УДК:519.7

НЕЧІТКЕ ДЕРЕВО РОЗВ'ЯЗКІВ ЯК МЕТОД АВТОМАТИЧНОГО АНАЛІЗУ ДАНИХ

Студ. Д.В. Сокур, гр. БМО1-17
Науковий керівник доц. О.Л. Блохін
Київський національний університет технологій і дизайну

Мета і завдання. Розглянути значення дерева рішень, розібрати різницю між чітким та нечітким деревом рішень, проаналізувати алгоритм побудови нечіткого дерева рішень, навчитись користуватись ним в автоматичному аналізі даних.

Об'єкт та предмет дослідження. Дерево рішень - широко відомий і популярний метод автоматичного аналізу даних, в основі якого лежить навчання на прикладах. Правила представлені у вигляді ієрархічної послідовної структури, де кожен об'єкт належить конкретному вузлу.

Методи та засоби дослідження. Використовуються загальні методи матеріалістичного дослідження: абстрактний, математичний, діалектико-матеріалістичний та інші.

Наукова новизна та практичне значення отриманих результатів. Важливість дослідження нечіткого дерева розв'язків як методу автоматичного аналізу даних зумовлена тим, що це є один з методів автоматичного аналізу величезних масивів даних. Незважаючи на досить велику кількість наукових досліджень і статистик з цієї проблематики, на сьогодні все триває розвиток автоматизації аналізу даних, що значно полегшує роботу людей, організацій, підприємств тощо.

Результати дослідження. Спочатку розглянемо значення: дерево рішень - це графічне зображення послідовності рішень і станів середовища з указівкою відповідних ймовірностей і виграшів для будь-яких комбінацій альтернатив і станів середовища; також широко відомий і популярний метод автоматичного аналізу даних, в основі якого лежить навчання на прикладах.

Відмінною рисою дерев рішень є те, що кожен приклад безумовно належить конкретному вузлу. У нечіткому випадку це не так. Для кожного атрибута необхідно виділити кілька його лінгвістичних значень і визначити ступеня приналежності прикладів до них. Замість кількості прикладів конкретного вузла нечітке дерево рішень групує їх ступінь приналежності прикладів до них.

Для прийняття рішення за допомогою дерева рішень необхідно виконати такі кроки:

- оцінити стан ринку вектором чинників X = (x1, x2, x3, x4) (цей крок виконується користувачем системи);
- визначити клас зростання прибули шляхом руху вектора X = (x1, x2, x3, x4) по дереву рішень з верхніх рівнів до нижніх (цей крок виконується системою).

У методиці використовується ієрархічна структурна схема. Для її побудови прийняті відповідні позначення елементів (подій) і логічних операцій.

В основу методу "дерева цілей" покладено підпорядкованість і ранжування цілей. Дерево цілей з кількісними показниками, що використовуються в якості одного із засобів при прийнятті рішень, і носить назву "дерева рішень".

Щоб достатньо ознайомитись з нечітким деревом розв'язків як методу автоматичного аналізу даних слід ознайомитись алгоритм його побудови. Замість кількості прикладів конкретного вузла нечітке дерево рішень групує їх ступінь приналежності.

Коефіцієнт - це співвідношення прикладів $D_j \in S^N$ вузла N для цільового значення ***i***, який обчислюється як:

$$P_i^N = \sum_{S^N} \min\left(\mu_N(D_j), \mu_i(D_j)\right),$$

де $\mu_N(D_i)$ - степінь приналежності прикладу D_i до вузла, $\mu_i(D_j)$ - степінь приналежності прикладу щодо цільового значення і S^N - множина всіх прикладів вузла N. Потім знаходимо коефіцієнт

$$E(S^N) = -\sum_i \frac{P_i^N}{P^N} \cdot \log_2 \frac{P_i^N}{P^N}, \quad (2)$$

, що позначає загальні характеристики прикладів вузла N. Ця формула дає оцінку середньої кількості інформації для визначення класу об'єкта P^N з множини

Наступним кроком побудови нечіткого дерева рішень алгоритм обчислює ентропію для розбиття по атрибуту A з значеннями a_j.

$$E(S^N, A) = \sum_j \frac{P^{N|j}}{P^N} \cdot E(S^{N|j}),$$

Належність до цільового класу для нового запису знаходиться за формулою:

$$\sigma_j = \frac{\sum_l \sum_k P_k^l \cdot \mu_l(D_j) \cdot \chi_k}{\sum_l (\mu_l(D_j) \cdot \sum_k P_k^l)},$$

Серед інших методів Data Mining, метод дерева прийняття рішень має кілька переваг:

- простий в розумінні та інтерпретації. Люди здатні інтерпретувати результати моделі дерева прийняття рішень після короткого пояснення
- не потребує підготовки даних. Інші техніки вимагають нормалізації даних, додавання фіктивних змінних, а також видалення пропущених даних.
- Дозволяє оцінити модель за допомогою статистичних тестів. Це дає можливість оцінити надійність моделі.
- Є надійним методом. Метод добре працює навіть в тому випадку, якщо були порушені початкові припущення, включені в модель.
- Дозволяє працювати з великим об'ємом інформації без спеціальних підготовчих процедур. Даний метод не вимагає спеціального обладнання для роботи з великими базами даних.

Висновки. Головна перевага "дерева рішень" перед іншими методами - можливість пов'язати ставлення цілі з діями, що підлягають реалізації в сьогоденні. При побудові багаторівневого "дерева рішень" досягнення мети кожного з рівнів моделі забезпечується комплексом заходів попереднього рівня. Кожен рівень "дерева рішень" повинен займати певне місце в ієрархічній послідовності, складеної на основі дотримання причинно-наслідкових зв'язків.

Ключові слова: *дерево рішень, нечітке дерево рішень, автоматичний аналіз даних, нечітке дерево рішень як метод автоматичного аналізу даних.*

ЛІТЕРАТУРА:

1. Паклин Н.Б., Орешков В.И. Бизнес-аналитика: от данных к знаниям . СПб: Питер, 2013.
2. Левитин А. В. Алгоритмы. Введение в разработку и анализ — М.: Вильямс, 2006. — 576 с.
3. Wang J. Encyclopedia of data warehousing and mining. 2006.