

УДК 621.326

Матичак А. – ст. гр. СПс-41

Тернопільський національний технічний університет імені Івана Пулюя

РОЗРОБКА ПРОГРАМНОЇ СИСТЕМИ МОНІТОРИНГУ ТА ПРОГНОЗУВАННЯ РІВНЯ ЗАБРУДНЕННЯ ПОВІТРЯ З ВИКОРИСТАННЯМ МЕТОДІВ МАШИННОГО НАВЧАННЯ

Науковий керівник: к.ф.-м.н., доцент Цебрій О. Р.

Matychak A.

Ternopil Ivan Puluj National Technical University

DEVELOPMENT OF A SOFTWARE SYSTEM FOR MONITORING AND PREDICTING AIR POLLUTION LEVELS USING MACHINE LEARNING METHODS

Supervisor: PhD in Physics and Mathematics, Associate Professor Tsebrii O.

Ключові слова: прогнозування, забруднення повітря, мультимодальні моделі
Keywords: prediction, air pollution, multimodal models

Прогнозування рівня забруднення повітря є важливою задачею для забезпечення екологічної безпеки та покращення якості життя населення. Забруднення атмосфери має складну природу та залежить від великої кількості факторів, серед яких метеорологічні умови, транспортне навантаження, промислові викиди та географічні особливості. Традиційні статистичні методи часто не здатні врахувати складні нелінійні залежності між цими факторами, що обмежує їх точність. У цьому контексті методи машинного навчання дозволяють будувати більш точні моделі, здатні враховувати багатовимірність та динамічність даних. Сучасні підходи до прогнозування якості повітря базуються на використанні різних типів даних, включаючи часові ряди показників забруднення (PM_{2.5}, PM₁₀, NO₂, CO), метеорологічні дані, а також зображення або супутникові дані. Мультимодальні моделі дозволяють інтегрувати ці різноманітні джерела інформації в єдину систему, що значно підвищує якість прогнозування. Зокрема, комбінація рекурентних нейронних мереж для аналізу часових рядів та згорткових мереж для обробки зображень дозволяє враховувати як часову, так і просторову структуру даних.

Важливим етапом є попередня обробка даних, яка включає очищення, нормалізацію, синхронізацію різних джерел та заповнення пропущених значень. Оскільки дані можуть надходити з різних сенсорів і мати різну частоту оновлення, необхідно забезпечити їх узгодження у часі та просторі. Це дозволяє сформулювати єдине представлення вхідних даних для подальшого навчання моделей.

Для побудови прогнозних моделей широко використовуються алгоритми, такі як LSTM, GRU, градієнтний бустинг та глибокі нейронні мережі. Особливу ефективність демонструють гібридні архітектури, які поєднують різні типи моделей для обробки окремих модальностей даних. Наприклад, LSTM може використовуватись для аналізу часових залежностей, тоді як CNN — для обробки супутникових зображень або карт забруднення.

Окрім цього, важливим є врахування просторових залежностей між різними точками спостереження. Для цього застосовуються підходи, що враховують

географічну структуру даних, зокрема графові нейронні мережі або просторові згорткові моделі. Це дозволяє враховувати вплив сусідніх регіонів та поширення забруднення у просторі.

Експериментальні результати показують, що використання мультимодальних підходів дозволяє підвищити точність прогнозування порівняно з моделями, які використовують лише один тип даних. Основними метриками оцінювання виступають MAE, RMSE та коефіцієнт детермінації R^2 , що дозволяє комплексно оцінити якість прогнозів [1]. Разом з тим, залишаються відкритими питання, пов'язані з якістю та доступністю даних, а також обчислювальною складністю моделей. Тому актуальними є задачі оптимізації моделей, зменшення їх розміру та забезпечення можливості роботи у реальному часі.

Додатково важливим є питання інтеграції різних джерел даних у єдину узгоджену модель. Мультимодальні підходи передбачають об'єднання структурованих числових даних із сенсорів, метеорологічних показників та неструктурованих даних, таких як супутникові зображення або карти. Для цього використовуються механізми злиття ознак на різних рівнях — ранньому, проміжному або пізньому, що дозволяє ефективно поєднувати інформацію з різних модальностей. Вибір стратегії злиття безпосередньо впливає на якість моделі та її здатність узагальнювати складні залежності між факторами забруднення.

Окрему увагу приділяють побудові адаптивних моделей, здатних враховувати змінність середовища у часі. Рівень забруднення повітря може суттєво змінюватися залежно від сезонних факторів, погодних умов та людської діяльності. Тому сучасні системи використовують механізми онлайн-навчання або періодичного донавчання моделей на нових даних. Це дозволяє підтримувати актуальність прогнозів і забезпечувати їх точність у довгостроковій перспективі без необхідності повного перенавчання моделі.

Також важливим напрямом є інтерпретованість моделей, що дозволяє зрозуміти внесок окремих факторів у формування прогнозу. Використання методів пояснюваного машинного навчання дає можливість визначити, які саме змінні мають найбільший вплив на рівень забруднення в конкретний момент часу. Це є критично важливим для прийняття управлінських рішень, оскільки дозволяє не лише прогнозувати стан повітря, але й ідентифікувати основні джерела забруднення та оцінювати ефективність екологічних заходів.

Таким чином, поєднання мультимодальних підходів, сучасних методів машинного навчання та ефективної обробки даних створює передумови для розробки точних і надійних систем прогнозування забруднення повітря. Такі системи можуть бути використані для підтримки екологічного моніторингу, прийняття управлінських рішень та інформування населення про стан довкілля.

1. Zhang, Z., Wang, J. (2022). Deep learning for air quality prediction: A survey. — *Environmental Modelling & Software*.
2. Li, X., Peng, L., Yao, X. (2023). Multimodal deep learning for air pollution forecasting. — *Atmospheric Environment*.
3. Gao, J., Sun, Y., Li, H. (2024). Explainable multimodal deep learning for air quality prediction. — *Environmental Science and Pollution Research*.