочій області визначення його функції;
- визначення оптимальних значень геометричних параметрів контейнерів класів, які відповідають глобальному максимуму КФЕ в робочій області визначення його функції.

Розглянемо оптимізацію геометричних параметрів контейнерів двох класів X_1^o і X_2^o , побудованих в радіальному базисі простору ознак, які характеризували функціональний стан технологічного процесу виробництва складного мінерального добрива NPK у ВАТ «Сумихімпром». При цьому клас X_1^o характеризував функціональний стан технологічного процесу нейтралізації кислот, який відповідав заданому технологічному режиму, а клас X_2^o , який відповідав вмісту азоту в пульпі «Менше норми». Навчальна матриця для кожного з цих класів включала по 40 реалізацій, одержаних за виробничих умов і кожна з яких складалася з 55 структурованих ознак. При цьому 39 ознак вводилося безпосередньо з відповідних датчиків інформації, маркування яких наведено в додатку А, і 16 ознак вводилося оператором за результатами хімічного аналізу згідно з існуючим регламентом. Тривалість процесу нейтралізації кислот дорівнює двом годинам, а всього процес виготовлення виробу дорівнює шести годинам, чотири з яких припадають на оброблення пульпи в барабані грануляції, де СМД набуває остаточних фізико-хімічних властивостей.

Оптимізація геометричних параметрів контейнерів класів X_1^o і X_2^o здійснювалася за ітераційним алгоритмом (3.1.1) як за інформаційним критерієм Кульбака (2.4.2), так і за ентропійним критерієм Шеннона (2.4.1).

На рис. 3.1 показано оптимізацію контейнерів цих класів за критерієм Кульбака.

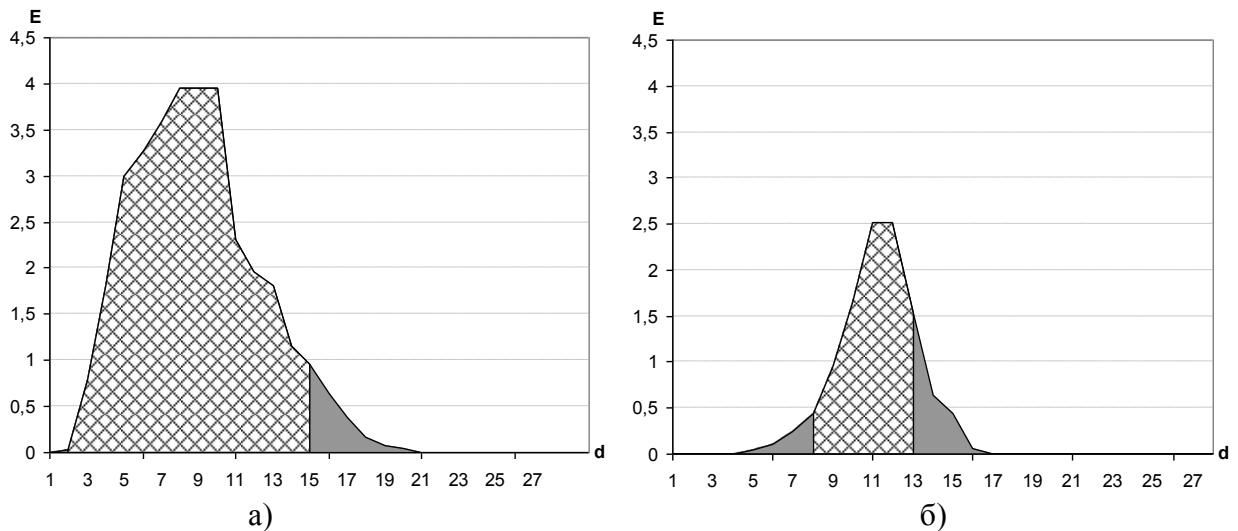


Рисунок 3.1 – Графіки залежності КФЕ за Кульбаком від радіусів контейнерів: а) клас X_1^o ; б) клас X_2^o

На рис 3.1 (і далі в роботі) робочу (допустиму) область визначення функції критерію позначено подвійною штриховкою. Аналіз рис. 3.1 показує, що оптимальний радіус для контейнера класу X_1^o дорівнює $d_1^* = 8$ кодовим одиницям в просторі Хеммінга, а – для контейнера класу X_2^o дорівнює $d_2^* = 11$ кодовим одиницям. При цьому максимальні значення КФЕ дорівнюють відповідно $E_1^* = 4$ і $E_2^* = 2,5$ при максимальному граничному значенні критерію $E_{гран} = 7,5$, яке обчислюється за формулою (2.4.2) при $r = 2$ і підстановці точнісних характеристик: $D_1 = 1$ і $\beta = 0$.

На рис.3.2 з метою порівняння результатів показано процес оптимізації контейнерів сусідніх класів X_1^o і X_2^o за тим самим алгоритмом (3.1.1), але за критерієм Шеннона (2.4.1).

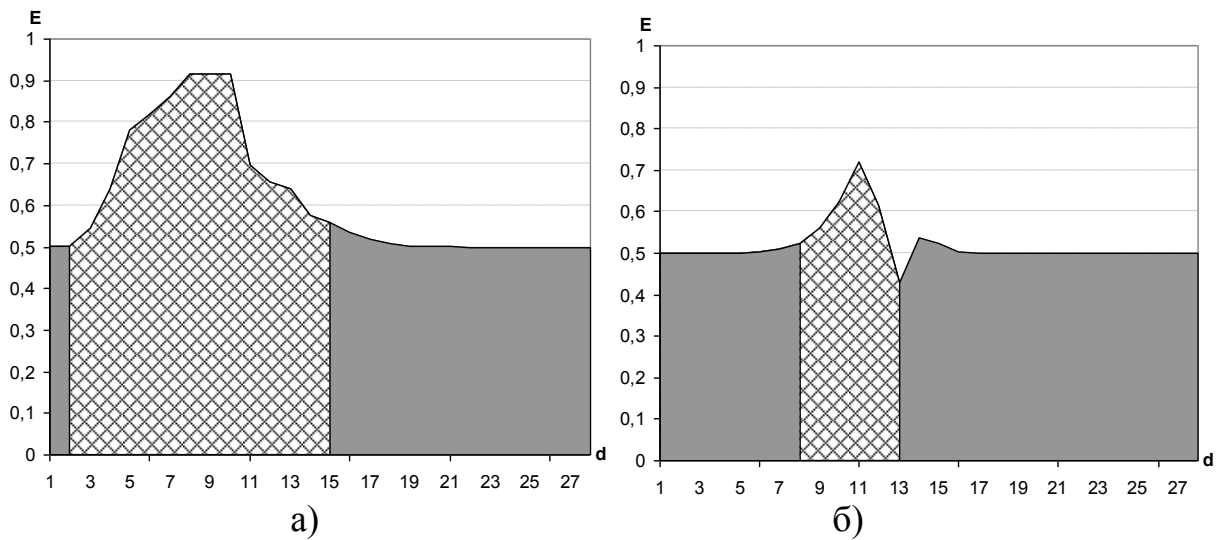


Рисунок 3.2 – Графіки залежності КФЕ за Шенноном (2.4.1) від радіусів контейнерів: а) клас X_1^o ; б) клас X_2^o

Аналіз рис. 3.2 показує, що оптимальні радіуси контейнерів класів X_1^o і X_2^o дорівнюють відповідно так само $d_1^* = 8$ кодовим одиницям і $d_2^* = 11$ кодовим одиницям. При цьому максимальні значення КФЕ дорівнюють відповідно $E_1^* = 0,9$ і $E_2^* = 0,7$ при максимальному граничному значенні $E_{гран} = 1,0$ нормованого критерію (2.4.1).

Таким чином, шляхом фізичного моделювання доведено, що всі інформаційні критерії, які відповідають відомим умовам, дозволяють визначити однакові за значенням оптимальні параметри функціонування СППР шляхом пошуку глобального максимуму функції критерію в робочій області її визначення.

На рис. 3.3 показано розподіл проєкцій реалізацій класів X_1^o і X_2^o на площі для вище побудованих за базовим алгоритмом навчання їх

оптимальних контейнерів. При цьому ромбами позначено проекції реалізацій класу X_1^o , а кружечками – проекції реалізацій класу X_2^o , контейнер якого знаходиться справа. Зрозуміло, що такий розподіл проекцій реалізацій на двовимірну площу є не ізоморфним, а гомоморфним.

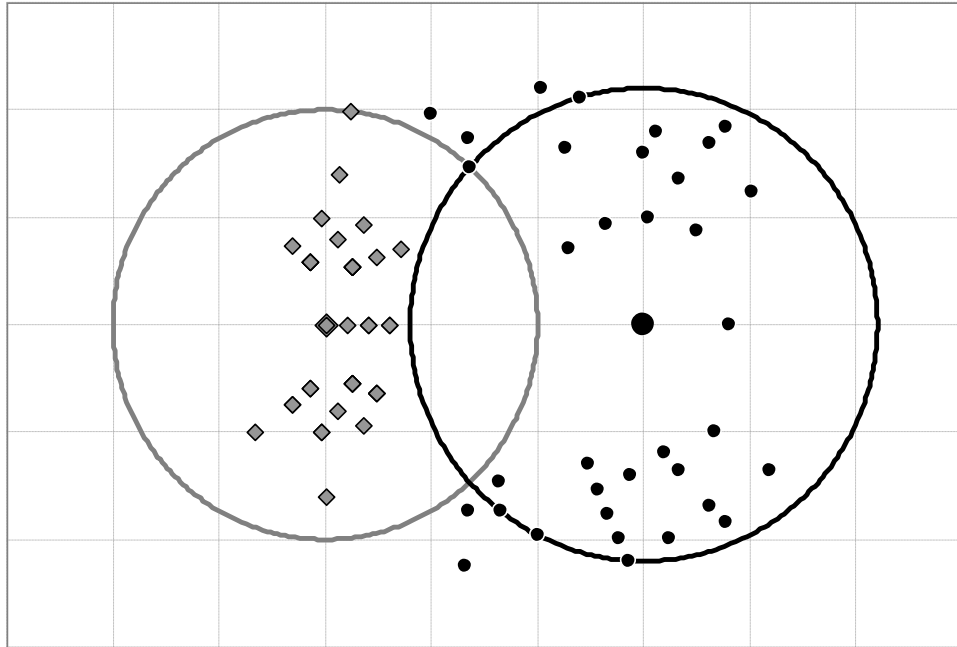


Рисунок 3.3 – Розподіл реалізацій класів X_1^o і X_2^o на площі для побудованого за базовим алгоритмом оптимального розбиття класів

На рис. 3.3 величина ромбів та кружечків пропорційна кількості проекцій вершин векторів-реалізацій в одну координату. Аналіз рис. 3.3 показує, що класи утворюють нечітке розбиття, оскільки суттєво перетинаються. Це є свідченням того, що побудовані за базовим алгоритмом навчання вирішальні правила не є безпомилковими за навчальною матрицею.

Таким чином, у загальному випадку застосування тільки базового алгоритму навчання не дозволяє одержати оптимальні геометричні параметри контейнерів класів, оскільки їх радіуси і геометричні центри, які збігаються з вершинами еталонних двійкових векторів-реалізацій відповідних класів, залежать від контрольних допусків на ознаки та інших

параметрів функціонування системи, які можуть бути неоптимальними. Крім того базовий алгоритм навчання не гарантує наявності робочої області визначення функції інформаційного критерію, який є функціоналом від точнісних характеристик рішень, що приймаються системою. Тому на практиці базовий алгоритм є внутрішнім циклом багатоциклічної структури алгоритму оптимізації параметрів функціонування в процесі навчання системи. Згідно з принципом відкладеного рішення [134] глибина багатоциклічної структури алгоритму оптимізації в методах ІЕІ-технології визначається параметром функціонування, при якому побудовано безпомилкові за навчальною матрицею вирішальні правила.

Розглянемо алгоритм відновлення оптимальних контейнерів класів за алгоритмом паралельної оптимізації контрольних допусків, який реалізується операторами контуру, показаному на рис.2.4. При цьому оптимізація СКД в процесі навчання СППР здійснювалася за ітераційною процедурою [12]

$$\delta^* = \arg \max_{G_\delta} \{ \max_{G_d} \bar{E}^{|M|} \}, \quad (3.1.3)$$

де G_δ – допустима область значень параметра поля допусків δ .

Допустима область значень параметра δ у виразі (3.1.3) визначається за умови $\delta \in [0; \delta_H / 2]$, де δ_H – нормоване поле допусків для відносної шкали вимірювання ознак, яке є областю значень для параметра контрольного поля допусків δ .

Оптимізація проводилася для чотирьох класів, які характеризували функціональний стан технологічного процесу виробництва складного мінерального добрива НРК у ВАТ «Сумихімпром». При цьому клас X_1^o так само характеризував функціональний стан технологічного процесу, який відповідав заданому технологічному режиму. Інші три класи характеризували відхилення вмісту хімічних компонентів (азот, фосфор і калій відповідно) від норми. Навчальна матриця для кожного з цих класів

мала так само розмір 55×40 . Динаміку зміни усередненого значення КФЕ навчання СППР з оптимізацією СКД показано на рис. 3.4.

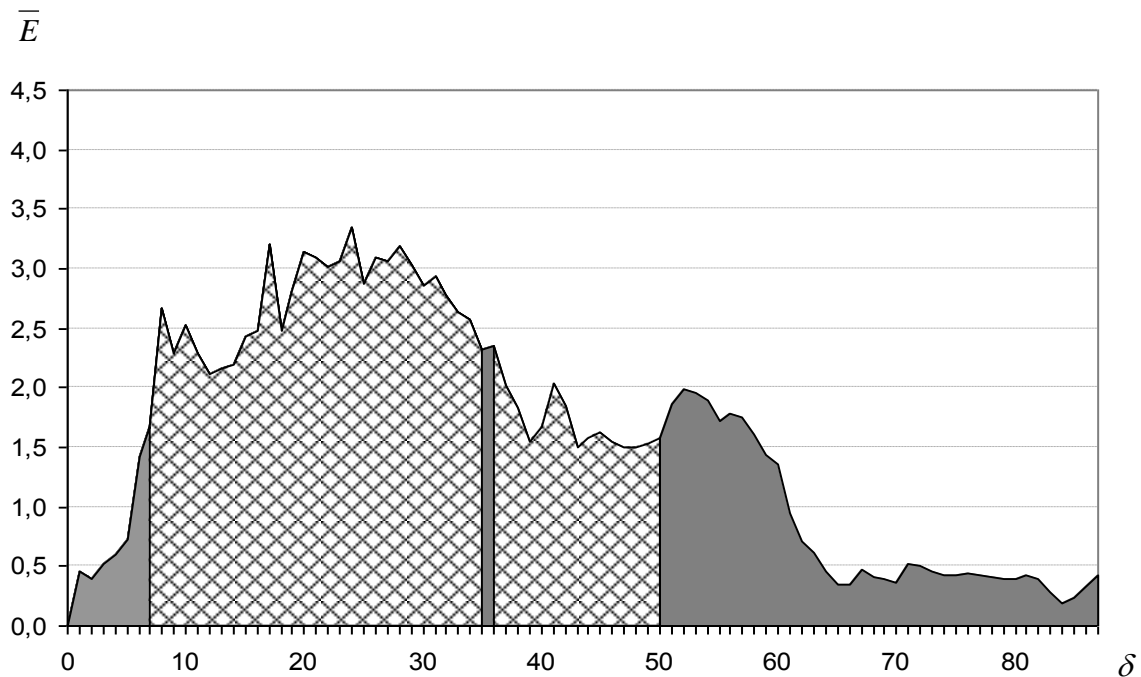


Рисунок 3.4 – Графік залежності усередненого КФЕ від параметра контрольного поля допусків при паралельній оптимізації СКД

Аналіз рис. 3.4 показує, що оптимальне значення параметра контрольного поля допусків дорівнює $\delta^* = \pm 24$ градаціям відносного симетричного нормованого поля допусків. При цьому максимальне значення усередненого КФЕ, при якому СКД приймає відповідні оптимальні значення, дорівнює $E^* = 3,34793$.

Відмінність графіка, наведеного на рис.3.4, від графіків, показаних на рис.3.1 і рис.3.2, одержаних при оптимізації контейнерів у частинному випадку ($M = 2$), полягає в тому, що він має у загальному випадку не одну робочу область, а декілька (світлі ділянки, позначені подвійною штриховкою), що обумовлено необхідністю виконання в процесі його реалізації обмежень $d_m < d(x_m \oplus x_c)$, тобто оптимальний радіус контейнера класу X_m^o повинен бути менше його міжцентрової кодової відстані з

сусіднім класом X_c^o , інакше клас X_m^o поглинав би ядро класу X_c^o , що не відповідає умові (2.2.6). При цьому ліва межа робочих областей визначається за таким предикатним виразом:

$$(\forall X_m^o \in \tilde{\mathfrak{R}}^{|M|}) [(D_1^{(k)}(d) > 0,5) \& (D_2^{(k)}(d) > 0,5)],$$

де $D_1^{(k)}(d)$ – перша достовірність, обчислена на k -му кроці навчання; $D_2^{(k)}(d)$ – друга достовірність.

