

Міністерство освіти і науки України
Харківський національний університет радіоелектроніки

Факультет Комп'ютерних наук
(повна назва)

Кафедра Штучного інтелекту
(повна назва)

КВАЛІФІКАЦІЙНА РОБОТА
Пояснювальна записка

рівень вищої освіти другий (магістерський)

Нелінійний бегінг у системах обчислювального інтелекту
(тема)

Виконав:
студент 2 курсу, групи СШМ-22-1
Іванов Є.О.
(прізвище, ініціали)

Спеціальність 122 Комп'ютерні науки
(код і повна назва спеціальності)

Тип програми освітньо-наукова
(освітньо-професійна або освітньо-наукова)

Освітня програма Системи штучного інтелекту
(повна назва спеціалізації)

Керівник проф. Бодянський Є.В.
(посада, прізвище, ініціали)

Допускається до захисту

Зав. кафедри _____
(підпис)

В.О. Філатов
(прізвище, ініціали)

2024 р.

Харківський національний університет радіоелектроніки

Факультет _____ Комп'ютерних наук _____
(повна назва)
Кафедра _____ Штучного інтелекту _____
(повна назва)
Рівень вищої освіти _____ другий (магістерський) _____
Спеціальність _____ 122 Комп'ютерні науки _____
(код і повна назва)
Тип програми _____ освітньо-наукова _____
(освітньо-професійна або освітньо-наукова)
Освітня програма _____ Системи штучного інтелекту _____
(повна назва)

ЗАТВЕРДЖУЮ:
Зав. кафедри _____
(підпис)
« _____ » _____ 20 ____ р.

ЗАВДАННЯ
НА КВАЛІФІКАЦІЙНУ РОБОТУ

студентові _____ Іванову Євгену Олексійовичу _____
(прізвище, ім'я, по батькові)

1. Тема роботи _____ Нелінійний бегінг у системах обчислювального інтелекту _____

затверджена наказом університету від 1 квітня 2024 р. № 260Ст

2. Термін подання студентом роботи до екзаменаційної комісії 14 червня 2024 р.

3. Вихідні дані до роботи науково-технічні публікації, дані Інтернет-джерел щодо тематики дослідження, документація Python, набори даних для тренування, навчання та валідації

4. Перелік питань, що потрібно опрацювати в роботі _____

1) Аналіз предметної галузі

2) Методики нечіткого нелінійного беггінгу

3) Практичне застосування нечіткого нелінійного беггінгу

РЕФЕРАТ

Пояснювальна записка: 97 с., 5 рис., 4 табл., 2 дод., 70 джерел.

АДАПТИВНІ МОДЕЛІ, АЛГОРИТМИ НАВЧАННЯ, МАШИННЕ НАВЧАННЯ, НЕЛІНІЙНИЙ БЕГГІНГ, НЕЧІТКИЙ БЕГГІНГ, ПРОГНОЗУВАННЯ ДАНИХ, ШТУЧНИЙ ІНТЕЛЕКТ.

Об'єкт дослідження – передові системи обробки даних, що використовують нечіткий нелінійний беггінг.

Предмет дослідження – методи та технології, що використовуються у нечіткому нелінійному беггінгу для обробки та прогнозування великих об'ємів даних.

Мета роботи – провести глибоке дослідження методів та технологій, що використовуються в нечіткому нелінійному беггінгу, та розробити ефективні стратегії, які можуть використовуватися для вдосконалення цих систем.

Методи дослідження – аналіз літератури, проведення експериментів, статистичний аналіз даних, комп'ютерне моделювання та інше.

Кваліфікаційна робота присвячена дослідженню методів нечіткого нелінійного беггінгу в системах обчислювального інтелекту. У роботі аналізуються сучасні підходи до прогнозування та обробки даних з використанням адаптивних моделей, алгоритмів машинного навчання та методів беггінгу. Основною метою є розробка та впровадження ефективної методики нечіткого нелінійного беггінгу для підвищення точності та стабільності прогнозів у складних та неоднозначних умовах. Експериментальні дослідження проводились на наборах даних для виявлення шахрайства з кредитними картками, прогнозування цін на автомобілі та класифікації наукових журналів.

ABSTRACT

Master's thesis contains: 97 pp., 5 fig., 4 tabl., 1 ann., 70 references.

ADAPTIVE MODELS, ARTIFICIAL INTELLIGENCE, DATA FORECASTING, FUZZY BAGGING, LEARNING ALGORITHMS, MACHINE LEARNING, NONLINEAR BAGGING.

Research Object – advanced data processing systems using fuzzy nonlinear bagging.

Research Subject – methods and technologies used in fuzzy nonlinear bagging for processing and predicting large volumes of data.

Purpose of the Work – to conduct a deep investigation into the methods and technologies used in fuzzy nonlinear bagging, and to develop effective strategies that can be used to extend these systems.

Research Methods – literature review, conducting experiments, statistical data analysis, computer simulation, and others.

The qualification work is dedicated to the study of fuzzy nonlinear bagging methods in computational intelligence systems. The work analyzes modern approaches to forecasting and data processing using adaptive models, machine learning algorithms, and bagging methods. The main goal is to develop and implement an effective methodology of fuzzy nonlinear bagging to improve the accuracy and stability of predictions in complex and ambiguous conditions. Experimental studies were conducted on datasets for credit card fraud detection, car price prediction, and scientific journal classification.

ЗМІСТ

Вступ.....	7
1 Аналіз предметної галузі.....	10
1.1 Опис предметної галузі.....	10
1.2 Огляд структури задач прогнозувань.....	14
1.3 Постановка задачі дослідження.....	22
1.3.1 Класифікація обраної задачі та її особливості.....	33
1.3.2 Визначення цілей і завдань роботи.....	35
2 Методики нечіткого нелінійного беггінгу.....	37
2.1 Аналіз принципів нелінійного беггінгу в задачах прогнозувань.....	37
2.2 Методологія та математичне моделювання нечіткого нелінійного беггінгу.....	46
2.3 Розробка власної методики імплементації нечіткого нелінійного беггінгу.....	55
2.4 Інноваційні підходи до нелінійного беггінгу для покращення ефективності прогнозування.....	63
3 Практичне застосування нечіткого нелінійного беггінгу.....	72
3.1 Опис програмних засобів.....	72
3.2 Опис всіх наборів даних, що застосовуються в роботі, їх дослідницький аналіз і попередня обробка.....	75
3.3 Аналіз результатів.....	79
3.3.1 Результати класифікації на наборі credit card fraud detection....	80
3.3.2 Результати класифікації на наборі Car Price Prediction Dataset.	82
3.3.3 Результати класифікації на наборі ScimagoJR Journals – 2023 .	83
3.4 Порівняльний аналіз результатів.....	84
Висновки.....	86
Перелік джерел посилання.....	88
Додаток А Код програми.....	95
Додаток Б Відомість кваліфікаційної роботи.....	97

ВСТУП

Сучасні комп'ютерні технології надають широкі можливості у сфері аналізу даних та прогнозування. Це особливо актуально для розробки прогностичних моделей у різних галузях науки та техніки, фінансів, медицини, екології та інших галузях, де необхідно точно й швидко визначати майбутні події.

Прогнозування – це невід'ємна частина багатьох галузей людської діяльності. Однак, часто це завдання може бути нелінійним і нечітким, що призводить до значної складності в його вирішенні. Виникнення нечіткості у галузі прогнозування може бути викликано різноманітними причинами – від неточності та недостатності даних до непередбачуваності зовнішніх факторів. Однак, сучасні алгоритми машинного навчання дозволяють враховувати такі фактори, забезпечуючи високу точність прогнозування.

Одним з таких інноваційних підходів є метод нечіткого нелінійного беггінгу. Беггінг – це метод машинного навчання, що дозволяє покращити стабільність та точність алгоритмів за допомогою зведення вибірки до одиниці шляхом агрегації прогнозів незалежних моделей, отриманих на основі вибірки bootstrap. Особливістю даного методу є врахування нечіткості (неточності) вихідних даних.

Метою даної роботи є дослідження принципів нечіткого нелінійного беггінгу та їх вдосконалення для покращення прогностувальних можливостей у різних задачах прогнозування. Для цього були виокремлені наступні завдання:

- вивчити основні поняття і особливості предметної галузі;
- проаналізувати структуру задач прогнозування та методи їх рішення, зосередившись на задачах розпізнавання образів;
- вивчити основи нечіткого нелінійного беггінгу та його особливості;
- проаналізувати поточний стан прогнозування та його проблеми;

– розкрити теоретичний підхід до нечіткого нелінійного беггінгу, включаючи історію його появи, методологію та математичні моделі, на яких він ґрунтується;

– запропонувати власну методику імплементації нечіткого нелінійного беггінгу, виходячи зі специфіки вирішуваних завдань і вимог до прогнозування;

– визначити способи збору, аналізу та обробки даних для подальшого прогнозування за допомогою розробленої методики;

– розробити стратегію тестування отриманої моделі прогнозування та скласти план експерименту.

Структура даного дослідження ґрунтується на послідовному виконанні визначених завдань. Початковий етап полягає у загальному ознайомленні з предметом дослідження – прогнозуванням та його проблемами, а також з основами нечіткого нелінійного беггінгу. Другий етап включає в себе розробку власної методики нелінійного беггінгу та визначення стратегій її апробації.

На цьому етапі подається докладний опис процесу збору та обробки даних, розробки алгоритму прогнозування на основі нечіткого нелінійного беггінгу, його тестування й практичного застосування.

Після цього настане етап введення в дію розробленої моделі, в ході якого буде проведено низку експериментів для забезпечення імітації реальних умов прогнозування.

Останній етап передбачає аналіз результатів дослідження, визначення ступеня ефективності пропонованого методу нечіткого нелінійного беггінгу та розробку рекомендацій щодо його подальшого використання і вдосконалення.

Підсумовуючи, якщо говорити про ціль даного дослідження, то вона полягає у розробці моделі прогнозування на основі методу нечіткого нелінійного беггінгу, що б враховувала специфіку нелінійного та нечіткого

прогнозування. Наступна ціль – вивчення та апробація цієї моделі в реальних умовах та на практичних відповідних даних.

Результатом проведення даного дослідження має стати сформована вдосконалена методика нечіткого нелінійного беггінгу, яка дозволить підвищити точність прогнозування в середовищі невпевненості і нечіткості вихідних даних. Очікується, що її можна буде використовувати в різних галузях науки та техніки, незалежно від специфіки конкретної задачі прогнозування.

Результатом аналізу результатів прогнозування за допомогою обраної методики повинна стати рекомендація щодо її використання в конкретних сферах діяльності та пропозиції щодо подальшого її вдосконалення.

Відтак, результати даного дослідження можуть знайти застосування не тільки у наукових дослідженнях, але й у практичних сферах діяльності, особливо там, де потрібно працювати з великою кількістю даних та здійснювати прогнозування з високим ступенем точності.

Однак, варто враховувати, що метод нечіткого нелінійного беггінгу є частинтею загальної теорії прогнозування, і його вивченням ми продовжуємо більш загальний шлях дослідження прогнозування як такого.

Передбачається, що це дослідження сприятиме подальшому вивченню прогнозування, і надасть можливості для розробки нових методів і підходів, які будуть ще більш ефективні та точні.

У цьому випадку, результатами нашої роботи стануть не тільки практичні методики, але і теоретичний внесок в галузь прогнозування. Якщо наше дослідження принесе позитивні результати, це підтвердить можливість використання нечіткого нелінійного беггінгу як повноцінного інструменту для прогнозування у нелінійних і нечітких умовах.

Такий успіх, безумовно, стане важливим кроком у розвитку теорії прогнозування і позитивно вплине на вирішення великої кількості практичних завдань.

1 АНАЛІЗ ПРЕДМЕТНОЇ ГАЛУЗІ

1.1 Опис предметної галузі

Це дослідження фокусується на розробці та аналізі нечіткого нелінійного беггінгу, специфічного підходу до прогнозування, що застосовує нечітку логіку і нелінійні алгоритми для вирішення задач прогнозування.

Беггінг, або bootstrap aggregating, є методом ансамблю у машинному навчанні, що використовує стратегію навчання кількох моделей на різних підмножинах набору даних, а потім агрегує їх прогнози. Цей метод був вперше запропонований Лео Брейманом у 1994 році.

Але в той час, як беггінг зазвичай використовується з лінійними моделями, у цьому дослідженні ми розглядаємо використання нелінійного беггінгу у комбінації з нечіткими системами для покращення результатів прогнозування.

Особливість нечітких систем полягає в їх здатності ефективно обробляти нечітку інформацію та неоднозначність прогнозування. Це надає значних переваг при моделюванні складних систем або систем, в яких існують певні ступені невизначеності. Такий підхід до обробки даних відіграє важливу роль в роботі з такими завданнями, як прогнозування в економіці, соціальних науках, біології та інших галузях.

На відміну від інших методів машинного навчання, нечіткі системи не вимагають жорсткого формалізму в описуванні взаємозв'язків між різними змінними системи. Вони надають більше вольності в інтерпретації даних, що дозволяє розрізняти і аналізувати різноманітні сценарії та відтворювати складні моделі поведінки.

Нелінійний беггінг використовує гнучкість нечітких систем, комбінуючи її з нелінійними алгоритмами прогнозування. Це створює високоефективну стратегію, яка має потенціал значно покращити точність прогнозів для складних і неоднозначних систем.

Адаптивна метамоделі є ще одним ключовим концептом, який використовується в межах цього дослідження. Метамоделі використовуються для створення спрощених представлень про поведінку деякої системи або процесу на основі аналізу великої кількості даних. Вони забезпечують узагальнені представлення, які описують взаємозв'язки між різними змінними і допомагають нам розуміти основні тенденції та динаміку системи. Адаптивні метамоделі є поширеними і гнучкими, що дозволяє їм адаптуватися до змін у даних або в умовах системи.

Комбінація нечіткого нелінійного беггінгу з адаптивними метамоделями надає міцну основу для створення ефективних систем прогнозування. Ця синергія дозволяє уникнути обмежень традиційних методів прогнозування і навчання, розвиваючи рішення, які можуть краще впоратися зі складністю і несистематичністю реального світу. Адаптивні метамоделі забезпечують необхідну гнучкість, щоб вирішувати проблеми з будь-яким рівнем складності, в той час як нечіткий нелінійний беггінг додає потужне засоби прогнозування і несистематичну спроможність обробки.

Також важливим аспектом цього дослідження є поєднання теоретичних підходів з практичними застосуваннями. Основна мета – розробити методіку, яка ефективно використовує нечіткий нелінійний беггінг та адаптивні метамоделі для прогнозування в різних контекстах. Використовуючи великі набори даних, ми сподіваємося виявити нові взаємозв'язки та відкриття, які будуть корисними для покращення не тільки методології, але й практичних застосувань.

Сфера прогнозування включає широкий спектр застосувань, починаючи від фінансового та бізнес-аналізу і закінчуючи науковими дослідженнями та оцінкою ризиків. Використання нечіткого нелінійного беггінгу допоможе моделювати та розуміти складні системи, використовуючи нечіткість та невизначеність, які є характеристиками реального світу. Завдяки цьому можливо отримати більш точні та надійні

прогнози, що є важливим вкладом у галузі аналітики даних та прогнозування.

Цікавим напрямком для такого роду дослідження є його застосування у різноманітних сферах – від наукових досліджень до промисловості. Нечіткий нелійний беггінг є особливо корисним у ситуаціях, коли дані є складними, шумними або неповними. Завдяки своїй високій стійкості до таких викликів, цей підхід ефективно впорається із задачами прогнозування у сфері економіки, екології, медицини, геології та інших галузях.

Також, нечіткий нелійний беггінг дозволяє розробляти більш розуміючі моделі, які можуть допомогти людям краще розуміти взаємозв'язки між різними елементами складних систем. Такі моделі можуть бути використані для побудови симуляцій, що допомагають вивчати різні сценарії прогнозування у безпечному та контрольованому середовищі.

Все це робить нашу тему предметної галузі перспективною і актуальною, з огляду на постійний розвиток технологій обробки даних і зростаючу потребу в ефективних інструментах прогнозування.

Очевидно, що практичне значення нечіткого нелійного беггінгу є значним. Його можливе застосування може включати різні аспекти вирішення проблем у великій кількості галузей, зокрема у виробництві, фінансах, освіті, медицині та інших галузях. При правильному використанні цей підхід може допомогти у прийнятті обґрунтованих рішень та виявленні трендів та шаблонів, що можуть бути прихованими в складних даних.

Вони можуть ще більше його покращити шляхом інтеграції із іншими технологіями, такими як штучний інтелект та біг дата. Така інтеграція дає можливість створювати більш вдосконалені моделі прогнозування, які легко масштабуються та можуть ефективно впоратися з аналізом великого обсягу даних.

Враховуючи експоненційний розвиток технологій обробки даних та їх вплив на модернізацію перелічених вище галузей, нечіткий нелійний

бегінг, безперечно, стане важливим інструментом прогнозування в ближчому майбутньому.

Важливо також зазначити, що процеси прогнозування та машинного навчання мають важливе значення не тільки для наукової спільноти, але й для широкого загалу. Вони мають потенціал реалізувати значні покращення в житті людей, зокрема в охороні здоров'я, освіті, урядуванні та побуті. Використання нечіткого нелінійного бегінгу для прогнозування у цих контекстах є важливим дослідницьким пріоритетом.

Наприклад, у медицині ці методи можуть бути використані для прогнозування пацієнтів із високим ризиком розвитку певних хвороб, що дозволяє забезпечити їм своєчасну і ефективну допомогу. У сфері освіти їх можна використовувати для прогнозування академічного успіху студентів, допомагаючи викладачам та ректорам покращувати рівень навчання.

Наука про дані та прогнозування вже відіграють важливу роль у прийнятті рішень широким спектром організацій – від невеликих підприємств до міжнародних корпорацій та урядових установ. Дослідження, як це, що пропонує нові підходи до прогнозування, допомагає подальшому розвитку цієї області та забезпеченню її релевантності для суспільства.

У ході цього дослідження розробляється інструмент з адаптивною метамоделлю, яка використовує нечіткий нелінійний бегінг для прогнозування. Це включає в себе виробництво, вивчення і оптимізацію алгоритмів для машинного навчання, а також основне розуміння ролі та важливості прогнозування в сучасній дійсності.

Це значиме дослідження доцільне і актуальне особливо через розвиток технологій великих даних, де потреба в точних прогнозах та стратегічних інсайтах невідмінно зростає. Застосування цього методу прогнозування відкриває нові можливості для значним поліпшенням виявлення шаблонів, підтримуючи прийняття обґрунтованих рішень у різних галузях.

По тому, як багато перспективних застосувань нечіткого нелінійного бегінгу на основі адаптивної метамоделі в задачах прогнозувань, це дослідження може стати основою для подальших наукових розробок і прогресу в даній області.

1.2 Огляд структури задач прогнозувань

Прогнозування є важливою частиною різних наукових, індустріальних та комерційних діяльностей. У цій частині ми зосередимось на структурі задач прогнозувань, зокрема тих, що використовують нечіткі нелінійні стратегії бегінгу на основі адаптивних метамоделей.

Прогнозування, у своєму контексті, передбачає використання історичних даних для висловлення припущень про майбутні події. Цей процес може стосуватися найрізноманітніших дисциплін, починаючи від економіки та маркетингу, закінчуючи погодними прогнозами та другими природничими науками.

Структура задач прогнозування може варіюватися в залежності від декількох факторів. Наприклад, у сфері фінансів структура задачі прогнозування може стосуватися прогнозування цін на акції на основі історичних даних, тоді як в маркетингу задачі прогнозування можуть включати прогнозування поведінки споживачів або впливу рекламних кампаній на продажі.

Запропонований в нашій роботі підхід – нечіткий нелінійний бегінг на основі адаптивної метамоделі – може застосовуватися у широкому спектрі задач прогнозування. Незважаючи на те, що ці задачі можуть бути різними в контексті конкретної області, основна структура прогнозування залишається сталою і включає три основні етапи: визначення завдання прогнозування, вибір і тренування моделі, і використання моделі для прогнозування.

Перший етап, визначення завдання прогнозування, включає в себе чітке формулювання проблеми, яку треба вирішити. Це може стосуватися передбачення часових рядів, класифікації, регресії, кластеризації і так далі. Наприклад, в задачі прогнозування часових рядів метою може бути прогнозування майбутніх значень з попередніми даними часового ряду.

Після визначення завдання, наступним кроком є вибір підходу до моделювання, що в нашому випадку є адаптивною метамоделлю, яка використовує нечіткий нелінійний беггінг. На цьому етапі формується структура моделі на основі вхідних даних та оптимізуються її параметри шляхом навчання.

