

НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ «ЛЬВІВСЬКА ПОЛІТЕХНІКА»
МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

Міщук Олександра Сергіївна



УДК 004.67+004.89+004.942

**НЕЙРОПОДІБНІ МЕТОДИ ТА ЗАСОБИ ПРОГНОЗУВАННЯ
ПАРАМЕТРІВ ЗАБРУДНЕННЯ АТМОСФЕРНОГО ПОВІТРЯ**

05.13.23 – системи та засоби штучного інтелекту

Автореферат

дисертації на здобуття наукового ступеня

кандидата технічних наук

Львів – 2021

Дисертацією є рукопис.

Робота виконана в Національному університеті “Львівська політехніка” Міністерства освіти і науки України.

Науковий керівник:

доктор технічних наук, професор

Ткаченко Роман Олексійович,

Національний університет “Львівська політехніка”,

Завідувач кафедри інформаційних технологій видавничої справи

Офіційні опоненти:

доктор технічних наук, професор

Березький Олег Миколайович

Західноукраїнський національний економічний університет

Міністерства освіти і науки України, м. Тернопіль

завідувач кафедри комп'ютерної інженерії

кандидат технічних наук, доцент

Коцовський Владислав Миронович

Державний вищий навчальний заклад «Ужгородський національний університет» Міністерства освіти і науки України, м. Ужгород

доцент кафедри інформаційних управляючих систем та технологій

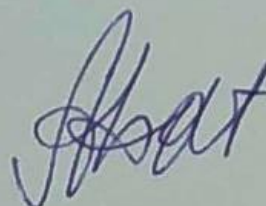
Захист відбудеться “09” квітня 2021 р. о 14:00 годині на засіданні спеціалізованої вченої ради Д 35.052.14 Національного університету “Львівська політехніка” за адресою: 79013, м. Львів, вул. Степана Бандери, 28а, ауд. 807, V навчальний корпус.

З дисертацією можна ознайомитися у науково-технічній бібліотеці Національного університету “Львівська політехніка” за адресою: 79013, м. Львів, вул. Професорська, 1.

Автореферат розісланий

“03” березня 2021 р.

Вчений секретар
спеціалізованої вченої ради
кандидат технічних наук, доцент



А.Є. Батюк

ЗАГАЛЬНА ХАРАКТЕРИСТИКА ДОСЛІДЖЕННЯ

Актуальність роботи.

Велика кількість небезпечних викидів у повітряне середовище викликана функціонуванням промислових виробництв та автотранспорту, котрі містять низку постійно діючих джерел забруднення атмосферного повітря, що становить реальну загрозу для людини та всієї екосистеми планети. Тому у більшості країн світу здійснюють моніторинг навколишнього середовища із застосуванням рекомендацій Організації Об'єднаних Націй (резолюція №2286), враховуючи національні особливості.

Одним із завдань моніторингу довкілля є прогнозування розвитку певного екологічного процесу для забезпечення мінімального ступеня негативного впливу людської діяльності на повітряне середовище. Прогнозування параметрів забруднення атмосферного повітря, котре повсякчасно впливає на здоров'я людей, дозволяє випереджувати та вчасно реагувати на підвищення рівня викидів шкідливих домішок у повітря. До шкідливих домішок відносяться такі параметри забруднення повітря, як оксиди більшості важких металів, оксиди вуглецю, сірки, азоту, сажа, пил, вуглеводні, з'єднання свинцю та інші. Перелік параметрів забруднення атмосферного повітря визначається спеціальним державним органом та у кожній країні може бути різним. Завдання прогнозування параметрів забруднення довкілля полягає у передбаченні кожної окремо визначеної забруднюючої речовини, що викидається у повітряне середовище.

Особливості методів прогнозування параметрів забруднення повітряного середовища, в тому числі за допомогою засобів штучного інтелекту, зокрема штучних нейронних мереж, розглядаються у наукових роботах: К. А. Мальцева, В. Д. Погребенника, В. С. Джигирея, С. М. Дзюби, Л. Е. Чорнобая, О. В. Ничика, С. С. Харинцева, А. А. Севастьянова, М. Х. Салахова, Ф. Меканіка (F. Mekanik), М. Брауера (M. Brauer) та інших.

При виконанні завдання прогнозування параметрів забруднення атмосферного повітря виникає проблема неповноти даних, внаслідок пропусків окремих атрибутів у векторах вимірних параметрів забруднення повітряного середовища під час виконання моніторингу довкілля. Це є перешкодою для прогнозування параметрів забруднення повітряного середовища та своєчасного реагування на збільшення рівня викидів шкідливих домішок у атмосферу.

Розв'язки задачі заповнення пропусків у даних, котрі передбачають інтелектуальний аналіз, підбір моделей та методів, та реалізацію цих методів сучасними інструментальними засобами подаються у роботах українських та зарубіжних авторів, серед яких: Р. М. Камінський, В. В. Пасічник, В. Б. Мокін, Н. Г. Загоруйко, О. О. Слабченко, Н. В. Кузнецова, І. К. Зангієва, Б. П. Бочаров, Р. Дж. Літл (R. J. Little), П. Елісон (P. Elison), Дж. В. Грехем (J. V. Graham), Д. Ньюман (D. Newman), А. Карахаліос (A. Karahalios) та багато інших.

Високоточне прогнозування може бути реалізоване шляхом використання нейромережових засобів та методів. Традиційні нейропарадигми, котрі базуються на основі методу рухомих часових вікон, не усувають негативний вплив шумоподібних коливань, що знижує точність прогнозів. Ситуативні зміни випадкових характеристик

даних моніторингу довкілля вимагають періодичного перенавчання, що складно реалізувати для традиційних нейромережових засобів з низькими характеристиками швидкодії навчання та складнощами налагодження параметрів. Нейроподібні структури на основі моделі послідовних геометричних перетворень (НС МПГП) мають певні біологічні аналогії лише в плані топології, але використовують математично обґрунтовані методи швидкого неітеративного навчання, що відкриває розширені можливості підвищення ефективності вирішення зазначеного завдання. Додаткові функції НС МПГП, зокрема можливість розкладу часових послідовностей на тренд і сукупність коливань, забезпечують додаткові можливості отримання важливої прогностичної інформації.

Концептуальна модель послідовних геометричних перетворень, що складає основу нейроподібної структури, має підвищену швидкодію та точність застосування. Однак вказана структура не використовувалася для розв'язку задач прогнозування в умовах невизначеності за рахунок наявних пропущених атрибутів у параметрах забруднення атмосферного повітря, для яких характерні такі особливості як швидка зміна в часі та наявність значних шумових складових. Також не здійснювалось виділення лише тренду, де тренд в сумі складових є часовою послідовністю, що може бути ефективним саме в задачах прогнозування параметрів забруднення атмосферного повітря.

Разом з тим, розвиток моніторингу стану довкілля є важливим для контролю якості атмосферного повітря та своєчасного реагування на динаміку поведінки джерел викидів забруднюючих речовин з використанням різних пристроїв комп'ютерної техніки. Тому необхідною є розробка нових методів та засобів прогнозування параметрів забруднення підвищеної точності та швидкодії для користування на мобільних пристроях та контролерах.

Враховуючи описане вище, розробка перспективних методів та засобів прогнозування параметрів забруднення атмосферного повітря, в тому числі в умовах пропусків у даних, за допомогою нейронних структур моделі послідовних геометричних перетворень є актуальним завданням, що потребує розв'язку.

Зв'язок роботи з науковими програмами, планами, темами.

Дисертація виконана згідно з науковим напрямом кафедри інформаційних технологій видавничої справи – «Синтез та технології інтелектуального аналізу даних в гібридних інформаційних середовищах». Результати дисертаційної роботи використано під час виконання:

- держбюджетної науково-дослідної роботи за темою: «Нейромережева технологія захисту та передачі даних у реальному часі з використанням шумоподібних кодів» - номер державної реєстрації № 0119U002256;
- держбюджетної науково-дослідної роботи за темою: «Інформаційна технологія опрацювання персоналізованої медичної інформації» - номер державної реєстрації № 0119U002257.

Розроблені в результаті дисертаційного дослідження методи впроваджені в Здолбунівському відділенні АТ "РІВНЕГАЗ" для використання у робочому процесі підприємства під час виконання завдань моніторингу повітряного середовища.

Мета і завдання дослідження.

Метою дослідження є розроблення програмного засобу та методів побудови нейроподібних структур з неітеративним навчанням для підвищення точності прогнозування параметрів забруднення атмосферного повітря на мобільних та вбудованих пристроях.

Об'єктом дослідження є процеси прогнозування, зокрема в умовах пропусків у даних, під час виконання екологічного моніторингу для контролю забруднення повітряного середовища та своєчасного реагування на можливе виникнення надзвичайних ситуацій.

Предметом дослідження є нейроподібні структури моделі послідовних геометричних перетворень, їх методи навчання і застосування в режимі однокрокового та багатокрокового прогнозування, в тому числі в умовах пропущених даних моніторингу забруднення атмосферного повітря.

Завдання дослідження. Для досягнення мети дисертаційного дослідження було поставлено такі завдання:

1. Проаналізувати актуальні завдання моніторингу довкілля та особливості існуючих методів та засобів прогнозування параметрів забруднення атмосферного повітря, в тому числі в умовах пропусків у даних моніторингу повітряного середовища.
2. Розробити та апробувати метод короткотермінового прогнозування за допомогою використання нейроподібних структур моделі послідовних геометричних перетворень для збільшення горизонту прогнозування.
3. Розробити та дослідити метод формування додаткових вхідних атрибутів векторів даних для підвищення точності заповнення пропущених концентрацій параметрів забруднення атмосферного повітря.
4. Розвинути метод нейромережевої ідентифікації коефіцієнтів полінома, що відтворює функції навченої нейроподібної структури у режимі застосування для зменшення затрат пам'яті мобільних та вбудованих пристроїв.
5. Удосконалити метод функціонального розширення вхідних векторів даних Йох-Хан Пао для зниження викидів в точках екстраполяції нелінійних поверхонь відгуку.
6. Розробити програмний засіб з набором бібліотек для виконання прогнозування параметрів забруднення повітряного середовища, зокрема в умовах пропусків у даних моніторингу атмосферного повітря.