Динаміка зміни відповідних значень КФЕ при оптимізації геометричних параметрів контейнерів відповідних класів показана на рис. 3.5. Аналіз рис. 3.5 показує, що оптимальними радіусами контейнерів класів є:

- для класу X_1^o оптимальний радіус контейнера дорівнює $d_1^* = 9$ (тут і далі в кодових одиницях) при максимальному значенні КФЕ $E_0^* = 2,98974$ і точнісних характеристиках $D_1^* = 0,95$ та $\beta = 0,05$ та кодовій відстані до центру найближчого класу X_2^o $d_c = d(x_1 \oplus x_2) = 11$;

- для класу X_2^o – $d_2^* = 10$, відповідне максимальне значення КФЕ $E_1^* = 3,26756$ при точнісних характеристиках $D_1^* = 0,95$ та $\beta^* = 0,03$ та кодовій відстані до центру найближчого класу X_1^o $d_c = 11$;

- для класу X_3^o – $d_3^* = 16$, відповідне максимальне значення КФЕ $E_3^* = 4,39232$ при точнісних характеристиках $D_1^* = 1,00$ та $\beta^* = 0,00$ та кодовій відстані до центру найближчого класу X_1^o $d_c = 18$;

- для класу X_4^o – $d_4^* = 10$, відповідне максимальне значення КФЕ $E_4^* = 2,74212$ при точнісних характеристиках $D_1^* = 0,93$ та $\beta^* = 0,05$ та кодовій відстані до центру найближчого класу X_2^o $d_c = 14$.

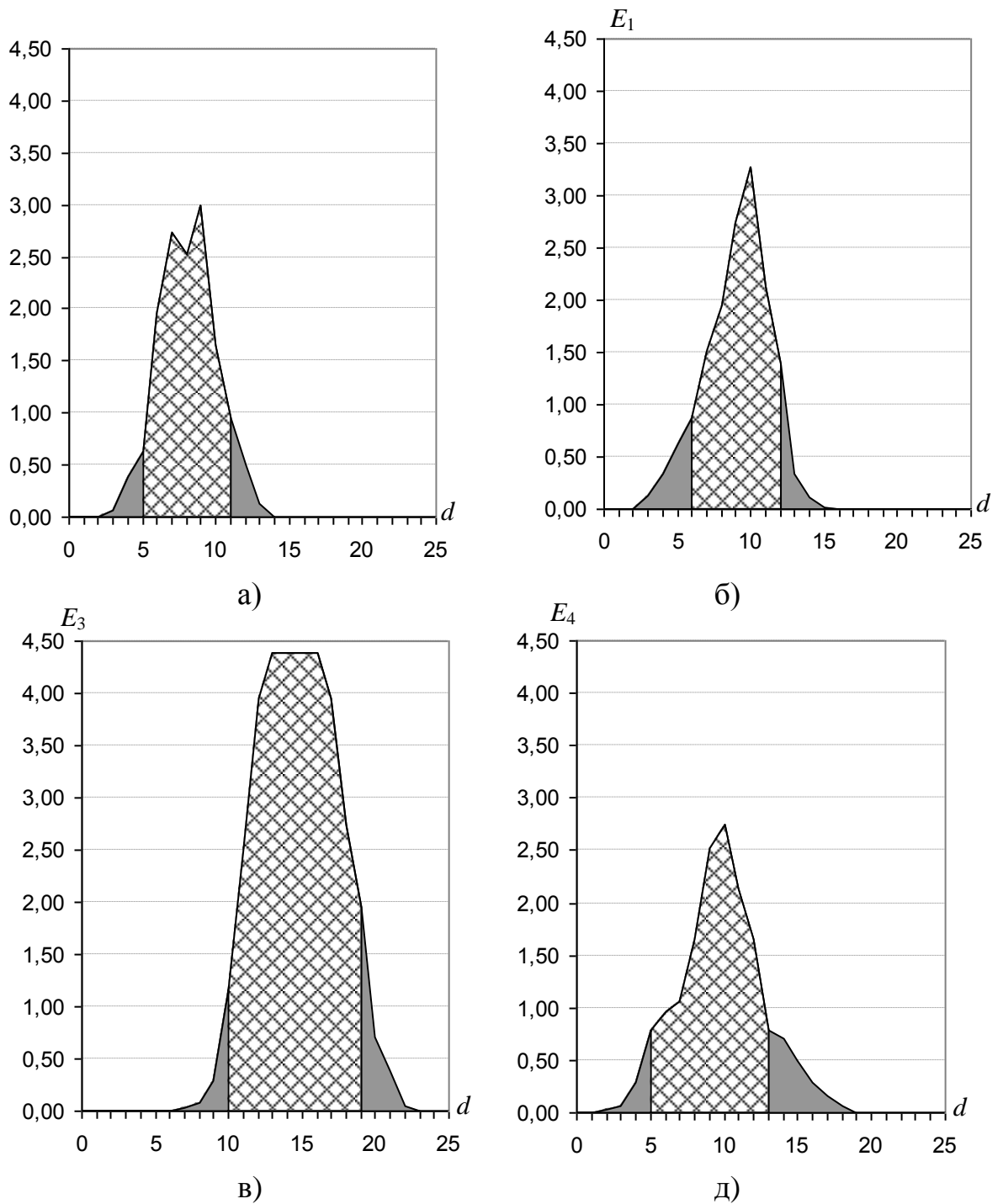


Рисунок 3.5 – Динаміка зміни КФЕ при оптимізації геометричних параметрів контейнерів: а) клас X_1^0 ; б) клас X_2^0 ; в) клас X_3^0 ; д) клас X_4^0

Розглянемо процес оптимізації радіуса контейнера базового класу X_1^0 детальніше. Для цього подамо відповідну інформацію у вигляді табл. 3.1.

Таблиця 3.1 –Визначення оптимального радіуса контейнера класу X_1^o

$\alpha_1^{(k)}$	$E_1^{(k)}$	$D_1^{(k)}$	$\beta^{(k)}$	$D_2^{(k)}$	$\alpha^{(k)}$	$K1^{(k)}$	$K2^{(k)}$	$K3^{(k)}$	$K4^{(k)}$
0	0,00000	0,00	0,00	1,00	1,00	0	0	40	40
1	0,00000	0,00	0,00	1,00	1,00	0	0	40	40
2	0,00656	0,05	0,00	1,00	0,95	2	0	38	40
3	0,05939	0,15	0,00	1,00	0,85	6	0	34	40
4	0,38425	0,38	0,00	1,00	0,63	15	0	25	40
5	0,63338	0,48	0,00	1,00	0,53	19	0	21	40
6	1,95949	0,78	0,00	1,00	0,23	31	0	9	40
7	2,74212	0,88	0,00	1,00	0,13	35	0	5	40
8	2,51895	0,90	0,05	0,95	0,10	36	2	4	38
9	2,98974	0,95	0,05	0,95	0,05	38	2	2	38
10	1,65513	0,98	0,25	0,75	0,03	39	10	1	30
11	0,96242	0,98	0,40	0,60	0,03	39	16	1	24
12	0,49974	0,98	0,55	0,45	0,03	39	22	1	18
13	0,13469	0,98	0,75	0,25	0,03	39	30	1	10
14	0,00164	1,00	0,98	0,03	0,00	40	39	0	1
15	0,00000	1,00	1,00	0,00	0,00	40	40	0	0
16	0,00000	1,00	1,00	0,00	0,00	40	40	0	0
17	0,00000	1,00	1,00	0,00	0,00	40	40	0	0
18	0,00000	1,00	1,00	0,00	0,00	40	40	0	0
19	0,00000	1,00	1,00	0,00	0,00	40	40	0	0
20	0,00000	1,00	1,00	0,00	0,00	40	40	0	0

Пошук оптимального радіуса контейнера класу X_1^o проводився в робочій області визначення КФЕ, на яку накладаються обмеження щодо точнісних характеристик $D_1^{(k)} \geq 0,5$ та $D_2^{(k)} \geq 0,5$ та безпосередньо до радіусу контейнера $d_1^{(k)} < d_c$. Таким чином, інтервал $d_1^{(k)} \in [6, 11]$ вважається робочою областю визначення функції КФЕ, а за оптимальне значення радіуса контейнера класу X_1^o слід прийняти значення $d_1^* = 9$ (код. од.), яке відповідає максимальному значенню КФЕ $E_1^* = 2,98974$.

Таким чином, показано, що використання базового алгоритму навчання не забезпечує необхідну достовірність вирішальних правил і згідно з принципом відкладених рішень щодо оптимального керування потребує

додаткової оптимізації інших параметрів функціонування СППР, що впливають на її функціональну ефективність. Шляхом фізичного моделювання за експериментальними даними, отриманими за цехових умов виробництва складного мінерального добрива НКР, доведено, що всі інформаційні критерії, які відповідають відомим умовам [73,74], дозволяють визначити однакові за значенням оптимальні параметри функціонування СППР шляхом пошуку глобального максимуму функції критерію в робочій області її визначення. Крім того у загальному випадку ($M > 2$) існує не одна робоча область визначення функції інформаційного КФЕ навчання СППР, як це має місце для випадку $M=2$, а декілька, що обумовлено необхідністю виконання в процесі відновлення в радіальному базисі простору ознак контейнерів класів геометричних обмежень, заданих виразом (2.2.11). При цьому за оптимальне приймається екстремальне значення параметра оптимізації в тій робочій області, де КФЕ має найбільший глобальний максимум. Наведений приклад оптимізації СКД показав, що вона може не забезпечити побудову безпомилкових за навчальною матрицею вирішальних правил, що потребує оптимізації інших параметрів функціонування СППР.

3.2 Оцінка інформативності ознак розпізнавання

Як було показано в підрозділі 1.2, більшість авторів оцінку інформативності ознак здійснюють на етапі формування словника за окремим алгоритмом обчислення інформаційної характеристики – як правило, кількості інформації [75]. За ІЕІ-технологією оцінку інформативності ознак доцільно здійснювати у рамках інформаційно-екстремального алгоритму навчання СППР. При цьому для розв'язання цієї задачі обмежимося використанням алгоритму паралельної оптимізації СКД на відповідні ознаки. Як генотипний параметр оптимізації розглянемо

параметр поля контрольних допусків δ . Тоді оптимізація цього параметру здійснюється за структурованою ітераційною процедурою [12]

$$\delta^* = \operatorname{argmax}_{G_\delta} \{\max_{\{d\}} E_1\}, \quad (3.2.1)$$

де E_1 – КФЕ навчання СППР розпізнавати реалізації класу X_1^o ; G_δ – область допустимих значень параметра δ ; $\{d\}$ – множина кроків навчання, $d \in [1; d(x_1 \oplus x_c) - 1]$. Тут $d(x_1 \oplus x_c)$ – міжцентрова кодова відстань між двійковим еталонним вектором $x_1 \in X_1^o$ і двійковим еталонним вектором $x_c \in X_c^o$, де X_c^o – найближчий сусідній клас до класу X_1^o .

За алгоритмом (3.2.1) оптимізація контрольних допусків здійснюється одночасно для всіх ознак. Оскільки параметр поля допусків $\pm \delta$, що оптимізується, є однаковим для всіх ознак еталонного вектора-реалізації класу X_1^o , то нижній A_{HK_i} і верхній A_{BK_i} контрольні допуски, наприклад, для i -ї ознаки обчислюються на кожному кроці навчання за формулами

$$A_{HK_i} = y_{1,i} - \delta * h_i; \quad A_{BK_i} = y_{1,i} + \delta * h_i,$$

де $y_{1,i}$ – i -а координата еталонного вектора-реалізації $y_1 \in X_1^o$; h_i – ціна градації поля допусків для i -ї ознаки.

При цьому вершини еталонних векторів-реалізацій визначають геометричні центри контейнерів відповідних класів розпізнавання, які відновлюються в процесі навчання в радіальному базисі простору ознак.

Вхідною інформацією для навчання СППР за алгоритмом паралельної оптимізації ознак є навчальна матриця $\|y_{m,i}^{(j)}\|_{m=\overline{1,M}; i=\overline{1,N}, j=\overline{1,n^*}}$, система полів нормованих допусків $\{\delta_{H,i}\}$, яка задає область визначення відповідних контрольних допусків $\{\delta_{K,i}\}$ на значення ознак. При цьому

параметр поля контрольних допусків δ для всіх ознак визначається з інтервалу $[0; \delta_H / 2]$, або $[0; \delta_{H,i} / 2h_i]$.

Цикл оптимізації СКД на ознаки розглядається як внутрішній по відношенню до ітераційної процедури оцінки інформативності ознак.

Розглянемо кроки реалізації алгоритму оцінки інформативності ознак.

1. Реалізується алгоритм паралельної оптимізації контрольних допусків на ознаки (3.2.1) і визначається усереднене значення КФЕ для повного словника ознак розпізнавання: $\bar{E}_{ALL}^*(P_1, P_2, \dots, P_{i-1}, P_i, P_{i+1}, \dots, P_N)$.

2. Формується лічильник ознак: $i=1$.

3. Реалізується алгоритм паралельної оптимізації ознак (3.2.1) і визначається усереднене значення КФЕ для словника Σ_i : $\bar{E}_i^*(P_1, P_2, \dots, P_{i-1}, P_{i+1}, \dots, P_N)$.

4. Обчислюється $\Delta_i = \bar{E}_i^* - \bar{E}_{ALL}^*$ і визначається тип ознаки за правилом: якщо $\Delta_i < 0$, то ознака інформативна, якщо $\Delta_i = 0$, то ознака неінформативна і якщо $\Delta_i > 0$, то ознака є “заважаючою”.

5. $i=i+1$.

6. Якщо $i \leq N$, виконується пункт 3, інакше – ЗУПИН реалізації алгоритму і вивід одержаних результатів.

У процесі оптимізації словника ознак на кожному кроці формувалися варіанти словників з потужністю менше на одиницю за потужність поточного словника ознак, для яких проводилася оптимізація параметрів навчання й обчислювалося максимальне усереднене значення КФЕ. На рис. 3.6 наведено ці значення для варіантів словників на першому кроці навчання. Безперервна пряма лінія на рис. 3.6 відповідає максимальному усередненому значенню ентропійного нормованого КФЕ \bar{E}_{ALL}^* для повного словника ознак. Односторонньою штриховкою позначено ознаки, при видаленні яких максимальне усереднене значення КФЕ \bar{E}_i^* збігається з \bar{E}_{ALL}^* ,

двосторонньою штриховкою – ознаки, для яких \bar{E}_i^* менше за \bar{E}_{ALL}^* , сірим кольором – варіанти, для яких \bar{E}_i^* більше за \bar{E}_{ALL}^* .

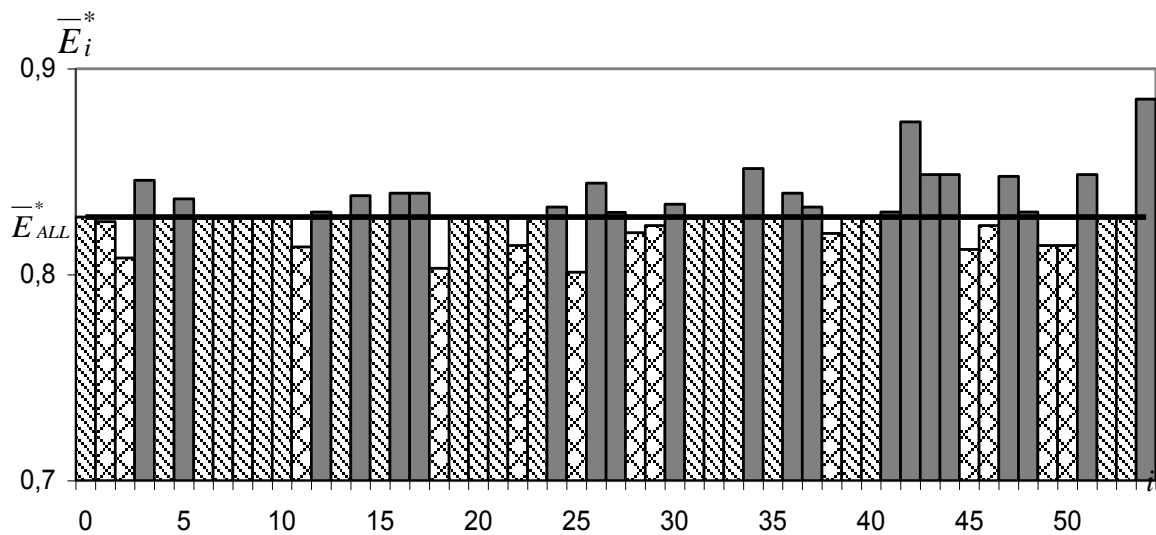


Рисунок 3.6 – Максимальне усереднене значення КФЕ для варіантів словників, що були сформовані в процесі оцінки інформативності ознак

Аналіз рис. 3.6 показує, що ознаки 0, 4, 6-10, 13, 15, 19-21, 23, 31-33, 35, 39, 40, 52, 53 є неінформативними, оскільки їх видалення не впливає на значення КФЕ навчання СППР ($\Delta_i = 0$). Ознаки, для яких $\Delta_i < 0$ є інформативними, а для яких $\Delta_i > 0$, то їх слід розглядати як “заважаючі”. Слід зазначити, що наведена класифікація ознак є виправданою тільки для скінченного алфавіту класів.

