

У статті розглядаються питання достовірності вхідних даних для задачі моделювання розповсюдження домішок в атмосферному повітрі. Наводяться джерела недостовірності даних, дано огляд підходів, що дозволяють зменшити її вплив. Запропоновано метод для оцінки імовірнісних характеристик недостовірності результатів моделювання

Ключові слова: недостовірність даних, штучний інтелект, Байєсова мережа довіри

В статье рассматриваются вопросы достоверности входных данных для задачи моделирования распространения примесей в атмосферном воздухе. Приводятся источники недостоверности данных, дан обзор подходов, позволяющих уменьшить её влияние. Предложен метод для оценки вероятностных характеристик недостоверности результатов моделирования

Ключевые слова: недостоверность данных, искусственный интеллект, Байесова сеть доверия

АНАЛИЗ СПОСОБОВ ОЦЕНКИ ДОСТОВЕРНОСТИ ВХОДНЫХ ДАННЫХ ДЛЯ ОЦЕНКИ СОСТОЯНИЯ ВОЗДУХА

Р. В. Криваковская

Аспирант

Отдел автоматизированного проектирования
энергетических установок
Институт проблем моделирования в энергетике
им. Г.Е. Пухова НАН Украины
ул. Генерала Наумова 15, г. Киев, Украина, 03164
Контактный тел.: 068-405-79-59
E-mail: deyatinor@ua.fm

1. Введение

В настоящее время вопросы обеспечения качества окружающей среды становятся важными и актуальными. Повышение антропогенной и техногенной нагрузки на окружающую среду приводит к её деградации, что, в свою очередь, приводит к негативным последствиям для здоровья населения. Таким образом, исследования в области разработки систем управления качеством окружающей среды являются актуальными.

При проектировании систем мониторинга или управления качеством воздуха, следует учесть тот факт, что входные данные, подаваемые на вход системы, могут быть неточными, недостоверными или неполными. Игнорирование этого факта при построении систем принятия решений, связанных с обеспечением надлежащего качества воздуха, может приводить к неверным решениям. Поэтому оценка и учёт неопределённости, возникающей в задаче при исследованиях в области обеспечения качества воздуха.

2. Постановка проблемы

Недостоверность и неполнота входных данных могут быть связаны с различными причинами. Методы оценки недостоверности и неполноты могут различаться для разных случаев. Поэтому представляется актуальным проведение анализа причин и методов оценки недостоверности и неполноты данных.

Недостоверность и неполнота данных может быть связана с несколькими причинами:

1. Неточность результатов работы моделей из-за несовершенства моделей
2. Неточность результатов работы моделей из-за недостоверности или неполноты данных о выбросах (например, использование только данных о выбросах от стационарных предприятий и игнорирование выбросов транспорта).
3. Неточность данных с постов наблюдения из-за погрешностей измерения и нерациональное размещение постов.
4. Усреднение данных по времени.

Важность проведения исследований, направленных на оценку степени недостоверности и неполноты входных данных, можно проиллюстрировать на примере. На рис. 1 приведен пример карт, созданных в информационно-аналитической системе AISEEM [1, 2], которые представляют собой результаты моделирования распространения загрязнения (слева) и интерполяции данных с постов наблюдения за состоянием атмосферного воздуха (справа). Можно увидеть, что результаты моделирования и наблюдений различны. Сравнение этих результатов показывает области сходства и различия данных. Также сравнение позволяет увидеть потенциальные источники недостоверных данных.

Обнаружение и исследование недостоверности и неполноты данных и выработка способов её преодоления являются важной задачей, которую необходимо решить при проектировании системы.

