

УДК 004.055

К. П. СТОРЧАК, канд. техн. наук, доцент;

А. М. ТУШИЧ,

К. С. КОЗЕЛКОВА, доктор техн. наук, професор;

М. М. СТЕПАНОВ, доктор техн. наук, ст. наук. співробітник,

Державний університет телекомунікацій, Київ

ІНТЕЛЕКТУАЛЬНИЙ АНАЛІЗ ДАНИХ ІЗ ВИКОРИСТАННЯМ НЕЙРОННИХ МЕРЕЖ

У статті розглянуто штучні нейронні мережі як засіб інтелектуального аналізу даних. Актуальність використання таких мереж зумовлено активним зростанням обсягів даних, що відбувається внаслідок автоматизації різного роду технологічних процесів. Визначено основні методи виконання аналізу даних, а також з'ясовано переваги і недоліки зазначених методів. У результаті обґрунтовано важливий висновок: застосування штучних нейронних мереж з метою інтелектуального аналізу накопичуваних даних цілком виправдане й доцільне, передусім із практичних міркувань. Наведено модель відповідної системи та описано процес інтелектуального аналізу даних на основі штучних нейронних мереж. Розглянуто етапи виконання цього процесу. Запропоновано алгоритм побудови зазначеної системи.

Ключові слова: інтелектуальний аналіз даних; штучна нейронна мережа; алгоритм.

Вступ

Розвиток автоматизованих систем управління технологічними процесами в поєднанні з подальшим підвищенням продуктивності автоматизованих робочих місць і серверного обладнання став одним із головних чинників появи великих обсягів даних, що зберігаються у відповідних базах. У процесі інтелектуального аналізу зазначених обсягів даних головним завданням є виявлення правил і закономірностей, яким підлягають існуючі дані.

Одним із ефективних інструментів інтелектуального аналізу даних дотепер була математична статистика. Проте такий аналіз має базуватися на перевірці визначеної статистичної гіпотези [1].

На сучасному етапі розвитку цієї сфери крім математичної статистики застосовують такі методи інтелектуального аналізу даних, як нейронні мережі, дерева рішень, генетичні алгоритми, нечітка логіка, алгоритми обмеженого перебору, еволюційне програмування, системи умовиводів на основі аналогічних випадків, індукція правил, аналіз із вибірковою дією, логічна регресія, алгоритми визначення асоціацій і послідовностей, візуалізація даних, комбіновані методи [1].

Для моделювання будь-яких систем аналізу даних необхідно створювати складні математичні структури, що надалі підлягають програмуванню. Системи можуть бути різні, наприклад дискретні чи неперервні або мішаного типу. Часто кожен із варіантів розв'язання поставленої проблеми має як переваги, так і недоліки, а тому вибір характеристик моделі зусиллями самої людини зробити досить складно.

Саме тут здатність комп'ютера самостійно підбирати параметри моделі допомогла б розв'язати цю проблему [2].

Допустимість високого рівня зашумленості даних і низька ймовірність помилки, а також безперервність процесу навчання мережі, постійне удосконалення її структури — усе це дає підстави вважати застосування нейронних мереж одним із найбільш перспективних методів аналізу даних. Адже використання нейронних мереж дозволяє аналізувати стан та ідентифікувати параметри об'єктів навіть у разі неповних даних [1].

Основна частина

Штучна нейронна мережа являє собою математичну модель людського мозку. Мозок складається з величезної кількості частинок — нейронів, з'єднаних між собою лініями зв'язку, які передають сигнали. Проходячи через нейрони, сигнали змінюються. Наявність незліченної кількості таких нейронів дозволяє людині розв'язувати складні задачі, запам'ятовувати великі обсяги інформації та приймати рішення [2].

Створення такої моделі дозволило навчити комп'ютер виконувати певні самостійні дії. Для навчання нейронної мережі надається навчальна вибірка, і мережа самостійно підстроює свою структуру для оптимального моделювання цієї вибірки. Після навчання мережа виявляє здатність видавати результат обчислення, маючи на вході дані, яких не було у вибірці, котру подавали на навчання мережі [2].

Здатність нейронної мережі діяти в такий спосіб істотно полегшує моделювання систем, зокрема й для аналізу даних. Необхідно лише визначитись із вхідними даними та з'ясувати, які мають бути вихідні дані, зібравши потрібні експериментальні дані та дати мережі змогу навчатись. Здатність мережі самостійно змінювати свою структуру позбавить необхідності керувати навчанням [2].