Методи дослідження.

У роботі використано такі методи: метод заповнення пропусків шляхом усереднення; методи машинного навчання на основі регресійного аналізу та нейромережеві методи для заповнення пропущених атрибутів оксиду азоту та карбону; методи наївного прогнозу та часових вікон для виконання однокрокового та багатокрокового прогнозування параметрів забруднення повітря на основі комітету нейроподібних структур різних типів; метод пришвидшеного прогнозування параметрів забруднення повітряного середовища на основі нейроподібної структури моделі послідовних геометричних перетворень та лінійних поліномів.

Наукова новизна одержаних результатів.

Під час розв'язання поставлених завдань отримано такі наукові результати:

вперше:

- розроблено метод уведення додаткових атрибутів – маркерів кластерів у вектори входів, що забезпечило підвищення точності заповнення пропущених показників параметрів забруднення атмосферного повітря;
- метод короткотермінового прогнозування параметрів забруднення атмосферного повітря за допомогою комітету лінійної та нелінійної нейроподібних структур для часткового коректування окремо додатних і від'ємних відхилень від точних значень, що забезпечило збільшення горизонту прогнозування часових послідовностей;

удосконалено:

- метод функційного розширення входів Йох-Хан Пао шляхом застосування раціональних дробів, що забезпечило підвищення точності заповнення пропущених концентрацій параметрів забруднення атмосферного повітря за рахунок зниження викидів в екстраполятивних точках;

отримав подальший розвиток:

- метод побудови матриці коефіцієнтів лінійних поліномів, створеної шляхом їх ідентифікації за результатами навчання лінійної нейроподібної структури моделі послідовних геометричних перетворень, що забезпечило зменшення затрат пам'яті під час прогнозування параметрів забруднення атмосферного повітря.

Практичне значення одержаних результатів.

Розроблений метод короткотермінового прогнозування часових послідовностей показників параметрів забруднення повітряного середовища за допомогою комітету лінійної та нелінійної нейроподібних структур забезпечив збільшення горизонту прогнозування тренду забруднення атмосферного повітря за рахунок часткового коректування окремо додатних і від'ємних відхилень від точних значень. Наприклад, для такого параметру забруднення повітряного середовища як чадний газ, похибка прогнозування розробленим методом відносно застосування нейроподібної структури моделі послідовних геометричних перетворень зменшилась на 15%, а горизонт прогнозування збільшився на два дні.

Розроблений метод заповнення пропущених атрибутів у параметрах забруднення повітряного середовища з розширенням входів за допомогою формування додаткових атрибутів вхідних векторів даних, забезпечив підвищення точності заповнення пропусків у даних моніторингу довкілля. Середня відносна похибка заповнення розробленим методом відносно застосування нейроподібної структури моделі послідовних геометричних перетворень зменшилась на 8,5 % (діоксиду азоту на 14,5 %, а оксиду вуглецю на 2,5 %).

Дослідження показали, що розроблений метод підвищення точності заповнення пропусків у параметрах забруднення атмосферного повітря на основі формування додаткових атрибутів вхідних векторів даних підвищує точність для всіх досліджених методів (нейроподібної структури моделі послідовних геометричних перетворень,

лінійної регресії зі стохастичним градієнтним спуском, дерева рішень та адаптивного бустингу), окрім методу на основі машини опорних векторів через особливості його побудови. Наприклад, при розширенні входних ознак векторів даних та при застосуванні методу на основі опорних векторів, точність прогнозування погіршилась на 1,5 % (для діоксиду азоту на 1%, а для оксиду вуглецю на 2 %).

Удосконалений метод функційного розширення входів Йох-Хан Пао за допомогою застосування раціональних дробів забезпечив підвищення точності заповнення пропущених значень параметрів забруднення атмосферного повітря на 2,6-6% в залежності від виду параметру відносно застосування нейроподібної структури моделі послідовних геометричних перетворень.

Експериментально доведено, що розвинутий метод побудови матриці коефіцієнтів лінійних поліномів, створеної шляхом їх ідентифікації за результатами навчання лінійної нейроподібної структури моделі послідовних геометричних перетворень забезпечив зменшення затрат пам'яті під час прогнозування параметрів забруднення атмосферного повітря в 12,75 разів.

Результати дисертаційної роботи використано в навчальному процесі кафедри інформаційних технологій видавничої справи при вивченні дисциплін «Інтелектуальний аналіз» для студентів спеціальності 186 «Видавництво та поліграфія» та при виконанні держбюджетних науково-дослідних робіт «Нейромережева технологія захисту та передачі даних у реальному часі з використанням шумоподібних кодів», № 0119U002256 (методи підвищення точності заповнення пропусків у даних) та «Інформаційна технологія опрацювання персоналізованої медичної інформації», № 0119U002257 (методи підвищення точності прогнозування).

Особистий внесок здобувача.

Авторка самостійно отримала усі результати дисертаційної роботи. В публікаціях, написаних одноосібно здобувачці належать: розроблення методів короткотермінового прогнозування параметрів забруднення атмосферного повітря на основі нейронних структур моделі послідовних геометричних перетворень [8, 15]; аналіз алгоритмів машинного навчання для прогнозування, в тому числі в умовах частково пропущених концентрацій параметрів забруднення атмосферного повітря [13]; дослідження методу введення додаткових ознак входних векторів, шляхом попереднього виділення компактних множин точок [11, 14]; реалізація методу розширення входів Йох-Хан Пао за допомогою введення обернено-пропорційних квадратичних функцій [12]; розроблення методу підвищення точності прогнозування на основі методу коректування похибки за допомогою комітету нейронних структур різних типів [1, 16, 17]. В публікаціях, написаних у співавторстві здобувачці належать: розроблення методу розширення входів за допомогою виділення компактних множин точок [5]; дослідження методу підвищення швидкості прогнозування параметрів забруднення повітряного середовища на основі лінійних поліномів [7]; реалізація та порівняння методів заповнення пропусків на основі нейроподібних структур [2, 3, 6, 9, 10]; виконання аналізу методу однокрокового прогнозування параметрів забруднення атмосферного повітря на основі нейронних структур моделі послідовних геометричних перетворень [4].

Апробація результатів дисертації.

Результати дисертаційної роботи були апробовані на наступних 12-ти міжнародних та всеукраїнських наукових конференціях: XIV-й міжнародній науковій конференції «Інтелектуальні системи прийняття рішень і проблеми обчислювального інтелекту» ISDMCI-2018 (м. Залізний Порт, 21-27 травня 2018 року); V-й науково-технічній конференції «Обчислювальні методи і системи перетворення інформації» ОМІСПІ-2018 (м. Львів, 4-5 жовтня 2018); IV-й міжнародній науково-технічній конференції «Комп'ютерне моделювання та оптимізація складних систем» КМОСС-2018 (м. Дніпро, 1-2 листопада 2018 року); IInd International Conference «Computer Science, Engineering and Education Applications» ICCSEEA-2019 (Kyiv, 26-27 January 2019); III-й міжнародній науково-технічній конференції «Комп'ютерні та інформаційні системи і технології» CSITIC-2019 (м. Харків, 23-24 квітня 2019); XVth International Work-Conference «Artificial Neural Networks» IWANN 2019 (Gran Canaria, 12-14 June 2019); XIVth International Conference «Future Networks and Communications» FNC-2019 (Halifax, 19-21 August 2019); XIth International Scientific and Practical Conference «Electronics and Information Technologies» ELIT-2019 (Lviv, 16-18 September 2019); XVIII-й міжнародній науково-практичній конференції «Побудова інформаційного суспільства: ресурси і технології» (м. Київ, 18-19 вересня 2019 року); XIIth International Scientific Conference «Information Technologies and Automation» ІТА-2019 (Odesa, 17-18 October 2019); V-й міжнародній науково-технічній конференції «Комп'ютерне моделювання та оптимізація складних систем» КМОСС-2019 (м. Дніпро, 6-8 листопада 2019); International Workshop «Digitalization and Servitization within Factory-Free Economy» D&SwFFE-2019 (Coimbra, 4-7 November 2019).

Також результати дослідження були представлені на наукових семінарах кафедри інформаційних технологій видавничої справи Національного університету «Львівська політехніка» (2013-2014, 2017-2019).

Публікації.

За тематикою дослідження опубліковано *сімнадцять* наукових публікацій, в тому числі *чотири* статті у наукових фахових виданнях України [1, 7, 8, 9]; *одна* стаття у науковому періодичному виданні іншої держави [3], *чотири* матеріали конференцій у наукових серійних закордонних виданнях, що включено до міжнародних наукометричних баз [2, 4, 5, 6]; *вісім* тез та матеріалів доповідей на наукових конференціях [10-17].

Структура та обсяг дисертаційної роботи.

Дисертація складається зі вступу, чотирьох розділів, висновків, списку використаних джерел та додатків. Основний текст роботи охоплює 151 сторінку, що містить 52 рисунки, 41 формулу та 10 таблиць. Загальний обсяг дисертації становить 218 сторінок, включаючи список літератури зі 161 найменування на 19 сторінках та 5 додатків на 32 сторінках.

ОСНОВНИЙ ЗМІСТ ДИСЕРТАЦІЙНОЇ РОБОТИ

У вступі обґрунтовано актуальність теми дослідження та зв'язок роботи з науковими програмами, планами, темами; сформульовано предмет та об'єкт дисертаційної роботи, мету та завдання дослідження для досягнення поставленої мети. Визначено наукову новизну та аргументовано практичне значення одержаних результатів дослідження. Також, наведено списки опублікованих праць за тематикою дослідження та конференцій, на котрих було апробовано основні результати дисертаційної роботи.

