

УДК 004.8

DOI: 10.26102/2310-6018/2019.26.3.023

А.В. Росляков¹, С.В. Пальмов^{1,2}, Е.В. Глушак¹
**ИССЛЕДОВАНИЕ МЕТОДОВ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА
ДЛЯ ПОСТРОЕНИЯ КЛАССИФИКАЦИОННОЙ МОДЕЛИ
УСПЕШНОСТИ РЕАЛИЗАЦИИ ТРАНЗАКЦИИ В СЕТИ
ОБЩЕКНАЛЬНОЙ СИГНАЛИЗАЦИИ**

¹ФГБОУ ВО «Поволжский государственный университет
телекоммуникаций и информатики», Самара

²ФГБОУ ВО «Самарский государственный технический университет»,
Самара

Телекоммуникации оказывают определяющее влияние на развитие человечества. Повышение эффективности их функционирования может быть реализовано различными способами. Искусственный интеллект позволяет поспособствовать решению упомянутой задачи. Однако существует проблема выбора метода искусственного интеллекта, в наибольшей степени подходящего для решения конкретной задачи в определенной предметной области. Это обусловлено большим числом существующих инструментов, доступных в рамках искусственного интеллекта, а также значительным разнообразием ситуаций, требующих учета тех или иных ограничений и нюансов при их анализе. Авторы статьи провели эксперимент, имевший своей целью упростить решение заявленной проблемы выбора в том случае, когда необходимо построить классификационную модель, определяющую успешность реализации транзакции в сети общеканальной сигнализации ОКС №7 при предоставлении голосовых услуг и услуг СМС в сети мобильной связи с использованием перенесенных мобильных абонентских номеров. Были проанализированы возможности пяти популярных методов: дерево решений, метод опорных векторов, случайный лес, нейронная сеть и наивный байесовский классификатор. Классификационные модели, создаваемые перечисленными методами, проверялись на соответствие двум требованиям: способность формировать достоверные прогнозы относительно возможности реализации транзакции и стабильность получаемых результатов. Качество оценивалось посредством метрик F-мера, специфичность и среднеквадратическое отклонение. В эксперименте использовались реальные обезличенные статистические данные, полученные в сети крупного оператора мобильной связи. После проведения соответствующих расчетов и сравнений, было установлено, что наиболее предпочтительным представляется использование метода «дерево решений», поскольку он формирует наиболее качественные классификационные модели.

Ключевые слова: искусственный интеллект, телекоммуникации, F-мера, дерево решений, классификационная модель, ОКС №7

Введение

Телекоммуникации стали неотъемлемой частью современного общества. Без них человеческая цивилизация не сможет существовать в принципе. Следовательно, вопросы, связанные с развитием и

совершенствованием телекоммуникационных систем и сетей, имеют крайне важное значение.

Одним из способов повышения эффективности работы телекоммуникационных систем и сетей является применение искусственного интеллекта в разрезе анализа статистической информации, результаты которого используются для создания классификационных моделей. Они помогают спрогнозировать возникновение нештатных ситуаций, что позволяет принять меры по предотвращению или снижению негативного воздействия на телекоммуникационную систему или сеть.

К настоящему времени разработано большое число методов искусственного интеллекта. На практике предпочтительность использования метода для решения конкретной задачи часто определяется исходя из субъективной информации: выбирается наиболее популярный подход. Однако такой путь не учитывает структуру исследуемых данных и особенности функционирования методов при их анализе, что чревато получением результатов, обладающих низким качеством. Следовательно, существует проблема выявления метода искусственного интеллекта, соответствующего определенной задаче в рамках заданной предметной области, в частности – в сети общеканальной сигнализации, используемой при предоставлении голосовых услуг и услуг СМС в сети мобильной связи с использованием перенесенных мобильных абонентских номеров.

Материалы и методы

В статье были проанализированы следующие методы искусственного интеллекта: деревья решений [1], нейронные сети [2], опорные векторы [3], случайный лес [4] и наивный байесовский классификатор [5], поскольку принципы их работы значительно отличаются друг от друга.

Исследование указанных методов выполнялось посредством аналитического пакета Orange [6].

Использовался реальный набор данных (обучающее множество, генеральная совокупность) объемом 5000 записей. Каждая запись содержит информацию о попытке реализации транзакции в сети общеканальной сигнализации ОКС №7 при предоставлении услуги в сети мобильной связи. Исходный файл содержал более сотни атрибутов. После выполнения предобработки (удалены атрибуты-константы и атрибуты, обладавшие уникальными значениями в каждой строке), структура записи генеральной совокупности приняла следующий вид (таблица 1).