Таким чином, із повного словника ознак можна умовно виділити 21 неінформативну, 21 заважаючу та 13 інформативних ознак. Найбільш інформативною ознакою є 25 ознака „ВА013” (додаток А). При її видаленні із словника ознак максимальне усереднене значення КФЕ зменшиться на 3% за рахунок погіршення точнісної характеристики – першої достовірності D_1 , що виникла при оптимізації геометричних параметрів контейнера класу X_4^o . Максимально заважаючою виявилася 54 ознака „ВА107”. При її видаленні із словника ознак максимальне усереднене значення КФЕ збільшилося на 7%

за рахунок поліпшення значення першої достовірності D_1 , що виникла при оптимізації геометричних параметрів контейнера класу X_3^o .

3.3. Оптимізація словника ознак з використанням випадкового пошуку

Перевірка ефективності алгоритмів оптимізації словника ознак із застосуванням методів випадкового пошуку [136] в рамках ІЕІ-технології здійснювалася для алфавіту із п'яти класів. Крім того як параметри оптимізації СППР розглядалися СКД на ознаки та геометричні параметри контейнерів відповідних класів. Зміну усередненого інформаційного критерію Кульбака (2.3.2), в процесі оптимізації СКД на ознаки для заданого алфавіту класів і апріорного словника ознак, який є надлишковим в інформаційному розумінні, показано на рис 3.7.

Аналіз рис. 3.7 показує, що при паралельній оптимізації за симетричною стратегією S_1 оптимальне значення параметра поля допусків, яке визначає оптимальні нижні та верхні контрольні допуски на відповідні ознаки, дорівнює $\delta^* = \pm 27$ відносних одиниць. При цьому глобальне значення усередненого КФЕ в робочій області визначення його функції дорівнює $\bar{E}^* = 2,76677$. На рис. 3.7 до робочої області відносяться тільки ті значення параметра δ , при яких можна побудувати для всіх класів заданого алфавіту контейнери з ненульовими радіусами за умови, що перша та друга достовірності перебільшують помилки першого та другого роду відповідно. При цьому оптимальні значення радіусів контейнерів класів розпізнавання дорівнюють: $d_1^* = 7$ для базового класу X_1^o при міжцентровій відстані $d_c = 9$; $d_2^* = 8$ для класу X_2^o при міжцентровій відстані $d_c = 10$; $d_3^* = 13$ для класу X_3^o при міжцентровій відстані $d_c = 17$; $d_4^* = 11$ для класу X_4^o при міжцентровій відстані $d_c = 13$; і $d_5^* = 8$ для класу X_5^o при міжцентровій відстані $d_c = 9$.

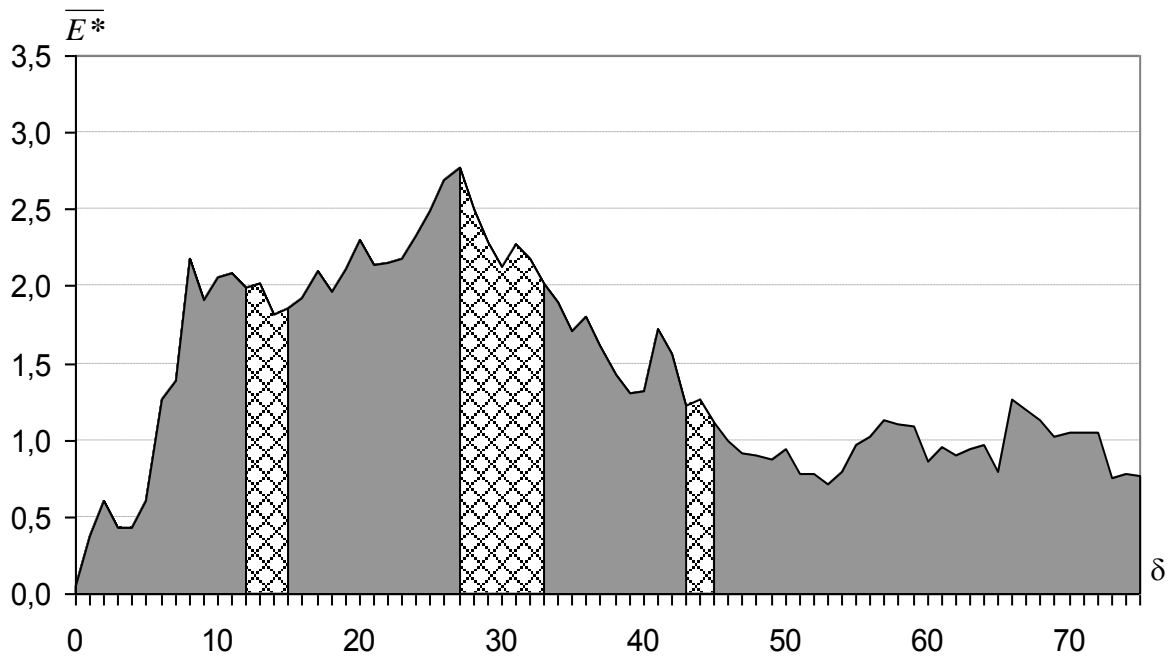


Рисунок 3.7 – Графік залежності КФЕ за Кульбаком від параметра поля допусків в процесі паралельної оптимізації СКД

Таким чином, у процесі оптимізації не вдалося побудувати безпомилковий класифікатор, на що вказує значна ступінь перетину контейнерів класів і значення точнісних характеристик: для класу X_1^o перша та друга достовірності дорівнюють відповідно $D_{1,1}=0,95$ і $D_{2,1}=1,0$; для класу X_2^o – $D_{1,2}=0,65$ і $D_{2,2}=0,95$; для класу X_3^o – $D_{1,3}=1,0$ і $D_{2,3}=1,0$; для класу X_4^o – $D_{1,4}=0,98$ і $D_{2,4}=1,0$ і для класу X_5^o – $D_{1,5}=0,63$ і $D_{2,5}=1,0$.

Для підвищення ефективності навчання СППР застосовувалася оптимізація словника ознак з використанням методу випадкового пошуку, який полягав у такому. Випадковим чином було сформовано 300 варіантів словників ознак і для кожного проводилася оптимізація тих самих параметрів функціонування, що і для нульової ітераційної процедури, яка відбувалася при апріорному надлишковому в інформаційному розумінні словнику: параметра параметра СКД як відхилення верхнього та нижнього контрольного допуску від центра розсіювання (математичного сподівання) значень кожної ознаки.

На рис. 3.8 показано динаміку зміни глобального значення КФЕ за Кульбаком в процесі випадкового генерування словників ознак.

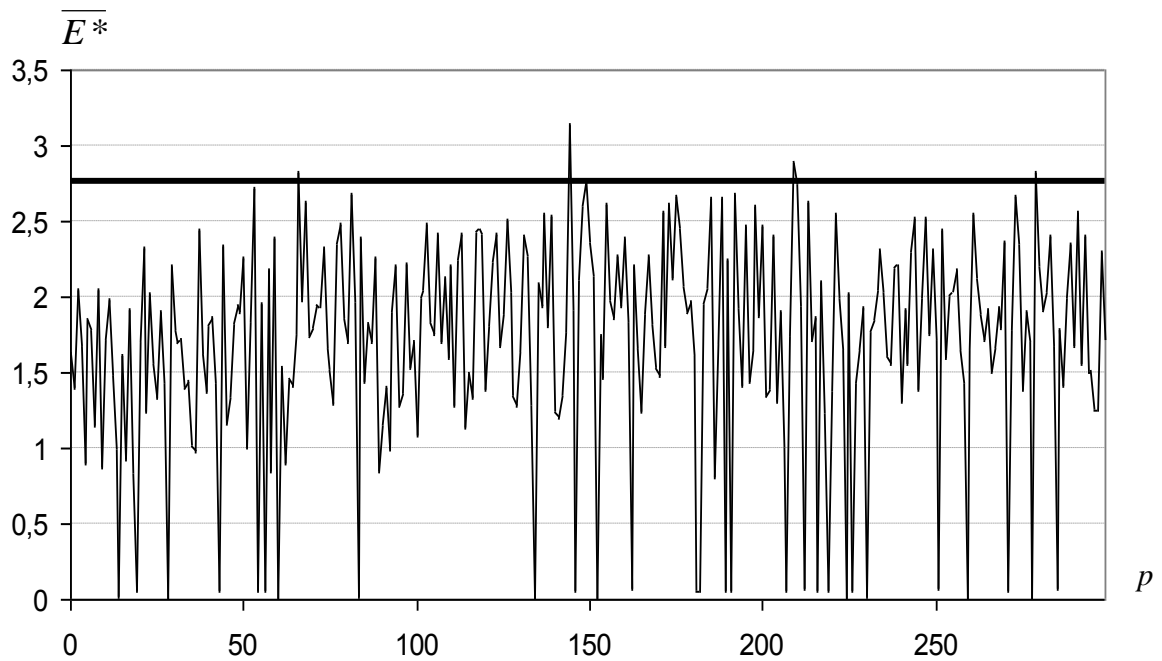


Рисунок 3.8 – Графік залежності глобального значення КФЕ за Кульбаком від варіантів словників ознак з використанням методу випадкового пошуку (без адаптації)

Серед всіх варіантів словника ознак оптимальним вважався такий, що характеризувався максимальним значенням усередненого КФЕ. Аналіз рис. 3.8 показує, що найбільш ефективним виявився варіант словника № 144, потужність якого дорівнювала $Card \Sigma^{|N|} = 32$ і який складався з 1– 4, 8, 10, 11, 13 – 16, 20, 24, 26, 28, 29, 31, 33, 37 – 39, 41, 42, 45, 46, 49-50 і 52-55 структурованих ознак.

Результати оптимізації словника за методом випадкового пошуку без адаптації подано на рис. 3.9. Аналіз рис. 3.9 показує, що оптимальне значення параметра поля контрольних допусків за алгоритмом їх паралельній оптимізації за симетричною стратегією S_1 і з випадковим генеруванням словників ознак не змінилося і дорівнює так само $\delta^* = \pm 27$.

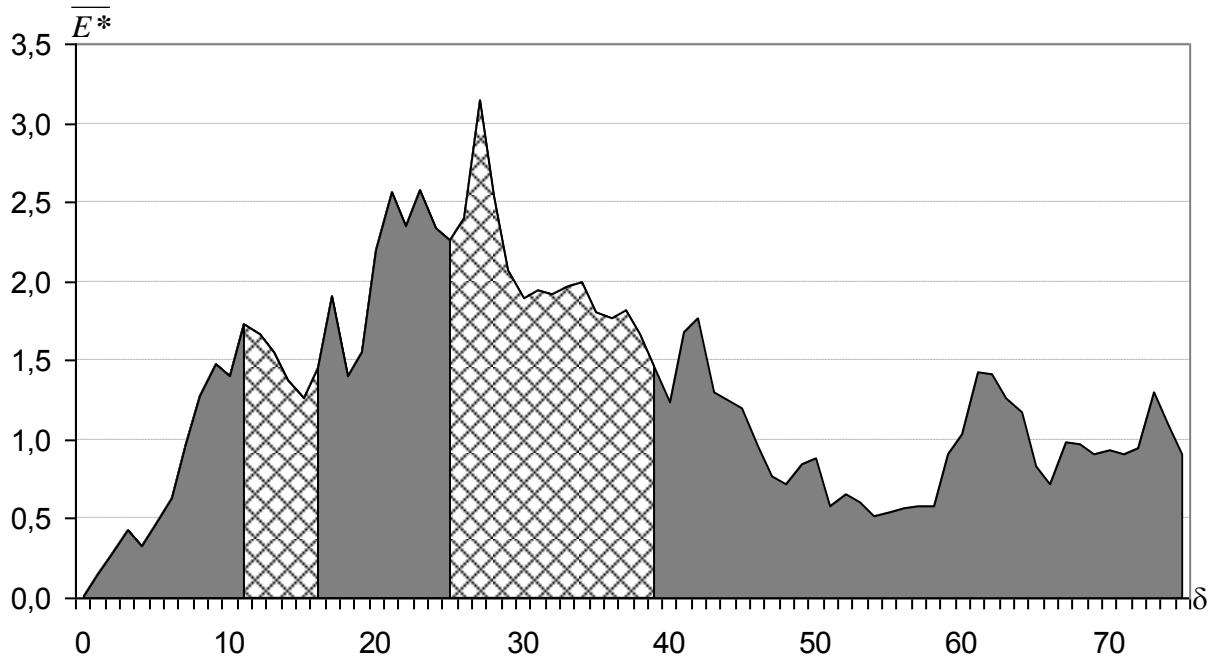


Рисунок 3.9 – Динаміка зміни КФЕ в процесі паралельної оптимізації СКД для словника ознак, сформованого з використанням випадкового пошуку

При цьому значення глобального усередненого КФЕ було більше за відповідне значення, отримане на нульовій ітерації і дорівнювало 3,1488. Як видно з рис. 3.9, оптимальні значення геометричних параметрів контейнерів класів дорівнювали: $d_1^* = 5$ для базового класу X_1^o при міжцентровій відстані в $d_c = 7$ кодових одиниць; $d_2^* = 5$ для класу X_2^o при міжцентровій відстані в $d_c = 7$ кодових одиниць; $d_3^* = 8$ для класу X_3^o при міжцентровій відстані в $d_c = 9$ кодових одиниць; $d_4^* = 7$ для класу X_4^o при міжцентровій відстані в $d_c = 10$ кодових одиниць і $d_5^* = 6$ для класу X_5^o при міжцентровій відстані в $d_c = 7$ кодових одиниць. Тобто після такої оптимізації не вдалося побудувати безпомилковий класифікатор, на що вказує значна ступінь перетину контейнерів класів і значення таких точнісних характеристик: $D_{1,1} = 0,95$ та $D_{2,1} = 1$; $D_{1,2} = 0,7$ та $D_{2,2} = 0,9$; $D_{1,3} = 1$ та $D_{2,3} = 1$; $D_{1,4} = 1$ та $D_{2,4} = 1$ і $D_{1,5} = 0,97$ та $D_{2,5} = 0,85$.

Таким чином, здійснена оптимізація дозволила значно підвищити ефективність системи, що навчається.

Аналіз рис. 3.8 вказує на певну проблему у формуванні варіантів словників, оскільки кількість варіантів словників, для яких середнє значення КФЕ було більше за відповідне значення КФЕ для повного словника (темна горизонтальна лінія), складає приблизно 1,7% від загальної кількості згенерованих. Для вирішення цієї проблеми було застосовано алгоритм оптимізації словника з використанням адаптивного випадкового пошуку. Аналогічно до оптимізації без адаптації формувалися 300 потенційних словників, для кожного з яких здійснювалася оптимізація відповідних параметрів функціонування та визначалося усереднене глобальне значення КФЕ. Після цього проводилася модифікація ймовірностей включення кожної ознаки до наступного словника в залежності від того, чи була присутня ця ознака в поточному словнику, а також чи дозволяє словник, що аналізується, підвищити ефективність у порівнянні з попередніми. Результат такої оптимізації подано на рис. 3.10. Аналіз рис. 3.10 показує, що найбільш ефективним виявився варіант словника № 136, який мав потужність $N=29$ і складався з 2, 3, 7 – 10, 12, 13, 15 – 20, 23 – 25, 28, 31, 32, 34, 35, 39, 44, 46, 50, 51, 53, 55 ознаки.

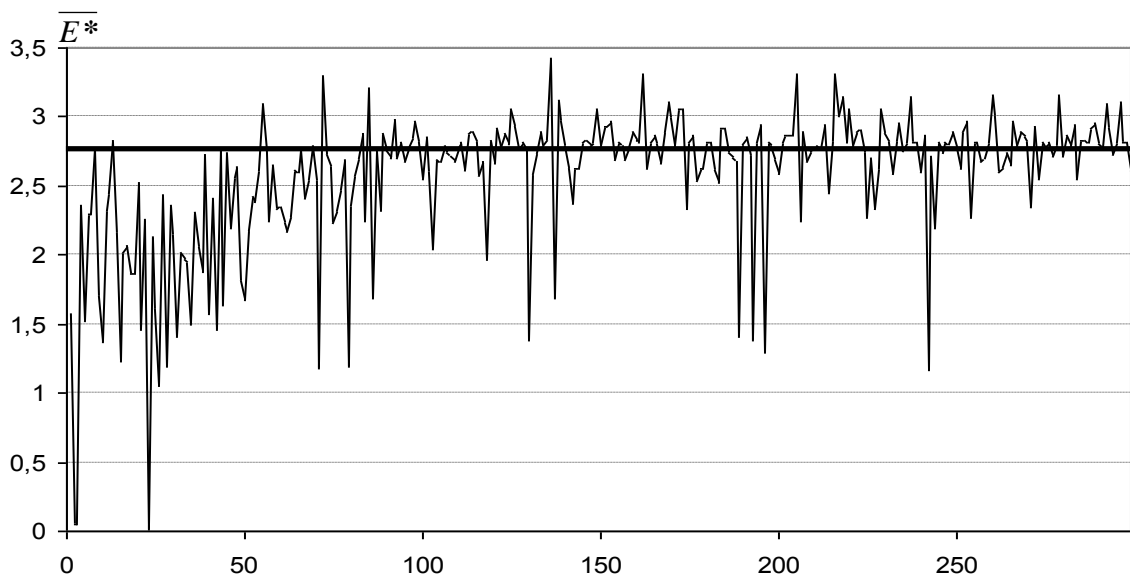


Рисунок 3.10 – Оптимізація словника ознак з використанням адаптивного випадкового пошуку

Динаміка зміни усередненого КФЕ при оптимізації СКД для сформованого словника ознак подана на рис. 3.11.

