

УДК 504.064.36; 004.8.032.26

О.С. Масич

Полтавський національний технічний університет імені Юрія Кондратюка, Полтава

НЕЙРОМЕРЕЖЕВІ МОДЕЛІ ЕКОЛОГІЧНОГО МОНІТОРИНГУ ТЕРИТОРІЇ

Обґрунтовано актуальність проблеми розпізнавання внутрішньої структури даних екологічного моніторингу і побудови математичної моделі техногенного навантаження регіону. Запропоновано методику побудови багатовимірної регресії для кожного кластера в форматі безлічі вагових коефіцієнтів навченої нейронної мережі. Знайдений ансамбль моделей з прийнятною для практики продуктивністю при обґрунтованій надійності і точності отриманих результатів. Показана реалізація запропонованих алгоритмів в середовищі стандартного пакета технічного аналізу, що дозволяє уникнути додаткових матеріальних і тимчасових витрат при створенні систем підтримки прийняття рішень.

Ключові слова: екологічний моніторинг, штучна нейронна мережа, багатовимірні регресія, розпізнавання, навчання.

Вступ

Екологічний моніторинг є комплексним моніторингом біосфери. Він включає в себе контроль за зміною стану навколишнього середовища під впливом як природних, так і антропогенних чинників [1].

Постановка проблеми. Для ефективного аналізу впливу техногенного навантаження на навколишнє середовище в регіоні необхідно знати внутрішню структуру екологічних факторів, їх вплив на показники якості і динаміку їх змін. Для цього доцільно знайти аналітичну залежність вхідних екологічних факторів з індикаторами екології в регіоні. Прийняття рішень щодо корекції екології залежить від значення детермінують сили кожного фактору окремо і його вкладу при спільному багатofакторному впливі всього масиву факторів на цільову функцію.

Якщо побудувати інформаційно-аналітичну модель аналізу даних екологічного моніторингу з оцінкою цих показників, то за інших рівних умов, отримані знання можуть істотно вплинути на домінуючі індикатори техногенного навантаження в регіоні [2 – 4].

Техногенний вплив на середовище і можна характеризувати рядом показників екологічного моніторингу, а індикатором цього впливу розглядати, наприклад, статистику захворювань серед жителів цього регіону. На прикладі [4] представлений масив даних, який можна розглядати як вибірку прикладів для навчання моделі.

Аналіз наявних даних дозволяє зробити висновок про різноманіття і різну природу факторів, що говорить про високий ступінь невизначеності в зв'язку індикаторів екології та статистикою захворювань в регіоні. Тому побудова подібних аналітичних моделей є актуальним завданням з практичної точки зору.

Основний матеріал дослідження та отримані наукові результати

Публікації в цій галузі досліджень дозволяють стверджувати, що ряд питань не дозволені і потрібно шукати механізми оцінки і прогнозу стану і динаміки навколишнього середовища. Перспективним відзначають напрямком досліджень на основі інтелектуальних технологій вилучення знань з масиву даних за результатами моніторингу екологічного навантаження [5]. Основу цього напрямку складає застосування інтелектуальних технологій, побудованих на принципах навчання і реалізованих в форматі пакетів технічного аналізу [4,5,6,7]. Аналіз публікацій дозволяє зробити висновок про доцільність постановки та вирішення задачі моделювання процесу оцінки і управління станом екології в регіоні з метою мінімізації техногенного навантаження. Якщо знайти адекватну математичну модель, яка б пов'язала дані екологічного моніторингу ґрунту, води, повітря з індикаторами стану здоров'я жителів регіону, то можна побудувати оптимальну систему прийняття рішень. Це особливо важливо при обмежених ресурсах і необхідності їх продуктивного точкового використання з максимальною ефективністю. Однак в умовах великої розмірності вектора вхідних даних (екологічні показники) і складності взаємозв'язків з індикаторами техногенного навантаження (статистика захворюваності) побудувати багатовимірну регресію для всього масиву даних з хорошою продуктивністю, як правило, не вдається [4, 6, 9]. Тому логічно припустити, що знаходженню багатofакторної залежності, як вирішення завдання багатовимірної регресії, повинна передувати процедура структурування вихідних даних з метою формування однорідних підмножин. Тоді значно легше домогтися заданих меж продуктивності при побудові моделі регресійного аналізу для кожної з отрима-

них підмножин. Вирішенню цього завдання присвячена ця стаття.