Таблица 1 – Структура записи

№	Наименование	Расшифровка наименования	Единицы измерения
1	Accept	Длительность подтверждения	мс
2	Transaction	Длительность транзакции	мс
3	DPC	Код пункта сигнализации узла назначения	–
4	Link	Название звена сигнализации между MSC и STP	–
5	Unit size	Размер сообщения	байт
6	Nb units	Число сообщений	шт.
7	SLS count	Количество пучков звеньев сигнализации	шт.
8	DR Status	Статус завершения транзакции	–
9	User error	Пользовательская ошибка	–
10	Provider cause	Ошибка на стороне сети	–
11	UDT Nb	Кол-во Unitdata-пакетов	шт.
12	Released by	Инициатор завершения транзакции	–
13	ITC	Тип вызова	–

Пояснения к Таблице 1:

- SLS (Signaling Link Set) – пучок звеньев сигнализации (SL). SL – звено сигнализации между двумя смежными пунктами сигнализации, через который происходит обмен сигнальными сообщениями ОКС №7.
- STP (Signaling Transfer Point) – узел сети ОКС №7, выполняющий функции маршрутизации сигнальных сообщений (транзитный пункт сигнализации) [7].
- MSC (Mobile Switching Center) – специализированная автоматическая телефонная станция, обеспечивающая возможность связи с коммутацией каналов, управления мобильностью и предоставления сервисов GSM для мобильных телефонов внутри зоны своего обслуживания [8].
- ITC (Тип вызова): Speech (голос) / Text (текст).

Цель работы формулируется следующим образом: выявить метод искусственного интеллекта, в наибольшей степени подходящий для создания классификационных моделей, предназначенных для формирования прогнозов относительно успешности реализации транзакции в сети общеканальной сигнализации ОКС №7 при предоставлении голосовых услуг и услуг СМС в сети мобильной связи с использованием перенесенных мобильных абонентских номеров.

Результаты и обсуждение

Для достижения поставленной цели, был проведен эксперимент, состоявший из двух частей. В первой части в качестве целевого показателя выступал «User error», во второй – «Provider cause». Часть состояла из четырех этапов. На каждом из этапов осуществлялась проверка качества работы классификационных моделей (КМ), построенных при помощи пяти, входящих в группу наиболее популярных, методов искусственного интеллекта: дерево решений, метод опорных векторов, случайный лес, нейронная сеть и наивный байесовский классификатор. Все КМ проверялись на пяти стратифицированных выборках с замещением, сгенерированных на основе генеральной совокупности. Критерии качества сформулированы следующим образом: $F1 \rightarrow 1$, $Spe \rightarrow 1$, $СКО F1 \rightarrow 0$, $СКО Spe \rightarrow 0$. Здесь $F1$ – это F -мера, Spe – специфичность [9] и $СКО$ – среднеквадратическое отклонение.

Таким образом, к КМ предъявлялись требования: а) формировать достоверные результаты (прогнозы) и б) обладать робастностью (стабильностью получаемых результатов).

Опишем подробнее каждую часть эксперимента:

Таблица 2 – Настройки эксперимента

Название	Этап 1	Этап 2	Этап 3	Этап 4
Целевой показатель	User error	User error	Provider cause	Provider cause
Атрибуты, исключенные из рассмотрения	Provider cause	Provider cause, Nb units	User error	User error, Nb units
Модификации данных	Первая часть эксперимента: дублирующиеся записи в обучающем множестве не удалены			
	Вторая часть эксперимента: дублирующиеся записи в обучающем множестве удалены			

Дадим пояснения, касающиеся информации, которая представлена в Таблице 2.

В ходе эксперимента оценивались возможности КМ формировать достоверные результаты при использовании:

1. «Оригинальных» (без модификаций) обучающего и тестового множеств (часть 1, этапы 1 и 3)

2. Модифицированных множеств; внесены следующие «искажения» в данные:

2а. При помощи MS Excel удалены дублирующиеся записи в обучающем множестве (часть 2, этапы 1 и 3).

2б. Средствами аналитической системы LogiNot выявлена сильная положительная корреляция между некоторыми атрибутами, (см. Таблицу 3) и один из них был исключен из рассмотрения (часть 1 и 2, этапы 2 и 4).

