

МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ РЕКОМЕНДАТЕЛЬНОЙ СИСТЕМЫ И ОБРАБОТКА ДАННЫХ ТЕЛЕКОММУНИКАЦИОННОЙ КОМПАНИИ С ПОМОЩЬЮ МОДЕЛЕЙ МАШИННОГО ОБУЧЕНИЯ

Н.А. Андриянов, naandriyanov@fa.ru, <https://orcid.org/0000-0003-0735-7697>

М.Р. Атаходжаева, atakhodzhaeva01@gmail.com

Е.И. Бородин, eiborodin2021@edu.fa.ru

Финансовый университет при Правительстве Российской Федерации, Москва, Россия

Аннотация. Цель исследования: разработка методов моделирования данных для разработки рекомендательных алгоритмов с использованием дважды стохастических авторегрессионных моделей случайных процессов и проверка их адекватности путем применения алгоритмов машинного обучения для кластеризации пользователей в имитируемом наборе данных и прогнозирования вероятностей интереса. **Методы исследования.** В статье рассмотрены методы, используемые при построении рекомендательных систем. При этом рассмотрена задача моделирования поведения пользователей с помощью дважды стохастической модели. Данная модель предложена для генерации искусственных данных. Дважды стохастическая модель позволяет генерировать нестационарные процессы, таким образом, создает пользователей с разными вероятностными свойствами в разных группах объектов интереса. После этого искусственно созданные пользователи (и их активность) кластеризуются на основе модифицированного алгоритма K-средних. Основная модификация заключается в необходимости автоматической предварительной оценки числа кластеров, а не его выбора человеком. Далее моделируется поведение представителей каждой группы пользователей для новых событий. На основе сгенерированной информации и обучающих данных решается задача прогнозирования и ранжирования предлагаемых услуг. При этом на первом этапе использования регрессионных моделей достаточно для отнесения пользователя к группе и формирования предложений данному пользователю. **Результаты исследования.** На обучающих данных в 2 кластерах были достигнуты высокие индексы детерминации, что говорит примерно о 90 % объясненной дисперсии при использовании предложенной дважды стохастической модели. Особое внимание уделено работе современных рекомендательных систем на примере системы «Диско» от Яндекс. Кроме того, выполнена предобработка и предварительный анализ данных реального сектора, а именно: исследуются данные телекоммуникационной компании. С целью выдачи релевантных предложений по услугам связи разработана тестовая рекомендательная система. **Заключение.** Таким образом, к основным результатам работы относится математическая модель, симулирующая реакцию пользователей на различные услуги, а также модель логистической регрессии, используемая для прогнозирования вероятности заинтересованности пользователя новой услугой. На основе прогнозируемых вероятностей не составляет труда ранжирование новых предложений. Апробация на синтезированных данных показала высокую эффективность модели.

Ключевые слова: рекомендательные системы, математическое моделирование, дважды стохастическая модель, логистическая регрессия, машинное обучение

Для цитирования: Андриянов Н.А., Атаходжаева М.Р., Бородин Е.И. Математическое моделирование рекомендательной системы и обработка данных телекоммуникационной компании с помощью моделей машинного обучения // Вестник ЮУрГУ. Серия «Компьютерные технологии, управление, радиоэлектроника». 2022. Т. 22, № 2. С. 17–28. DOI: 10.14529/ctcr220202

MATHEMATICAL MODELING OF RECOMMENDER SYSTEM AND DATA PROCESSING OF A TELECOMMUNICATIONS COMPANY USING MACHINE LEARNING MODELS

N.A. Andriyanov, naandriyanov@fa.ru, <https://orcid.org/0000-0003-0735-7697>

M.R. Atakhodzhaeva, atakhodzhaeva01@gmail.com

E.I. Borodin, eiborodin2021@edu.fa.ru

Financial University under the Government of the Russian Federation, Moscow, Russia

Abstract. The purpose of the study is to develop data modeling methods for projecting recommender algorithms using doubly stochastic autoregressive models of random processes and checking their adequacy by applying machine learning algorithms to cluster users in a simulated data set and predict probabilities of interest. **Research methods.** The article discusses the methods used in the construction of recommender systems. At the same time, the problem of modeling user behavior using a doubly stochastic model is considered. This model is proposed for generating artificial data. The doubly stochastic model allows generating non-stationary processes, thus creating users with different probabilistic properties in different groups of objects of interest. After that, artificially created users (and their activity) are clustered based on a modified K-means algorithm. The main modification is the need for automatic pre-estimation of the number of clusters, and not its choice by a person. Next, the behavior of representatives of each user group for new events is modeled. Based on the generated information and training data, the problem of predicting and ranking the services offered is solved. At the same time, at the first stage, the use of regression models is sufficient to assign users to a group and form offers for this user. **Results of the study.** On the training data in 2 clusters, high determination indices were achieved, which indicates approximately 90% of the explained variance when using the proposed doubly stochastic model. Particular attention is paid to the work of modern recommender systems on the example of the Disco system developed by Yandex. In addition, pre-processing and preliminary analysis of data from the real sector was performed, namely, the data of a telecommunications company are being studied. For the purpose of issuing relevant proposals for communication services, a test recommender system has been developed. **Conclusion.** Thus, the main results of the work include a mathematical model that simulates the reaction of users to various services, as well as a logistic regression model used to predict the probability of a user's interest in a new service. Based on predicted probabilities, it is not difficult to rank new proposals. Approbation on the synthesized data showed the high efficiency of the model.

