

студентів для занесення в журнал результатів тестування, а також розширення використання елементів штучного інтелекту, зокрема для формування варіативних відкритих тестів.

Література

1. Грабко В.В., Романюк О.Н., Бісікало О.В., Боцула М.П., Паламарчук Є.А., Коваленко О.О. Система інтеграції електронних ресурсів вищого навчального закладу "Інтегровані електронні ресурси ВНТУ JetIQ «Концепція інтеграції електронних ресурсів ВНТУ» службовий твір (№70723 від 21.11.16). [Електронний ресурс] . - <https://iq.vntu.edu.ua/Назва> з екрану.
2. Паламарчук Є.А., Бісікало О.В., Коваленко О.О. КП «Навігатор навчальних ресурсів» службовий твір (№70590 від 21.11.16). [Електронний ресурс] . - <https://iq.vntu.edu.ua>. - Назва з екрану.
3. Супертест [Електронний ресурс] / Доступ: <http://iq.vntu.edu.ua/help/superinstr.pdf>. - Назва з екрану.

УДК 004.9:530.1

Наталя Шаховська, Роман Камінський, Є.О.Засоба

Національний університет «Львівська політехніка», м. Львів

МЕТОД ПОШУКУ АСОЦІАТИВНИХ ЗАЛЕЖНОСТЕЙ У ВЕЛИКИХ ДАНИХ

В роботі запропоновано метод аналізу Великих даних в умовах наявності різних джерел даних та різних методів опрацювання цих даних. Уведено поняття асоціативної залежності, розроблено метод пошуку залежностей, визначено ефективність та можливості його розпаралелення.

Ключові слова: великі дані, асоціативне правило, залежність даних, складність алгоритму, паралельне опрацювання

The paper proposes a method for analyzing large data in the presence of various data sources and various methods for processing these data. The concept of associative dependence was introduced, the method of finding dependencies was developed, efficiency and possibilities of its parallelism were determined.

Keywords: large data, associative rule, dependence of data, complexity of algorithm, parallel processing

Вступ. Сьогодні задача опрацювання різноманітних неузгоджених інформаційних ресурсів (пошуку, системної інтеграції тощо) виникає досить часто. Так, для університету прикладом інтеграції є формування наукових звітів, визначення показників успішності та якості навчання, формування рейтингу кафедри тощо.

Тому виникає необхідність управління розрізненою інформацією, а саме її подання у зрозумілому для користувачів вигляді (навіть якщо вони не знають особливостей організації структур цього джерела даних) та опрацювання (пошуку, інтеграції, видобуванні нових знань тощо).

Одним із базових завдань опрацювання різнотипних даних є їхня інтеграція. Розроблені на сьогодні методи інтеграції даних за своєю функціональністю поділяються на два типи: інтеграція веб-застосунків та інтеграція на основі сховищ даних. Проте проведений аналіз літературних джерел показав, що для опрацювання інформації від усіх об'єктів галузі необхідно поєднати обидва типи інтеграції та вдосконалити наявні моделі даних.

Сучасний стан проблеми. Складові проблеми опрацювання різнотипних даних подані на рис. 1 [7]. Окремі предметні області, для яких актуальні задачі опрацювання даних з різнотипних джерел [8]:

- науково-навчальний процес (аналіз успішності студентів, облік публікацій, участі у конференціях, олімпіадах);
- аналіз наукових джерел (формування показника якості кафедр; визначення перспективності наукового напрямку, унікальності робіт науковця, використання даних для міжгалузевих досліджень).
- системи анотування та реферування множини документів (визначення ваги терміну та речення, яке входить у кінцевий реферат).

Наведені вище приклади перетинаються з іншим поширеним поняттям – Великі дані [9].

Багато існуючих методів аналізу даних непридатні до застосування для Великих даних, оскільки:

- Розмірність даних надзвичайно велика.
- За своєю структурою та через різноманітні джерела походження дані є неочищеними, є багато відхилень.
- Необхідно застосовувати паралельну обробку даних.

Тому необхідно модифікувати метод пошуку асоціативних правил для роботи з Великими даними. Правила асоціації та генерація правил широко використовуються, і вони стикаються з багатьма проблемами, головним з яких

є наявність великих даних та багатовимірних наборів даних [4].

