

УДК 004.896

МЕТОДИ ТА ЗАСОБИ ІНТЕЛЕКТУАЛЬНОЇ ОБРОБКИ ІНФОРМАЦІЇ

к.е.н., доц. О.А. Журан, К.В. Донченко

Державний університет «Одеська політехніка»

В даній роботі проводиться дослідження та аналіз особливостей технологій Data Mining. Описуються головні задачі які вирішує даний метод обробки інформації. Наводяться приклади сфер, які активно використовують дані технології. Особлива увага приділяється найбільш поширеним методам, які використовує Data Mining, вказуються їхні переваги та недоліки. Перераховуються програми інтелектуального аналізу, які є найбільш поширеними на даний час.

Ключові слова: генетичні алгоритми, дерево прийняття рішень, інтелектуальна обробка даних, нейронні мережі, статистичні методи, Data mining

З розвитком технологій запису і зберігання даних людство спіткнулося з колосальними потоками інформації в найрізноманітніших сферах своєї діяльності – виробництві, науці, медицині, державному управлінні, тощо. Інформаційний бум призвів до збільшення кількості даних, накопичених у багатьох предметних галузях у сотні та тисячі разів. І стало зрозуміло, що без їх продуктивної переробки, потоки знань швидко утворюють нікому не потрібну звалище і не здатні будуть застосовуватися у подальшому для розвитку цих галузей. Однак методів традиційної статистики виявлялося явно недостатньо для якісного аналізу великих обсягів даних. Тому головна мета даної роботи – це аналіз методів та засобів інтелектуальної обробки інформації.

Дана проблема швидкого аналізу даних з метою формування на їх основі рішень гостро постала ще в часи 2-ї світової війни. Її вивченням займалися фон Нейман, вчені школи Лебедева С.О, Глушкова В.М. Проте найбільша кількість досліджень методів та засобів інтелектуальної обробки даних припадає на дев'яності роки ХХ ст. та на наш час у зв'язку з розвитком можливостей сховищ даних.

Інтелектуальна обробка інформації – це використання цифрових технологій для вирішення завдань, що виникають на різних етапах управління даними, тобто розпізнавання, перетворення в структурований вигляд, класифікація, аналіз інформації та її пошук в різних джерелах [4, с.185]. Аналіз таких різнопланових даних згодом використовується для прогнозування ситуацій та розвитку подій, а також в системах, що допомагають здійснювати прийняття рішень.

Для аналізу потоку даних прийнято використовувати Data Mining (англ. добування знань). Data Mining – це міждисциплінарна галузь, що збрала у собі методи штучного інтелекту, комп'ютерного навчання, алгоритмів, прикладної статистики, розпізнавання образів, теорії баз даних та інших. Г. Пятецкий-Шапиро, який був одним з відомих спеціалістів в сфері Data Mining, охарактеризував метод інтелектуальної обробки інформації як аналіз штучним інтелектом, алгоритмами «необроблених даних», знаходження у них прихованих залежностей [2, с.17]. Тобто це новітня технологія інтелектуального дослідження зібраної інформації, яка допомагає виявити закономірності, які не є видимими з першого погляду, у вигляді тенденцій, знайти їх кореляцію ітд. Існує досить велика кількість визначень Data Mining, проте всі вони наголошують на чотирьох найважливіших ознаках даного методу, а саме – отримані результати:

- 1) раніше були невідомі;
- 2) є незвичайними;
- 3) корисні в практичному застосуванні;

4) здатні інтерпретуватися людиною та являються корисні для вибору рішень в різних галузях.

Головною особливістю Data Mining є об'єднання в собі великої кількості математичних інструментів, починаючи зі звичайних статистичних і закінчуючи сучасними кібернетичними, і навіть останніми досягненнями в сфері інноваційних технологій. Велика маса методів Data Mining була розроблена в рамках теорії штучного інтелекту. Зазвичай їх прийнято класифікувати на наступні:

1. Методи моделювання, прогнозування, класифікації на основі нейронних мереж, дерева рішень, еволюційного програмування, генетичних алгоритмів;
2. Статистичні методи, такі як факторний аналіз, кореляційно-регресійний аналіз, дисперсійний, лінійна регресія, аналіз часових рядів ;
3. Методи кластерного аналізу;
4. Методи обмеженого перебору.

Головні задачі, які вирішує Data Mining діляться на задачі прогнозування та описові моделі. Головна мета описових моделей – це покращення інтерпретації аналізованих даних. Такі моделі досить просто сприймаються людиною. Що ж до задач прогнозування, то для їх створення спочатку будується модель, яка потім буде здатна передбачати результати вже з новими наборами даних [1, с.71].

До базових методів Data Mining відносять засновані на алгоритмах обмеженого перебору. Однак при великому обсязі даних застосувати простий перебір неможливо. Переваги такого методу – це простота розуміння і реалізації.