Keywords: recommender systems, mathematical modeling, doubly stochastic model, logistic regression, machine learning

For citation: Andriyanov N.A., Atakhodzhaeva M.R., Borodin E.I. Mathematical modeling of recommender system and data processing of a telecommunications company using machine learning models. *Bulletin of the South Ural State University. Ser. Computer Technologies, Automatic Control, Radio Electronics*. 2022;22(2):17–28. (In Russ.) DOI: 10.14529/ctcr220202

Введение

Рынок рекомендательных систем в Российской Федерации (РФ) безостановочно растет. Рынок искусственного интеллекта (ИИ) в целом в РФ в 2018 г. составил 189 млрд руб., а к 2024 г. предсказывают, что он увеличится до 907 млрд руб. по данным CNews [1]. Данный показатель включает в себя выручку компаний в сфере искусственного интеллекта, выручку прочих ИТ-компаний, которые разрабатывают продукты благодаря ИИ, и прирост выручки компаний из различных отраслей экономики, который был получен благодаря данной технологии.

По мнению специалистов Сбербанк, в ближайшее время востребованность рекомендательных систем и интеллектуальных систем поддержки принятия решений будет высокой в рыночном, инфраструктурном, социальном секторах, а также в секторах государственного управления и безопасности. В табл. 1 представлена оценка технологий Сбербанком [2].

Таблица 1

Востребованность технологий

Table 1

Demand for technologies

Суб-СЦТ	Рыночные сектора	Инфраструктурные сектора	Социальная сфера	Госуправление и безопасность
Компьютерное зрение	Высокая	Высокая	Средняя	Средняя
Обработка естественного языка	Высокая	Средняя	Средняя	Средняя
Распознавание и синтез речи	Высокая	Средняя	Средняя	Средняя
Рекомендательные системы и интеллектуальные системы поддержки принятия решений	Высокая	Высокая	Высокая	Высокая
Перспективные методы развития технологий ИИ	Средняя	Низкая	Низкая	Низкая
Нейростимуляция и нейросенсинг	Средняя	Низкая	Высокая	Низкая
Нейропротезирование и нейроинтерфейсы	Высокая	Низкая	Высокая	Средняя

Однако в компьютерном зрении и обработке естественного языка существует множество данных, и их сбор зачастую не составляет особого труда. В то же время для разработки рекомендательных систем требуются специфические наборы данных, получение которых, как правило, весьма проблематично. В первую очередь, это связано с тем, что сбор таких данных доступен в крупных компаниях с множеством пользователей и услуг.

Существует три основных подхода, при комбинации которых формируется пул предложений для пользователя. Рассмотрим их на примере системы «Диско» от Яндекс.

Первый связан с анализом контента и основан на выявлении связей определенного объекта с производителем, брендом, страной распространения, целевой группой, группой товаров или на выявлении иных характеристик [3]. Приводится допущение, что заинтересованность в каком-либо жанре, виде продукта или компании влияет на возможную приверженность к объектам из данных кластеров. Система «Диско» предлагает продукты, услуги или контент тех брендов, типов или форм, к которым ранее уже обращался данный пользователь. Такой подход применим к тем потребителям, которые уже имеют свой трек в системе.

Второй подход основан на анализе пользователей (в Яндексе – из Крипты), сгруппированных по возрасту, полу, местоположению или роду занятий [4]. В данной ситуации делается допущение, что схожие пользователи будут интересоваться одинаковыми объектами в сервисах. Следовательно, «Диско» предлагает клиенту то, что уже было употреблено пользователями со схожими параметрами. Метод распространен не только среди устоявшихся пользователей сервиса, но и применим для так называемого холодного старта: новые клиенты еще не оставили информацию о своих предпочтениях в сервисе. Поэтому рекомендательная система генерирует предложения для новых пользователей даже с учетом того, что они еще не известны сервису.

Последний метод связан с персональными оценками пользователей потребляемого контента, товаров или услуг [5]. Оценкой может служить лайк, факт наличия комментария или время просмотра ролика. «Диско» формирует представление о будущей реакции пользователей на контент еще до его просмотра. Такой подход учитывает не только связь пользователей между собой и связи объектов по характеристикам, но и позволяет выделить нетривиальные закономерности между разрозненными группами на основе оценок потребителей.

