

УДК 658.012.1

А. А. Гаврилова

Харківський національний економічний університет, Харків

ВИКОРИСТАННЯ АЛГОРИТМІВ КЛАСТЕРИЗАЦІЇ ДЛЯ ПЕРВИННОГО АНАЛІЗУ СТАНУ КОРПОРАТИВНОГО УПРАВЛІННЯ

Для дослідження великої сукупності даних, що дозволяють скласти висновок про рівень корпоративного управління у відповідних структурах вітчизняної економіки, велике значення приділяється застосуванню нових інформаційних технологій в галузі математичної статистики та аналізу. Кластеризація, як перший етап аналізу даних, здійснюється на підставі певних методів, які мають відображення в алгоритмах розбиття груп об'єктів. Виділення груп дозволяє спростити роботу з даними, після кластеризації застосовуються інші методи, для кожної групи будується окрема модель [1].

Тому метою даної статті є обґрунтування вибору такого алгоритму класифікації, застосування якого ефективно б реалізовувало процедуру первинного аналізу стану корпоративного управління в державному корпоративному секторі.

На сьогоднішній момент кількість методів розбиття груп об'єктів на кластери досить великий – кілька десятків алгоритмів і ще більше їх модифікацій. За ознакою розбиття на кластери розрізняють алгоритми трьох типів: ієрархічні, неієрархічні і нейромережеві [2]. Відображення реалізації деяких з них можна знайти в таких статистичних та аналітичних пакетах, як STATISTIKA, DEDUCTOR та ін.

Класичні ієрархічні алгоритми працюють тільки з категорійними атрибутами, коли будується повне дерево вкладених кластерів. У них проводиться послідовне об'єднання вихідних об'єктів і відповідне зменшення числа кластерів.

Неієрархічні алгоритми ґрунтуються на оптимізації деякої цільової функції, що визначає оптимальне, в певному сенсі, розбиття множини об'єктів на кластери. У цій групі популярні алгоритми сімейства *k*-середніх (*k*-means), які в якості цільової функції використовують суму квадратів зважених відхилень координат об'єктів від центрів кластерів, які шукають. Кластери шукають сферичної або еліпсоїдної форми.

Окремої уваги заслуговують самоорганізуючі карти – це один з різновидів нейромережевих алгоритмів.

Основною відмінністю даної технології є те, що при навчанні використовується метод навчання без вчителя, тобто результат навчання залежить тільки від структури вхідних даних.

Алгоритм функціонування карт, що самонавчаються, (Self Organizing Maps – SOM) являє собою один з варіантів кластеризації багатовимірних векторів. До таких алгоритмів відноситься й неієрархічний алгоритм *k*-найближчих середніх (*k*-means). Важливою відмінністю алгоритму SOM є те, що в ньому всі нейрони (вузли, центри класів) впорядковані в деяку структуру (звичайно двовимірну сітку). При цьому в ході навчання модифікується не тільки нейрон-переможець, але і його сусіди, але в меншому ступені. За рахунок цього SOM можна вважати одним з методів проектування багатовимірного простору в простір з більш низькою розмірністю. При використанні цього алгоритму вектора, схожі у вихідному просторі, виявляються поруч і на отриманій карті.

Так як алгоритм SOM поєднує в собі два основних напрямки – векторне квантування і проектування, то можна знайти і основні застосування цьому алгоритму при вирішенні задачі розбиття на групи досліджуваної сукупності підприємств за заданими показниками корпоративного управління. Дану методику можна використовувати для пошуку та аналізу закономірностей у вихідних даних.

Таким чином, не існує єдиного універсального алгоритму кластеризації. При використанні будь-якого з розглянутих алгоритмів необхідно враховувати його основні переваги та недоліки для вирішення конкретної задачі, враховувати природу даних, з якими він краще працює і здатність до обробки великої кількості цих даних.

Список літератури

1. Паклин Н.Б. Бизнес-аналитика: от данных к знаниям (+ CD): учеб. пособие / Н.Б. Паклин, В.И. Орешков. – 2-е изд., перераб. и доп. – СПб.: Питер, 2010. – 704 с.
2. Кацко И.А. Практикум по анализу данных на компьютере: Учеб. пособие для вузов // И.А. Кацко, Н.Б. Паклин – М.: Колос, 2009. – 278 с.