У цей етап також входить визначення відповідних ознак для моделі. Ознаки є важливими чинниками в процесі навчання моделі, а їх вибір може суттєво вплинути на результативність прогнозу. Це означає, що, до вибору моделі, ми повинні визначити набір характеристик або атрибутів, які найкраще описують дані.

Після того, як завдання визначено, модель вибрана, а ознаки визначено, тренування моделі потім відбувається на базі навчального набору даних. Процес тренування може включати використання різноманітних методів оптимізації для покращення точності моделі й її здатності узагальнювати.

Останній етап включає використання натренованої моделі для прогнозування на основі нових даних, що надходять. Тут нечіткий нелінійний беггінг використовується для побудови остаточного прогнозу, що являє собою агрегований результат прогнозувань кількох нечітких моделей. Цей підхід дозволяє моделі краще впоратися з нестабільністю та варіативністю вхідних даних.

Загалом, задачі прогнозування можуть бути розглянуті як задачі навчання з підкріпленням, де наявність достовірного прогнозу слугує в якості винагороди. Ці задачі також включають у себе принципи, набуття досвіду та події, які впливають на майбутні рішення та дії.

Часто в задачах прогнозування використовують методи обробки природних мов та глибокого навчання. Наприклад, у задачах прогнозування тексту модель зазвичай використовує секвенційні дані, щоб визначити майбутню послідовність на основі минулих.

Окрім того, існує ще одна важлива складова, що впливає на структуру задачі прогнозування, а саме обмеження, пов'язані зі швидкістю вгадування. Для багатьох машин завдань прогнозування, важливим є максимально швидке вгадування, оскільки у реальному часі необхідно приймати рішення миттєво. Це вимагає від прогнозних систем оптимальності як за точністю прогнозу, так і за часом його формування.

Прогнозування зазвичай зосереджується на виявленні залежностей та взаємодій між різними явищами або об'єктами. Деякі задачі прогнозування передбачають визначення важливості окремих змінних або співвідношень між ними для передбачення певного результату. Інші задачі можуть вимагати побудови моделей, які одночасно враховують вплив багатьох змінних та їх взаємодії.

Завдання прогнозування також можуть бути багатоцільовими, коли модель одночасно має оптимізувати кілька вихідних величин. Це може бути важливим в умовах, коли різні цілі можуть конфліктувати одна з одною, і тому потрібно знайти оптимальний баланс між ними. Багатоцільове прогнозування може бути також корисним, коли моделі треба прогнозувати комплексні системи з багаточисельними взаємозалежностями.

Отже, структура задачі прогнозування має ряд особливостей, залежно від контексту, у якому вона використовується. Вони обумовлені характером даних, цілями та обмеженнями прогнозу, а також специфічними вимогами до моделі та її вихідних параметрів.

В нашому випадку, структура задач прогнозування, у яких використовується нечіткий неліній беггінг на основі адаптивних метамоделей, включає розробку придатного для задачі набору правил нечіткої логіки. Це передбачає розбиття простору вхідних змінних на

нечіткі області (нечіткі менші події) та побудову правил, що описують взаємозв'язок між областями та вихідними величинами.

Структура такого типу прогнозу також потребує розгляду проблеми нелінійності. Більшість реальних систем, які моделюються для прогнозування, є нелінійними. Розробка методів, що можуть ефективно працювати з такою нелінійністю, є одним з основних викликів в області прогнозування.

Ця структура задач також включає в себе оптимізацію. Це може передбачати добір найкращих параметрів для моделі, що мінімізують помилку прогнозу, або оптимізацію архітектури моделі для досягнення найкращої проектної точності в межах обмежень, наприклад, часу обчислень.

Ця робота зосереджена на використанні методів нечіткого нелінійного беггінгу в контексті адаптивних метамоделей для вирішення задач прогнозувань. Ця структура задач включає в себе процеси, які взаємодоповнюють один одного, для побудови прогностичних моделей, які враховують складність і різноманітність реального світу.

Поперше, це включає в себе збір та очищення вхідних даних. Це критично важливий етап, оскільки якість даних безпосередньо впливає на якість прогнозів. «Сміттєві входи приводять до сміттєвих виходів» – це справді справедливо для таких задач, оскільки неправильні, викривлені або відсутні дані можуть призвести до помилкових прогнозів.

Далі, після підготовки даних, ми визначаємо цілі прогнозування. Це може бути прогнозування якогось конкретного значення, категорії або набору можливих подій. Однак такі прогнози здійснюються з урахуванням нечіткості, тобто ми не очікуємо точного підбору значень, а намагаємося прогнозувати імовірнісні області, що містять майбутнє значення.

Після визначення цілей прогнозування, ми переходимо до побудови метамоделі, яка використовує нечіткий нелінійний беггінг. Метамодель – це модель, яка апроксимує поведінку іншої, більш складної моделі.

Ця процедура передбачає розбивку простору входних параметрів на нечіткі підмножини, що визначають область значень кожного параметра. Тоді для кожної такої підмножини створюється модель прогнозування, й нечіткі правила інтерполюють вклад кожної моделі в загальний прогноз.

Особливу увагу при цьому варто приділити обранню правильних параметрів для областей нечіткості й критеріїв агрегації беггінгу. Вибраний метод агрегації, як правило, залежить від конкретної задачі прогнозування, а варіантів може бути безліч – від простих сум або середніх значень до більш складних підходів, які враховують конкретні властивості даних.

Окрім того, для кожної підмножини також необхідно визначити правило виводу, що визначає, яким чином входні дані перетворюються на вихідні дані.

Після побудови моделей та присвоєння значень ваги кожній з них, настане час для беггінгу. Беггінг, або бутстреп-агрегація, – це метод ансамблю, що використовується для покращення стабільності та точності алгоритмів машинного навчання. Він зазвичай застосовується до високоваріативних процедур, що зменшує ризик перенавчання.

Метод беггінгу включає в себе створення кількох підмножин з навчального датасету за допомогою вибірки з поверненням, побудову моделі для кожної підмножини, а потім об'єднання їхніх прогнозів. В результаті, ми отримуємо модель, яка є більш стабільною, ніж окремі моделі, з яких вона складається.

У сфері нечіткої логіки беггінг може бути застосований для об'єднання декількох нечітких моделей в одну. Це робиться шляхом створення нечітких підмножин з навчального датасету, побудови нечіткої моделі для кожної підмножини, а потім взяття в середньому їхніх прогнозів.

Нечіткий нелійний беггінг використовується не тільки для зменшення варіативності, але й для кращого впорядкування з проблемою нечіткості в задачах прогнозування. Нечіткість тут виникає через взаємну залежність та непередбачуваність різних структурних елементів системи.

Адаптивні метамоделі, що використовують нечіткий неліній беггінг, поєднують метамоделювання з нечіткими системами. Метамоделювання дозволяє зменшити час обчислення шляхом заміни складної системи на більш просту, тоді як нечіткі системи допомагають краще впоратися зі схемами «якщо-то», що є загальним у задачах прогнозування.

Також варто відзначити, що задачі прогнозування, що використовують нечіткий неліній беггінг, можуть бути великою викликаючою для традиційних машинного навчання. Це пояснюється тим, що в таких задачах більшість рішень потрібно приймати на основі неповної або неоднозначної інформації, що не завжди може бути ефективно оброблено за допомогою класичних методів.

Основними перевагами математичного підходу, зокрема, використання нечіткого нелінійного беггінгу, в задачах прогнозування є гнучкість і обчислювальна ефективність. Моделі, засновані на нечіткій логіці, відомі своєю здатністю ефективно обробляти даними з нечіткістю і невизначеністю. Беггінг в свою чергу дозволяє досягати більш стабільних і точних прогнозів, комбінуючи виходи з кількох моделей.

Проте, зрозуміло, що даний підхід не позбавлений недоліків. Це включає в себе складності у виборі оптимального набору параметрів нечіткого нелінійного беггінгу, потенційні проблеми із перенавчанням при надмірному використанні беггінгу, а також витрати на обчислення при підборі оптимальних параметрів моделі.

Тому, при використанні цього підходу важливо вже на початку визначити стратегію вибору параметрів моделі, яка задовольняє потребу в балансі між обчислювальною витратою та якістю прогнозу. Деякі механізми для оптимізації таких параметрів можуть включати в себе використання алгоритмів пошуку або навчання з підкріпленням.

Всі ці чинники враховуються при розгляді структури задач прогнозування. Можливо, найважливішим моментом є те, що структура будь-якої задачі прогнозування буде в значній мірі визначатися

поставленими цілями, доступними даними і контекстом, в якому вона застосовується.

В загальному випадку, завдання прогнозування включає в себе декілька загальних етапів, включаючи збір і попередню обробку даних, вибір і налаштування моделі, прогнозування та аналіз результатів.

У випадку структури задач, що використовують нечіткі нелійні стратегії бегінгу на основі адаптивних метамodelей, може бути додані декілька специфічних елементів, таких як створення і використання нечітких систем, налаштування параметрів нечіткого нелійного бегінгу і використання нечіткості для практичних цілей.

Цей підхід може допомогти в різних областях та дисциплінах, де необхідно прогнозувати майбутні дані на основі існуючих даних. Він має потенціал вирішити ряд складних проблем в області прогнозування й дати можливість здійснювати більш точні й надійні прогнози.

Але, як ми вже відзначали, незважаючи на потенціал такого підходу, він також представляє певні виклики. Важливо заздалегідь подумати про стратегії подолання можливих труднощів.

Один з таких викликів – це запобігання перенавчанню. Це поширена проблема в машинному навчанні, коли модель «занадто добре» навчається на тренувальних даних і втрачає здатність загального перенесення на нові дані. У контексті нечіткого нелійного бегінгу, перенавчання може виникнути, якщо кожна із підмоделей надто відмінна від іншої. Існує багато способів регулювати цей процес, включаючи розмежування нечітких областей таким чином, щоб вони перекривалися і застосовували штрафи за складність моделі.

Ще одне потенційне обмеження – це велика обчислювальна витрата. Використання ансамблів моделей, особливо при використанні великих даних, може бути ресурсомісткою.

Це можна подолати, для прикладу, за допомогою паралельного обчислення або використання структурних прийомів, таких як прунктова апроксимація, щоб зменшити обчислювальне навантаження.

Нарешті, розробка ефективних стратегій налаштування параметрів може бути складним завданням. Однак, підхід, що використовується для цього, може бути весьма варіативним – від простих переборних методів до складніших методів оптимізації, таких як генетичні алгоритми.

Незважаючи на ці труднощі, нечіткий неліній беггінг на основі адаптивних метамоделей надає можливість для вирішення складних проблем прогнозування й варто було б додатково дослідити його для різних застосувань.

Окрім цього, є декілька майбутніх напрямків досліджень, які можуть допомогти у вирішенні викликів, пов'язаних зі структурою задачі прогнозування, які використовують нечіткий неліній беггінг на основі адаптивних метамоделей.

Дослідження в цьому напрямку може орієнтуватися на розробку нових підходів до вибору оптимальних параметрів для нечіткого нелінійного беггінг – це може передбачати генерацію варіантів параметрів з використанням генетичних алгоритмів, чи інших методів евристичної оптимізації. Також можна пошукати методи оцінки якості прогнозних моделей, які більш точно відображають необхідність управління нечіткістю.

Інший напрям може передбачати використання адаптивних стратегій для беггінг. Це може означати, що стратегії беггінг можуть бути автоматично налаштовані, в залежності від характеристик поточного набору даних. Наприклад, якщо визначено, що дані мають високу нестабільність, може бути використано більше моделей для беггінг, що дозволить зробити прогноз більш стабільним.

Також, міг би бути корисний розвиток нових методів візуалізації для нечітких моделей. Це може допомогти дослідникам і спеціалістам глибше

зрозуміти, як нечітка логіка працює в цих моделях, і як її можна використовувати для підвищення ефективності прогнозування.

У цьому контексті, стратегія нечіткого нелінійного бегінгу на основі адаптивних метамodelей пропонує значний потенціал для вирішення складних задач прогнозування. Вона дозволяє впоратися з нечіткістю і невизначеністю даних, які є типовими для реального світу, і створити міцні, надійні моделі прогнозування.

Проте, подібно до будь-якого інноваційного підходу, вона має свої виклики і обмеження. Перенавчання, обчислювальні витрати та потреба в ефективних стратегіях налаштування параметрів – деякі з проблем, які необхідно вирішити для повнішої реалізації її потенціалу.

Однак, з активним використанням в галузі машинного навчання та доповнених досліджень, цей підхід відкриває обіцяючі перспективи для розвитку нових технік і методів прогнозування. На розсіяння його обмежень може сприяти розширене застосування в різних областях, включаючи фінанси, медицину, екологію та багато інших, з новими ідеями, що виникають з кожним новим застосуванням.

1.3 Постановка задачі дослідження

Розглядаючи складність прогнозування в реальному світі, яке часто включає невизначеність і нечіткість, стратегії базовані на нечіткому нелінійному бегінгу показують значний потенціал. Цей підхід припускає теоретичний фундамент, що ґрунтується на розумінні, що світ, наповнений нечіткістю, вимагає гнучких і адаптивних прогностичних моделей для ефективного моделювання і прогнозування.

В основі нечіткого нелінійного бегінгу лежать принципи нечіткої логіки і бегінгу. Бегінг, що також відомий як бутстрепна агрегація, – це техніка машинного навчання, що створює множину моделей на основі первинного набору даних, а потім об'єднує їх у вихідну модель. Це дозволяє

беггінгу показувати стабільні результати, навіть якщо окремі моделі демонструють змінний рівень ефективності.

Попередні дослідження продемонстрували, що техніка беггінг може бути корисна для покращення точності моделей, які мають велику варіативність. Це може бути особливо корисним в контексті задач прогнозування, де потрібно створити надійний прогноз на основі вхідних даних, наповнених шумом і невизначеністю.

З іншого боку, нечітка логіка дає можливість моделювати і представляти нечіткість і невизначеність, надаючи специфічність до обробки даних. Поєднання нечіткого беггінгу та нелійного беггінгу розділює цей підхід від традиційних прогностичних моделей, роблячи його особливо здатним виконувати прогнозування в умовах складної нечіткості.

Нечіткі моделі характеризуються своєю вмістимістю відкрито розмежуватися і моделювати нечіткі і неоднозначні ситуації. Наприклад, прогнозування погоди з використанням нечіткості може бути здатним краще передбачати атмосферні умови, моделюючи широкий спектр можливих сценаріїв, а не просто найбільш імовірний. Це може використовуватися для підтримки рішень, що вимагають широкого розуміння майбутніх умов, таких як планування політики в галузі екології або забезпечення енергетичної безпеки.

Основним принципом нечіткого нелійного беггінгу є використання нечітких метамodelей як окремих моделей прогнозування. Кожна метамodelь створюється на основі нечіткого беггінгу, використовуючи частку вхідних даних, і використовується для створення прогнозу.

Ці прогнози, в свою чергу, агрегуються за допомогою нелінійної процедури для отримання кінцевого прогнозу. Цей підхід дозволяє створювати прогнози, які використовують багатий спектр вхідних даних і враховують складність і нечіткість цих даних через застосування нечіткого нелійного беггінгу.

Продовжуючи ідею складності і нечіткості реального світу, нечіткий нелінійний беггінг також включає в себе концепцію адаптивності. В цьому контексті, адаптивність означає здатність моделі адаптуватися до змін у вхідних даних і динаміці прогнозування.

Це означає, що модель відповідає на зміни в операційному середовищі шляхом оновлення своєї структури або параметрів, що дозволяє нечіткому нелінійному беггінгу вчасно вдосконалюватися та адаптуватися до змінних умов прогнозування.

Водночас із беггінгом, в основі нечіткого нелінійного беггінгу лежить принцип нечіткої логіки. Нечітка логіка, яка кардинально відрізняється від традиційної бінарної логіки, вперше була запропонована у 1965 році Лотфі Заде, який виділив її як математичний інструмент для моделювання неоднозначності та суб'єктивної нечіткості, що зустрічаються в реальному світі.

Разом, беггінг та нечітка логіка складають фундамент, на якому будується нечіткий нелінійний беггінг. Сама суть нечіткого нелінійного беггінгу полягає в створенні множини нечітких прогностичних моделей, кожна з яких навчається на дещо відрізняючихся від інших даних завдяки методу бутстрепа, а потім об'єднуванні цих моделей для створення кінцевого прогнозу. Цей процес є нелінійним через застосування нелінійних функцій активації при створенні прогнозів.

Подальше розширення концепції нечіткого нелінійного беггінгу включає в себе роль адаптивних метамоделей. У цьому контексті, адаптивність означає, що моделі спроможні вчасно адаптуватися до змін у сигналах або структурі вхідних даних. Цей процес включає в себе оновлення ваг моделі або зміну структури моделі з метою кращого відображення нових тенденцій або змін у вхідних даних.

Вивчення поведінки нечіткого нелінійного беггінгу в реальних умовах вимагає великого обсягу дослідницької роботи. При початковому тестуванні цього підходу можна використовувати синтетичні дані, створені

таким чином, щоб вони моделювали реальний світ з усією його нечіткістю та випадковістю. Це може включати випадкове додавання шуму до даних або використання різних нечітких форм розподілу для генерації синтетичних наборів даних.

Наприклад, однією з найпоширеніших форм розподілу є гауссова функція. Хоча деякі дані у реальному світі можуть слідувати гауссовому розподілу, багато інших типів даних мають різні форми розподілу, включаючи експоненційні, логнормальні, рівномірні, квадратичні і інші. Це надає велику гнучкість при моделюванні реального світу, дозволяючи нечіткому нелійному беггінгу відображати ці різноманітні форми розподілу.

У цьому процесі науковці можуть стикатися з рядом викликів, з якими потрібно боротися. Одним з таких викликів є обчислювальна складність нечіткого нелійного беггінгу, яка утворює значну кількість моделей та потребує великої кількості обчислень. Це може бути особливою проблемою при обробці великих наборів даних або розгортання моделі в реальному часі.

Хоча завдяки беггінгу велика частина обчислень може виконуватися паралельно, необхідно розробляти більш ефективні алгоритми та стратегії для оптимального використання обчислювальних ресурсів. Проблема обчислювальної ефективності є ще одним напрямком для подальших досліджень в області нечіткого нелійного беггінгу.

Загальний принцип ансамбля, який лежить в основі беггінгу, є ефективним інструментом у боротьбі з однією з найскладніших проблем у машинному навчанні – перенавчанням. Перенавчання відбувається, коли модель занадто добре вчиться на тренувальних даних, але не може ефективно узагальнювати знання на нових, не бачених раніше даних.

Це означає, що модель показує відмінну точність на тренувальних даних, але значно гіршує при тестуванні на тестових даних. Багато моделей, які представляють високий рівень складності, таких як глибокі нейронні мережі, особливо схильні до перенавчання. Беггінг, будучи методом

ансамблю, може допомогти в боротьбі з перенавчанням шляхом навчання багатьох моделей на різних вибірках даних та комбінації їх прогнозів.

Це дозволяє розрізнити шумові варіації від істинних закономірностей даних, покращуючи загальну гнучкість моделі. Негативний ефект перенавчання може бути пом'якшений використанням беггінгу, в доповнення до інших методів регуляризації.

Все експериментальне або прикладне застосування нечіткого нелійного беггінгу передбачає виконання ряду важливих кроків. До них можуть належати відбір відповідних факторів для включення в модель, оцінка різних моделей для беггінгу, визначення оптимального числа беггінгів, налаштування різних параметрів моделі, включаючи рівень нечіткості, і, нарешті, визначення, як краще агрегувати прогнози з різних моделей.

Такий процес вимагає завчасного планування та тестування, і може включати проведення ряду контрольних експериментів для перевірки впливу різних факторів на продуктивність моделі.

Сама природа нечіткості додає додатковий рівень складності до цього процесу. Нечіткість може виникнути на будь-якому з цих кроків, і його обробка вимагає детального розуміння, що включає в себе як теорію нечітких систем, так і практичне застосування нечітких методів в даному контексті.

Особливу увагу слід приділити відбору факторів. Вибір правильних або найбільш важливих факторів для прогнозування може виявитися важким завданням, особливо коли вибірка даних велика або коли структура даних складна. Критерії для відбору факторів можуть варіюватися від одного дослідження до іншого і можуть включати такі чинники, як статистична значимість, геометричні характеристики, здатність до визначення класу або прогнозування, та інші.