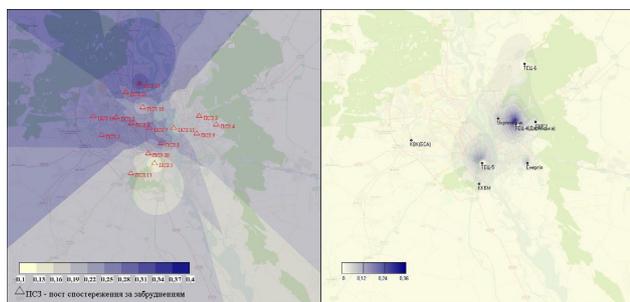


Рис. 1. Карты загрязнения с постов (слева) и с модели (справа)

3. Цели и задачи исследования

При исследованиях, связанных с недостоверностью и неполнотой входных данных, можно выделить следующие цели:

1. Определение ограничений использования применяемых моделей.
2. Выбор моделей распространения примесей в атмосферном воздухе с учётом ограничений входных данных.
3. Определение источников недостоверности данных.
4. Обоснование правильности выбора стратегий оптимизации выбросов (и других решений).

Исследования в области недостоверности и неполноты входных данных могут идти в следующих направлениях:

1. Числовая оценка вклада различных факторов в степень недостоверности данных.
2. Разработка методов, позволяющих улучшать достоверность результатов (получать правильные результаты) при недостоверных входных данных.
3. Определение областей устойчивости моделей.
4. Ранжирование моделей распространения примесей в зависимости от чувствительности к различным видам недостоверности и неполноты входных данных. Это может позволить производить выбор модели (из имеющихся) в зависимости от качества имеющихся данных.

4. Анализ литературы

Для исследования применяются различные способы:

1. Сравнение результатов моделирования с данными наблюдений позволяет обнаружить дополнительные неучтённые воздействия.
2. Анализ чувствительности моделей позволяет определить степени неточности входных данных, которые можно использовать для определения возможности использования данной модели в данной ситуации.
3. Использование методов идентификации модели по данным наблюдений (усвоение данных наблюдений) позволяет улучшить качество моделирования.

4. Применение ансамблей моделей в некоторых ситуациях позволяет улучшить качество моделирования [3].
5. Решение обратных задач позволяет уточнить данные инвентаризации источников загрязнения, а также идентифицировать источники выбросов [4].

Эти методы могут применяться по отдельности или вместе.

При сравнении результатов моделирования с данными наблюдений можно определить недостоверность данных о выбросах и неточность данных постов наблюдения. Недостатком этого подхода является то, что для работы модели имеют значение как несовершенство модели, так и неполнота и недостоверность данных о выбросах. Таким образом, мы не можем отделить степень недостоверности данных о выбросах от ограничений самой модели.

Снять этот недостаток помогает обратное моделирование. Обратное моделирование помогает оценить числовые характеристики источников загрязнения, зная результаты наблюдений. Существует несколько способов обратного моделирования. Следует понимать, что обратное моделирование также имеет погрешность и это нужно учитывать. Использование обратного моделирования для усвоения данных может повысить качество прямого моделирования. Также в случае использования обратных моделей полученные данные возможно сравнивать между собой и таким образом оценить степень недостоверности исходных данных о выбросах.

При выборе моделей для использования следует учитывать их ограничения на качество входных данных. Для этого можно использовать результаты анализа чувствительности. Чаще всего исследуется влияние небольших возмущений входных данных на результаты моделирования. Если анализ чувствительности показал, что модель чувствительна к небольшим возмущениям входных данных, а предполагаемая степень недостоверности входных данных велика, то такую модель в данном случае использовать нельзя. Возможен также случай различной чувствительности модели к разным типам входных данных. В таких случаях использование модели возможно с оговорками.

Исследования, направленные на обнаружение области устойчивости модели, позволяют точнее идентифицировать степень недостоверности входных данных, при которых возможно использование моделей.

Результаты проведения анализа чувствительности также можно использовать для обнаружения дополнительных неучтённых воздействий. Например, в [5] при проведении оценки адекватности моделей автор отмечает, что для постов, которые в меньшей мере испытывают влияние транспорта, его модель является адекватной. Если сравнить данные с других постов с данными моделирования, можно получить оценку интенсивности выбросов от транспорта на других постах.

Для получения вероятностных оценок недостоверности данных можно использовать байесовы сети доверия (БСД). Эта технология используется для проведения вероятностных рассуждений [6]. Байесовы сети широко используются в экологии в задачах выбора стратегий, диагностики и прогнозирования [7-9]. Похожая по принципу работы фильтрация Калмана

и её модификации широко используется в экологических задачах для улучшения качества моделирования путём усвоения данных наблюдений. Также, достоинством байесовых сетей является их способность к обучению.

