

АНОТАЦІЯ

Тушич А.М. Методика побудови інтелектуальної системи аналізу даних на основі нейронних мереж. – Кваліфікаційна наукова праця на правах рукопису.

Дисертація на здобуття наукового ступеня доктора філософії за спеціальністю 123 – комп'ютерна інженерія (галузь знань 12 – інформаційні технології). – Державний університет телекомунікацій Міністерства освіти і науки України, Київ, 2021.

Дисертаційна робота присвячена актуальній науковій задачі розробки методики побудови інтелектуальної системи аналізу даних на основі нейронних мереж. Тематика дисертаційного дослідження відповідає тимчасовому стандарту та фаховим компетентностям освітньо-наукової програми підготовки докторів філософії з комп'ютерної інженерії Державного університету телекомунікацій Міністерства освіти і науки України, а саме: фундаментальним науковим дослідженням теоретико-методологічних, науково-методичних та прикладних засад підвищення ефективності інноваційної та виробничої діяльності підприємства, а також вдосконаленню процесу забезпечення впровадження новітніх інформаційних технологій на об'єктах інформаційної діяльності.

Для забезпечення заданих високих вимог до кількісних і якісних показників обробки інформації, що є процесом інформаційної взаємодії технологічних процесів сучасних підприємств і організацій, необхідна реалізація інтелектуальної системи аналізу даних.

Для досягнення цього завдання необхідно спроектувати таку систему, яка буде стійкою до відмов, мати можливість до автоматичного відновлення, бути байдужою до потрапляння зашумлених даних та показувати результати у прийнятному для користувача вигляді.

У роботі проведено порівняльний аналіз технологій, які вирішують задачу аналізу накопичених даних, що отримують підприємства у процесі своєї

діяльності. Однак розробці методів і програм для аналізу даних, які можуть ідентифікувати потенційно корисну, але неявну інформацію, приділяється не багато уваги. Отримання цієї інформації може дати життєво важливий імпульс науковим дослідженням в інших областях. Цей нетривіальний витяг неявної, раніше невідомої і потенційно корисної інформації з великих баз даних відомий як інтелектуальний аналіз даних або виявлення знань.

Одна з вимог до інтелектуальних систем виявлення знань – ефективність і масштабованість. Робота з дуже великими базами даних вимагає ефективних алгоритмів, а неточність і часто неповнота даних створює додаткові проблеми при отриманні прихованих шаблонів. Навчені нейронні мережі здатні генерувати приховані знання з даних: створюється можливість прогнозування, класифікації та розпізнавання образів, але їх логічна структура з традиційним підходом залишається прихованою від користувача.

У дисертаційній роботі сформульовано вимоги до сучасних систем аналізу даних для обґрунтування вибору математичного апарату ядра аналітичної системи, розроблено новий метод нелінійної нормалізації даних, який полягає у поступовій реалізації перетворень над даними із змінним типом нелінійності, досліджено, які нейронні мережі мають перевагу над традиційними методами математичної статистики та технічного аналізу для визначеної задачі, обрано оптимальну топологію і параметри обраної технології, розроблено новий підхід до явного отримання правил від навченої нейронної мережі, який полягає у групуванні вхідних параметрів та активності нейронів, розроблено інформаційну технологію визначення пошуку правил функціонування досліджуваного об'єкта в явній формі на основі обробки ряду емпіричних даних за допомогою апарату нейронної мережі.

Метою даної дисертації є оптимізація процесу виявлення прихованих закономірностей, що містяться в базах даних за допомогою вдосконаленої методики обробки даних.

Для досягнення цієї мети необхідно вирішити такі *задачі*:

1. Визначити вимоги до сучасних аналітичних систем, виходячи з представлених вимог, для обґрунтування вибору математичного апарату ядра аналітичної системи.
2. Розробити метод попередньої обробки даних, що аналізуються.
3. Провести дослідження на основі побудови алгоритму аналізу даних, використовуючи обрану нейромережеву технологію.
4. Обрати оптимальну топологію і параметри обраної топології.
5. Розробити алгоритм виявлення знань в аналізованих даних.
6. Розробити програмну реалізацію інформаційної технології для імітаційного моделювання процесів системи.

