ISSN XXXX-XXXX (Print)
2019. № 1. С. 32–40

СИИТ
СИСТЕМНАЯ ИНЖЕНЕРИЯ И ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

ISSN XXXX-XXXX (Online)
http://siit.ugatu.su

УДК 004.65

АЛГОРИТМЫ ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА ДАННЫХ БАНКОВСКИХ ТРАНЗАКЦИЙ В СОСТАВЕ СИСТЕМЫ ПРОТИВОДЕЙСТВИЯ ФИНАНСОВОМУ МОШЕННИЧЕСТВУ

А. В. Никонов[1], А. М. Вульфин[2], М. М. Гаянова[3], М. Ю. Сапожникова[4]

[1] nikonovandrey1994@gmail.com, [2] vulfin.alexey@gmail.com, [3] maya.gayanova@gmail.com, [4]sapozhnikova.maria.pro@yandex.ru

ФГБОУ ВО «Уфимский государственный авиационный технический университет» (УГАТУ)

Поступила в редакцию 15 декабря 2018 г.

Аннотация. Статья посвящена вопросам повышения эффективности систем мониторинга транзакций (СМТ). Рассмотрены основные виды СМТ с указанием их достоинств и недостатков, также были рассмотрены различные алгоритмы интеллектуального анализа данных, применяемых в этой сфере. На основании проведенного анализа существующих систем и алгоритмов, были выбраны три классификатора: многослойный персептрон (МСП), классификатор на основе случайного леса и метод опорных векторов. Выбранные классификаторы были протестированы на натурных данных. Наилучшие показатели были отмечены для классификатора на основе случайного леса. Перспективными являются СМТ, использующие при анализе модель поведения пользователя, таким образом, следующим шагом должен стать сбор статистических данных о поведении и построение модели пользователя.

Ключевые слова: интеллектуальный анализ данных; многослойный персептрон; случайный лес; метод опорных векторов.

ВВЕДЕНИЕ

Активное развитие интернет-банкинга в последние годы привело к существенному росту количества финансовых киберпреступлений в этой сфере. По данным Центрального банка РФ на 2014 год доля мошеннических операций в интернет-банкинге составила 63%, а за последние 2 года – выросла в 5,5 раз и составила 93% всех преступлений, связанных с хищением средств со счетов держателей карт [1]. Таким образом, построение, развитие и совершенствование систем анализа и мониторинга банковских транзакций является одной из важнейших задач борьбы с киберпреступлениями в финансовой сфере.

АНАЛИЗ СИСТЕМ МОНИТОРИНГА БАНКОВСКИХ ТРАНЗАКЦИЙ

На сегодняшний день существует достаточно большое количество разнообразных по своей структуре, используемым алгоритмам и прочим архитектурным параметрам, систем мониторинга транзакций (СМТ). Все системы можно условно классифицировать по следующей схеме (рис. 1) [2].

Оперативность реагирования в таких системах – это время, за которое система анализирует транзакцию и принимает решение о ее дальнейшей обработке в случае, если относит ее к мошенническим транзакциям. Системы реального времени анализируют транзакции в режиме онлайн и могут влиять на результат принятия транзакции. Системы псевдо-реального времени также проводят анализ транзакций в режиме онлайн, но решение принимается только после завершения подозрительной транзакции. В системах отложенного режима анализ осуществляется периодических (ежедневно, еженедельно и т.д.), в результате формируются отчеты, на основе которых принимаются решения. По мнению экспертов, наиболее эффективны системы реального времени [3, 4], но и алгоритмическая база подобных систем наиболее сложна.

Работа поддержана грантом РФФИ 19-07-00780.

Рис. 1. Классификация систем мониторинг транзакций

По способу взаимодействия с процессинговым центром СМТ делятся на интегрированные системы и SaaS (Software as a Service «программное обеспечение как сервис»). «ПО как сервис» позволяет сократить затраты на установку ПО на каждый компьютер и покупку дополнительного оборудования.

По типу принятия решений СМТ делятся на автоматические, в которых решение по транзакции принимается без участия человека, и автоматизированные, в которых система предоставляет уполномоченному сотруднику информацию для принятия решения. Автоматические системы обладают преимуществом в виду отсутствия значительных задержек при реагировании на подозрительные транзакции, но, в то же время, отсутствие оператора не всегда позволяет системе корректно принять решение в нестандартной ситуации.