В случае наличия больших массивов входных данных, для получения результатов можно использовать такие методы, как нейронные сети. Эти методы широко используются в различных задачах распознавания образов. Недостатком нейронных сетей является то, что им необходимы большие объемы данных для обучения. Эти методы также требуют обязательной верификации на тестовых данных.

В настоящее время широко используются методы усвоения данных, анализа чувствительности моделей, развиваются ансамблевые методы моделирования. В то же время исследований в области моделирования в условиях недовольности входных данных недостаточно. Поэтому было принято решение сосредоточиться на работе в этом направлении.

5. Особенности ситуации в Украине

В случае работы по оценке качества атмосферного воздуха в Украине, появляются дополнительные требования, налагаемые на методы. Данные наблюдений и выбросов часто являются фрагментарными, а также часто доступны только усреднённые значения числовых характеристик. Поэтому к методу оценки достоверности входных данных предъявляются следующие требования:

1. Простота.
2. Возможность работы с небольшими наборами данных.
3. Ориентация на работу с информацией в условиях неопределённости.

6. Использование байесовых сетей для оценки степени недовольности данных

Для решения этой задачи перспективным представляется использование байесовых сетей доверия для оценки степени недовольности и неполноты данных. Их преимущества для решения задач такого типа таковы:

1. Условные вероятности можно вычислять один раз или обучать, в зависимости от количества имеющихся данных.
2. Это довольно простая модель, для которой существуют правила вывода.
3. Эта модель предназначена для работы в условиях неопределённости.

При построении байесовой сети необходимо пройти несколько основных этапов. Это такие этапы:

1. Построение структуры сети. На этом этапе происходит изучение предметной области, выделение факторов и нахождение зависимостей между ними.
2. Определение значений узлов сети. На этом этапе производится выбор значений переменных узлов сети и определяется их тип (дискретные, непрерывные).
3. Заполнение таблиц условных вероятностей состояний. Значения условных вероятностей можно находить следующими способами:

1. С помощью экспертных оценок.
2. С помощью статистической информации, полученной ранее.

После выполнения основных этапов мы получаем готовую сеть, которую можно использовать для проведения экспериментов.

Адекватность моделей, построенных с помощью байесовых сетей доверия, можно проверить на основе имеющихся данных. При необходимости после проверки в сеть могут быть внесены изменения.

Предлагается структура сети, приведенная на рис. 2.

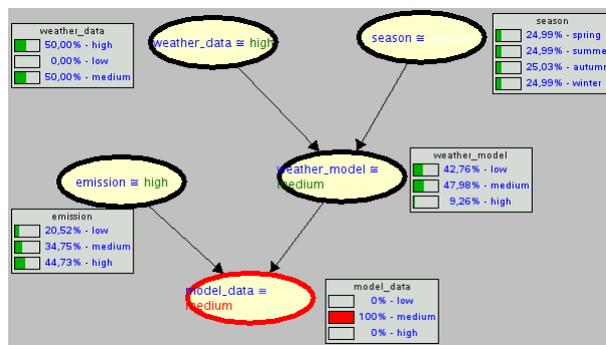


Рис. 2. Структура сети доверия

В табл. 1 приведены значения переменных.

Таблица 1

Переменные, используемые в БСД

Переменная	season	weather_data	weather_model	emission	model_data
Список значений	spring, summer, autumn, winter	low, medium, high	low, medium, high	low, medium, high	low, medium, high
Описание	Время года	Достоверность входных данных погодной модели	Достоверность выходных данных погодной модели (в зависимости от степени достоверности входных данных)	Достоверность данных о выбросах	Достоверность данных моделирования

Следующим этапом в работе будет заполнение сети значениями условных вероятностей.

Для получения значений условных вероятностей можно использовать данные, полученные при анализе чувствительности моделей. Также, для нахождения значений условных вероятностей для байесовой сети может быть использован такой подход:

1. Проводится сравнение результатов моделирования с данными с постов наблюдения. При этом принимаются такие допущения.
2. Степень достоверности данных на постах наблюдения принимается равной процентному отношению погрешности измерения к результату наблюдения.
3. На основе сравнения данных производится нахождение значений условной вероятности и заносится в таблицу.

