

АНАЛИЗ ПРИМЕНЕНИЯ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ ДЛЯ ЗАДАЧ ОПЕРАТОРОВ СВЯЗИ

А. М. Белозор, А. Б. Гольдштейн*

Санкт-Петербургский государственный университет телекоммуникаций
им. проф. М. А. Бонч-Бруевича,
Санкт-Петербург, 193232, Российская Федерация

*Адрес для переписки: agold@niits.ru

Аннотация

При переходе к сетям пятого поколения у операторов связи возникает все больше задач, требующих решения с применением сложных алгоритмов обработки данных, в частности – задачи прогнозирования. Машинное обучение становится все более востребованным для применения, но ввиду многообразия его алгоритмов и невозможности их универсального использования, появляется необходимость определения области применения для каждого из них. Для операторов связи одной из первоочередных задач является прогнозирование различных параметров (нагрузки на контакт-центр, оттока абонентов, показателя лояльности) в условиях лавинообразного роста количества взаимодействующих объектов, разнообразия услуг, увеличения объемов данных. Авторам доклада кажется интересным найти соответствие между задачами прогнозирования оператора связи и методами машинного обучения, которые могут быть использованы для эффективного решения этих задач.

Ключевые слова

Прогнозирование, машинное обучение, классическое обучение, OSS/BSS.

Информация о статье

УДК 004.02

Язык статьи – русский.

Поступила в редакцию 03.04.2019, принята к печати 30.12.19.

Ссылка для цитирования: Белозор А. М., Гольдштейн А. Б. Анализ применения методов машинного обучения для задач операторов связи // Информационные технологии и телекоммуникации. 2019. Том 7. № 3. С. 26–30. DOI 10.31854/2307-1303-2019-7-3-26-30.

ANALYSIS OF THE APPLICATION OF MACHINE LEARNING FOR THE TASKS OF TELECOM OPERATORS

A. Belozor, A. Goldstein*

The Bonch-Bruevich Saint-Petersburg State University of Telecommunications,
St. Petersburg, 193232, Russian Federation

*Corresponding author: agold@niits.ru

Abstract—In the transition to fifth generation networks, telecom operators are faced with more and more tasks that require solutions using complex data processing algorithms, in particular – forecasting tasks. Machine learning is becoming increasingly popular for application, but due to the variety of its algorithms and the impossibility of their universal use, there is a need to determine the area of application for each of them. For telecom operators one of the priority tasks is to predict various parameters (contact center load, subscriber churn, loyalty index) in an enormous growth of the number of interacting objects, diversity of services, increasing data volumes. The authors of the report find it interesting to find the connection between the tasks of telecoms operator and the methods of machine learning, which can be used to effectively solve these problems.

Keywords—Forecasting, Machine Learning, classical training, OSS/BSS.

Article info

Article in Russian.

Received 03.04.2019, accepted 30.12.19.

For citation: Belozor A., Goldstein A.: Analysis of the Application of Machine Learning for the Tasks of Telecom Operators // Telecom IT. 2019. Vol. 7. Iss. 3. pp. 26–30 (in Russian). DOI 10.31854/2307-1303-2019-7-3-26-30.

Очевидно, что перечислить все задачи операторов связи невозможно ввиду разнообразия их типов, сути, областей применения. Однако в условиях усложнения сети, роста числа передающих устройств, увеличения разнообразия используемых абонентами услуг, работы с большими объемами данных, все большее влияние на результативность работы оператора связи оказывает его способность прогнозировать дальнейшие события, а также строить свою деятельность на основании этих прогнозов.

Теперь задачи прогнозирования оттока абонентов, прогнозирования продаж, которые раньше воспринимались как второстепенные, становятся первоочередными для оператора. А задачи прогнозирования нагрузки на сеть или количества заявок, которые поступят в контакт-центр, не просто не теряют своей важности, но и значительно усложняются качественно. Нельзя не заметить так же, что растет важность точности и детальности прогноза: недостаточно знать, сколько абонентов уйдет из сети оператора связи, важно уметь выделить конкретных пользователей, чтобы иметь возможность воздействовать непосредственно на них. Понятно, что методы прогнозирования, которые использовались раньше (формулы Эрланга, экстраполяция величин), перестают удовлетворять возникающим требованиям, поэтому появляется потребность в новых подходах.

Поиск новых алгоритмов прогнозирования в условиях усложнения рассматриваемых систем, работы с Большими Данными и развития новых математических аппаратов – актуальная тема, которая интересна не только для применения на практике, но и в рамках научного исследования.