Известны также и гибридные алгоритмы [6]. Таким образом, все указанные подходы опираются на реальные данные. Недостатком является необходимость накопления данных для разработки качественных алгоритмов и проверки их адекватности. С другой стороны, такие системы и их ошибки не являются критическими. При этом, по сути, задача предложения рекомендации пользователю сводится к задаче ранжирования. Последняя же задача может быть сведена к про-

гнозированию вероятности интереса к продукту. В связи с этим возникает задача моделирования действий пользователей на основе некоторого вероятностного распределения. Однако важно оставить корреляционные связи между вероятностями, поскольку шумовые независимые данные не могут адекватно описывать поведение пользователей. Другими словами, мы с большой вероятностью ожидаем от пользователя похожих действий для похожих продуктов и услуг.

Для моделирования вероятностей интереса к продукту можно использовать авторегрессионные модели [7–9]. К сожалению, их недостатком является достаточно высокая «колючесть» генерируемых данных и, как правило, небольшой интервал корреляции. Еще одним классом моделей прогнозирования вероятности являются модели с кратными корнями характеристических уравнений [10, 11]. Такие модели обеспечивают более гладкие изменения в генерируемых процессах, однако не могут адекватно спрогнозировать изменение свойств в таких данных. С целью генерации неоднородных и нестационарных массивов были предложены дважды стохастические модели [12–14]. Такие модели также хорошо себя зарекомендовали в задачах аугментации данных, в том числе многомерных [15]. Однако такие модели могут генерировать данные в произвольном диапазоне. С одной стороны, можно ввести пороговые ограничения, но тогда будет отбрасываться важная информация, характеризующая разницу между генерируемыми числами за порогами. В таком случае можно использовать модель функции softmax [16] для перевода в интервал вероятностей – (0; 1).

Далее рассмотрим используемую модель, сгенерируем обучающие и тестовые данные и предложим варианты прогнозирования рекомендаций. В конце статьи будут представлены результаты предварительной обработки для реальных данных телекоммуникационной компании.

1. Моделирование активности пользователей

Пусть имеется $N = 20$ пользователей и $M = 100$ различных услуг. При этом услуги принадлежат 5 категориям, описываемым интервалами от 1 до 20, от 21 до 40 и т. д. Таким образом, требуется коррелированное поведение пользователя внутри каждой категории и неоднородность при смене категорий.

Для заполнения матрицы случайными числами, имитирующими активность пользователей, воспользуемся дважды стохастической моделью вида

$$A_{ik} = A_{i-1k}\rho_k + \sigma_{Ak}\sqrt{1-\rho_k^2}\xi_i, \quad i=1, \dots, M, \quad k=1, \dots, M/20, \quad \hat{A}_{ik} = A_{ik} + m_{Ak}, \quad (1)$$

где \hat{A} – вектор реакций пользователя на услуги; ρ_k – корреляция в k -й категории; σ_{Ak} – стандартное отклонение реакции в k -й категории; ξ_i – случайная нормальная добавка с нулевым средним и единичной дисперсией; m_{Ak} – средняя реакция пользователя в k -й категории.

Следует отметить, что выражение (1) описывает поведение одного пользователя, а система таких выражений будет описывать действие N пользователей.

Генерация такой модели для произвольно выбранных пользователей представлена на рис. 1. Для различных категорий параметры модели были заданы, как

$$\sigma_A = [1; 2; 3; 5; 0,5], \quad r_A = [0,9; 0,65; 0,92; 0,99; 0,8], \quad m_A = [0; -1; 1; -1; 2].$$

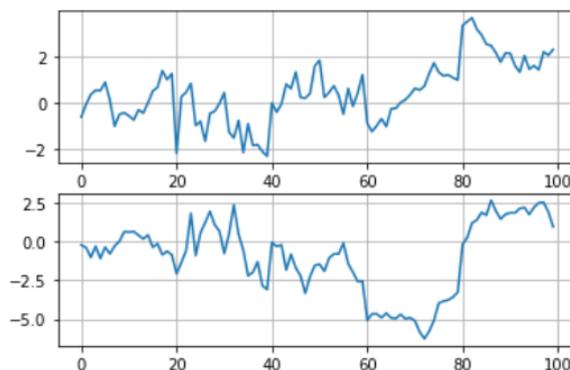


Рис. 1. Реализация дважды стохастической модели для имитации активности пользователей
Fig. 1. Implementation of a doubly stochastic model to simulate user activity

Анализ представленного рис. 1 показывает, что по форме реализаций можно сделать вывод, что сгенерированы похожие вероятности реакции пользователей для большинства категорий. Однако для услуг, например, 4-й категории значения получились сильно отличающимися. Это свидетельствует об адекватности моделирования поведения, поскольку добавляет индивидуальности пользователям.

На втором этапе для перехода к вероятностям необходимо сделать преобразование вида

$$p_i = \frac{1}{1 + \exp(-A_i - bn_i)}, \quad (2)$$

где b – произвольный коэффициент, а n_i – независимые случайные величины с нормальным законом распределения, нулевым средним и единичной дисперсией. Наличие данного слагаемого позволяет контролировать уровень «случайности поведения» при преобразовании полученных значений дважды стохастических моделей в значения вероятности. В показателе экспоненты присутствует смесь информационного сигнала с шумом, имеющим определенный вес.