Еще одна из важнейших характеристик системы – данные, используемые при анализе и принятии решения о типе транзакции. Самые простые и менее эффективные системы используют для анализа только параметры самой транзакции [6]. Более продвинутые системы учитывают не только параметры текущей транзакции, но и информацию по прошедшим операциям данной карты [7]. Наиболее перспективные системы используют модели поведения держателей карт, т.е., помимо данных о транзакциях, анализируется информация об устройствах, с которых пользователь

инициирует транзакции, и их программно-аппаратные характеристики [3, 8].

Системы, использующие в качестве основы для принятия решений только лишь сигнатурный подход, отличаются простотой реализации, но не способны выявить сложные, неявные закономерности, присущие современной сфере интернет-банкинга. Более сложные системы, использующие статические методы, такие как методы описательной статистики, корреляционного анализа, регрессионного анализа, также не всегда способны предоставить достаточный объем информации для принятия решения по результатам анализа истории транзакций клиента, но эти методы могут применяться на этапе предварительного анализа истории транзакций, генерации и селекции признаков [9].

СМТ на основе методов интеллектуального анализа данных хорошо зарекомендовали себя в сфере обработки экономических и банковских данных: различные реализации нейронные сетей [10, 11], метод опорных векторов (support vector machine – SVM) [11, 12], скрытые марковские модели (hidden markov model – HMM) [13–15], генетические алгоритмы и эволюционное программирование [3, 4, 10, 11], деревья

решений (decision trees) [11, 16], нечеткая логика (fuzzy logic) [11].

Однако в чистом виде эти алгоритмы уже не способны решить существующие задачи в виду возрастающих объемов обрабатываемых данных. Возникает потребность модифицировать эти алгоритмы, а также комбинировать их для получения приемлемого результата. Технологии построения ансамблей классификаторов (бэггинг (bagging) [3, 17]), комитетов слабых классификаторов (бустинг (boosting) [17, 18]) и композиции гетерогенных классификаторов (стэкинг (stacking) [3, 17]) позволяют достичь высоких показателей в задаче анализа банковских данных. Кроме того, в сфере интернет-банкинга актуальным становится применение технологий больших данных (Bid Data) [19, 20], позволяющие обрабатывать огромные объемы накапливаемой информации.

Таким образом, цель исследования – повышение эффективности системы мониторинга транзакций и обнаружения кибермошенничества путем совершенствования алгоритмов анализа данных банковских транзакций. Для достижения поставленной цели были сформулированы следующие задачи:

- Анализ существующих систем мониторинга кибермошенничества;
- Выбор алгоритмов для анализа данных банковских транзакций;
- Реализация выбранных алгоритмов предобработки и анализа;
- Оценка полученных результатов на натурных данных.

ВЫЧИСЛИТЕЛЬНЫЙ ЭКСПЕРИМЕНТ

Для анализа эффективности методов интеллектуального анализа данных была использована база банковских транзакций UCSD-FICO data mining contest 2009 [21, 22]. Набор данных создан Калифорнийским университетом Сан-Диего на основе лога электронных банковских транзакций в 2009 году. Набор содержит размеченные обучающие данные и неразмеченные тестовые, в этом эксперименте использованы только размеченные данные, которые содержат информацию о более 100000 транзакций 74000 клиентов за период 98 дней. Набор содержит 20 анонимизированных полей (рис. 2), включая поле с размеченными транзакциями, 4 из которых символьные.

Amount	Hour1	State1	Zip1	CustAttr1	Field1	CustrAttr2	Field2	Hour2	Flag1	total	Field3	Field4	Indicator1	Indicator2	Flag2	Flag3	Flag4	Flag5	Class

Рис. 2. Структура набора данных до предобработки

В ходе анализа исходных данных были удалены следующие поля:

Поля custAttr1, поле номера карты клиента, и custAttr2, поле электронной почты клиента. Оба этих поля идентифицируют клиента, поэтому можно оставить только поле custAtr1;

Поля total и amount имеют одинаковые значения, поэтому остается только поле total;

Также как поля hour1 и hour2 имеют идентичные значение, остается только поле hour2;

- Из полей state1 и zip1 остается только поле state1, поскольку эти поля содержат одну и ту же информацию для всех имеющихся записей.