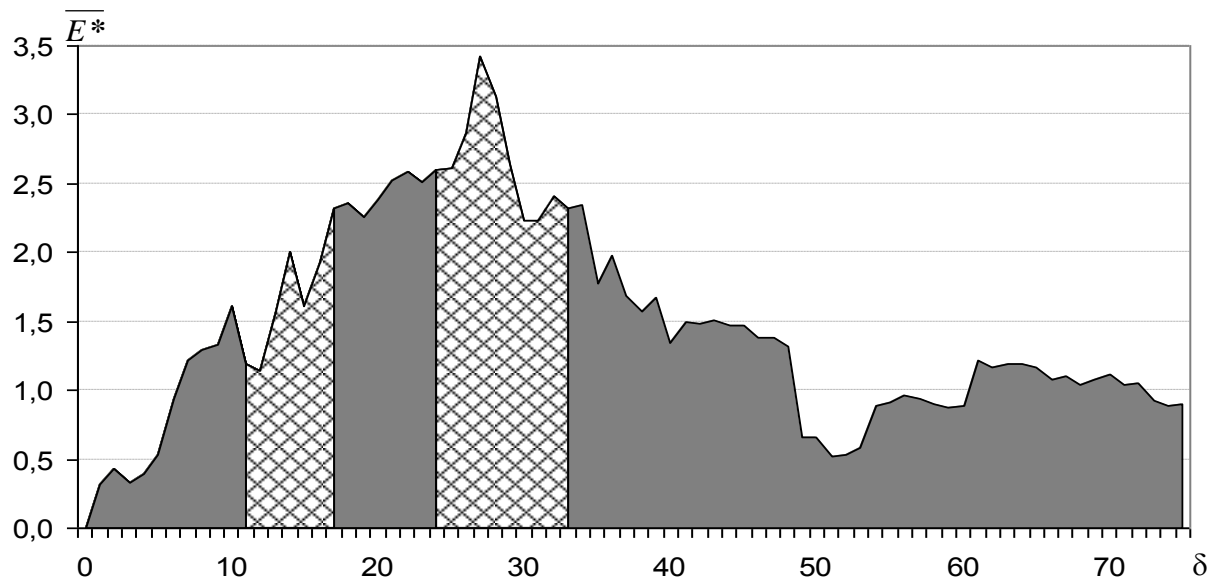


Рисунок 3.11 –Зміна КФЕ в процесі паралельної оптимізації СКД для словника ознак, сформованого з використанням адаптивного випадкового пошуку

Аналіз рис. 3.11 показує, що оптимальне значення відхилення верхнього та нижнього контрольного допуску від центра розсіювання значень кожної ознаки при паралельній оптимізації за симетричною стратегією S_1 δ^* не змінилося і дорівнює 27. При цьому значення усередненого КФЕ було більше за значення, отримане на нульовій ітерації і дорівнювало 3,42673. Оптимальні значення геометричних параметрів контейнерів класів розпізнавання дорівнювали $d_0^*=4$ для базового класу X_0^o при міжцентровій відстані в $d_c=6$ кодових одиниць, $d_1^*=4$ для класу X_1^o при міжцентровій відстані в $d_c=6$ кодових одиниць, $d_2^*=8$ для класу X_2^o при міжцентровій відстані в $d_c=10$ кодових одиниць, $d_3^*=8$ для класу X_3^o при міжцентровій відстані в $d_c=10$ кодових одиниць, $d_4^*=5$ для класу X_4^o при міжцентровій відстані в $d_c=6$ кодових одиниць. Тобто після такої

оптимізації не вдалося побудувати безпомилковий класифікатор, на що вказують значення точнісних характеристик: $D_{1,0}=0,97$ та $D_{2,0}=1$, $D_{1,1}=0,78$ та $D_{2,1}=0,95$, $D_{1,2}=1$ та $D_{2,2}=1$, $D_{1,3}=1$ та $D_{2,3}=1$, $D_{1,4}=0,95$ та $D_{2,4}=0,93$. Аналіз рис. 3.10 вказує на те, що кількість варіантів словників, для яких середнє значення КФЕ було більше за відповідне значення КФЕ для повного словника (темна горизонтальна лінія), складає 46,7% від загальної кількості згенерованих.

Таким чином, додавання алгоритму оптимізації словника ознак за методами випадкового пошуку властивості адаптивності дозволяє не тільки підвищити ефективність системи, але й перевірити значно більшу кількість наближених до оптимального вигляду словників ознак розпізнавання.

3.4 Оптимізація словника ознак за послідовно-спадковою селекцією

Реалізацію інформаційно-екстремального методу оптимізації словника ознак, будемо здійснювати за схемою алгоритму ПСС, який належить до групи евристичних точнісних методів [105,106] і є найбільш поширеним через його логічність і простоту. Розглянемо кроки реалізації алгоритму ПСС за категорійною моделлю, наведеною на рис. 2.4:

1. За алгоритмом паралельної оптимізації контрольних допусків (3.2.1) визначаються оптимальні параметри навчання та максимальне усереднене значення КФЕ для повного словник ознак.

2. Формується лічильник кількості ознак у словнику: $i := 0$, $i \in [1; N]$.

3. Формується множина варіантів словників $\{\Sigma_{i,h}\}$, $h = \overline{1, i}$, потужність яких буде на одиницю менше від поточних.

4. За алгоритмом оптимізації контрольних допусків визначаються оптимальні параметри навчання та максимальне усереднене значення КФЕ для кожного варіанту словника $\overline{E}_{\Sigma_{i,h}}$.

5. Визначається оптимальний поточний словник Σ_i^* , для якого $\bar{E}_{\Sigma_i^*} = \max_h \bar{E}_{\Sigma_{i,h}^*}$.
6. $i=i-1$.
7. Якщо $i>0$, виконуємо пункт 3, інакше – 8.
8. Визначаємо оптимальний словник Σ^* , для якого $\bar{E}_{\Sigma^*} = \max_i \bar{E}_{\Sigma_i^*}$.
9. Зупин.

На рис 3.12 показано залежність усередненого за алфавітом класів ентропійного КФЕ (2.3.1) від параметра поля контрольних допусків на ознаки, , обчисленого при нульовій ітерації за алгоритмом (3.2.1).

Аналіз рис. 3.12 показує, що оптимальне значення параметра поля контрольних допусків дорівнює $\delta^* = \pm 28$. При цьому оптимальні значення радіусів контейнерів дорівнюють: для базового класу $d_1^* = 7$ (тут і далі розмірність відрізків, побудованих в просторі Хеммінга, подається у кодових одиницях) при максимальному значенні КФЕ $E_1^* = 2,98974$; для класу X_2^o – $d_2^* = 8$, для класу X_3^o – $d_3^* = 16$ і для класу X_4^o – $d_4^* = 11$, а середнє значення радіусів контейнерів для початкового (надлишкового неоптимального словника ознак дорівнює $\bar{d}_0 = 10,5$.

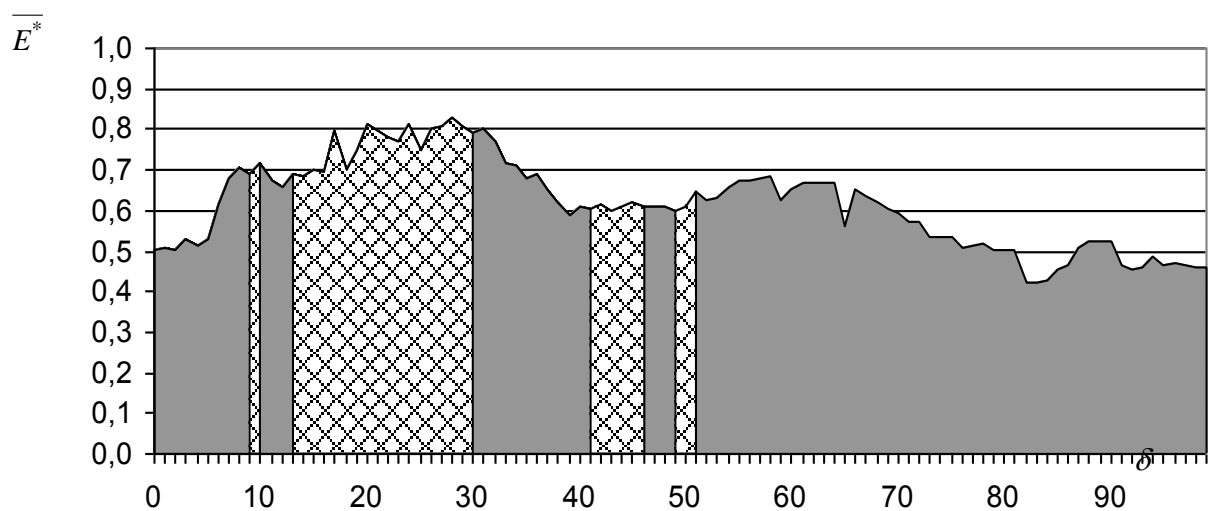


Рисунок 3.12 – Графік зміни ентропійного критерію при оптимізації параметра поля допусків для початкового словника ознак

Найближчим класом для базового класу X_1^o є клас X_2^o (міжцентрова відстань $d_c=9$), для класу X_2^o – базовий клас X_1^o ($d_c=9$), для класу X_3^o – так само базовий клас ($d_c=16$) і для класу X_4^o – клас X_1^o ($d_c=13$). Тобто середня міжцентрова відстань для заданого алфавіту дорівнює $\bar{d}_{0c}=11,75$. Таким чином, можна зробити висновок, що класи суттєво перетинаються, що підтверджують максимальні значення КФЕ та значення точнісних характеристик системи, які не всі досягають своїх граничних значень. Наприклад, для базового класу X_1^o максимальне значення КФЕ дорівнює $E_1^*=0,81747$ ($E_{гран}=1,0$), а перша достовірність і помилка другого роду відповідно $D_1=0,93$ і $\beta=0$; для класу X_2^o – $E_2^*=0,57732$, $D_1=0,63$ і $\beta=0$; для класу X_3^o – $E_3^*=1$, $D_1=1$, $\beta=0$ і для класу X_4^o – $E_4^*=0,91729$, $D_1=0,98$ і $\beta=0$.

Динаміку зміни значення максимуму усередненого критерію \bar{E}^* при оптимізації словника ознак за ППС показано на рис. 3.13.

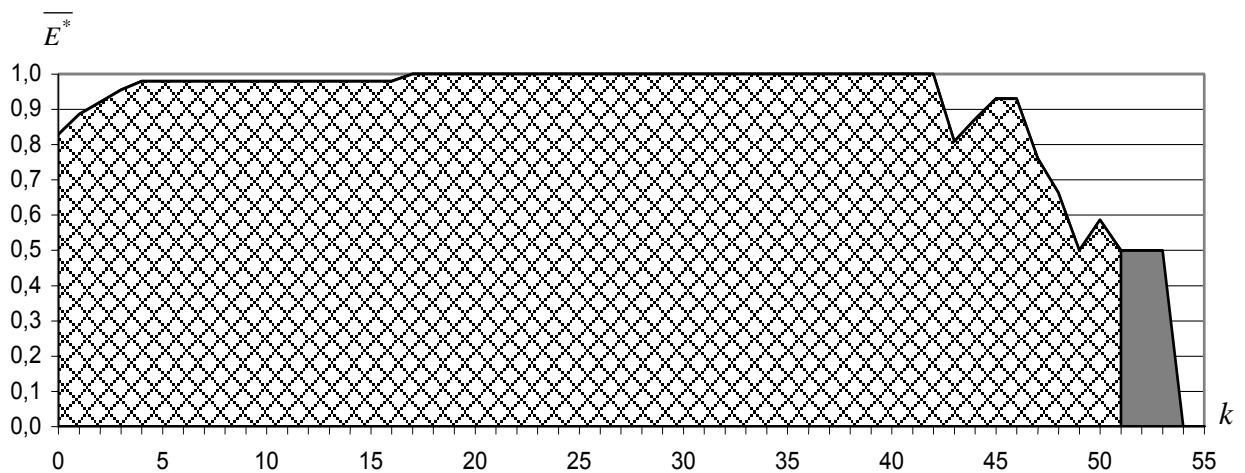


Рисунок 3.13 – Графік зміни максимального усередненого значення ентропійного КФЕ при оптимізації словника ознак за ППС

Аналізуючи рис. 3.14, можна зробити висновок, що оптимальний словник ознак було одержано на 17 кроці навчання. Після чого значення \bar{E}^* набуло свого максимального значення і залишалося без змін протягом

наступних 25 кроків, а потім значно зменшилося, хоча цей процес був немонотонним. Останній факт вказує на те, що процес оптимізації словників необхідно проводити навіть тоді, коли варіанти словників, одержані на поточному кроці характеризуються меншими значеннями \overline{E}^* . Оскільки існує ймовірність того, що на наступному кроці \overline{E}^* зросте і буде досягнуто нового максимуму. Процес оптимізації СКД для оптимального словника Σ^* ілюструє рис. 3.15.

Аналіз рис. 3.15 показує, що оптимальним значенням параметра поля контрольних допусків для всіх ознак дорівнює $\delta^* = \pm 19$ (тут і далі у відносних одиницях). При цьому оптимальні значення геометричних параметрів контейнерів відповідних класів приймають такі значення: для базового класу X_1^o радіус контейнера дорівнює $d_1^* = 4$, для класу X_2^o – $d_2^* = 3$, для класу X_3^o – $d_3^* = 4$, для класу X_4^o – $d_4^* = 4$.

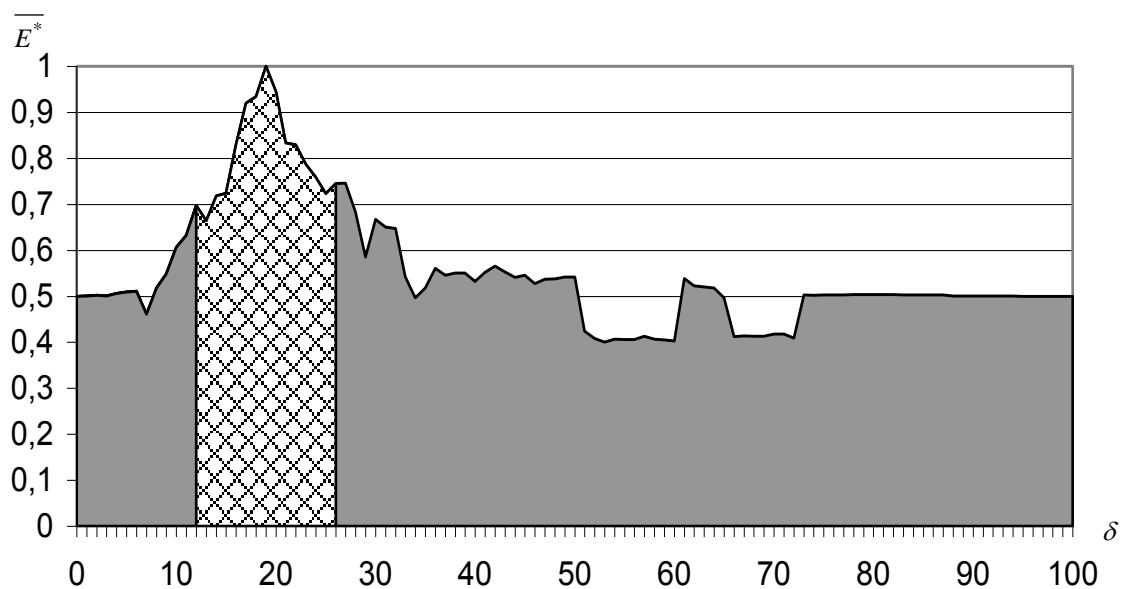


Рисунок 3.15 – Графік залежності ентропійного КФЕ від параметра поля контрольних допусків для оптимального словника ознак

При цьому середнє значення радіусів контейнерів дорівнює $\bar{d} = 3,75$. Крім того найближчим класом до базового є клас X_2^o (міжцентрова відстань $d_c=6$), до класу X_2^o – клас X_4^o ($d_c=6$), до класу X_3^o – клас X_4^o ($d_c=7$), до класу X_4^o – клас X_2^o ($d_c=6$). Тобто середня міжцентрова відстань для заданого алфавіту дорівнює $\bar{d}_c = 6,25$.