Таблица 3 – Результаты корреляционного анализа

Название атрибута 1	Название атрибута 2	Значение коэффициента корреляции Пирсона
UDT Nb	Nb units	0,99999
Nb units	Units size	0,94232

Эксперимент (включая генерацию выборок) был выполнен при помощи свободно распространяемой аналитической системы Orange.

Эксперимент проводился при следующих настройках (Таблицы 4 -7).

Таблица 4 – Параметры дерева решений

Название параметра	Значение
Строить бинарное дерево	нет
Минимальное число записей в узлах, шт.	1
Ограничить максимальный размер дерева	нет

Таблица 5 – Параметры нейронной сети

Название параметра	Значение
Число нейронов в скрытом слое, шт.	100
Активационная функция	ReLU
Тип решателя	L-BFGS-B
Штраф регуляризации Тихонова	0,0001
Максимальное число итераций, шт.	200

Таблица 6 – Параметры метода опорных векторов (МОП)

Название параметра	Значение
Тип МОП	SVM
Значение цены (cost)	1
Значение функции потерь	0,1
Тип ядра	линейный
Numerical tolerance*	0,001
Ограничить число итераций	нет

* – определяет, насколько результат должен быть близок к правильному, чтобы считаться правильным

Таблица 7 – Параметры случайного леса

Название параметра	Значение
Число деревьев, шт.	500
Ограничить число атрибутов, рассматриваемых при каждом разбиении	нет
Ограничить размер отдельных деревьев	нет
Минимальное число записей в узлах, шт.	1

Возможность вносить изменения в настройки наивного байесовского классификатора разработчиками Orange не предусмотрена.

Указанные выше настройки позволили получить наиболее качественные классификационные модели при минимальных временных затратах на их построение [10].

Наиболее эффективным методом будем считать тот, который обладает наивысшими значениями метрик $F1$ и Spe в среднем по всем частям эксперимента, а также наибольшей стабильностью, т.е. наименьшим значением СКО $F1$ и Spe .

Прежде, чем перейти к оцениванию методов, необходимо сделать общее замечание: все методы, кроме дерева решений и нейронной сети, хотя бы единожды (усреднение по пяти выборкам по определенным значениям целевых переменных), строили низкокачественные КМ (см. Таблицу 8). Низкокачественная КМ – модель, у которой $F1$ или Spe равны нулю.

Таблица 8 – Низкокачественные КМ

Метод	Число КМ с $F1 = 0$, шт.	Число КМ с $Spe = 0$, шт.
Дерево решений	0	0
Метод опорных векторов	7	3
Случайный лес	1	0
Нейронная сеть	0	0
Наивный байесовский классификатор	7	1

В Таблице 9 представлены только лучшие и худшие КМ.

Влияние удаления малозначимого атрибута на качество КМ (Часть 1 Этапы 2, 4) приведено в Таблице 10.

Влияние удаления дублирующихся записей на качество КМ (Часть 2 Этапы 1, 3) приведено в Таблице 11.

Влияние удаления на качество КМ дублирующихся записей и малозначимого атрибута (Часть 2 Этапы 2, 4) приведено в Таблице 12.

Таблица 9 – Результаты (усреднение по пяти тестовым множествам) Часть 1 Этапы 1, 3

Часть 1 Этап 1				
Метод	<i>F1</i>	<i>Spe</i>	СКО <i>F1</i>	СКО <i>Spe</i>
Дерево решений	max	max	min	min
Наивный байесовский классификатор	min	min	–	–
Нейронная сеть	–	–	max	max
Часть 1 Этап 3				
Метод	<i>F1</i>	<i>Spe</i>	СКО <i>F1</i>	СКО <i>Spe</i>
Дерево решений	max	max	min	min
Наивный байесовский классификатор	min	min	max	max

* – max – лучшее значение, min – худшее, «–» – другое

Таблица 10 – Влияние удаления малозначимого атрибута (Часть 1 Этапы 2, 4)

Часть 1 Этап 2				
Метод	<i>F1</i>	<i>Spe</i>	СКО <i>F1</i>	СКО <i>Spe</i>
Дерево решений	–	–	–	–
Метод опорных векторов	↓	↓	↓	↓
Случайный лес	↓	↑	↑	↑
Нейронная сеть	↑	↑	↑	↑
Наивный байесовский классификатор	↑	↑	↓	↑
Часть 1 Этап 4				
Метод	<i>F1</i>	<i>Spe</i>	СКО <i>F1</i>	СКО <i>Spe</i>
Дерево решений	–	–	–	–
Метод опорных векторов	↑	–	↓	↓
Случайный лес	–	↓	↓	↓
Нейронная сеть	↑	↑	↑	↑
Наивный байесовский классификатор	↓	↓	↓	↓

*«–» – без изменений, ↑ – значение увеличилось, ↓ – значение уменьшилось.

