

ОСОБЛИВОСТІ ЗАСТОСУВАННЯ ІНСТРУМЕНТАЛЬНИХ ЗАСОБІВ ПІД ЧАС ІНТЕЛЕКТУАЛЬНОЇ ОБРОБКИ ІНФОРМАЦІЇ

к.т.н., доцент Творошенко І.С.

Харківський національний університет радіоелектроніки
(61166, Харків, просп. Науки, 14, каф. Інформатики, тел. (057) 702-13-35)
e-mail: iryna.tvoroshenko@nure.ua

The report reveals the peculiarities of the use of instrumental means during the intellectual processing of information. It is established that for extraction of significant information from data warehouses there are special methods of OLAP-analysis, Data Mining or Knowledge Discovery, based on the application of methods of mathematical statistics, neural networks, and inductive methods of constructing decision trees.

Інтелектуальна обробка інформації (Data Mining, інтелектуальний аналіз даних) – це мультидисциплінарна область, що виникла та розвивається на базі таких наук, як: розпізнавання образів, штучний інтелект, прикладна статистика, теорія баз даних та інших.

Крім того, інтелектуальна обробка інформації – це виявлення в сирих даних раніше невідомих, нетривіальних, практично корисних та доступних інтерпретацій знань, необхідних для прийняття рішень в різних сферах людської діяльності.

Специфіка сучасних вимог до такого виявлення така:

- дані мають необмежений обсяг;
- дані є різномірними (кількісними, якісними, текстовими);
- результати повинні бути конкретні та зрозумілі;
- інструменти обробки даних повинні бути прості у використанні.

Технологія Data Mining призначена для пошуку у великих обсягах даних неочевидних, об'єктивних та практично корисних закономірностей:

- неочевидних (закономірності не виявляються відомими методами обробки інформації або експертним шляхом);
- об'єктивних (закономірності повністю відповідають дійсності, на відміну від експертної думки, яка завжди є суб'єктивною);
- практично корисних (висновки мають конкретне значення, якому можна знайти практичне застосування).

Таким чином, Data Mining – автоматичний пошук прихованих закономірностей між змінними у великих масивах необроблених даних, який передбачає задачі: класифікації, моделювання та прогнозування.

Data Mining здійснює багаторазове виконання над даними:

- відбору ознак;
- стратифікації;
- кластеризації;
- візуалізації;
- регресії.

Під час аналізу інструментальних засобів інтелектуальної обробки інформації встановлено, що Data Mining може видавати великий відсоток

помилкових або недостовірних результатів. Щоб цього уникнути, необхідна перевірка адекватності отриманих моделей на тестових даних.

Виявлено, що засоби інтелектуальної обробки інформації теоретично не вимагають наявності певної кількості ретроспективних даних. Дана особливість може бути причиною появи недостовірних, хибних моделей і, як результат, прийняття на їх основі невірних рішень. Отже, необхідно здійснювати контроль статистичної значущості виявлених знань.

Слід зазначити, що існуючі технології вилучення знань зі сховищ даних засновані на методах статистичного аналізу і моделювання, а також орієнтовані на пошук моделей та відношень, прихованих у множині даних.

Для вилучення значимої інформації зі сховищ даних є спеціальні методи OLAP-аналізу, Data Mining або Knowledge Discovery, що засновані на застосуванні методів математичної статистики, нейронних мереж, індуктивних методів побудови дерев рішень.

Засоби Data Mining відрізняються від OLAP тим, що крім перевірки передбачуваних залежностей вони здатні самостійно генерувати гіпотези про закономірності, що існують в даних, будувати моделі, що дозволяють кількісно оцінити ступінь взаємовпливу досліджуваних факторів.

Потреба в адаптивних інформаційних системах виникає тоді, коли підтримувані ними проблемні області постійно розвиваються, тому дані системи повинні задовольняти ряд специфічних вимог, а саме:

- адекватно надавати знання проблемної області в кожен момент часу;
- бути придатними для легкої реконструкції під час зміни середовища.

Адаптивні властивості інформаційних систем забезпечуються за рахунок інтелектуалізації їх архітектури [1, 2].

Розробка даних систем передбачає оригінальне проектування (засноване на використанні систем автоматизованого проектування чи CASE-технологій) або типове проектування (засноване на програмних засобах компонентного проектування інформаційних систем).

Таким чином, головна відмінність зазначених підходів полягає у тому, що під час використання CASE-технологій на основі сховища даних при зміні проблемної області кожного разу виконується генерація програмного забезпечення, а при використанні складальної технології – конфігурація програм і тільки в рідкісних випадках – їх переробка.

Список використаних джерел:

1. Творошенко И.С. Структура и функции интеллектуальных средств принятия решений в сложных системах / И.С. Творошенко // Искусственный интеллект. – 2004. – № 4. – С. 462-470.

2. Творошенко И.С. Анализ процессов принятия решений в интеллектуальных системах / И.С. Творошенко // Системы обработки информации. – 2010. – Вып. 2 (83) – С. 248-253.