У першому розділі проаналізовано завдання моніторингу навколишнього середовища: організацію спостережень, оцінювання та прогнозування забруднюючих речовин, розробку обґрунтованих рекомендацій для прийняття управлінських природоохоронних рішень. Обґрунтовано дослідження прогнозування параметрів забруднення атмосферного повітря на стаціонарних постах спостереження.

Основними об'єктами викидів забруднюючих речовин є теплове та енергетичне устаткування, промислові підприємства, сільське господарство, всі види транспорту, а особливо автотранспорт через постійне збільшення автотранспортних перевезень. Крім того, однією з основних причин забруднення атмосферного повітря є низький рівень оснащення вже працюючих джерел викидів пилогазоочисним обладнанням та відсутність установок по вимірюванню параметрів забруднення.

Регулювання викидів забруднюючих речовин у повітряне середовище стаціонарними джерелами здійснюється для найбільш поширених і небезпечних параметрів забруднення, серед яких шкідливими є викиди діоксиду сірки, діоксиду азоту, оксиду вуглецю, метану, аміаку, пилу, неметанових органічних сполук та інші. Перелік параметрів забруднення атмосферного повітря встановлюється Кабінетом Міністрів України і переглядається не рідше, ніж один раз на п'ять років.

Однією з важливих умов успішного прогнозування стану атмосферного повітря є наявність достовірної і повної інформації про виміряні протягом певного інтервалу часу параметри забруднення повітряного середовища. Та достатньо часто параметри забруднення атмосферного повітря містять пропуски, поява яких може бути зумовлена шумами (несприятливими погодними умовами); призупиненням вимірювань під час вихідних днів, свят; помилкою приладів вимірювання; пошкодженням носіїв інформації; виконанням недостатньої кількості вимірювань чи зумисним приховуванням. Тому необхідним є заповнення пропущених концентрацій параметрів забруднення атмосферного повітря.

Вибір методу заповнення пропусків істотно залежить від методу аналізу даних, який буде використовуватися надалі, оскільки заповнення пропусків може змінити структуру вибірки. Проаналізовано сучасні прості та складні методи заповнення пропущених атрибутів параметрів забруднення атмосферного повітря та обґрунтовано реалізацію простих методів з метою менших затрат часу на відновлення пропусків. Серед простих методів заповнення пропусків виділено методи регресійного моделювання, які містять: метод на основі машини опорних векторів (SVR), метод на основі лінійної регресії зі стохастичним градієнтним спуском (SGDr), адаптивний бустинг, дерево рішень та нейроподібну структуру моделі послідовних геометричних перетворень (НС МПГП), що зображені на рисунку 1.

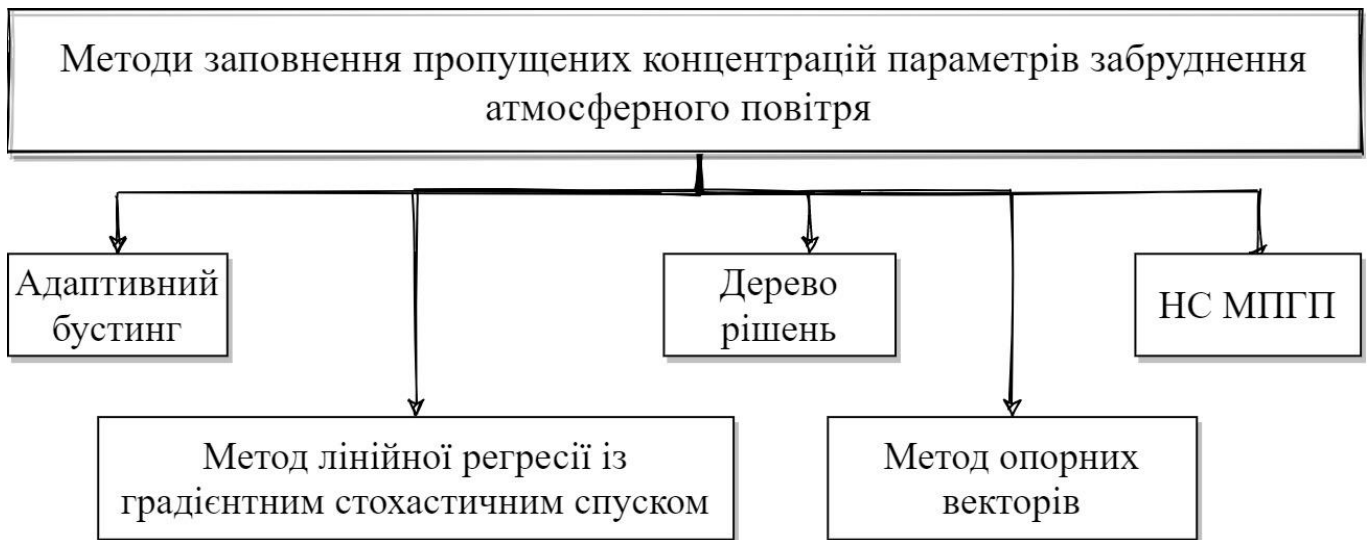


Рисунок 1 – Досліджені методи заповнення пропусків у даних моніторингу забруднення атмосферного повітря

Оскільки аналіз системи моніторингу довкілля показав, що однією із його функцій є прогнозування перспектив викидів забруднюючих речовин у атмосферу, виконано порівняння евристичних, аналітичних та статистичних підходів до прогнозування забруднення атмосферного повітря. Серед найпростіших методів розглянуто найвіщий прогноз, для порівняння зі складними, котрі можуть мати низьку квадратичну похибку, але все ж таки поступатися простому прогнозуванню найвіщим методом.

Також, проаналізувавши моделі прогнозування параметрів забруднення атмосферного повітря встановлено, що розробка адекватних математичних моделей повинна відображати зміни, які відбуваються в природному середовищі під впливом людської діяльності; своєчасно забезпечувати підсистеми моделювання якісною інформацією про стан природного середовища і про параметри функціонування техносфери; включати в себе ретроспективний аналіз існуючих прогнозів з метою коригування математичних моделей, на основі яких вони були виконані.

Теоретичною базою, що визначає концептуальну можливість прогнозування параметрів забруднення атмосферного повітря на підставі відомої передісторії є теорема Такенса, що стверджує: якщо часовий ряд породжується динамічною системою, тобто значення x_t є довільною функцією системи, існує така глибина занурення d , яка забезпечує однозначне передбачення наступного значення часового ряду. Отже, прогнозування часового ряду зводиться до задачі інтерполяції функції багатьох змінних. Тому прогнозування параметрів забруднення атмосферного повітря можна виконати за допомогою нейронних мереж завдяки відновленню невідомої функції по набору прикладів, заданих історією даного часового ряду. Згідно з теоремою Такенса, для деяких значень m існує залежність (1):

$$x_{t+m} = F(x_t, x_{t+1}, \dots, x_{t+m-1}) \quad (1)$$

Таким чином, після завершення процесу навчання нейронна мережа імовірно "вміє" обчислювати функцію F і для продовження часового ряду потрібно подати на вхід мережі h_{N-m+1}, \dots, h_N компоненти часового ряду і отримати на виході шукану величину - h_{N+1} .

Дані моніторингу стану атмосферного повітря, і довкілля в цілому, складаються з вимірних параметрів забруднення упродовж певного періоду з деякими однаковими часовими проміжками (мінімальний проміжок становить вимірювання двічі на день). Тому у дисертаційній роботі виконується дослідження методів прогнозування часових рядів.

До моделей часових рядів належать: нейронні мережі Adaline та Madaline, нейромережі зустрічного поширення, багатошарові перцептрони, нейронні мережі радіальних базисних функцій, рекурентні нейронні мережі, нейронні мережі узагальненої регресії (GRNN – General regression neural network) та нейроподібні структури, зокрема моделі послідовних геометричних перетворень (НС МПГП). Методи прогнозування часових послідовностей також включають в себе методи регресійного моделювання, як і в завданні заповнення пропусків у параметрах забруднення атмосферного повітря, які показані на рисунку 2.

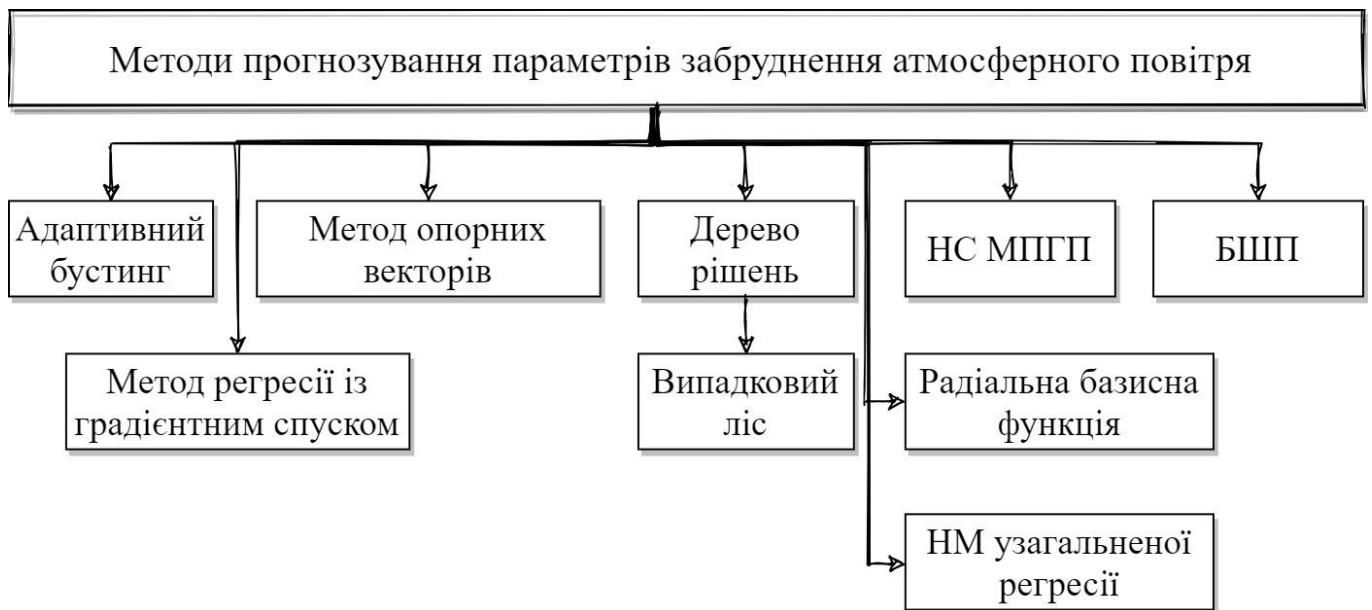


Рисунок 2 – Моделі та методи прогнозування параметрів забруднення повітря

Для даних моніторингу атмосферного повітря характерні такі особливості як швидка зміна в часі, не лінійність, наявність значних шумових складових та наявність пропущених компонент у параметрах забруднення. Всі ці особливості враховуються під час використання НС МПГП, оскільки її додаткові функції, зокрема можливість розкладу часових послідовностей на тренд і сукупність коливань, забезпечує додаткові можливості отримання важливої прогнозної інформації.