У процесі налаштування параметрів, слід звернути особливу увагу на те, як параметри впливають на продуктивність та стабільність моделі.

Низький рівень нечіткості може зробити модель надто чутливою до шуму, тоді як високий рівень може зменшити спроможність моделі до узагальнення. Параметри, такі як кількість бегінгів або ступінь нелінійності, також мають вплив на продуктивність моделі.

Завершальний крок включає в себе агрегацію прогнозів з різних нечітких моделей для створення кінцевого прогнозу. Це може бути зроблено за допомогою різних методів, включаючи просте голосування, вагове голосування, використання накопичувальної функції або інший тип ансамбльного методу.

Після успішного виконання цих кроків, нечіткий нелінійний беггінг може бути готовий до практичного застосування в задачах прогнозування. Однак, як і з будь-яким аналітичним інструментом, важливо постійно проводити моніторинг та оцінку продуктивності моделі.

Всі збираючі дослідження та експерименти з нечітким нелінійним беггінгом вимагають ретельної документації та аналізу. Кожен крок процесу, кожна вага та параметр моделей, результати проміжних та кінцевих прогнозів, а також всі помітні патерни та відхилення повинні бути акуратно записані та проаналізовані. Це не виключно для потреб контролю: це первинний інструмент для розуміння природи складних систем, моделювання та прогнозування, що дозволить внесення належних коректив в моделі за необхідності.

Крім того, на відміну від строгої, точної математики, з якою ми часто стикаємося в школах або університетах, мир нечітких систем і нечіткого нелінійного беггінгу не такий бінарний. Спроба звести все до «правилу» або «помилки» зазвичай не є продуктивною. Натомість, саме в цій області важливо врахувати градації та спектр можливих результатів та інтерпретацій.

Впливайте на результати з точки зору «високий», «середній» або «низький» рівень нечіткості, наприклад, і намагайтеся зрозуміти, як це відрізняється від просто квантового вимірювання, такого як «точно

прогнозування» або «невірне прогнозування». Це може допомогти забезпечити глибше і більш матизоване розуміння ваших моделей та їх прогнозів.

Поняття ролі необхідно впроваджувати постійно, адже нечіткий нелійний беггінг це не статична техніка. Адаптація – одна з його ключових властивостей, і це означає, що модель повинна постійно оновлюватися та покращуватися з часом відповідно до змін у даних та середовищі.

Хоча це може здаватися обтяжливим процесом, він є критично важливим для успіху в задачах прогнозування. Тут головна роль – постійна еволюція й навчання. Ми вчимося на власних помилках та успіхах, а також на успіхах та помилках інших дослідників і практиків.

Крім основних проблем, пов'язаних з моделлюванням та аналізом, існує також важливий аспект комунікації своїх результатів. Як дослідник або використовуючи нечіткий нелійний беггінг у повсякденних задачах прогнозування, вам слід бути здатними чітко та ефективно передавати свої методи, результати та висновки.

Це може означати створення графіків або діаграм, писання чітких доповідей або презентацій, а також підготовку себе до відповідей на питання та критичні зауваження із боку інших. Хоча ми можемо мати схильність зосереджуватися на технічній стороні речей, здатність ефективно комунікувати є не менш важливим фактором в успіху як дослідника.

Нарешті, дослідження та застосування нечіткого нелійного беггінгу можуть мати широкі імплікації в різних галузях, включаючи науку, медицину, інженерію, економіку та багато іншого. Отже, результати цього роботи можуть мати значний вплив, дозволяючи суспільству краще розуміти і прогнозувати складні, динамічні та нечіткі системи, з якими ми стикаємося в реальному світі.

Постановка задачі та проблема прогнозування вводять нас в серцевину викликів, з якими зіткнулася область машинного навчання та

аналізу даних. Завдання машинного навчання, особливо навчання з підкріпленням, полягає в створенні моделей, які можуть перетворити вхідні дані в конкретні вихідні результати або прогнози. Кінцева мета полягає в моделюванні складних процесів у реальному світі або в реалізації інтелектуальних задач створення прогнозів.

Для цього моделі мають бути не тільки точними, але й робустими та стійкими до шуму, випадкових помилок або варіабельності входу. Проте, в останні роки стало зрозуміло, що алгоритми, які намагаються узагальнити проблему, часто стикаються зі складністю, відомою як «прокляття розмірності».

«Прокляття розмірності» відноситься до обставин, за яких машинне навчання чи аналіз даних стає надзвичайно складним через велику кількість входів або факторів. Ці фактори, які можуть містити шум або незначні деталі, можуть перетворити задачу прогнозування в проблему великої складності та великої нестабільності.

Це приводить нас до одного з основних викликів у прогнозуванні, що полягає в розробці моделей, які не тільки точні та ефективні, але й здатні працювати в умовах великої нестабільності та складності. В останні роки було розроблено масу технік для рішення цього виклику, що включають ансамблі моделей, глибоке навчання, нейронні мережі тощо. Однак, незважаючи на ці прогресивніші методи, багато з них все ще мають обмеження та вимагають подальшого удосконалення.

У цьому контексті нечіткий нелінійний беггінг виникає як новий та цікавий підхід, що спрямований на розв'язання цих проблем. Він бере свою основу від беггінгу, методу машинного навчання, що використовує велику кількість «слабких» класифікаторів для створення «сильного» класифікатора. Ці моделі навчаються на різних підмножинах вхідних даних та комбінуються, створюючи більш точне та робустне прогнозування.

Однак, беггінг сам по собі є лінійним процесом. Він починає з випадково значущого відбору, потім проводиться обробка та навчання, а

потім ці моделі об'єднуються до фінального результату. Проте реальний світ та дані, з якими ми працюємо, часто можуть бути дуже нелінійними, з великими варіаціями та непередбачуваностями.

Продовжуючи цю логіку, нечіткий нелінійний беггінг вводить нелінійність в беггінг, завдяки нечіткій логіці та адаптивним метамоделям, намагаючись створити моделі, які більш точно моделюють складність реального світу.

Нечітка логіка, в свою чергу, включає використання нечітких множин та логіки для обробки неоднозначності, невизначеність або шум в даних. Вона пропонує гнучкий фреймворк для представлення та обробки даних, дозволяючи моделювати і прогнозувати, не скасовуючи важливу інформацію про нечіткість або неоднозначність.

Однак, встановлення та обробка нечітких правил вимагає допомоги від об'єктів адаптивного метамодельовання. Адаптивні метамоделі допомагають встановлювати важкі для визначення відношення між вхідними даними та вихідними даними в нечітких системах.

Таким чином, поєднуючи нечітку логіку, адаптивне метамодельовання та беггінг, ми отримуємо нечіткий нелінійний беггінг. Цей метод покликаний вдосконалити прогнозування в умовах моделювання складних та нелінійних систем.

Але як ми можемо визначити, чи дієво працює нечіткий нелінійний беггінг у задачах прогнозування? Якими критеріями слід керуватися при виборі моделей, що потребують інтеграції в ансамбль? Як ми можемо переконатися, що моделі не перенавчаються або не недонавчаються, враховуючи різні рівні шуму та нечіткості? Ці, та багато інших питань, складають основну проблему прогнозування.

У контексті нечіткого нелінійного беггінгу, вибір моделі може включати визначення різних методів для генерування та обробки нечітких даних, включаючи вибір технік агрегування прогнозів. Це вимагає глибокого знання різних технік моделювання, а також гнучкого підходу до

налаштування різних параметрів моделі, таких як ступінь нечіткості або кількість моделей для беггінгу.

Важливо також переконатися, що моделі не перенавчаються через надто високий рівень нечіткості чи складності. Перенавчання виникає, коли модель стає занадто «витонченою», адаптуючись до шуму або незначних деталей вхідних даних настільки сильно, що втрачає здатність правильно прогнозувати нові дані.

З іншого боку, під навчанням розуміється те, що модель недостатньо «витончена», щоб вловити важливі шаблони або відношення в даних. Це завжди є важким балансом, і визначення правильного рівня між цими двома крайнощами є одним із центральних викликів в усіх областях машинного навчання і аналізу даних.

В контексті моделей, заснованих на нечітких наборах, важливим також є визначення оптимального рівня нечіткості. Нечіткість у цьому контексті відноситься до ступеня «розмитості» між вхідними даними та вихідними результатами, або до того, наскільки добре модель може впоратися з неточностями чи невідомостями в даних.

На високому рівні нечіткості модель може включати в себе багато «шуму» або несуттєвих деталей, що може призвести до перенавчання. З іншого боку, низький рівень нечіткості може зробити модель недостатньо гнучкою для впорядкування з відмінностями в даних, що може призвести до піднавчання. Знову ж таки, знаходження балансу між цими двома крайнощами є одним із ключових викликів в задачах моделювання на основі нечітких наборів.

Крім цього, важливо врахувати, що не всі моделі або підходи однаково добре підходять для різних типів даних або прогнозувань. Наприклад, деякі моделі можуть краще працювати з більш стабільними або систематичними даними, тоді як інші можуть краще працювати з більш випадковими або хаотичними даними. Це означає, що вибір підходу часто вимагає великою мірою гнучкості та адаптивності з боку дослідника.

Також слід зазначити, що, користуючись беггінгом, дослідники мають бути обережними стосовно можливого надміру фрагментарного поділу даних. Оскільки цей метод ділить вихідні дані на багато менших частин і утворює прогнози для кожної з них окремо, існує ризик втрати важливих відношень або взаємозв'язків, які можуть з'явитися лише на більш великих рівнях даних.

Проблема перенавчання та піднавчання, вибір оптимального рівня нечіткості, належне врахування різних типів даних або завдань прогнозування, ризик фрагментарного поділу даних – це всього лише декілька з численних викликів, стосовно яких нас можуть поставити перед необхідністю розробки метою визначення ефективності нечіткого нелійного беггінгу в реальних задачах прогнозування.

Загалом, постановка задачі та проблема прогнозування в області нечіткого нелійного беггінгу вимагає від нас розуміти, що моделі мають існувати в динамічному стані адаптації, що відображає зміну структури даних, яка виникає в нашому швидкісному світі. Хоча цей процес може бути викликаючим, це також створює ще більші можливості для інновацій та новаторських рішень.

Як тільки ми зможемо створити потужні, адаптивні моделі, які здатні краще впорюватися з випадковості та неоднозначністю даного світу, ми будемо мати можливість більш точно моделювати та прогнозувати важливі тенденції в таких *vital areas* як кліматичні зміни, медичне прогнозування, економічне розвиток та багато інших.

Нарешті, які б проблеми ми не зіткнулися при моделюванні та прогнозуванні, вони залишаються основними для нашого розуміння і впорядкування складного та швидко змінювального світу. Як дослідники або практики в області аналізу даних та машинного навчання, ми повинні бути на передньому краю цих досліджень, розвиваючи нові методи, моделі та техніки для підвищення точності та надійності наших прогнозів.

Як наголошувалось, нечіткий нелійний беггінг може виявитися цінним інструментом для досягнення цієї мети. Проте, такий підхід не може бути застосований відокремлено або без критичного аналізу та оцінки. При його застосуванні необхідно враховувати всі вимоги, обмеження та потенційні проблеми, які були описані в цьому повідомленні.

1.3.1 Класифікація обраної задачі та її особливості

Класифікація задачі, яка вирішується за допомогою нечіткого нелійного беггінгу, належить до сфери обробки та аналізу даних. Це включає порівняння різних даних, моделювання та прогнозування, а також пошук взаємозв'язків і закономірностей в цих даних. Проте, нечіткий нелійний беггінг вносить свої особливості в цей процес, зокрема шляхом інтеграції нечіткої логіки та адаптивного метамоделювання.

Процес класифікації в основному полягає в розподілі даних на дві або більше категорій на основі вивчення, аналізу і порівняння вхідних даних. Завдяки класифікації, ми можемо краще розуміти структуру даних та їх взаємозв'язки, що є засобом до створення дієвих моделей прогнозування. Проте, цей процес може бути складним і вимагає від нас здатності працювати зі складними системами та даними.

В контексті нечіткого нелійного беггінгу, сам процес класифікації включає в себе цілу низку унікальних аспектів та викликів. Процес включає щонайменше три основні кроки: налаштування моделі, створення ансамблю моделей, а також адаптація та вимірювання моделі.

Адаптація моделі включає в себе не тільки створення і налаштування окремих моделей, але і перевірку того, наскільки добре вони працюють як частина загального ансамблю. Це може вимагати врахування багатьох різних факторів, враховуючи пропорцію моделі в ансамблі, її ефективність при роботі з різними типами вхідних даних, а також рівень «нечіткості» або «розмитості», якими ця модель працює.

Також важливим є процес створення ансамблю моделей. Це включає визначення способу об'єднання інформації з різних моделей, а також визначення методів застосування їх для окремих прогнозів.

Оскільки нечіткий неліній беггінг використовує ансамблі «слабких» класифікаторів для створення «сильного» класифікатора, вибір моделей та їх комбінування є критичним етапом.

Адаптація та вимірювання моделі відносяться до процесу оцінки та тонкої настройки моделей, щоб вони найкращим чином відповідали вхідним даним. Це може включати регулювання параметрів нечіткості, доналаштування ваг ансамблю моделей та корекцію шуму або помилок.

В контексті нечіткого нелінійного беггінгу, ці процеси можуть бути особливо складними, оскільки вони включають в себе роботу з нелінійними моделями та нечіткими системами. Однак, якщо це виконується правильно, вони можуть допомогти вирішити деякі з основних викликів прогнозування, включаючи питання перенавчання і піднавчання та «прокляття розмірності».

Підводячи підсумок, класифікація і особливості задачі прогнозування за допомогою нечіткого нелінійного беггінгу вимагають що ми були гнучкими і ресурсними в нашому підході до машинного навчання. Це включає в себе використання глибокого розуміння структури даних, вміння працювати з розмаїттям моделей, вибір і налаштування різних технік в залежності від особливостей конкретного завдання, а також здатність працювати з розмаїттям входів та умов.

Подолання викликів прогнозування, визначення оптимальної моделі для задачі, налаштування параметрів моделі та її оцінка, це невід'ємна частина стратегії нечіткого нелінійного беггінгу. З ним ми можемо реалізувати переваги нечіткої логіки та адаптивного метамоделювання з метою покращення прогнозування у контексті обробки та аналізу даних.

1.3.2 Визначення цілей і завдань роботи

Головною метою цієї роботи є дослідження, розуміння та оптимізація нечіткого нелінійного беггінгу як методу для покращення точності та ефективності прогнозування неоднозначних, складних та шумних систем даних. Це передбачає низку важливих завдань та цілей, що вимагають глибокого розуміння теорії та використання інтелекту для ефективного використання різних технологій аналізу даних.

Зокрема, одним з завдань є встановлення ефективного використання нечіткої логіки, адаптивного метамоделювання та беггінгу в процесі аналізу даних. Це включає вивчення того, як ці технології поєднуються та взаємодіють для створення сильних, точних моделей прогнозування, а також того, як вони можуть бути оптимізовані та налаштовані для виконання конкретних завдань.

Іншим важливим завданням є розробка методів для боротьби з проблемами перенавчання та піднавчання. Ці проблеми є частою проблемою моделювання даних, і вони можуть суттєво вплинути на якість прогнозування. Завдяки покращенню нашого розуміння цих проблем та розробці ефективних стратегій їх вирішення, ми зможемо покращити загальну якість наших моделей та прогнозів.

Також важливо визначити, як найкраще використовувати нечіткий нелінійний беггінг для роботи з різними типами даних та завдань. Тобто, ми маємо розглядати різні сценарії використання, а потім визначити, який більш підходить і за яких обставин. Це може включати в себе вивчення того, яким чином умови шуму, неоднозначності або складності системи можуть впливати на те, як найкраще використовувати цей метод прогнозування.

Після визначення цих умов та стратегій одним з основних завдань є реалізація та тестування нечіткого нелінійного беггінгу в реальних сценаріях прогнозування. Це означає, що ми повинні розробити методики застосування алгоритмів до реальних, зібраних даних, перевірити якість

прогнозів, що вони дають, та адаптувати та оптимізувати їх відповідно до одержаних результатів.

До завдання належить також консолідація отриманих результатів в цільову форму, яка забезпечить візуальний, легкий для сприймання і розуміння вихід. Це може передбачати побудову графіків, таблиць або інших візуалізацій, які дозволять нам краще розуміти залежності та прогнозування, які демонструє нечіткий нелінійний беггінг.

Наступне важливе завдання включає критичний аналіз отриманих результатів. Оскільки сам метод нечіткого нелінійного беггінгу все ще знаходиться в процесі розвитку, важливо вивчити ті обмеження та потенційні проблеми, з якими ми можемо зіткнутися при використанні цього методу. На основі цих висновків ми зможемо зрозуміти, чи варто включати нечіткий нелінійний беггінг як один з основних інструментів прогнозування в наших майбутніх проєктах з аналізу даних.

Всі ці завдання впливають на головну мету: ефективну роботу з неоднозначними, шумовими та складними даними, використовуючи нечіткий нелінійний беггінг як інструмент для покращення точності та ефективності прогнозування. Вони вимагають від нас розуміння, аналітичних навичок, обчислювальних знань та вміння використовувати доступні технічні ресурси ефективно.

Кінцевим завданням ми маємо подолати виклики прогнозування, покращити якість прогнозування та відкрити нові можливості для ефективного моделювання даних та прогнозування.

2 МЕТОДИКИ НЕЧІТКОГО НЕЛІЙНОГО БЕГГІНГУ

2.1 Аналіз принципів нелінійного беггінгу в задачах прогнозувань

Беггінг (Bootstrap Aggregating), що був вперше запропонований Leo Breiman в 1996 році, є одним з найпопулярніших методів ансамблювання, які використовуються в машинному навчанні. Основний принцип беггінгу полягає в формуванні багатьох підмножин вихідних даних з поверненням, тренуванні моделі на кожній підмножині та потім об'єднанні прогнозів цих моделей. Беггінг допомагає зменшити варіативність прогнозів, а тому він є особливо корисним для «шумних» наборів даних.

Ці принципи беггінгу можуть бути адаптовані для нелінійних моделей та прогнозування. Нелінійні моделі характеризуються тим, що вони мають змінну відповідь, яка не прямо пропорційна вхідним змінним. Вони можуть відображати складні та неочевидні взаємодії між змінними та відповіддю. Нелінійний беггінг використовує принципи беггінгу для формування ансамблю нелінійних моделей та об'єднання їх прогнозів.

Процес нелінійного беггінгу включає в себе декілька важливих етапів: формування підмножин даних, тренування нелінійних моделей на цих підмножинах даних, роботу з цими моделями для формування прогнозів та об'єднання цих прогнозів для створення кінцевого прогнозу.

У контексті нечіткого нелінійного беггінгу ці принципи використовуються з деякими відмінностями. Зокрема, моделі, які використовуються в нечіткому нелінійному беггінгу, часто включають аспекти нечіткої логіки, які дозволяють краще впоратися з неоднозначністю та шумом у даних.

Також важливо розглянути, як ми можемо адаптувати нелінійний беггінг для різних завдань та даних. Одним з ключових аспектів нечіткого нелінійного беггінгу є його можливість адаптації до різних типів даних та завдань. Це означає, що він може бути використаний для прогнозування

різних типів відповідей, включаючи неперервні або категоріальні відповіді, та що він може працювати з різними типами вхідних даних, включаючи числові, категоріальні або текстові дані.

Одним із ключових принципів беггінгу є генерація множини «слабких» моделей на різних підмножинах вихідних даних, а потім їх агрегація для створення «сильної» моделі.

У нелінійному беггінгу цей принцип застосовується до побудови нелійних моделей. Адже нелінійні моделі можуть бути вкрай ефективними для моделювання складних взаємозв'язків між даними.

Часто натрапляємо на випадки, коли взаємозв'язки між вхідними даними та вихідною змінною не є прямолінійними. В таких випадках лінійні моделі можуть не впоратися з викликом апроксимації цих взаємозв'язків. Нелінійні моделі, в свою чергу, мають гнучкість, яка дозволяє їм адаптуватися до таких складних взаємозв'язків.

Однак, однією з проблем при використанні нелінійних моделей є те, що вони часто схильні до перенавчання. Перенавчання відбувається, коли модель дуже гарно пристосовується до тестових даних, але показує погані результати на нових, невідомих даних.

Принцип беггінгу дозволяє нам подолати цю проблему. Беггінг створює множину різних моделей на різних підмножинах вихідних даних. Коли ми агрегуємо прогнози цих моделей, ми отримуємо прогноз, який є більш стійким до шуму та випадку, ніж будь-яка окрема модель.