Оптимізація початкового словника ознак за модифікованим алгоритмом ППС проводилося шляхом формування на кожному кроці такого варіанта словника, який включав лише групу ознак, що були задіяні при виконанні максимально-дистанційного принципу оптимізації простору розбиття (рис. 2.7). При цьому значення відповідних координат еталонних векторів контейнерів класів розпізнавання не збігалися, що свідчить про збільшення кодової відстані між центрами класів. Для сформованого таким чином словника проводилася оптимізація параметрів навчання й обчислювалося максимальне усереднене значення КФЕ. Результати оптимізації показано на рис. 3.16:

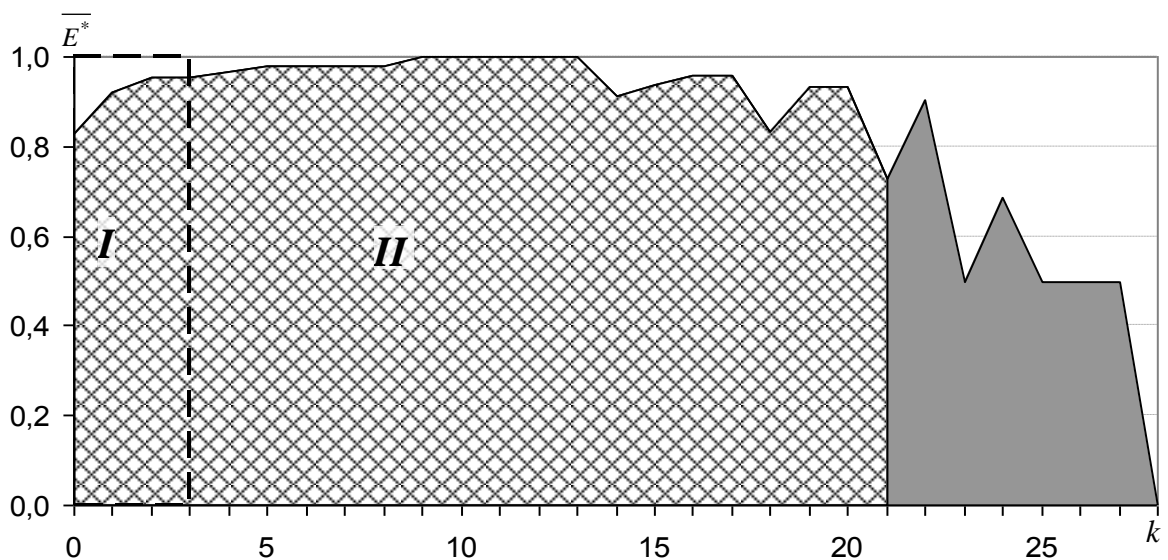


Рисунок 3.16 – Графік зміни максимального усередненого значення КФЕ при оптимізації словника за модифікованим алгоритмом ПСС

На рис 3.16 область I відповідає етапу оптимізації, на якому словник формується лише з тих ознак, що були задіяні для виконання максимально-

дистанційного принципу оптимізації геометричних параметрів розбиття простору ознак. При цьому було виконано три ітерації, що скоротили кількість ознак до 26 із 55 початкових. Область II відповідає оптимізації за алгоритмом ПСС, що проводилася для сформованого на першому етапі словника. Аналіз рис. 3.16 показує, що оптимальний словник сформовано на 10-му кроці навчання. При цьому оптимальний словник Σ^{**} містить 16 ознак початкового словника. Процес оптимізації СКД для словника Σ^{**} показано на рис. 3.17.

Аналіз рис. 3.17 показує, що оптимальне значенням параметра СКД дорівнює $\delta^* = \pm 20$. При цьому оптимальні значення геометричних параметрів контейнерів відповідних класів приймають такі значення: для базового класу X_1^o радіус контейнера дорівнює $d_1^* = 4$, для класу X_2^o – $d_2^* = 5$, для класу X_3^o – $d_3^* = 5$, для класу X_4^o – $d_4^* = 5$. Тобто середнє значення радіусів контейнерів дорівнює $\bar{d} = 4,75$.

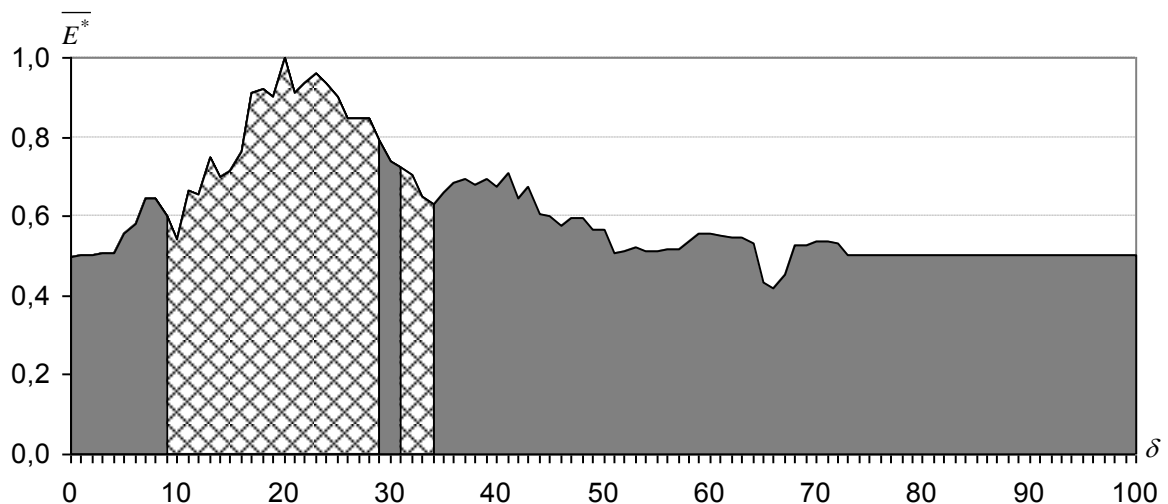


Рисунок 3.17 – Графік зміни максимального усередненого значення КФЕ при оптимізації СКД для оптимального словника

Крім того найближчим класом до базового є клас X_2^o (міжцентрова відстань $d_c=6$), до класу X_2^o – клас X_1^o ($d_c=6$), до класу X_3^o – клас X_4^o ($d_c=8$), до класу X_4^o – клас X_2^o ($d_c=7$). Тобто середня міжцентрова відстань

для заданого алфавіту дорівнює $\bar{d}_c = 6,75$, що більше у порівнянні з алгоритмом ПСС, який не враховує особливості ІЕІ-технології.

Таким чином, запропонована модифікація алгоритму ПСС з урахуванням особливостей навчання СППР у рамках ІЕІ-технологія дозволяє значно скоротити потужність словника ознак без зменшення усередненого значення КФЕ і без помітного збільшення часових витрат.

За результатами фізичного моделювання, виконаного за даними, одержаними за цехових умов виробництва складного мінерального добрива НРК у ВАТ „Сумхімпром”, доведено таке:

- оптимізація словника ознак у рамках запропонованого інформаційно-екстремального алгоритму навчання СППР призводить до зменшення середнього значення радіусів контейнерів класів у 2,8 рази, а середньої міжцентрової відстані для заданого алфавіту в 1,9 рази;

- запропонований інформаційно-екстремальний алгоритм навчання СППР дозволяє побудувати безпомилкові за багатовимірною навчальною матрицею вирішальні правила, оскільки в процесі оптимізації словника ознак максимальне значення ентропійного інформаційного критерію оптимізації параметрів навчання досягає в робочій області визначення його функції свого граничного значення $\bar{E} = E_{гран} = 1,0$, що показано на рис. 3.15.

Варто зазначити, що оскільки після оптимізації словника ознак зменшується його потужність через усунення неінформативних ознак, то вимірність простору ознак зменшується, а відповідно зменшуються і відстані між вершинами еталонних векторів-реалізацій класів. Тому можна стверджувати, що відносна середня міжцентрова відстань для заданого алфавіту після оптимізації збільшилася. Таким чином, в процесі навчання за розглянутим алгоритмом виконуються умови як мінімально-дистанційного принципу, який полягає в мінімізації середньої відстані векторів-реалізацій відповідних класів від ядра класу, так і максимально-дистанційного принципу, який полягає в максимізації середньої міжкласової відстані для заданого алфавіту.

3.5 Гібридний алгоритм оптимізації словника ознак розпізнавання

Застосування генетичних алгоритмів в ІЕІ-технології для пошуку глобального максимуму багато екстремальної функції інформаційного КФЕ дозволяє підвищити оперативність навчання системи за умови зниження у прийнятних межах точності обчислення КФЕ [138-140]. Це досягається шляхом заміни прямого перебору значень параметрів навчання на процедуру випадкового пошуку. Математичну модель гібридного інформаційно-екстремального алгоритму навчання СППР показано на рис. 3.18.

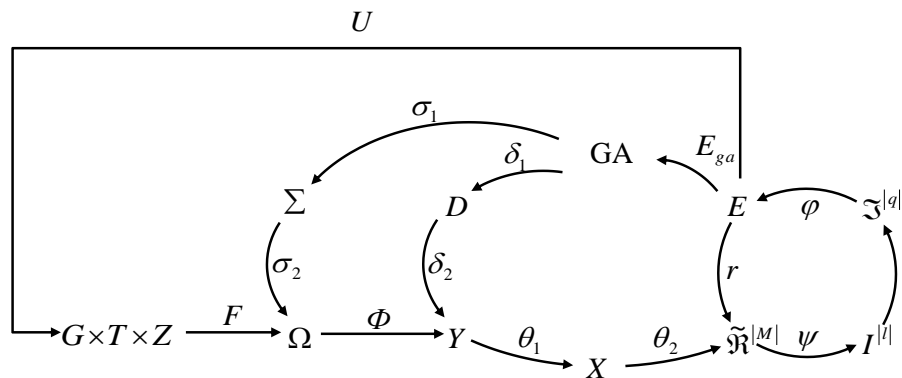


Рисунок 3.18 – Категорійна модель самонавчання СППР за гібридним алгоритмом

У діаграмі відображення множин (рис.3.18) оператор $E_{ga} : E \rightarrow GA$ формує популяцію GA певних значень параметрів функціонування системи. При цьому внутрішній цикл гібридного алгоритму оптимізує контейнери відповідних класів шляхом реалізації операторів базового контура.

Оскільки оптимізація словника ознак і системи контрольних допусків відбувається одночасно, то генетичний алгоритм дозволяє об'єднати кілька контурів оптимізації параметрів функціонування СППР в один, що збільшує оперативність навчання та дозволяє використовувати інші переваги генетичних алгоритмів. Крім того змінюються відповідні оператори, що

використовувалися в інших контурах, такі як σ_1 та δ_1 . Вони модифікуються шляхом додавання функції виділення значень відповідних параметрів із популяції.

Гібридний алгоритм оптимізації параметрів функціонування СППР, що самонавчається реалізується за такою схемою.

1. На першому кроці алгоритму випадковим чином формується початкова популяція $S(0) = \{s_1, s_2, \dots, s_m\}$, що складається з M істот. Кожна істота представляється геномом, що складається із бінарних генів, які описують словник ознак, та дійсночислових генів, що описують СКД на ознаки.

2. Проводиться навчання за інформаційно-екстремальним алгоритмом для словника ознак та СКД, які описуються геномом даної особини, та знаходиться значення фітнес-функції для цієї особини у вигляді

$$f(v) = \overline{E_v^*} - k \frac{|\Omega_v|}{|\Omega|}, \quad (3.5.1)$$

де $\overline{E_v^*}$ – середнє арифметичне значення інформаційного КФЕ навчання СППР для всіх класів із заданого алфавіту; $|\Omega_v|$ – кількість ознак розпізнавання у словнику; k – коефіцієнт пріоритету якості навчання над кількістю ознак розпізнавання, що впливає на швидкість всіх процесів оптимізації; $|\Omega|$ – кількість ознак у вихідному словнику.

Розмір популяції вибирався експериментально, а результати роботи гібридного алгоритму для різних розмірів популяцій показано на рис. 3.19-3.21.

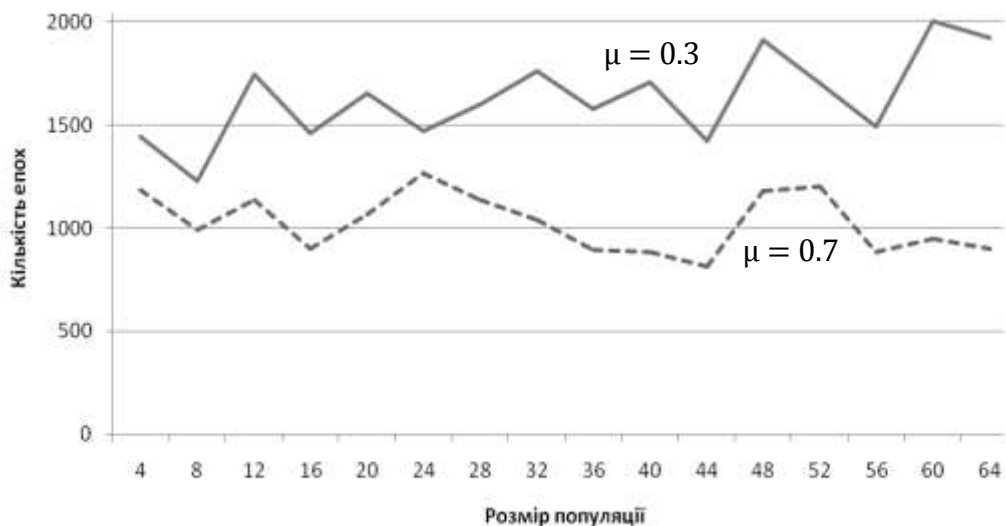


Рисунок 3.19 – Графіки залежності кількості епох від розміру популяції

Аналіз рис.3.19 показує, що кількість епох несуттєво збільшується при збільшенні розміру популяції при $\mu = 0,3$ і майже не змінюється при $\mu = 0,7$.

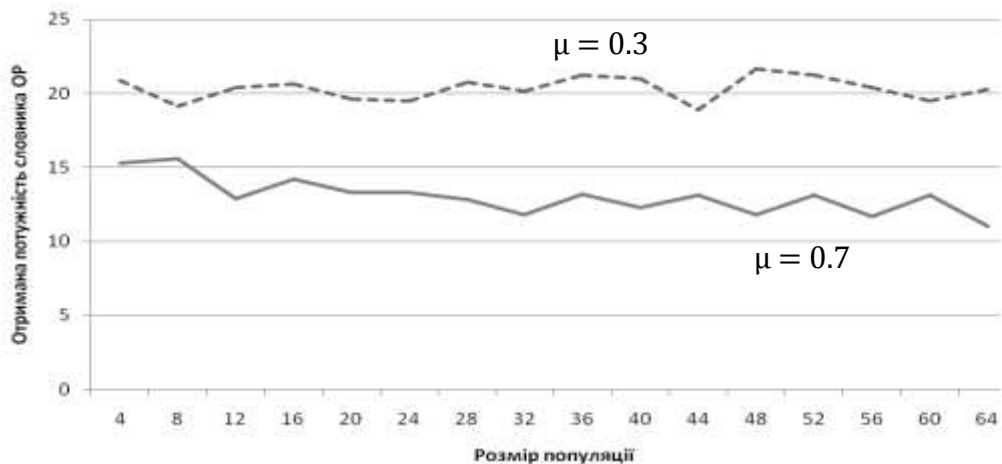


Рисунок 3.20 – Графіки залежності потужності словників ознак від розміру популяції

Аналіз рис.3.20 показує, що потужність словників ознак при $\mu = 0,3$ практично не змінюється, а при $\mu = 0,7$ незначно зменшується.

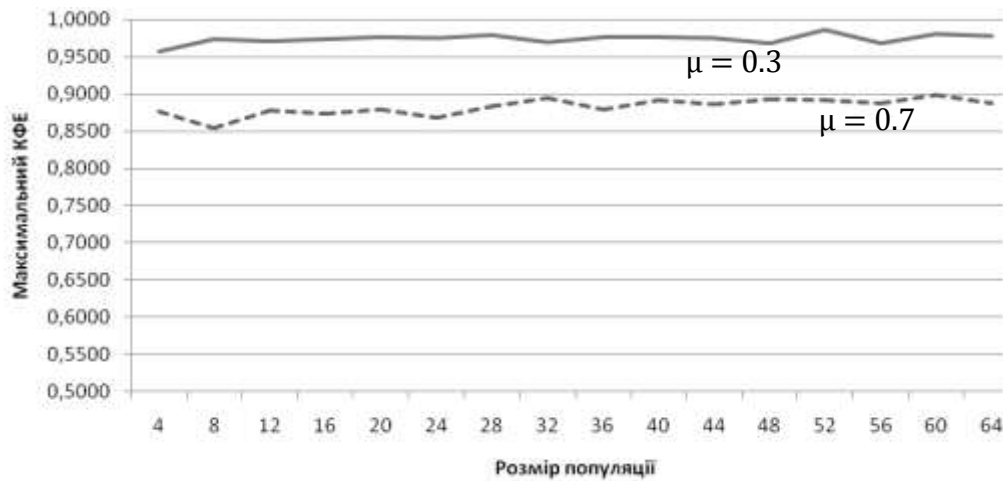


Рисунок 3.21 – Графіки залежності значень глобального максимуму критерію (3.4.1) від розміру популяції

Аналіз рис.3.21 показує, що значення глобального максимуму КФЕ навчання при збільшенні розміру популяції мають незначну тенденцію до збільшення, але суттєво залежать від імовірності мутації μ . Оскільки розмір популяції впливає на оперативність роботи алгоритму, то вибирається за оптимальне значення розміру популяції 32 особи.