Концептуальна модель послідовних геометричних перетворень, що складає основу НС, має підвищену швидкодію та точність застосування. Однак вказана структура не використовувалася для розв'язку задач прогнозування в умовах невизначеності за рахунок наявних пропущених даних. Також не здійснювалось виділення лише тренду, де тренд в сумі складових є часовою послідовністю, що може бути ефективним саме в задачах прогнозування параметрів забруднення атмосферного повітря.

Отже, перспективними є дослідження існуючих та розробка нових методів прогнозування параметрів забруднення повітря використовуючи НС МПГП, в тому числі в умовах пропусків у даних моніторингу повітряного середовища.

У другому розділі розроблено методи розширення розмірності векторів вхідних вибірок даних (відповідно до висновків з теореми Ковера), які досліджено на реальних вимірних параметрах забруднення атмосферного повітря зі стаціонарного посту промислових викидів у Італії. Дані взяті з репозиторію UCI Machine Learning Repository, – одного з джерел наборів даних для використання в машинному навчанні.

Вибірка параметрів забруднення, що виміряні на датчиках, вбудованих в хімічний мультисенсорний пристрій для визначення якості повітря, включає в себе: концентрації оксидів вуглецю, не метанових вуглеводнів, бензолів, титанів, оксидів та діоксидів азоту, оксидів та діоксидів вольфраму, оксидів індію, температури повітря, абсолютних та відносних вологостей повітря. Вибірка складається з 9358 погодинних усереднених спостережень, з них 2373 вектори містять пропуски у параметрах забруднення атмосферного повітря, позначені значеннями «-200». Оскільки 25% векторів даних містять пропущені концентрації параметрів забруднення повітря, розроблено та досліджено методи заповнення пропусків на основі розширення векторів даних за допомогою введення додаткових атрибутів.

Вперше розроблено метод розширення вхідних ознак векторів даних моніторингу забруднення атмосферного повітря за допомогою виконання попереднього виділення кластерів, що відповідає гіпотезі компактності. Відповідно до неї, близьким за спільними характеристиками об'єктам в геометричному просторі ознак відповідають відокремлені множини точок, котрі мають властивості чіткого розподілу. Досліджувані дані моніторингу повітря відповідають таким положенням гіпотези компактності: множини різних векторів концентрацій пересікаються в порівняно невеликому числі точок, а границі класів мають порівняно плавну форму та не мають глибоких виступів в межі інших класів. В результаті різні вектори можуть розділятися досить простими гіперповерхнями за спільними в геометричному просторі ознаками: кутом нахилу вздовж прямої, на колі, по спіралі, на решітці і т.п.

Під час встановлення належності кожного вектора до відповідної компактної множини виникає можливість надання кожному з векторів тренувальної та тестової вибірок даних додаткового атрибуту – номера відповідного кластера. Згідно з теоремою Ковера нелінійне перетворення завдання класифікації векторів в простір більш високої розмірності підвищує ймовірність лінійної роздільності образів. Отже, досягнення ефекту кращого розділення класів виникає за рахунок лінеаризації простору реалізацій за рахунок розширення його розмірності. Очевидно, що часткова лінеаризація простору підвищує точність розв'язку не лише завдань класифікації, але і регресії, до яких належить відновлення пропущених концентрацій параметрів забруднення атмосферного повітря.

Таким чином, в основі методу розширення входів є попереднє виділення кластерів та встановлення належності вхідних векторів даних до цих кластерів для лінеаризації простору. Метод формування додаткових атрибутів складається з набору послідовності кроків, де першим кроком є аналіз методів кластеризації та вибір оптимального методу для виконання поставленого завдання. Для зменшення впливу «екстремальних» викидів на результат заповнення пропусків у даних виконується пошук аномалій за рахунок обчислення Евклідових відстаней до центрів кластерів та їх фільтрування.

Вибравши оптимальну кількість кластерів, виконується розподіл векторів вибірок параметрів забруднення атмосферного повітря по кластерах. Кластеризація векторів навчальної матриці виконується методом “к середніх” без врахування виходів на k компактних множин точок (кластерів). Слід зазначити, що відстані від векторів до центрів кластерів розраховуються за формулою Евклідової відстані (2):

$$\rho(x, x') = \sqrt{\sum_i^n (x_i - x'_i)^2}. \quad (2)$$

Вектори тестової вибірки підлягають кластеризації за допомогою їх віднесення до того кластеру навчальної вибірки, евклідова відстань до центрів мас якого є найменшою. Розширення розмірності векторів навчальної та тестової вибірок параметрів забруднення атмосферного повітря виконується за рахунок введення k додаткових входів, де k – кількість кластерів визначена раніше. Якщо вектор належить до k -го кластера, то k -ий додатковий вхід отримує значення 1. Решта додаткових ознак вхідних векторів заповнюються нулями.

Розроблений метод розширення входів застосовано на вибірці параметрів забруднення атмосферного повітря та апробовано на таких регресійних методах: Adaptive Boosting, лінійна регресія із застосуванням стохастичного градієнтного спуску (SGDr), машина опорних векторів із застосуванням регресії (SVR) та лінійна НС МПГП. Для кожного методу визначено похибки заповнення пропущених концентрацій діоксиду азоту (рис. 3.).

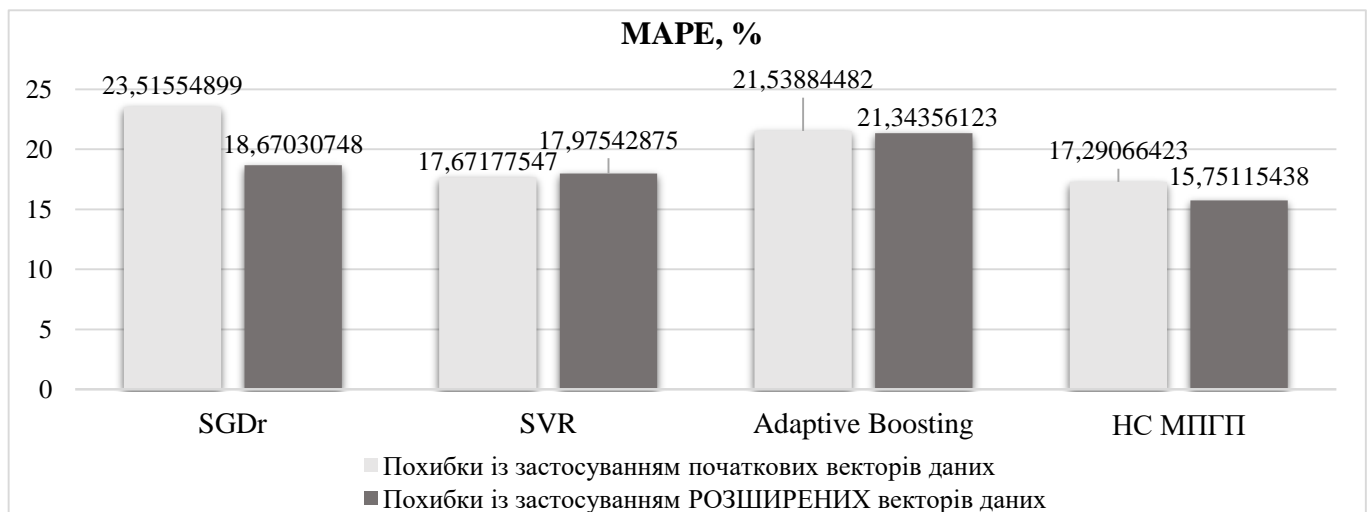


Рисунок 3 – Середні значення відносних похибок заповнення пропущених компонент параметру забруднення – діоксиду азоту

Реалізований метод відновлення пропущених концентрацій параметрів забруднення повітряного середовища на основі попереднього виділення кластерів забезпечив підвищення точності заповнення пропущених концентрацій для діоксиду азоту на 14,5 % відносно НС МПГП (рис. 5), а чадного газу на 2,5%.

Однак розроблений метод підвищення точності заповнення пропусків у даних моніторингу атмосферного повітря показав ефективність не у всіх випадках застосування, що пояснюється особливостями розподілу векторів даних окремих параметрів в просторі реалізацій. Наприклад, при розширенні вхідних ознак векторів параметрів забруднення повітря та при застосуванні методу на основі машини опорних векторів, середня точність прогнозування навіть погіршилась на 1,5 %.

