

## АНАЛІЗ ОСНОВНИХ МЕТОДІВ ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ ДАНИХ ДЛЯ ЗАДАЧІ ПІДБОРУ ПОКУПОК

Вінницький національний технічний університет;

### **Анотація**

*Розкрито поняття «інтелектуального аналізу даних». Розглянуто алгоритм дерева прийняття рішень, алгоритм Байєса, експертні системи. Досліджено переваги та недоліки алгоритмів. Обрано найдоцільніший алгоритм для задачі інтелектуального підбору покупок.*

**Ключові слова:** інтелектуальний аналіз даних, статистика, дані, дерево прийняття рішень, спрощений алгоритм Байєса, експертні системи.

### **Abstract**

*Disclosed the concept of data mining. The decision tree algorithm, Bayes algorithm, expert systems are considered. The advantages and disadvantages of algorithms are explored. The most suitable algorithm for the smart purchasing task is chosen.*

**Keywords:** data mining, statistics, data, decision tree, simplified Bayesian algorithm, expert systems.

### **Вступ**

В наш час все частіше застосовуються інтелектуальні помічники для підбору різних видів покупок. Подібні системи дозволяють користувачам здійснити всі необхідні покупки для їхнього девайсу чи чогось іншого, а виробнику чи посереднику такі заходи дають додатковий прибуток, що складає помітний відсоток від загального.

Однією з таких систем може бути «розумний» холодильник, який може перевіряти наявність певних продуктів. Завдяки впровадженню системи інтелектуального підбору покупок, «розумний» холодильник може скласти список покупок, та замовляти їх через мережу Інтернет. Завдяки цій системі власники «розумного» холодильника завжди будуть мати перелік продуктів, які звикли споживати.

### **Результати дослідження**

Основними методами для інтелектуального аналізу даних для задачі підбору покупок є

- експертні системи;
- дерево прийняття рішень;
- спрощений алгоритм Баєса.

Експертні системи (ЕС) використовуються для вирішення так званих неформалізованих задач, загальним для яких є те, що:

- задачі не можуть бути задані в числовій формі;
- мету не можна виразити в термінах точно визначеної цільової функції;
- не існує алгоритмічного рішення задачі;
- якщо алгоритмічне рішення є, то його не можна використовувати через обмеженість ресурсів (час, пам'ять) [1].

Експертна система - це програмний засіб, що використовує експертні знання для забезпечення високоефективного рішення неформалізованих задач у вузькій предметній області. Основу ЕС складає база знань (БЗ) про предметну область, яка накопичується в процесі побудови та експлуатації ЕС. Накопичення і організація знань - найважливіша властивість усіх ЕС.

Недоліки експертних систем:

- здоровий глузд. На додаток до широкого технічного знання, людина-експерт має здоровий глузд. Ще не відомо, як закласти здоровий глузд в експертні системи;
- творчий потенціал. Людина-експерт може реагувати творчо на незвичайні ситуації, експертні системи не можуть;
- навчання. Людина-експерт автоматично адаптується до зміни середовища; експертні системи потрібно явно модифікувати;
- сенсорний досвід. Людина-експерт має широким діапазоном сенсорного досвіду; експертні системи в даний час засновані на введених символах;
- експертні системи не хороші, якщо рішення не існує або коли проблема лежить поза області їх компетенції.

Дерево прийняття рішень — використовується в галузі статистики та аналізу даних для прогнозованих моделей. Структура дерева містить такі елементи: «листя» і «гілки». На ребрах («гілках») дерева прийняття рішення записані атрибути, від яких залежить цільова функція, в «листі» записані значення цільової функції, а в інших вузлах — атрибути, за якими розрізняються випадки[1]. Щоб класифікувати новий випадок, треба спуститися по дереву до листа і видати відповідне значення. Подібні дерева рішень широко використовуються в інтелектуальному аналізі даних для задачі підбору покупок. Мета полягає в тому, щоб створити модель, яка прогнозує значення цільової змінної на основі декількох змінних на вході. Кожен лист являє собою значення цільової змінної, зміненої в ході руху від кореня по листа. Кожен внутрішній вузол відповідає одній з вхідних змінних. Дерево може бути також «вивчено» поділом вихідних наборів змінних на підмножини, що засновані на тестуванні значень атрибутів. Це процес, який повторюється на кожній з отриманих підмножин. Рекурсія завершується тоді, коли підмножина в вузлі має ті ж значення цільової змінної, таким чином, воно не додає цінності для прогнозування. Процес, що йде «згори донизу», індукція дерев рішень (TDIDT), є прикладом поглинаючого «жадібного» алгоритму, і на сьогодні є найбільш поширеною стратегією дерев рішень для даних, але це не єдина можлива стратегія. В інтелектуальному аналізі даних, для розв'язання задачі підбору покупок, дерева рішень можуть бути використані як математичні та обчислювальні методи, щоб допомогти описати, класифікувати і узагальнити набір даних, які можуть бути записані таким чином:

