

АНАЛІЗ ВЗАЄМОВПЛИВІВ РЯДІВ ДАНИХ МОНІТОРИНГУ СТАНУ АТМОСФЕРНОГО ПОВІТРЯ З ВИКОРИСТАННЯМ МЕТОДУ БАЙЄСІВСЬКОГО МОДЕЛЮВАННЯ

Вінницький національний технічний університет

Анотація

Продemonстровано ефективність застосування методу байєсівського моделювання для задачі аналізу взаємовпливів даних моніторингу стану атмосферного повітря, отриманих з різних станцій моніторингу певного регіону. Запропоновано алгоритм застосування цього методу для визначення як зміни показників на одній із станцій моніторингу можуть впливати на відповідні дані, отриманих з сусідніх станцій. Наведено приклад застосування для даних моніторингу стану атмосферного повітря з станцій у Вінницькій області за 2022-2024 роки від мережі громадського моніторингу EcoCity.

Ключові слова: машинне навчання, моделювання, стан атмосферного повітря, системний аналіз, байєсівське моделювання, часові ряди, природне середовище, EcoCity.

Abstract

This paper provides an overview of the effectiveness of Bayesian modeling-based method for identifying the interdependencies of air quality indicators obtained from various monitoring stations within a specific region. An algorithm for applying this method is proposed. It allows to determine how data changes at one station may influence the corresponding data changes at neighboring stations. An example of its application is provided. It is based on the air quality monitoring data from stations in the Vinnytsia region for 2022–2024 period, provided by the EcoCity public monitoring network.

Keywords: machine learning, modelling, air quality, system analysis, Bayesian modelling, time series, natural environment, EcoCity.

Вступ

Під час аналізу часових рядів стану атмосферного повітря і різних природних середовищ загалом існує складність моделювання впливу забруднюючих речовин та їх розповсюдження в межах певної географічної зони. Зазвичай, для розв'язання подібних задач застосовують моделі дифузії забруднення: Ейлерові, Гауссові або Лагранжеві моделі, що використовуються для опису розповсюдження забруднюючих речовини у повітрі та дозволяють визначити межі забруднення в часі та просторі [1, 2]. Проте, ці традиційні підходи не завжди дозволяють отримати точну інформацію щодо того як зміна значень показників на одній станції моніторингу можуть впливати на зміну відповідних показників на сусідніх станціях.

Байєсівське моделювання є ефективним методом для аналізу взаємозв'язків, що дозволяє виявляти причинно-наслідкові залежності на основі існуючих даних і, завдяки цьому, може допомогти визначати як зміни стану атмосферного повітря розповсюджуються в межах певної географічної області [3, 4].

Метою цього дослідження є перевірка ефективності застосування методу байєсівського моделювання для пошуку взаємовпливів рядів даних стану атмосферного повітря, отриманих з різних станцій моніторингу в межах певного регіону. Цей метод може бути корисним як для моделювання розповсюдження забруднення, так і слугувати інструментом попередньої обробки даних. Даним моніторингу властива наявність пропусків, а інформація щодо взаємовпливу значень показника на сусідніх станціях може бути використана для заповнення цих пропусків – дані в конкретній часовій точці можуть бути обчислені, використовуючи дані сусідніх станцій.

Розв'язання задачі

В одній з робіт авторів було застосовано цей метод для моделювання чутливості людей до певних грибкових алергенів [5]. Запропонована в тій роботі методологія може бути адаптована для розв'язання поточної задачі оцінки впливу змін показників якості атмосферного повітря залежно від змін відповідних показників на сусідніх станціях моніторингу. Пропонується наступний алгоритм:

1. Підготовка даних. Дані моніторингу, зазвичай, представлені в абсолютних величинах, але для

байєсівського моделювання переважно використовують бінарні значення. Набір даних стану атмосферного повітря необхідно привести до бінарного формату за рахунок порівняння значення показника у певний момент часу з його попереднім спостереженням – якщо поточне значення більше попереднього, то воно замінюється на «1», якщо ж значення менше або рівне – «0».

2. Побудова спрямованого ациклічного графа. Для отриманих даних проводиться вивчення структури, що дозволяє виявити залежності між існуючими змінними та сформувати спрямований ациклічний граф.