Також, у другому розділі удосконалено метод функційного розширення входів Йох-Хан Пао. Оскільки, застосування запропонованих нелінійних функцій типу (ко)синусоїдальних не відповідає характеру нелінійності поверхні відгуку для завдань заповнення пропущених концентрацій у параметрах забруднення атмосферного повітря, розглянуто варіант функціонального розширення входів типу раціональних дробів. Їх використання забезпечує побудову нелінійних поверхонь в області вузлів інтерполяції і зменшує появу можливих викидів за їх межами.

Удосконалений метод функційного розширення вхідних ознак векторів Йох-Хан Пао передбачає виконання розрахунку раціональних дробів для кожної компоненти вхідних векторів вибірок параметрів забруднення атмосферного повітря за формулою (3):

$$c_i = 1 / (1 + x_i^2) \quad (3)$$

де x_i – початкові входи матриці.

Розширення векторів навчальної та тестової вибірок параметрів забруднення повітряного середовища виконується за допомогою введення обрахованих функцій (раціональних дробів) як додаткових вхідних ознак векторів даних, кількість яких є рівною кількості вхідних атрибутів векторів.

Удосконалений метод розширення входів також застосовано на вибірці параметрів забруднення атмосферного повітря та апробовано на тих регресійних методах, що і попередній розроблений метод. Для кожного методу визначено похибки заповнення пропущених концентрацій параметрів забруднення (рис. 4.).

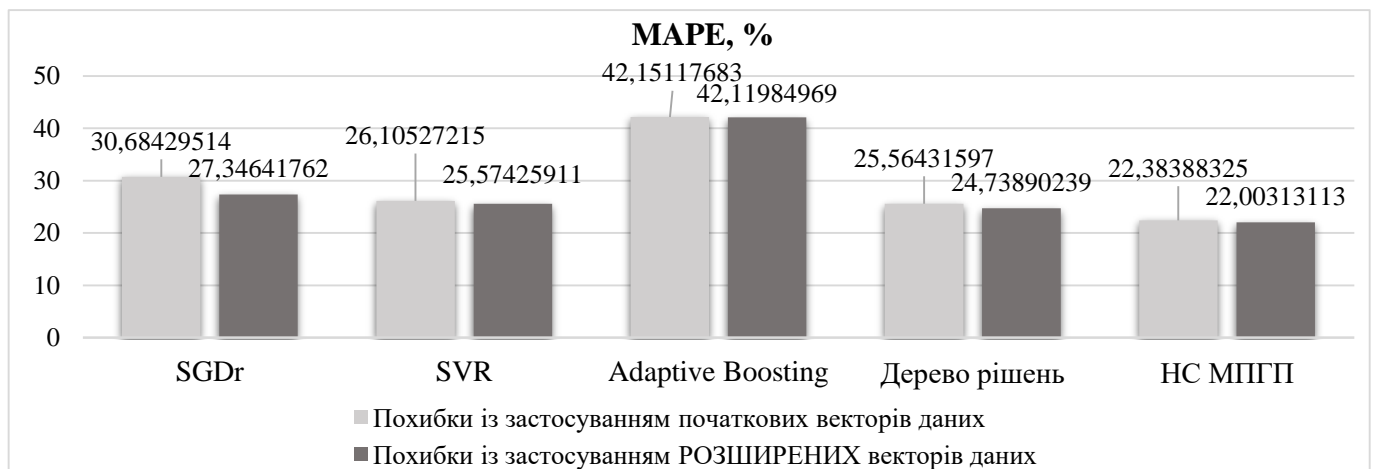


Рисунок 4 – Похибки заповнення пропущених компонент чадного газу

Удосконалений метод функційного розширення входів Йох-Хан Пао шляхом застосування раціональних дробів забезпечив зниження похибок заповнення пропущених концентрацій оксиду вуглецю на 2,6 % (рис. 4.), а діоксиду азоту на 6 %. Також досліджено, що цей метод розширення входів підвищує точність заповнення пропусків у даних моніторингу повітряного середовища для всіх апробованих регресійних методів, в тому числі для методу на основі машини опорних векторів.

Отже, у другому розділі, досліджено, що розроблені методи розширення входів підвищують точність заповнення пропущених концентрацій параметрів забруднення атмосферного повітря під час виконання всіх регресійних методів окрім машини опорних векторів через особливості побудови даного методу.

У третьому розділі реалізовано метод короткотермінового прогнозування параметрів забруднення атмосферного повітря на основі розробленого методу корекції похибки застосування за допомогою комітету нейроподібних структур різних типів. Також розвинуто метод побудови апроксимаційних поліномів шляхом ідентифікації їх коефіцієнтів за результатами навчання НС МПГП.

Для досліджень прогнозування параметрів забруднення повітряного середовища використано дані екологічного моніторингу зі стаціонарного посту на Бесарабській площі в м. Києві, взяті з електронного ресурсу центральної геофізичної обсерваторії імені Б. Срезневського. Систематичні спостереження виконані з періодичністю вимірювань 3-4 рази на добу. Вибірка параметрів забруднення повітря складається з п'яти шкідливих домішок, що входять в програму обов'язкового моніторингу якості повітряного середовища: оксид вуглецю (CO), діоксид азоту (NO₂), діоксид сірки (SO₂), формальдегід (CH₂O), пил.

Метод корекції похибки прогнозування побудований на врахуванні передумови, що відхилення від точних значень для нелінійної апроксимації в більшості точок відгуку є меншими ніж для лінійної. Для збільшення точності прогнозування розроблюваний метод реалізується на застосуванні комітету трьох НС різних типів (НС1, НС2 та НС3), кожна з яких послідовно виконує покладені на неї функції. Першим кроком, на якому базується реалізація методу, є передобробка даних за допомогою НС1 МПГП лінійного типу. Для цього виконується навчання НС1 з використанням векторів тренувальної вибірки виду $x_{i1}, \dots, x_{ij}, \dots, x_{in}, Y_i^{Real} \rightarrow Y_i^{Real}$, де Y_i^{Real} – відомий вихід тренувального вектора.

У випадку, якщо число нейронних елементів (НЕ) прихованого шару є рівним кількості входів вектора n , в режимі застосування на виході НС1 завжди отримується точне значення сигналу Y_i^{Real} . Зменшуючи число НЕ прихованого шару НС1 на одиницю, отримується так зване згладжене значення сигналу Y_i^{sReal} , яке не враховує останньої головної компоненти при перетвореннях з початкової системи координат в систему головних компонент. З отриманого згладженого сигналу Y_i^{sReal} формується нова матриця векторів за допомогою використання методу ковзних часових вікон. Величина сигналу Y_i^{sReal} може бути передбачена за допомогою НС2 МПГП нелінійного типу, зокрема з використанням поліноміальних, або RBF входів розширення. Після виконання навчання і застосування НС2 МПГП (лінійного типу) та НС3 RBF (нелінійного типу), розраховуються похибки прогнозування.

Якщо виконується умова, що похибка прогнозування нелінійною НС3 є меншою за похибку прогнозування лінійною НС2, тоді для довільної точки виконується співвідношення за допомогою наступної формули (4):

$$y_i = y_i^{RBF} - \alpha(y_i^{RBF} - y_i^{SGTM}) \quad (4)$$

де y_i^{SGTM} – значення сигналу, котре отримується на виході навченої НС2 МПГП лінійного типу;

α - коефіцієнт пропорційності, який підбирається експериментально і є однаковим для всіх точок вибірки.

Ілюстрація принципу корекції похибки для випадку одновимірного сигналу подається на рис. 5.

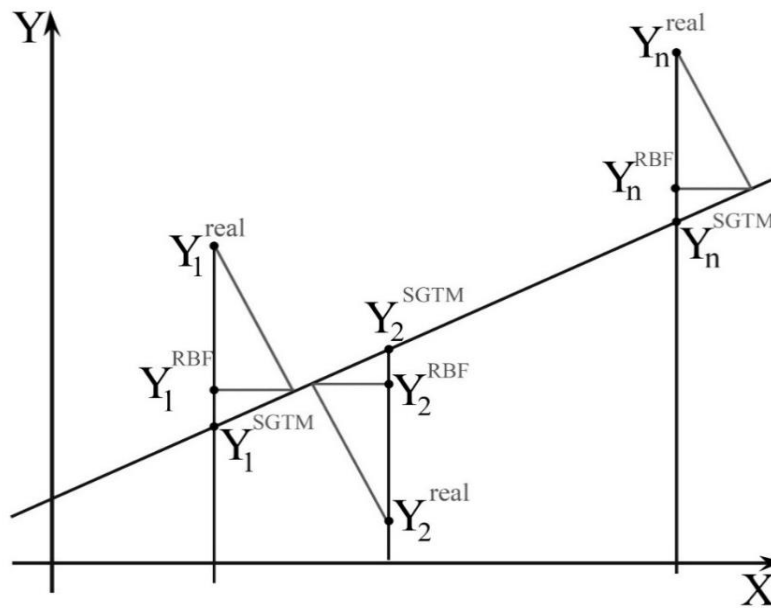


Рисунок 5 – Графічне зображення методу корекції похибки

Так як передбачається згладжене значення виходу Y_i^{sReal} , в якому в навчальній вибірці усувається остання, як правило шумова складова, точність передбачення вдається помітно підвищити. В результаті виникає можливість окремо зменшувати похибки різних знаків і помітно підвищити точність прогнозів. Ефективність розроблюваного методу підтверджується експериментально на даних моніторингу забруднення атмосферного повітря, взятих із стаціонарного посту спостереження у місті Києві, що описані раніше. Після розрахунку середньої похибки прогнозування на основі застосування розробленого методу доцільним є виконання оцінки його ефективності порівняно з іншими методами. Результати порівняння методу прогнозування на основі корекції похибки за рахунок використання комітет НС різних типів наведені в таблиці 1.