Далі в роботі алгоритму починається цикл зміни епох. На кожному кроці всі особи популяції впорядковуються у порядку зменшення значення фітнес-функції (3.5.1) і гірша половина особин знищується. Особи, що залишилися, схрещуються попарно, породжуючи нові особи до заповнення вільних місць у популяції. Схрещування відбувається за допомогою оператора кросинговеру $s_t = \text{Crossing}(s_{c_1}, s_{c_2})$. Результуюча істота має геном, що формується із геномів батьківських особин таким чином, що кожен бінарний ген може бути як від однієї так і від другої особи рівно ймовірно, а кожен дійсно числовий ген результуючої особи є випадковою лінійною комбінацією:

$$g_c(i) = g_a(i) \cdot r + g_b(i) \cdot (1-r), \quad r \in [0;1].$$

При цьому кожен ген нової особи може бути вибраний випадковим чином із імовірністю мутації μ , яка вибирається за результатом роботи генетичного алгоритму від обраного значення. Результати експерименту наведено на рис. 3.22-3.24.

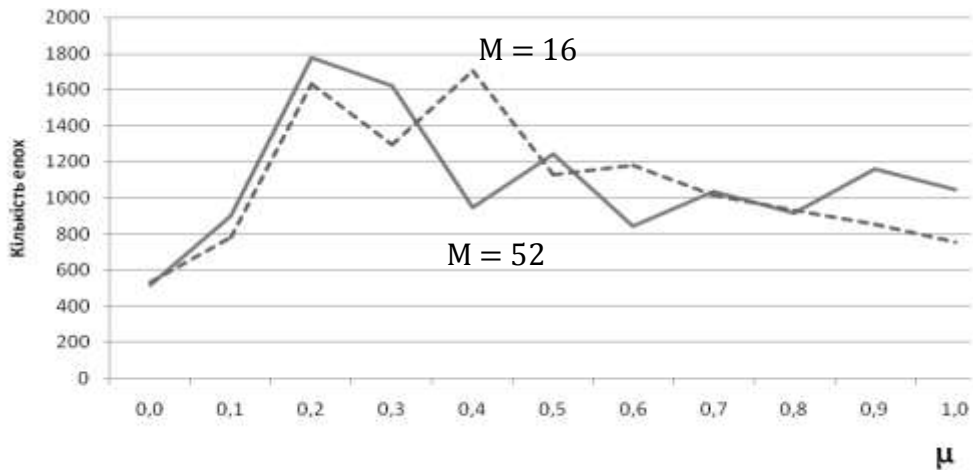


Рисунок 3.22– Графіки залежності кількості епох від імовірності мутації

Аналіз рис.3.22 показує, що кількість епох досягає сумарного максимального значення для різних розмірів популяції, а потім із збільшенням імовірності мутації має тенденцію до зменшення.

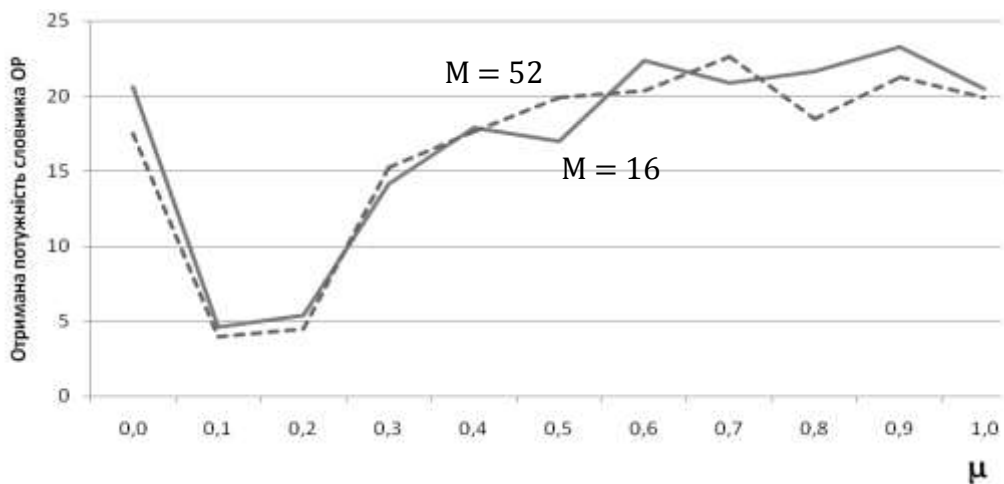


Рисунок 3.23– Графіки залежності потужності словників від імовірності мутації

Аналіз рис.3.23 показує, що мінімальна потужність словників ознак для різних розмірів популяцій має місце при ймовірності мутації в інтервалі $[0,1;0,2]$, а потім при збільшенні ймовірності мутації має тенденцію до збільшення.

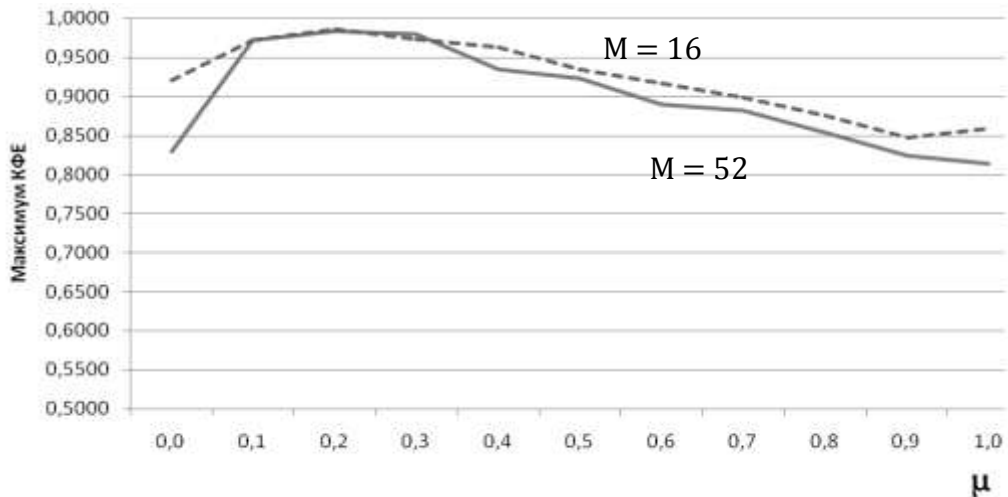


Рисунок 3.24– Графіки залежності максимуму критерію (3.4.1) від ймовірності мутації

Аналіз рис.3.24 показує, що значення глобального максимуму КФЕ навчання при зазначених розмірах популяції є найбільш наближеними до точного при ймовірності мутації $\mu = 0,2$ і при її збільшенні мають тенденцію до збільшення похибки

Таким чином, аналізуючи рис. 3.22-3.24 можна стверджувати, що ймовірність мутації μ значно впливає на оптимальність результату, тому слід вибирати $\mu = 0,2$.

Алгоритм вважається закінченим, якщо впродовж W епох не було покращено максимальне значення фітнес-функції (3.5.1) серед усіх особин.

На рис. 3.25-3.27 показана залежність результатів роботи генетичного алгоритму від періоду очікування W .

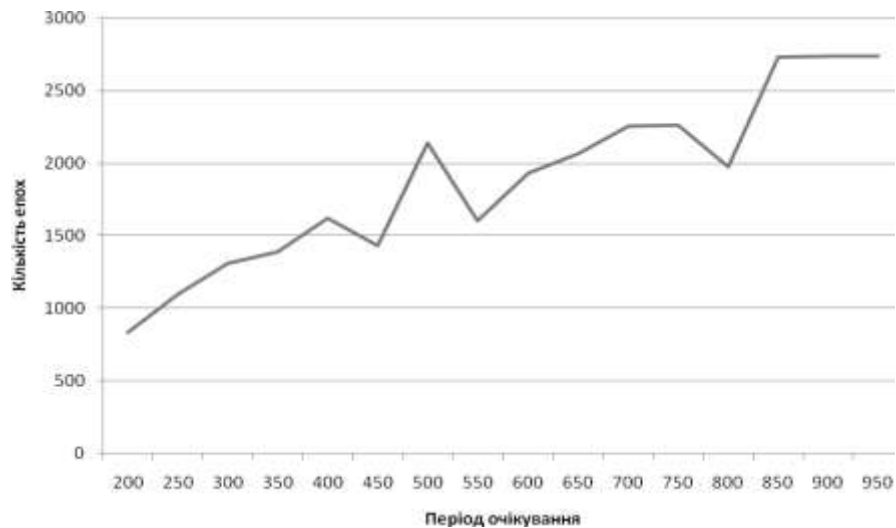


Рисунок 3.25 – Графіки залежності кількості епох від періоду очікування

Аналіз рис.3.25 показує, що кількість епох збільшується при збільшенні періоду очікування.

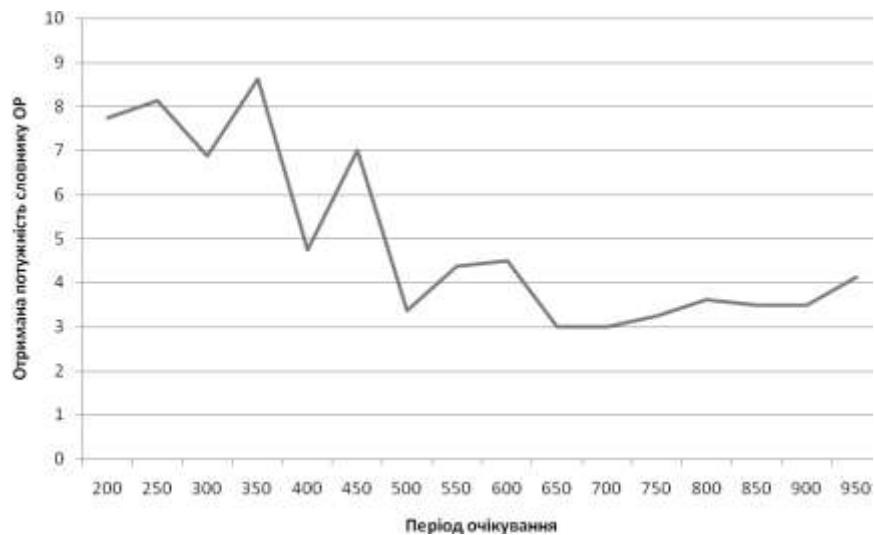


Рисунок 3.26– Графіки залежності потужності словників від періоду очікування

Аналіз рис.3.26 показує, що потужності словників зменшуються при збільшенні періоду очікування, досягаючи мінімальних значень в інтервалі [650; 720].

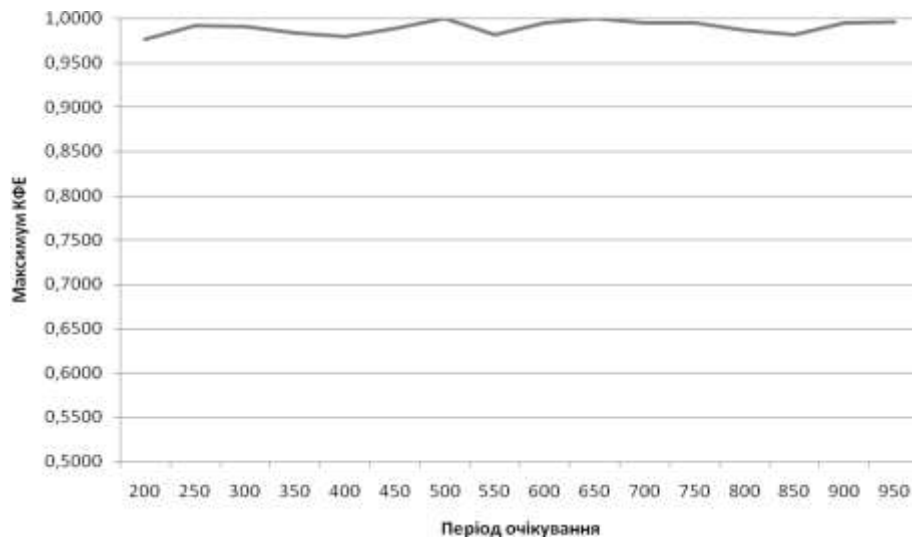


Рисунок 3.27– Графіки залежності значень глобального максимуму критерію (3.4.1) від періоду очікування

Аналізуючи рис. 3.25-3.27 можна зробити висновок, що період очікування функціонально пов'язаний із кількістю епох, незначно впливає на КФЕ та дає словники ознак мінімальної потужності при значенні $W = 700$.

Таким чином, специфікою використання запропонованого гібридного інформаційно-екстремального алгоритму є як оптимізація параметрів навчання системи, так і оптимізація параметрів генетичного алгоритму.

3.6 Унімодальний класифікатор

У працях [6-8] розглядалося питання оптимізації у рамках ІЕІ-технології просторово-часових параметрів функціонування, які впливають на функціональну ефективність СППР для мультимодального класифікатора, що характеризується наявністю декількох центрів розсіювання реалізацій класів із заданого алфавіту. Але на практиці існують задачі, які потребують застосування унімодального класифікатора з одним загальним центром

розсіювання для реалізацій усіх класів розпізнавання. До цих задач відносяться, наприклад, задачі керування із трьохальтернативною системою оцінок типу «Норма», «Менше норми» і «Більше норми».

Розглянемо алгоритм навчання СППР з оптимізацією контрольних допусків на ознаки розпізнавання для унімодального класифікатора

Нехай для алфавіту нечітких класів розпізнавання $\{X_m^o \mid m = \overline{1, M}\}$, які характеризують M допустимих функціональних станів складного організаційно-технічного комплексу, сформовано апріорно класифіковану нечітку навчальну матрицю типу «об'єкт-властивість» $\|y_{m,i}^{(j)}\|$, $i = \overline{1, N}$, $j = \overline{1, n}$, де N, n – кількість ознак розпізнавання і випробувань відповідно. Крім того дано вектор параметрів функціонування СППР $g = \langle x_1, d_{m,1}, d_{m,2}, \delta \rangle$, де x_1 δ – еталонний вектор-реалізація базового класу X_1^o , який визначає геометричний центр унімодального класифікатора – центр розсіювання векторів-реалізацій усіх класів розпізнавання; $d_{m,1}$, $d_{m,2}$ – менший і більший радіуси контейнера класу X_m^o , що відновлюється в радіальному базисі простору ознак розпізнавання, відповідно; δ – параметр поля контрольних допусків. При цьому задано такі обмеження: $x_1 \in \Omega_B$, де Ω_B – бінарний простір ознак розпізнавання (Хеммінга); $d_{m,1} = d_{m-1,2} < [0; d(x_m \oplus x_c) - 1]$, де $d(x_m \oplus x_c)$ – кодова відстань центру класу x_m^o від центру найближчого (сусіднього) до нього класу X_c^o і параметр $\delta \in [0; \delta_H / 2]$, де δ_H – нормоване (експлуатаційне) поле допусків для відносної шкали вимірювання ознак, яке є областю значень для параметра контрольного поля допусків δ . При цьому нижній допуск на i -ту ознаку дорівнює $A_{H_i} = y_{1,i} - \delta$, а верхній допуск $-A_B = y_{1,i} + \delta$. Тут $y_{1,i}$ – i -та ознака еталонного вектора-реалізації y_1 базового класу X_1^o . На етапі навчання необхідно оптимізувати параметри навчання за усередненим інформаційним критерієм функціональної ефективності (КФЕ)

$$\bar{E} = \frac{1}{M} \sum_{m=1}^M E_m^*, \quad (3.6.1)$$

де E_m^* – глобальний максимум інформаційного КФЕ навчання СППР розпізнавати реалізації класу X_m^o , що шукається в робочій області визначення його функції, і на етапі екзамену прийняти високодостовірне рішення про належність вектора-реалізації образу, що розпізнається, до деякого класу із заданого алфавіту $\{X_m^o\}$.

Категорійну модель унімодального класифікатора для побудови оптимального розбиття простору ознак на класи розпізнавання, яке утворює сукупність вирішальних правил, або класифікатор, подамо у вигляді діаграми відображень множин. Так само, як і для мультимодального класифікатора, діаграма відображення множин, що формується при навчанні унімодальної СППР, в зворотному зв'язку містить побудовані за принципом повної композиції контури оптимізації відповідних параметрів навчання. Основною відмінністю моделі унімодальної СППР є наявність додаткового контуру вибору відповідного вирішального правила для реалізації оператора перевірки основної статистичної гіпотези. На рис. 3.28 показана діаграма відображення множин, що формуються в процесі навчання унімодальної СППР з оптимізацією системи контрольних допусків (СКД) на ознаки розпізнавання.

