

ВИКОРИСТАННЯ БАЙЄСОВИХ МЕРЕЖ ДЛЯ ОЦІНЮВАННЯ ТОЧНОСТІ РЕЗУЛЬТАТІВ МОДЕЛЮВАННЯ ПОШИРЕННЯ ДОМШОК У ПОВІТРІ ЗА УМОВИ НЕТОЧНИХ ВХІДНИХ ДАНИХ

Р.В. КРИВАКОВСЬКА

Анотація. Розглянуто питання оцінювання точності результатів моделювання поширення домішок у повітрі в разі подання на вхід моделей неточних вхідних даних. Наведено обмеження на методи оцінювання точності для України. Для отримання оцінок запропоновано використання байєсових мереж з дискретними вхідними змінними. Подано структуру мережі та запропоновано методи заповнення таблиць імовірностей.

Ключові слова: оцінювання стану атмосферного повітря, байєсові мережі довіри.

ВСТУП

Питання забезпечення якості навколишнього середовища стають дедалі більш важливими. Підвищення антропогенного і техногенного навантаження на навколишнє середовище призводить до його деградації, що, у свою чергу, може спричинити негативні наслідки для здоров'я населення.

Проектуючи системи оцінювання стану повітря, слід враховувати той факт, що вхідні дані, які подаються на вхід системи, як правило, неточні. Ігнорування цього факту у процесі побудови систем прийняття рішень, пов'язаних із забезпеченням належної якості повітря, може зумовити неправильні рішення.

Постає проблема визначення точності моделювання у випадку неточних вхідних даних. Неточність результатів моделювання може виникати через недосконалість моделей.

У разі роботи з оцінювання якості атмосферного повітря в Україні з'являються додаткові вимоги, що накладаються на методи. Дані спостережень і потужності викидів часто подаються не повністю, і часто доступні тільки усереднені значення числових характеристик. Тому до методу оцінювання достовірності вхідних даних ставляться такі вимоги:

- 1) простота;
- 2) можливість роботи з невеликими наборами даних;
- 3) орієнтація на спрощування інформації в умовах невизначеності.

Байєсові мережі широко використовують в екології для вибору стратегій, діагностики і прогнозування [2, 3]. Схожа за принципом роботи фільтрація Калмана і її модифікації широко застосовуються в екологічних завданнях для поліпшення якості моделювання шляхом засвоєння даних спостережень [4–6]. Перевагою байєсових мереж є також їх здатність до навчання [7].

ПОСТАНОВКА ЗАВДАННЯ

Розглянемо інформаційну систему оцінювання якості повітря в умовах України. Нехай її частиною буде підсистема моделювання, на вхід якої надходять невеликі набори неточних вхідних даних. Необхідно отримати оцінки точності результатів моделювання поширення домішок у приземному шарі атмосферного повітря, якщо відомі відносні відхилення вхідних даних для моделі. Також належить оцінити точність вхідних даних моделі, якщо відомі відносні похибки результатів моделювання.

СТВОРЕННЯ МОДЕЛІ

Для отримання ймовірнісних оцінок достовірності результатів моделювання за неточних вхідних даних можна використовувати байєсові мережі довіри. Цю технологію використовують для проведення ймовірнісних міркувань [1].

Їх переваги для вирішення завдань цього типу такі:

1. Умовні ймовірності можна обчислювати один раз або отримувати ітеративно за допомогою навчання залежно від кількості наявних даних.
2. Це досить проста модель, для якої існують правила виведення.
3. Ця модель призначена для роботи в умовах невизначеності.

Основні етапи побудови байєсової мережі:

1. Побудова структури мережі: вивчення предметної галузі, виділення факторів і знаходження залежності між ними.
2. Визначення значень вузлів мережі: вибір значень змінних-вузлів мережі, визначення їх типу (дискретні, неперервні).
3. Заповнення таблиць умовних імовірностей станів. Значення умовних імовірностей можна знаходити за допомогою:

- експертних оцінок;
- статистичної інформації, отриманої раніше.

Після виконання основних етапів отримуємо готову мережу, яку можна використовувати для проведення експериментів.

Наступним етапом у роботі буде заповнення мережі значеннями умовних імовірностей. Заповнення таблиць умовної ймовірності можна починати з перевірки адекватності моделі. Для цього зазвичай використовують дані високої (або принаймні визначеної) точності.

Для отримання значень умовних імовірностей можна використовувати дані, отримані від експертів, або інші методи. Наприклад, для байєсової мережі це може бути такий підхід:

1. Порівняння результатів моделювання з даними з постів спостереження. Припускається, що ступінь достовірності даних на постах спостереження дорівнює відсотковому відношенню похибки вимірювання до результату спостереження.
2. На основі порівняння даних знаходження значень умовної ймовірності і занесення їх у таблицю.

Якщо адекватність перевіряється шляхом порівняння даних моделювання з даними з пунктів спостереження за станом атмосферного повітря, то одразу можна визначити відносне відхилення результатів моделювання від даних спостережень для всіх точок спостережень. Відхилення будуть коливатися у певних межах.

