

Chistyakov Alexander Evgenjevich

E-mail: cheese_05@mail.ru.

Phone: +78634371606.

The Department of Higher Mathematics; Assistant.

Khachunts Dianna Samvelovna

E-mail: diana-hachunts@mail.ru.

Phone: +79287786737.

Postgraduate Student.

УДК 681.3.06:502.504

В.К. Гадельшин, Ю.А. Мещерякова, Т.И. Погорелова

ПРОГНОЗИРОВАНИЕ КОНЦЕНТРАЦИЙ ВРЕДНЫХ ВЕЩЕСТВ В ВОЗДУШНОЙ СРЕДЕ ПО СТАТИСТИЧЕСКИМ ДАННЫМ

Рассматривается задача контроля и оперативного прогнозирования состояния воздушной среды на основе статистических данных по выбросам загрязняющих веществ автотранспортом в экологически проблемных местах г. Таганрога. Предлагается программный комплекс, включающий базу данных измерений и программное обеспечение, осуществляющий группировку данных на основе алгоритмов кластеризации, реализующий процедуру оперативного прогнозирования концентрации вредных веществ в атмосфере при заданных параметрах методами распознавания образов. Приводятся примеры контрольных расчетов.

Прогнозирование; кластеризация; автомобильные выбросы; загрязнение атмосферы; распознавание образов.

V.K. Gadelshin, J.A. Meshcheryakova, T.I. Pogorelova

THE FORECASTING PROCEDURE THE CONCENTRATION OF HARMFUL SUBSTANCES INTO THE ATMOSPHERE OF STATISTICAL DATA

This is a study of air pollution harmful emissions from vehicles. It addresses the problem of organizing the procedure of forecasting the use of an environmental database, formed according to the results of field experiments. In the grouping of data carried out on the basis of algorithms for clustering, using the method of pattern recognition is implemented forecasting procedure the concentration of harmful substances into the atmosphere during the given parameters, the forecast is implemented in a software package, generate verification program, performed control calculations.

Forecasting procedure; clustering; emissions from vehicles; air pollution; of pattern recognition.

Одной из главных экологических проблем в городах является высокий уровень загрязнения воздушной среды. Основные источники выбросов вредных, загрязняющих атмосферу веществ – автотранспорт и промышленные предприятия. Например, в 2009 г. на долю автотранспорта в г. Таганроге приходилось 79,1 % от общего объема выбросов загрязняющих веществ, а по данным УГИБДД ГУВД по Ростовской области количество автомобилей в Таганроге с 35 тысяч в 2001 г. к 2010 г. увеличилось до 61 тысячи. Вблизи транспортных магистралей с интенсивным движением при неблагоприятных метеоусловиях и заторах содержание вредных примесей в воздухе значительно превышает допустимый уровень.

Сложность проведения регулярных трудоемких натурных экспериментов для оперативной и долговременной оценки и прогнозирования состояния воздушной

среды приводит к необходимости применения вычислительных экспериментов методами математического моделирования и обработки статистических данных.

В данной работе предлагается метод и программный комплекс, позволяющие на основе статистических данных по выбросам загрязняющих веществ автотранспортом в экологически проблемных местах г. Таганрога, выполнять оперативный прогноз концентрации вредных веществ в атмосфере при заданных метеорологических условиях и составе транспортного потока.

Для оперативного прогноза концентрации загрязняющих веществ разработан алгоритм, состоящий из следующих процедур: создание базы данных, выборка данных из БД по точечным и интервальным запросам; выявление решающих параметров; нормализация данных; классификация замеров по месту и метеоусловиям; группирование замеров на основе алгоритмов кластеризации; выявление принадлежности исходной точки к тому или иному кластеру; формирование прогноза концентрации веществ, характерных для данной группы замеров.

Исходными данными являются замеры концентрации загрязняющих веществ в воздухе, сведения по составу автомобилей в потоке, метеоусловия с привязкой к месту и времени проводимых измерений, выполняемых ФГУЗ «Центр гигиены и эпидемиологии» в г. Таганроге в течение более десяти лет. Они позволяют характеризовать обстановку в городе в целом, а также служат для всестороннего анализа оценок и прогнозов. Пример исходных данных приведен в табл. 1.