На рис. 2 представлены вероятности, соответствующие реализациям из рис. 1. Реализации получены в условиях отсутствия шума, т. е. при $b = 0$.

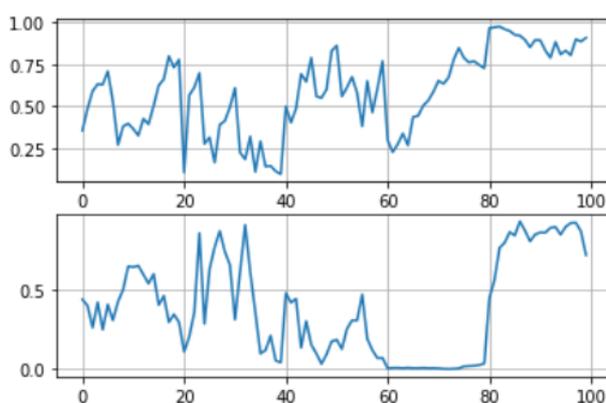


Рис. 2. Вероятности активности пользователя ($b = 0$)
Fig. 2. Probabilities of user activity ($b = 0$)

На рис. 3 представлены вероятности, соответствующие реализациям из рис. 1, при $b = 1$.

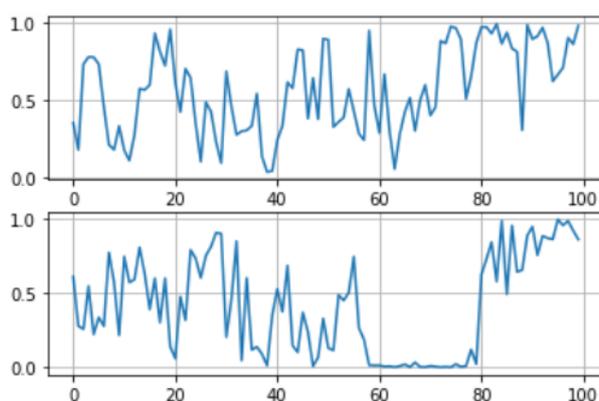


Рис. 3. Вероятности активности пользователя ($b = 1$)
Fig. 3. User activity probabilities ($b = 1$)

Видим, что реализации на рис. 3 имеют уже большую дисперсию и изменяются резко. Варьируя параметры модели, можно моделировать различные данные для заполнения матрицы обучающих систем. При этом саму покупку услуги можно моделировать, используя данную вероятность либо просто выбирая порог. Например, для вероятностей больше 0,7 на пересечении

пользователя и услуги ставится «покупка». Это позволит перейти к бинарному представлению данных.

Далее рассмотрим идею обучения модели для прогнозирования реакции на новые услуги.

2. Обучение и ранжирование новых данных

Для начала выполним кластеризацию пользователей для случая $b = 0$. Модификация алгоритма К-средних [17] подразумевает наличие предварительной стадии, на которой определяется оптимальное число кластеров.

На рис. 4 показаны результаты применения метода локтя, на рис. 5 – метода силуэта [18].

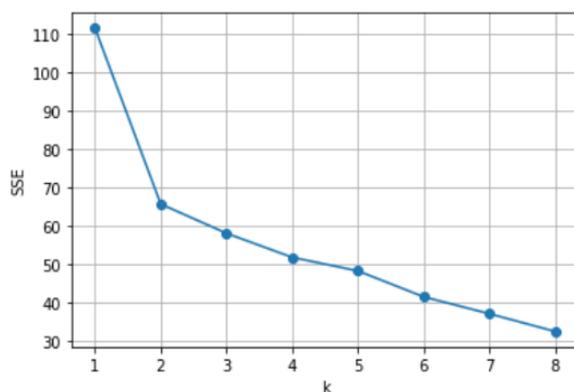


Рис. 4. Метод локтя
Fig. 4. Elbow Method

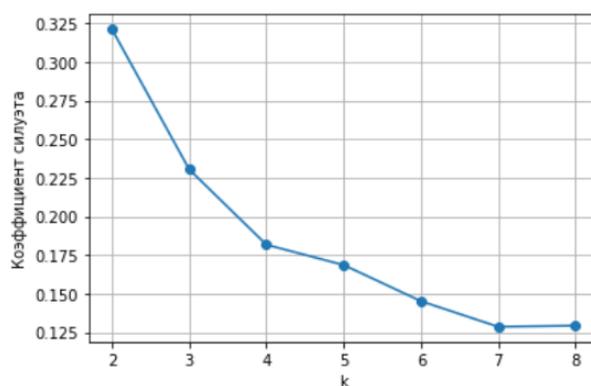


Рис. 5. Метод силуэта
Fig. 5. Silhouette method

Таким образом, наиболее предпочтительно разделение пользователей на 2 кластера. Разделение происходит в пропорции 9 на 11.