Таким образом, теперь набор данных содержит 16 полей, все из которых числовые (Рис. 3).

State1	CustAttr1	Field1	Field2	Hour2	Flag1	Total	Field3	Field4	Indicator	Indicator	Flag2	Flag3	Flag4	Flag5	Class

Рис. 3. Структура набора данных после предобработки

Как было сказано ранее, набор содержит 100000 транзакций, и только около 1000 из них – мошеннические транзакции. В такой ситуации алгоритмы классификации могут давать некорректные результаты, поскольку размеры классов слишком различаются. Для устранения перекоса в размере классов, число немошеннических транзакций случайным образом было сокращено до 5000.

После этих действий на сформированном наборе данных был проведен вычисли-

тельный эксперимент. Для анализа были выбраны следующие классификаторы:

Многослоный персептрон (Multilayer P ный персептрон (Multilayer Perceptron – MLP);

Классификатор на основе случайного леса (Random forest committee on Bag – RFC);

Классификатор на основе машины опорных векторов (Support vector machine – SVM).

Многослойный персептрон (МСП) – нейронная сеть прямого распространения, отличается высокой связностью, в силу чего этот классификатор обладает высоким потенциалом в построении сложной разделяющей гиперповерхности. МСП, используемый в работе, состоит из четырех слоев: один входной слой, два скрытых слоя по 12 и 10 нейронов соответственно и один выходной слой. Функция активации нейронов скрытого слоя – гиперболический тангенс, выходного слоя – линейная функция.

Бэггинг (Bag) – один из самых простых видов построения ансамбля классификаторов. При этом методе k раз выбирается случайно часть экземпляров выборки размером N (в данной реализации – k=10). При этом часть значений набора данных может не попасть в выборку для обучения классификатора, а некоторые могут попасть несколько раз, что и обуславливает эффективность метода, поскольку «объекты-выбросы» (аномальные объекты, находящиеся на границе класса и расположенные достаточно далеко от центра и основного множества объектов, составляющих ядро класса) могут не попасть в обучающие подвыборки. Для каждой подвыборки происходит построения классификатора, в данном случае, построение дерева решений. Итоговый классификатор будет усреднять значения всех классификаторов, уменьшая таким образом дисперсию обучаемого классификатора.

Метод опорных векторов (SVM) строит оптимальную разделяющую гиперплоскость, разделяющую два или несколько классов. В данном случае разделение происходит на два класса – бинарная классификация – «one-vs-all». Гиперплоскость строится таким образом, что расстояние от нее до каждого класса максимально. Точность классификации зависит прежде всего от выбора ядра классификатора – классифицирующей функции. В данном случае использована радиально-базисная функция:

$$k(x, x') = e^{-\gamma\|x-x'\|^2}, \gamma > 0. \quad (1)$$

Все рассмотренные классификаторы и алгоритмы тестирования реализованы в среде Matlab. Для всех классификаторов выполнена одиночная проверка на обучающем и тестовом множествах и затем, для оценки обобщающей способности, 10-кратная перекрестная проверка (k-fold cross validation) [23]. В такой схеме оценки исходный набор данных разбивается на K (в данном случае 10) одинаковых по размеру блоков. Из всех блоков один блок используется для тестирования модели, а остальные в качестве тренировочных наборов. Этот процесс повторяется K раз.

Для оценки результатов классификации, как правило, применяются метрики: точность (precision), полнота (recall), F1-мера и коэффициент корреляции Мэтьюса (MCC), которые основываются на основных показателях классификации [6, 21]:

- Истинно положительные (TP). Мошеннические операции, классифицированные, как мошеннические.
- Истинно отрицательные (TN). Добросовестные операции, классифицированные, как добросовестные.
- Ложно положительные (FP). Добросовестные операции, классифицированные, как мошеннические.
- Ложно отрицательные (FN). Мошеннические операции, классифицированные, как добросовестные.