Таблица 1

Перекресток: Александровская – Гоголевский		Дата: 12.04.05		Время: 9 ¹⁰ – 10 ⁴⁵	
Тип машин	Кол-во	Метеоусловия			
Грузовые + диз.	36	Давление		756	
Легковые	620	$T_{сух}$		14	
Дизельные		$T_{влаж}$		14	
Автобусы	10	Влажность		60	
Мотоциклы		Напр. ветра		Ю-З	
Трактора		Скорость		2-3	
Маршрутки	64	Облачность		Переменная облачность	
Газы	Пробы				Норма
NO_2	0,02	0,02	0,08	0,02	0,085
SO_2	0,04	0,04	0,04	0,04	0,5
Форм	0,01	0,01	0,01	0,01	0,035
Сажа	0,025	0,025	0,025	0,025	0,15
Озон	0,001	0,004	0,002	0,002	0,16
СО	3,0	1,0	3,0	2,0	5,0

На основе анализа исходных данных и различных СУБД для поддержки программного комплекса используется: **Microsoft SQL Server 2008 R2 Express (для базы данных)** и **Microsoft Visual Studio Express 2010** (для программного продукта) (рис. 1).

Импорт данных из БД в программу осуществляется с помощью стандартных функций ADO.NET. Решающими параметрами выбраны: место проводимых измерений (на концентрацию вредных веществ в атмосфере имеет влияние рельеф местности и застройка); тип погоды; температура; давление; скорость ветра; количество машин по типам.

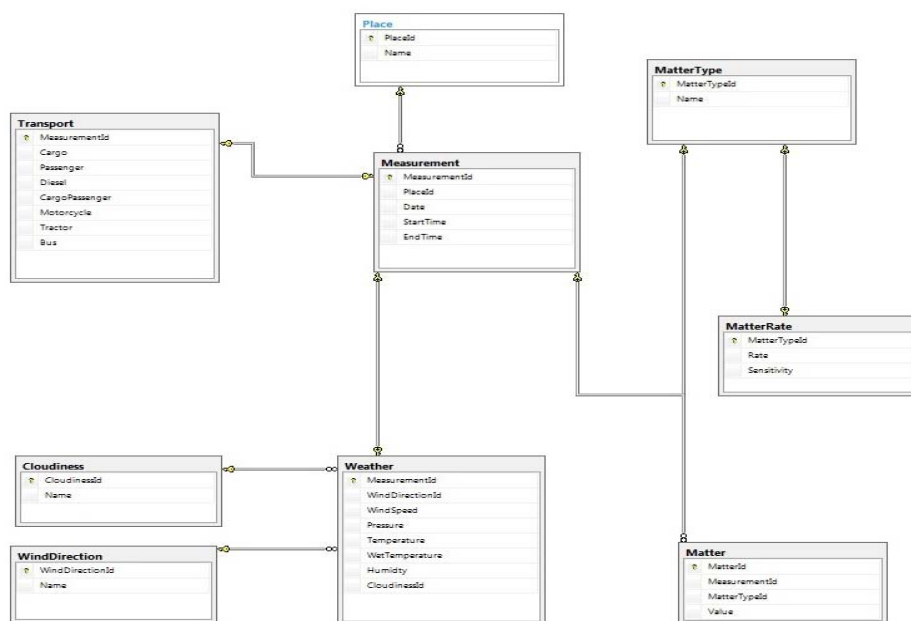


Рис. 1. Состав и структура базы данных

Общий вклад загрязняющих веществ, вносимых каждым элементом, определяется расчетным путем с использованием коэффициентов, полученных с применением методики, изложенной в [9]. При разных значениях коэффициентов проведена верификация базы данных и выбраны те, для которых получено наименьшее расхождение между исходными и прогнозируемыми значениями концентраций. Подобная практика использования коэффициентов встречается в работах [1], в которых оценивалось влияние каждого метеопараметра на степень загрязнения, и полученные результаты использовали для составления прогноза загрязнения.

После нормализации данные представляются в удобном для использования виде.

Результатом нормализации является модель данных, которую легко поддерживать, и она не содержит неопределенностей в данных и повторений. В данном случае все замеры для типов автомобилей находятся на некотором отрезке $[a,b]$, который нормализуется в $[0,1]$, и значение $x \in [a,b]$ вычисляется по формуле $x_1 = \frac{x-a}{l}$, $x_1 \in [0,1]$.