3. Вивчення параметрів. Вивчення параметрів даних – це оцінка значень умовних розподілів ймовірностей, яка проводиться на графі, отриманому на попередньому етапі. Результатом є набір ймовірностей зміни значення певної вершини графа, що залежать від її батьківських вершин.

Приклад застосування методу

Для перевірки цього алгоритму було використано дані моніторингу якості атмосферного повітря, а саме показник пилу PM10, від мережі громадського моніторингу EсоCity, що було отримано із кількох станцій моніторингу у Вінницькій області в період 2022-2024 років [6]. Графік значень цього показника для однієї з станцій наведено на рис. 1.

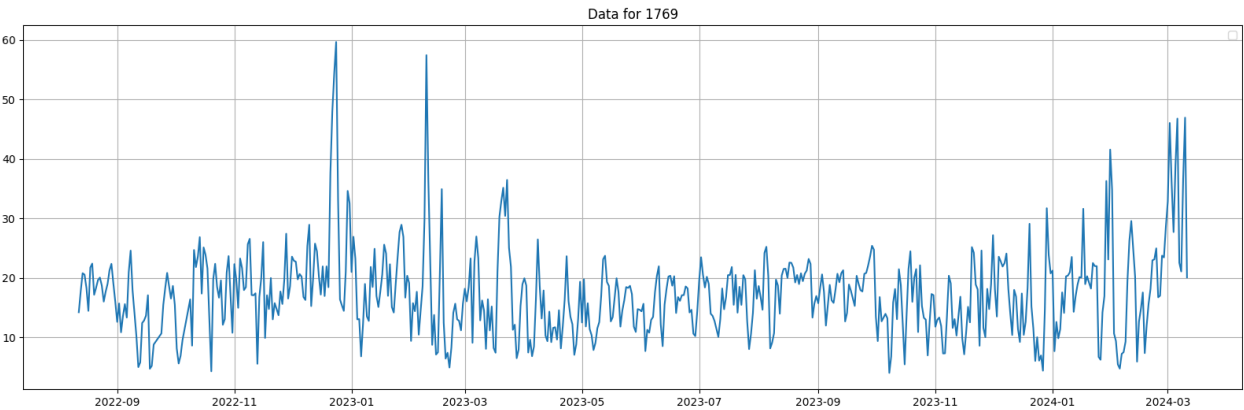


Рис. 1. Дані показника PM10 однієї з станцій моніторингу за 2022-2024 роки [7]

Згідно з описаним алгоритмом, сформовано набір, де кожен рядок це дата проведення вимірювання, а значення в колонках – відповідні значення показника PM10 на конкретній станції моніторингу. Згідно з кроком №2 алгоритму, дані перетворено у бінарний формат і отримано набір вхідних даних для моделювання, що наведено на рис. 2.

	st_650	st_1769	st_1811	st_1813	st_1825	st_1830	st_1834
0	0	0	0	0	0	0	0
1	1	1	0	1	1	1	1
2	0	1	1	1	1	0	1
3	0	1	1	1	1	1	1
4	1	0	1	0	0	0	0
...	...	...	...	...	...	...	...
279	1	0	0	0	0	0	0
280	0	0	1	1	1	1	0
281	1	1	1	1	1	0	1
282	0	1	0	0	0	1	1
283	0	0	0	0	0	0	0

Рис. 2. Вхідні дані для моделювання взаємовпливу показника PM10 [7]

Використовуючи Python бібліотеку bnlearn, для отриманих даних було побудовано спрямований ациклічний граф, який наведено на рис. 3.