Таблица 11 – Влияние удаления дублирующихся записей (Часть 2 Этапы 1, 3)

Часть 2 Этап 1				
Метод	<i>F1</i>	<i>Spe</i>	СКО <i>F1</i>	СКО <i>Spe</i>
Дерево решений	–	–	–	–
Метод опорных векторов	–	–	–	–
Случайный лес	↓	↓	↓	↓
Нейронная сеть	↓	↑	↓	↓
Наивный байесовский классификатор	↑	↑	↑	↑
Часть 2 Этап 3				
Метод	<i>F1</i>	<i>Spe</i>	СКО <i>F1</i>	СКО <i>Spe</i>
Дерево решений	–	–	–	–
Метод опорных векторов	↑	–	↓	–
Случайный лес	↓	↓	↓	–
Нейронная сеть	↓	↑	↑	↓
Наивный байесовский классификатор	↑	↓	↑	↑

Таблица 12 – Влияние удаления дублирующихся записей и малозначимого атрибута (Часть 2 Этапы 2, 4)

Часть 2 Этап 2				
Метод	<i>F1</i>	<i>Spe</i>	СКО <i>F1</i>	СКО <i>Spe</i>
Дерево решений	–	–	–	–
Метод опорных векторов	–	–	↑	↑
Случайный лес	–	–	↑	↑
Нейронная сеть	–	–	↑	↑
Наивный байесовский классификатор	–	–	↑	↑
Часть 2 Этап 4				
Метод	<i>F1</i>	<i>Spe</i>	СКО <i>F1</i>	СКО <i>Spe</i>
Дерево решений	–	–	–	–
Метод опорных векторов	↓	–	↓	↓
Случайный лес	↑	↑	↑	↑
Нейронная сеть	↓	↑	↓	↑
Наивный байесовский классификатор	↓	↑	↑	↑

Замечание по Таблицам 10-12: увеличение или уменьшение носит сильно ограниченный характер. Изменений относительно информации, представленной в Таблице 9, нет.

Заключение

Как видно из приведенных результатов (Таблицы 8-12), дерево решений в наибольшей степени удовлетворяет принятым критериям качества. Следовательно, этот метод является предпочтительным для решения задачи повышения эффективности функционирования сети общеканальной сигнализации посредством создания классификационных моделей, предназначенных для формирования прогнозов относительно успешности реализации транзакции в сети ОКС №7 при предоставлении голосовых услуг и услуг СМС в сети мобильной связи с использованием перенесенных мобильных абонентских номеров.

ЛИТЕРАТУРА

1. Sharma, Himani & Kumar, Sunil. (2016). A Survey on Decision Tree Algorithms of Classification in Data Mining // International Journal of Science and Research. 2016. Vol. 5. No. 4. – Pp. 2094–2097.
2. Pagariya, Rani & Bartere, Mahip. Review Paper on Artificial Neural Networks // International Journal of Advanced Research in Computer Science. 2013. Vol. 6. No. 6. – Pp. 49–53.
3. Bhavsar, Himani & Panchal, Mahesh. A Review on Support Vector Machine for Data Classification // International Journal of Advanced Research in Computer Engineering & Technology. 2012. Vol. 1. No. 10. – Pp. 185–189
4. Goel, Eesha & Abhilasha, Er. Random Forest: A Review // International Journal of Advanced Research in Computer Engineering & Technology. 2017. Vol. 7. No. 1. – Pp. 251–257.
5. Kaviani1, Pouria & Dhotre, Sunita. Short Survey on Naive Bayes Algorithm // International Journal of Advance Engineering and Research Development. 2017. Vol. 4. No. 11. – Pp. 607–611.
6. Kodati, Sarangam & Vivekanandam, R. Analysis of Heart Disease using in Data Mining Tools Orange and Weka // Global Journal of Computer Science and Technology. 2018. Vol. 18. No. 1. – Pp. 17–21.
7. Росляков, А.В. Общеканальная система сигнализации №7 / А.В. Росляков. – М.: Эко-Трендз, 1999. – 176 с.
8. Росляков, А.В. ОКС № 7: архитектура, протоколы, применение / А.В. Росляков. – М.: Эко-Трендз, 2008. – 320 с.
9. Hossin, M. & Sulaiman, M.N. A Review on Evaluation Metrics for Data Classification Evaluations// International Journal of Data Mining & Knowledge Management Process. 2015. Vol. 5. No. 2. – Pp. 1–11.