Рис. 6 демонстрирует разделение на кластеры по 1-й и 20-й услугам.

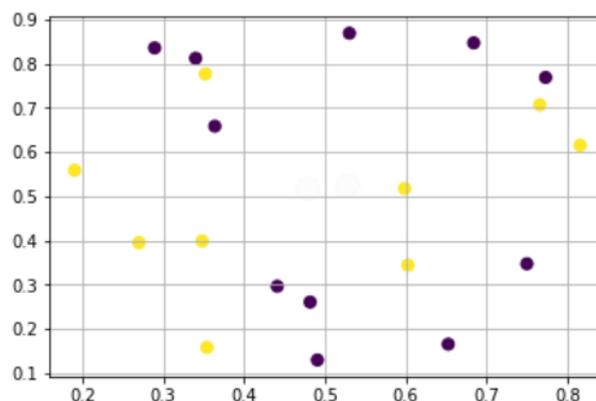


Рис. 6. Пример кластеризации в пространстве 2 признаков
Fig. 6. An example of clustering in the space of 2 features

Далее для обучения можно выбрать по 3–4 представителя каждого кластера. Сгенерировать их поведение для 5 новых услуг аналогично выражениям (1) и (2). Для прогнозов других пользователей в кластере можно построить модель логистической множественной регрессии на основе обучающего набора данных. Затем в соответствии с полученной моделью генерируется вероятность интереса к новой услуге, а сами услуги ранжируются по вероятности. Предлагаются те, которые заинтересуют пользователя с наибольшей вероятностью. Отметим, что для сгенерированных обучающих данных индексы детерминации в первом и втором кластерах составили $R_1^2 = 0,91$ и $R_2^2 = 0,86$ соответственно при $b = 0$. В табл. 2 рассмотрены оценки индексов детерминации при различных значениях b .

Таблица 2
Зависимость предсказательной способности модели
от коэффициента случайности действий пользователя

Table 2

Dependence of the predictive ability of the model
on the coefficient of randomness of user actions

b	0	1	2
R_1^2	0,91	0,73	0,52
R_2^2	0,86	0,71	0,54

Анализ данных табл. 2 показывает, что при увеличении «шума» при действиях пользователя качество модели падает.

Далее рассмотрим предобработку данных в рекомендательной системе телекоммуникационной компании.

3. Предобработка данных телекоммуникационной компании

В данном разделе выполняется предварительная обработка данных для разработки рекомендательной системы услуг для абонентов Tele2.

Первый шаг – выяснить, насколько активный пользователь похож на других пользователей. В данном случае был датасет с важной информацией об абоненте, а именно где он находится, какой у него тариф и каково среднее значение arpu (average revenue per user – «средний доход на пользователя»)

Формат исходных данных представлен на рис. 7.

```
[6] df.head()
```

	ID	TP_NAME	BSEGMENT	CORE_LIVE_SUBS	ARPU	BRANCH_NAME	MACROREGION_NAME	MINUTIES	MB	SMS	SERVICE_NAME
0	11111	Мой онлайн+	B2C	1	419.166583	Омск	Сибирь	20.0	7300.085539	0	Безлимит на ВКонтакте
1	11112	Мой разговор	B2C	1	275.000000	Волгоград	Юг	95.0	12071.448762	0	Скидка 100% на мой помощник
2	11113	Мой онлайн+	B2C	1	266.666673	Оренбург	Урал	29.0	2028.019118	0	Семейство Up-Sale
3	11114	Везде онлайн	B2C	1	416.666583	Курск	Черноземье	84.0	30398.311928	0	Безлимит на ВКонтакте
4	11115	Мой онлайн+	B2C	1	454.166677	Ульяновск	Волга	24.0	60.057621	0	Безлимит на Viber

Рис. 7. Структура исходного датасета
Fig. 7. The structure of the original dataset

Для начала нужно подготовить датасет. Выберем все возможные уникальные тарифы (рис. 8).

```
df.TP_NAME.unique()
```

```
array(['Мой онлайн+', 'Мой разговор', 'Везде онлайн', '1500 мин_30 Гб',  
'Мой онлайн', 'Новогодний', 'Прозрачные границы М', 'Мой Бизнес S',  
'200 мин_3 Гб', '200 мин_2 Гб', '250 мин_3 Гб', '500 мин_15 Гб',  
'Безлимит', '350 мин_5 Гб', 'Лайт', '350 мин_15 Гб',  
'500 мин_20 Гб', 'Социальный', '600 мин_20 Гб', '500 мин_5 Гб',  
'400 мин_5 Гб', '700 мин_30 Гб', '500 мин_25 Гб', '300 мин_12 Гб',  
'400 мин_25 Гб', '500 мин_40 Гб', 'Компаньон М', 'Супер онлайн+',  
'350 мин_12 Гб', '0 мин_2 Гб', '700 мин_5 Гб', '1000 мин_10 Гб',  
'Интернет для вещей', 'Федеральный', '200 мин_7 Гб'], dtype=object)
```

Рис. 8. Уникальные тарифы
Fig. 8. Unique tariff plans

Как можно увидеть на рис. 8, есть множество действующих тарифов. Было решено сгруппировать их в 3 группы: Старый тариф B2C, Новый тариф B2C и Тариф B2B.