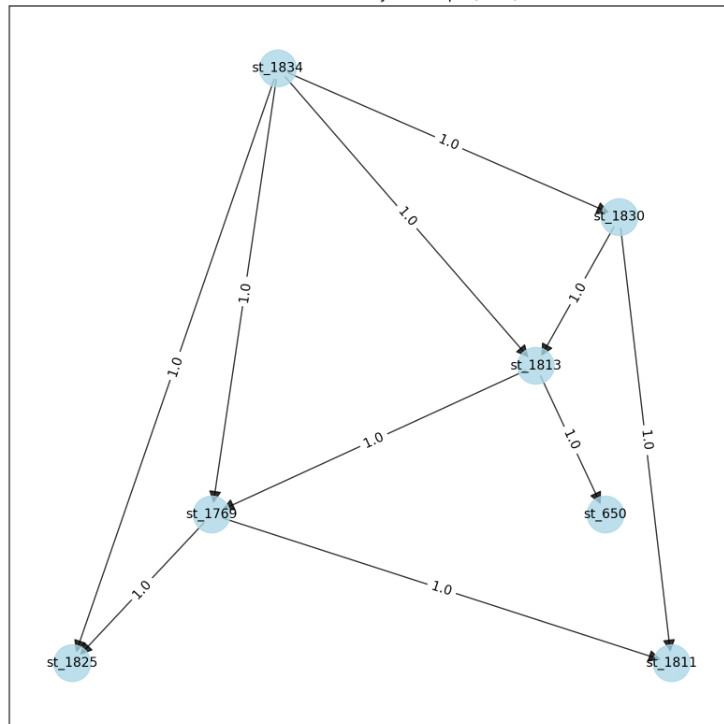


Рис. 3. Спрямований ациклічний граф взаємовпливів показника РМ10 [7]

Як можна побачити на рисунку 3, у наборі даних присутні взаємозв'язки між кількома станціями моніторингу атмосферного повітря, наприклад, значення на станції «1825» залежать від значень на станціях «1834» та «1769». Скористаємося вже згаданою бібліотекою `bnlearn` та знайдемо значення умовних розподілів ймовірностей. Частина отриманих результатів, а саме розподілів ймовірностей для станцій «1769» та «1811», наведено на рисунку 4.

```

[bnlearn] >CPD of st_1769:
+-----+-----+-----+-----+
| st_1813 | st_1813(0) | ... | st_1813(1) |
+-----+-----+-----+-----+
| st_1834 | st_1834(0) | ... | st_1834(1) |
+-----+-----+-----+-----+
| st_1769(0) | 0.5813253012048193 | ... | 0.3994252873563218 |
+-----+-----+-----+-----+
| st_1769(1) | 0.4186746987951807 | ... | 0.6005747126436781 |
+-----+-----+-----+-----+

[bnlearn] >CPD of st_1811:
+-----+-----+-----+-----+
| st_1769 | st_1769(0) | ... | st_1769(1) |
+-----+-----+-----+-----+
| st_1830 | st_1830(0) | ... | st_1830(1) |
+-----+-----+-----+-----+
| st_1811(0) | 0.5772870662460567 | ... | 0.4057971014492754 |
+-----+-----+-----+-----+
| st_1811(1) | 0.4227129337539432 | ... | 0.5942028985507246 |
+-----+-----+-----+-----+

```

Рис. 4. Значення умовних розподілів ймовірностей для станції «1769» та «1811» відповідно [7]

Як можна побачити на рисунку 4, при одночасному підвищенні рівня забруднення на станціях «1813» (с. Якушинці) та «1834» (с. Стрижавка), імовірність підвищення рівня забруднення на станції

«1769» (м. Вінниця) зростає в 1.5 рази. Подібну ситуацію можна спостерігати і на станції «1811» (м. Хмільник) – зростання імовірності підвищення рівня забруднення майже в 1.5 разів за умови підвищення рівня забруднення на станціях «1769» (м. Вінниця) та «1830» (с. Війтівці). Крім того, було перевірено точність роботи даної моделі – отримати значення точності 77% (рис. 5), що є досить хорошим показником.

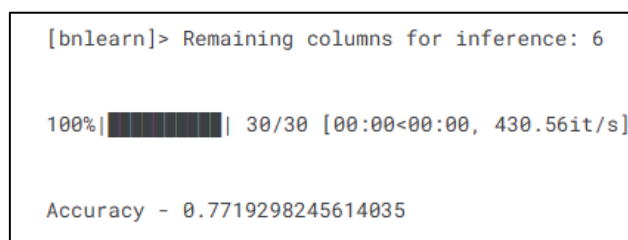


Рис. 5. Точність роботи моделі [7]

Висновки

В роботі було проаналізовано ефективність застосування методу байєсівського моделювання для аналізу взаємовпливів між даними рядів даних показників стану атмосферного повітря, отриманих з різних станцій моніторингу в межах певного району. Було запропоновано алгоритм застосування методу байєсівського моделювання для визначення того, як зміна рівня забруднення на одній станції моніторингу впливає на зміну рівнів забруднення на сусідніх станціях.