Кластерный анализ – задача разбиения множества замеров на подмножества, называемые кластерами, так, чтобы каждый кластер состоял из схожих объектов (замеров), а объекты разных кластеров существенно отличались. Кластер–группа элементов, характеризующихся общим свойством, главная цель кластерного анализа – нахождение групп схожих объектов в выборке. Решение задачи кластеризации зависит от параметров: критерия «похожести» элементов (после нормализации данные-замеры представляются в виде точек в евклидовом пространстве с набором координат-параметров, критерий похожести – это сходство координат, т.е. насколько близко в пространстве находятся точки); используемой метрики d , измеряющей расстояние между векторами-образами (пространство является евклидовым, метрику вычисляем стандартно); оцениваемого числа кластеров (оптимальное число кластеров ориентировочно предполагается равным b , что основывается на исходном количестве замеров); наличия нетипичных элементов (т.е. элементов, не попадающих ни в один из кластеров).

Задача кластеризации формулируется следующим образом. Дана обучающая выборка – множество всех замеров $\Omega = \{x_1, \dots, x_n\}$. Требуется найти такую функцию кластеризации f , которая каждой точке (замер с решающими параметрами) $x \in \Omega$ ставила бы в однозначное соответствие некоторый элемент – метку $z \in Z$ из множества меток $Z = \{z_1, \dots, z_n\}$. Обучающая выборка, используя один из алгоритмов кластеризации, разделяется на необходимые подгруппы внутри уже полученных классов по месту измерения и подклассов типа погоды.

Анализ алгоритмов кластеризации показывает, что для решения данной задачи рационально использовать максиминный алгоритм, поскольку он прост для реализации, обучается быстрее и не чувствителен к размерности (по сравнению с алгоритмом k -means). На рис. 2 приведен пример кластеризации.

1. В качестве первого центра кластера выбирается элемент $c_1 = x_1$.

2. Вторым центром кластера выбирается тот элемент $c_2 = x_{j_2}$, который находится на наибольшем расстоянии от c_1 , т.е. $\|x_{j_2} - c_1\| = \max_{x \in \Xi} \|x - c_1\|$.

3. Предположим, что выбраны k центров $C^{(k)} = \{c_1, \dots, c_k\}$ кластеров. В качестве очередного $(k+1)$ -го центра кластера выбирается тот элемент $x_{j_{k+1}}$, который находится на наибольшем расстоянии от ближайшего из центров c_1, \dots, c_k (рис. 2), т.е.

$$\min_{c \in C^{(k)}} \|x_{j_{k+1}} - c\| = \max_{x \in \Xi / C^{(k)}} \min_{c \in C^{(k)}} \|x - c\|.$$

4. Проверяется условие останова (число полученных кластеров равно шести).

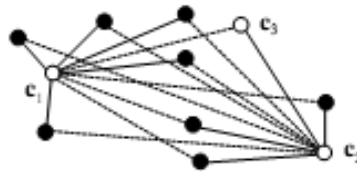


Рис. 2. Определение центров кластеров и относящихся к ним элементов (c_1, c_2, c_3 – центры кластеров)

Данные о концентрации примеси и относящиеся к ним ситуации группируются так, чтобы выделить 6 групп (кластеров) значений потока, причем сначала формируются группы по месту замера (количество – 6), затем подгруппы по типу погоды (количество – 3 для каждого класса). Пусть замеры описываются векторными характеристиками $\bar{x}_i = (x_{i1}, x_{i2}, \dots, x_{in})$, $i=1, 2, \dots, n$. Для группы, состоящей из M_i ситуаций, определяются центр тяжести i -й характеристики X_i и средняя дисперсия σ^2 .

Например, для 1-го кластера

$$x_i^{(1)} = \frac{1}{M_1} \sum_{j=1}^{M_1} x_{ij}^{(1)}, \quad \sigma_i^{(1)2} = \frac{1}{M_1} \sum_{j=1}^{M_1} [x_{ij}^{(1)} - x_i^{(1)}]^2.$$

Для каждой из групп выделяется по одному вектору-характеристике таким образом, чтобы эти векторы имели наиболее низкие степени близости между собой. При разбиении совокупности на два кластера ($m=2$) выбираются два вектора-представителя с минимальной близостью между ними, т.е. определяются номера i^*, j^* , из условия: $\alpha_{i^*j^*} = \min_{i,j} \alpha_{ij}$, где α_{ij} – числа, характеризующие подобие и определяющиеся формулой

$$\alpha_{ij} = \frac{1}{\|x_j - x_i\|}.$$

При $m=3$ к двум уже выбранным векторам V_1, V_2 добавляется третий вектор $V_3 = x_{k'}$, где k' выбирается из условия

$$\max_{i^*, j^* \in I} (\alpha_{i^* k'}, \alpha_{j^* k'}) = \min_k \max_{i^*, j^* \in I} (\alpha_{i^* k}, \alpha_{j^* k}),$$

которое позволяет выбрать третий вектор с наименьшей степенью близости совокупности первых двух.

