УДК 004.93'14:004.048

ОБНАРУЖЕНИЕ АНОМАЛИЙ И ФАЛЬСИФИКАЦИЙ В ДАННЫХ СОЦИАЛЬНЫХ СЕРВИСОВ В РАМКАХ ЦИФРОВОЙ ЭКОНОМИКИ

© 2019 П.В. Хрипунов[1], Е.Ю. Минаев[1], В.И. Проценко[1,2], Н.С. Давыдов[1], А.В. Никоноров[1,2]

[1] Самарский национальный исследовательский университет имени академика С.П. Королёва
[2] Институт систем обработки изображений -
филиал ФНИЦ «Кристаллография и фотоника» РАН, г.Самара

Статья поступила в редакцию 21.08.2019

Статья посвящена исследованию проблемы выявлений аномалий и фальсификаций в данных, поступающих от социальных сервисов. Задача выявления аномалий крайне актуальна для основанных на данных процессов цифровой экономики. В настоящей работе предложен подход поэтапного детектирования аномалий с применением автокодировщиков и критерия сопряженности. Экспериментальное исследование эффективности предложенных алгоритмов проведено на открытом тестовом наборе данных.
Ключевые слова: обнаружение аномалий, фальсификации в данных, автокодировщики, сверточные нейронные сети, критерий сопряженности

Методы и алгоритмы разработаны при поддержке грантов РФФИ (19-29-01235-мк, 16-29-11744-офи-м, № 16-29-09528-офи-м, № 17-29-03112-офи-м, № 18-07-01390-А), экспериментальные исследования – в рамках госзадания ИСОИ РАН - филиал ФНИЦ «Кристаллография и Фотоника» РАН (соглашение № 007-ГЗ/Ч3363/26).

1. ВВЕДЕНИЕ

В связи с активным проникновением информационных технологий в социально-экономическую сферу, государственное управление и бизнес, задача поиска возможных фальсификаций и непреднамеренных искажений данных в корпоративных и государственных информационных системах в настоящее время приобретает особую актуальность[1]. С одной стороны, цифровизация государственных услуг и бизнес-процессов позволяет существенно ускорить и упростить доступ к необходимой информации, с другой стороны, с ростом количества информационных систем и связей между ними, зачастую ослабляется контроль над правомерностью доступа к данным, что может привести к нежелательным искажениям. Увеличение количества пользователей информационных систем, в свою очередь, приводит к увеличению возможных ошибок и неточностей на этапе сбора информации[2].

Хрипунов Павел Владимирович, аспирант-экстерн кафедры суперкомпьютеров и общей информатики Самарского университета. E-mail: odissey-iip@yandex.ru
Минаев Евгений Юрьевич, кандидат технических наук, доцент кафедры суперкомпьютеров и общей информатики Самарского университета.
E-mail: e.minaev@gmail.com
Проценко Владимир Игоревич, аспирант, м.н.с. лаборатории интеллектуального анализа видеоданных ИСОИ РАН. E-mail: protsenkovi@gmail.com
Давыдов Никита Сергеевич, аспирант.
E-mail: amail9496@gmail.com
Никоноров Артём Владимирович, доктор технических наук, профессор кафедры суперкомпьютеров и общей информатики Самарского университета, руководитель лаборатории интеллектуального анализа видеоданных ИСОИ РАН. E-mail: artniko@gmail.com

2. АНАЛИЗ АНОМАЛИЙ В СОЦИАЛЬНЫХ И ГОСУДАРСТВЕННЫХ СЕРВИСОВ В ЦИФРОВОЙ ЭКОНОМИКЕ

Методы интеллектуального анализа данных для выявления аномалий в информационных сервисах делятся на шесть основных категорий: методы классификации, кластеризации, регрессии, обнаружения выбросов, визуализации и прогнозирования [3]. Каждая из этих категорий включается в себя конкретные методы. Например, нейронные сети и метод опорных векторов используются для классификации данных, метод К-средних используется для кластеризации данных. Кроме того, интеллектуальный анализ данных включает в себя множество методов из других областей, таких как статистика, машинное обучение, распознавание образов, базы данных и хранилища данных, поиск информации, визуализация, высокопроизводительные вычисления и других прикладных областей [4]. В последнее время обнаружение фальсификаций объединяет подход обнаружения аномалий и подход, основанный на использовании методов интеллектуального анализа данных [5].

