

## ЗАСТОСУВАННЯ МЕТОДІВ DATA MINING В ТЕЛЕКОМУНІКАЦІЙНИХ СИСТЕМАХ

**Мороз А.М., Глоба Л.С.**

*Інститут телекомунікаційних систем КПІ ім. Ігоря Сікорського, Україна*

*E-mail: a.moroz2708@gmail.com*

### **Application of Data Mining Methods in Telecommunication Systems**

Currently, there is an intensive development of telecommunication systems, accompanied by tough competition for the client. Data analysis aimed at identifying hidden patterns is implemented using the methods of modern information technology Data Mining.

Раніше в сфері обробки знань застосовувалась прикладна статистика. Вона оцінювала, перевіряла гіпотези, але давала грубі і усереднені результати. Завдяки технічному прогресу люди почали зберігати величезні обсяги інформації, яка була різнорідна, і яку треба було обробляти. Прикладна статистика не справлялась зі своїми задачами в повній мірі. Стало зрозуміло, що необхідна нова технологія обробки даних, яка зможе забезпечити глибинний аналіз і дати точніші результати. Оператор телекомунікацій більше не може вирішувати завдання, застосовуючи статистику. Необхідно структурувати великі обсяги даних і витягувати з них корисну інформацію для того, щоб:

- успішно проаналізувати ці дані в майбутньому
- спроектувати бази нечітких знань за допомогою нечіткої логіки
- візуалізувати ці дані для легкого сприйняття людиною

Метою даної роботи є виявлення особливостей існуючих математичних методів Data Mining, що використовуються операторами телекомунікацій для вирішення проблем, що виникають внаслідок нових інформаційних технологій, таких як IoT, M2M та інші. Більш докладно розглянемо метод обробки великих даних - Data Mining.

На сьогоднішній день рішення Data Mining широко застосовуються операторами телекомунікацій, операторами мережі та Інтернет-провайдерами, з метою поліпшення обробки неструктурованих великих обсягів даних для прийняття більш вдалих рішень у своєму бізнесі. Аналітика дозволяє провайдерам телекомунікацій значно підвищити економічну ефективність у сфері надання послуг.

DataMining - це інтелектуальний аналіз даних. Це технологія, яка допомагає виявляти приховані зв'язки в базах даних дуже великих розмірів [1].

Метою DataMining є знаходження таких моделей, які не можуть бути знайдені звичайними методами. Існує два види моделей: передбачувані і описові.

Передбачувані моделі: позиціонуються на наборі даних з відомими результатами. Їх використовують для передбачення результатів на підставі інших наборів даних. Це моделі класифікації (описують правила, за якими

можна віднести опис об'єкта до одного з класів) і моделі послідовностей (вони описують функції, за якими можна прогнозувати зміну безперервних числових параметрів).

Описові моделі: вони приділяють особливу увагу суті залежностей в наборі даних, взаємного впливу різних чинників, побудови емпіричних моделей. Є легкими для сприйняття людиною [5].

Типи закономірностей, які може знайти технологія DataMining:

- Асоціація - застосовується, коли кілька подій пов'язані між собою. Наприклад, трапляється, що абонент перестає користуватися телекомунікаційними послугами оператора, оскільки вся його родина користується послугами іншої телекомунікаційної компанії. Таким чином, абонент починає користуватися телекомунікаційними послугами компанії, які використовує його сім'я.
- Класифікація - виявлення рис, які будуть характеризувати групу, до якої належить об'єкт, на основі навчання на вже класифікованих об'єктах.
- Кластеризація - відрізняється від класифікації тим, що групи заздалегідь не відомі і засоби DataMining самостійно виявляють різні однорідні групи даних.
- Послідовність - застосовується при існуванні ланцюжка подій, пов'язаних у часі.
- Прогнозування - створення або знаходження шаблонів, які будуть істинно показувати тенденцію поведінки необхідних показників по часових рядах. За допомогою них можна передбачити поведінку системи в майбутньому [2].

