

Савчук Т.О., Петришин С.І.

Вінницький національний технічний університет

Порівняльний аналіз використання методів кластеризації для ідентифікації надзвичайних ситуацій на залізничному транспорті

Всі дослідження, які використовують кластерний аналіз, зокрема і аналіз надзвичайних ситуацій використовують п'ять основних кроків: пошук вибірки таких ситуацій для кластеризації; визначення множини ознак, за якими будуть вони оцінюватись; обрахування значень мір схожості; застосування методу кластерного аналізу; перевірка достовірності результатів кластерного аналізу. Перед проведенням кластеризації надзвичайних ситуацій потрібно оцінити різні її методи, та визначитись, якому із них надати перевагу, щоб отриманий результат був найбільш достовірним. Обираючи між ієрархічними і неієрархічними методами, необхідно враховувати такі їх особливості:

- Неієрархічні методи виявляють більш високу стійкість по відношенню до шумів і викидів, некоректного вибору метрики, включенню незначущих змінних в набір, який бере участь в кластеризації, за їх допомогою можливо опрацювати потужні бази даних, що є потрібним для опрацювання інформації про надзвичайні ситуації. Недоліком таких методів є те, що на вхід потрібно задавати кількість кластерів або кількість ітерацій, що є не актуальним для даної предметної області, оскільки кількість кластерів можливо визначити [1].

- Якщо немає припущень щодо числа кластерів або кількості ітерацій, потрібно використовувати ієрархічні алгоритми. Ієрархічні методи, на відміну від неієрархічних будують повне дерево вкладених кластерів, що не є особливо актуальним для аналізу надзвичайних ситуацій. Недоліки таких методів кластеризації: обмеження обсягу набору даних, вибір мір близькості, негнучкість отриманих класифікацій. Перевага цієї групи методів у порівнянні з неієрархічними методами – їх наочність і можливість отримати детальне уявлення про структуру даних, що також є важливим для аналізу надзвичайних ситуацій на залізничному транспорті, але дана особливість не є принциповою для користувачів, так як важливим є лише результат [2].

Для генерації правил, за якими буде проводитись кластеризація надзвичайних ситуацій потрібно виконати такі дії: визначити набір X характеристик, який буде властивий певному кластера; для кожної з характеристик ввести величину J_i (інформативність); впорядкувати множину X за ознакою зменшення інформативності її елементів; якщо інформативність J_i певної характеристики X_i більше Sa_J , включити її до набору X ; для отриманої множини характеристик провести аналогічну процедуру для кожного з її значень x_j , досліджуючи інформативність всіх можливих інтервалів; для отриманого набору X за значеннями характеристик відібрати множину X^I надзвичайних ситуацій, що мають принаймні одну характеристику, числове значення якої належить визначеному інтервалу; отриману множину X^I упорядкувати за “значущістю” μ ; якщо “значущість” μ_i надзвичайної ситуації менша Sa_μ множини μ , результатом буде шукана множина надзвичайних ситуацій для даного набору значень характеристик; аналогічно для будь-якої множини надзвичайних ситуацій буде утворено набір ключових характеристик та їх значень; визначити правила формування кластерів з урахуванням операцій типів “Г” та “АБО”; побудувати матриці операцій – “Г” та “АБО” – та обрати з них члени з найбільшими значеннями показників операцій “Г” та “АБО”; сформувати новий набір ознак, які є у надзвичайних ситуацій; повторити три останні пункти, але вже із характеристиками на вищому рівні ієрархії; сформувати правила для визначення кластерів.

Отже, основними перевагами представленого алгоритму є висока швидкодія, висока якість, що дає можливість його ефективного використання для ідентифікації надзвичайних ситуацій в реальному режимі часу.

1. Т.О. Савчук, С.І. Петришин Використання ієрархічних методів кластеризації для аналізу надзвичайних ситуацій на залізничному транспорті // Стаття, Міжнародний науково-технічний журнал “Виміррювальна та обчислювальна техніка в технологічних процесах”. – Хмельницький, 2009. – №1, С. 193–198.
2. Методы анализа структуры http://www.sati.archaeology.nsc.ru/stat/methods_info.php.