Метод обнаружения аномалий или выбросов опирается на методы поведенческого профилирования, в которых моделируется поведение каждого человека, и отслеживаются любые отклонения от нормы. У методов обнаружения фальсификаций на основе аномалий есть потенциал для обнаружения новых мошеннических действий. Такие методы можно дополнительно классифицировать по трем типам [6]: автоматические, полуавтоматические, с предварительным обучением.

Для методов с предварительным обучением требуется набор данных, который заранее классифицирован на «мошеннические» и «не мошеннические» метки и предполагает обучение классификатора. Основным преимуществом методов с предварительным обучением является то, что все результаты классификации имеют понятное значение для человека, и их можно легко использовать для классификации различных шаблонов и регрессионного анализа. Однако методы с предварительным обучением имеют несколько ограничений. Первое из них связано с трудностью предварительной классификации данных на «мошеннические» и «не мошеннические». Когда имеется огромный объем входных данных, маркировка является очень трудоемкой задачей, и не всегда выполнимой в реальных условиях. Во-вторых, не всегда можно чётко маркировать те или иные данные, возникают неопределенности и двусмысленности. В некоторых случаях эти ограничения могут препятствовать реализации подходов с предварительным обучением. Поэтому для преодоления этих недостатков используются автоматическое обучение и полуавтоматическое обучение. Автоматические методы обучения без учителя позволяют выявлять фальсификации, в тестовых немаркированных данных, основываясь на предположении, что большинство образцов данных в наборе не являются фальсифицированными. В отличие от методов с обучением, не требуется маркировка данных на классы при построении модели. Основное преимущество использования неконтролируемого подхода заключается в том, что он не опирается на точную идентификацию данных по классам, которые зачастую невозможно определить заранее.

Полуавтоматические методы представляют собой гибрид описанных выше подходов. Основной целью полуавтоматического подхода является обучение классификатора как по маркированным, так и по немаркированным данным. Полуавтоматические методы имеют больше преимуществ по сравнению с методами с предварительным обучением, поскольку они обеспечивает лучшую производительность за счет одновременного использования как маркированных, так и немаркированных данных. Кроме того, полуавтоматические методы предоставляют вычислительные модели для изучения данных, в которых большая часть информации не маркирована.

3. ВЫЯВЛЕНИЕ АНОМАЛИЙ НА ОСНОВЕ СЕТЕЙ АВТОКОДИРОВЩИКОВ

Автокодировщики - нейронные сети, цель которых выучить тождественное отображение при условии ряда ограничений, накладываемых на её архитектуру. Одним из таких ограничений может быть многослойная сеть с меньшим количеством внутренних нейронов, чем внешних. Активации, получаемые на наименьшем слое, позволяют представлять исходные данные в сжатом виде и широко используются в приложениях для дальнейшей машинной обработки. Такие активации называют выходными значениями кодировщика. Автокодировщики имеющие одинаковое или большее количество нейронов во внутренних слоях также представляют интерес, при определённой регуляризации или наличии штрафа в функции минимизации. Сжимающие автокодировщики (contractive autoencoders) позволяют получить кодировщики менее чувствительные к слабым изменениям данных обучающей выборки благодаря регуляризации, согласованной с Фробениусовой нормой матрицы Якоби активаций кодировщика. При ограничении количества ненулевых активаций автокодировщика на элемент входной выборки, процесс обучения позволяет получить энкодер, возвращающий разреженные вектора значений активаций. Добавление гауссова шума к входным векторам, или к активациями внутренних слоёв приводит к обучению весов с более плавными градиентами. Такие, устойчивые к определённому шуму, автокодировщики называют шумоподавляющими.

На основе автокодировщика может быть построен классификатор, определяющий принадлежность к классу, на данных которого он был обучен. Правило такого классификатора зависит от выбранного порогового значения, применяемого к ошибке между входными и реконструированными данными. Изменение порогового значения позволяет получить нужное в решаемой задаче соотношение количества ложно-позитивных и ложно-негативных срабатываний, которые возникают при невозможности строгого разделения классов. Автокодировщик может быть обучен ради получения кодировщика, отображающего исходное пространство данных в пространство, в котором качество классических методов классификации повышается, например данные становятся линейно разделимыми. Архитектура и анализ результатов экспериментов подобного двухэтапного классификатора будут представлены далее в работе.