Алгоритм було перевірено на даних моніторингу стану атмосферного повітря Вінницькій області за 2022-2024 роки, отриманих від мережі громадського моніторингу EcoCity. Дослідження показало, що існує взаємозв'язок між значеннями показника РМ10 на сусідніх станціях і спостерігається зростання імовірності підвищення рівня забруднення за умови зростання цього рівня на сусідніх станціях в межах географічної зони.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Адаменко Микола, Кацман Михайло, Білецька Євгенія. Аналіз існуючих математичних моделей і комп'ютерних програм для прогнозування розповсюдження забруднюючих речовин в атмосфері. Системи обробки інформації. 2018. № 1(152). С. 155-162. <https://doi.org/10.30748/soi.2018.152.22>.
2. Popov, O., Iatsyshyn, A., Kovach, V., Artemchuk, V., Taraduda, D., Sobyna, V., Sokolov, D., Dement, M., Hurkovskyi, V., Nikolaiev, K., Yatsyshyn, T., & Dimitrieva, D. (2019). Фізичні особливості розповсюдження забруднюючих речовин в атмосферному повітрі за умов надзвичайної ситуації на АЕС. *Ядерна та радіаційна безпека*, (4(84)), 88-98. [https://doi.org/10.32918/nrs.2019.4\(84\).11](https://doi.org/10.32918/nrs.2019.4(84).11).
3. Nagarajan R, Scutari M, Lèbre S. Bayesian network inference algorithms. In: *Bayesian networks in R*. New York: Springer, 2013. doi: 10.1007/978-1-4614-6446-4
4. Трофимчук, О. М., Бідюк, П. І., Просянкін-Жарова, Т. І. і Терент'єв, О. М. (2023) «Байєсівський аналіз даних у моделюванні та прогнозуванні нелінійних нестационарних фінансово-економічних процесів», Міжнародний науково-технічний журнал "Проблеми керування та інформатики", 68(4), с. 71–83. doi: 10.34229/1028-0979-2023-4-6.
5. Victoria Rodinkova, Serhii Yuriev, Vitalii Mokin, Mariia Kryvopustova, Dmytro Shmundiak, Mykyta Bortnyk, Yevhenii Kryzhanovskiy, Andrii Kurchenko, Bayesian analysis suggests independent development of sensitization to different fungal allergens, *World Allergy Organization Journal*, Volume 17, Issue 5, 2024, 100908, ISSN 1939-4551, <https://doi.org/10.1016/j.waojou.2024.100908>
6. Vitalii Mokin, Dima Shmundiak, Volodymyr Kopniak "Air Quality Monitoring from EcoCity" Kaggle Dataset, 2021. [Електронний ресурс]. Режим доступу: <https://www.kaggle.com/datasets/vbmokin/air-quality-monitoring-from-ecocity>.
7. Dmytro Shmundiak, "Bayes for air monitoring data" Kaggle Notebook. [Електронний ресурс]. Режим доступу: <https://www.kaggle.com/code/dimashmundiak/bayes-for-air-monitoring-data>.

Шмундіак Дмитро Олександрович – аспірант кафедри системного аналізу та інформаційних технологій, e-mail: dimashmund@gmail.com

Науковий керівник: Мокін Віталій Борисович – д. т. н., професор, завідувач кафедри системного аналізу та інформаційних технологій, e-mail vbmokin@vntu.edu.ua

Shmundiak Dmytro O. - PhD student of the Chair of System Analysis and Information Technologies, e-mail: dimashmund@gmail.com

Scientific supervisor: Mokin Vitalii B. - Doctor of Technical Sciences, Professor, Head of the Chair of System Analysis and Information Technologies, e-mail vbmokin@vntu.edu.ua