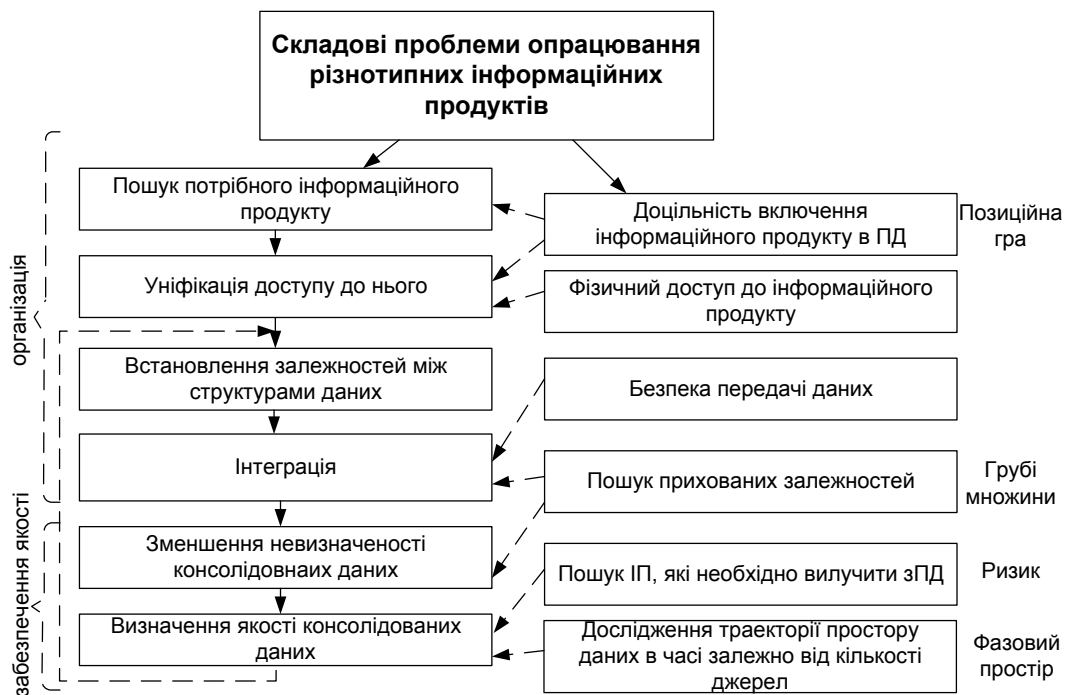


Рис. 1. Складові проблеми опрацювання різнотипних даних

Визначення асоціативної залежності. Введемо поняття асоціативної залежності. Шукатимемо залежність на відношенні r . Це відношення може бути сформоване як для реляційних джерел даних, так і для нереляційних (NoSQL) шляхом формування пари значень - назви об'єкта та його характеристик. Для різних об'єктів кількість характеристик може бути різною. У такому випадку відношення r формуватиметься як $CROSS(r)$. Далі по тексту вживатимемо тільки позначення r .

Асоціативна залежність (A3) – це продукційне правило в селекції відношення r , яке справджується для значущої кількості об'єктів цієї селекції. Поріг значущості повинен визначатись експертним шляхом, або виходячи з розрахунків імовірності помилкового виділення цієї залежності.

Основними параметрами традиційних асоціативних правил є рівень довіри та рівень підтримки.

Рівень довіри – відношення кількості об'єктів, для яких має місце така A3 до кількості об'єктів в селекції:

Рівень підтримки – характеристика предиката селекції на відношенні, що обчислюється як відношення кількості об'єктів, які задовольняють предикат P до загальної кількості об'єктів у відношенні.

Рівень покращення обчислюється, як відношення рівнів довіри та підтримки A3.

Метод пошуку асоціативної залежності. Побудуємо метод пошуку АЗ [3]. Метод складається з трьох етапів:

- I етап. Аналіз вхідних даних.
- II етап. Агрегування залежностей.
- III етап. Побудова АЗ.

Вхідними даними є:

1. Відношення $CROSS(r)$, схема R визначена лише на аналізованій селекції.
2. Хеш-функції для кожного атрибуту відношення R : $h_j(A_j)$.
3. Порогове значення рівня довіри залежностей, що шукаються – p_0 . Замість даного параметра рівноцінно може використовуватись кількість кортежів, на яких повинна бути визначена шукана залежність – $minSupport$.
4. Порогове значення рівня довіри залежностей, що враховуються при утворенні нових залежностей = p^* .

Вихідними даними є множина асоціативних залежностей, що відповідає вказаним критеріям: $\{S_i \rightarrow T_i\}, \forall i: Conf(S_i \rightarrow T_i) \geq p_0$.

Результат аналізу розробленим методом подано на рис. 2.

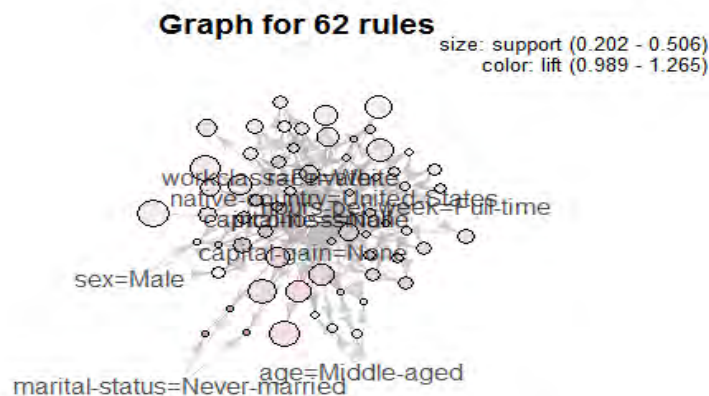


Рис. 2. Візуалізація знайдених асоціативних залежностей

Порівняльний аналіз алгоритмів виявлення асоціативних залежностей у даних за кількісними критеріями. Для порівняння ефективності розробленого методу аналізу даних було обрано три найближчих за призначенням методи: Apriori, HybridApriori та метод FP-tree.