Эти значения можно получить из матрицы неточностей M, которая представляется собой в случае разбиения на два класса следующую матрицу:

$$M = \begin{pmatrix} TP & FP \\ FN & TN \end{pmatrix}.$$

Исходя из этих показателей точность (precision) рассчитывается как

$$Precision = \frac{TP}{TP + FP}, \quad (2)$$

полнота (recall)

$$Recall = \frac{TP}{TP + FN}, \quad (3)$$

F1-мера

$$f1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall}, \quad (4)$$

и коэффициент корреляции Мэтьюса

$$MCC = \frac{(TP * TN) - (FP * FN)}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}. \quad (5)$$

Кроме этих параметров были использованы другие характеристики: чувствительность (Sensitivity), специфичность (Specificity), корректность классификации (correctRate), число ошибочно распознанных транзакций (numError), а также разброс значения корректности классификации (scatter).

Метрика Sensitivity вычисляется также, как и полнота (Recall). Другая схожая метрика Specificity вычисляется на основе значений матрицы неточности:

$$Spesificity = \frac{TN}{TN + FP}. \quad (6)$$

Число ошибочно распознанных транзакций (numError) – это сумма элементов побочной диагонали:

$$numError = FN + FP. \quad (7)$$

Корректность классификации (correctRate) – среднее значение коэффициента корректной классификации образов для всех запусков перекрестной проверки. Разброс значений корректности классификации (scatter) – разброс значений для всех запусков перекрестной проверки.

Все описанные характеристики использованы для оценки результатов работ алгоритмов. Все значения сведены в следующие таблицы (табл. 1 – 4).

По завершению перекрестной проверки рассчитываются итоговые значения матрицы неточностей:

M=

		Номер класса	
		1	2
Номер класса	1	4879/98 %	204/17 %
	2	121/2 %	992/83 %

на основе этих данных рассчитывается correctRate = 95,1 и scatter = ±0,004.

Табл. 1. Результат работы МСП с 10-кратной перекрестной проверкой

k	**Sensitivity Чувствительность**	**Specificity Специфичность**	**Precision Точность**	**Recall Полнота**	**F1-мера**	**MCC Коэффициент корреляции Мэтьюса**
1	0,997	0,899	0,976	0,997	0,986	0,980
2	0,998	0,905	0,977	0,998	0,988	0,937
3	0,997	0,898	0,976	0,997	0,986	0,928
4	0,999	0,917	0,980	0,999	0,989	0,945
5	0,998	0,919	0,981	0,998	0,989	0,94
6	0,998	0,926	0,982	0,998	0,990	0,950
7	0,998	0,894	0,975	0,998	0,986	0,930
8	0,998	0,909	0,978	0,998	0,988	0,940
9	0,998	0,885	0,973	0,998	0,985	0,923
10	0,996	0,9	0,976	0,996	0,986	0,928

Табл. 2. Результат работы RFC с 10-кратной перекрестной проверкой

k	Sensitivity Чувствительность	Specificity Специфичность	Precision Точность	Recall Полнота	F1-мера	MCC Коэффициент корреляции Мэтьюса
1	0,998	0,981	0,995	0,998	0,997	0,984
2	0,998	0,976	0,994	0,998	0,996	0,982
3	0,998	0,972	0,993	0,998	0,995	0,978
4	0,998	0,973	0,993	0,999	0,996	0,981
5	0,998	0,975	0,994	0,998	0,996	0,982
6	0,998	0,981	0,995	0,998	0,997	0,984
7	0,998	0,982	0,995	0,998	0,996	0,983
8	0,998	0,979	0,995	0,998	0,997	0,984
9	0,998	0,979	0,995	0,998	0,996	0,983
10	0,998	0,974	0,994	0,998	0,996	0,981

Табл. 3. Результат работы SVM с перекрестной проверкой

	correctRate	numError	Sensitivity	Specificity	Precision	Recall	F1	MCC
Значение	0,935	398	0,995	0,684	0,929	0,995	0,961	0,784

Табл. 4. Результат работ алгоритмов при одиночном запуске

	correctRate	errorRate	Sensitivity	Specificity
МСП	0,9785	0,0215	0,9974	0,8997
SVM	0,5239	0,4761	0,5704	0,3294
RF	0,995	0,005	0,998	0,9808

Аналогично МСП в конце работы определяет итоговые значения матрицы неточностей

M=

		Номер класса	
		1	2
Номер класса	1	4979/99 %	142/12 %
	2	21/1 %	1054/88 %

на основе значений которой рассчитывается correctRate = 97,4 и scatter = ±0,001.