Следующими этапами данного исследования в будущем нахождение значений условных вероятностей и проведение моделирования на сети. Пути развития исследования может быть получение результатов на

других входных данных и сравнение БСД с другими методиками, в первую очередь с методиками на основе обратного моделирования.

7. Выводы

В данной статье был проведен анализ источников недостоверности и неполноты входных данных, возникающих в задаче оценки состояния атмосферного воздуха. Также, был сделан анализ основных методов работы в условиях недостоверности входных данных, в результате чего были выделены области применения различных методов. Проанализированы особенности и ограничения применения этих методов в условиях Украины. Автором предлагается оценивать степень недостоверности данных моделирования распространения примесей в атмосферном воздухе при помощи байесовых сетей доверия. Выделены направления дальнейших исследований.

Литература

1. Артемчук, В.А. Информационно-аналитическая система эколога-энергетического мониторинга [Текст] / В.А. Артемчук, О.А. Грибан // Моделювання та інформаційні технології. – 2010. – Т. 1, спец. вип. С. 120-128
2. Яцишин, А.В. Комп'ютерні засоби прогнозування техногенних навантажень на атмосферу [Текст] / А.В. Яцишин, О.О. Попов, В.О. Артемчук // Східно-Європейський журнал передових технологій - 2009. – Т. 5, № 2 (41). – С. 33-36.
3. Mallet. V. Ensemble-based air quality forecasts: A multimodel approach applied to ozone [Текст] / V. Mallet, B. Sportisse // J. Geophys. Res. - 2006, 111, D18302, doi:10.1029/2005JD006675.
4. Кожевникова, М.Ф. Идентификация источников загрязнения: вычислительные методы [Текст] / М.Ф. Кожевникова, В.В. Левенец, И.Л. Ролик // Вопросы атомной науки и техники. - 2011. - №6. Серия: Вакуум, чистые материалы, сверхпроводники (19), с. 149-156.
5. Попов, О.О. Математичне та комп'ютерне моделювання техногенних навантажень на атмосферу міста від стаціонарних точкових джерел забруднення [Текст] : дис. ... канд. техн. наук : 01.05.02 / О. О. Попов – К., 2010. – 198 с.
6. Рассел С., Норвиг П. Искусственный интеллект: современный подход [Текст] : пер. с англ. М: "Вильямс", 2006 – 1408 с.
7. Али Мансур Номан Мархуб. Экспертная система поддержки принятия решений в интеллектуальной системе экологического мониторинга атмосферного воздуха промышленного региона (на примере г.Новомосковска Тульской области) [Текст] / автореф. дис. ... канд. техн. наук. 05.13.06, 03.02.08 / Али Мансур Номан Мархуб - М., 2011 – 18 с.
8. Johnson S. Integrated Bayesian Network frameworks for modeling complex ecological issues [Текст] / PhD thesis / S. Johnson - 2009.
9. He L. A probabilistic reasoning-based decision support system for selection of remediation technologies for petroleum-contaminated sites [Текст] / L. He., C.W. Chan, G.H. Huang, G.M. Zeng. // Expert Systems with Applications – 2006 – Вып. 30 – с. 783–795

Abstract

The article concerns the issues of reliability of input data for modeling of admixtures propagation in the air, and possible results of unreliable input data application at further optimization and control of the air quality. The main sources of unreliable data are given. There are objectives of the research in the sphere of input data limitation, review of approaches, decreasing its effect. The requirements to the methods of overcoming of input data limitation, as well as certain demands determined by Ukrainian realities, were formulated. The method based on the Bayes networks was suggested to assess the probabilistic properties of the modeling results unreliability during the application of unreliable input data. The article reveals the advantages of Bayes networks and the stages of their plotting. The structure of the network was suggested and its variables were described to assess the results of modeling. There is an approach to find out the values of conditional probabilities for the network.

Keywords: data unreliability, artificial intelligence, Bayes network