Математические методы редко создаются для решения конкретных инженерных задач – они достаточно универсальны. Но, тем не менее, один и тот же подход не может быть применен для разных типов прогнозирования. Понятно, что прогнозирование нагрузки на сеть и нахождение паттерна допродаж не может быть осуществлено одним и тем же способом. Поэтому достаточно важной и интересной задачей является подбор алгоритмов прогнозирования для разных потребностей операторов связи.

Одними из трендов, возникающих при существующем усложнении сети и использовании Больших Данных, являются тренды на Искусственный Интеллект, машинное обучение, нейронные сети. На данном этапе исследования было принято решение рассмотреть возможности машинного обучения.

Машинное обучение – это методики анализа данных, которые позволяют аналитической системе обучаться в ходе решения множества сходных задач, извлекать информацию из исходных данных, выявлять закономерности и принимать решения с минимальным участием человека [1]. Именно минимальное вовлечение человека и возможность эффективной обработки данных делает методы машинного обучения настолько актуальными при экспоненциальном росте объемов данных в наше время.

Одной из ветвей машинного обучения является классическое обучение, методы которого были выбраны для дальнейшего рассмотрения. Классическое обучение подходит для работы с простыми данными, характеристики которых понятны. Информация, которая хранится у оператора связи, как правило, представляет собой числа или текст (простые данные), скомпонованные в таблицы (название колонки является ее характеристикой), что полностью подходит для классического обучения. Классическое обучение, в свою очередь, разделяется на обучение с учителем и без учителя. В первом случае данные должны быть каким-либо образом предварительно размечены, а во втором – могут быть никак не категоризованы [2].

Для апробации были выбраны три метода: регрессия (обучение с учителем), кластеризация и ассоциация (без учителя).

Регрессия относится к функциям, пытающимся предсказать вещественное значение. Функции такого типа оценивают зависимую величину Y , зная независимую X .

Задача регрессии – отыскать функцию (линейную или нелинейную), описывающую связь между X и Y так, чтобы для известных значений x предсказанные ей значения y были верны. Если Y зависит от ряда параметров $x_1 \dots x_i$, говорят о многомерной регрессии [3].

Возможность определения искомых значений параметра в будущем позволяет применять регрессию в задачах, связанных с предсказанием численных значений временных рядов. Например, для прогнозирования количества заявок, поступающих в контакт-центр, количества абонентов, пришедших в сеть оператора связи или ушедших из нее.

Кластеризация – метод обучения без учителя, задача которого – классификация объектов на основе их сходства друг с другом, когда принадлежность обучающих объектов каким-либо классам не задаётся. Ее востребованность обосновывается тем, что во многих прикладных задачах измерять степень сходства объектов существенно проще, чем формировать признаковые описания [1].

Задача кластеризации – на заданном пространстве объектов X найти множество кластеров Y таких, что каждый кластер будет состоять из близких объектов, а объекты разных кластеров будут существенно различны. То есть среднее внутри кластерное расстояние стремится к минимуму, а среднее межкластерное расстояние – к максимуму. Это правило можно выразить формулами [4, 5]:

$$F_0 = \frac{\sum_{i < j} [a_i = a_j] \rho(x_i, x_j)}{\sum_{i < j} [a_i = a_j]} \rightarrow \min,$$

$$F_1 = \frac{\sum_{i < j} [a_i \neq a_j] \rho(x_i, x_j)}{\sum_{i < j} [a_i \neq a_j]} \rightarrow \max,$$

где F_0 , F_1 – средние расстояния между элементами в одном и разных кластерах соответственно,

$\rho(x_i, x_j)$ – расстояния между i -м и j -м элементами,

a_i , a_j – кластеры, в которых расположены i -й и j -й элементы соответственно.

Кластеризация позволяет разделить объекты на группы, что может быть использовано при выявлении группы абонентов, находящихся в зоне риска оттока, категоризации заявок, поступающих в контакт центр.

Ассоциация – метод, цель которого заключается в обнаружении зависимостей между атрибутами внутри крупных объемов данных [6].

Задача поиска ассоциативного правила заключается в нахождении непересекающихся наборов объектов ϕ и y таких, что совместно эти наборы встречаются часто и, если встречается ϕ , то часто встречается и y [7]:

$$P(\phi \cup y) \geq \delta,$$

$$P(\phi|y) \equiv \frac{P(\phi \cup y)}{P(\phi)} \geq \mu,$$

где $P(\phi)$ – вероятность появления события ϕ ,

$P(\phi \cup y)$ – вероятность совместного появления событий ϕ и y ,

$P(\phi|y)$ – вероятность появления события ϕ при условии возникновения y ,

δ – минимальная вероятность совместного появления событий («минимальная поддержка»),

μ – минимальная значимость правила $P(\phi|y)$.