Для метода опорных векторов приведены только итоговое значение матрицы неточностей ввиду особенностей реализации.

Итоговые значения матрицы неточностей:

M=

		Номер класса	
		1	2
Номер класса	1	4979/99 %	377/32 %
	2	21/1 %	819/68 %

Табл. 5. Сравнение результатов с работами других авторов

Название статьи	**Алгоритмы**	**MCC**
A novel credit card fraud detection model based on frequent itemset mining [21]	Алгоритм Apriori	0,78
	Метод опорных векторов (SVM)	0,39
	Метод k-ближайших соседей (k-nn)	0,62
	Метод наивного Байеса (NB)	-0,21
	Алгоритм случайного леса (RF)	0,64
Data mining techniques for credit card detection: empirical study [22]	Метод k-ближайших соседей (k-nn)	0,47
	Деревья решений	0,41
	Метод наивного Байеса (NB)	0,51
	Метод опорных векторов (SVM)	0,39
Алгоритмы интеллектуального анализа данных банковских транзакций в составе системы противодействия финансовому мошенничеству	МСП с 10-кратной перекрестной проверкой	0,82
	RF с 10-кратной перекрестной проверкой	**0,91**
	Метод опорных векторов (SVM) с 10-кратной перекрестной проверкой	0,78

ОБСУЖДЕНИЕ РЕЗУЛЬТАТОВ

Сравним полученные результаты с работами других авторов, которые также использовали выбранный набор данных [21, 22].

Предобработка данных в этих статьях осуществлялась схожим образом. Сравнение результатов выполняется по характеристике MCC (табл. 5).

Характеристики рассмотренных классификаторов сравнимы и в ряде случаев превосходят значения классификаторов, реализованных авторами статей [21, 22].

Это объясняется тем, что для серии экспериментов устранен перекос в размерности классов. Классификатор на основе случайного леса показал наилучшие результаты среди рассмотренных классификаторов.
Он позволяет получить меньшее количество ложно отрицательных и ложно положительных ошибок.

К достоинствам данного алгоритма можно отнести простоту используемой модели и эффективность параллельной реализации вычислительной схемы [17].

ЗАКЛЮЧЕНИЕ

Основной проблемой повышения эффективности СМТ является недостаточный объем фиксируемых параметров, передаваемых с клиентской стороны онлайн-банкинга в процессинговый центр, и несовершенство методов и алгоритмов сигнатурного анализа в силу низких возможностей по адаптации и гибкой настройке.

В данной работе был использован ограниченный по объему набор данных банковских транзакций. Для повышения практической значимости исследования необходимо увеличивать базу данных банковских транзакций, и оценивать эффективность систем обнаружения мошеннических транзакций на реальных данных.

В работе проанализированы возможности классификаторов различного типа: многослойный персептрон, классификатор на основе случайного леса, классификатор на основе машины опорных векторов. Установлено, что наилучшие результаты по совокупности критериев (сложность программно-аппаратной реализации, корректность классификации, оценки ошибок первого и второго рода, F1-мере) продемонстрировал классификатор на основе случайного леса.

СМТ, использующие при анализе модель поведения пользователя, гораздо эффективнее других систем, таким образом, следующим шагом должен стать сбор статистических данных для построения адаптивной модели поведения пользователя.

Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта № 19-07-00780

СПИСОК ЛИТЕРАТУРЫ

1. **Евдокимов К. Н.** Структура и состояние компьютерной преступности Российской Федерации // Юридическая наука и правоохранительная практика. 2016. Т 1. № 35. С. 86–94. [K. N. Evdokimov “Structure and state of computer crime in the Russian Federation”, (in Russian), *in Yuridicheskaya nauka i pravookhranitel'naya praktika*, vol. 1, no. 35, pp. 86-94, 2016.]