Для побудови програмної моделі аналізу структури вхідних даних використовується середовище стандартного пакету технічного аналізу Statsoft (версія Statistika 6.1) і забезпечується адаптація емуляторів базових функцій до задачі регресійного аналізу техногенного навантаження регіону в рамках встановлених обмежень.

Необхідно побудувати математичну модель внутрішньої структури факторів техногенного навантаження на основі вихідних даних екологічного моніторингу, яка по заданому вхідному вектору спостережень видавала б відгук із заданою точністю:

$$F: X \rightarrow Y_{opt}, X \in \mathfrak{R}^m, Y_{opt} \in \mathfrak{R}, \quad (1)$$

де X - множина вибірок ознак опису екологічного стану;

Y_{opt} - верифіковане вихідне значення функції регресії.

У нейромережевому форматі ця задача може бути представлена у такому вигляді:

$$y^k(x^m) = \sum_{j=1}^n v_j (w_{j1}x_1 + w_{j2}x_2 + \dots + w_{jm}x_m + u_j), \quad (2)$$

де y^k - вектор станів;

k -номер індикатора станів (дихання, кров, % звернень до лікаря і т.д.);

(x^m) - вектор факторів;

m - розмірність вхідного вектору даних - (NO, SO₂, CO і т.д.);

n - потужність навчальної вибірки;

v_j - параметри нейромережі;

$w_{j1}, w_{j2}, \dots, w_{jm}$ - вагові (синаптичні) коефіцієнти нейромережевої моделі.

Залежність (2) буде представляти собою безперервну функцію [9].

Збудуємо мережу для багатовимірної регресії, отримаємо фізичну модель техногенного навантаження як реалізацію функції (1). Якщо продуктивність мережі і помилки на навчальній, контрольній і тестовій множині в допустимих межах, то модель стає інструментом підтримки прийняття рішень в оцінці техногенного навантаження регіону. Прогноз динаміки захворювань і вибір оптимальних керуючих факторів для досягнення цільового стану мають детермінований зв'язок через масив синаптичних ваг навченої нейронної мережі.

Знаходження однорідних підмножин у вхідному масиві даних розглянемо як завдання кластерного аналізу. Є кінцевий масив навчальних вибірок $X^n = \{x_1, x_2, \dots, x_n\} \in X$. Відомий вид функції відстані між об'єктами $\rho(x, x^*)$ (евклідова відстань). Необхідно розділити масив даних на непересічні підмножини, так, щоб кожна складалася з об'єктів, близь-

ких за метрикою $\rho(x, x^*)$, а об'єкти різних підмножин істотно відрізнялися між собою.

Вирішивши завдання кластерного аналізу і домігшись однорідності даних вибірок в кожній підмножині, залежність «вхід-вихід», буде являти собою безперервну функцію [7].

Необхідно побудувати для кожного з кластерів окрему нейронну мережу для багатовимірної регресії, отримати фізичну модель техногенного навантаження як реалізацію функції (1).

Якщо продуктивність мережі і помилки на навчальній, контрольній і тестовій множині в допустимих межах, то модель стає інструментом підтримки прийняття рішень в оцінці техногенного навантаження регіону.

Прогноз динаміки захворювань і вибір оптимальних керуючих факторів для досягнення цільового стану мають детермінований зв'язок через масив синаптичних ваг навченої нейронної мережі.

Рішення завдання ґрунтується на добуванні знань з масиву прикладів шляхом навчання нейромережевої моделі аналізу даних. В якості програмного інструментарію використовуємо пакет технічного аналізу Statistika Neural Network.

З огляду на те, що число розглянутих факторів аналізу техногенного навантаження досягає декількох десятків [4], ці фактори, як правило, впливають один на одного, частина з них зашумлена або відсутня, рішення доводиться приймати в умовах високої апріорної невизначеності.

У цих умовах найбільш ефективні методи і алгоритми нейромережевого розпізнавання образів [5, 6, 8].