Анализ данных после преобразования показывает, что преобладают новые тарифы B2C (рис. 9).

```
df.pivot_table('ID', 'TP_NAME', 'MACROREGION_NAME', 'count').plot(kind='bar', stacked=True)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f63100e26d0>
```

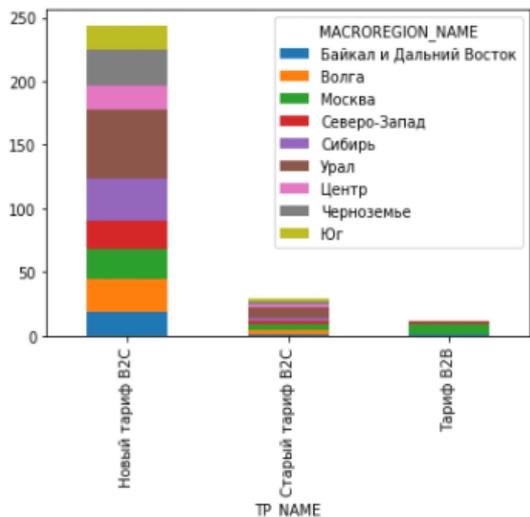


Рис. 9. Распределение тарифов
Fig. 9. Distribution of tariff plans

Посмотрим на распределение ARPU (рис. 10).

```
[44] sns.scatterplot(data=df.ARPUs)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f630ffbadd0>
```

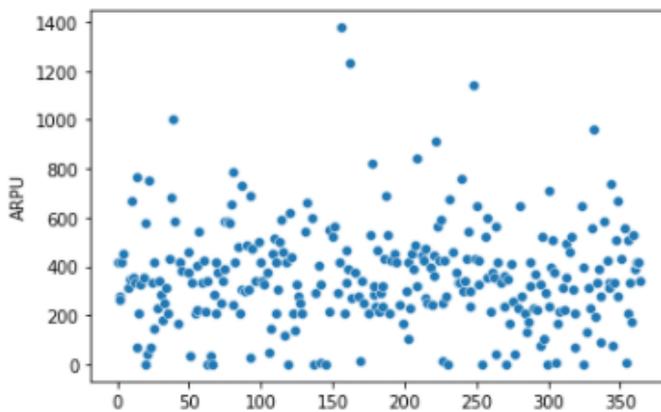


Рис. 10. Распределение значений внутри данных ARPU
(по оси X – все уникальные строки исходной таблицы)
Fig. 10. Distribution of values within ARPU data
(along the X-axis – all unique rows of the source table)

Разделим ARPU также на 3 группы: low, middle и high. Результаты представлены на рис. 11.

Практический опыт показывает, что основными параметрами для построения моделей будут являться макрорегион, тариф и ARPU. Все пользователи были сгруппированы по этим параметрам.

Итоговый набор предобработанных данных (первые 5 строк) с добавлением столбца group (объединение ключевых параметров) представлен на рис. 12.