Методи дослідження. Результати проведених і представлених в дисертації досліджень отримані з використанням методів системного аналізу, теорії інформації, теорії ймовірностей, методів статистичного аналізу, комбінаторики, нейромережевого моделювання, кластерного аналізу, оптимізації, формалізації та методів динамічного програмування.

Наукова новизна дисертації:

1. Вперше розроблено метод нелінійної нормалізації даних, який ґрунтується на послідовному виконанні перетворень змінного типу нелінійності.
2. Удосконалено алгоритм навчання нейронної мережі, який відрізняється від існуючих застосуванням методу задання адаптивних параметрів.
3. Набув подальшого розвитку алгоритм виявлення принципів роботи тестової системи на навченій нейронній мережі, який відрізняється від існуючих групуванням вхідних параметрів і активності нейронів мережі.
4. Розроблено інформаційну технологію визначення закономірностей в накопичених даних на основі методу нелінійної нормалізації, методик навчання та виявлення принципів роботи системи.

Практична цінність роботи полягає в розробці формальної методології, яка дозволяє використовувати її в багатьох організаціях. Дисертація містить конкретні практичні рекомендації з використання запропонованої в ній методології і прийомів в різних прикладних областях. Впровадження програмного забезпечення аналітичної системи, створеного в рамках дисертації, може бути використано для автоматизації аналітичної роботи.

Застосування результатів роботи. Результати наукових досліджень були використані на кафедрі інформаційних систем та технологій Навчально-наукового інституту інформаційних технологій Державного університету телекомунікацій під час виконання науково-дослідної роботи на тему «Розробка системи активного управління чергою пакетів в мережах TCP/IP з використанням REM-регуляторів» (№ 0119U101284, ДУТ, м. Київ), госпдоговірних робіт «Аналіз ринку та дослідження перспектив розвитку комп'ютерно-інформаційного обладнання» (ДУТ, м. Київ) та «Дослідження обробки траєкторної інформації в вимірювально-обчислювальних системах» (ДУТ, м. Київ), впроваджені у виробничий процес на підприємствах ТОВ «ХУАВЕЙ Україна», ТОВ «ППЛ УА», ТОВ «Ай Ті Джи» та в навчальному процесі Державного університету телекомунікацій (Київ).

Основні положення для захисту:

1. Вперше розроблено метод нелінійної нормалізації даних, який ґрунтується на послідовному виконанні перетворень змінного типу нелінійності.
2. Удосконалено алгоритм навчання нейронної мережі, який відрізняється від існуючих застосуванням методу задання адаптивних параметрів.
3. Набув подальшого розвитку алгоритм виявлення принципів роботи тестової системи на навченій нейронній мережі, який відрізняється від існуючих групуванням вхідних параметрів і активності нейронів мережі.

4. Розроблено інформаційну технологію визначення закономірностей в накопичених даних на основі методу нелінійної нормалізації, методик навчання та виявлення принципів роботи системи.

Тематика дисертаційної роботи і отримані результати безпосередньо відповідають пріоритетності розвитку інформаційних та комунікаційних технологій в Україні до 2021 р. згідно із Законом України «Про пріоритетні напрями розвитку науки і техніки», від 11.07.2001 № 2623-III, зі змінами внесеними згідно із Законом України «Про наукову та науково-технічну діяльність» від 26.11.2015 № 848-VIII. Дисертаційна робота виконана відповідно до планів наукової і науково-технічної діяльності Державного університету телекомунікацій і є частиною досліджень в рамках науково-дослідної роботи «Розробка системи активного управління чергою пакетів в мережах TCP/IP з використанням REM-регуляторів» (№ 0119U101284, ДУТ, м. Київ), а також госпдоговірних робіт «Аналіз ринку та дослідження перспектив розвитку комп'ютерно-інформаційного обладнання» (ДУТ, м. Київ) та «Дослідження обробки траєкторної інформації в вимірювально-обчислювальних системах» (ДУТ, м. Київ).