Усередині кластерів дані однорідні, а, значить, поведінка фізичного процесу (техногенного навантаження) в рамках одного кластера більш передбачувана, ніж динаміка цього процесу у всьому просторі спостережень наявних прикладів.

Таким чином, якщо будуть знайдені угруповання в даних моніторингу як значення факторів компактних в багатовимірному просторі на основі деякої метрики близькості, то нові спостереження можна класифікувати за належністю до того чи іншого кластеру.

Якщо на вхід навченої моделі подати вибірку значень факторів заданої розмірності, модель віднесе її до кластеру з мінімальним значенням міри близькості. Це еквівалентно присвоєнню номера класу техногенного стану по досліджуваній вибірці даних.

Маніпуляції розмірністю вектора і інформативністю його елементів з багаторазовим повторенням цієї процедури кластеризації, дозволяють емпіричним шляхом знайти сукупність факторів відповідних потрібного класу екологічного навантаження регіону [5 – 7].

Скористаємося узагальненими показниками екологічного моніторингу на прикладі завдання, складемо вихідну таблицю даних в пакеті технічного аналізу Statistika 6.1.

Проведемо кластеризацію даних в просторі спостережень методом К-середніх.

Мета методу — розділити n спостережень на k кластерів, так щоб кожне спостереження належало до кластера з найближчим до нього середнім значенням. Метод базується на мінімізації суми квадратів відстаней між кожним спостереженням та центром його кластера.

Таким чином, можна стверджувати, що всі множини даних добре розділяються на три кластери. Варто відзначити, що дані таблиць добре ілюструють їх стійке угруповання всередині кластерів і надійний поділ спостережень на однорідні групи між кластерами. Алгоритми, що реалізують процедури угруповання даних, забезпечують однорідність даних усередині кластерів і максимальну віддаленість самих кластерів. Отже, умови (1) і (2) при виконанні процедури формування однорідних груп у вихідних даних дотримані.

На наступному етапі реалізуємо процедуру (1) як завдання багатовимірної регресії. Для цього побудуємо алгоритм аналізу даних в режимі «навчання з учителем» [3].

Процедура (2) в постановці завдання виконується ітераційно в процесі навчання мережі за алгоритмом зворотного поширення помилки [7] стосовно до даних формату екологічного навантаження регіону:

$$\frac{1}{mn} \sum_{j=1}^n \sum_{i=1}^m (x_{ij} - d_{ij})^2 \Rightarrow \min,$$

де u_{ij} - вектор індикаторів вихідних станів;

d_{ij} - результат навчання мережі на j - виході, при i - му прикладі навчальної вибірки.

$j=1, n$ - номер виходу мережі;

$i=1, m$ - номер прикладу;

n, m - розмірність масиву прикладів і числа вихідних елементів мережі.

Після виконання навчання отримаємо ансамбль нейромережевих моделей (рис. 1 – 3).

