УДК 004.021

ИСПОЛЬЗОВАНИЕ МЕТОДОВ МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ И ТЕОРИИ ГРАФОВ С ЦЕЛЬЮ ПРЕДОТВРАЩЕНИЯ ФИНАНСОВОГО МОШЕННИЧЕСТВА

А.С. Портнова portnova-a@mail.ru

МГТУ им. Н.Э. Баумана, Москва, Российская Федерация

Аннотация

Рассмотрены методы математической статистики и теории графов, позволяющие обнаруживать мошенников в информационной сфере. Первый метод позволяет предотвратить противозаконную деятельность всей мошеннической структуры, второй — определить, кто именно занимается незаконной деятельностью на бирже, третий — обнаружить фальсификацию данных

Ключевые слова

Антифрод, закон Бенфорда, марковская сеть, инсайдерская сеть

Поступила в редакцию 20.07.2016

Введение. В сфере информационных технологий вследствие мошеннических действий компании и частные пользователи ежегодно несут убытки в сотни млрд долл. США. В целях защиты от онлайн-мошенничества применяют методы с привлечением математического аппарата, в частности методы математической статистики и теории графов. К задачам безопасности относят обнаружение инсайдерских сделок, отслеживание незаконных транзакций в банках, выявление мошенников на онлайн-аукционах, обнаружение поддельных аккаунтов в социальных сетях и проч. В настоящей работе рассмотрены методы на основе:

1) марковских сетей и их модификаций, которые позволяют обнаруживать мошеннические учетные записи;

2) анализа модели инсайдерской сети, построенной с помощью неориентированного графа с целью определения состава этой сети;

3) статистических выводов закона Бенфорда для проверки подлинности данных.

Обнаружение мошеннических учетных записей с использованием марковских сетей. Различные интернет-магазины и онлайн-аукционы набирают все большую популярность среди пользователей. Так, на ресурсе «eBay» количество пользователей приблизилось к нескольким миллионам, а денежный оборот ресурса исчисляется в млрд долл. США, что весьма привлекательно для мошенников [1]. Они действуют по известной и отработанной схеме. Сначала пособники, которые имеют хорошую репутацию (поскольку ведут себя, как честные пользователи), разными способами повышают рейтинг злоумышленника, например оставляют положительные отзывы. Затем злоумышленник становится активен и совершает ряд противоправных действий. Даже если администрация сайта быстро обнаружит мошенника и заблокирует его учетную запись, предъявить претензии к подельникам все равно не удастся. Такая безнаказанность позволяет злоумышленникам повторить весь процесс в кратчайшие сроки.

С помощью теории графов можно обнаружить криминальную сеть и заблокировать аккаунты всех ее участников. В этом случае для восстановления сети преступникам потребуется гораздо больше времени и усилий, чем при блокировании только мошенника.

Предположим, имеются данные о пользователях и сделках одного онлайн-аукциона. Представим эти данные в виде графической модели. Для этого построим неориентированный граф, например марковскую сеть, вершинами которой будут учетные записи онлайн-аукциона, а ребрами — вероятные связи между пользователями (сделки, покупки и т. д.).

На начальном этапе все вершины — честные пользователи. Если аккаунты провели хотя бы одну сделку, соответствующие им вершины свяжет дуга. Далее понадобится алгоритм, чтобы определить, кто именно является мошенниками и пособниками. Алгоритм идентификации мошеннических вершин с помощью анализа связей его соседей называют алгоритмом распространения доверия (Belief Propagation) [2]:

$$m_{ij}(\sigma) \leftarrow \sum_{\sigma'} \psi(\sigma', \sigma) \prod_{n \in \mathrm{N}(i) \setminus j} m_{nj}(\sigma'); \quad (1)$$

$$b_i(\sigma) \leftarrow k \prod_{j \in \mathrm{N}(i)} m_{ij}(\sigma), \quad (2)$$