Ключові слова: інформаційні технології, штучний інтелект, нейронні мережі, кластеризація, інтелектуальний аналіз даних, виявлення знань, великі дані.

ANNOTATION

Tushych A.M. Methods for building an intelligent data analysis system based on neural networks. – Qualifying scientific work as a manuscript.

Dissertation for the degree of Doctor of Philosophy in specialty 123 – computer engineering (area of knowledge 12 – information technology). – State University of Telecommunications of the Ministry of Education and Science of Ukraine, Kyiv, 2021.

The dissertation work is devoted to the urgent scientific task of developing a methodology for constructing an intelligent data analysis system based on neural networks. The topic of the dissertation research corresponds to the temporary standard and professional competence of the educational and scientific program for the training of doctors of philosophy in computer engineering of the State University of Telecommunications of the Ministry of Education and Science of Ukraine, namely: fundamental scientific research of theoretical and methodological, scientific and methodological and applied foundations of increasing the efficiency of innovative and production activities enterprises, as well as improving the process of ensuring the introduction of the latest information technologies at the objects of information activities.

To ensure the specified high requirements for quantitative and qualitative indicators of information processing, is the process of information interaction of technological processes of modern enterprises and organizations, it is necessary to implement an intelligent data analysis system.

To achieve this goal, it is necessary to design a system that will be resilient to failures, be able to automatically recover, be indifferent to noisy data and show the results in a form acceptable to the user.

The paper provides a comparative analysis of technologies that solve the problem of analyzing the accumulated data that enterprises receive in the course of their activities. However, little attention has been paid to the development of methods and

programs for data analysis that can identify potentially useful but implicit information. Obtaining this information can provide a vital impetus for research in other areas. This nontrivial extraction of implicit, previously unknown and potentially useful information from large databases is known as data mining or knowledge discovery.

One of the requirements for intelligent knowledge discovery systems is efficiency and scalability. Working with very large databases requires efficient algorithms, and inaccuracies and often incomplete data creates additional problems when retrieving hidden patterns. Taught neural networks are able to generate hidden knowledge from data: the ability to predict, classify and recognize patterns is created, but their logical structure with the traditional approach remains hidden from the user.

The dissertation work formulates the requirements for modern data analysis systems to justify the choice of the mathematical apparatus of the core of the analytical system, a new method of nonlinear data normalization is developed, which consists in the gradual implementation of transformations on data with a variable type of nonlinearity, it is investigated which neural networks have an advantage over traditional methods of mathematical statistics and technical analysis for a specific task, the optimal topology and parameters of the selected technology have been chosen, a new approach has been developed to explicitly obtaining rules from a trained neural network, consists in grouping input parameters and neuron activity, an information technology has been developed for determining the search for the rules of functioning of the object under study in an explicit form on based on processing a number of empirical data using a neural network apparatus.

The *purpose* of this dissertation is to optimize the process of revealing hidden patterns contained in databases using improved data processing techniques.

To achieve this goal, it is necessary to solve the following *tasks*:

1. Determine the requirements for modern analytical systems, based on the presented requirements, to justify the choice of the mathematical apparatus of the core of the analytical system.

2. Develop a method for preprocessing the data being analyzed.
3. Conduct a study based on the construction of a data analysis algorithm using the selected neural network technology.
4. Select the optimal topology and parameters of the selected topology.
5. Develop an algorithm for identifying knowledge in the analyzed data.
6. Develop a software implementation of information technology for simulation of system processes.

Research methods. The results of the research carried out and presented in the dissertation were obtained using the methods of systems analysis, information theory, probability theory, methods of statistical analysis, combinatorics, neural network modeling, cluster analysis, optimization, formalization and dynamic programming methods.

Scientific novelty of the thesis:

1. For the first time, a method of nonlinear data normalization based on the sequential execution of transformations of a variable type of nonlinearity has been developed.
2. The algorithm for learning the neural network has been improved, it differs from the existing ones by the application of the method for setting adaptive parameters.
3. Entered the further development of the algorithm for identifying the principles of the test system on a trained neural network, which differs from the existing grouping of input parameters and the activity of neurons in the network.
4. An information technology has been developed for determining patterns in the accumulated data based on the nonlinear normalization method, teaching methods and identifying the principles of the system's operation.