Если уже выбраны $m-1$ векторов V_1, V_2, \dots, V_{m-1} , то вектор V_m выбирается из оставшихся векторов-характеристик x_i по условию

$$\begin{aligned} & \max\{\alpha(V_m, V_1), \alpha(V_m, V_2), \dots, \alpha(V_m, V_{m-1})\} = \\ & = \min_i \max\{\alpha(x_i, V_1), \alpha(x_i, V_2), \dots, \alpha(x_i, V_{m-1})\}. \end{aligned}$$

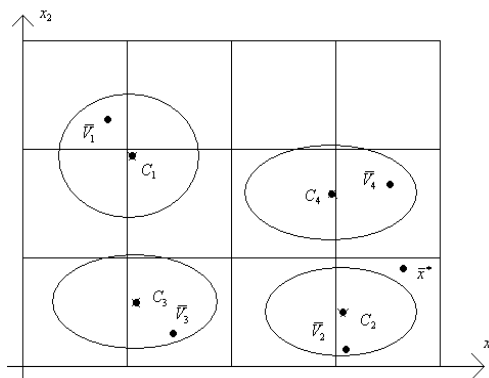


Рис. 3. Кластеры

Распределение на кластеры заканчивается, если

$$\max \|V_i - V_j\| \leq \varepsilon,$$

где ε – достаточно малое число.

Если считать, что среди проведенных наблюдений зафиксированы все возможные метеорологические ситуации, то прогноз методом распознавания образов связан с выбором из множества кластеров таких ситуаций (кластеров), которые близки к «прогнозируемой». В основе применения метода распознавания образов лежит представление о том, что большие концентрации примесей связаны с вполне определенными метеорологическими ситуациями и характером предшествующего загрязнения воздуха. Конкретный комплекс характеристик используется для определения группы, к которой можно отнести прогнозируемую ситуацию (процедура «обучения»). Для каждой конкретной ситуации, характеризуемой в фазовом пространстве точкой y с координатами $y_1, y_2 \dots y_N$, определяется расстояние до центра тяжести кластера:

$$\rho_i^2 = \sum_{i=1}^N \left(\frac{y_i - x_i^{(1)}}{\sigma_i^{(1)}} \right)^2.$$

Аналогично находится расстояние до остальных групп.

Ситуация u относится к той группе, расстояние до которой минимально. Далее считается средняя концентрация каждого вещества, характерного для данного кластера:

$$x_i = \frac{\sum_{j=1}^n x_j}{n},$$

где n – количество элементов в кластере, x_j – концентрация вещества.

Чтобы убедиться в том, что программа работает корректно, проведен ряд вычислительных экспериментов. Из БД случайным образом извлекается замер, который задается в качестве прогнозируемой ситуации, и далее сравниваются прогнозируемая концентрация и исходные данные.

Пример:

В качестве прогнозируемой ситуации вводится существующий замер, удаленный из базы (рис. 4). Работа программы отражена на рис. 4–6.

Рис. 4. Параметры прогнозируемой ситуации

ТипПогоды	Местоположение	СкоростьВетра	Температура	Грузовые	Легковые	Дизельные	Мот
Ясно	Гоголевский - Александровская	2	13	21	610	46	0
Ясно	Гоголевский - Александровская	2	13	48	500	0	0
Ясно	Гоголевский - Александровская	1.5	13	44	756	0	2

Рис. 5. Кластер, к которому отнесена прогнозируемая ситуация

Название	Значение	Норма
СО	7,184	5
Мп	0.0001	0.01
NO2	0.0093	0.085
Рь	0.0002	0.001
SO2	0.1307	0.5
Бензин	2.2724	5
Озон	0.0163	0.16
Пыль	0.3453	0.5
Сажа	0	0.15
Формальдегид	0.0231	0.035

Рис. 6. Прогнозируемая концентрация веществ для заданных условий

Вещество	NO2	SO2	Формальдегид	СО
Значение	0	0,12	0,029	7,5
ПДК	0,085	0,5	0,035	5

Рис. 7. Исходные данные по концентрации вредных веществ при заданных условиях

На рис. 6 и 7 результаты для сравнения: прогнозируемая концентрация близка к действительной. Проверка показала, что 11 из 15-ти замеров совпали по оценке (отклонение до 5 %) с тем прогнозом, что рассчитан программным комплексом, а интервал разброса для оставшихся 4-х замеров составляет от 18 до 25 %, что свидетельствует о достаточно успешной работе программного комплекса.