4. ПОВЫШЕНИЕ ТОЧНОСТИ С ПРИМЕНЕНИЕМ КРИТЕРИЯ ПО ПОКАЗАТЕЛЮ СОПРЯЖЕННОСТИ

В качестве второго этапа классификации предлагается использовать так называемый показатель сопряженности. Для построения классификатора будем использовать подход, описанный в работе [7]. Для каждого (*к*-го) класса из принадлежащих ему M обучающих векторов $\mathbf{x}_j(k), \quad j=\overline{1,M}, k=\overline{1,K}$ составляется – $N\times M$ -матрица:

$$\mathbf{X}_k=\left[\mathbf{x}_1(k),\mathbf{x}_2(k),...,\mathbf{x}_j(k),...,\mathbf{x}_M(k)\right], k=\overline{1,K} \quad (1)$$

и вычисляется $N\times N$ -матрица k -го класса:

$$\mathbf{Q}_k=\mathbf{X}_k\left[\mathbf{X}_k^T\mathbf{X}_k\right]^{-1}\mathbf{X}_k^T, \quad k=\overline{1,K}, \quad (2)$$

которую далее мы будем называть *решающей*.

На этапе распознавания решение о принадлежности вектора $\mathbf{x}_j$ к m -му классу принимается, если:

$$R_m\left(\mathbf{x}_j\right)=\max_{k=1,K}R_k\left(\mathbf{x}_j\right), \quad (3)$$

где $R_k\left(\mathbf{x}_j\right)=\mathbf{x}_j^T\mathbf{Q}_k\mathbf{x}_j\left(\mathbf{x}_j^T\mathbf{x}_j\right)^{-1}, k=\overline{1,K}$ – показатели сопряжённости текущего вектора $\mathbf{x}_j$ с каждым из распознаваемых классов.

Нетрудно заметить, что в данном методе информация о классах содержится в матрицах $\mathbf{Q}_k \quad k=\overline{1,K}$, вычисленных по матрицам $\mathbf{X}_k(M)$. В работе [8] для формирования этих матриц предложено использовать небольшое число обучающих векторов, образующих так называемые опорные подпространства классов, за счет данного свойства можно использовать показатель сопряженности в случае небольшой обучающей выборки, что часто сопутствует задаче классификации аномалий.

В данной работе в качестве исходных векторов по которым проводится обучение классификатора используются промежуточные выходы внутреннего слоя автокодировщика.

5. РЕЗУЛЬТАТЫ ЭКСПЕРИМЕНТОВ

В данной работе использовался автокодировщик с 4 слоями, два внутренних слоя которого были в два раза больше по количеству нейронов – 58, чем два внешних – 29. В ходе анализа набора данных было найдено небольшое количество мошеннических транзакций в данных, очень тесно расположенных рядом с обычными транзакциями. Повышение размерности на первом этапе двухэтапной классификации позволило легче разделить два класса. В качестве функций активации слоёв были использованы чередующиеся от первого к последнему слою функции tanh и LeakyRELU.

Для обучения были выбраны следующие параметры: метод оптимизации – RMSProp [9], количество эпох – 50, размер выборки итерации метода оптимизации – 32, функция потерь – среднее квадратическое отклонение. Обучение автокодировщика производилось на обычных транзакциях, составляющих 80% из всего количество транзакций. В тестовых данных присутствовали все мошеннические транзакции и около 20% обычных транзакций.

Перед обучением мы провели предобработку данных. Была исключена колонка со временем транзакции, так как её наличие не давало улучшения качества классификации и ухудшало разделимость групп мошеннических транзакций в двумерной t-sne проекции. Была также произведена нормировка колонки с суммой перевода Amount для сопоставимого диапазона значений с другими признаками.

В качестве метрик оценки качества классификации были использованы: AUC и PR-AUC. В силу дисбаланса набора данных и приоритета в выявлении слабо представленного класса мошеннических транзакций в качестве основной метрики будет использоваться PR-AUC. Высокие значения AUC покажут насколько хорошо разделяются два класса.