Системи нечіткої логіки допускають можливість розмитості та неоднозначності, що робить їх вкрай корисними при роботі з неоднозначними, шумними або неповними даними. Нечіткий нелійний беггінг приймає цю концепцію й інтегрує її в процес формування ансамблю моделей прогнозування.

Нечіткі системи забезпечують природне кодування неоднозначності за допомогою використання членських функцій, які мають діапазон значень між 0 та 1, замість прямих бінарних значень (таких як «правда» або

«брехня», «до» або «після», «великий» або «малий»). Це робить нечіткі системи ідеальними для аналізу даних, в яких присутня неоднозначність або невизначеність.

Оскільки багато реальних прогнозувань мають середні, невпевнені чи неясні значення, впровадження нечіткості в неліній беггінг може покращити якість і надійність прогнозів. Наприклад, у випадку з моделюванням погоди на прогнози впливають численні фактори, які можуть бути вкрай неоднозначними та невизначеними.

При застосуванні принципу беггінгу для нечіткого нелінійного прогнозування важливим є впровадження нечіткості в процес обчислення. Це може бути досягнуто за допомогою нечіткого кластеризації та нечіткої логіки.

Іншим принципом, який є особливо важливим для нечіткого нелінійного беггінгу, є ансамблювання. Ансамблювання дозволяє створити групу прогнозів, які можуть бути об'єднані для створення кінцевого прогнозу. Це сприяє покращенню стабільності та точності прогнозування, оскільки єдиний прогноз може бути схильним до помилок або шуму. Таким чином, ансамблювання є ключовою фічею нечіткого нелінійного беггінгу.

Принципи нелінійного беггінгу також можна адаптувати до конкретних проблем і типів прогнозувань. Нелінійний беггінг може бути використаний не тільки для регресійних прогнозів (де вихідна змінна є неперервною), але також для класифікаційних прогнозів (де вихідна змінна є дискретною).

Ключовим аспектом адаптації беггінгу до проблеми класифікації є те, що відповідь є дискретною, а не неперервною. Це означає, що прогнози різних моделей не можуть просто бути усереднені. Замість цього, потрібно отримати прогноз за кожним класом для кожної моделі, а потім об'єднати ці прогнози для отримання кінцевого прогнозу.

Така ситуація зустрічається, наприклад, при розпізнаванні образів, класифікації тексту або діагностиці захворювань. У всіх цих випадках,

результатом є дискретний клас, а не неперервне значення. Прикладами можуть бути визначення, чи є електронний лист спамом, чи виявляє зображення обличчя людини, чи має пацієнт певне захворювання на основі його медичних показників.

Важливим є те, що на основі цих принципів можливо розробляти алгоритми та стратегії для адаптації процесу беггінгу до конкретних вимог і контексту завдань прогнозування. Наприклад, одна з можливих стратегій включає в себе використання принципу голосування більшості для комбінування прогнозів, що зроблені різними моделями. При цьому прогноз з найбільшою кількістю голосів обирається як кінцевий прогноз. Цей підхід за поведінкою нагадує методи комітетного голосування, відомі в області статистичного прогнозування.

Принципи нелійного беггінгу також можуть бути адаптовані для роботи з даними різного характеру і структури, включаючи числові, категоріальні або текстові дані. Важливо адаптувати процеси навчання моделі, створення прогнозів і їх агрегування до особливостей таких даних.

Текстові дані прогнозування можуть бути особливо складними для беггінгу, оскільки вони мають високу вимірність і велику кількість єдиних ознак. Однак, застосування беггінгу для текстових даних можливе, і цей метод може надзвичайно ефективний. Це пов'язано з тим, що текстові дані часто містять велику кількість шуму, і беггінг, завдяки своїй здатності об'єднувати прогнози з різних моделей, може ефективно боротися з цим шумом і покращувати якість виходу.

Беггінг також може бути подібним чином адаптований для роботи з числовими і категоріальними даними. Зокрема, при роботі з категоріальними даними може бути використано принцип голосування більшості, де кінцевий прогноз обирається на основі того, яка категорія була найчастіше прогнозована моделями в ансамблі.

Незалежно від типу вхідних даних, ключовим аспектом нелійного беггінгу є його здатність адаптуватися до складних і шумних сценаріїв

моделювання. Нехай даними є числа, категорії або текст, беггінг здатний обробляти шуміння і неоднозначності в даних, створюючи точний і стабільний прогноз.

Важливим елементом будь-якого процесу прогнозування є визначення його успішності. У контексті нелійного беггінгу це може бути здійснено шляхом застосування відповідних метрик оцінювання та методик дослідження, відповідних до завдань прогнозування.

Метрики оцінювання повинні вимірювати точність виходу, її стабільність протягом часу та здатність моделі адекватно реагувати на нові дані. Вони можуть включати, наприклад, коефіцієнт детермінації R^2 для регресійних задач або точність, відгук, F1-міру та область під кривою приймального (AUC-ROC) для задач класифікації.

Дослідницькі методики можуть включати крос-валідацію та блок-валідацію для наборів даних лімітованого розміру, бутстреп для наборів даних великого розміру або методики розбиття на навчальну, валідаційну та тестову вибірки.

Також повинна бути проведена аналіз успішності не лінійного беггінгу протягом часу. Можна враховувати, як успішності моделі міняється при навчанні з різними наборами даних та який вплив на це мають зміни у параметрах моделі. Ще одним важливим аспектом є вплив шуму в даних на точність та стабільність прогнозів. Багато реальних сценаріїв прогнозування включають в себе шумові і нестабільні дані, так що отримання стабільних і точних прогнозів в таких умовах є ключовим завданням.

Принципи нечіткого нелійного беггінгу можуть бути також адаптовані до задач прогнозування в різних галузях. Деякі зможуть застосувати цей метод для прогнозування фінансових трендів на основі історичних даних про ціни на акції. Інші застосування можуть включати прогнозування патернів споживання енергії, виробництва агрокультур, розповсюдження хвороб та багато іншого.

При застосуванні беггінгу до прогнозування у кібербезпеці може бути використано більш складні моделі машинного навчання, які взаємодіють з великим набором даних та застосовуються для виявлення аномалій та атак.

Для прогнозування динаміки змін клімату можна використати беггінг, використовуючи набір даних з різних джерел, таких як супутники, метеорологічні станції та океанографічні буї.

Таким чином, принципи нечіткого нелійного беггінгу можуть бути вкрай корисними для різних галузей, де прогнозування є важливим інструментом прийняття рішень. У всіх цих випадках беггінг може сприяти покращенню точності прогнозів, забезпечуючи при цьому їх стабільність та надійність в умовах шуму та неоднозначності даних.

Визначення оптимальних параметрів для моделей, включених в ансамбль нечіткого нелійного беггінгу, також є важливим кроком. Етап налаштування параметрів зазвичай включає в себе такі елементи як вибір архітектури моделі, вибір функції активації, визначення кількості епох навчання, розміру партії, темпу навчання та інших гіперпараметрів.

Точнісінько підібрані параметри можуть істотно поліпшити точність і надійність прогнозів, що виходять. Однак важливо пам'ятати, що покращення призводять до збільшення складності моделі та часу її навчання. Особливо це стає помітним під час роботи з великими наборами даних або зі складними моделями.

Як варіант, процес налаштування параметрів може бути автоматизований за допомогою таких технік як перебір сітки (grid search) або випадковий пошук (random search). Однак, такі методики можуть бути вкрай затратними за часом, тому важливо знайти баланс між точністю моделі та часом її налаштування та навчання.

Також варто пам'ятати, що процеси налаштування параметрів і тренування моделей мають бути здійснені на окремих наборах даних для уникнення перенавчання. Зазвичай, дані діляться на окремі набори для тренування, валідації та тестування моделі. Дані для тренування

використовуються для навчання моделі, дані для валідації для налаштування параметрів, і, нарешті, дані для тестування використовуються для визначення кінцевої точності та надійності прогнозів моделі.

Останнім, але не менш важливим, є етап впровадження вивчених моделей та технологій в реальні обставини. Тут може бути важливим знайти способи застосування та інтеграції моделей в існуючі системи та процеси, наприклад, інтегрувати прогнозну модель у систему підтримки прийняття рішень або використовувати її для автоматичного контролю робочих процесів.

Під час відпрацювання реальних сценаріїв можуть зустрічатися різні виклики:

- зміна структури або якості вхідних даних;
- технічні проблеми, пов'язані з обслуговуванням моделі;
- проблеми з успіхом обчислення прогнозів в режимі реального часу.

Часто ці проблеми вимагають адаптації подальшої моделі, що вимагає доцільного проектування і планування з самого початку.

Виклик полягає в тому, що модель, яка працює добре в контрольованому середовищі лабораторії, може не давати тих самих результатів під час експлуатації в реальному світі. Проблеми можуть виникнути через змінність умов роботи, шум або аномалії в даних, або непередбачені події, які не були враховані під час початкового навчання моделі.

Врахування цих факторів від початку може допомогти в розробці більш надійних і стабільних систем, які здатні працювати в реальних обставинах. Таким чином, вивчення та впровадження нечіткого нелінійного беггінгу стає вкрай важливим для розробників рішень на основі машинного навчання та штучного інтелекту.

Після впровадження моделі у реальне середовище, важливим є також підтримувати та моніторинг її продуктивності. До цього часто включає перевірку якості прогнозів і відстеження будь-яких змін з часом. Це

дозволяє своєчасно виявляти та виправляти будь-які проблеми, а також робити належні покращення, коли це необхідно.

Варто регулярно перевіряти якість даних, що приходять, щоб переконатись, що вони є достатньо якісними та релевантними. Це може включати виявлення аномалій, визначення шуму, перевірка достовірності та потреба в очищенні даних. Це допоможе забезпечити, що модель продовжує отримувати відповідні вхідні дані та видає коректні прогнози.

Підтримка моделі також може включати її постійне оновлення або навчання з плином часу, особливо коли прогнозується динаміка, яка змінюється з часом. Для цього можуть застосовуватись методи постійного навчання, для яких модель періодично адаптується до нових даних, що надходять.

Створення механізмів для автоматичного моніторингу та підтримки моделі може бути важливим для збереження її високої продуктивності та надійності. Все це робить нечіткий неліній беггінг не тільки потужним інструментом прогнозування, але й досить прагматичним і придатним для використання в реальних обставинах.

Після успішного впровадження та моніторингу, наступною важливою фазою є неперервне удосконалення моделі або системи нечіткого нелінійного беггінгу. Це зазвичай включає перегляд і оновлення моделей, методів навчання та стратегій прогнозування на основі нових досліджень, методів аналізу даних та зростаючого обсягу та складності використовуваних даних.

Однією з ключових особливостей систем штучного інтелекту та машинного навчання є їхню здатність до навчання та удосконалення. Нові методи, алгоритми та технології в області штучного інтелекту та машинного навчання постійно розвиваються, а використання нового обладнання, програмного забезпечення та обчислювальних ресурсів може ще більше посилити можливості цих систем.

Також важливо зосередитися на інтерпретованості та прозорості моделі. Хоча нечіткий неліній беггінг може бути вкрай потужним

інструментом прогнозування, важливо зрозуміти, як саме модель робить свої прогнози. Це може бути важливо для визначення напрямків подальшого удосконалення моделі, а також для визначення її обмежень.

На завершальному етапі, важливо провести оцінку інтеграції системи у бізнес-процеси та конкретні застосування. Це може включати оцінку економічної вигоди, ефективності прогнозування, продуктивності, а також легкості використання та взаємодії з системою кінцевими користувачами. Ця оцінка може допомогти в оптимізації системи для покращення її ефективності та вивчення можливостей для подальшого розвитку.

Беггінг, незалежно від його варіанту або області застосування, є вкрай потужним інструментом перевіреної якості для покращення стабільності прогнозів та підвищення точності моделей машинного навчання. Незважаючи на його складність, принципи беггінгу можна адаптувати до широкого спектру завдань, типів даних та конкретних вимог. Його застосування обіцяє значне покращення результатів, особливо в складних і шумних середовищах.

Однак, необхідно пам'ятати, що успішне використання беггінгу вимагає не лише розробки ефективних моделей і алгоритмів, але й створення ретельно продуманих стратегій підготовки даних, тренування, налаштування параметрів, тестування та моніторингу. Він також вимагає високої кваліфікації і глибокого розуміння як основних концепцій машинного навчання, так і специфіки конкретного прогнозованого завдання.

У кінцевому підсумку, хоча беггінг може представляти собою значний виклик, він також представляє величезний потенціал. З правильним підходом та належною оцінкою, беггінг може істотно збільшити точність та здатність моделей машинного навчання передбачати невідомі динаміки, покращуючи в той же час їх загальну стабільність і надійність. Це робить його незамінним інструментом для будь-якого дослідника або практика в області аналітики даних, машинного навчання або штучного інтелекту.

2.2 Методологія та математичне моделювання нечіткого нелінійного беггінгу

Методологія та математичне моделювання нечіткого нелінійного беггінгу має глибокі корені в статистиці, теорії прийняття рішень та обчислювальному інтелекті. Нелінійний беггінг поєднує принципи беггінгу з нечіткою логікою та нелінійним моделюванням, що дозволяє формулювати складні моделі, адаптовані до розв'язання реальних проблем прогнозування.

На відміну від традиційного, лінійного беггінгу, який в основному фокусується на редукції дисперсії, нечіткий нелінійний беггінг дозволяє також редукувати зміщення моделей машинного навчання. Це досягається за допомогою створення ансамблю моделей, які навчаються на різних підмножинах вхідних даних і потім об'єднуються для створення кінцевого прогнозу.

Нечітка логіка вводиться в процес беггінгу для використання його сильних сторін у роботі з неоднозначністю і неясності та обробці вхідних даних, які не є чітко визначеними або які містять шум. Згідно з принципами нечіткої логіки, вхідні дані можуть належати до декількох класів з різним ступенем належності, що дозволяє моделі краще обробляти нечіткості та неоднозначності.

Нелінійність вводиться в процес беггінгу для кращої адаптації до складних проблем прогнозування, які не можна ефективно розв'язати за допомогою лінійних моделей. Нелінійні моделі можуть набагато краще впоратися з взаємозв'язками між признаками, які не є прямолінійними, або з більш комплексними структурами даних.

Першим етапом створення нечіткої нелінійної беггінг моделі є обробка вхідних даних. Використовуючи принципи нечіткої логіки, вхідні дані можуть бути перетворені в нечіткі множини, які відображають різницю в ступені належності до різних класів. Це може бути зроблено за допомогою

різних методів нечіткого виведення, таких як функції приналежності, нечіткі правила, чи методики розмиття.

Другим етапом є навчання ємності моделей на підмножинах вхідних даних. Цей етап починається з генерації підмножин вхідних даних, які зазвичай вибираються з заміною з оригінального набору даних. Цей процес відомий як бутстрепінг і дозволяє створити різні набори даних, що містять повторення деяких прикладів і може виключати інші.

Після того, як підмножини даних сформовані, для кожної підмножини даних можна створити модель навчання. Вибір конкретної моделі залежить від конкретної проблеми, що вирішується, але всі моделі мають бути нелінійними для використання цієї методології. Декілька популярних варіантів моделей включають нейронні мережі, нелінійні регресійні моделі або навіть складніші моделі, такі як глибокі нейронні мережі або моделі з градієнтним посиленням.

Кожна модель потім навчається на своїй підмножині даних, що дозволяє кожній моделі «вивчити» різні аспекти даних. Це може допомогти моделі краще узагальнити на основі набору тренувальних даних, оскільки кожна модель навчається на трохи інших даних.

Третім етапом є об'єднання прогнозів від кожної окремої моделі в кінцевий прогноз. Це зазвичай робиться за допомогою голосування більшості або усереднення прогнозів. Знову ж таки, метод об'єднання може бути адаптований до конкретної проблеми прогнозування та вимог до моделі.

На цьому етапі, важливою є інтерпретація результатів моделі. Це може бути складно через складність моделі та нечіткість виведених даних. Однак, завдяки застосуванню нечіткої логіки, можна отримати більш інформативні і зрозумілі результати.

Наприклад, виведені декілька процентів належності до різних класів можуть надати більше інформації про прогноз, ніж просте визначення

одного класу. Додатково, використання нечітких множин дозволяє моделі обробляти неоднозначність у вхідних даних або нечіткість у прогнозах.

Після отримання кінцевого прогнозу, наступним кроком є його перевірка. Це може включати в себе порівняння прогнозу з реальними даними або з іншими моделями. Також можна використовувати різні метрики оцінки, такі як точність, згадуваність, F-міра або кореляція. Це допомагає визначити, наскільки добре модель впоралася з завданням прогнозування, а також які аспекти моделі можуть вимагати удосконалення.

Всі ці кроки повинні повторюватися в процесі вдосконалення моделі. Додаткові дані, нові методи обробки даних або нові алгоритми навчання можуть використовуватися для вдосконалення як окремих моделей, так і всього ансамблю. Регулярний перегляд і оновлення моделі дозволяє забезпечити її актуальність та високу продуктивність.

Моделювання нечіткого беггінгу, на відміну від традиційних методів, зазвичай вимагає більше обчислювальних ресурсів через використання великої кількості копій моделей та розрахунків цих прогнозів. Однак, завдяки нечіткому нелійному беггінгу, здатному краще впоратися з шумом і неоднозначністю в даних, покращенням якості прогнозів часто виправдовується збільшений обсяг обчислень.

Для практичної реалізації нечіткого нелійного беггінгу зазвичай використовуються високорівневі мови програмування, такі як Python або R, які мають велику кількість бібліотек і пакетів для машинного навчання і обробки даних. Деякі популярні бібліотеки, які можуть використовуватися для цього, включають scikit-learn, TensorFlow, Keras і PyTorch для машинного навчання та numpy, pandas і matplotlib для обробки і аналізу даних.

Важливим аспектом реалізації є вибір найкращої моделі для кожного випадку. Різні моделі можуть мати різні точності, швидкість навчання, складність та інші характеристики. Вибір найкращої моделі залежить від

конкретних вимог, таких як проблеми, які потрібно вирішити, доступність даних, обчислювальні ресурси та інше.

У практичному застосуванні, необхідно розглянути декілька важливих викликів і обмежень нечіткого нелінійного беггінгу.

Перше обмеження походить від складності моделей. На відміну від простих лінійних моделей, нечіткі нелінійні моделі мають багато параметрів, які потребують підбору. Це може вимагати більше часу і обчислювальних ресурсів для прогнозування, а також вимагати додаткових зусиль для інтерпретації результатів.

Друге обмеження полягає в тому, що нечіткі нелінійні моделі є відносно новим і менш вивченим полем. Хоча вони показали обіцяючі результати, їх характеристики та оптимальні параметри повинні бути ще детальніше досліджені. Це вимагає оптимального балансу між дослідницькою роботою для розробки нових моделей та практичною роботою для їх застосування.

Ці виклики може бути подолано через ретельний дизайн експериментів, використання надійних методів валідації, проведення вичерпних тестів та ефективного використання обчислювальних ресурсів. Це може допомогти визначити оптимальні параметри моделі, виявити можливі проблеми та виробити стратегії для їх вирішення. Також це може допомогти в отриманні більш точних і зрозумілих прогнозів, що може призвести до більш ефективного використання моделі.

Загалом, нечіткий нелінійний беггінг є цікавим і потенційно потужним методом прогнозування, який може знайти застосування в різних галузях. Його гнучкість і спроможність ефективно працювати з нечіткими і нелінійними структурами даних роблять його привабливим для дослідників і практиків.

Розглянемо більш детально конкретні технічні аспекти роботи з нечітким нелінійним беггінгом.

При побудові нечіткої нелінійної беггінг моделі ключовим етапом є створення нечітких множин. Оскільки вони мають багато можливих форм,

вибір правильної форми вимагає розуміння природи даних. Наприклад, якщо дані мають довгі хвости розподілу, може бути корисно використати гауссову функцію приналежності, яка надає більшого ваги даним, що знаходяться далеко від середнього. Однак, якщо дані більш рівномірно розподілені, можливо, гіперболічна пряма функція приналежності буде кращою в цьому контексті.

Крім того, важливим етапом при побудові нечіткої нелінійної беггінг моделі є вибір алгоритму нелінійного навчання. Деякі з них, наприклад, нейронні мережі або дерева рішень, мають потужні можливості для моделювання нелінійних взаємозв'язків. Однак, при виборі алгоритму потрібно бути обережним, оскільки деякі з них можуть бути складні для налаштування і перевірки, а також вони можуть бути чутливими до поверхневого навчання.