Таблиця 1. Похибки прогнозування показників чадного газу за допомогою НС

Похибка прогнозування	Метод 1 (лінійна НС)	Метод 2 (нелінійна НС + RBF)	Метод 3 (комітет НС)
MAPE, %	3,960715 %	3,671264 %	3,317151 %

Також, у третьому розділі отримав подальший розвиток метод побудови апроксимаційних поліномів шляхом ідентифікації їх коефіцієнтів за результатами навчання НС МПГП з багатьма виходами. Розвиток методу полягає у розширенні тестової матриці на один додатковий вектор та використанні НС МПГП з одним виходом для застосування на мобільних пристроях. Тестова матриця складається з векторів, що мають всі значення рівні нулю окрім одного атрибуту по діагоналі, значення якого рівне одиниці. Розмірність тестової матриці рівна розмірності вхідного вектора для навчання та одного додаткового горизонтального вектора, елементами якого є нулі. Слід зазначити, що додатковий вектор є першим рядком тестової матриці.

Передбачені під час навчання виходи дозволяють отримати матричний оператор коефіцієнтів ваг синаптичних зв'язків, котрі використовуються як коефіцієнти a_i для лінійного полінома (5):

$$y_i = a_0 + a_1 * x_1 + \dots + a_i * x_i \quad (5)$$

де коефіцієнти $a_0 = y_0$, $a_i = y_i - y_0$, а x_i – вхідні параметри тестової матриці, для яких потрібно спрогнозувати вихідний сигнал.

Отже, матричний оператор коефіцієнтів ваг синаптичних зв'язків забезпечує реалізацію прогнозування параметрів забруднення атмосферного повітря без застосування НС МПГП, що пришвидшує результати передбачення за рахунок застосування лінійного полінома, що зменшує навантаження операційної системи додатковими обчисленнями.

Таким чином, відбувається зменшення затрат оперативної пам'яті вбудованих та мобільних пристроїв під час прогнозування параметрів забруднення атмосферного повітря у 12,75 разів та зменшення часових затримок прогнозування у 5 разів порівняно з використанням НС МПГП (рис. 6.).



Рисунок 6 – Час застосування досліджуваних методів прогнозування

Отже, у третьому розділі розроблено метод корекції похибки короткотермінового прогнозування параметрів забруднення повітряного середовища в реальному часі на основі комбінету нейроподібних структур різних типів, що забезпечило зменшення похибки однокрокового прогнозування на 15% відносно НС МПГП та збільшило горизонт короткотермінового прогнозування на два дні.

Також розвинуто метод прогнозування параметрів забруднення атмосферного повітря, котрий базується на навчання лінійних нейроподібних структур моделі послідовних геометричних перетворень і побудови на їх основі матриці коефіцієнтів лінійних поліномів. Використання цього методу залежно від виду параметру забруднення атмосферного повітря та вибраного алгоритму для здійснення прогнозування забезпечує зниження часових затримок в режимі застосування, що є важливим для виконання прогнозів на мобільних пристроях та мікроконтролерах.

У четвертому розділі розроблено програмний засіб для прогнозування параметрів забруднення повітряного середовища, зокрема в умовах пропусків у даних, який складається з набору бібліотек реалізацій розроблених, удосконалених та розвинутих у роботі методів; системи збору, обробки та аналізу даних; бази даних для збереження результатів; та користувацького інтерфейсу для використання на мобільних пристроях. Розроблений програмний засіб для прогнозування та заповнення пропущених параметрів забруднення атмосферного повітря реалізовано двома об'єктно-орієнтованими мовами програмування, які здатні продуктивно виконувати вихідний код програми - Python та Java.

Реалізація нейромережових та нейроподібних методів машинного навчання для заповнення пропусків та прогнозування даних (описаних у попередніх розділах) здійснена мовою Python, котра є високорівневою мовою програмування з мінімалістичним синтаксисом та низкою безкоштовних бібліотек методів машинного навчання. У розробленому Python-фреймворку виконується налаштування параметрів реалізованих методів машинного навчання для прогнозування концентрацій забруднення атмосферного повітря в тому числі в умовах пропусків у даних, їх навчання та застосування. Методи розроблені використовуючи пакет Anaconda Python – повністю налаштованому середовищі програмування з попередньо установленими бібліотеками різних версій. Python-фреймворк створено за допомогою таких сучасних бібліотек машинного навчання, як: Theano та scikit-learn (якими реалізовано методи регресії та всю внутрішню та низькорівневу логіку роботи нейромереж та нейроподібних структур).

Python-фреймворк складається з трьох модулів: розробка моделей методів машинного навчання з налаштуванням їх параметрів, виконання навчання та прогнозування чи заповнення пропусків. У Python-проекті реалізовано такі моделі нейромережових та нейроподібних методів: RandomForestRegressor, AdaBoostRegressor, SGDRegressor, MLPRegressor, DecisionTreeRegressor, SVR, AverageFill, RBFSSample та NaivePredict. Також розроблено моделі нейроподібних методів SGTModel, SGTModel_1, SGTModel_x_1, котрі відрізняються налаштованими параметрами для використання в режимах навчання та застосування. SGTModel_1 та SGTModel_x_1 мають налаштовані всі ті ж самі параметри, що і SGTModel, окрім кількості головних компонент, що повинні бути використані під час здійснення навчання алгоритму. Для моделі SGTModel кількість головних компонент дорівнює кількості вхідних векторів параметрів забруднення атмосферного повітря. Для моделі SGTModel_1 кількість головних компонент задається рівною одиниці, а модель SGTModel_x_1 містить кількість головних компонент меншу на одиницю від кількості вхідних векторів даних. Всі три розроблені моделі відрізняються призначенням навчання та застосування. Модель SGTModel використовується для здійснення прогнозування параметрів забруднення повітряного середовища в умовах пропущених даних. Модель SGTModel_1 використовується для виділення тренду забруднення атмосферного повітря деяким параметром, а SGTModel_x_1 здійснює згладжування вихідного вектора значень для лінеаризації поверхні відгуку, що підвищує точність прогнозування.

Основним недоліком мови Python є відсутність підтримки багатопотоковості на рівні процесорів, що може збільшувати часові затримки прогнозування параметрів

забруднення атмосферного повітря на мобільних пристроях та мікроконтролерах. Тому для реалізації розроблених методів прогнозування параметрів забруднення повітряного середовища в умовах пропусків у даних використано мову Java, котра підтримує багатопотоковість та є універсальною для програмування комп'ютерних пристроїв (мобільних телефонів, контролерів, планшетів і т.д.). Для здійснення комунікації між розробленими методами двох мов, у Java-фреймворку реалізовано клас з функціями виклику методів з Python-фреймворку. Загальна архітектура розробленого програмного засобу зображена на рис. 7.

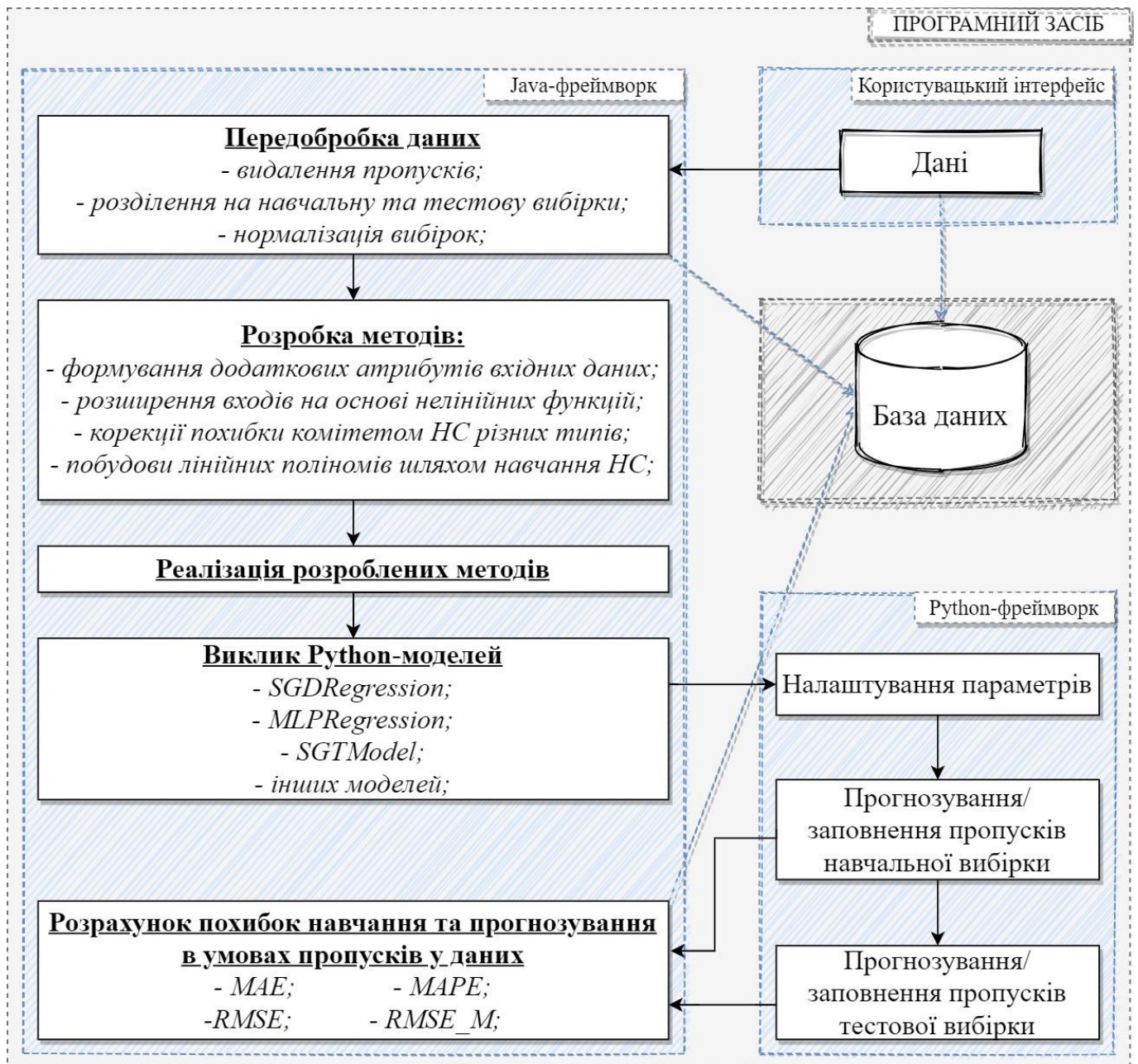


Рисунок 7 – Загальна архітектура розробленого програмного засобу

Java-фреймворк включає в себе реалізацію методу формування додаткових атрибутів вхідних векторів даних на основі попереднього виділення компактних множин точок; методу розширення входів Йох-Хан Пао на основі обернено-пропорційних функцій; методу короткотермінового прогнозування параметрів

забруднення повітряного середовища на основі корекції похибки комітетом нейроподібних структур різних типів та методу побудови апроксимаційних поліномів шляхом ідентифікації їх коефіцієнтів за результатами навчання НС МППП.

