УДК 004.056.5

Андрій Яровий, Тетяна Польгуль, Людмила Крилик (Вінниця)

РОЗРОБКА МЕТОДУ ВИЯВЛЕННЯ ШАХРАЙСТВА ПРИ ІНСТАЛЮВАННІ МОБІЛЬНИХ ДОДАТКІВ З ВИКОРИСТАННЯМ ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ ДАНИХ

У наш час компанії витрачають великі кошти на маркетингові кампанії для просування своїх продуктів, зокрема з метою збільшення інсталювань мобільних додатків. На цьому кроці варто знати, що частина або вся множина інсталювань мобільних додатків могла бути здійснена шахрайським способом. Задача ефективного виявлення шахрайства, яке є одним з типом аномалій [1], є **актуальною**, оскільки дозволить мінімізувати затрати компаній.

Постановка задачі. Необхідно формалізувати процес виявлення аномалій з метою мінімізації затрат компаній.Для реалізації поставленої задачі здійснено аналіз традиційних способів виявлення аномалій та представлено класифікацію методів, в якій виділено 3 основні групи: методи класифікації, методи кластеризації, статистичні методи. Слід зазначити, що більшість розглянутих методів не працюють із різнорідними даними, а використовують методи переведення різнорідних даних в однорідні або ж відкидають дані, через що погіршується точність результатів. Під різнорідністю розуміємо дані різних типів (числові, категоріальні, бінарні) та розмірності, які не можливо рівноцінно порівняти між собою.Нейронні мережі у свою чергу працюють з різнорідними даними, проте на даний момент неможливо чітко обґрунтувати рішення, прийняте ними, а в досліджуваній області таке обґрунтування є обов'язковим.Тому у даній роботі запропоновано наступний алгоритм дій, який працює з різнорідними даними та дозволяє обґрунтувати прийняте рішення – він не відкидає дані (для підвищення точності) та зменшує їх розмірність (для підвищення швидкодії). Основні кроки алгоритму наступні:подолання різнорідності даних та зменшення розмірності даних, оптимізація даних, класифікація даних.

Оскільки виявлення аномалій відноситься до задач інтелектуального аналізу даних, для якого традиційно використовують теорію множин, то для формалізації представленогопроцесу виявлення аномалій використаємо теорію множин (формула 1.1).

$$\begin{cases} \bar{X}\begin{pmatrix} U1(x1,\dots,xn) \\ U2(x1,\dots,xn) \\ \dots \\ Us(x1,\dots,xn) \end{pmatrix} \rightarrow F1\,(\bar{X}) \rightarrow \begin{pmatrix} U1(k01) \\ U2(k02) \\ \dots \\ Us(k0s) \end{pmatrix} \\ \overline{B}\begin{pmatrix} U1(b1,\dots,bk) \\ U2(b1,\dots,bk) \\ \dots \\ Us(b1,\dots,bk) \end{pmatrix} \rightarrow F2(\overline{B}) \rightarrow \begin{pmatrix} U1(k11) \\ U2(k11) \\ \dots \\ Us(k1s) \end{pmatrix} \rightarrow F4\begin{pmatrix} U1(k01,k11,\dots,k21) \\ U2(k02,k12,\dots,k22) \\ \dots \\ Us(k0s,k1s,\dots,k2s) \end{pmatrix} \rightarrow F5 \rightarrow \begin{pmatrix} U1(C1) \\ U2(C2) \\ \dots \\ Us(C1) \end{pmatrix}, \\ \dots \\ \overline{W}\begin{pmatrix} U1(w1,\dots,wm) \\ U2(w1,\dots,wm) \\ \dots \\ Us(w1,\dots,wm) \end{pmatrix} \rightarrow F3(\overline{W}) \rightarrow \begin{pmatrix} U1(k21) \\ U2(k22) \\ \dots \\ Us(k2s) \end{pmatrix} \end{cases} \quad (1.1)$$

де $\overline{X}, \overline{B}, \dots, \overline{W}$ – вектори з множинами вхідних даних по користувачам.Кожен вектор містить множини з однорідними даними, що дозволяє уникнути кроку приведення різнорідних даних в однорідні;$F1(\overline{X}), F2(\overline{B}), \dots, F3(\overline{W})$– операції над множинами, які вирішують задачу подолання різнорідності даних та зменшення їх розмірності;$F4, F5$ – операції над множинами, які оптимізують дані та класифікують їх. З використанням Python та пакету scikit-learnбуло протестовано основні методи виявлення аномалій а також запропонований алгоритм на основі вибірки з [2], чим підтверджено адекватність розробленого алгоритму.

Висновки. Здійснено класифікацію існуючих методів виявлення аномалій. Формалізовано процес виявлення аномалій при виявлені шахрайства щодо інсталювань мобільних додатків. Здійснено програмну реалізацію запропонованогометоду та порівняно його з існуючими.

Список літературнних джерел

1. V. Chandola. Anomaly Detection : A Survey / V. Chandola, A. Banerjee, V. Kumar – ACM Computing Surveys (CSUR), Volume 41, Issue 3, Article No. 15, New York, NY, USA, July 2009.
2. Credit Card Applications [Електроннийресурс]. – Режим доступу: https://www.kaggle.com/ujjwal9/credit-card-applications