2. **Бизнес** энциклопедия: платежные карты /И. М. Голдовский и др. Москва: ЦИПСиР, 2014. 560 с. **[**I. M. Goldovskiy et al. “Business encyclopedia: payment cards”, Moscow (in Russian) / *TsIPSiR*, 2014, P. 560.**]**

3. **Huang R., Tawfik H., Nagar A. K.** A novel Hybrid Artificial Immune Inspired Approach for Online Break-in Fraud Detection // Procedia Computer Science. 2012. Vol. 1. pp. 2733–2742. **[**R. Huang, H. Tawfik, A. K. Nagar “A novel Hybrid Artificial Immune Inspired Approach for Online Break-in Fraud Detection”, *in Procedia Computer Science*, vol. 1, pp. 2733-2742, 2012.**]**

4. **Schaidnagel M., Petrov I., Laux F.** DNA: An Online Algorithm for Credit Card Fraud Detection for Games Merchants // Data analytics 2013: The Second International Conference on Data Analytics, 2013. pp. 1–6. **[**M. Schaidnagel, I. Petrov, F. Laux “DNA: An Online Algorithm for Credit Card Fraud Detection for Games Merchants”, 2013, pp. 1-6.**]**

5. **Patil S., Somavanshi H., Gaikward J.** Credit Card Fraud Detection Using Decision Tree Induction Algorithm // International Journal of Computer Science and Mobile Computing. 2015. Vol.4. pp. 92–95. **[**S. Patil, H. Somavanshi, J. Gaikward “Credit Card Fraud Detection Using Decision Tree Induction Algorithm*”, in International Journal of Computer Science and Mobile Computing*, vol. 4, pp. 92-55, 2015.**]**

6. **Delamaire L., Abdou H., Pointon J.** Credit card fraud and detection techniques: a review // Bank and Bank Systems. 2009. Vol. 4. pp. 56–68. **[**L. Delamaire, H. Abdou, J. Pointon “Credit card fraud and detection techniques: a review*”, in Bank and Bank Systems,* vol. 4, pp. 56-68, 2009.**]**

7. **Detecting** Credit Card Fraud using Periodic Features / A. C. Bahnsen, et al // Computer Science. 2015. №. 3. pp. 37–43. **[**A. C. Bahnsen, et al., “Detecting Credit Card Fraud using Periodic Features”, *in Computer Science*, no. 3, pp. 37-43, 2015.**]**

8. **A Novel** Approach for Automated Credit Card Transaction Fraud Detection using Network-Based Extensions V.V. Vlasselaer et al // Decision Support Systems. 2015. p. 38–48. **[**V. V. Vlasselaer, C. Bravo, O. Caelen, L. Akoglu “A Novel Approach for Automated Credit Card Transaction Fraud Detection using Network-Based Extensions”, *in Decision Support Systems*, p. 38-48, 2015.**]**

9. **Турков П. А., Красоткина О. В., Моттль В. В.** Отбор признаков в задаче классификации при смещении решающего правила // Известия Тульского государственного университета: Естественные науки. 2015. № 4. С. 67–78. **[**A. Turkov, O. V. Krasotkina, V. V. Mottl' “Selection of signs in the classification problem when the decision rule is shifted ”, (in Russian), *in Izvestiya Tul'skogo gosudarstvennogo universiteta: Estestvennye nauki*, no. 4, pp. 67-68, 2015.**]**

10. **Patidar R., Sharma L.** Credit Card Fraud Detection Using Neural Network // International Journal of Soft Computing and Engineering. 2011. Vol.1. pp. 32–38. **[**R. Patidar, L. Sharma “Credit Card Fraud Detection Using Neural Network”, *in International Journal of Soft Computing and Engineering*, vol. 1, pp. 32-38, 2011.**]**

11. **West J., Bhattacharya M.** Some Experimental Issues in Financial Fraud Mining // Procedia Computer Science. 2016. Vol. 80. pp. 1734–1744. **[**J. West, M. Bhattacharya “Some Experimental Issues in Financial Fraud Mining”, *in Procedia Computer Science*, vol. 80, pp. 1734-1744, 2016.**]**