Розроблений програмний засіб для підвищення точності та швидкодії прогнозування в умовах пропусків у даних складається з наступних модулів:

- Модуль 1.* Передобробка даних;
- Модуль 2.* Набір розроблених методів взаємодії двох фреймворків;
- Модуль 3.* Класи з методами для роботи з таблицями даних;
- Модуль 4.* Реалізація методів прогнозування;
- Модуль 5.* Класи з розробленими методами підвищення точності та швидкодії прогнозування параметрів забруднення атмосферного повітря;
- Модуль 6.* Класи з методами обрахунку точності та швидкодії реалізованих нейромережових та нейроподібних методів машинного навчання для здійснення прогнозування параметрів забруднення атмосферного повітря, в тому числі в умовах пропусків у даних моніторингу повітряного середовища.

Модуль передобробки даних включає в себе такі класи: *DataFilter* (виконує фільтрування вибірки вхідних даних та видаляє вектори даних моніторингу атмосферного повітря, де є пропущені концентрації параметрів забруднення повітряного середовища); *Normalization* (здійснює пошук максимального значення по стовпцях навчальної вибірки даних та нормалізує тренувальну та тестову матриці); *Scaling* (клас з методами виконання дій над таблицями, котрі використовуються під час здійснення нормалізації вибірок даних) та *Splitting* (містить методи розділення вхідної вибірки даних на навчальну та тестову). Модуль взаємодії двох програмних проектів містить в собі клас *PythonWrapper*, котрий складається з методів виклику нейромережових та нейроподібних алгоритмів прогнозування параметрів забруднення повітряного середовища та заповнення пропусків у даних моніторингу атмосферного повітря з Python-фреймворку.

Модуль набору розроблених методів для підвищення точності та швидкодії прогнозування параметрів забруднення атмосферного повітря, в тому числі в умовах пропусків у даних моніторингу повітряного середовища складається з трьох пакетів *errorcorrection* (корекція похибки прогнозування параметрів забруднення атмосферного повітря), *acceleration* (зменшення часу затримки застосування нейромережових алгоритмів прогнозування/заповнення пропусків концентрацій параметрів забруднення повітря) та *inputextension* (розширення вхідних ознак векторів параметрів забруднення), котрий включає в себе два підпакети – по одному для кожного методу розширення: *clusterizedextension* та *functionextension*. Підпакет *functionextension* має клас *nonlinearextension*. Підпакет *clusterizedextension* містить в собі такі класи: *Clusterization*, *DistanceCalculation* та *ClusterizedExtension*.

Отже, в останньому розділі дисертаційної роботи реалізовано програмний засіб, котрий включає в себе два фреймворки, що складаються з нейроподібних та нейромережових методів машинного навчання (Python-фреймворк) та бібліотек програмних засобів розроблених методів підвищення точності та швидкодії прогнозування параметрів забруднення повітряного середовища, в тому числі в умовах неповноти даних (Java-фреймворк).

ВИСНОВКИ

У роботі на основі виконаних досліджень розв'язано актуальне наукове завдання, – розроблено нейроподібні методи та засоби з неітеративним навчанням та підвищеною точністю прогнозування параметрів забруднення атмосферного повітря, в тому числі в умовах пропусків у даних моніторингу повітряного середовища.

Під час вирішення завдань дисертаційної роботи отримано наступні основні результати:

1. Проаналізовано актуальні задачі моніторингу навколишнього середовища та особливості існуючих методів і засобів для здійснення функцій прогнозування параметрів забруднення атмосферного повітря, зокрема в умовах пропусків у даних моніторингу повітряного середовища.
2. Вперше розроблено метод короткотермінового прогнозування параметрів забруднення атмосферного повітря за допомогою комітету лінійної та нелінійної нейроподібних структур для часткового коректування окремо додатних і від'ємних відхилень від точних значень, що забезпечило зменшення похибки прогнозування на 15% та збільшило горизонт прогнозування на два дні.
3. Вперше розроблено метод формування додаткових ознак вхідних векторів даних, що забезпечило підвищення точності заповнення пропущених значень параметрів забруднення атмосферного повітря за допомогою нейроподібної структури моделі послідовних геометричних перетворень на 2,5-14,5% в залежності від виду параметру. Встановлено, що розроблений метод під час заповнення пропусків у даних моніторингу атмосферного повітря за допомогою машини опорних векторів зменшує точність заповнення пропущених значень на тестових даних на 1-2% залежно від виду параметру забруднення повітряного середовища.
4. Розвинуто метод побудови апроксимаційних поліномів шляхом ідентифікації їх коефіцієнтів за результатами навчання відповідних нейроподібних структур, що забезпечило зменшення затрат пам'яті під час прогнозування параметрів забруднення атмосферного повітря в 12,75 разів.
5. Удосконалено метод функційного розширення входів Йох-Хан Пао шляхом застосування раціональних дробів, що забезпечило зниження похибок заповнення пропущених значень параметрів забруднення атмосферного повітря за допомогою нейроподібної структури моделі послідовних геометричних перетворень в режимі використання на невідомих при навчанні даних на 2,6-6% в залежності від виду параметру.
6. Розроблено програмний засіб з набором бібліотек реалізацій методів прогнозування параметрів забруднення повітряного середовища на мобільних та вбудованих пристроях, зокрема в умовах пропусків у даних моніторингу атмосферного повітря.

СПИСОК ПУБЛІКАЦІЙ ЗА ТЕМОЮ ДИСЕРТАЦІЙНОЇ РОБОТИ

1. Mishchuk O. Development of the method of forecasting the atmospheric air pollution parameters based on error correction by neural-like structures of the model of successive geometric transformations // *Technology Audit and Production Reserves*. — 2019. — № 6/2(50). — P. 22–26.
2. Mishchuk O. The Accelerated Method of Filling Gaps in Data Using a Linear SGTM Neural-Like Structure / O. Mishchuk, R. Tkachenko, V. Pohrebennyk // *International Journal of Science and Engineering Investigations (IJSEI)*. – 2019. – №8(91). – P. 154-159.
3. Міщук О. С. Нейронна мережа з комбінованою апроксимацією поверхні відгуку / О. С. Міщук, П. Б. Вітинський // *Наукові вісті КПП: міжнародний науково-технічний журнал*. – 2018. – № 2. – С. 18-24.
4. Міщук О. С. Методи оброблення та заповнення пропущених параметрів у даних екологічного моніторингу / О.С. Міщук, Р.О. Ткаченко // *Науковий вісник НЛТУ України*. – 2019. – №29(6). – С. 119-122. doi: 10.15421/40290623
5. Міщук О. С. Багатокрокове прогнозування тренду показників забруднення атмосферного повітря // *Науковий вісник НЛТУ України*. – 2019. – №29(8). – С. 142-146.
6. Mishchuk O. Missing Data Imputation Through SGTM Neural-Like Structure for Environmental Monitoring Tasks / O. Mishchuk, R. Tkachenko, I. Izonin // *Advances in Computer Science for Engineering and Education II. – International Conference on Computer Science, Engineering and Education Applications, ICCSEEA 2019. – Springer. – Vol. 938. – P. 142-151. doi: 10.1007/978-3-030-16621-2_13 (Scopus)*
7. Izonin I. SGD-Based Wiener Polynomial Approximation for Missing Data Recovery in Air Pollution Monitoring Dataset / I. Izonin, M. Greguš, R. Tkachenko, M. Logoyda, O. Mishchuk, Y. Kynash // *Advances in Computational Intelligence. – 15th International Work-Conference on Artificial Neural Networks, IWANN 2019. – Springer. – Vol. 11506. – P. 781-793. doi: 10.1007/978-3-030-20521-8_64 (Scopus)*
8. Tkachenko R. A Non-Iterative Neural-Like Framework for Missing Data Imputation / R. Tkachenko, O. Mishchuk, I. Izonin, N. Kryvinska, R. Stoliarchuk // *Procedia Computer Science. – The 14th International Conference on Future Networks and Communications, FNC 2019. – Springer. – Vol. 155. – P. 319-326. doi: 10.1016/j.procs.2019.08.046 (Scopus)*
9. Izonin I. Recovery of Incomplete IoT Sensed Data using High-Performance Extended-Input Neural-Like Structure / I. Izonin, R. Tkachenko, N. Kryvinska, K. Zub, O. Mishchuk, T. Lisovych // *Procedia Computer Science. — International Workshop on Digitalization and Servitization within Factory-Free Economy, D&SwFFE-2019. — Elsevier. — Vol. 160. — P. 521-526. (Scopus)*