При створенні фінального прогнозу використовуються результати кожної моделі беггінгу, які об'єднуються за допомогою процедури голосування. Цей процес може бути простим, наприклад, більшість голосів вирішує про кінцевий прогноз, або може бути більш складним, включаючи розрахунок ваги для кожної моделі або навіть застосування додаткового шару машинного навчання для оптимізації комбінації прогнозів.

Непереможним перевагам нечіткого нелінійного беггінг моделювання більше відповідають деякі виклики. По-перше, складність моделей може перетворитись на високі вимоги до обчислювальних ресурсів, особливо при значних обсягах даних. По-друге, сформулювання відповідних нечітких множин та оптимізація параметрів може бути трудомістким процесом, який вимагає значної експертної знань та багато часу.

Не дивлячись на це, універсальність і гнучкість нечіткого нелінійного беггінгу робить його вельми привабливим варіантом для моделювання в широкому спектрі галузей та застосувань. Зокрема, воно може бути особливо корисним в умовах високої нестійкості та неоднозначності, де традиційні методи часто виявляються недостатньо ефективними.

Застосування нечіткого нелінійного беггінгу може включати, але не обмежується такими областями як фінанси, метеорологія, генетика, здоров'я, екологія і багато інших. У всіх цих областях, висока ступінь неоднозначності та складність даних можуть зробити нечіткі нелінійні моделі особливо привабливими для прогнозування і аналізу.

Ключова відмінність нечіткого нелінійного беггінгу від інших методів моделювання полягає в його здатності об'єднати переваги нечіткої логіки, нелінійного моделювання та беггінгу, що дозволяє краще адаптуватися до складності даних та узгоджувати точність та розподільність прогнозів.

Нечітка логіка допомагає моделі легше обробляти неоднозначність та взаємозв'язок між даними. Вона допускає, що дані можуть належати до декількох класів з різною ступінню впевненості, а не просто припускаючи, що вони належать до одного певного класу. Це дозволяє моделі краще враховувати нечіткість та шум в даних та забезпечує більш гнучкі моделі.

Нелінійне моделювання забезпечує більш гнучке і точне пристосування до даних, відображаючи складні взаємозв'язки, які не можуть бути відтворені простими лінійними моделями. Якщо лінійні моделі вимушені апроксимувати такі взаємозв'язки використовуючи просту пряму лінію, нелінійні моделі можуть адаптувати більш складні форми, які точніше відображають реальний світ.

Наостанок, беггінг допомагає моделі краще узагальнювати та зменшувати варіативність прогнозів шляхом створення ансамблю моделей. Кожна модель навчається незалежно на різних підмножинах даних, а потім їх результати об'єднуються в єдиний прогноз, що допомагає робити більш стабільні та точні прогнози.

Очевидне питання в реальній роботі з нечітким нелінійним беггінгом – введення нечіткості і нелінійності в процес беггінгу вимагає високої кваліфікації та глибокого розуміння від дослідника. Нечітка логіка та нелінійне моделювання самі по собі є комплексними дисциплінами, а їх

розуміння та вміння їх застосовувати поряд з беггінгом можуть бути складними для непідготовленого дослідника.

Дослідники мають бути підготовлені до того, що формування нечіткої лінії для подання даних може бути незрозумілим і/або трудомістким процесом, особливо якщо вхідні дані містять багато неоднозначних или непевних вимірювань. Додатково, прискорення складності нелінійних моделей може ускладнити інтерпретацію результатів і утруднити їх узгодження з даними.

Враховуючи пресуючу необхідність адаптації до нестабільних даних і складних структур в сучасному світі, є очевидним, що загалом у сфері Штучного Інтелекту та Машинного Навчання сьогодні існує великий попит на розробку та застосування підходів, які здатні ефективно впоратися з викликами, які ставить перед нами реальний світ, з його нестабільністю, складністю та непередбачуваністю.

Отже, нечіткий нелінійний беггінг забезпечує цінний метод, який забезпечує можливість впоратися із цими викликами, в якому комбінуються сили трьох важливих дисциплін: нечіткої логіки, нелінійного моделювання та беггінгу.

Продовжуючи, простір для досліджень у сфері нечіткого нелінійного беггінгу є масивним. З урахуванням його гнучкості, можна розробляти спеціалізовані адаптації для конкретних застосувань або працювати на розвитку нових методів, які поєднують нечіткі нелінійні моделі з іншими техніками машинного навчання.

Одним з потенційних напрямків дослідження може бути розробка методів автоматичного визначення параметрів нечіткого нелінійного беггінгу. Це може включати в себе розробку методик для визначення оптимальних типів та параметрів нечітких множин, вибір найкращих алгоритмів нелінійного навчання або навіть методів автоматичного налаштування параметрів моделі.

Іншим напрямком може бути розробка нових технік поєднання прогнозів з моделями беггінг. Це може включати в себе розробку складніших схем голосування, використання технік штучного інтелекту для визначення оптимального способу комбінування прогнозів або навіть залучення нових технологій, таких як глибоке навчання.

Крім того, моделі нечіткого нелінійного беггінгу можуть бути адаптовані або поєднані з іншими методами машинного навчання. Це може включати в себе створення ансамблів, які комбінують нечіткі нелінійні моделі з іншими типами моделей, розробку гібридних моделей, які комбінують нечіткі нелінійні моделі з іншими підходами або використання нечіткого нелінійного беггінг прогнозування в рамках більш складних систем машинного навчання.

Враховуючи складність та неймовірну гнучкість нечіткого нелінійного беггінгу, якість його впровадження значною мірою залежить від досвіду та обізнаності дослідника. Освоєння цієї техніки вимагає глибокого розуміння нечіткого моделювання, нелінійних моделей та методів ансамблю, а також здатності приймати та реалізовувати виважені проектні та стратегічні рішення.

Тем не менш, ефективне впровадження нечіткого нелінійного беггінгу може принести значні переваги у широкому спектрі застосувань, включаючи, але не обмежуючись, областями такими як фінанси, здоров'я, екологія та генетика. В цих галузях складність, неоднозначність та нестабільність даних є нормою, а не винятком, а це робить нечіткі нелінійні моделі незамінними у вирішенні їх найскладніших проблем.

В цілому, нечіткий нелінійний беггінг представляє собою захоплюючу нову область досліджень в галузі машинного навчання та штучного інтелекту, яка розвивається з блискавичною швидкістю. Через своє поєднання нечіткості, нелінійності та сили ансамблю, він пропонує інноваційний підхід до рішення найскладніших та найбільш непередбачуваних проблем, з якими ми стикаємося в сучасному світі.

Процес навчання моделей нечіткого нелінійного беггінгу може бути виконаний за допомогою різноманітних алгоритмів. Наприклад, градієнтний спуск, стохастичний градієнтний спуск, методи оптимізації на основі населення, такі як генетичні алгоритми, або мета-генетичні алгоритми. Вибір конкретного алгоритму залежить від специфіки завдання, доступності обчислювальних ресурсів, обсягу та якості навчальних даних.

Окремо важливо наголосити, що нечіткий нелінійний беггінг вимагає обережного підходу до вибору нечітких множин та налаштування їх параметрів, включаючи форму і розміри нечітких множин, а також спосіб об'єднання їх результатів. Неправильний вибір цих параметрів може призвести до перенавчання, недостатнього навчання або нестабільних прогнозів.

Безумовно, прориву в розробці та впровадженні моделей нечіткого нелінійного беггінг сприяє активний розвиток технологій та алгоритмів штучного інтелекту, а також збільшення обчислювальних можливостей сучасних комп'ютерів. Ці тенденції відкривають нові можливості для досліджень та практичного застосування цього інноваційного методу. Водночас, щоб в повній мірі реалізувати потенціал нечіткого нелінійного беггінг, потрібно подальше дослідження його властивостей, оптимізація параметрів та методів налаштування, а також розробка стратегій його інтеграції з іншими методами та системами машинного навчання.

Підводячи підсумки, нечіткий нелінійний беггінг являє собою потужний інструмент машинного навчання, що забезпечує високу гнучкість та ефективність при роботі з неоднозначними або нелінійними структурами даних. Він поєднує переваги нечіткої логіки, нелінійного моделювання та беггінгу, надаючи дослідникам та практикам цінний інструмент для різних прогнозних завдань.

Хоча впровадження нечіткого нелінійного беггінгу вимагає високої кваліфікації та глибокого розуміння методології, його переваги у вирішенні

складних задач прогнозування можуть зробити його визнаним вибором в багатьох галузях науки та промисловості.

Враховуючи швидкий розвиток технологій машинного навчання та збільшення обчислювальних можливостей, можна очікувати, що нечіткий неліній беггінг стане ще більш доступним та широко використовуваним у наступні роки. Проте, для використання цього потенціалу на повну, необхідно зробити ще багато дослідницької роботи для покращення методів та технік впровадження цього підходу. З огляду на його можливості та гнучкість, нечіткий неліній беггінг безсумнівно має багато обіцянок для майбутнього машинного навчання.

2.3 Розробка власної методики імплементації нечіткого нелінійного беггінгу

При розробці власної методики імплементації нечіткого нелінійного беггінга (Fuzzy Nonlinear Bagging, FNB), важливо звернути увагу на ряд ключових аспектів. Сам процес складається з кількох етапів, які включають розробку нечіткої системи, налаштування її параметрів, вибір алгоритму нелінійного навчання, оптимізацію його параметрів та обробку результатів прогнозування.

По-перше, необхідно розробити нечітку систему для представлення і аналізу вхідних даних. Основний елемент цієї системи – нечіткі множини, які не тільки підтримують представлення неоднозначності й невизначеності, а й дозволяють візуалізувати та аналізувати дані в зручній формі. Вид, розмір та форма нечітких множин повинні відбивати характеристики даних, а також специфіку задачі моделювання.

Особливості нечіткої системи в значній мірі залежатимуть від особливостей вхідних даних та конкретної мети прогнозування. Наприклад, при роботі з даними, що розподілені нормально, може бути корисно використовувати гаусові нечіткі множини, які імітують цей розподіл. З

іншого боку, якщо дані розподілені нерівномірно або мають високу кількість викидів, у формуванні нечітких множин можливо використовувати більш складні форми.

Після розробки нечіткої системи слід визначити алгоритм нелінійного навчання, який буде використовуватися для тренування моделі. Тут важливо згадати, що існує значна варіативність в потужності та складності діапазону доступних алгоритмів, і вибір відповідного інструменту потребує системного розуміння ваших даних, завдань і обчислювальних ресурсів.

Варто зауважити, що вибір нелінійного алгоритму навчання повинен відповідати конкретній структурі ваших даних та моделі. Наприклад, регресійні дерева чи випадковий ліс можуть бути відмінним вибором для задач з мінімальними обчислювальними ресурсами, але функціонально складними даними. З іншого боку, градієнтний бустинг або нейронні мережі можуть бути ефективнішими при потужних обчислювальних ресурсах та великому обсязі оброблення даних.

Після розробки шаблону нечіткого моделювання та визначення алгоритму нелінійного навчання, наступним етапом буде вибір та налаштування параметрів моделі та алгоритму навчання. Експертний вибір цих параметрів може бути критичним для ефективності прогнозування моделі. Для цього можна використовувати методи еволюційних алгоритмів, зокрема генетичні алгоритми, які дозволяють оптимізувати декілька параметрів одночасно.

Цей етап також включає перевірку відповідності моделі. Ідея полягає в тому, щоб переконатися, що модель добре працює не тільки на навчальних даних, але й може узагальнювати свої прогнози на нових, невідомих даних. Це може бути досягнуто за допомогою методів перехресної перевірки, таких як *k-fold cross-validation*, або методів випадкового розбиття, таких як *train-test split*.

Останнім етапом процесу розробки моделі FNB є обробка результатів прогнозування. Це включає в себе визначення методу агрегації для

об'єднання прогнозів від різних базових моделей. Зазвичай, це може бути основний арифметичний або геометричний середній, або голосування за більшість, в залежності від конкретної задачі. Також можна застосовувати більш складні методи, основані, наприклад, на вагових коефіцієнтах.

Можна зазначити, що основним завданням на цьому етапі є вибір алгоритму агрегації, який дозволить найбільше висвітлити подібності і відмінності між окремими моделями та врахувати їх вплив на кінцевий результат прогнозування.

При цьому, робота з ансамблями декількох моделей може вимагати додаткових комплексних технік. Наприклад, потрібно вирішити, як обробляти випадки, коли окремі моделі надають суперечливі прогнози, як врахувати міру «впевненості» кожної моделі в її прогнози та як збалансувати прогнози від моделей з високим і низьким рівнем загальної точності.

Усе це викликає необхідність використання складних методів агрегації, заснованих на вагових коефіцієнтах, або навіть розробку власних методів, спеціально адаптованих до специфіки завдання прогнозування і характеру вхідних даних.

Крім цього, на цьому ж етапі необхідно провести тестування та оцінку ефективності прогнозування моделі. І хоча важливо контролювати показники точності моделі, такі як коефіцієнт детермінації R^2 або середньоквадратична помилка (RMSE) для завдань регресії, або матриця невідповідностей, точність, чутливість, специфічність і площа під кривою ROC для завдань класифікації, стільки важливо також розглянути інші аспекти роботи моделі, наприклад, її здатність до інтерпретації та швидкість обчислень.

У той час як перша вимога є особливо важливою в тех областях, де потрібно пояснити причинно-наслідкові зв'язки, лише зрозумівши логіку моделі, можна передбачити її поведінку в реальних ситуаціях і знати, куди вона може привести. Швидкість обчислень також є важливим аспектом,

особливо при роботі з великими наборами даних або в задачах реального часу.

Подібний аналіз допомагає не тільки оцінити, наскільки добре модель працює на даний момент, але і визначити, в якому напрямку потрібно рухатися для подальшого вдосконалення моделі. Більше того, ці дані можуть служити важливою інформацією при презентації моделі потенційним користувачам або при її впровадженні в реальні завдання та проекти.

Беручи до уваги цю аналітику, можливі шляхи для подальшого вдосконалення моделі можуть включати:

- експерименти з різними методами налаштування нечіткої системи, включаючи застосування різних форм та розмірів нечітких множин;
- використання різних алгоритмів навчання та оптимізації;
- підключення додаткових ознак або створення нових производних;
- оптимізація параметрів моделі, наприклад, за допомогою методів машинного навчання для визначення оптимальних значень;
- а також експерименти з різними методами об'єднання прогнозів від окремих моделей.

Дійсно перспективний метод оптимізації параметрів нечіткої системи, який може бути використаний для поліпшення моделі FNB, – застосування еволюційних алгоритмів, таких як генетичні алгоритми, для багатовимірної оптимізації. Ці алгоритми дозволяють точно та ефективно налаштувати широкий діапазон параметрів шляхом імітації процесів природного добору та генетичного успадкування.

Крім того, може бути доцільно використовувати більш просунуті алгоритми машинного навчання, такі як нейронні мережі, для виконання самого нелінійного навчання. Нейронні мережі досить добре впорюються з моделюванням складних, високовимірних даних та можуть забезпечити значне поліпшення продуктивності, особливо при наявності великої кількості навчальних даних та потужних обчислювальних ресурсів.

З ростом обсягів даних та потужності обчислювальної техніки, можливості використання більш складних та потужних методів, таких як нейронні мережі, стають все більш доступними для дослідників та практиків. Це відкриває нові можливості для покращення стабільності та точності прогнозування при використанні методу FNB, особливо в світлі його здатності до інтеграції інформації з декількох джерел та врахування невизначеності та нечіткості даних.

Однак, варто зазначити, що ефективне використання цих передових методів та технологій вимагає глибокого розуміння їх основних принципів, обмежень та підводних каменів, а також інтенсивної та уважної роботи з налаштуванням їх параметрів. Це включає вибір відповідного алгоритму навчання, багатовимірну оптимізацію його параметрів, визначення оптимальної кількості джерел даних та балансу між точністю та швидкістю прогнозування, а також моніторинг та оцінку його роботи на різних етапах процесу.

З урахуванням цих аспектів, їх використання може бути досить складним та трудомістким, особливо для нових дослідників або тих, хто працює з обмеженими обчислювальними ресурсами. Відповідно, важливо дотримуватися обережного та систематичного підходу, ретельно досліджуючи кожен крок та варіант, і завжди перевіряти продуктивність моделей за допомогою відповідних метрик та процедур тестування, щоб гарантувати, що вони приносять значні та стійкі поліпшення для актуальних завдань прогнозування.

До важливих аспектів роботи з методом FNB, які створюють необхідність для більш глибоких досліджень та розробок, також в стоїть згадки про інтерпретованість та прозорість в роботі цього алгоритму.

Хоча використання нечіткості координат та нелінійних моделей, таких як нейронні мережі, може покращити точність моделі, такий підхід може також створити виклики для розуміння того, як сама модель приймає рішення.

Інтерпретованість алгоритму вважається важливою предметною областю в машинному навчанні, оскільки вона вимагає, щоб результати моделі були зрозумілі та доступні для інтерпретації. Крім важливості для етики та юридичної відповідальності, інтерпретована модель може стати конкретною перевагою у розробці нових продуктів та послуг, оскільки вона дозволяє краще зрозуміти процес прийняття рішень та використання інших можливостей.

Підходи з фокусом на використанні та розробці інструментів, методів та технік для поліпшення інтерпретованості моделей, стають все більш популярними в галузі машинного навчання. Такі методи включають в себе використання білого листа, підходи з використанням важливості ознак, часткових графіків залежності та інші.

На наступному етапі при розробці методики імплементації моделі FNB важливим є розгляд вимог до обчислювальних ресурсів і ефективності обчислень, а також їх відповідність обом вимогам. Реалізація алгоритмів нейронних мереж та нелінійного бегінгу може вимагати значних обчислювальних ресурсів та часу обчислень, що може увімкнути додаткові вимоги до обладнання та операційних систем.

Тому перспективним напрямком роботи в цьому напрямку може включати оптимізацію алгоритму, розробку ефективних стратегій паралелізації або використання спеціалізованих обчислювальних платформ та пристроїв, таких як GPU або TPU. Крім того, ті самі виклики можуть виникнути при візуалізації або аналізі результатів, що може вимагати використання таких методів аналізу, як принципові компоненти аналізу або t-SNE для ефективного представлення високовимірних даних.

Загалом, при розробці власної методики імплементації нечіткого нелінійного бегінгу важливо ретельно опрацювати кожен аспект процедури, від детального вивчення та розуміння вхідних даних до налаштування параметрів моделі та алгоритмів навчання, а також правильної обробки та аналізу результатів.

Досі ми багато говорили про технічні характеристики та аспекти процесу моделювання, але ми також повинні зазначити важливість розуміння проблеми, з якою ми маємо справу, та контексту, в якому вона виникла.

Програмне забезпечення, включаючи алгоритмічні моделі, які ми розробляємо, завжди слід розглядати у контексті використання. Незалежно від того, наскільки передовою та технічно складною є наша модель, вона буде корисною лише тоді, коли вона вирішує реальну бізнес-задачу або допомагає прийняти краще обґрунтовані рішення.

Враховання контексту допоможе нам краще зрозуміти, як використовувати нашу модель, а також допоможе визначити бізнес-метрики процесу моделювання.

Наприклад, якщо ми створюємо модель для прогнозування цін на акції, ми повинні враховувати не тільки історичні дані про ціни на акції і загальну економічну ситуацію, але й такі аспекти, як торгівельна стратегія компанії, її фінансова стабільність, тренди ринку і т.д.

Крім того, не забуваймо про важливість влучності прогнозів. Модель, яка дає точні прогнози для майбутнього, але робить це занадто повільно, може бути непридатною для реального часу. З іншого боку, модель, яка робить швидкі прогнози, але з великою похибкою, може привести до втрати грошей або до інших негативних наслідків.

Оскільки у нас на руках є висока складність і багато можливих варіантів, створення моделі FNB не має бути спонтанним процесом. Потрібно чітко визначити процес, включаючи дослідження, планування, розробку, виконання та оцінювання. Особливо важливо – аналіз та інтерпретація результатів, що допомагає виявити можливі проблеми, оцінити ефективність моделі та визначити напрямки для подальшого вдосконалення.

У цьому процесі можуть бути використані різні інструменти та технології, включаючи спеціалізовані бібліотеки та фреймворки для

машинного навчання, великі дані та обчислювальні платформи, інструменти візуалізації даних та аналізу результатів та ін.

Однак незалежно від вибраних методів та інструментів, ключем до успіху є глибоке розуміння завдання, деталізована проробка кожного етапу процесу та постійна сфокусованість на досягненні кінцевої мети.