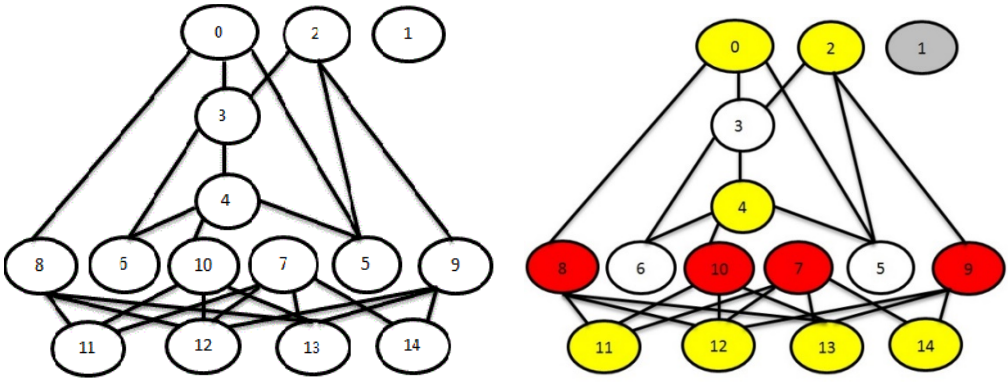
где m_{ij} — сообщение, отправленное узлом i узлу j; $\psi(\sigma', \sigma)$ — вход в матрицу распространения, дающий вероятность нахождения узла в состоянии σ' при наличии соседнего узла также в состоянии σ; $N(i)$ — совокупность узлов, соседних i; $b_i(\sigma)$ — уровень доверия узла i в состоянии σ; k — константа нормировки.

Представленный алгоритм — это алгоритм маргинализации с помощью двунаправленной передачи сообщений на графе, применяемый для вывода на графических вероятностных моделях. Суть его заключается в следующем: вершина «выясняет» свое состояние, основываясь на информации о состоянии своих соседей и собственном текущем состоянии. Далее, учитывая обновленные данные, с помощью матрицы распространения доверия (см. таблицу) сообщает о своем новом текущем состоянии. В матрице представлены вероятности отношений вершин с различными состояниями, которых может быть три: мошенник, пособник, честный пользователь.

Матрица распространения доверия

Соседние состояния	Состояние узла		
	Мошенник	Пособник	Честный пользователь
Мошенник	0,05	0,9	0,05
Пособник	0,5	0,1	0,4
Честный пользователь	0,05	0,475	0,475

Вероятность того, что мошенник вступит в контакт с другим мошенником очень мала. Также мала вероятность того, что честный пользователь обманет мошенника. Пособники же взаимодействуют и с мошенниками, и с честными пользователями, но, как и мошенники, редко контактируют друг с другом. При этом честные пользователи с равной вероятностью могут иметь отношения и с пособниками, и с другими законопослушными пользователями. Все эти обстоятельства учтены в приведенной матрице.



а *б*

Модель графа до (*а*) и после применения алгоритма (*б*)

Сначала все вершины — честные пользователи (рисунок, часть *а*), далее запускается описанный выше алгоритм распространения доверия (1) и (2), на каждой итерации которого вершины «обмениваются» информацией о своем текущем состоянии и меняют свой статус относительно других. После завершения работы алгоритма (рисунок, часть *б*) видно, кто мошенники (красные номера), пособники (желтые номера), честные пользователи (неокрашенные номера), а также вершины с неопределенным состоянием (серые номера).

Модель инсайдерской сети, построенная с помощью неориентированного графа. Другой сферой деятельности мошенников является инсайдерская торговля — сделки между должностными лицами, директорами, крупными акционерами или частными лицами, обладающими конфиденциальной внутренней информацией, позволяющей получать прибыль от покупки или продажи акций [3]. Разумеется, законом не запрещено служащим, имеющим доступ к важной информации о компании, торговать на бирже. Однако если их решение о покупке или продаже акций основано исключительно на конфиденциальных данных компании, это может привести к нестабильности и потерям, не только компании, но и всей экономики.