12. **Patel S., Gond S.** Supervised Machine (SVM) Learning for Credit Card Fraud Detection // International Journal of Engineering Trends and Technology. 2014. Vol. 8. pp. 137–139. **[**S. Patel, S. Gond “Supervised Machine (SVM) Learning for Credit Card Fraud Detection”, *in International Journal of Engineering Trends and Technology*, vol. 8, pp. 137-139, 2014.**]**

13. **Bhusari V., Patil S.** International Journal of Engineering Trends and Technology // International Journal of Distributed and Parallel Systems. 2011. Vol 2. No.6. pp. 203–211. **[**V. Bhusari, S. Patil “International Journal of Engineering Trends and Technology”, *in International Journal of Distributed and Parallel Systems*, vol. 2 no. 6, pp. 203-211, 2011.**]**

14. **Prakash A. Chandrasekar C.** An Optimized Multiple Semi-Hidden Markov Model for Credit Card Fraud Detection // Indian Journal of Science and Technology. 2015. Vol. 8. No. 2. pp. 164–171. **[**A. Prakash, C. Chandrasekar “An Optimized Multiple Semi-Hidden Markov Model for Credit Card Fraud Detection”, *in Indian Journal of Science and Technology*, vol. 8, no. 2, pp. 163-171, 2015.**]**

15. **Matheswaran P., Siva E., Rajesh R.** Fraud Detection in Credit Card Using Data Mining Techniques // International Journal of Distributed and Parallel Systems. 2015. Vol. 2. P. 11–18. **[**P. Matheswarn, E. Siva, R. Rajesh “Fraud Detection in Credit Card Using Data Mining Techniques*”, in International Journal of Distributed and Parallel Systems*, vol. 2, pp. 11-18, 2015.**]**

16. **Patil S., Somavanshi H., Gaikward J., Deshmane A.** Credit Card Fraud Detection Using Decision Tree Induction Algorithm // International Journal of Computer Science and Mobile Computing. 2015. Vol. 4. pp. 92–95. **[**S. Patil, H. Somavanshi, J. Gaikward, A. Deshmane “Credit Card Fraud Detection Using Decision Tree Induction Algorithm”, *in International Journal of Computer Science and Mobile Computing*, vol. 4, pp. 92-95, 2015.**]**

17. **Чистяков С. П.** Случайные леса: обзор // Труды Карельского научного центра РАН. 2013. № 1. С. 117–136. **[**S. P. Chistyakov “Random forests: a review”, (in Russian), *in Trudy Karel'skogo nauchnogo tsentra RAN,* no. 1, pp. 117-136, 2013.**]**

18. **Salvatore J., Fan W., Lee W.** Cost-based Modeling for Fraud and Intrusion Detection: Results from the JAM Project // International Journal of Computer Science and Mobile Computing. 2015. Vol. 1. Pp. 1-15. **[**J. Salvatore, W. Fan, W. Lee “Cost-based Modeling for Fraud and Intrusion Detection: Results

from the JAM Project", *in International Journal of Computer Science and Mobile Computing*, vol. 1. pp. 1-15, 2015.**]**

19. **McDonald C.** Real time credit card fraud detection with Apache Spark and event streaming. [Электронный ресурс] URL: https: // mapr.com/blog/real-time-credit-card-fraud-detection-apache-spark-and-event-streaming / (дата обращения 14.12.2018). **[**C. McDonald (2018, Dec. 14) *Real time credit card fraud detection with Apache Spark and event streaming* [Online]. Available: https: // mapr.com/blog/real-time-credit-card-fraud-detection-apache-spark-and-event-streaming/**]**

20. **Ghosh P.** Real time fraud detection with sequence mining **[**Электронный ресурс**]** URL: https://pkghosh.wordpress.com/2013/10/21/real-time-fraud-detection-with-sequence-mining/ (дата обращения 14.12.2018). **[**P. Ghosh (2018, Dec. 14) *Real time fraud detection with sequence mining* [Online]. Available: https://pkghosh.wordpress.com/2013/10/21/real-time-fraud-detection-with-sequence-mining/**]**