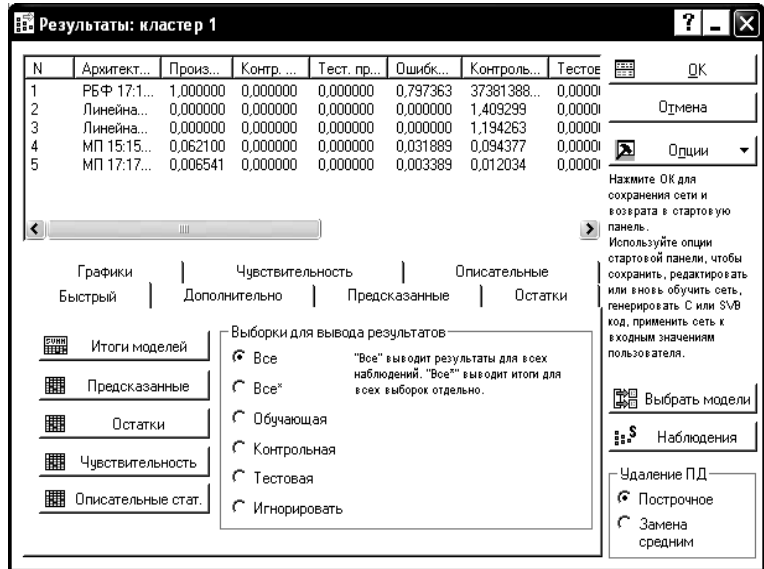


Рис. 1. Результаты для 1-го кластеру

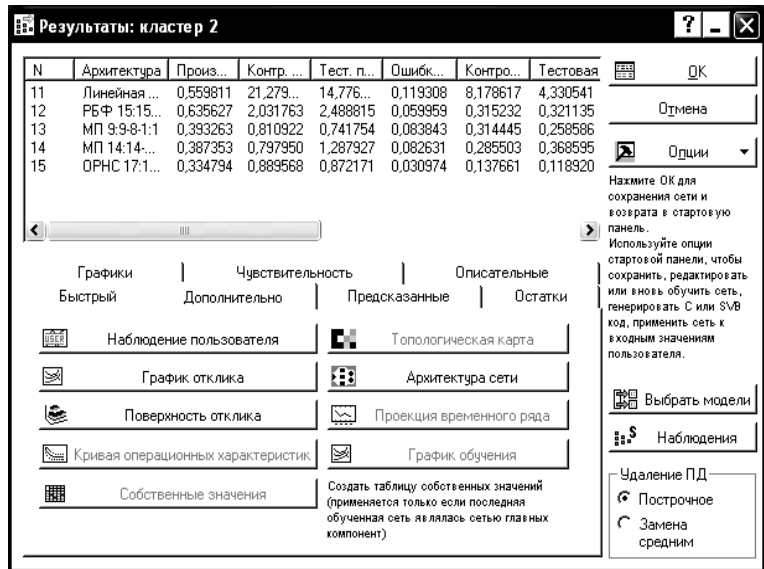


Рис. 2. Результаты для 2-го кластеру

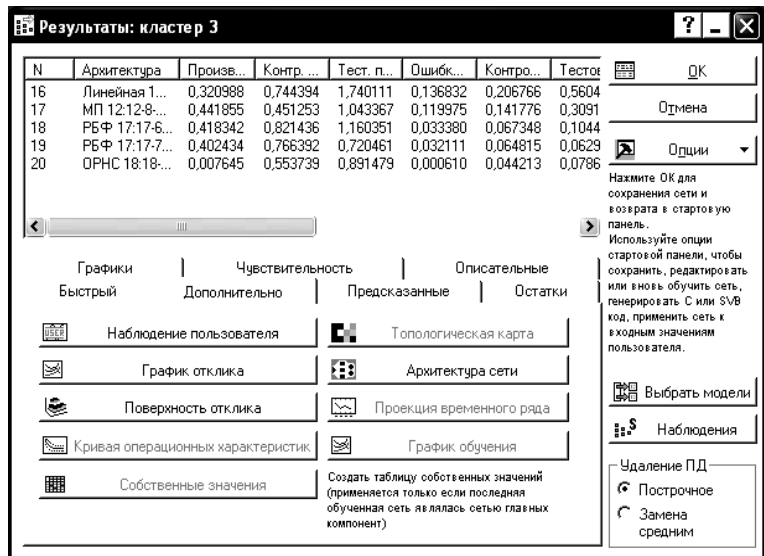


Рис. 3. Результаты для 3-го кластеру

Аналіз профілів результатів моделювання дозволяє виділити мережі з прийнятними для практики продуктивністю і помилками на навчальних, контрольних та тестових множинах. Це свідчить про адекватність математичної моделі фізичному змісту досліджуваного процесу, який формалізований виразом (2) в постановці завдання.

Знайдена регресійна залежність встановлює зв'язок входних факторів і вихідних індикаторів системи через масиви синаптичних коефіцієнтів п'яти навчених нейромережевих моделей для кожного з кластерів.

Мережа успішно навчилася і придатна для використання в оцінці та прогнозуванні значень залежної змінної. Таким чином, математична модель екологічного моніторингу регіону побудована. Постановлена задача вирішена.

Висновки

Отже, моделі інтелектуального вилучення знань з бази регіональних даних екологічного навантаження на навколишнє середовище в регіоні були успішно побудовані. Була проведена кластеризація даних методом К-середніх, що якнайкраще підходить для наведеного типу даних. Результати кластеризації також дають можливість в майбутньому, при надходженні нових спостережень, на основі деякої метрики близькості класифікувати їх за належністю до того чи іншого кластеру.