Якщо модель є адекватною, можна визначити відсоток точок спостереження, у яких відхилення матимуть різні межі. Далі цей відсоток можна перевести в апостеріорну (умовну) ймовірність. Так можна отримати значення умовних імовірностей для даних високої достовірності.

Якщо дані інвентаризації емісій неповні і модель вважається адекватною для спостережень у точках, у яких вплив неврахованих викидів мінімальний, то для точок зі значним впливом неврахованих викидів можна припустити сумарну недостовірність викидів середньою або високою. Виходячи з цього, можна знайти частоту появи даних середнього або високого ступеня недостовірності і обрахувати відповідні значення умовних імовірностей.

Для врахування впливу погодних умов на результати моделювання можна вдатися до пробних запусків моделі або скористатися даними інших учених, які проводили подібні дослідження.

Якщо використовуються відомі моделі, для них зазвичай проводилися дослідження впливу неточності даних на результати моделювання і тому можна побудувати таблиці умовних імовірностей за цими даними. Для невідомої моделі можна брати дані дослідження найбільш близьких до неї моделей.

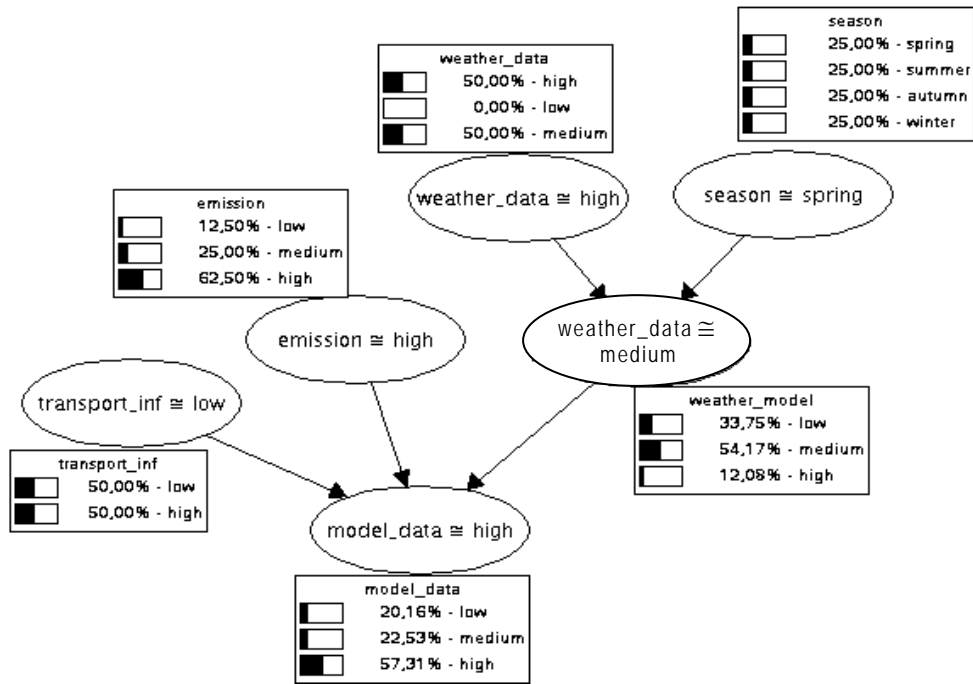
Якщо використовується моделювання для вирішення завдань, нагромаджується статистика запусків, яка придатна для уточнення значень у таблицях умовних імовірностей.

Якщо відомі значення умовних імовірностей, то можна досліджувати ймовірності у зворотному напрямку. Фіксуючи значення результуючої змінної, можна обчислити ймовірності значень вхідних змінних згідно з таблицями умовних імовірностей.

Пропонується структура мережі, показана на рисунку.

Після визначення структури мережі можна визначати значення змінних. У байєсовій мережі довіри використовують як дискретні, так і неперервні змінні. У розгляданому випадку використано дискретні значення змінних, оскільки з ними працювати простіше.

Один блок мережі відповідає одній змінній. Кожна змінна являє собою джерело неточності, яке впливає на точність результатів моделювання поширення домішок. Кожна змінна мережі (крім змінних *season* і *transport_inf*) може набувати трьох значень: *low*, *medium*, *high* з міркувань балансу між зручністю використання і точністю, що відповідають рівням якості вхідних даних або результатів моделювання. Ці рівні вибрано на основі аналізу відносних відхилень результатів моделювання від даних з пунктів спостереження. Числові значення, що відповідають рівням якості, задаються користувачем системи.



Структура мережі довіри: weather_data — достовірність метеорологічних даних; season — пора року; weather_model — достовірність результатів метеорологічної моделі; emission — достовірність значень викидів від точкових джерел; transport_inf — вплив транспорту; model_data — значення достовірності результатів моделювання поширення домішок у приземному шарі атмосфери

У таблиці наведено значення змінних.