The practical value of the work lies in the development of a formal methodology that allows it to be used in many organizations. The dissertation contains specific practical recommendations on the use of the proposed methodology and techniques in

various applied fields. The implementation of the analytical system software, created within the framework of the dissertation, can be used to automate analytical work.

Application of work results. The results of scientific research were used at the Department of Information Systems and Technologies of the Educational and Scientific Institute of Information Technologies of the State University of Telecommunications when performing research work on the topic "Development of a system for active management of packet queue in TCP / IP networks using REM regulators" (No. 0119U101284, FLS., Kyiv), contractual works "Analysis of the market and research of prospects for the development of computer information equipment" (FLS., Kyiv) and "Study of processing trajectory information in measuring and computing systems" (FLS., Kyiv) , introduced into the production process at the enterprises of Huawei Ukraine LLC, PIPL UA LLC, ITG LLC and in the educational process of the State University of Telecommunications (Kyiv).

Basic provisions for protection:

1. For the first time, a method of nonlinear data normalization based on the sequential execution of transformations of a variable type of nonlinearity has been developed.
2. The algorithm for learning the neural network has been improved, it differs from the existing ones by the application of the method for setting adaptive parameters.
3. Entered the further development of the algorithm for identifying the principles of the test system on a trained neural network, which differs from the existing grouping of input parameters and the activity of neurons in the network.
4. An information technology has been developed for determining patterns in the accumulated data based on the nonlinear normalization method, teaching methods and identifying the principles of the system's operation.

The topic of the dissertation work and the results obtained directly correspond to the priority of the development of information and communication technologies in Ukraine until 2021. According to the Law of Ukraine "On priority directions of

development of science and technology", dated July 11, 2001 No. 2623-III, as amended in accordance with the Law of Ukraine "On scientific and scientific and technical activities" dated November 26, 2015 No. 848-VIII. The dissertation was completed in accordance with the plans of scientific and scientific and technical activities of the State University of Telecommunications and is part of the research within the framework of the research work "Development of an active packet queue management system in TCP / IP networks using REM regulators" (No. 0119U101284, FLS., Kyiv), as well as contractual works "Analysis of the market and research of prospects for the development of computer-information equipment" (FLS., Kyiv) and "Study of processing trajectory information in measuring and computing systems" (FLS., Kyiv).

Keywords: information technology, artificial intelligence, neural networks, clustering, data mining, knowledge discovery, big data.

СПИСОК ПУБЛІКАЦІЙ ЗДОБУВАЧА

Наукові праці, в яких опубліковані основні наукові результати дисертації:

1. V. Savchenko, V. Akhramovych, A. Tushych, I. Sribna, and I. Vlasov, "Analysis of Social Network Parameters and the Likelihood of its Construction", *International Journal of Emerging Trends in Engineering Research*, vol. 8, no 2, pp. 271-276, 2020.
2. K. Storchak, N. Yakovenko, A. Tushych, I. Sribnaya, and O. Polonevich, "Improving the material quality of network equipment due to a mechanism of surface hardening", *Scientific discussion*, vol. 1, no. 49, pp. 34-39, 2020.
3. А.М. Тушич, К.П. Сторчак, А.П. Бондарчук, та А.О. Макаренко, "Вимоги до інтелектуальних систем аналізу даних та їх класифікацій", *Науково-технічний журнал "Телекомунікаційні та інформаційні технології"*, №1, с. 31-36, 2019.