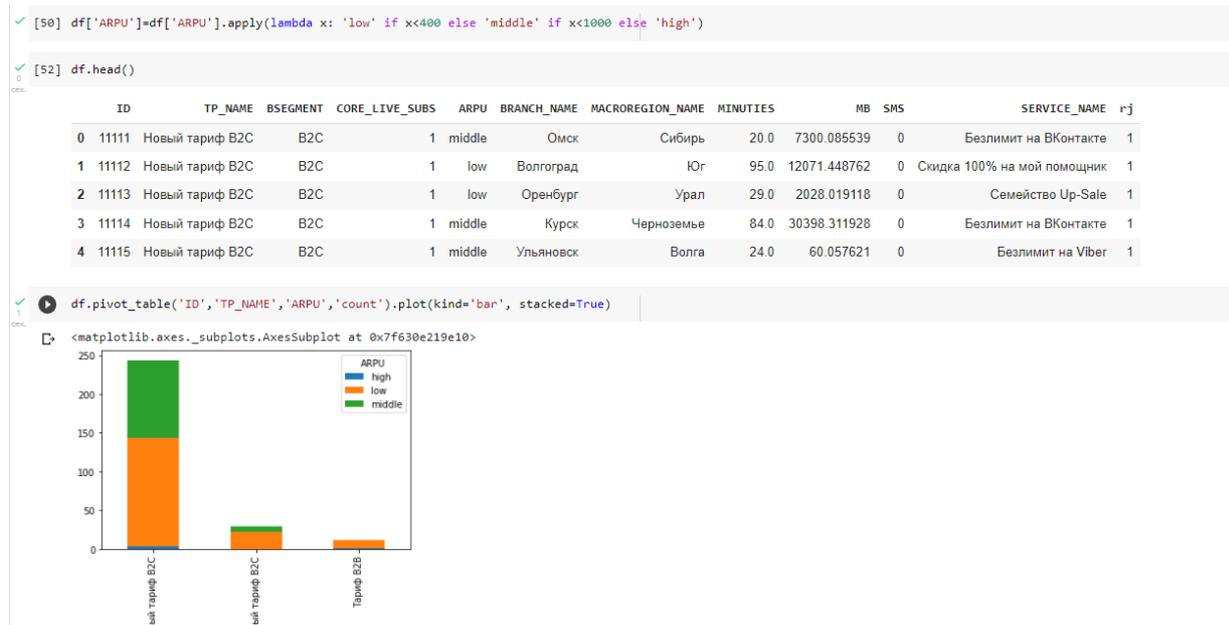


Рис. 11. Распределение по тарифам и ARPU
Fig. 11. Distribution by tariff plans and ARPU

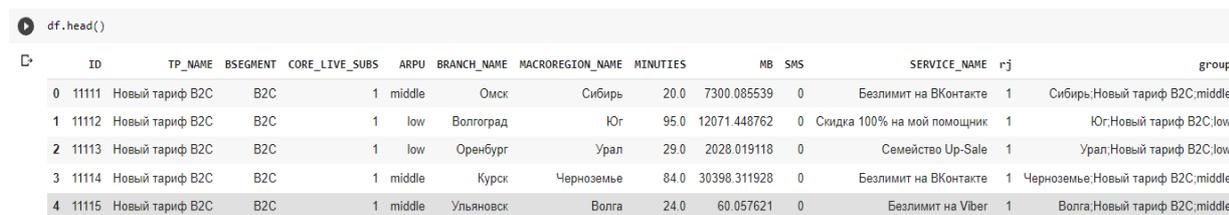


Рис. 12. Распределение по тарифам и ARPU
Fig. 12. Distribution by tariff plans and ARPU

Таким образом, абоненты Tele2 были сгруппированы по их схожим параметрам. В будущем планируется разработка рекомендательной системы на основе полученных данных.

Заключение

В статье отмечены 2 важных этапа для разработки рекомендательных систем. Первый этап связан с отсутствием данных и большого количества технологий для разработки и апробирования алгоритмов рекомендательных систем. В этом случае предложены дважды стохастические модели, позволяющие имитировать действия пользователей. На этапе преобразования предложено использовать дополнительную случайную составляющую. Однако чем больше её вес, тем хуже получается модель предсказания. На искусственных данных в отсутствие дополнительного шума в модели получены индексы детерминации порядка 90 %. Этого удалось добиться за счёт предварительной кластеризации пользователей. Второй этап связан с подготовкой данных для рекомендательных систем, когда такие данные имеются. В статье рассмотрены операции, которые позволили структурировать данные телекоммуникационной компании Tele2 по тарифным планам. В будущем планируется апробация предложенной модели на реальных данных и разработка рекомендательной системы по данным Tele2.

Список литературы

1. Будущее искусственного интеллекта в России: как технологии превратятся в решения. URL: news.ru/articles/2019-10-02_budushchee_iskusstvennogo_intellekta (дата обращения: 02.03.2022).
2. Сбербанк заработает на искусственном интеллекте 450 миллиардов рублей. URL: www.vedomosti.ru/technology/articles/2020/02/19/823464-sberbank-zarabotaet (дата обращения: 06.03.2022).