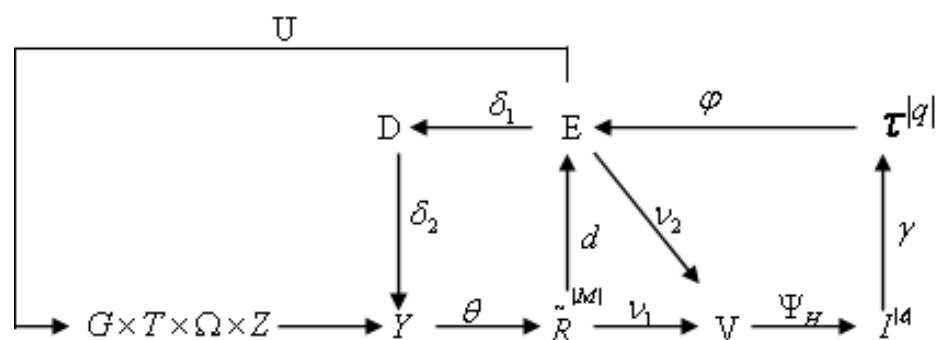


Рисунок 3.28– Категорійна модель навчання унімодальної СППР

У діаграмі (рис.3.28) контур, показаний на рис 3.29 оптимізує геометричні параметри побудованих в радіальному базисі простору ознак розпізнавання контейнерів класів розпізнавання

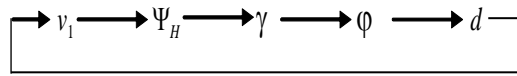


Рисунок 3.29–Контур оптимізації контейнерів класів розпізнавання для унімодальної СППР

У контурі (рисю3.29)) оператор v_1 відображає розбиття $\tilde{\mathfrak{R}}^{|M|}$ на відповідні вирішальні правила, оператор Ψ_H перевіряє основну статистичну гіпотезу за відповідним вирішальним правилом і оператор d змінює геометрію контейнерів відповідних класів розпізнавання. Таким чином, для унімодального класифікатора оператор перевірки основної статистичної гіпотези Ψ розглядається як композиція операторів $\Psi = v_1 \circ \Psi_H$.

На рис 3.30 показано контур оптимізації контрольних допусків на ознаки розпізнавання

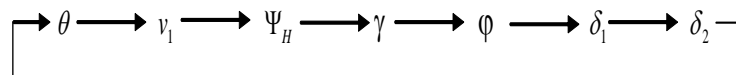


Рисунок 3.30–Контур оптимізації контрольних допусків на ознаки розпізнавання для унімодальної СППР

У контурі (рис. 3.30) оператор δ_1 дає дозвіл на формування наступних кроків навчання СППР в процесі пошуку глобального максимуму КФЕ в робочій області його визначення, а оператор δ_2 формує відповідні контрольні допуски на ознаки розпізнавання.

Специфічним контуром для діаграми відображення множин при побудові унімодального класифікатора є контур, показаний на рис. 3.31

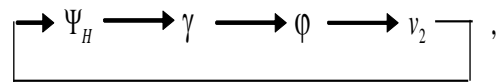


Рисунок 3.31–Контур оптимізації радіусів контейнерів класів розпізнавання для унімодальної СППР

У контурі (рис. 3.31) оператор ν_2 – здійснює перехід до наступного вирішального правила після оптимізації контейнера попереднього класу розпізнавання.

Алгоритм навчання унімодальної СППР з оптимізацією СКД на ознаки розпізнавання, як і для мультимодальної, полягає у реалізації структурованої двоциклическої ітераційної процедури пошуку глобального максимуму інформаційного КФЕ (3.6.1) в робочій області визначення його функції. Для інформаційно-екстремального алгоритму паралельної оптимізації СКД, при якому параметр поля контрольних допусків δ змінюється одночасно для всіх ознак розпізнавання, така процедура має вигляд

$$\delta^* = \arg \max_{G_\delta} \{ \max_{G_E} \bar{E} \}, \quad (3.6.2)$$

де G_δ – область допустимих значень відповідних контрольних допусків на ознаки розпізнавання; G_E – область допустимих значень інформаційного КФЕ (3.6.1) навчання СППР.

У процедурі (3.6.2) внутрішній цикл реалізує базовий алгоритм навчання, основними задачами якого є:

- обчислення інформаційного КФЕ навчання СППР;

- пошук глобального максимуму КФЕ в робочій (допустимій) області визначення його функції;
- оптимізація геометричних параметрів контейнерів класів розпізнавання..

При цьому специфіка базового алгоритму навчання унімодальної СППР полягає у відсутності процедури визначення для кожного класу найближчого сусіда, оскільки класи розпізнавання апріорно є впорядкованими, що суттєво підвищує оперативність навчання.

Розглянемо у рамках ІЕІ-технології алгоритм навчання унімодальної СППР з паралельною оптимізацією контрольних допусків на ознаки розпізнавання. Вхідні дані: масив реалізацій класів розпізнавання $\{y_{m,i}^{(j)} \mid m = \overline{1, M}; i = \overline{1, N}; j = \overline{1, n}\}$; система нормованих допусків $\{\delta_{H,i} \mid i = \overline{1, N}\}$, що визначає область значень відповідних контрольних допусків. Попередньо для кожної ознаки визначається ціна градації її шкали виміру, що дозволяє обчислювати на кожному кроці навчання нижній і верхній контрольні допуски відповідно:

$$A_{KH,i} = y_{1,i} - \delta; \quad A_{KB,i} = y_{1,i} + \delta, \quad (3.6.3)$$

де $y_{1,i}$ – вибіркове середнє значення i -ї ознаки у векторах-реалізаціях базового класу X_1^o .

Реалізація алгоритму навчання унімодальної СППР з паралельною оптимізацією контрольних допусків на ознаки розпізнавання здійснюється за такою схемою:

- 1) обнулюється лічильник кроків зміни параметра δ (кроків навчання): $l:=0$;
- 2) $\delta: l:=l+1$;

3) на кожному кроці навчання за формулами (3.6.3) обчислюються нижній $A_{HK,i}[l]$ і верхній $A_{HK,i}[l]$ контрольні допуски для всіх ознак розпізнавання;

4) реалізується базовий алгоритм навчання і визначається поточний глобальний максимум усередненого за алфавітом класів розпізнавання інформаційного критерію $\bar{E}[l]$ в робочій області визначення його функції.

5) якщо в робочій області визначення функції інформаційного критерію має місце $\bar{E}[l] \leq \text{extr max } \bar{E}$, де $\text{extr max } \bar{E}$ – граничний максимум усередненого за алфавітом класів розпізнавання інформаційного критерію (3.6.1), то виконується пункт 6, інакше – пункт 8 ($\bar{E}[0] = 0$);

6) якщо $\delta \leq \delta_H / 2$, то виконується пункт 2, інакше – пункт 7;

7) $\text{extr max } \bar{E} := \max_{\{l\}} \bar{E}^*[l]$

8) $\delta^* := \arg \text{extr max } \bar{E}$

9) для параметра δ^* обчислюються оптимальні нижні $\{A_{HK,i}^*\}$ і верхні $\{A_{BK,i}^*\}$ контрольні допуски на ознаки розпізнавання;

10) виведення оптимальних двійкових еталонних векторів-реалізацій $\{x_m^*\}$, вершини яких визначають геометричні центри контейнерів класів розпізнавання;

11) ЗУПИН.

Як КФЕ навчання для двохальтернативних рішень застосуємо модифікований критерій Кульбака (2.4.2). При цьому суми $K_{1,m}^{(k)}$ і $K_{2,m}^{(k)}$ ($K_{1,m}^{(k)}[0] = 0$, $K_{2,m}^{(k)}[0] = 0$) обчислюються на k -му кроці навчання унімодальної СППР відповідно

$$\text{if } x_m^{(j)} \notin X_m^o \text{ then } K_{1,m}^{(k)}[j] := K_{1,m}^{(k)}[j-1] + 1;$$

$$\text{if } x_c^{(j)} \in X_m^o \text{ then } K_{2,m}^{(k)}[j] := K_{2,m}^{(k)}[j-1] + 1,$$

де $x_c^{(j)}$ – j -та реалізація найближчого (сусіднього) класу X_c^o .

Належність деякої реалізації $x^{(j)}$, наприклад, класу X_m^o для унімодального класифікатора здійснюється за правилом

$$\text{if } d_{m-1}^* < d[x_1 \oplus x^{(j)}] < d_m \text{ then } x^{(j)} \in X_m^o \text{ else } x^{(j)} \notin X_m^o,$$

де d_{m-1}^* – визначений в процесі навчання оптимальний радіус контейнера лівого сусіднього класу; $d[x_1 \oplus x^{(j)}]$ – кодова відстань вектора $x^{(j)}$ від центра розсіювання реалізацій, що визначається вершиною усередненого вектора x_1 ; \oplus – символ операції складання за модулем два; d_m – поточний радіус контейнера класу X_m^o , що відновлюється в бінарному просторі ознак розпізнавання.

Аналіз вище наведеного алгоритму показує, що він дозволяє навчання системи без досягнення граничного максимум інформаційного критерію (3.6.1), що відповідає принципу відкладених рішень і надає можливість здійснювати оптимізацію інших параметрів навчання з метою побудови безпомилкового за навчальною матрицею класифікатора.

Розглянемо результати реалізації запропонованого алгоритму на прикладі паралельної оптимізації контрольних допусків на ознаки розпізнавання при навчанні СППР, що використовується в АСК технологічним процесом виробництва фосфорної кислоти у ВАТ «Сумихімпром». Вхідну апріорно класифіковану навчальну матрицю було сформовано за експериментальними даними, одержаними для трьох класів, що характеризували зміст P_2O_5 на виході п'ятої ступені екстрактора за такою трьохалтернативною системою оцінок: «Норма», «Менше норми» і «Більше норми». При цьому кількість ознак розпізнавання дорівнювала 65, із них 15 поступали за результатами хімічного аналізу, а інші значення різних за природою ознак розпізнавання поступали в СППР безпосередньо з датчиків

інформації. Як базовий було обрано клас x_1^o , що характеризував функціональний стан технологічного процесу «Більше норми». Клас x_2^o характеризував заданий режим технологічного процесу «Норма» і клас x_3^o – «Менше норми».

На рис. 3.32 показано графік залежності критерію Кульбака (12) від параметра поля контрольних допусків, одержаний в процесі паралельної оптимізації контрольних допусків на ознаки розпізнавання. На графіку темні ділянки позначають робочі області визначення функції інформаційного критерію, в яких перша і друга достовірності перевершують відповідно похибки першого та другого роду і одночасно виконується умова $d_m > d_{m-1}$. При цьому у загальному випадку при кількості класів більше двох можуть існувати декілька робочих областей.

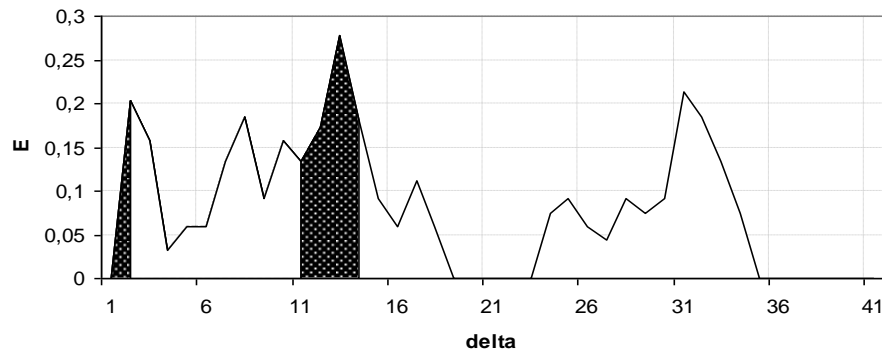


Рисунок 3.32– Графік залежності критерію Кульбака від параметра поля контрольних допусків

Аналіз рис.3.32 показує, що оптимальний параметр поля контрольних допусків дорівнює $\delta \pm 13$ (у відносних одиницях) при значенні максимуму усередненого критерію $\bar{E}^* = 0,28$.

При цьому побудова мультимодального класифікатора з паралельною оптимізацією контрольних допусків на ознаки розпізнавання для цих класів і такої самої вхідної навчальної матриці забезпечує вдвічі менше значення максимуму критерію \bar{E}^* .

Оскільки метою навчання СППР є відновлення в просторі ознак оптимальних контейнерів класів розпізнавання, то на рис.3.33 і рис.3.34 наведено графіки залежності КФЕ від радіусів контейнерів класів X_1^o і X_2^o .

Аналіз рис. 3.33 і рис. 3.34 показує, що оптимальні радіуси контейнерів класів X_1^o і X_2^o дорівнюють двом кодовим одиницям. Оскільки центр ваги реалізацій класу X_2^o зміщений від центру розсіювання реалізацій всіх класів на відстань $d(x_1 \oplus x_2) = 3$, де x_2 – усереднений вектор-реалізація для класу X_2^o , оптимальний радіус для класу X_1^o , дорівнює $d_1^* = 2$ і для класу X_2^o – $d_2^* = 5$.

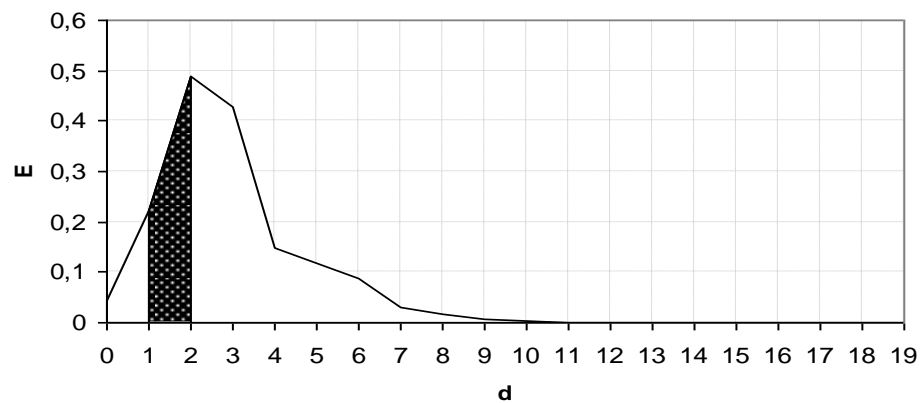


Рисунок 3.33–Графік залежності критерію Кульбака від радіусу контейнера класу X_1^o

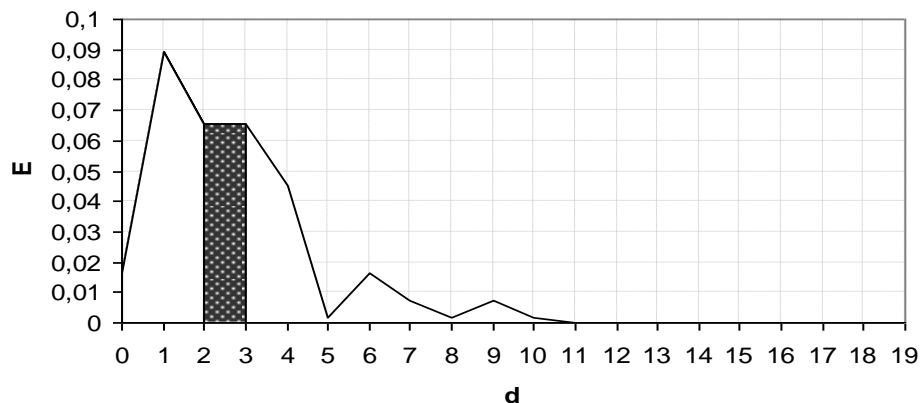


Рисунок 3.34– Графік залежності критерію Кульбака від радіусу контейнера класу X_2^o

Таким чином, аналіз значень КФЕ показує, що одержані в прикладі оптимальні параметри навчання не забезпечують побудову безпомилкового за навчальною матрицею класифікатора, тобто вони можуть розглядатися як квазіоптимальні. Для побудови безпомилкового за навчальною матрицею класифікатора згідно з принципом відкладених рішень необхідна оптимізація інших параметрів навчання.

ВИСНОВКИ

У роботі розв'язано важливу науково-технічну прикладну задачу аналізу і синтезу здатної навчатися в режимі ФКА СППР для автоматизації технологічного процесу за умови апріорної невизначеності через відсутність з технічних проблем неперервного вхідного контролю сировини та матеріалів природного походження. Головні наукові та практичні результати роботи полягають у такому:

1. За результатами аналітичного огляду джерел інформації проаналізовано сучасний стан та тенденцію розвитку АСК на базі інтелектуальних СППР і показано, що використання сучасних засобів автоматизації, наприклад, у вигляді цифрових ПІД-регуляторів не дозволяє ефективно розв'язувати задачі автоматизації технологічних процесів за умов відсутності через технічні проблеми неперервного вхідного контролю сировини та матеріалів природного забезпечення і зроблено висновок про необхідність надання АСК властивості адаптивності шляхом самонавчання і автоматичної класифікації.

2. Запропоновано новий інформаційно-екстремальний метод аналізу і синтезу інтелектуальної СППР, що самонавчається в режимі ФКА, який дозволяє побудувати безпомилкові за навчальною та контрольною багатовимірними матрицями вирішальні правила шляхом оптимізації просторово-часових параметрів функціонування, в тому числі і словника ознак

3. Запропоновано комплекс логічно пов'язаних у рамках ІЕІТ категорійних моделей у вигляді діаграм відображень множин, що застосовуються при функціонуванні СППР в режимі ФКА з оптимізацією словника ознак, які дозволяють спрощувати синтез алгоритмів функціонування інтелектуальних АСКТП.