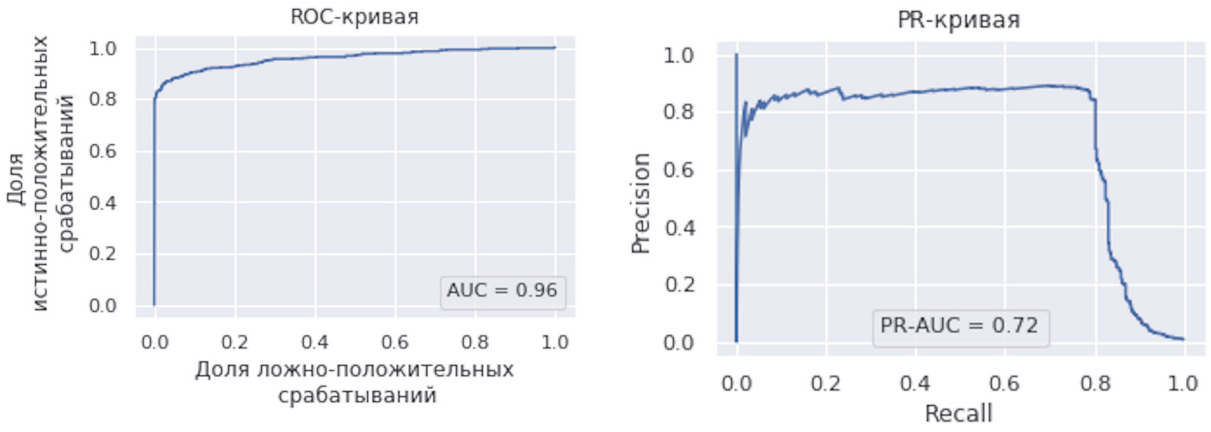
5.1. Описание экспериментального набора данных

Набор данных [10] содержит последовательность транзакций европейских владельцев карт за два дня в сентябре 2013 года. Транзакции разделяются на два класса – 492 мошеннических и 284315 легальных. Набор данных является сильно несбалансированным – мошеннические транзакции составляют всего 0,172% от общего количества. Из соображений приватности персональных данных таблица содержит только числовые значения, являющиеся результатом преобразования методом главных компонент исходных логов транзакций. Исключение составляют две колонки: сумма перевода и время в секундах, отсчитываемое с начала первой транзакции из набора данных.

5.2. Результаты

Использование нелинейной активации LeakyReLU с коэффициентом наклона 0.3 позволило увеличить PR-AUC по сравнению с классическим ReLU с 0,47 до 0,72.

На рисунках 1а и 1б видны результаты классификатора, построенный исключительно на обученном автокодировщике. Метрики классификации AUC и PR-AUC были равно 0,96 и 0,72 соответственно. На рисунке 2 приведены визуализация корректно и некорректно клас-



(а) (б)

Рис. 1. Результаты работы одноэтапного классификатора, построенного на основе автокодировщика:
а – ROC-кривая; б – PR-кривая