Выявление закономерностей в последовательностях действий абонентов при помощи алгоритмов метода ассоциации позволяет использовать его для определения моделей поведения абонентов, склонных к оттоку, или паттерна подключения дополнительных услуг абонентами для увеличения допродаж.

Задачей следующего этапа исследований является проверка возможности применения обозначенных методов для ранее перечисленных задач и выбор наиболее эффективного алгоритма для каждого случая.

Литература

1. Паттерсон Дж., Гибсон А. Глубокое обучение с точки зрения практика. М.: ДМК Пресс, 2018. 418 с.
2. Бринк Х., Ричардс Дж., Феверолф М. Машинное обучение. СПб.: Питер, 2017. 336 с.
3. Воронцов К. В. Лекции по алгоритмам восстановления регрессии [Электронный ресурс] // МГУ. 2007. URL: <http://www.ccas.ru/voron/download/Regression.pdf> (дата обращения 16.03.2019).
4. Воронцов К. В. Лекции по алгоритмам кластеризации и многомерного шкалирования [Электронный ресурс] // МГУ. 2007. URL: <http://www.ccas.ru/voron/download/Clustering.pdf> (дата обращения 16.03.2019).
5. Воронцов К. В. Кластеризация и частичное обучение [Электронный ресурс] // МФТИ. 2019. URL: <http://www.machinelearning.ru/wiki/images/5/52/Voron-ML-Clustering-SSL-slides.pdf> (дата обращения 16.03.2019).
6. Жерон О. Прикладное машинное обучение с помощью Scikit-Learn и TensorFlow: концепции, инструменты и техники для создания интеллектуальных систем: пер. с англ. – СПб.: ООО «Альфа-книга», 2018. 688 с.
7. Воронцов К. В. Методы поиска ассоциативных правил (презентация) [Электронный ресурс] // МФТИ. 2019. URL: <http://www.machinelearning.ru/wiki/images/7/7c/Voron-ML-AssocRules-slides.pdf> (дата обращения 16.03.2019).

References

1. Patterson Dzh., Gibson A. Glubokoye obucheniye s tochki zreniya praktika. M.: DMK Press. 2018. 418 s.
2. Brink Kh., Richards Dzh., Feverolf M. Mashinnoye obucheniye. SPb.: Piter. 2017. 336 s.
3. Vorontsov K. V. Lektsii po algoritmam vosstanovleniya regressii [Elektronnyy resurs] // MGU. 2007. URL: <http://www.ccas.ru/voron/download/Regression.pdf> (data obrashcheniya 16.03.2019).
4. Vorontsov K. V. Lektsii po algoritmam klasterizatsii i mnogomernogo shkalirovaniya [Elektronnyy resurs] // MGU. 2007. URL: <http://www.ccas.ru/voron/download/Clustering.pdf> (data obrashcheniya 16.03.2019).
5. Vorontsov K. V. Klasterizatsiya i chasticchnoye obucheniye [Elektronnyy resurs] // MFTI. 2019. URL: <http://www.machinelearning.ru/wiki/images/5/52/Voron-ML-Clustering-SSL-slides.pdf> (data obrashcheniya 16.03.2019).
6. Zheron O. Prikladnoye mashinnoye obucheniye s pomoshchyu Scikit-Learn i TensorFlow: konseptsiy, instrumenty i tekhniki dlya sozdaniya intellektualnykh sistem: per. s angl. – SPb.: OOO «Alfa-kniga». 2018. 688 s.
7. Vorontsov K. V. Metody poiska assotsiativnykh pravil (prezentatsiya) [Elektronnyy resurs] // MFTI. 2019. URL: <http://www.machinelearning.ru/wiki/images/7/7c/Voron-ML-AssocRules-slides.pdf> (data obrashcheniya 16.03.2019).

**Белозор
Ангелина Михайловна**

– студентка, СПбГУТ, Санкт-Петербург, 193232, Российская Федерация, belozor.am@yandex.ru

**Гольдштейн
Александр Борисович**

– кандидат технических наук, доцент, СПбГУТ, Санкт-Петербург, 193232, Российская Федерация, agold@niits.ru

Belozor Angelina

– Student, SUT, St. Petersburg, 193232, Russian Federation, belozor.am@yandex.ru

Goldstein Alexander

– Candidate of Engineering Sciences, Associate Professor, SUT, St. Petersburg, 193232, agold@niits.ru