Сфери застосування технології інтелектуального аналізу даних обмежується виключно наявністю великої кількості інформації. Тому керівники і аналітики, які користуються даною технологією, можуть отримувати відчутні переваги в конкурентній боротьбі.

В телекомунікації дуже важливо виявляти категорії клієнтів зі схожими стереотипами використання послуг, як наслідок – йде розробка цінової політики; проведення аналізу відмов; передбачення пікових навантажень.

В силу того, що DataMining було розвинуто на стику багатьох областей, то можна привести класи і методи, які використовують цю технологію: нейронні мережі, дерева рішень, генетичні алгоритми, алгоритми обмеженого перебору, асоціативні правила, кластерний аналіз та інші [4].

Більш того, методи інтелектуального аналізу даних дозволяють ефективно вирішувати завдання структурного інженерного проектування інноваційних технічних систем в телекомунікаціях [3]. Дані методи мають багато спільного з методами вирішення проблем класифікації, діагностики та розпізнавання образів. Тим не менш, однією з їхніх головних відмінних ознак є функція інтерпретації закономірностей, що формують основні правила для включення об'єктів у класи еквівалентності. Тому логічні методи стають все більш поширеними сьогодні.

Розробка інтелектуальних методів обробки даних для телекомунікаційних компаній необхідна для:

- Зменшення обчислювальної складності методів обробки великих даних для надання послуг оператором абоненту з заданою якістю обслуговування;
- Передбачення ризиків, які можуть виникнути внаслідок роботи телекомунікаційної системи;
- Щоб мати можливість ідентифікувати несправності в системі та з'ясувати причину їх виникнення.

Деякі з існуючих алгоритмів можуть бути адаптовані для обчислення великих розподілених інформаційних масивів. У той же час серйозні труднощі можуть виникнути при візуальному поданні результатів - через величезну кількість інформації, що надходить на вхід, кількість різних звітів на виході різко зростає. Для їх зручного подання необхідні нові математичні методи, які принципово відрізняються від генераторів звітів, що використовуються для традиційного зберігання.

*Висновок.* У цьому огляді проаналізовано методи обробки великих даних, що стосуються оператора зв'язку. Існує багато проблем для вибору методів обробки даних:

- Проблема великих даних - потребують математичних методів для структурування даних і знань (у подальшій роботі пропонується використовувати онтологію).
- Проблема прогнозування - пропонується використовувати моделі бази знань і нечітку логіку.

Всі перераховані вище проблеми можна усунути шляхом розробки нових методів обробки великих даних, які: успішно аналізують ці дані, розробляють нечітку базу знань і візуалізують ці дані для легкого сприйняття людини.

Подальша робота полягає у вивченні методів структурування даних та проектування нечіткої бази знань для оператора зв'язку.

### Література

1. Dr. Mehmet Ulema. (Professor of Computer Information Systems Manhattan College, Riverdale New York, USA) "Big Data and Telecommunications" [Electronic resource], 4th International Black Sea Conference on Communications and Networking, 2016.
2. Technologies of data analysis: DataMining, VisualMining, TextMining, OLAP training. allowance. 2 nd ed. / A. A. Barsegyan, M. S. Kupriyanov, V. V. Stepanenko, I. I. Kholod. SPb .: 2007. 59 p.
3. Educational-methodical complex on the discipline "Information Technologies" of the Institute of Economics and Management [Electronic resource]. URL: <http://www.sergeeva-i.narod.ru/inform/page11.htm>.
4. Chubukova I.A. Course DataMining [Electronic resource]. URL: <http://www.intuit.ru/department/database/datamining/5>. *Ильченко М.Е.* Электродинамика диэлектрических резонаторов / М.Е. Ильченко, А.А. Трубин. - К.: Наук. думка, 2004. — 264 с.
5. On the solution of the problem of the optimal choice of feeding parameters of a linear induction machine by a genetic algorithm with local search / S.S. Bezhitsky, E.A. Golovenko, V.A. Goremykin, M.V. Pervukhin // Bulletin of Siberian State University of Automation. 2010. Issue. 4. С. 23-27.