

УДК 343.346.8 : 004.056.53

Володимир Михайлович СТРУКОВ,

*кандидат технічних наук, доцент, завідувач кафедри
інформаційних технологій факультету № 4 Харківського
національного університету внутрішніх справ*

МОДИФІКАЦІЯ АЛГОРИТМУ КЛАСТЕРИЗАЦІЇ K-MEANS

В даний час спостерігається тенденція все більш частого використання кримінальними елементами Інтернету для злочинних цілей - розповсюдження порнографії, наркотиків, зброї, втягнення неповнолітніх у сферу своєї діяльності тощо. У постійно зростаючому потоці даних шукати кримінально значиму інформацію вручну неефективно та й недоцільно.

У зв'язку з стрімким наростанням потоків неструктурованої інформації у віртуальному просторі і неможливістю накопичувати і зберігати всі дані, що поступають, в пам'яті, все більш актуальним є застосування методів аналітичної обробки потоків даних Data Mining, Text Mining, Web Mining. В основі цих наукоємних технологій лежать методи класифікації-кластеризації.

Існує досить велика кількість різних підходів, методів та алгоритмів кластеризації [1], які застосовуються і ефективні в різних конкретних ситуаціях. Розглянемо випадок, коли передбачається, що кластери добре виражені, тобто є угруповання даних, які досить чітко відокремлені одне від одного. Крім того, дані мають велику розмірність (тисячі і мільйони записів) і вимірюються у числовій шкалі.

Одними з простих та широко використовуваних алгоритмів кластеризації є алгоритми K-means і C-means [1]. Їх основними перевагами є простота реалізації, результативність і висока швидкодія. Головний недолік – заздалегідь задана кількість кластерів, на які необхідно розбити вихідний набір об'єктів. Крім того, в алгоритмах цього класу на першій стадії при визначенні первинних центрів шуканих кластерів, як правило, випадковим чином обираються K об'єктів. При певних обставинах це певною мірою прийнятно - наприклад, при відносно невеликій кількості вихідних об'єктів (до декількох сотень) і передбаченої в подальшому процедурі оптимізації знайденого рішення. Однак у тих випадках, коли кількість досліджуваних об'єктів досягає сотень тисяч і мільйонів записів, такий підхід є завідомо неприйнятним. У цьому випадку є доцільним мати

досить "хорошу" початкову вибірку первинних центрів, яка в подальшому забезпечить рішення, близьке до локального (або глобального) екстремуму.

Пропонується підзадачу визначення первинних центрів кластерів вирішувати як задачу оптимізації деякого критерію, що забезпечує максимальне розсіювання первинних центрів кластерів. Обґрунтуванням вибору такого критерію є припущення про те, що кластери добре виражені (виділені і відокремлені), отже, їх центри розташовані в такий спосіб, що вони доставляють екстремум функції розсіювання деяких параметрів первинних центрів кластерів. В якості такого параметру пропонується використовувати показник середньоквадратичного відхилення від середньої відстані між первинними центрами.

Для формалізації задачі введемо наступні позначення:

$X = \{x_{ij}\} \in R^{m \times n}$ - матриця «об'єкт-властивість», де x_{ij} - значення j -ї властивості i -го об'єкта a , $i=1,2, \dots, m$; $j=1,2, \dots, n$, при цьому кожний об'єкт може бути описаний вектором-рядком властивостей $x_i = (x_{i1}, x_{i2}, \dots, x_{in}) \in R^n$. Вважатимемо, що значення всіх n ознак задані у звичайній числовій шкалі, хоча, як показано в [2], до такого випадку можна звести і ситуацію, коли значення ознак об'єктів задані в інших шкалах.

K - кількість кластерів у задачі;

$X_c = \{x_i^c\}$, $i=1, 2, \dots, K$ - множина шуканих центрів кластерів.

Нехай також у просторі R^n об'єктів задана відстань між ними у звичайній евклідовій метриці:

$$d_{il} = \sqrt{\sum_{j=1}^n (x_{ij} - x_{lj})^2}. \quad (1)$$

Тоді через d_{ij}^c позначимо відстань між центрами кластерів x_i^c і x_j^c . Відповідно, середня відстань від об'єкта x_i^c до усіх інших центрів (об'єктів множини X_c) може бути записана у вигляді:

$$d_i^c = \frac{1}{K-1} \sum_{x_j^c \in X_c, x_j^c \neq x_i^c} d_{ij}^c. \quad (2)$$

Таким чином, цільова функція у підзадачі пошуку первинних центрів K кластерів буде мати наступний вигляд:

$$f^c = \sum_{j=1}^K \left(d_j^c - \frac{1}{K} \sum_{i=1}^K d_i^c \right)^2 \rightarrow \min. \quad (3)$$

Експериментальне дослідження програмної реалізації запропонованого підходу на тестових прикладах підтвердило його ефективність.

Список використаних джерел:

1. Aggarwal C.C., Reddy C.K. Data clustering. Algorithms and Applications. Cham: Springer Ltd. Publ., Switzerland, 2015. 734 p.
2. Бодянский Е.В., Струков В.М., Узлов Д.Ю. Задача оценки близости многомерных объектов анализа данных. *Управляющие системы и машины*. 2016. № 6. С. 67-72.

Одержано 19.10.2017

УДК 004.658.6

Юлія Володимирівна ТРЕГУБ,

*студентка Національного аерокосмічного університету ім.
Н. С. Жуковського «ХАІ»*

**ВИКОРИСТАННЯ ТЕХНОЛОГІЙ БЛОКЧЕЙН ДЛЯ ПРОТИДІЇ
КІБЕРЗЛОЧИННОСТІ У СФЕРІ ДЕРЖАВНОГО УПРАВЛІННЯ**

З розвитком інформаційного суспільства, державні послуги зазнають великих змін і переходять до обслуговування за допомогою мережі Інтернет. Саме тому, наша держава пропонує велику кількість онлайн послуг, а саме: електронний уряд, послуги центру державного земельного кадастру, тощо. Цим державним сервісам приділяють велику увагу кіберзлочинці. Тому постає питання про адекватний захист інтересів держави у кіберпросторі.

5-го жовтня 2017 року у Верховній Раді України був прийнятий законопроект №2126а від 19.06.2015 «Про основні засади забезпечення кібербезпеки в території України».

Згідно з Законом в Україні створюється національна система кібербезпеки, як "сукупність суб'єктів забезпечення кібербезпеки та взаємопов'язаних заходів політичного, науково-технічного, інформаційного характеру, організаційних, правових, оборонних, інженерно-технічних заходів, а також заходів криптографічного і технічного захисту національних інформаційних ресурсів, кіберзахисту об'єктів критичної інформаційної інфраструктури." [1].

Основними суб'єктами національної системи кібербезпеки є Державна служба спеціального зв'язку та захисту інформації України, Національна поліція України, Служба безпеки України, Міністерство оборони України та Генеральний штаб Збройних Сил України, Національний банк України, які відпо-