21. **Seeja K. R.** FraudMiner: A Novel Credit Card Fraud Detection Model Based on Frequent Itemset Mining. // The Scientific World Journal. 2014. Vol. 1. pp. 1–10. **[**K. R. Seeja "FraudMiner: A Novel Credit Card Fraud Detection Model Based on Frequent Itemset Mining", *in The Scientific World Journal*, vol. 1, pp. 1-10, 2014.**]**

22. **Fahmi M., Hamdy A., Nagati K.** Data Mining Techniques for Credit Card Fraud Detection: Empirical Study. // Sustainable Vital Technologies in Engineering & Informatics. 2016. pp.1–9. **[**M. Fahmi, A. Hamdy, K. Nagati "Data Mining Techniques for Credit Card Fraud Detection: Empirical Study", *in Sustainable Vital Technologies in Engineering & Informatics*, pp. 1-9, 2016.**]**

23. **Watkins C. J. C.** Combining cross-validation and search. // Progress in Machine Learning-Proceedings of EWSL: 10nd European Working Session on Learning. 2009. pp. 79–90. **[**C. J. S. Watkins "Combining cross-validation and search", *in Progress in Machine Learning-Proceedings of EWSL: 10nd European Working Session on Learning,* pp. 79-90, 2009.**]**

ОБ АВТОРАХ

НИКОНОВ Андрей Владимирович, асп. каф. вычислительной техники и защиты информации. Дипл. магистра по направлению Программная инженерия (УГАТУ, 2018).

ВУЛЬФИН Алексей Михайлович, доцент каф. вычислительной техники и защиты информации. Дипл. инженер-программист (УГНТУ, 2008). Канд. техн. наук по сист. анализу и управлению (УГАТУ, 2012). Иссл. в обл. технологий интеллектуального анализа данных.

ГАЯНОВА Майя Марсовна, Дипл. математик (Башкирский гос. ун-т, 1997). Канд. техн. наук по упр. в соц. и экон. системах (УГАТУ, 2006). Иссл. в обл. моделей и методов искусственного интеллекта в сложных системах обработки информации и управления.

САПОЖНИКОВА Мария Юрьевна, Дипл. магистра по направлению Программная инженерия (УГАТУ, 2018).

METADATA

Title: Data mining algorithms of bank transactions data as a part anti-fraud system.

Authors: A. V. Nikonov[1], A. M. Vulfin[2], M. M. Gaynova[3], M. Y. Sapozhnikova[4]

Affiliation:
Ufa State Aviation Technical University (UGATU), Russia.

Email:[1]nikonovandrey1994@gmail.com, [2]vulfin.alexey@gmail.com,[3]maya.gayanova@gmail.com, [4]sapozhnikova.maria.pro@yandex.ru

Language: Russian.

Source: SIIT, no. 1, pp. 32-40, 2019. ISSN XXXX-XXXX (Online), ISSN XXXX-XXXX (Print).

Abstract: The article is devoted to improving the effectiveness of transaction monitoring systems (TMS). The main types of TMS were considered with their advantages and disadvantages, various data mining algorithms used in this area were also considered. Based on the analysis of existing systems and algorithms, three classifiers were chosen: a multilayer perceptron (MLP), a classifier based on a random forest and method of the support vector machine. Selected classifiers were tested on full-scale data. The best result were noted for a classifier based on a random forest. Promising are TMS that use a model of user behavior in the analysis, so the next step should be to collect statistics on behavior and build a user model.

Key words: data mining; multilayer perceptron; random forest; support vector machine.

About authors:

NIKONOV, Andrey Vladimirovich, Postgrad. (PhD) Student. Dept. of Computer Engineering and Information Security, Master of Technics & Technology (UGATU, 2018).

VULFIN, Aleksey Mikhaylovich, Associate Prof. of Computer Engineering and Information Securirty Dept. Certified software engineer (UGNTU, 2008). Candidate of Technical Sciences in Systems Analysis and Management (USATU, 2012). Research in the field of data mining technologies.

GAYNOVA, Maya Marsovna, Certified mathematician (Bashkir State University, 1997). Candidate of Technical Sciences in Management in Social and Economic Systems (USATU, 2006). Research in the field of models and methods of artificial intelligence in complex information processing systems and management.

SAPOZHNIKOVA, Maria Yurevna, Master of Technics & Technology (UGATU, 2018).