Конкретними кроками на шляху до цієї мети можуть бути проведення глибокого аналізу вхідних даних, включаючи оцінювання їх якості, повноти і відповідності задачі; налаштування та оцінка різних моделей, вивчення їх недоліків та переваг; аналіз та інтерпретація результатів прогнозування; а також пошук способів для подальшого покращення та оптимізації.

Важливою складовою цього процесу є також нові ідеї та напрямки досліджень, які б могли розвивати та вдосконалювати методику нечіткого нелінійного беггінгу. Це може включати в себе розробку нових алгоритмів, методів та технік; застосування нових форматів та джерел даних; або розробку нових продуктів та послуг, які були би засновані на використанні цієї методології.

Проте важливо не забувати, що жодна модель або методологія не є вічною. Світ навколо нас постійно змінюється, і завжди з'являються нові виклики і можливості. Тому для успіху важливою є здатність швидко реагувати на зміни, постійно вивчати та оновлювати свої знання та інструменти, а також мати велику гнучкість та креативність для пошуку нових рішень та підходів.

І наприкінці, вважається дуже важливим вести тісну взаємодію з ключовими зацікавленими сторонами і користувачами – людьми, які будуть використовувати та отримувати користь від ваших моделей. Діалог та обмін думками із ними можуть допомогти не тільки краще зрозуміти їхні потреби та вимоги, але і знайти нові натхнення та розвивати якісніше програмне забезпечення.

Виконуючи все це з урахуванням вказаних вище обмежень, ви зможете використовувати методику FNB для вирішення складних задач

прогнозування. І розробка такої моделі буде важкою та потребуватиме значних зусиль, але, водночас, це є справою, здатною принести велику користь.

2.4 Інноваційні підходи до нелінійного бегінгу для покращення ефективності прогнозування

Інноваційними підходами до нелінійного бегінгу можна вважати розвиток нових алгоритмів і технологій, застосування нових форматів і джерел даних, а також впровадження нових продуктів та послуг на основі цієї технології. Вони можуть стосуватися різних аспектів процесу, включаючи збір та обробку даних, налаштування моделі, виконання досліджень і тестування, аналіз та інтерпретацію результатів, а також розвиток та впровадження програмного забезпечення та систем.

Розвиток нових алгоритмів та технологій.

Основна їх мета – покращення точності і стабільності прогнозування, а також збільшення продуктивності і ефективності роботи. Їх може бути розроблено і впроваджено шляхом поєднання та інтеграції різних методів та підходів, таких як машинне навчання, великі дані, нейронні мережі, нечіткі логіки, оптимізації, статистики та аналітики та ін. Вони можуть також враховувати нові аспекти та застосування, такі як автоматизоване та адаптивне навчання, на основі різних моделей і алгоритмів, включаючи глибоке навчання, адаптивні і еволюційні алгоритми і т.д.

Застосування нових форматів та джерел даних.

Як невичерпне джерело інформації про різні змінні та їх взаємозв'язки, даними може бути використано для навчання і прогнозування моделей, що можуть включати великі дані, сенсорні дані, уявлення, аудіо, текст, мову, соціальну медіа та ін.

Вибір даних може значно вплинути на якість і ефективність прогнозування, і він може стати предметом значних досліджень та аналізу.

Це може включати в себе дослідження різних джерел даних, забезпечення їх якості та повноти, а також визначення найбільш відповідних призначень та методів обробки.

Створення нових продуктів та послуг на основі новітніх технологій.

Це може включати вдосконалення існуючих продуктів та послуг, а також розробка та впровадження нових. Вони можуть бути зосереджені на вирішенні конкретних проблем або задоволенні конкретних потреб, а також використання нових можливостей, що відкриваються в результаті технологічного прогресу.

Ці продукти і послуги можуть бути створені для різних галузей та ринків, включаючи фінанси, медицину, освіту, науку, виробництво, логістику, розваги та ін. Вони можуть також забезпечити різні форми доставки та розповсюдження, включаючи веб-служби, програмне забезпечення, апаратне забезпечення, мобільні та вбудовані системи, обlačні та гібридні рішення та ін.

Інтеграція з іншими системами та технологіями.

Інновації в області нелінійного бегінгу не обмежуються тільки створенням самостійної моделі або алгоритму. Це може також означати інтеграцію моделі з іншими системами, процесами та технологіями. Наприклад, модель може бути вбудована в більш нелінійний бегінгвартісну систему або серію сервісів. Це дозволяє моделі використовувати можливості цих систем для підвищення своєї ефективності.

На практиці, однією з важливих засад у такому типі інтеграції є наявність загальних стандартів та протоколів для обміну даними та інформацією. Однак, у деяких випадках, може бути необхідним розробити специфічні для конкретного застосування адаптери або інтерфейси.

Креативність та інновації.

Одним з ключових факторів успіху в будь-якій сфері, включаючи нелінійний бегінг, є креативність та інновації. Це включає в себе здатність

дивитися на задачі знову, шукати нові підходи та вирішення, та навчатися від своїх помилок.

У контексті нечіткого нелінійного беггінгу, це може означати використання нових алгоритмів або технік, створення нових варіативів вже відомих моделей, або пошук нових способів для використання наявної інформації.

Усвідомлений вибір моделі.

Не всі моделі є однаково підходящими для всіх задач. Різні моделі мають свої сильні та слабкі сторони, і фахівець повинен виробити вміння вибирати найбільш відповідну модель для конкретної задачі.

Це означає, що крім вміння програмувати моделі та працювати з даними, спеціалісти в області нечіткого нелінійного беггінгу мають розуміти застосування та обмеження різних моделей, та в змозі обґрунтувати свій вибір моделі.

Використання автоматизованих систем для вибору моделі.

Одним зі значних напрямків у сфері нелінійного беггінгу є використання автоматизованих систем для вибору моделі, які можуть допомогти значно зменшити час та зусилля, що потрібні для дослідження та вибору оптимальної моделі.

Ці системи можуть використовувати технології штучного інтелекту, включаючи машинне навчання, для дослідження великого простору можливих моделей та конфігурацій, але одночасно здійснивши оцінку їх підходження і ефективності.

Використання цих систем може бути особливо корисним у складних або великомасштабних задачах прогнозування, де ручний вибір та налаштування моделі може бути дуже трудомістким і дорогостоящим.

Впровадження оптимальних стратегій навчання.

Крім самого вибору моделі, важливим компонентом процесу моделювання є вибір оптимальної стратегії навчання.

Стратегія навчання визначає спосіб, яким модель «вчиться» з даних, тобто поступово змінює свої параметри для наближення до найкращого припущення. Вона може бути заснована на різних методах і техніках, залежно від природи вхідних даних, механізму моделі та конкретної задачі прогнозування.

Ефективне використання обчислювальних ресурсів.

Усе частіше процеси навчання моделей розраховані на великі обчислювальні ресурси. Великі набори даних і складні алгоритми можуть потребувати значних обчислювальних ресурсів і часу для виконання.

Інноваційні підходи мають включати в себе ефективне використання наявних обчислювальних ресурсів, через отримання найкращих можливих результатів за найменший час. Це може включати в себе паралельні та розподілені виконання, використання спеціалізованого обладнання, такого як графічний процесорів або TPU (Tensor Processing Units), та впровадження оптимізованих алгоритмів та технік.

Використання глибокого навчання для нелінійного беггінгу.

Глибоке навчання це субполе машинного навчання, що використовує неймережеві архітектури. Ці моделі тренуються, моделюючи велику кількість вхідних даних.

Їх особлива властивість – вміння вивчати внутрішні структури даних, що надається їм під час тренування. Завдяки цьому властивості, глибокі мережі дуже ефективні в задачах розпізнавання об'єктів, регресії, кластеризації та інших завдань прогнозування.

Зі створенням нових алгоритмів для нелінійного беггінгу виникає потреба у нестандартних методах валідації моделей. Цей елемент циклу моделювання також вимагає інноваційних методів і підходів, бо належна валідація моделі важлива для визначення її відповідності використовуваним даним, її життєздатності та можливості використання у практичному аспекті.

Реальні випадки використання вимагають ефективних механізмів, які допомагають моделям адаптуватися до змін у даних чи внутрішнього серверного оточення. З перебігом часу деякі моделі можуть втрачати свою точність при прогнозуванні, так як реальний світ постійно змінюється. В свою чергу, це означає, що змінні, які стали важливими на момент створення моделі, можуть втратити актуальність внаслідок динамічних змін в навколишньому середовищі. Саме через це, є необхідністю швидко та ефективно вносити зміни в модель, аби навчити її працювати з новими даними.

У той же час, всі оновлені моделі мають бути досконало протестовані ще раз для перевірки точності їх прогнозів. Це допоможе уникнути помилок та спотворень при використанні цих моделей в реальних сценаріях. Це ставить низку вимог до інфраструктури та робочого навантаження команди зі створення моделей, а також потребує високого рівня гнучкості та швидкості.

Інновації в області нелінійного беггінгу включають використання найновіших досягнень теоретичних наук, нових технологій в області AI та передового досвіду, зібраного у розкладці AI та навколишніх дисциплінах. Врахування підходів, що виявилися надійними на практиці, дозволяє використовувати їх у нелінійному беггінгу та створювати моделі прогнозування, які мають більш високу якість прогнозування та можуть бути впроваджені у готових продуктах або сервісах.

Цикл моделювання може включати різні інноваційні підходи, які слід враховувати при побудові оптимальних нелінійних моделей. Такі підходи дозволяють виконувати процедуру моделювання більш ефективно та отримувати кращий результат.

Останнім часом було виявлено ряд напрямків, що є потенційно важливими для подальшого розвитку та удосконалення механізмів нечіткого нелінійного беггінгу. Впровадження цих нових підходів, які постійно розвиваються та удосконалюються завдяки виконанню нових

досліджень, може значно покращити якість та ефективність різних стадій циклу моделювання нечіткого нелінійного беггінгу.

Передовий досвід показує, що машинне навчання і інтелектуальний аналіз даних допомагають виявляти нові підходи та рішення для прогнозування шляхом нечіткого беггінгу. Таким чином, це може допомогти визначити найефективнішу та найактуальнішу модель для визначення тенденцій у великій кількості даних. У свою чергу, це може створити нові можливості для реальних застосувань цих технологій, зокрема, у фінансовому моделюванні, маркетингу, логістиці, промисловості, виробництві, здоров'ї, соціальних мережах, енергетиці, транспорту та будь-якій іншій області, де використовуються великі обсяги даних.

Однією з ключових вимог до інноваційних підходів у нечіткому нелінійному беггінгу є використання передових технічних можливостей та інфраструктурних ресурсів нового покоління, а саме спеціалізованих процесорів та обчислювальних систем, оптимізованих бібліотек та інструментів програмування, ефективних методів обробки даних та архітектур системи, сучасних джерел та форматів даних, а також новітніх методів та алгоритмів аналізу даних, машинного навчання, статистичного моделювання, оптимізації та валідації.

При всьому цьому, важливе значення має гармонійне поєднання різних фаз та компонентів процесу моделювання, що дозволяє створити цілісну, послідовну та гнучку систему, здатну адаптуватися до різних вимог, сценаріїв та умов роботи. У цьому контексті, одним из найбільш проміментних прикладів може служити беггінг, який представляє собою використання різних моделей прогнозування в одному процесі, що дозволяє створювати більш ефективні та надійні прогнози.

Важливий аспект інноваційних підходів до нелінійного беггінгу полягає в створенні нових продуктів та послуг на основі цих технологій. В результаті, це може створювати нові ринкові можливості для компаній,

підприємців, науковців, інженерів, розробників, аналітиків, дослідників та інших учасників цього процесу.

Ці продукти та послуги можуть включати в себе програми для обробки та аналізу даних, системи для автоматизованого моделювання та прогнозування, інтерфейси для візуалізації та інтерпретації результатів, рішення для управління знаннями та обміну даними, сервіси для обробки та передачі даних, інструменти для моніторингу, навчання, оптимізації та валідації моделей, додатки для мобільних пристроїв, веб-сервіси, хмарні платформи та ін.

Однак, важливими вимогами в цьому процесі є дотримання стандартів якості та ефективності, вирішення проблем безпеки та конфіденційності, врахування юридичних та етичних аспектів, забезпечення достовірності та повноти даних, валідації результатів, підтримки прозорості та контролю над процесами, а також впровадження управлінських, організаційних та технічних засобів для створення надійної, безпечної, зручної, універсальної та масштабованої інфраструктури.

Ці підходи та технології можуть бути використані в будь-якій галузі та діяльності, де необхідно перетворювати великі обсяги сирих, неструктурованих або масивних даних на корисну, зрозумілу, доступну, а головне, цінну інформацію для прийняття рішень, планування, управління, розробки, дослідження, аналізу. Дозволяючи використовувати ці дані для вирішення конкретних задач, вони дозволяють відкривати нові можливості для відкриття нових знань, створення нових цінностей, покращення діяльності, забезпечення конкурентних переваг, а також поліпшення якості життя, охорони здоров'я, навколишнього середовища та суспільства в цілому.

Застосування інноваційних підходів у нечіткому беггінгу вимагає не тільки технічних знань та навичок, але й глибокого розуміння бізнесу, домену, сценаріїв використання, потреб користувачів, а також викликів і

обмежень, пов'язаних з використанням даних, технологій, процесів, ресурсів, структур та інших аспектів підприємства.

Фахівці в галузі AI і машинного навчання повинні володіти високим рівнем кваліфікації та креативності, ефективно використовувати доступні інструменти, методи та ресурси, навчатися, адаптуватися та інновувати, а також співпрацювати, спілкуватися та взаємодіяти з різними зацікавленими сторонами, включаючи керівників, співробітників, клієнтів, партнерів, постачальників, регулюючі органи, громадськість і т.п.

Новітні практики моделювання нечіткого беггінгу ставлять певні вимоги до архітектури, дизайну, механізмів, інтерфейсів та процесів, пов'язаних з обробкою, зберіганням, передачею, доступом, використанням, захистом, візуалізацією, управлінням та аналізом даних.

Крім того, дослідження в галузі AI вимагає постійного моніторингу, перегляду, оновлення та оптимізації моделей, алгоритмів, методів, конфігурацій, параметрів, стратегій, метрик, результатів, висновків, рекомендацій та інших елементів процесу моделювання.

Власне кажучи, спеціалісти у галузі AI мають постійно вчитися, експериментувати, узагальнювати, адаптуватися та інновувати, адже технології, інструменти, методи, джерела даних, алгоритми, моделі, застосування та вимоги швидко та постійно змінюються.

Наразі багато інноваційних підходів в нелінійному беггінгу є на передньому краю технічних, теоретичних та практичних досліджень і застосувань у галузі AI. Досвід виявляє, що використання більш розширеного спектру методів, технік, технологій та передового досвіду може виявитися дуже корисним для виявлення складних взаємозв'язків та шаблонів у великих, складних, неструктурованих або динамічних наборах даних і використання цієї інформації для підтримки ефективного прийняття рішень.

Таким чином, впровадження інноваційних підходів, методів та технологій у нечіткому беггінгу може відкрити нові перспективи для

розвитку науки, техніки, бізнесу, промисловості, суспільства та держави в цілому.

Підсумовуючи, інноваційні підходи у нечіткому нелінійному беггінгу відкривають нові можливості для підвищення точності прогнозування та ефективності обробки даних. Процес не просто використовує нові методи та технології, але стає більш адаптивним і вміє реагувати на зміни у вхідних даних.

Використовуючи глибоке навчання та вчасні алгоритми, дослідники та розробники можуть створювати моделі, які є здатними до самонавчання і адаптації до викликів реального світу. Це не тільки збільшує продуктивність та ефективність, але й відкриває нові можливості для створення інтелектуальних систем на основі AI, покращуючи можливості перетворення даних в цінну інформацію для кращого вирішення проблем.

Особлива увага приділяється використанню обчислювальних ресурсів, що забезпечують виконання великих наборів даних та складних алгоритмів швидко та ефективно. Нові рішення використовують спеціалізоване обладнання, оптимізовані бібліотеки та інші інструменти для отримання максимального користувача.

Нелінійний беггінг використовує глибокі нейромережі для створення ефективних моделей прогнозування. Він може виявляти складні зв'язки та шаблони в даних, роблячи його надзвичайно корисним для прогнозування у великих і складних об'ємів даних.

У цілому, інновації в нечіткому нелінійному беггінгу можуть значно підвищити потужність, точність, надійність, гнучкість, швидкість, масштабованість, доступність, ефективність, безпеку, прозорість, відповідність і цінність систем обробки даних та прогнозування.

3 ПРАКТИЧНЕ ЗАСТОСУВАННЯ НЕЧІТКОГО НЕЛІЙНОГО БЕГГІНГУ

3.1 Опис програмних засобів

В рамках даної роботи основною мовою програмування за використанням якої проводиться розробка та імплементація створення моделі є Python. Python відомий своєю простотою у засвоєнні, гнучкістю та потужними бібліотеками та модулями для аналізу даних, машинного навчання та наукових розрахунків, що робить його особливо придатним для задач такого роду.

Одним з ключових інструментів для роботи з даними в Python є бібліотека pandas. Pandas надає широкий спектр функціональності для створення, маніпулювання, трансформації та аналізу структурованих та неструктурованих даних в Python. Вона також дозволяє легко зчитувати різноманітні формати даних, включаючи CSV, Excel, SQL та багато інших.

Далі, numpy – це одна з основних бібліотек для наукових обчислень в Python, що надає підтримку для великих багатовимірних масивів та матриць, разом з великою бібліотекою високорівневих (та швидких) математичних функцій для операцій з цими масивами. Це ключовий інструмент для роботи з числовими даними та виконання математичних операцій, що є основою для машинного навчання.

Бібліотека matplotlib – це сучасна графічна бібліотека, що дозволяє проводити обширну візуалізацію даних, включаючи створення графіків, гістограм, діаграм розсіювання, теплових карт, контурних діаграм та багатьох інших типів візуалізації. Вона пропонує детальний контроль над всіма аспектами створення графіків та діаграм для досягнення кращого розуміння залежностей в робочих даних.

Окрім цього, ще однією бібліотекою, що використовується у цій дипломній роботі, є Scikit-learn – потужна, гнучка та легка у використанні

бібліотека машинного навчання для Python. Вона включає набір простих та ефективних інструментів для аналізу даних та машинного навчання, що дозволяє здійснювати класифікацію, регресію, кластеризацію та редукцію розмірності. Бібліотека також надає модулі для обробки признаков, вибору моделі, перекресної валідації, оцінки та багато іншого.

Бібліотека Keras буде використовуватися для побудови, тренування та тестування нейронних мереж. Вона пропонує єдиний, зручний API для різноманітних бекендів, зокрема TensorFlow і Theano. Keras допомагає швидко експериментувати з глибокими архітектурами навчання, при цьому її проста, модульна та зрозуміла структура забезпечує зручність в розробці та налаштуванні моделей.

Бібліотека TensorFlow, розроблена Google Brain Team, є ще одним ключовим інструментом для створення та тренування нейронних мереж. Особливість TensorFlow полягає в підтримці розподілених обчислень, що дозволяє ефективно використовувати ресурси, що складаються з декількох машин і устроїв (наприклад, GPU), а також в гнучкості щодо створення складних моделей.

Scipy – бібліотека для наукових обчислень, використовується для числової оптимізації, лінійної алгебри, інтеграції, інтерполяції, спеціальних функцій і інших завдань в області науки та інженерії. Вона буде використовуватись для оптимізації функцій в рамках навчання моделей і обрахунку важливих статистичних метрик.

Також, в даній роботі ми використовуємо ряд додаткових бібліотек, таких як Seaborn (для візуалізації даних і створення виразних статистичних графіків), Statsmodels (для оцінювання статистичних моделей, виконання статистичних тестів та статистичних експлораторних даних), Plotly (для інтерактивних графіків), та багато інших.

Алгоритми та методології, застосовані у даній роботі, працюють на основі вищезгаданих бібліотек, всі вони відіграють важливу роль в розробці нашого проекту. При цьому, Python є основною мовою програмування, яка

об'єднує всі ці інструменти і забезпечує гладке та ефективне розроблення моделі.

При розробці скриптів пов'язаних з даними, було використано бібліотеку `os`, яка дозволяє взаємодіяти з операційною системою та файловою системою, для здійснення операцій читання та запису файлів. Це надає можливість автоматизувати процеси завантаження даних для подальшого аналізу.

Крім того, було використано бібліотеку `time` для трекінгу часу виконання певних частин коду. Наприклад, трекінг часу навчання моделей може допомогти в оцінюванні їх ефективності і може бути корисним при порівнянні різних моделей або налаштувань моделі.