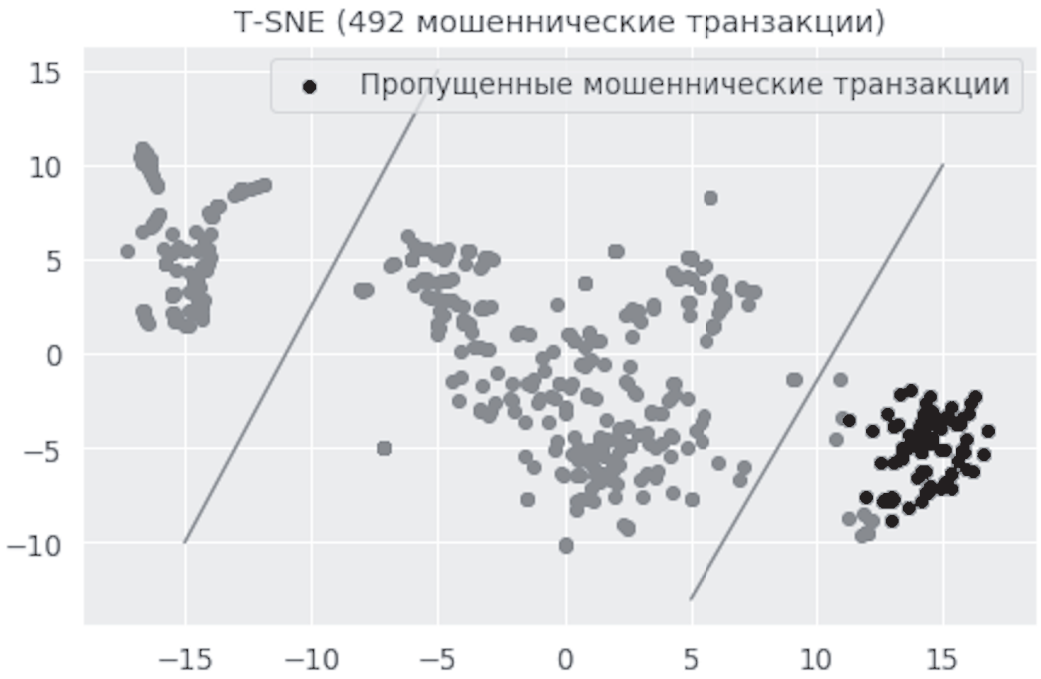


Рис. 2. Корректно и ошибочно классифицированные случаи фальсификации в пространстве предобработанных признаков

сифицированных случаев фальсификации в пространстве предобработанных признаков. Можно видеть, что ошибки классификации автокодировщика составляют достаточно плотный кластер. Дополнительный анализ показал, что пропущенные фальсификации имеют сильное пересечение распределений значений признаков с распределениями соответствующих признаков обычных транзакций. На рисунке 3 показана матрица ошибок.

Для улучшения распознавания 83 аномалий, образующих кластер на рисунке 2, которые не удалось классифицировать автокодировщиком и уменьшения ложных 534 срабатываний, был применен второй этап распознавания с применением критерия по показателю сопряженности. В качестве исходных векторов по которым проводилось обучение классификатора использовались промежуточные выходы внутреннего слоя автокодировщика из 58 элементов. За счет построения опорных подпространств на 10% от неверно распознанных векторов, в результате второго этапа классификации удалось уменьшить количество нераспознанных мошеннических транзакций до 25 и количество ложных срабатываний до 157 транзакций. Таким образом, предложенный подход, заключающийся в уточнении детектора на основе автокодировщиков за счет применения критерия сопряженности доказал свою эффективность на тестовом наборе данных.

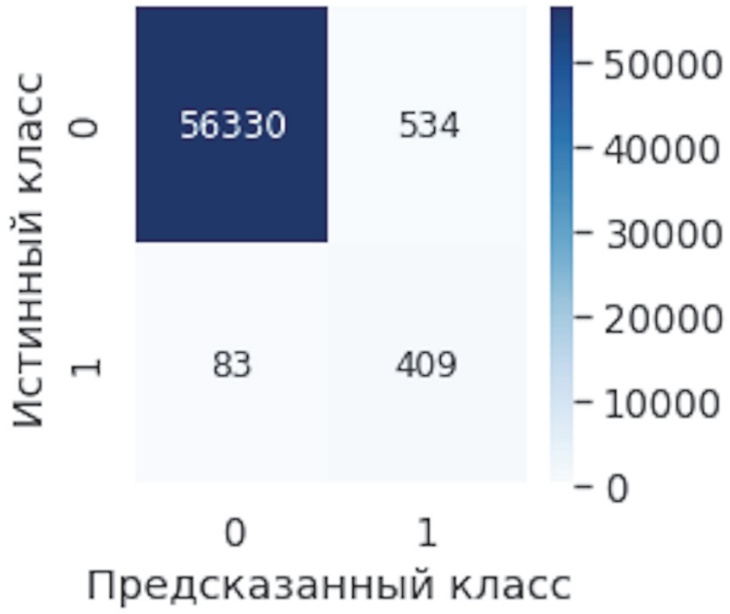


Рис. 3. Матрица ошибок классификации одноэтапным классификатором. 0 соответствует классу обычных, а 1 классу мошеннических транзакций

6. ЗАКЛЮЧЕНИЕ

В настоящей работе проведено исследование задачи выявления фальсификаций в данных полученных в результате функционирования социальных сервисов. Несмотря на актуальность создания эффективных алгоритмов обнаружения фальсификаций в данных в рамках цифровой экономики, в открытом доступе присутствует крайне небольшое число наборов данных, позволяющих провести валидацию таких алгоритмов. Исследования показали, что классический подход на основе автокодировщиков позволяет выполнять детектирование фальсификаций, однако точность такого детектирования невысокая. Поднять точность позволил проведенный нами дополнительный этап классификации на основе критерия сопряженности. Проверка эффективности предложенного подхода к выявлению аномалий на других наборах данных, полученных в социальных системах, а также валидация алгоритма на синтезированных данных с заранее известными характеристиками являются предметом дальнейших исследований.