Инсайдеры редко работают в одиночку. Зачастую они организуют сеть, в которой обмениваются необходимыми данными. Если имеется информация о служащих и их сделках на бирже, то, построив граф, несложно вычислить инсайдерскую сеть. Вершинами графа будут предположительные инсайдеры, а ребрами соединятся вершины тех, кто следует похожим паттернам в торговле. Это

означает, что их продажи и покупки будут происходить в один день или что они будут покупать акции только одной компании на определенном отрезке времени. Чтобы понять, нужно ли создавать между вершинами ребро, рассмотрим функцию (3), которая позволит сравнивать информацию о двух вершинах.

$$S\left(X_c, Y_c\right) = \frac{\left(\sum_{i=1}^{|X_c|}\sum_{j=1}^{|Y_c|} I(x_i, y_j)\right)^2}{|X_c| \times |Y_c|}, \quad (3)$$

здесь для каждой пары служащих X и Y некоторой компании сравним набор транзакций с датами X_c и Y_c. Функция $I\left(x_i, y_j\right)$ равна 1, если $x_i = y_j$, в противном случае — 0. Если инсайдеры торгуют в одну и ту же дату постоянно, то $S\left(X_c, Y_c\right) = 1$. Далее включим эту пару вершин в граф и создадим между ними ребро. Функция равна 0, если нет общих дат сделок.

Необходимо построить два графа: один, связанный с покупками, другой — с продажами на бирже. Если после анализа видно, что для одних и тех же связок вершин характерны одинаковые покупки и продажи, то с высокой степенью вероятности можно говорить об инсайдерской сети.

Статистические выводы закона Бенфорда для проверки подлинности данных. Финансовая сфера связана с огромным количеством статистических данных: бухгалтерия, счета, транзакции, налоговые декларации, страховые выплаты и др. Специалисты в сфере финансовой безопасности давно пытаются решить задачи, связанные с выявлением мошеннической активности и фальсификациями в данных с помощью различных математических моделей и распределений.

Одним из способов выявления аномалий в данных является применение закона Бенфорда. В 1938 г. Фрэнк Бенфорд доказал, что цифры в номерах естественных систем встречаются не с равной вероятностью [4]. Статистические выводы настоящего наблюдения можно использовать при анализе финансовых и других данных в сфере информационной безопасности. Применение данного метода заключается в исследовании имеющихся данных и сравнении результатов с известными вероятностями появления определенной первой значащей в них цифры. Данные из финансовых отчетов также подчиняются закону Бенфорда, но с некоторыми условиями:

- отсутствие системы нумерации (например, в номерах банковских карт первые шесть цифр отражают информацию об организации, выдавшей карту и типе карты в рамках данной платежной системы);
- объем проверяемой информации должен быть достаточно большим;
- отсутствие ограничений по максимуму и минимуму (например, минимальная и максимальная сумма для снятия наличных денег).

В том случае, когда данные появляются вследствие естественных причин, например переводимые клиентами суммы денежных средств или номера платежных поручений от различных покупателей, то в таких данных с высокой степенью вероятности можно вычислить аномалии при помощи закона Бенфорда.

Аналитики используют различные тесты на основе закона Бенфорда для разных задач (анализа частоты первой цифры, анализа частоты первой и второй цифры, анализа округлений и т. д.), однако алгоритм для всех вариаций одинаков.

Алгоритм выявления аномалии в данных, основанный на законе Бенфорда.

1. Формирование данных для анализа.
2. Определение «эталона» для данной ситуации.
3. Исследование частоты появлений цифр из выбранного диапазона (в зависимости от выбранного теста, она будет разной).
4. Сравнение рассчитанной частоты появления цифр и эталонной.
5. Наложение критерия аномальности (отклонение от эталона), с применением регрессивного анализа.
6. Формирование данных, которые следует проверить вручную.
7. Выводы об аномальности проверенной информации.