Змінні, що використовуються в байсовій мережі довіри

Змінна	season	weather_data	weather_model	emission	transport_inf	model_data
Список значень	spring, summer, autumn, winter	low, medium, high	low, medium, high	low, medium, high	low, high	low, medium, high
Опис	Пора року	Достовірність вхідних даних погодної моделі	Достовірність вхідних даних погодної моделі (залежно від ступеня достовірності вхідних даних)	Достовірність даних про викиди	Дані про забруднення від транспорту	Достовірність результатів моделювання

ВИСНОВКИ

Дослідження присвячено оцінюванню точності моделювання поширення домішок у приземному шарі атмосферного повітря за неточних вхідних да-

них. Розглянуто обмеження, що виникають у цій задачі для умов України. Задача розв'язується за допомогою побудови байесової мережі довіри і подальшого заповнення таблиць умовних імовірностей експертом. Така модель дозволяє отримати ймовірнісні характеристики для точності результатів моделювання залежно від точності вхідних даних, а також отримати ймовірності різних ступенів точності вхідних даних, якщо відомий рівень точності результатів моделювання.

ЛІТЕРАТУРА

1. С. Рассел, П. Норвиг, *Искусственный интеллект: современный подход*, пер. с англ. Москва: «Вильямс», 2006.
2. S. Johnson, "Integrated Bayesian Network frameworks for modeling complex ecological issues", *PhD thesis*, 2009.
3. Eugene Santos Jr., "Validation and verification of Bayesian knowledge-bases", *Data & Knowledge Engineering*, vol. 37, pp. 307–329, 2001.
4. R.E. Kalman, "A New Approach to Linear Filtering and Prediction Problems", *Journal of Basic Engineering*, vol. 82, pp. 35–45, 1960.
5. D.Q. Zheng, "Online update of model state and parameters of a Monte Carlo atmospheric dispersion model by using ensemble Kalman filter", *Atmospheric Environment*, vol. 43, no. 12, pp. 2005, 2009.
6. М.А. Толстых, *Применение ансамблевых фильтров Калмана в численном прогнозе погоды* [Электронный ресурс]. Доступно: ftp://geophyslab.srcc.msu.ru/Events/CITES2009/Tolstykh_090709.pdf
7. D. Heckerman, "A Tutorial on Learning with Bayesian Networks", *Technical Report MSR-TR-95-06, Microsoft Research*, 1995.

Надійшла 04.04.2019

INFORMATION ON THE ARTICLE

R.V. Kryvakovska, programmer, "Mama Products" Ltd, Ukraine, e-mail: deyatynor@gmail.com.

APPLICATION OF BAYESIAN NETWORKS FOR ACCURACY ESTIMATION OF MODELING RESULTS OF THE AIR POLLUTION DISPERSION GIVEN INACCURATE INPUT DATA / R.V. Kryvakovska

Abstract. The article deals with estimating the results accuracy of the modeling of air pollution dispersion when introducing inaccurate input data. Restrictions on accuracy estimation methods for Ukraine are considered. It is suggested to use Bayesian networks with discrete input variables to obtain the estimates. The structure of the network is presented, and the methods of filling the probability tables are proposed.

Keywords: atmospheric air assessment, Bayesian networks.

ИСПОЛЬЗОВАНИЕ БАЙЕСОВСКИХ СЕТЕЙ ДЛЯ ОЦЕНКИ ТОЧНОСТИ РЕЗУЛЬТАТОВ МОДЕЛИРОВАНИЯ РАСПРОСТРАНЕНИЯ ПРИМЕСЕЙ В ВОЗДУХЕ ПРИ УСЛОВИИ НЕТОЧНЫХ ВХОДНЫХ ДАННЫХ / Р.В. Криваковская

Аннотация. Рассмотрен вопрос оценки точности результатов моделирования распространения примесей в воздухе при подаче на вход моделей неточных входных данных. Приведены ограничения на методы оценки точности для Украины. Для получения оценок предложено использование байесовских сетей с дискретными входными переменными. Подана структура сети и предложены методы заполнения таблиц вероятностей.

Ключевые слова: оценка состояния атмосферного воздуха, байесовы сети доверия.

REFERENCES

1. Stuart Russell and Peter Norvig, *Artificial Intelligence: A Modern Approach*, 4th Edition, Hoboken, USA: Pearson, 2020.
2. S. Johnson, "Integrated Bayesian Network frameworks for modeling complex ecological issues", PhD thesis, 2009.
3. Eugene Santos Jr., "Validation and verification of Bayesian knowledge-bases", *Data & Knowledge Engineering*, vol. 37, pp. 307–329, 2001.
4. R.E. Kalman, "A New Approach to Linear Filtering and Prediction Problems", *Journal of Basic Engineering*, vol. 82, pp. 35–45, 1960.
5. D.Q. Zheng, "Online update of model state and parameters of a Monte Carlo atmospheric dispersion model by using ensemble Kalman filter", *Atmospheric Environment*, vol. 43, no. 12, pp. 2005, 2009.
6. M.A. Tolstykh, Application of ensemble Kalman filters in numerical weather forecast [Online]. Available: ftp://geophyslab.srcc.msu.ru/Events/CITES2009/Tolstykh_090709.pdf Accessed on: January 16, 2016.
7. D. Heckerman, "A Tutorial on Learning with Bayesian Networks", *Technical Report MSR-TR-95-06, Microsoft Research*, 1995.