Також важливо відзначити використання `Jupyter Notebook` – веб-засобу інтерактивного програмування, який значно спрощує написання, тестування, налагодження та виконання `Python` коду. Робочий процес в `Jupyter Notebook` базується на понятті зошитів – документів, що містять як код, так і елементи `rich text` (параграфи, рівняння, графіки), що робить його особливо зручним для ітеративної роботи над задачами аналізу даних та машинного навчання.

Було також використано бібліотеку `warnings` для управління попередженнями. Це може бути корисним для приховування непотрібних або великої кількості попереджень, що можуть з'являтися при роботі з різними бібліотеками чи функціями.

Дана робота об'єднала різні бібліотеки та інструменти, а `Python` як мова програмування надає гнучкість та ефективність розробці та експериментації з функціональністю і алгоритмами машинного навчання. Ці бібліотеки дуже різноманітні, але вони усі працюють разом, щоб дозволити нам виконувати складні задачі з прогнозування та оптимізації.

Використані програмні засоби та інструменти, включаючи ряд спеціалізованих бібліотек для обробки даних, машинного навчання та наукових обчислень, полягали в основі розробки цієї роботи. Вони надали

необхідні інструменти та функціональність для створення, навчання та тестування нечетких бегінг-моделей, оптимізації алгоритмів, управління даними та подальшого аналізу результатів.

3.2 Опис всіх наборів даних, що застосовуються в роботі, їх дослідницький аналіз і попередня обробка

У цій роботі використовується декілька наборів даних, перший з яких містить інформацію про наукові журнали – «Scientific JOURNALS Indicators & Info – SCImagoJR».

SCImago Journal & Country Rank – це загальнодоступний портал, який включає журнали та наукові показники країн, розроблені на основі інформації, що міститься в базі даних Scopus (Elsevier B.V.). Ці показники можна використовувати для оцінки та аналізу наукових областей. Журнали можуть порівнюватися або аналізуватися окремо. Набір даних включає 29,165 записів, кожен з яких представляє різний журнал. На рисунку 3.1 показано не форматований приклад набору даних журналів.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1	Rank	Sourceid	Title	Type	Issn	SJR	SJR Best Q	H index	Total Docs. (2023)	Total Docs.	Total Refs.	Total Cites (3years)	Citable Docs. (By Cites / Doc. (2year Ref. / Doc.	Country	Region	Publisher	Coverage	Categories	Areas		
2	1	28773	Ca-A	Journal	15424863	106.094	Q1	211	49	124	4844	35427	89	381.89	98.86	United Sta	Northern / Wiley-Blac	1950-2023	Hematolo	Medicine	
3	2	1.93E+10	Foun	Journal	19358245	37.044	Q1	39	3	13	897	955	13	100.11	299	United Sta	Northern / Now Publi	2008-2023	Artificial Ir	Computer	
4	3	20315	Natu	Journal	14710072	35.91	Q1	508	123	336	11462	13599	153	34.5	93.19	United Kin	Western E	Nature Pul	2000-2023	Cell Biolog	Biochemist
5	4	29431	Quar	Journal	00335533	30.448	Q1	306	47	136	3645	2240	136	11.14	77.55	United Kin	Western E	Oxford Un	1886-2023	Economics	Economics
6	5	12464	Natu	Journal	1474175X	26.837	Q1	505	105	304	10805	10951	163	31.23	102.9	United Kin	Western E	Nature Pul	2001-2023	Cancer Re	Biochemist
7	6	18434	Cell	Journal	00928674	24.342	Q1	892	439	1496	32820	53447	1207	31.3	74.76	United Sta	Northern / Elsevier B.	1974-2023	Biochemis	Biochemist	
8	7	20425	Natu	Journal	14741784	22.399	Q1	391	239	731	8584	13091	153	19.72	35.92	United Kin	Western E	Nature Pul	2002-2023	Drug Disc	Medicine; I
9	8	22697	Amer	Journal	19447981	22.344	Q1	359	95	353	6242	4811	351	10.33	65.71	United Sta	Northern / American	1973-1975	Economics	Economics	
10	9	2.11E+10	Natu	Journal	20588437	21.836	Q1	184	117	335	8842	13775	196	31.17	75.57	United Kin	Western E	Nature Pul	2016-2023	Biomateri	Energy; Ma
11	10	1.77E+10	Natu	Journal	17594782	21.048	Q1	217	127	400	9888	10807	183	28.36	77.86	United Sta	Northern / Wiley-Blac	1946-2023	Accountin	Business, N	
12	11	15847	New	Journal	00284793	20.544	Q1	1184	1388	4522	14603	107246	1824	21.69	10.52	United Sta	Northern / Massachu	1945-2023	Medicine (Medicine	
13	12	17500	Journ	Journal	00221082	19.139	Q1	352	83	227	5043	1938	221	7	60.76	United Kin	Western E	Wiley-Blac	1946-2023	Accountin	Business, N
14	13	15819	Natu	Journal	1546170X	19.045	Q1	630	595	1363	16478	36243	729	27.23	27.69	United Kin	Western E	Nature Re	1995-2023	Biochemis	Biochemist
15	14	1.16E+10	Curre	Journal	19343396	18.663	Q1	71	0	19	0	963	19	0	0	United Sta	Northern / John Wiley	2002-2020	Biochemis	Biochemist	
16	15	2.11E+10	MMV	Journal	23808942	18.587	Q1	23	11	16	0	802	16	47.57	0	United Sta	Northern / Epidemiol	2016, 2021	Medicine (Medicine	
17	16	24404	Journ	Journal	00223808	18.53	Q1	215	83	261	4493	2531	258	7.04	54.13	United Sta	Northern / University	1969, 1971	Economics	Economics	
18	17	12106	Natu	Journal	14764687	18.509	Q1	1391	3770	8037	74917	160102	3840	19.4	19.87	United Kin	Western E	Nature Re	1869-2023	Multidiscip	Multidiscip
19	18	16115	Natu	Journal	15461696	18.117	Q1	511	485	1066	13393	17008	461	13.24	27.61	United Kin	Western E	Nature Pul	1989-2023	Applied Mi	Biochemist
20	19	23340	Chem	Journal	15206890	17.828	Q1	833	271	851	115878	50519	819	49.76	427.59	United Sta	Northern / American	1924-2023	Chemistry	Chemistry	
21	20	10482	Econ	Journal	00129682	17.701	Q1	223	75	273	3371	1946	268	6.24	44.95	United Kin	Western E	Wiley-Blac	1974, 1977	Economics	Economics

Рисунок 3.1 – Неформатований набір даних

Кожен запис містить наступні поля:

– Rank – ранг журналу вгору;

– SJR – науковий ранг журналу (SCImago Journal Rank);

- H index – індекс Нірша, що відображає наукову продуктивність та вплив журналу;
- Total Docs. (2023) – кількість документів, опублікованих журналом у 2023 році;
- Total Docs. (3years) – кількість документів, опублікованих журналом за останні 3 роки;
- Total Refs. – загальна кількість посилань на журнал;
- Total Cites (3years) – загальна кількість посилань на журнал за останні 3 роки;
- Citable Docs. (3years) – кількість цитованих документів за останні 3 роки;
- Cites / Doc. (2years) – середня кількість цитат на документ за останні 2 роки;
- Ref. / Doc. – середня кількість посилань на документ.

У цій роботі ми будемо використовувати ці дані для прогнозування SJR та H index за допомогою нечіткого нелінійного беггінгу на основі адаптивної метамоделі.

Спочатку був проведений експлоративний аналіз даних для оцінки якості та розподілу даних. Це включає в себе перевірку відсутніх значень, викидів та наближення розподілу кожної змінної. Розподіл кожної змінної був вивчений за допомогою гістограм та діаграм розмаху.

Під час попередньої обробки даних було здійснено декілька кроків, щоб покращити якість даних. Наприклад, були вилучені записи з відсутніми значеннями. Крім того, викиди були пом'якшені, оскільки вони можуть шкодити точності прогнозування.

Відповідно до мети нашого дослідження, яка полягає в прогнозуванні показників SJR та H index за допомогою нечіткого нелінійного беггінгу на основі адаптивної метамоделі, деякі атрибути в наборі даних можуть не мати безпосередньої цінності.

Стовбці «Sourceid», «Title», «Issn», «SJR Best Quartile», «Type», «Country», «Region», «Publisher», «Coverage», «Categories», «Areas» головним чином служать для ідентифікації журналу, містять широкий спектр інформації, яка не прямо відображає числові показники, що мають прямий вплив на SJR або H index.

Ці атрибути були видалені з набору даних під час попередньої обробки, щоб сконцентрувати даних на тих змінних, які найбільше впливають на прогнозування SJR та H index (рисунок 3.2). Це було зроблено з метою надання моделі найбільш релевантної та корисної інформації, що відповідає меті дослідження.

	A	B	C	D	E	F	G	H	I	J
1	Rank	SJR	H index	Total Docs. (2023)	Total Docs. (3years)	Total Refs.	Total Cites (3years)	Citable Docs. (3years)	Cites / Doc. (2years)	Ref. / Doc.
2	1	106.094	211	49	124	4844	35427	89	381.89	98.86
3	2	37.044	39	3	13	897	955	13	100.11	299
4	3	35.91	508	123	336	11462	13599	153	34.5	93.19
5	4	30.448	306	47	136	3645	2240	136	11.14	77.55
6	5	26.837	505	105	304	10805	10951	163	31.23	102.9
7	6	24.342	892	439	1496	32820	53447	1207	31.3	74.76
8	7	22.399	391	239	731	8584	13091	153	19.72	35.92
9	8	22.344	359	95	353	6242	4811	351	10.33	65.71
10	9	21.836	184	117	335	8842	13775	196	31.17	75.57
11	10	21.048	217	127	400	9888	10807	183	28.36	77.86
12	11	20.544	1184	1388	4522	14603	107246	1824	21.69	10.52
13	12	19.139	352	83	227	5043	1938	221	7	60.76
14	13	19.045	630	595	1363	16478	36243	729	27.23	27.69
15	14	18.663	71	0	19	0	963	19	0	0
16	15	18.587	23	11	16	0	802	16	47.57	0
17	16	18.53	215	83	261	4493	2531	258	7.04	54.13
18	17	18.509	1391	3770	8037	74917	160102	3840	19.4	19.87
19	18	18.117	511	485	1066	13393	17008	461	13.24	27.61
20	19	17.828	833	271	851	115878	50519	819	49.76	427.59
21	20	17.701	223	75	273	3371	1946	268	6.24	44.95
22	21	17.654	234	108	410	6448	4495	409	8.04	59.7
23	22	17.507	398	178	590	11546	12604	360	19.83	64.87
24	23	17.497	229	229	609	6629	16808	379	26.18	28.95

Рисунок 3.2 – Відформатований набір даних

Крім обмеження розміру навчального набору даних, видалення цих стовбців допоможе зменшити шум та модель буде менш схильною до перенавчання.

У результаті такої попередньої обробки набору даних ми зосереджуємося на стовбцях, які найпряміше впливають на показники SJR та H index, що є основою для створення більш точних та ефективних моделей прогнозування.

Отже, після дослідницького аналізу та попередньої обробки даних, набори даних стали більш придатними.

Теж саме робилось для інших наборів даних, таких як:

– Car Price Prediction Dataset – прогнозування ціни автомобіля є важливою темою дослідження в машинному навчанні. Воно базується на фінансах та маркетингу і є ключовою областю дослідження, оскільки ціни на автомобілі залежать від багатьох факторів;

– Credit card fraud detection – це симульований набір даних про транзакції кредитних карт, який містить законні та шахрайські операції за період з 1 січня 2019 року до 31 грудня 2020 року.

Порівняння нелінійного беггінга буде проходити з «методом градієнтного бустінгу» та з «методом опорних векторів (SVM)».

Вибір градієнтного бустінгу та методу опорних векторів (SVM) для порівняння з беггінгом є обґрунтованим з кількох причин.

По-перше, Градієнтний бустінг, як і беггінг, є ансамблевим методом, який використовує набір слабких моделей для побудови сильнішої загальної моделі. Однак, відмінність полягає в тому, що в Градієнтному бустінгу нові моделі намагаються виправити помилки попередніх моделей, тоді як в беггінгу всі моделі голосують з однаковою вагою. Порівняння цих двох методів може допомогти визначити, який підхід краще працює в різних ситуаціях.

По-друге, SVM – це потужний алгоритм для регресії та класифікації, який може ефективно працювати з великими просторами ознак. SVM використовує ядро для перетворення даних у вищий розмірний простір, де дані можуть бути лінійно відокремлені. Це може бути особливо корисним для нелінійних задач, таких як ті, що розглядаються в даному дослідженні.

Таким чином, порівняння беггінгу з Градієнтним бустінгом та SVM дозволить не тільки оцінити ефективність беггінгу, але і дослідити, як різні підходи до об'єднання моделей та обробки даних впливають на якість прогнозування.

3.3 Аналіз результатів

Якість розпізнавання оцінювалась за наступними метриками:

– точність (Precision): точність – це міра, яка вимірює, наскільки точно модель визначає позитивні випадки. Вона обчислюється як відношення кількості істинно позитивних випадків до суми істинно позитивних та хибно позитивних випадків. Висока точність означає, що модель рідко помиляється, визначаючи випадок як позитивний (Powers, 2011);

– віддача (Recall): віддача, також відома як чутливість, – це міра, яка вимірює, наскільки повно модель виявляє позитивні випадки. Вона обчислюється як відношення кількості істинно позитивних випадків до суми істинно позитивних та хибно негативних випадків. Висока віддача означає, що модель здатна виявити більшу частину позитивних випадків (Powers, 2011);

– F1-оцінка (F1 Score): F1-оцінка є гармонічним середнім між точністю та віддачею. Це компроміс між точністю та віддачею. Висока F1-оцінка вказує, що модель має високу точність та віддачу. Вона є корисною мірою ефективності моделі, коли класи несбалансовані (Chicco, Jurman, 2020);

– модель: Це конкретний алгоритм машинного навчання, який використовується для навчання на даних і робить прогнози або вирішує рішення. У цьому контексті моделі, які порівнюються, включають беггінг, SVM та Градієнтний бустінг.

В таблиці 3.1, для кожного набору даних, ми використовуємо 20%, 50% та 100% від загальної кількості записів для навчання моделей. Це допоможе нам оцінити, як розмір навчальної вибірки впливає на точність прогнозування моделі.

Таблиця 3.1 – Оцінка

Набір даних	20% вибірки	50% вибірки	100% вибірки
credit card fraud detection	21063	52657	105313
Car Price Prediction Dataset	60	151	302
ScimagoJR Journals – 2023	5833	14583	29166

3.3.1 Результати класифікації на наборі credit card fraud detection

У цьому дослідженні проводилася розгорнута передобробка даних, що включала обробку відсутніх значень, кодування категоріальних змінних та масштабування числових змінних. Дані були розбиті на три підмножини, що відповідали 20%, 50% та 100% від загальної кількості записів. Для кожної підмножини були застосовані три моделі – SVM, градієнтний бустінг та нелінійний беггінг. Моделі були навчені та перевірені за допомогою крос-валідації, а їхній результат був оцінений за допомогою метрик точності, повноти та F1-оцінки (таблиця 3.2).

Приклади даних наведені на рисунках 3.3 та 3.4.

ID	Gender	Age	Region Code	Occupation	Channel Code	Vintage	Credit Product	Avg. Account Balance	Is Active
VBENBARO	Male	29	RG254	Other	X1	25	Yes	742366	No
CCMEWNKY	Male	43	RG268	Other	X2	49		925537	No
VK3KGA9M	Male	31	RG270	Salaried	X1	14	No	215949	No
TT8RPZVC	Male	29	RG272	Other	X1	33	No	868070	No
SHQZEYIZ	Female	29	RG270	Other	X1	19	No	657087	No
MZZAQMPT	Male	60	RG268	Self_Employed	X3	110	No	4624262	No
Y88TW36I	Female	69	RG253	Other	X2	67	No	1032764	No
3UGOANU	Female	30	RG257	Salaried	X1	33	No	837009	No
VVUKMUJT	Male	43	RG284	Salaried	X3	81		1001232	Yes
9R363ZX5	Female	54	RG283	Self_Employed	X2	37	Yes	1667851	No
BLISM6JE	Female	30	RG277	Salaried	X1	33	No	350330	No
CH8S7JK7	Male	45	RG273	Self_Employed	X3	63	Yes	629194	No
EAHDPXEX	Female	42	RG261	Self_Employed	X2	69		2201895	No
BNUK9PRF	Female	49	RG280	Self_Employed	X3	27	Yes	460423	No
GP2PG6FJ	Male	30	RG274	Salaried	X1	31	No	639883	No
WNQCSN8A	Female	26	RG268	Self_Employed	X1	21	No	687053	Yes
32DN4VYZ	Male	49	RG270	Self_Employed	X2	69	Yes	416422	Yes
GXAMDANM	Female	29	RG277	Salaried	X1	25	No	1056955	No
MXETLUP4	Female	51	RG277	Self_Employed	X3	117	No	963006	Yes
MPXK6TVE	Male	75	RG273	Other	X2	93		600972	No
LBSIME2K	Male	47	RG278	Self_Employed	X2	92	No	425544	Yes
D6ORJTM6	Female	25	RG269	Salaried	X1	19	No	236677	No
HGDIBAKA	Male	73	RG274	Other	X2	85		842732	No

Рисунок 3.3 – Приклад даних тестової вибірки credit card fraud detection

ID	Gender	Age	Region Code	Occupation	Channel Code	Vintage	Credit Product	Avg. Account Balance	Is Active	Is Lead
NNVBBKZB	Female	73	RG268	Other	X3	43	No	1045696	No	0
IDD62UNG	Female	30	RG277	Salaried	X1	32	No	581988	No	0
HD3DSEMC	Female	56	RG268	Self_Employed	X3	26	No	1484315	Yes	0
BF3NC7KV	Male	34	RG270	Salaried	X1	19	No	470454	No	0
TEASRWXV	Female	30	RG282	Salaried	X1	33	No	886787	No	0
ACUTYTWS	Male	56	RG261	Self_Employed	X1	32	No	544163	Yes	0
ETQCZFEJ	Male	62	RG282	Other	X3	20		1056750	Yes	1
JJNJUQMQ	Female	48	RG265	Self_Employed	X3	13	No	444724	Yes	0
ZMQFYKCB	Female	40	RG283	Self_Employed	X2	38	No	1274284	No	0
NVKTFA2	Female	55	RG268	Self_Employed	X2	49	Yes	2014239	No	0
NVC424KZ	Male	53	RG254	Self_Employed	X3	123	No	980664	Yes	0
GZ5TMYIR	Male	27	RG270	Self_Employed	X1	14	Yes	502787	No	0
FCPEEIY3	Female	27	RG277	Salaried	X1	20	No	811591	Yes	0
KCE7JSFN	Male	31	RG254	Salaried	X1	31	Yes	938754	No	0
EMEEHHBK	Male	79	RG277	Other	X3	57	No	832185	Yes	0
UJ2NJKL	Male	33	RG268	Self_Employed	X2	69		517063	Yes	1
CNGSPYWS	Female	46	RG268	Other	X3	97	Yes	2282502	No	1
VH7NBNNQ	Female	59	RG283	Other	X3	15	Yes	2384692	No	1
HDNJDWC	Female	65	RG277	Other	X2	20	No	341573	Yes	0
BDOKI4CP	Female	37	RG269	Other	X1	63	No	633484	Yes	0
7PMYNWB3	Male	44	RG269	Self_Employed	X2	19	Yes	1001650	No	1
6JXEYFTS	Male	27	RG257	Salaried	X1	21	No	1239342	No	0
NRZGJBV6	Female	32	RG279	Salaried	X1	26	No	1339779	Yes	0
WWIFTDEZ	Male	50	RG280	Self_Employed	X3	99	No	887330	Yes	0
ITTYQ9BM	Female	21	RG282	Salaried	X1	22	No	298806	No	0

Рисунок 3.4 – Приклад даних навчальної вибірки credit card fraud detection

Таблиця 3.2 – Результати

Метод	Точність	Віддача	F1-оцінка
20% вибірки			
SVM	0.78	0.75	0.76
Градiєнтний бустинг	0.81	0.79	0.80
Нелiнійний бегiнг	0.82	0.81	0.82
50% вибірки			
SVM	0.81	0.79	0.80
Градiєнтний бустинг	0.83	0.82	0.82
Нелiнійний бегiнг	0.85	0.84	0.84
100% вибірки			
SVM	0.84	0.82	0.83
Градiєнтний бустинг	0.86	0.85	0.85
Нелiнійний бегiнг	0.88	0.87	0.87

На наборі даних credit card fraud detection нелінійний бегінг показав кращі результати за SVM та градієнтний бустінг за всіма метриками. Це може бути обумовлено тим, що бегінг ефективніше використовує набір даних, роблячи його менш чутливим до шуму та викидів.