Иначе говоря, для сформированных данных требуется определить эталон согласно закону Бенфорда, рассчитать реальную частоту появления цифр и сравнить ее с эталонной, затем исследовать отклонения от эталона.

Представленные методы широко применяют в сфере информационной безопасности. Например, тесты на основе закона Бенфорда включены в программные комплексы, разработанные для аудиторов или ревизоров. Известны следующие пакеты: ACL компании ACL ServicesLtd, AuditNET, ActiveData компании InformationActiveLtd и др. Также с помощью теории графов и закона Бенфорда вычисляют ботов в социальных сетях [4].

Выводы. Представленные методы не всегда работают совершенно, прежде всего, в виду сложностей реализации в режиме реального времени. Поэтому в дальнейших исследованиях планируется модифицировать представленные алгоритмы, а также найти новые сферы применения.

Литература

1. D. Kavu1 T., Rugube T., Kawondera F., Chifamba N. A fraud detection tool in e-auctions // African journal of mathematics and computer science research. 2016. Vol. 9. No.1. P. 1–11. DOI: 10.5897/AJMCSR2015.0593

2. Pandit S., HorngChau D., Wang S., Faloutsos C. NetProbe: a fast and scalable system for fraud detection in online auction networks // Sixteenth International World Wide Web Conference (WWW2007). Banff, Alberta, CANADA. 2007. May 8–12.

3. Словарь финансово-экономических терминов / Под ред. И.З. Ярыгиной, Н.Г. Кондрахиной. М.: Финансовый университет. 2012. 172 с.

4. Golbeck J. Benford's law applies to online social networks // PLoS ONE. 2015. Vol. 10. No.8. DOI: 10.1371/journal.pone.0135169

Портнова Анна Сергеевна — студентка кафедры «Информационная безопасность», МГТУ им. Н.Э. Баумана, Москва, Российская Федерация.

Научный руководитель — Н.С. Коннова, старший преподаватель кафедры «Информационная безопасность», МГТУ им. Н.Э. Баумана, Москва, Российская Федерация.

MATHEMATICAL STATISTICS AND GRAPH THEORY METHODS PREVENTING FINANCIAL FRAUD

A.S. Portnova portnova-a@mail.ru

Bauman Moscow State Technical University, Moscow, Russian Federation

Abstract

The study tested the mathematical statistics and graph theory methods allowing us to detect fraudsters in the information environment. The first method makes it possible to prevent illegal activities of the whole fraudulent structure, the second method enables us to determine who is engaged in illegal activity on the stock exchange, the third one is good for detecting the data falsification

Keywords

Antifraud, Benford's Law, Markov network, insider network

References

[1] D. Kavu1 T., Rugube T., Kawondera F., Chifamba N. A fraud detection tool in e-auctions. African journal of mathematics and computer science research, 2016, vol. 9, no. 1, pp. 1–11. DOI: 10.5897/AJMCSR2015.0593

[2] Pandit S., HorngChau D., Wang S., Faloutsos C. NetProbe: a fast and scalable system for fraud detection in online auction networks. Sixteenth International World Wide Web Conference (WWW2007), Banff, Alberta, CANADA. 2007. May 8–12.

[3] Yarygina I.Z., Kondrakhina N.G., eds. Slovar' finansovo-ekonomicheskikh terminov [Financial and economic terms dictionary]. Moscow, FU Publ., 2012. 172 p. (in Russ.).

[4] Golbeck J. Benford's law applies to online social networks. PLoS ONE, 2015, vol. 10, no. 8. DOI: 10.1371/journal.pone.0135169

Portnova A.S. — student of the Department of Information Security, Bauman Moscow State Technical University, Moscow, Russian Federation.

Scientific advisor — N.S. Konnova, Assist. Professor of the Department of Information Security, Bauman Moscow State Technical University, Moscow, Russian Federation.