За кількісні критерії було обрано кількість корисних залежностей, відсоток корисних залежностей у множині знайдених залежностей що виявив метод, а також час аналізу даних.

Отримана оцінка часу виконання алгоритму є субполіноміальною, а отже, розроблений алгоритм є ефективним паралельним алгоритмом.

Розроблений алгоритм дає змогу стверджувати, що задача виявлення асоціативних залежностей у розподілених базах даних належить до класу P-

задач. Отже, алгоритм пошуку асоціативних залежностей добре вирішувати з допомогою MapReduce.

Крім кількох послідовних реалізацій, паралельні реалії для роботи з великими даними не є широко доступними. Одним із прикладів серійної реалізації добре відомий статистичний обчислення з пакетом R, що називається "arules". Паралельна реалізація програми FP-Growth доступна в бібліотеці для вивчення комп'ютера з відкритим вихідним кодом (MLlib) Apache Spark та Apache Mahout.

Висновки. Асоціативні правила асоціації широко використовуються у великій кількості додатків. Наш алгоритм може бути використаний в деяких з цих програм, в яких використання традиційних алгоритмів не є життєздатним через величезну кількість даних, які потрібно обробляти. Це може бути дуже корисним, наприклад, у сенсорних мережах, які генерують величезну кількість даних за короткі проміжки часу, або в соціальних мережах, у яких є мільйони користувачів. Нарешті, слід зазначити, що ці процедури можна узагальнити на інші методи видобування даних, які використовують правила асоціації, такі як винятки та аномалії [4], поступові залежності [5, 6] та ін.

Література

1. Zaki, M. J. (2000). *Scalable algorithms for association mining*. *IEEE Transactions on Knowledge and Data Engineering* 12(3): 372–390с.
2. J. Han, H. Pei, and Y. Yin. *Mining Frequent Patterns without Candidate Generation*. In: *Proc. Conf. on the Management of Data (SIGMOD'00, Dallas, TX)*. ACM Press, New York, NY, USA 2000..
3. Пшеничний О. Ю. *Математичне та програмне забезпечення виявлення кон'юнктивних асоціативних залежностей у великих масивах даних : автореферат дисертації на здобуття наукового ступеня кандидата технічних наук : 01.05.03 – математичне та програмне забезпечення обчислювальних машин і систем / Олександр Юрійович Пшеничний ; Національний університет «Львівська політехніка» . - Львів, 2012. - 24 с..*
4. Delgado, M., Ruiz, M.D. & Sánchez, D., *New approaches for discovering exception and anomalous rules*. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 19(02), pp. 361–399, 2011.
5. Berzal, Fernando, et al., *A new framework to assess association rules*. In *Advances in Intelligent Data Analysis*, Springer Berlin: Heidelberg, pp. 95–104, 2001.
6. Hüllermeier, E., *Association rules for expressing gradual dependencies*. In *Principles of Data Mining and Knowledge Discovery*, Springer Berlin:

- Heidelberg, pp. 200–211, 2002.
7. Шаховська Н.Б. Програмне та алгоритмічне забезпечення сховищ та просторів даних: монографія / Міністерство освіти і науки України, Національний університет «Львівська політехніка». – Львів, 2010. – 194 с.
 8. Шаховська Н.Б. Структура та задачі простору даних // Складні системи і процеси. — 2008. — № 1. — С. 73—86.
 9. Big Data Dimensions, <http://www.klarity-analytics.com/392-dimensions-of-big-data.html>
 10. Agrawal, Rakesh, Imielinski, Tomasz & Swami, Arun, Mining association rules between sets of items in large databases. ACM SIGMOD Record, 22, pp. 207–216, 1993.

УДК 004.9

**Наталія Піндус, Степан Чеховський, Галина Сенів, Старко І.Ю.,
Юлія Ковальчук**

Івано-Франківський національний технічний університет нафти і газу

ПРАКТИЧНІ АСПЕКТИ РОЗРОБЛЕННЯ ЕЛЕКТРОННИХ КУРСІВ В ІФНТУНГ

© Піндус Н.М., Чеховський С.А., Сенів Г.О., Старко І.Ю., Ковальчук Ю.В. 2017

У статті висвітлено організаційно-методичні аспекти розроблення, впровадження і стандартизації електронних курсів в ІФНТУНГ шляхом використання навчальної платформи Moodle.

Ключові слова – електронний курс, методичне забезпечення, дистанційне навчання, інформаційно-комунікаційні технології.

This article deals with the organizational and methodical aspects of the development, implementation and standardization of electronic courses in IFNTUOG through the use of the Moodle training platform.

Keywords - electronic course, methodological support, distance learning, information and communication technologies.

Вступ. Актуальною проблемою сьогодення освіти, зокрема дистанційної форми навчання, в Україні є надання освітніх послуг шляхом застосування сучасних інформаційно-комунікаційних технологій. З цією метою в ІФНТУНГ започатковано очно-дистанційні курси підвищення кваліфікації науково-педагогічних працівників [1,2].