Розв'язання задачі інтелектуального аналізу даних пропонується проводити в три етапи. На першому необхідно визначитись із типом штучної нейронної мережі, на другому — із параметрами архітектури мережі (кількість нейронних шарів тощо), на третьому здійснити навчання мережі, тобто задати мережі необхідні параметри, що визначають правильний вихідний результат, або скоригувати на початкових етапах, якщо навчання відбувається «з учителем».

Основні **моделі нейронних мереж**, здатних реалізувати поставлене завдання, поділяють на три групи.

1. Мережі прямого поширення. Одні з найбільш розповсюджених. Найчастіше використовуються для прогнозування та розпізнавання образів, зокрема для аналізу даних.

2. Мережі зі зворотним зв'язком, такі як дискретна модель Хопфілда. Здебільшого використовуються для оптимізації розрахунків та задіяння асоціативної пам'яті.

3. Самоорганізовані мережі, скажімо моделі адаптивної резонансної теорії та моделі Кохонера. Переважно використовуються для кластерного аналізу [1; 2].

Однією з головних переваг нейронних мереж є можливість апроксимації з будь-якою наперед заданою точністю на компактній множині довільної неперервної функції. Це дозволяє дослідникові не приймати заздалегідь будь-які гіпотези стосовно моделі.

До недоліків відносять той факт, що остаточний розв'язок залежить від початкових налаштувань мережі, а через це його практично неможливо інтерпретувати у традиційних аналітичних термінах [1; 2].

Процес аналізу даних відбувається також у три етапи: 1) підготовка даних; 2) аналіз даних; 3) вираження та інтерпретація результатів (див. рисунок).

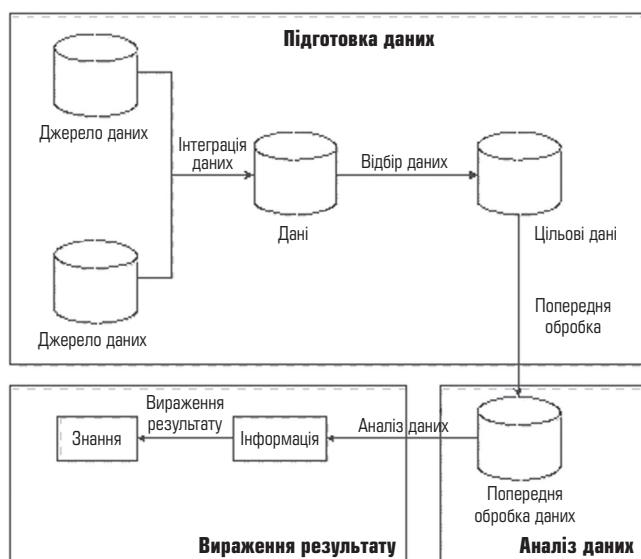


Схема етапів аналізу даних

Підготовка даних у цій задачі не потребує складних маніпуляцій, пов'язаних із перетворенням у цифрову форму.

Попередня обробка даних є розширенням процесу очищення даних, які було вибрано. Місце отримання некоректних або зашумлених даних також визначається нейронною мережею.

Вираження даних має перетворити дані після попередньої обробки в ту форму, яка може бути прийнята за умовою алгоритму аналізу даних, здійснюваного на основі нейронної мережі.

Аналіз даних, базований на нейронній мережі, може виконуватись тільки з числовими даними, а отже, необхідно символічні дані перетворювати в числові. Найпростіший спосіб такого перетворення полягає в побудові таблиці відповідностей між символічними та числовими даними. Інший, складніший підхід полягає в прийнятті хеш-функцій для створення унікальних числових даних, які відповідають узятому рядку.

Незважаючи на те, що в реляційній базі даних є безліч типів даних, усі вони здебільшого можуть бути зведені до символічних, дискретних числових і неперервних числових даних, тобто до трьох логічних типів даних.

Далі необхідно виявити правила навчання підготовленої мережі. Найчастіше для цього використовують метод LRE (*Limited Relative Error*), метод чорного ящика, метод вилучення нечітких правил, метод вилучення правил із рекурсивної мережі, алгоритм вилучення правил двійкового входу і виходу (BIO-RE), алгоритм часткового вилучення правил (Partial-RE) і алгоритм повного вилучення правил (Full-RE).

І, нарешті, необхідно здійснити оцінювання роботи мережі. Розрізняють чотири види такого оцінювання:

- 1) знайти оптимальну послідовність вилучення правил;
- 2) перевірити точність витягнутих правил;
- 3) визначити кількість знань у нейронній мережі, які не були сприйняті;
- 4) визначити суперечності між витягнутими правилами і навченою нейронною мережею [1].