4. Розроблено алгоритм оцінки функціональної ефективності навчання і досліджено вплив параметрів словника ознак на функціональну ефективність СППР, що самонавчається в режимі ФКА.

5. Розроблено алгоритми оптимізації геометричних параметрів контейнерів класів розпізнавання, системи контрольних допусків на ознаки розпізнавання, та параметрів оптимізації словника ознак, що дозволило побудувати безпомилкові за навчальною вибіркою вирішальні правила, підвищити достовірність та оперативність класифікації функціональних станів технологічного процесу.

6. Запропоновано алгоритм оптимізації словника з використанням адаптивного випадкового пошуку, для якого після оптимізації відповідних параметрів функціонування проводилася модифікація ймовірностей включення кожної ознаки до наступного словника в залежності від того, чи була присутня ця ознака в поточному словнику, а також чи дозволяє словник, що аналізується, підвищити ефективність у порівнянні з попередніми. Результати фізичного моделювання адаптивного алгоритму навчання показують, що кількість варіантів словників, для яких середнє значення КФЕ було більше за відповідне значення КФЕ для повного словника, складає приблизно 46,7% від загальної кількості згенерованих у порівнянні з 1,7% при реалізації алгоритму традиційного випадкового пошуку.

7. У рамках ІЕІ-технології оптимізація словника ознак за схемою алгоритму ПСС дозволяє побудувати безпомилкові за багатовимірною навчальною матрицею вирішальні правила, оскільки максимальне значення ентропійного КФЕ досягає в робочій області визначення його функції досягає свого граничного значення $\bar{E} = E_{гран} = 1,0$.

8. Запропонований гібридний алгоритм навчання у рамках ІЕІ-технології з використанням генетичного алгоритму дозволив підвищити оперативність навчання СППР шляхом заміни прямого перебору значень параметрів навчання на процедуру паралельної оптимізації параметрів функціонування.

9. Розроблено алгоритми навчання та екзамену для унімодального класифікатора з метою підвищення достовірності рішень, що приймаються СППР, при впорядкованому алфавіті класів.

ПЕРЕЛІК ПОСИЛАНЬ

1. Довбиш А.С. Основи проектування інтелектуальних систем: Навчальний посібник.– Суми: Видавництво Сум ДУ, 2009.– 171 с.
2. Шелехов І.В., Дзюба О.О. Дослідження впливу параметрів гібридного алгоритму на функціональну ефективність навчання системи прийняття рішень // Радіоелектронні і комп'ютерні системи.–2009.–№3.–С. 88-91.
3. Довбиш А.С., Чекалов О.П., Мартиненко С.С. Інтелектуальна система діагностування онкопатологій // Радіоелектронні і комп'ютерні системи.– 2009.–№3.–С.92-96.
4. Довбиш А.С., Мартиненко С.С. Інформаційно-екстремальний метод розпізнавання електронограм // Вісник СумДУ. Серія: Технічні науки.–2009.– №2.– С. 85-91.
5. Довбиш А.С., Тронь В.А. Прогностичне навчання автоматизованої системи керування технологічним процесом виробництва фосфорної кислоти // Вісник СумДУ. Серія: Технічні науки.–2009.–№2.– С. 85-91.
6. Довбиш А.С., Жаловага В.О. Інтелектуальна система підтримки прийняття рішень для діагностування онкопатологій // Вісник СумДУ. Серія: Технічні науки.–2009.–№3.– С.
7. Довбиш А.С., Барило Р.Б., Барило О.Б. Визначення та корекція індивідуальних траєкторій навчання із застосуванням індивідуально орієнтованих електронних освітніх ресурсів // Вісник СумДУ. Серія: Технічні науки.–2009.–№3.– С.
8. Довбиш А.С., Якушев О.А. Оптимізація кроку квантування акустичних сигналів при розпізнаванні голосових команд // Вісник СумДУ. Серія: Технічні науки.–2009.–№3.– С.
9. Довбиш А.С. Стан та тенденція розвитку інтелектуальних технологій в інформаційному суспільстві // Тези доповідей Другої

міжнародної науково-технічної конференції «Інтелектуальні системи в промисловості та освіті», 3-5 грудня 2009 р.– Суми, 2009.– С. 52-56.

10. Дзюба О.О. Оцінка оперативності гібридного алгоритму навчання інтелектуальної системи підтримки прийняття рішень // Тези доповідей Другої міжнародної науково-технічної конференції «Інтелектуальні системи в промисловості та освіті», 3-5 грудня 2009 р.– Суми, 2009.– С. 23-24.

11. Довбиш А.С., Козинець М.В., Котенко С.М. Оптимізація контрольних допусків на ознаки розпізнавання в інформаційно-екстремальних методах автоматичної класифікації // Вісник Сумського державного університету. Серія «Техніка», №1, 2007.– С. 169-178.

12. Краснополюсовський А.С. Інформаційний синтез інтелектуальних систем керування: Підхід, що ґрунтується на методі функціонально-статистичних випробувань. – Суми: Видавництво Сумського державного університету, 2004. – 261с.

13. Востоцький В.О. Алгоритм автоматичної класифікації // Тези доповідей Другої міжнародної науково-технічної конференції «Інтелектуальні системи в промисловості та освіті», 3-5 грудня 2009 р.– Суми, 2009.– С. 37-38.

14. Шелехов І.В., Барило О.Б. Інформаційно-екстремальна корекція генотипних параметрів навчання // Тези доповідей Другої міжнародної науково-технічної конференції «Інтелектуальні системи в промисловості та освіті», 3-5 грудня 2009 р.– Суми, 2009.– С. 18-19.

15. Довбиш А.С., Котенко С.М. Побудова адаптивної системи керування слабоформалізованим процесом за методом факторного кластер-аналізу // Вісник СумДУ. Серія: Технічні науки.–2010.–№1.– С. 7-15.

16. Довбиш А.С., Алтиннікова К.В. Ієрархічний алгоритм розпізнавання електронограм // Автоматизированные системы управления и приборы автоматики.–2009.– Вып. 148.– С.20-25.

17. Довбиш А.С., Дзюба О.О. Аналіз алгоритмів оптимізації контрольних допусків на ознаки розпізнавання // Адаптивні системи автоматичного управління.–2010.–№16(36).–С.11-15.

18. Довбиш А.С., Руденко М.С. Оптимізація параметрів навчання СППР для діагностування онкопатологій // Вісник СумДУ. Серія: Технічні науки.– 2010.–№2.– С.-

19. Саад Джулгам, Востоцький В.О. Оптимізація параметрів навчання системи підтримки прийняття рішень // Вісник Сумського державного університету.- Серія: Технічні науки.– 2010.-№3. –С.

20. Проектирование систем управления на ЭВМ / [А.Ю. Соколов, Ю.Н. Соколов, В.М. Илюшко и др.]; под ред. Ю.Н. Соколова.–Х.: «ХАИ», 2005.– 590 с.

21. Zadeh L. A. Fuzzi sets // Information and Control. – 1965. –№8. – Р. р. 338–353.

22. Hybrid Methods in Pattern Recognition / Edited by H.Runke & A.Kandel. – Verlag Shpringer. – 2002. – 336 p.

23. Турбович И. Т., Гитис В. Г., Маслов В. К. Оpozнание образов. Детерминир.-статист. Поход /. –М.: Наука, 1971.– 246 с.

24. Фельдбаум А.А. Основы теории оптимальных автоматических систем. Изд. 2-е, испр. и доп.– М: Наука, 1966.– 623 с.

25. Васильев В.И. Теория редукции в проблемах экстраполяции // Проблемы управления и информатики. – 1996.– № 1,2.– С 239-251.

26. Devijver P.A., Kittler J. Pattern recognition: a statistical approach.- Prentice Hall, 1982.

27. Jerome H. Friedman, Werner Stutzle. Projection pursuit regression // Journal of the American Statistical Association.– December 1981.–№ 76(376).– Р. 817-823.

28. Hastie T. J., Tibshirani R. J.. Generalized Additive Models // Monographs on Statistics and Applied Probability.- Chapman & Hall.– 1990.– vol.43.

29. Jerome H. Friedman. Multivariate adaptive regression splines // *The Annals of Statistics*.– 1991.– №19 (1).– P. 1-67.
30. Dash M. and Liu H. Feature selection for classification // *Intelligent Data Analysis - An International Journal*.–1997.–№ 1(3).
31. Siedlecki W. and Sklansky J. On automatic feature selection // *International Journal of Pattern Recognition and Artificial Intelligence*.– 1988.– v. 2.– P. 197-220.
32. Doak J. An evaluation of feature selection methods and their application to computer security- Technical report. Davis CA University of California. Department of Computer Science, 1992.
33. Narendra P. M. and Fukunaga K. A branch and bound algorithm for feature subset selection // *IEEE Transactions on Computers*.– 1977.– №26.– P. 917-922.
34. Schlimmer J. C. Efficiently inducing determinations A complete and systematic search algorithm that uses optimal pruning // *In Proceedings of Tenth International Conference on Machine Learning*.– 1993.– P. 284-290.
35. Juzhen Z. Dong, Skowron Andrzej, and Ohsuga Setsuo. Using rough sets with heuristics to feature selection // *Proceedings of the 7th International Workshop on New Directions in Rough Sets, Data Mining, and Granular-Soft Computing (RSFDGrC-99)*.– Berlin, November 9-11 1999.– Springer.–1999.– Volume 1711 of *Lecture Notes in Artificial Intelligence*.– P. 178-187.
36. Duran B. S. and Odell P. K.. *Cluster analysis - a survey* - Springer,Verlay, 1974.
37. Xu. L., Yan P. and Chang T. Best first strategy for feature selection // *In Proceedings of Ninth International Conference on Pattern Recognition*.– 1988.– P. 706-708.
38. Yu B. and Yuan B. A more efficient branch and bound algorithm for feature selection // *Pattern Recognition*.–1993.–V. 26(6) .– P. 883-889.

39. Evaluation of the branch and bound algorithm for feature selection/ [Hamamoto Y., Uchimura S., Matsuura Y., Kanaoka T. and Tomita S.] // -Pattern Recognition Letters. – July 1990.– V. 11.–P. 453-456.

40. Bobrowski L. Feature selection based on some homogeneity coefficient // In Proceedings of Ninth International Conference on Pattern Recognition.– 1988.– P. 544-546.

41. Sheinvald J., Dom B. and Niblack W. A modelling approach to feature selection // In Proceedings of Tenth International Conference on Pattern Recognition.–1990.– V. 1.–P. 535-539.

42. Almuallim Hussein and Dietterich Thomas G. Learning Boolean concepts in the presence of many irrelevant features // Artificial Intelligence.– 1994.– V. 69(1-2).–P. 279-305.

43. Oliveira A. L. and Vincentelli A. S. Constructive induction using a nongreedy strategy for feature selection // In Proceedings of Ninth International Conference on Machine Learning. Aberdeen, Scotland: Morgan Kaufmann.– 1992.–P. 355-360.

44. Ichino M. and Sklansky J. Feature selection for linear classifier // In Proceedings of the Seventh International Conference on Pattern Recognition.– 1984.– V.1.–P. 124-127.

45. Ichino M., Sklansky J. Optimum feature selection by zero-one programming // IEEE Trans. on Systems, Man and Cybernetics, SMC.—1984.– V. 14(5).–P. 737-746.

46. Geoffrion A. M. Integer programming by implicit enumeration and balas, method // SIAM Review.– April 1967.– V. 9.–P. 178-190.

47. Kenji Kira and Rendell Larry A. The feature selection problem: Traditional methods and a new algorithm // In William Swartout, editor, Proceedings of the 10th National Conference on Artificial Intelligence.– San Jose, CA: MIT Press.– July 1992.–P. 129-134.

48. Kononenko I. Estimating attributes: Analysis and extensions of RELIEF. In Francesco Bergadano and Luc de Raedt, editors, Proceedings of the European

Conference on Machine Learning.– Berlin, April 1994.– Springer.–1994.–V. 784 of LNAI.– P. 171-182.

49.Robnik-Sikonja Marko and Kononenko Igor. Attribute dependencies, understandability and split selection in tree based models // In Proc. 16th International Conf. on Machine Learning .– San Francisco, CA: Morgan Kaufmann, 1999.– P. 344-353.

50.Segen J. Feature selection and constructive inference // In Proceedings of Seventh International Conference on Pattern Recognition.– 1984.– pages 1344 – 1345.

51. Cardie C. Using decision trees to improve case-based learning // In Proceedings of Tenth International Conference on Machine Learning.– 1993.– P. 25-32.

52. Koller Daphne and Sahami Mehran. Toward optimal feature selection // In Proceedings of ICML-96, the Thirteenth International Conference on Machine Learning.– Bari, Italy.– 1996.– P. 284-292.

53.Mucciardi A. N. and Gose E. E. A comparison of seven techniques for choosing subsets of pattern recognition // IEEE Transactions on Computers.– 1971.– C-20.– P. 1023-1031.

54. Shen Qiang, Chouchoulas Alexios. Rough set-based dimensionality reduction for supervised and unsupervised learning // Applied Mathematics and Computer Science, Special Issue on Rough Sets and their Applications, 2001.– №11(3).–P. 101-119.

55. Liu H and Setiono R. A probabilistic approach to feature selection - a filter solution // In Proceedings of International Conference on Machine Learning. –pages 319-327, 1996.– P. 319-327.

56. Kittler J. : In C. H. Chen, editor. Feature set search algorithms // Pattern Recognition and Signal Processing, Sijthoff and Noordhoff Alphen aan den Rijn, Netherlands, 1978.–P. 41-60.

57.Pudil P., Novovicova J. and Kittler J. Floating search methods in feature selection // Pattern Recognition Letters.– November 1994.– V.15.–P. 1119-1125.

58. Caruana Rich, Dayne Freitag, Morgan Kaufmann. Greedy attribute selection // In Proceedings of the 11th International Conference on Machine Learning, 1994.–P. 28-36.

59. Moore A. W. and Lee M. S. Efficient algorithms for minimizing cross validation error // In Proceedings of Eleventh International Conference on Machine Learning. pages 190-198. New Brunswick. New Jersey: Morgan Kaufmann.–1994.– P. 190-198.

60. Domingos P. Context_sensitive feature selection for lazy learners – To appear in Artificial Intelligence Review.– 1996.

61. Queiros C. E. and Gelsema E. S. On feature selection // In Proceedings of Seventh International Conference on Pattern Recognition.– July-Aug 1984.–1984.– V. 1.–P. 128-130.

62. Skalak D. B. Prototype and feature selection by sampling and random mutation hill_climbing algorithms // In Proceedings of Eleventh International Conference on Machine Learning.– New Brunswick: Morgan Kaufmann.–1994.– P. 293-301.

63. Siedlecki W. and Sklansky J. A note on genetic algorithms for large-scale feature selection // Pattern Recognition Letters.– November 1988.– V. 10.–P. 335-347.

64. Ferri F., Pudil P., Hatef M. and Kittler J.: In E. Gelsema and L. Kanal, editors Comparative study of techniques for large-scale feature selection // Pattern Recognition in Practice IV, Elsevier Science B.V.– 1994.– P. 403-413.

65. Mao J., Mohiuddin K. and Jain A. K. Parsimonious network design and feature selection through node pruning // In Proceedings of 12th ICPR.– Jerusalem.–1994.– P. 622-624.

66. Rumelhart D. E., Hinton G. E. and Williams R. J. In Rumelhart D. E. and McClelland J. L. Editors. Learning internal representations by error propagation // Parallel Distributed Processing, Explorations in the Microstructure of Cognition.– MIT Press.– 1986.– V. 1.– Chapter 8.–P. 318-362.

67. Краснопоясовський А.С., Козинець М. В. Факторний класифікаційний аналіз за методом функціонально-статистичних випробувань // Радіоелектронні та комп'ютерні системи.– 2004. – №4. – С. 46-50.

68. Козинець М.В. Система інтелектуальної підтримки прийняття рішень для автоматизації розподілених технологічних процесів: автореф. дис. на здобуття наук. ступеня канд. техн. наук: спец. 05.13.07 „Автоматизація процесів керування”.– К, 2009.– 20 с.

69. Колмогоров А.А. Три подхода к определению понятия «количество информации» // Проблемы передачи информации.– 1965.–Т.1. –Вып.1. –С. 3–11.

70. Большев Л. Н., Смирнов Н. В. Таблицы математической статистики / Л. Н. Большев, Н. В. Смирнов.– М.: Наука.– 1983. – 416 с.

71. Ивахненко А. Г. О принципах построения обучающихся систем управления сложными процессами – М.: Наука, 1970.–252 с.

72. Буч Г. Объектно-ориентированное проектирование с примерами применения – Киев – Москва: Діалектика, И.В.К.– 1992.– 519 с.

73. Растрингин Л.Д. Адаптация сложных систем .– Рига: Зинатне, 1981.– 376 с.

74. Кузьмин И. В. Оценка эффективности и оптимизация автоматизированных систем контроля и управления. – М.: Сов. радио, 1971. – 296 с.