4. К.П. Сторчак, А.М. Тушич, та А.П. Бондарчук, “Кластерний аналіз даних з використанням штучних нейронних мереж”, *Науковий журнал “Зв’язок”*, №6, с. 36-38, 2018.
5. К.П. Сторчак, А.М. Тушич, К.С. Козелкова, та М.М. Степанов, “Інтелектуальний аналіз даних з використанням нейронних мереж”, *Науковий журнал “Зв’язок”*, №4, с. 17-19, 2018.
6. К.П. Сторчак, А.М. Тушич, О.М. Ткаленко, В.М. Чорна, та Т.М. Жила, “Аналіз вимог до проектування хмарної платформи для інтернету речей”, *Науковий журнал “Зв’язок”*, №6, с. 26-34, 2019.
7. О.А. Золотухіна, О.М. Ткаленко, А.М. Тушич, В.М. Чорна, та О.Р. Нікітенко, “Концепція розвитку підсистеми передавання мультимедійних повідомлень IMS”, *Науково-технічний журнал “Телекомунікаційні та інформаційні технології”*, №4, с. 81-89, 2019.
8. І.С. Сиротенко, І.С. Щербина, К.П. Сторчак, та А.М. Тушич, “Аналіз ефективності використання нейронних мереж на прикладі багат шарового перцептронну та мережі Кохонена”, *Науковий журнал “Зв’язок”*, №5, с. 22-26, 2020.
9. Б.В. Шефкін, І.В. Красюк, В.О. Хоменчук, К.П. Сторчак, та А.М. Тушич, “Дослідження та впровадження нейронної мережі на основі TENSORFLOW”, *Науковий журнал “Зв’язок”*, №6, с. 20-25, 2020.
10. К.П. Сторчак, Д.В. Кравець, А.М. Тушич, та Д.В. Сорокін, “Аналіз методів організації прав користувачів у GNU/Linux системах”, *Науковий журнал “Зв’язок”*, №4, с. 38-40, 2020.

Наукові праці, які засвідчують апробацію матеріалів дисертації:

11. А.М. Тушич, “Використання штучних нейронних мереж для створення IoT рішення фільтрації VPN трафіку”, на *XI Міжнар. наук.-техн. конф.*

студентів та аспірантів “Перспективи розвитку інформаційно-телекомунікаційних технологій та систем”, Київ, 2019, с. 361.

12. А.М. Тушич, “Машинне навчання з використанням TENSORFLOW”, на XII Міжнар. наук.-техн. конф. студентів та аспірантів “Перспективи розвитку інформаційно-телекомунікаційних технологій та систем”, Київ, 2020, с. 379.

13. М. Пелепей, та А. Тушич, “Штучний інтелект – друг, ворог чи помічник людини”, на IX Міжнар. наук.-техн. конф. студентства та молоді “Світ інформації та телекомунікацій”, Київ, 2019, с. 303-304.

14. А.М. Тушич, “Аналіз доцільності використання автоматизованої системи інтелектуального аналізу даних на основі штучних нейронних мереж”, на VII Всеукр. наук.-практ. конф. студентів, аспірантів та молодих вчених з автоматичного управління присвячена Дню космонавтики, Херсон, 2019, с. 76-77.

15. Д.Я. Алтинніков, та А.М. Тушич, “Як штучний інтелект та ІОТ доповнюють один одного”, на Всеукр. наук.-техн. конф. “Сучасний стан та перспективи розвитку ІОТ”, Київ, 2020, с. 303-304.

16. А.М. Тушич, та П.В. Лебединець, “Дослідження системи моніторингу Zabbix для ІТ-інфраструктури підприємства”, на VII Наук.-техн. конф. “Сучасні інфокомунікаційні технології”, Київ, 2018, с. 20-21.

17. А.М. Тушич, та О.М. Скрипаль. “Безпека функціонування телекомунікаційних систем та мереж”, на XII Міжнар. наук.-техн. конф. “Проблеми інформатизації”, Київ, 2018р. – С.75-76.

18. А.М. Тушич, та П.В. Лебединець, “Дослідження відкритого програмного забезпечення поштового сервера та клієнта”, на XII Міжнар. наук.-техн. конф “Проблеми інформатизації”, Київ, 2018, с. 105.

19. Б. Свєрдлюк, Ю.Каграманова, та А. Тушич, “Інтернет речей”, на *IX Міжнар. наук.-техн. конф. студєнтства та молоді “Світ інформації та телекомунікацій”*, Київ, 2019, с. 107-108.