Висновки

Основною перевагою застосування штучних нейронних мереж є можливість розв'язувати різні неформалізовані задачі. Використання таких мереж для аналізу даних доцільне, оскільки ці мережі здатні до апроксимації функцій, до навчання та вдосконалення власної структури, забезпечують низьку ймовірність помилки за умов коректного початкового налаштування параметрів мережі, а також можливості аналізу навіть за наявності неповних і зашумлених даних. При цьому є змога

досить просто моделювати ситуації, подаючи на вхід мережі дані та оцінюючи результат, який вона видає.

Запропонований алгоритм розв'язання поставленого завдання достатньо повно описує процес виконання необхідного аналізу.

Список використаної літератури

1. Кулаков П. А. Основные классы нейронных сетей в задаче диагностики технологического оборудования // Молодежный научный вестник: электронный науч.-практ. журн. 2016. № 10. С. 74–77.

2. Манжула В. Г., Федяшов Д. С. Нейронные сети Кохонена и нечеткие нейронные сети в интеллектуальном анализе данных // Fundamental research. 2014. №4. С. 108–114.

3. Бирюков Е. В., Корнев М. С. Практическая реализация нечеткой нейронной сети при краткосрочном прогнозировании электрической нагрузки // Портал магистров ДонНТУ. URL:

<http://masters.donntu.org/2006/kita/chuykov/links/index.htm> (дата звернення 02.10.2018).

4. Дюк В. А. Data Mining — интеллектуальный анализ данных // Информационные технологии. URL:

<http://www.inftech.webservis.ru/it/database/datamining/ar2.html> (дата звернення 01.11.2018)

5. Назаров А. В., Лоскутов А. И. Нейросетевые алгоритмы прогнозирования и оптимизации систем. СПб.: Наука и Техника, 2015. 384 с.

Рецензент: канд. техн. наук, ст. наук. співробітник Ю. В. Мельник, Державний університет телекомунікацій, Київ.

К. П. Сторчак, А. Н. Тушич, Е. С. Козелкова, М. Н. Степанов

ИНТЕЛЛЕКТУАЛЬНЫЙ АНАЛИЗ ДАННЫХ С ИСПОЛЬЗОВАНИЕМ НЕЙРОННЫХ СЕТЕЙ

В статье рассмотрены искусственные нейронные сети как средство интеллектуального анализа данных. Актуальность использования указанных сетей обусловлена стремительным возрастанием объемов данных в результате автоматизации различного рода технологических процессов. Определены основные методы анализа данных, а также раскрыты их важнейшие преимущества и недостатки. В результате получен принципиальный вывод: использование искусственных нейронных сетей для интеллектуального анализа данных весьма целесообразно, прежде всего из практических соображений. Построена модель системы и описан процесс интеллектуального анализа данных, основанный на искусственных нейронных сетях. Рассмотрены этапы выполнения этого процесса. Предложен алгоритм построения указанной системы.

Ключевые слова: интеллектуальный анализ данных; искусственная нейронная сеть; алгоритм.

K. Storchak, A. Tushych, K. Kozelkova, M. Stepanov

INTELLECTUAL ANALYSIS OF DATA USING NEURAL NETWORKS

The article considers artificial neural networks as a means of intellectual data analysis. The urgency of using the proposed data analysis means lies in the active growth of data as a result of automation of various technological processes. The main methods for performing data analysis are analyzed in the paper, and their advantages and disadvantages are identified. As a result, it was concluded that the use of artificial neural networks is expedient. The main advantage of artificial neural networks usage is the ability to solve various informational tasks. The use of such networks for data analysis is advisable because they have properties of function approximation, learning ability, improvement of their own structure, low probability of error with correct initial setup of network parameters, analysis capabilities even in the presence of incomplete and noisy data. At the same time, it is possible to simulate the situation simply by submitting the data to the input of the network and evaluating the result that it issues. In the work the model of the system is constructed and the process of the intellectual data analysis is based on artificial neural networks and stages of its implementation. The proposed algorithm for solving the problem adequately describes the process of performing an analysis that goes through three stages: data preparation (does not require complex manipulation associated with digitization, preprocessing data is an extension of the data purification process), data analysis (the neural network can only work with numerical data, which implies that it is necessary to convert symbol data into numerical ones by creating correspondence tables between symbolic data and numeric or accepting hash-functions for creation unique numeric data), Expression (transformation data after pretreatment in the form which can be adopted subject to the data analysis algorithm) and interpretation of results (in user-friendly form).

Keywords: data mining; artificial neural network; algorithm.