10. Пальмов, С.В. Исследование применимости методов технологии Data Mining для анализа клиентской базы телекоммуникационной компании / С.В. Пальмов, А.А. Крюкова // Прикладная информатика. – М.: – 2019. – Т.14. – №1(79). – С. 17–28.

A.V. Roslyakov¹, S.V. Palmov^{1,2}, E.V. Glushak¹

**STUDY OF ARTIFICIAL INTELLIGENCE METHODS FOR
CONSTRUCTING A CLASSIFICATION MODEL OF SUCCESSFUL
TRANSACTION IMPLEMENTATION IN THE SS7 NETWORK**

¹*Povolzhskiy State University of Telecommunications and Informatics, Samara*

²*Samara State Technical University, Samara*

Telecommunications have a decisive influence on the development of human society. The type of communication can be improved in various ways. Artificial intelligence allows to contribute to the solution of the abovementioned problem. However, there is the problem of choosing the method that is most suitable for solving a specific task in a particular subject area. This is due to the large number of existing artificial intelligence tools, as well as a significant variety of situations that require consideration of certain restrictions and nuances when analyzing them. The authors of the paper conduct an experiment that aimed to simplify the solution of the stated problem when it is necessary to build a classification model that determines the success of a SS7 network transaction implementation in the provision of voice and SMS services in a mobile communication network using transferred mobile subscriber numbers. The capabilities of the five methods are analyzed: decision tree, support vector machines, random forest, neural network and naive Bayes classifier. The classification models generated by the abovementioned methods test for compliance with two requirements: the reliable predictions generation and the results stability. Models quality evaluated by three metrics: F-measure, specificity and standard deviation. The experiment used real depersonalized statistics obtained in the network of a large mobile operator. After carrying out the relevant calculations and comparisons, it was found that the decision tree method usage seems to be the most preferable, since it forms the highest quality classification models.

Keywords: artificial intelligence, telecommunications, F-measure, decision tree, classification model, SS7

REFERENCES

1. Sharma, Himani & Kumar, Sunil. (2016). A Survey on Decision Tree Algorithms of Classification in Data Mining // International Journal of Science and Research. 2016. Vol. 5. No. 4. – Pp. 2094 – 2097.
2. Pagariya, Rani & Bartere, Mahip. Review Paper on Artificial Neural Networks // International Journal of Advanced Research in Computer Science. 2013. Vol. 6. No. 6. – Pp. 49–53.
3. Bhavsar, Himani & Panchal, Mahesh. A Review on Support Vector Machine for Data Classification // International Journal of Advanced Research in Computer Engineering & Technology. 2012. Vol. 1. No. 10. – Pp. 185–189

4. Goel, Eesha & Abhilasha, Er. Random Forest: A Review // International Journal of Advanced Research in Computer Engineering & Technology. 2017. Vol. 7. No. 1. – Pp. 251–257.
5. Kaviani1, Pouria & Dhotre, Sunita. Short Survey on Naive Bayes Algorithm // International Journal of Advance Engineering and Research Development. 2017. Vol. 4. No. 11. – Pp. 607–611.
6. Kodati, Sarangam & Vivekanandam, R. Analysis of Heart Disease using in Data Mining Tools Orange and Weka // Global Journal of Computer Science and Technology. 2018. Vol. 18. No. 1. – Pp. 17–21.
7. Roslyakov, A.V. Signaling System no.7. – Russia, Ed. House «Eco-Trendz», 1999, 176 p. (In Russ.)
8. Roslyakov, A.V. SS7: architecture, protocols, application. – Russia, Ed. House «Eco-Trendz», 2008, 320 p. (In Russ.)
9. Hossin, M. & Sulaiman, M.N. A Review on Evaluation Metrics for Data Classification Evaluations// International Journal of Data Mining & Knowledge Management Process. 2015. Vol. 5. No. 2. – Pp. 1–11.
10. Kriukova, A.A. & Palmov, S.V. [Data Mining applicability study for the telecommunications company's customers analysis] // Prikladnaya informatika = Applied Informatics. 2019. Vol.14. №1(79). – Pp. 17–28. (In Russ.)