Після отримання трьох кластерів було проведено навчання нейронної мережі за допомогою регресії через майстер рішень. В результаті навчання були отримані архітектури, продуктивності яких не перевищували 0,3, або були близькі до даного числа, що при багатовимірному регресійному аналізі вірно.

В кінцевому результаті була побудована математична модель техногенного навантаження регіону. Аналіз екологічного навантаження регіону як рі-

шення задачі розпізнавання реалізується в пакеті технічного аналізу StatSoft успішно. Новизна підходу полягає в розробці методики вилучення знань для прийняття рішень з масиву даних екологічного моніторингу регіону шляхом використання стандартного нейромулятор Statistika Neural Network.

Список літератури

1. Хотунцев Ю.Л. Экология и экологическая безопасность. Учебное пособие для студентов высших педагогических учебных заведений / Ю.Л. Хотунцев. – М.: Academia, 2004. – 480 с.
2. Бусленко Н.П. Моделирование сложных систем. 2 – изд. – М.: Наука, 1978. – 400с.
3. Хайкин С. Нейронные сети: полный курс, 2-е изд. Пер. с англ. - М.: Изд. дом «Вильямс», 2006. – 1104 с.
4. Морозов А.А. Состояние и перспективы нейросетевого моделирования СППР в сложных социотехнических системах / В.П. Клименко, А.Л. Ляхов, С.П. Алёшин // Математичні машини і системи. – 2010. - № 1.- С. 127 – 149.
5. Ляхов А.Л. Интеллектуальный анализ данных в прикладных экономических задачах / А.Л. Ляхов, С.П. Алёшин // Научный вестник Полтавского национального технического университета. Экономика і регіон – 2009. - № 4(23). – С. 140 – 147.
6. Барабаи Ю. Л. Коллективные статистические решения при распознавании.- М.: Радио и связь , 1983.- 224с.
7. Боровиков В.П. STATISTICA NN - Техническое описание. М.: Мир, 1999. – 239 с.
8. Колмогоров А.Н. О представлении непрерывных функций нескольких переменных в виде суперпозиций непрерывных функций одного переменного и сложения / Колмогоров А.Н. // Доклады АН СССР. – 1957. – Т. 114. – С. 953 – 956.
9. Ляхов А.Л. Сложная социотехническая система как объект управления искусственной нейронной сетью / А.Л. Ляхов, С.П. Алёшин // Вісник інженерної академії України. – 2010. – № 1. – С. 93 – 97.

Надійшла до редколегії 14.10.2015

Рецензент: канд. техн. наук, доц. С.П. Альошин, Полтавський технічний університет ім. Ю. Кондратюка, Полтава.

НЕЙРОСЕТЕВЫЕ МОДЕЛИ ЭКОЛОГИЧЕСКОГО МОНИТОРИНГА ТЕРРИТОРИИ

А.С. Масыч

Обосновано актуальность проблемы распознавания внутренней структуры данных экологического мониторинга и построения математической модели техногенной нагрузки региона. Предложена методика построения многомерной регрессии для каждого кластера в формате множества весовых коэффициентов обученной нейронной сети. Найденный ансамбль моделей с приемлемой для практики производительностью при обоснованной надежности и точности полученных результатов. Показана реализация предложенных алгоритмов в среде стандартного пакета технического анализа, что позволяет избежать дополнительных материальных и временных затрат при создании систем поддержки принятия решений.

Ключевые слова: экологический мониторинг, искусственная нейронная сеть, многомерная регрессия, распознавание, обучение.

NEURAL NETWORK MODEL OF ENVIRONMENTAL MONITORING OF AREA

O.S. Masych

There was justified by the relevance of the problem of recognition of internal data structures for environmental monitoring and constructing a mathematical model of anthropogenic impact in the region. Constructing multivariate regression technique has been proposed for each cluster in the format of the plurality of weights trained neural network. Found ensemble models with an acceptable practice performance with reasonable accuracy and reliability of the results. Shown the implementation of the proposed algorithms in the environment of the standard technical analysis package, which avoids additional material and time costs at creation of decision support systems.

Keywords: environmental monitoring, artificial neural networks, multivariate regression, pattern recognition, learning.