Також до базових методів Data Mining відносяться елементи теорії статистики. Це, як було сказано раніше – кореляція, регресія та дисперсія та інші види статистичного аналізу даних. Однак дані методи мають масу недоліків і один з найбільших це недостатньо висока точність отриманих результатів, підбиття середніх тенденцій. Дерево прийняття рішень є також дуже широко використовуваним методом. Проте він має високу точність інтерпретування результатів тільки в чітко визначених ситуаціях.

Авторами в попередніх публікаціях розглядалися питання використання інтелектуального аналізу даних Data Mining для побудови та розрахунку конкурентних моделей, для оцінки конкурентоспроможності підприємств [5 с.43].

В наш час в технології інтелектуального аналізу часто використовуються методи нечіткої логіки, генетичні алгоритми та нейронні мережі. Дуже часто вихідні дані з тих чи інших причин містять неповну або неточну інформацію. Однак вище перелічені методи здатні наближувати свій машинний аналіз до людського. Адже людина не використовує формальне моделювання на основі математичних виразів, вона користується нечіткими природними мовами, і в процесі прийняття рішення вона розділяє ситуацію на окремі події, завдяки великій кількості різних критеріїв. Саме таким чином працюють методи нечіткої логіки, дозволяючи оперувати з безліччю індивідуальних правил замість одного чіткого узагальненого правила.

Ще один інноваційний метод аналізу даних – нейронні мережі. Вони розроблені по біологічній аналогії мозку людини[3, с.32]. Нейронні мережі, подібно людині проходять етапи навчання і на основі уже отриманих даних, використовують набуті знання для вирішення різних завдань аналізу даних. Нейронні мережі є найбільш точними і можуть видавати прогноз дуже високої якості.

На ринку інформаційних продуктів всього налічується близько сорока програм для інтелектуального аналізу даних, до найбільш відомих із них відносяться: SAS Data mining, Teradata, Board, Dundas BI, Inetsoft, Qlik, Oracle BI, KNIME, Solmin's XLminer, Sisense, ELKI, Alteryx, Enterprise Miner, Datawatch Desktop, PolyAnalyst, Civis, Viscosity. Дані програми є досить дорогими і тому на даний час використовуються тільки в великих компаніях. Найбільшого поширення вони отримали в торгівлі – часто дані компанії

аналізують купівельні корзини своєї потенційної аудиторії з метою виявлення товарів-комplementів та створення моделей, які здатні спрогнозувати оптимальну кількість запасів; в банківській справі – це розкриття шахрайських дій, які пов'язані з кредитними картками; в медицині – для створення експертних систем, які здатні ставити медичні діагнози; молекулярній генетиці та генній інженерії – для виявлення захворювань, які передаються спадково; прикладна хімія – для створення сполук з певними хімічними властивостями, наприклад – ліки, парфуми, тощо.

Отже, інтелектуальна обробка даних – це процес, що допомагає проаналізувати дані і на основі цього прийняти певні рішення. Головна її особливість – це пошук певних залежностей у великій кількості інформації з метою виявлення їх невидимих закономірностей. Ця технологія має як переваги так і недоліки, але безумовно є перспективною сферою розвитку. Проте на даний час засоби Data Mining відносяться до дуже дорогим програмних інструментів, тому основні користувачі – це великі компанії. Однак, поступова популяризація технології повинна привести до появи більш бюджетних програмних засобів, якими зможе користуватися кожен.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. N. Shakhovska, O. Veres, Y. Bolubash. Big Data information technology and data space architecture // Sensors & Transducers. – 2015. – Vol. 195, – P. 69–77
2. Ланде Д.В., Субач І.Ю., Бояринова Ю.Є. Основи теорії і практики інтелектуального аналізу даних у сфері кібербезпеки: навчальний посібник. — Київ: ІСЗІ КПІ ім. Ігоря Сікорського», 2018. — 297 с
3. Марченко О.О., Россада Т.В. Актуальні проблеми Data Mining: навчальний посібник для студентів факультету комп'ютерних наук та кібернетики. — Київ. — 2017. — 150 с.
4. Ленков С. В., Джулій В. М., Горбатюк О. М. Берназ Н. М. Концептуальна схема системи інтелектуальної обробки даних // Збірник наукових праць Військового інституту Київського національного університету імені Тараса Шевченка. - 2014. - Вип. 46. - С. 181-190.
5. Журан О.А. Використання сучасних інформаційних технологій для оцінки конкурентоспроможності підприємств // Матеріали науково-практичної інтернет-конференції «Економічна кібернетика: теорія, практика та напрямки розвитку». – Одеса, Україна. – 2019. - С.41-44

Zhuran Olena, Donchenko Kateryna
State University "Odessa Polytechnic", Ukraine

Methods and means of intellectual processing of information

The article examines and analyzes the features of Data Mining technologies. It describes the main problems that solved by this method of information processing. Examples of areas that actively use these technologies are given. Particular attention is paid to the most common methods used by Data Mining pointing out their advantages and disadvantages. The intellectual analysis programs which are the most widespread at present are listed.

Keywords: genetic algorithms, decision tree, intelligent data processing, neural networks, statistical methods, Data mining