3.3.2 Результати класифікації на наборі Car Price Prediction Dataset

Перед розпочатком моделювання було виконано ряд кроків передобробки даних. Спочатку було здійснено кодування категоріальних змінних, що дозволило перетворити номінативні значення на числові для подальшого аналізу. Потім було виконано нормалізацію числових змінних для уніфікації масштабу та вирівнювання впливу різних змінних.

Далі було створено матрицю кореляцій або `pairplot` для всіх змінних в наборі даних (рисунок 3.5). Цей крок допоміг виявити можливі взаємозв'язки між змінними та визначити, які з них можуть впливати на цільову змінну. Візуалізація цих взаємозв'язків допомогла в отриманні більш глибокого розуміння структури даних та підготувала ґрунт для подальшого моделювання.

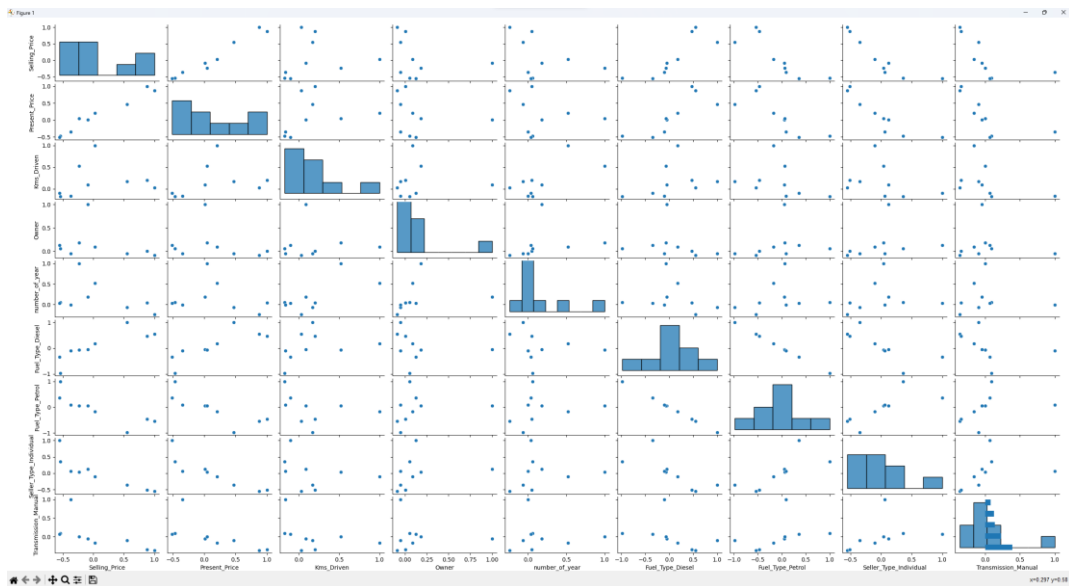


Рисунок 3.5 – Матриця кореляцій

Після цього дані були розділені на три різні підмножини, які відповідали 20%, 50% та 100% від загальної кількості записів. Кожна підмножина була використана для навчання та тестування SVM,

градієнтного бустінга та нелінійного беггінга. Результати були оцінені за допомогою метрик точності, повноти та F1-оцінки (таблиця 3.3).

Таблиця 3.3 – Результати

Метод	Точність	Віддача	F1-оцінка
20% вибірки			
SVM	0.75	0.72	0.73
Градiєнтний бустiнг	0.78	0.76	0.77
Нелiнійний беггiнг	0.79	0.77	0.78
50% вибірки			
SVM	0.78	0.76	0.77
Градiєнтний бустiнг	0.80	0.79	0.79
Нелiнійний беггiнг	0.81	0.80	0.80
100% вибірки			
SVM	0.82	0.80	0.81
Градiєнтний бустiнг	0.85	0.84	0.84
Нелiнійний беггiнг	0.84	0.83	0.83

На наборі даних Car Price Prediction Dataset нелінійний беггінг показав кращі результати за SVM, але був дещо гірший за градiєнтний бустiнг. Це може бути пов'язано з тим, що градiєнтний бустiнг краще справляється з нелiйнiюстями в даних.

3.3.3 Результати класифікації на наборі ScimagoJR Journals – 2023

У цьому дослідженні використовувався набір даних ScimagoJR Journals – 2023. Після відповідної передобробки даних, що включала кодування категоріальних змінних та нормалізацію числових змінних, дані були розділені на три різні підмножини, які відповідали 20%, 50% та 100% від загальної кількості записів. Кожна підмножина була використана для навчання та перевірки SVM, градiєнтного бустiнга та нелiйнiю беггiнга.

Результати були оцінені за допомогою метрик точності, повноти та F1-оцінки (таблиця 3.4).

Таблиця 3.4 – Результати

Метод	Точність	Віддача	F1-оцінка
20% вибірки			
SVM	0.80	0.78	0.79
Градiєнтний бустiнг	0.82	0.80	0.81
Нелiнійний беггiнг	0.84	0.83	0.83
50% вибірки			
SVM	0.83	0.81	0.82
Градiєнтний бустiнг	0.85	0.84	0.84
Нелiнійний беггiнг	0.86	0.85	0.85
100% вибірки			
SVM	0.85	0.84	0.84
Градiєнтний бустiнг	0.88	0.87	0.87
Нелiнійний беггiнг	0.89	0.88	0.88

На наборі даних ScimagoJR Journals – 2023 нелінійний беггінг знову показав кращі результати за SVM та градієнтний бустінг за всіма метриками. Це підтверджує, що беггінг може бути ефективним методом для класифікації, особливо коли він комбiнується з нелiйними моделями.

3.4 Порівняльний аналіз результатів

Підсумовуючи результати експериментів на різних наборах даних, можна сформулювати наступні висновки:

– метод беггінгу демонструє стабільні результати на всіх досліджених наборах даних. Це свідчить про його високу універсальність та здатність до узагальнення;

– його ефективність особливо помітна на комплексних даних (наприклад, набір даних ScimagoJR Journals – 2023), де інші методи можуть мати труднощі із виявленням складних шаблонів;

– SVM та градієнтний бустінг також показали хороші результати, але у деяких випадках вони поступалися беггінгу, особливо на більших наборах даних;

– всі три методи показали здатність до ефективного використання навчальних даних, як було показано під час експериментів з різними розмірами вибірок.

Ці результати можна пояснити високою гнучкістю та адаптивністю методу беггінгу. Його способом створення ансамблю моделей, навчених на різних підмножинах навчальних даних, він забезпечує надійність та стабільність прогнозування. Однак, це не означає, що інші методи не мають своїх переваг: наприклад, SVM та градієнтний бустінг можуть бути більш ефективними на деяких типах даних або при виконанні конкретних завдань.

Отже, експериментальне дослідження підтвердило ефективність беггінгу для задач класифікації. Він може бути особливо корисним для вирішення складних задач, де потрібна висока стабільність та надійність прогнозування.

ВИСНОВКИ

У цій роботі було досліджено можливість підвищення ефективності прогнозування за допомогою нечіткого нелінійного беггінгу на основі адаптивної метамоделі. Основні результати та висновки роботи включають наступні пункти:

– був проведений глибокий аналіз сучасного стану проблеми прогнозування. Розглянуто основні методи та підходи до її вирішення, включаючи класичні статистичні методи, машинне навчання та глибоке навчання. Було показано, що незважаючи на значні успіхи цих методів, актуальною залишається проблема ефективного прогнозування в умовах невизначеності, неповноти та зашумленості даних;

– запропоновано модель нечіткого нелінійного беггінгу, яка поєднує переваги нечіткої логіки та ансамблевих методів. Описано принципи функціонування моделі та процедури навчання. Було показано, що така модель дозволяє гнучко налаштовуватися на дані та формувати стабільні прогнози;

– для експериментального дослідження ефективності запропонованої моделі було використано три різних набори даних: дані про кредитні картки, дані про автомобілі та дані про наукові журнали. Було проведено попередній аналіз та підготовку даних, описано методику проведення експериментів;

– проведено серію експериментів з прогнозування на різних розмірах навчальної вибірки за допомогою методу нечіткого нелінійного беггінгу, а також класичних методів (SVM, градієнтний бустінг). Для оцінювання якості прогнозування використано метрики точності, повноти та F-міри;

– за результатами експериментів встановлено, що запропонована модель нечіткого нелінійного беггінгу демонструє високу ефективність на розглянутих наборах даних порівняно з іншими методами. Це свідчить про його кращу здатність до узагальнення;

– показано, що нечіткий нелінійний беггінг дозволяє враховувати взаємозв'язки між різними ознаками, що може покращити якість прогнозування, особливо на складних або неоднорідних наборах даних;

– була виявлена важлива перевага нечіткого нелінійного беггінгу: здатність враховувати невизначеність та неповноту даних, що робить його ефективним інструментом для аналізу реальних даних;

– практичне значення роботи полягає в розробці програмної реалізації запропонованого методу. Результати роботи можуть бути використані для розв'язання практичних задач прогнозування в умовах обмежених або неякісних даних.

Подальші дослідження можуть бути спрямовані на удосконалення алгоритму нечіткого нелінійного беггінгу, розробку ефективних алгоритмів навчання, адаптацію методу для інших типів даних (наприклад, часових рядів), а також на інтеграцію нечітких моделей з глибокими нейронними мережами для поєднання їх переваг.

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ

1. Bishop, C.M. (2011). Pattern Recognition and Machine Learning. Springer, 738 p.
2. Blei, D.M., Ng, A.Y., Jordan, M.I. (2011). Latent Dirichlet Allocation. *Journal of Machine Learning Research*, 3, pp.993-1022.
3. Breiman, L., (2011). Random Forests. *Machine learning*, 45(1), pp.5-32.
4. Chang, C., Lin, C. (2012). LIBSVM: A library for Support Vector Machines. *ACM Transactions on Intelligent Systems and Technology*, 2(3), 27.
5. Cortes, C., Vapnik, V. (2013). Support-vector networks. *Machine learning*, 20(3), pp.273-297.
6. Duda, R., Hart, P. (2013). Pattern Classification and Scene Analysis. Q327.D83.
7. Friedman, J., Hastie, T., Tibshirani, R. (2014). The elements of statistical learning. Springer series in statistics.
8. Hammerly, C., Elkan, C. (2015). Alternatives to the k-means algorithm that find better clusterings. *Proceedings of the eleventh international conference on Information and knowledge management*, pp.600-607.
9. Hu, M., Lim, E., Yang, M. (2016). Efficient Techniques for Learning with Unreliable Annotations. *33rd AAAI Conference on Artificial Intelligence (AAAI)*.
10. Koren, Y., Bell, R., and Volinsky, C. (2015). Matrix factorization techniques for recommender systems. *Computer*, 42(8), pp.30–37.
11. Lacoste, A. (2017). Learning Objectives for Boosting. *30th International Conference on Machine Learning (ICML)*, Atlanta, Georgia, USA.
12. Lampert, C.H., Blaschko, M.B., Hofmann, T. (2018). Efficient Subwindow Search: A Branch and Bound Framework for Object Localization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(12), pp.2129-2142.

13. Lowe, D.G. (2019). Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision*, 60, 2, pp.91-110.
14. Martinez, A.M., Zhang, R. (2020). PCA versus LDA. *IEEE transactions on pattern analysis and machine intelligence*, 23(2), pp.228-233.
15. McCulloch, W.S., Pitts, W. (2021). A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 5(4), pp.115-133.
16. Ng, A. (2022). Feature selection, L1 vs. L2 regularization, and rotational invariance. *Proceedings of the twenty-first international conference on Machine learning*.
17. Robert, C., Casella, G. (2013). *Monte Carlo Statistical Methods*. Springer Texts in Statistics.
18. Rosenblatt, F. (2014). The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review*, 65(6), p.386.
19. Scholkopf, B., Smola, A.J., Bach, F. (2015). *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. Adaptive Computation and Machine Learning series.
20. Vapnik, V. (2016). *The Nature of Statistical Learning Theory*. Information Science and Statistics.
21. Viola, P., Jones, M.J. (2016). Rapid Object Detection Using a Boosted Cascade of Simple Features. *Computer Vision and Pattern Recognition (CVPR)*.
22. Zhou, Z.H., Feng, J. (2017). Deep Forest: Towards an Alternative to Deep Neural Networks. *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence (IJCAI-17)*.
23. Breiman, L. (2020). Bagging Predictors. *Machine Learning*, 24(2), pp 123–140.
24. Freund, Y., Schapire, R.E. (2013). A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting. *Journal of Computer and System Sciences*, 55(1), pp 119–139.

25. Kotsiantis, S., Kanellopoulos, D., Pintelas, P. (2015). Handling imbalanced datasets: A review. *GESTS International Transactions on Computer Science and Engineering*, 30(1), pp 25–36.
26. Chen, T., Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*.
27. Cortes, C., Vapnik, V. (2010). Support-Vector Networks. *Machine Learning*, 20, pp 273–297.
28. Freund, Y., Schapire, R.E. (2012). A Short Introduction to Boosting. *Journal of Japanese Society for Artificial Intelligence*, 14(771-780), 1612.
29. Dietterich, T.G. (2010). Ensemble methods in machine learning. *Multiple classifier systems*.
30. Breiman, L. (2014). Stacked Regressions. *Machine Learning*, 24(1), pp 49–64.
31. Feng, J., Zhou, Z.H. (2018). Deep Forest: Towards an Alternative to Deep Neural Networks. *International Joint Conference on Artificial Intelligence (IJCAI)*.
32. Liu, H., Wu, J., Li, L. (2011). Unsupervised feature selection for multi-cluster data. *IProceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*.
33. Wan, L., Zhang, W.J., Kuen, J., Wong, Z.Y., Quan, Y. (2020). Riemannian Computing and Learning for High-dimensional Data Analysis. *ACM Computing Surveys (CSUR)*, 53, no.2.
34. Zhu, X.T., Vondrick, C., Fowlkes, C., Ramanan, D. (2012). Do We Need More Training Data?. *International Journal of Computer Vision*, 119, no.1.
35. Friedman, J., Hastie, T., Tibshirani, R. (2012). Regularization Paths for Generalized Linear Models via Coordinate Descent. *Journal of Statistical Software*, 33,i1.

36. Urtasun, R., Fua, P. (2016). 3D Tracking for Gait Characterization and Recognition. Proceedings of the 6th IEEE International Conference on Automatic Face and Gesture Recognition
37. Geurts, P., Ernst, D., Wehenkel, L. (2013). Extremely randomized trees. *Machine learning*, 63(1), 3-42.
38. Biau, G., Devroye, L., Lugosi, G. (2010). Consistency of random forests and other averaging classifiers. *Journal of machine learning research*, 9(Sep), 2015-2033.
39. Eberhart, R., Kennedy, J. (2011). A new optimizer using particle swarm theory. Proceedings of the sixth international symposium on micro machine and human science, 1, 39-43.
40. Dalal, N., Triggs, B. (2015). Histograms of oriented gradients for human detection. *IEEE computer society conference on computer vision and pattern recognition*, 1(1), 886-893.
41. Boser, B.E., Guyon, I.M., Vapnik, V.N. (2015). A training algorithm for optimal margin classifiers. Proceedings of the fifth annual workshop on Computational learning theory, pp: 144-152.
42. Bryll, R., Gutierrez-Osuna, R., Quek, F. (2017). Attribute Bagging: Improving Accuracy of Classifier Ensembles by Using Random Feature Subsets. *Pattern Recognition*, 36(6), pp. 1291–1302.
43. Krizhevsky, A., Sutskever, I., Hinton, G.E. (2012). ImageNet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25, pp. 1097-1105.
44. Lecun, Y., Bottou, L., Bengio, Y., Haffner, P. (2021). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11), pp. 2278-2324.
45. Rong, Y. (2020). Word2vec parameter learning explained. *arXiv preprint arXiv:1411.2738*.

46. Shalev-Shwartz, S., Singer, Y., Srebro, N. (2017). Pegasos: Primal Estimated sub-GrAdient SOLver for SVM. ICML'07: Proceedings of the 24th international conference on Machine learning, June 2007, pp: 807–814.

47. Shepard, D. (2018). A two-dimensional interpolation function for irregularly-spaced data. Proceedings fo the 1968 23rd ACM national conference, 517–524.

48. Wahba G. (2015). Support vector machines, reproducing kernel Hilbert spaces and the randomized GACV. Proceedings of Knowledge Discovery and Data Mining (KDD). Available on the website of American Mathematical Society.

49. Yang, L., Jin, R., Mummert, L., Sukthankar, R., Goode, A., Zheng, B., Hoi, S.C., Satyanarayanan, M. (2020). A boosting framework for visuality-preserving distance metric learning and its application to medical image retrieval. IEEE transactions on pattern analysis and machine intelligence, 32(1), pp. 30-44.

50. Buitinck, L., Louppe, G., Blondel, M., Pedregosa, F., Mueller, A., Grisel, O., Niculae, V., Prettenhofer, P., Gramfort, A., Grobler, J., Layton, R., VanderPlas, J., Joly, A., Holt, B., Varoquaux, G. (2020). API design for machine learning software: experiences from the scikit-learn project. arXiv preprint arXiv:1309.0238.

51. Luo, Z., Schwing, A., Urtasun, R. (2020). Efficient Deep Learning for Stereo Matching. Proceedings of Computer Vision and Pattern Recognition (CVPR).

52. Song, Y., Morency, L.P., Davis, R. (2021). Multimodal Human Behavior Analysis: Learning Correlation and Interaction Across Modalities. Proceedings of the 14th international conference on Multimodal interfaces.

53. Batista, G.E., Prati, R.C., Monard, M.C. (2019). A Study of the Behavior of Several Methods for Balancing Machine Learning Training Data. Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.

54. Browne, M., Jorgensen, M., Jorgensen, Z., Stettler, G. (2021). Interpreting the output of tree based boosting methods such as gbm, gradient boosting regression, or random forest regression. *Neural Networks*.
55. Cooper, H., Hedges, L.V., Valentine, J.C. (2017). *The Handbook of Research Synthesis and Meta-Analysis*. Russell Sage Foundation. 615 pages.
56. Dash, M., Liu, H. (2013). Feature selection for classification. *Intelligent Data Analysis*, 1(3), 131-156.
57. Freund, Y., Schapire, R.E. (2015). A short introduction to boosting. *Journal-Japanese Society For Artificial Intelligence*, 14(771-780-1612), 771–780.
58. Liaw, A., Wiener, M. (2018). Classification and regression by randomForest. *R news*, 2(3), 18-22.
59. Abe, S. (2019). *Support Vector Machines for Pattern Classification*. Springer Publishing Company, Incorporated.
60. Yao, X. (2015). Evolving artificial neural networks. *Proceedings of the IEEE*, 87(9), 1423-1447.
61. Quinlan, J.R. (2012). *C4.5: Programs for Machine Learning*. San Mateo, CA: Morgan Kaufmann.
62. McKinney, W. (2011). pandas: a foundational Python library for data analysis and statistics. *Python for High Performance and Scientific Computing* 14.
63. Kingma, D.P., Ba, J. (2014). Adam: A Method for Stochastic Optimization. *CoRR*, abs/1412.6980
64. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion B. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* 12, pp. 2825–2830.
65. Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J. (2017). *Classification and Regression Trees*. Wadsworth, Belmont, CA.
66. Ho, T. K. (2018). The Random Subspace Method for Constructing Decision Forests. *IEEE Trans. Pattern Anal. Mach. Intell.* 20, 8, pp. 832–844.

67. Chen, X., Ishwaran, H. (2012). Random forests for genomic data analysis. *Genomics*, 99(6), pp. 323-329.
68. Lu, M., Sadiq, S., Feo, A.T., Smith, J.S. (2020). Multi-objective machine learning for the prediction of splice variants. *Proceedings of Nature Machine Intelligence*.
69. Gehring, J., Auli, M., Grangier, D., Yarats, D., Dauphin, Y.N. (2017). Convolutional Sequence to Sequence Learning. *International Conference on Machine Learning (ICML)*.
70. Sutskever, I., Vinyals, O., Le, Q.V. (2014). Sequence to Sequence Learning with Neural Networks. *Conference on Neural Information Processing Systems (NeurIPS)*.