$$(x, Y) = (x_1, x_2, x_3 \dots x_k, Y)$$

Залежна змінна  $Y$  є цільовою змінною, яку необхідно проаналізувати, класифікувати й узагальнити. Вектор  $x$  складається з вхідних змінних  $x_1, x_2, x_3$  тощо, які використовуються для виконання цього завдання[2].

Спрощений алгоритм Байеса є алгоритмом класифікації, на підставі теореми Байеса і використовується в прогнозуючому моделюванні. Слово «спрощений» в його назві вказує на те, що алгоритм використовує методи Байеса, але не враховує можливі залежності. Даний алгоритм вимагає меншої кількості обчислень, ніж інші алгоритми, і може бути використаним для швидкого формування моделей інтелектуального аналізу даних для виявлення відносин між вхідними і прогнозованими стовпцями, також цей алгоритм підходить для розв'язання задачі підбору покупок.

Абстрактно імовірнісна модель для класифікатора — це умовна модель

$$p(C | F_1, \dots, F_n)$$

над залежною змінною класу  $C$  з малою кількістю результатів або класів, залежна від кількох змінних  $F_1, \dots, F_n$ [3]. Проблема полягає в тому, що коли кількість властивостей  $n$ , тобто кількість можливих комбінацій покупок, дуже велика або коли властивість може приймати велику кількість значень, тоді будувати таку модель на імовірнісних таблицях стає неможливо. Тому ми переформулюємо модель, щоб зробити її такою, яка легко піддається обробці.

Використовуючи теорему Баєса, запишемо

$$p(C | F_1, \dots, F_n) = p(C)p(F_1, \dots, F_n | C) / p(F_1, \dots, F_n)$$

На практиці цікавий лише чисельник цього дробу, так як знаменник не залежить від  $C$  і значення властивостей дані, так що знаменник — константа. Цей алгоритм можна використовувати для початкового дослідження даних, а потім застосувати результати для створення додаткових моделей інтелектуального аналізу з іншими алгоритмами, які вимагають більшої кількості обчислень і є більш точними, що в кінцевому рахунку приведе до розв'язання задачі підбору покупок

## Висновки

У ході оцінювання методів інтелектуального аналізу даних для задачі підбору покупок було

розглянуто такі методи, як дерево прийняття рішень, спрощений алгоритм Баєса, експертні системи. Оскільки покупки зазвичай пов'язані між собою, то для задачі підбору покупок доцільно використовувати алгоритм дерева прийняття рішень в комбінації з спрощеним алгоритмом Баєса, оскільки це забезпечить можливість перегляду значення цільової функції з додаванням нових параметрів, та зміною уже існуючих параметрів.

#### СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Методы экспертных оценок. - Методические указания к лабораторной работе по курсу "Системы поддержки принятий решений". - МГИЭМ. Сост.: А.В.Вишнеков, И.Е.Сафонова, Н.С.Курилова, В.И.Бадулин. - М., 2001. - 24 с.

2. Дерево принятия решений - <http://stud.com.ua/31896/menedzhment>.

3. Левитин А. Алгоритмы. Введение в разработку и анализ. Вильямс, 2006. Р. 160.

*Замковий Олександр Дмитрович* — студент групи ІКН-146, факультет інформаційних технологій та комп'ютерної інженерії, Вінницький національний технічний університет, м. Вінниця.

*Петришин Сергій Іванович*, – асистент кафедри комп'ютерних наук ВНТУ, Вінницький національний технічний університет, м. Вінниця.

*Zamkovti Oleksandr D.* — Faculty of Information Technologies and Computer Engineering, Vinnytsia National Technical University, Vinnytsia.

*Sergiy I. Petrishyn* — assistant of the Computer Sciences Chair, Vinnytsia National Technical University, Vinnytsia