На основе полученной прогнозированием концентрации можно контролировать ситуацию загрязнения на автотранспортных магистралях.

Программный комплекс позволяет делать прогноз концентрации вредных веществ в воздушной среде в городских условиях и может использоваться в экологических службах, в частности в «Центре гигиены и эпидемиологии в Ростовской области».

Для повышения эффективности работы программного комплекса в дальнейшем предполагается добавление в него процедуры «переобучения». Приведенные в данной работе результаты подтверждают возможность построения таких схем на базе методов распознавания образов.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Берлянд М.Е. Прогноз и регулирование загрязнения атмосферы. – Л.: Гидрометеоздат, 1985. – 271 с.
2. «Руководящий документ. Охрана природы. Атмосфера. Руководство по прогнозу загрязнения воздуха». РД 52.04.306-92
3. Генихович Е.Л., Сонькин Л.Р., Гуцин В.А. О возможности прогноза загрязнения городского воздуха методом распознавания образов // Труды ГГО, 1969. – Вып. 238.

4. Интернет-ресурс: www.taganrog.su.
5. Интернет-ресурс: wikipedia.ru.
6. *Петрова В.С.* Диссертация по геоэкологии «Диагноз и прогноз уровня загрязнения атмосферы», интернет-ресурс www.nauka-shop.com.
7. *Лепский А.Е., Броневиц А.Г.* Математические методы распознавания образов: Курс лекций. – Таганрог: Изд-во ГТИ ЮФУ, 2009. – 155 с.
8. *Авдеева Т.П.* Расчет выброса загрязняющих веществ: Учеб. пособие. – Пенза, 1997. – 87 с.
9. Методика расчетов выбросов в атмосферу загрязняющих веществ автотранспортом на городских магистралях. – М.: НИИАТ, 1997. – 54 с.
10. *Берлянд М.Е.* Современные проблемы атмосферной диффузии и загрязнения атмосферы. – Л.: Гидрометеиздат, 1975. – 448 с.
11. *Семенченко Б.А., Белов П.Н.* Метеорологические аспекты охраны природной среды. – М.: Изд-во МГУ, 1984. – 96 с.
12. *Назаров И.М., Николаев А.Н., Фридман Ш.Д.* Основы дистанционных методов мониторинга загрязнения природной среды. – Л.: Гидрометеиздат, 1983. – 213 с.
13. *Беккер А.А., Агаев Т.Б.* Охрана и контроль загрязнения природной среды. – Л.: Гидрометеиздат, 1989. – 67 с.
14. Метеорологические аспекты загрязнения атмосферы // Сб. докладов на международном симпозиуме в Ленинграде / Под ред. Берлянда М.Е. – Л.: Гидрометеиздат, 1971. – 51 с.

Статью рекомендовал к опубликованию д.т.н., профессор Н.Н. Чернов.

Гадельшин Валерий Камельянович

Технологический институт федерального государственного автономного образовательного учреждения высшего профессионального образования «Южный федеральный университет» в г. Таганроге.

E-mail: valikam@gmail.com.

347928, г. Таганрог, пер. Некрасовский, 44.

Тел.: 88634371606.

Кафедра высшей математики; к.т.н.; доцент.

Мещерякова Юлия Александровна

E-mail: life-super@yandex.ru.

Тел.: +79043464554.

Кафедра высшей математики; бакалавр.

Погорелова Татьяна Ивановна

Тел.: +79185947452.

Кафедра высшей математики; бакалавр.

Gadelshin Valeriy Kamelianovich

Taganrog Institute of Technology – Federal State-Owned Autonomy Educational Establishment of Higher Vocational Education “Southern Federal University”.

E-mail: valikam@gmail.com.

44, Nekrasovskiy, Taganrog, 347928, Russia.

Phone: +78634371606.

The Department of High Mathematic; Cand. of Eng. Sc.; Associate Professor.

Meshcheryakova Julia Aleksandrovna

E-mail: life-super@yandex.ru.

Phone: +79043464554.

The Department of High Mathematic; Bachelor.

Pogorelova Tatiana Ivanovna

Phone: +79185947452.

The Department of High Mathematic; Bachelor.