10. Міщук О. С. Нейроподібні структури моделі геометричних перетворень з комбінованою апроксимацією поверхні відгуку // Інтелектуальні системи прийняття рішень і проблеми обчислювального інтелекту (ISDMCI-2018) : матеріали XIV-ї міжнародної наукової конференції, Залізний Порт, 21-27 травня 2018. – Херсон. – С. 87-89.
11. Міщук О. С. Нелінійне розширення входів нейронної структури моделі послідовних геометричних перетворень // Обчислювальні методи і системи перетворення інформації (ОМІСПІ-2018) : збірник праць V-ї науково-технічної конференції, Львів, 4-5 жовтня 2018. – Львів. – С. 126-129.
12. Міщук О. С. Відновлення пропусків у даних моніторингу забруднення повітря за допомогою нейронної структури моделі послідовних геометричних перетворень // Комп'ютерне моделювання та оптимізація складних систем (КМОСС-2018) : матеріали IV-ї міжнародної науково-технічної конференції, Дніпро, 1-2 листопада 2018. – Дніпро. – С. 260-261.
13. Mishchuk O. Expansion of Neural-like structures inputs using combined approximation / O. Mishchuk, R. Tkachenko // Computer and Information Systems and Technologies (CSITIC-2019) : Proceedings of the IIIrd International Scientific Conference, Kharkiv, 23-24 April 2019. – Kharkiv. – P. 29-32.
14. Mishchuk O. One-step Prediction of Air Pollution Control Parameters using Neural-Like Structure Based on Geometric Data Transformations / O. Mishchuk, R. Tkachenko // Electronics and Information Technologies (ELIT-2019) : Proceedings of the XIth International Scientific and Practical Conference, Lviv, 16-18 September 2019. – Lviv. – P. 192-197.
15. Міщук О. С. Прогнозування параметрів забруднення атмосферного повітря за допомогою лінійних нейроподібних структур // Побудова інформаційного суспільства: ресурси і технології: матеріали XVIII-ї міжнародної науково-практичної конференції, Київ, 19-20 вересня 2019. – Київ. – С. 275-278.
16. Mishchuk O. Neural network method of forecasting the air pollution trend by carbon monoxide // Information Technologies and Automation (ITA-2019) : Proceedings of the XIIth International Scientific Conference, Odesa, 17-18 October, 2019. – Odesa. – P. 101-102.
17. Міщук О. С. Підвищення точності прогнозування параметрів забруднення повітря // Комп'ютерне моделювання та оптимізація складних систем (КМОСС-2019) : матеріали V-ї міжнародної науково-технічної конференції, Дніпро, 6-8 листопада 2019. – Дніпро. – С.129-130.

АНОТАЦІЇ

Мищук Олександра Сергіївна. Нейроподібні методи та засоби прогнозування параметрів забруднення атмосферного повітря. На правах рукопису.

Дисертаційна робота на здобуття наукового ступеня кандидата технічних наук за спеціальністю 05.13.23 – системи та засоби штучного інтелекту. Національний університет “Львівська політехніка”, Львів, 2020.

У роботі розв’язано актуальне наукове завдання розроблення методів та засобів прогнозування параметрів забруднення повітряного середовища підвищеної точності та швидкодії, зокрема в умовах частково пропущених їх концентрацій на основі нейроподібних структур моделі послідовних геометричних перетворень.

Для підвищення точності та розширення горизонту прогнозування параметрів забруднення атмосферного повітря розроблено метод прогнозування часових послідовностей на основі комітету трьох нейроподібних структур різних типів. Для підвищення точності заповнення пропущених концентрацій параметрів забруднення атмосферного повітря розроблено метод введення додаткових ознак за допомогою попередньої кластеризації вхідних векторів та розвинуто метод функціонального розширення входів Йох-Хан Пао із використанням раціональних дробів для зниження похибок застосування на невідомих при навчанні даних.

Також для підвищення швидкодії прогнозування параметрів забруднення атмосферного повітря розроблено метод побудови апроксимаційних поліномів шляхом ідентифікації їх коефіцієнтів за результатами навчання лінійної нейроподібної структури моделі послідовних геометричних перетворень. Розроблено програмний засіб, що складається з реалізації методів підвищення точності та швидкодії прогнозування параметрів забруднення повітряного середовища та допоміжних пакетів обробки даних, завантаження і запису масивів даних, їх масштабування, визначення похибок навчання та застосування.

Ключові слова: моніторинг довкілля, параметри забруднення, часові ряди, нейроподібні структури моделі послідовних геометричних перетворень, пропущені параметри у даних, кластеризація, прогнозування тренду.

Мищук Александра Сергеевна. Нейроподобные методы и средства прогнозирования параметров загрязнения атмосферного воздуха. На правах рукописи.

Диссертационная работа на соискание ученой степени кандидата технических наук по специальности 05.13.23 – системы и средства искусственного интеллекта. Национальный университет "Львовская политехника", Львов, 2020.

В работе решена актуальная научная задача разработки методов и средств прогнозирования параметров загрязнения воздушной среды повышенной точности и быстродействия, в частности в условиях частично пропущенных концентраций на основе нейроподобных структур модели последовательных геометрических преобразований.

Для повышения точности и расширения горизонта прогнозирования параметров загрязнения атмосферного воздуха разработан метод прогнозирования временных последовательностей на основе комитета трех нейроподобных структур различных типов. Для повышения точности заполнения пропущенных концентраций параметров загрязнения атмосферного воздуха разработан метод введения дополнительных признаков с помощью предварительной кластеризации входных векторов и развит метод функционального расширения входов Йох-Хан Пао с использованием рациональных дробей для снижения погрешностей применения на неизвестных при обучении данных.

Также для повышения быстродействия прогнозирования параметров загрязнения атмосферного воздуха разработан метод построения аппроксимационных полиномов путем идентификации их коэффициентов по результатам обучения линейной нейроподобной структуры модели последовательных геометрических преобразований. Разработано программное средство, которое состоит из реализации методов повышения точности и быстродействия прогнозирования параметров загрязнения воздушной среды и вспомогательных пакетов обработки данных, загрузки и записи массивов данных, их масштабирование, определения погрешностей обучения и применения.

Ключевые слова: мониторинг окружающей среды, параметры загрязнения, временные ряды, нейроподобные структуры модели последовательных геометрических преобразований, пропущенные параметры в данных, кластеризация, прогнозирование тренда.

Mishchuk Oleksandra Sergijivna. Neural-like methods and tools to forecast the parameters of atmospheric air pollution. *Manuscript copyright.*

Thesis for a Candidate Degree in Engineering in the specialty 05.13.23 – systems and tools of artificial intelligence. Lviv Polytechnic National University, Lviv, 2020.

The current scientific problem of developing methods and means for predicting the parameters of air pollution, in particular, under partially missed parameters in atmospheric air monitoring data based on neural-like structures of a successive geometric transformations model for making informed recommendations on environmental management decisions, was solved.

Since filling in the missed concentrations of pollution parameters in atmospheric air monitoring data using artificial neural networks or neural-like structures is due to the performance of their predictions, so the second section proposed developing methods to improve the accuracy of the output values by expanding the inputs of the data samples. To increase the likelihood of recovering missing parameters in atmospheric air monitoring data, a method has been developed for introducing additional features of input vectors by preliminary isolating compact sets of points, and a method for expanding Yokhan-Pao inputs with additional introducing inversely proportional quadratic functions to reduce errors in the application mode for unknown for learning mode data has been developed.

The algorithm of recovery of the missed concentrations of air pollution parameters using the developed method of introducing additional features of the input vectors by pre-allocation of the clusters ensured the increase of the accuracy of the filling of the missed concentrations by 8,5 % relative to the neural-like structure of SGTm depending on the type of parameter. But it is investigated that the developed method does not work for all the analyzed algorithms. For example, by expanding the input characteristics of the vectors of air pollution parameters and applying the algorithm of reference vectors, the prediction accuracy has deteriorated for nitrogen dioxide by 1 % and for carbon monoxide by 2 %. The method of widening Yokhan-Pao inputs by applying inverse proportional quadratic functions was also developed, which reduced the error in the mode of use on unknown data by 4,3 %, and investigated that this method also reduces the error of prediction of air pollution parameters for all methods of recovery except for the vector reference method. In the case of this method with extended input inverse-proportional quadratic functions, the prediction error increases..

To increase accuracy and the forecast horizon of atmospheric air pollution parameters, a method for real-time prediction of time sequences based on a committee of three neural-like structures of various types using error correction of multi-step forecasting has been developed. The error correction was performed provided that the prediction error of the nonlinear neural structure was less than the prediction error of the linear neural structure. The use of error correction reduced the error by 15% and increased the forecast horizon by two days.

Also, to increase the speed of functioning in real time, in the dissertation research was developed a method for restoring gaps of individual components of multidimensional data vectors by identifying the coefficients of linear polynomials identical in their functions to the trained neural-like structure of a sequential geometric transformations model. It is proved experimentally that the method of fast filling of the gaps, which is based on the application of the committee of different linear neural structures of the model of successive geometric transformations and construction on their basis of the matrix of coefficients of linear polynomials shows a better time of application than the other methods analyzed. On the example of nitrogen dioxide, it is found that the time of filling in the missing parameters by the developed method is 1.23 milliseconds, which is 2-10 times faster than other methods taken for comparison.

A software tool was developed that consists of packages of methods for predicting atmospheric air pollution parameters that were implemented during research, including partially missed pollution parameters in air monitoring data and auxiliary packages. Auxiliary packages consist of data processing methods that contain classes for loading and writing data arrays, scaling them, working with tables determining prediction and restoration errors, etc.

Key words: environmental monitoring, pollution parameters, time series, neurosimilar structures of the sequential geometric transformations model, missing parameters in the data, clustering, trend forecasting.

Підписано до друку 24.02.2021 р.
Формат 60×84/16. Папір офсетний.
Друк цифровий. Умовн. друк. арк. 0,89.
Наклад 110 прим.

Видавництво Тараса Сороки
79026, м.Львів, вул. Володимира Великого, 2
Свідоцтво державного реєстру: серія ЛВ №17

Надруковано:
ФОП Сорока Т.Б.