СПИСОК ЛИТЕРАТУРЫ

1. *Dong W., Liao S., Zhang Z.* Leveraging financial social media data for corporate fraud detection // Journal of Management Information Systems. 2018. V. 35(2). P. 461-487.
2. A comprehensive survey of data mining-based fraud detection research / *C. Phua, V. Lee, K. Smith, R. Gayler* // arXiv (preprint arXiv:1009.6119). 2010. P. 1-14.
3. *Abdallah A., Maarof M. A., Zainal A.* Fraud detection system: A survey // Journal of Network and Computer Applications. 2016. № 68. P. 90–113. doi:10.1016/j.jnca.2016.04.007
4. *Han J., Pei J., Kamber M.* Data mining: concepts and techniques. Elsevier, 2011. 744 p.
5. *Sasirekha M.* A defense mechanism for credit card fraud detection // Int. J. Cryptogr. Inf. Secur. 2012. V.2. №3. P. 89-100.
6. *Akhilomen J.* Data mining application for cyber credit-card fraud detection system // Industrial Conference on Data Mining. Berlin: Springer, 2013. P. 218-228.
7. *Жердев Д. А., Казанский Н. Л., Фурсов В. А.* Распознавание объектов на радиолокационных изображениях с использованием показателей сопряжённости и опорных подпространств // Компьютерная оптика. 2015. Т. 39. № 2. С. 255-264.
8. *Жердев Д. А., Казанский Н. Л., Фурсов В. А.* Распознавание объектов по диаграммам рассеяния электромагнитного излучения на основе метода опорных подпространств // Компьютерная оптика. 2014. Т. 38. № 3. С. 503-510.
9. *Hinton G.* RMSProp [Electronic resource] – URL: https://www.cs.toronto.edu/~tijmen/csc321/slides/lecture_slides_lec6.pdf (online; accessed: 2019-09-06).
10. Credit Card Fraud Detection Dataset [Electronic resource] – URL: https://www.kaggle.com/mlg-ulb/creditcardfraud (online; accessed: 2019-09-06).

ANOMALY AND FRAUD DETECTION BASED ON SOCIAL SERVICES DATA IN THE SPHERE OF DIGITAL ECONOMY

© 2019 P.V. Khripunov[1], E.Y. Minaev[1], V.I. Protsenko[1,2], N.S. Davydov[1], A.V. Nikonorov[1,2]

[1] Samara National Research University named after Academician S.P. Korolyov
[2] IPSI RAS - Branch of the FSRC «Crystallography and Photonics» RAS, Samara

The article is devoted to the study of the anomaly and fraud detection problem in the data from social services. The problem of detecting anomalies is extremely relevant for data-based processes in the digital economy. In this paper, we propose a two-step approach for the phased detection of anomalies using auto-encoders and the contingency criterion. An experimental study of the efficiency of the proposed algorithms was conducted on an open test data set.
Keywords: anomaly detection, data fraud, auto-encoders, convolutional neural networks, contingency criterion.

Pavel Khripunov, Postgraduate Student at the Department of Supercomputers and General Informatics, Samara University. E-mail: odissey-iip@yandex.ru
Evgeny Minaev, Candidate of Technical Sciences, Associate Professor at the Department of Supercomputers and General Informatics, Samara University. E-mail: e.minaev@gmail.com
Vladimir Protsenko, Postgraduate Student, Junior Researcher of the Laboratory for Video Intelligent Analysis. E-mail: protsenkovi@gmail.com
Nikita Davydov, Postgraduate Student. E-mail: amail9496@gmail.com
Artem Nikonorov, Doctor of Technical Sciences, Professor at the Department of Supercomputers and General Informatics, Samara University, Head of the Laboratory for Video Intelligent Analysis, IPSI RAS. E-mail: artniko@gmail.com