3. Авхадеев Б.Р., Воронова Л.И., Охапкина Е.П. Разработка рекомендательной системы на основе данных из профиля социальной сети «ВКонтакте» // Вестник НВГУ. 2014. № 3. URL: <https://cyberleninka.ru/article/n/razrabotka-rekomendatelnoy-sistemy-na-osnove-dannyh-iz-profilya-sotsialnoy-seti-vkontakte> (дата обращения: 08.03.2022).
4. Кластеризация профилей пользователей в рекомендательных системах поддержки жизнеобеспечения на основе реальных неявных данных / С.А. Филиппов, В.Н. Захаров, С.А. Ступников, Д.Ю. Ковалев // Труды XVIII Международной конференции DAMDID/RCDL'2016 «Аналитика и управление данными в областях с интенсивным использованием данных». 2016. С. 98–103.
5. Isinkaye F.O., Folajimi Y.O., Ojokoh B.A. Recommendation systems: Principles, methods and evaluation // Egyptian Informatics Journal. 2015. Vol. 16 (3). P. 261–273.
6. Нефедова Ю.С. Архитектура гибридной рекомендательной системы GEFEST (Generation–Expansion–Filtering–Sorting–Truncation) // Системы и средства информатики. 2012. Т. 22 (2). С. 176–196.
7. Ullrich T. On the Autoregressive Time Series Model Using Real and Complex Analysis // Forecasting. 2021. Vol. 3. P. 716–728. DOI: 10.3390/forecast3040044
8. Neural autoregressive distribution estimation / B. Uria, M.-A. Côté, K. Gregor et al. // JMLR. 2016. Vol. 17 (1). P. 7184–7220.
9. Модели систем квазипериодических процессов на основе цилиндрических и круговых изображений / В.Р. Крашенинников, Ю.Е. Кувайскова, О.Е. Маленова, А.Ю. Субботин // Известия Самарского научного центра Российской академии наук. 2021. Т. 23, № 1. С. 103–110. DOI: 10.37313/1990-5378-2021-23-1-103-110
10. Андриянов Н.А., Васильев К.К. Свойства авторегрессий с кратными корнями характеристических уравнений // Вестник УлГТУ. 2019. № 1 (85). URL: <https://cyberleninka.ru/article/n/svoystva-avtoregressiy-s-kratnymi-kornyami-harakteristicheskikh-uravneniy> (дата обращения: 08.03.2022).
11. Васильев К.К., Попов О.В. Авторегрессионные модели случайных полей с кратными корнями // Труды 4-й конференции «РОАИ: новые информационные технологии». 1998. Т. 4 (1). С. 258–260.
12. Krasheninnikov V.R., Subbotin A.Yu. Doubly stochastic model of a quasi-periodic process as an image on a cylinder // Proceedings of the International Scientific and Technical Conference “Advanced Information Technologies”. 2018. P. 1017–1021.
13. Васильев К.К., Дементьев В.Е., Андриянов Н.А. Оценивание параметров дважды стохастических случайных полей // Радиотехника. 2014. № 7. С. 103–106.
14. Vasil'ev K.K., Dement'ev V.E., Andriyanov N.A. Doubly stochastic models of images // Pattern Recognition and Image Analysis. 2015. Vol. 25 (1). P. 105–110. DOI: 10.1134/S1054661815010204
15. Demytyiev V.E., Andriyanov N.A., Vasilyev K.K. Use of Images Augmentation and Implementation of Doubly Stochastic Models for Improving Accuracy of Recognition Algorithms Based on Convolutional Neural Networks // 2020 Systems of Signal Synchronization, Generating and Processing in Telecommunications (SYNCHROINFO). 2020. P. 1–4. DOI: 10.1109/SYNCHROINFO49631.2020.9166000
16. Alzen J.L., Langdon L.S., Otero V.K. A logistic regression investigation of the relationship between the Learning Assistant model and failure rates in introductory STEM courses // IJ STEM. 2018. Vol. 5. P. 56–63. DOI: 10.1186/s40594-018-0152-1
17. Coates A., Ng A.Y. Learning Feature Representations with K-means. Stanford University Press. 2012. 318 p.
18. Севастьянова М.Д., Желябин Д.В., Андриянов Н.А. Применение прикладных методов обработки данных в задаче кластеризации многомерных данных в сфере образования // Современные проблемы проектирования, производства и эксплуатации радиотехнических систем: сб. науч. тр. 2021. С. 172–177.

References

1. The Future of Artificial Intelligence in Russia: how technologies turn into solutions. Available at: https://cnews.ru/articles/2019-10-02_budushchee_iskusstvennogo_intellekta (accessed 02.03.2022). (In Russ.)
2. Sberbank will earn 450 billion rubles on artificial intelligence. Available at: <https://www.vedomosti.ru/technology/articles/2020/02/19/823464-sberbank-zarabotaet> (accessed 06.03.2022) (In Russ.)

3. Avkhadeev B.R., Voronova L.I., Okhapkina E.P. [Development of a recommendation system based on data from the profile of the social network “VKontakte”]. *Bulletin of Nizhnevartovsk state university*. 2014, no. 3. Available at: <https://cyberleninka.ru/article/n/razrabotka-rekomendatelnoy-sistemy-na-osnove-dannyh-iz-profilya-sotsialnoy-seti-vkontakte> (accessed 03.08.2022). (In Russ.)

4. Philippov S.A., Zaharov V.N., Stupnikov S.A., Kovalev D.Yu. Clustering of user profiles based on real implicit data in e-commerce recommender systems. In: *Trudy XVIII Mezhdunarodnoy konferentsii DAMDID/RCDL'2016 “Analitika i upravleniye dannymi v oblastiakh s intensivnym ispol'zovaniyem dannykh”* [Proceedings of the XVIII International Conference DAMDID/RCDL'2016 “Analytics and data management in data-intensive areas”]; 2016. P. 98–103. (In Russ.)

5. Isinkaye F.O., Folajimi Y.O., Ojokoh B.A. Recommendation systems: Principles, methods and evaluation. *Egyptian Informatics Journal*. 2015;16(3):261–273.

6. Nefedova Yu.S. The architecture of hybrid recommender system GEFEST (Generation–Expansion–Filtering–Sorting–Truncation). *Systems and Means of Informatics*. 2012;22(2):176–196. (In Russ.)

7. Ullrich T. On the Autoregressive Time Series Model Using Real and Complex Analysis. *Forecasting*. 2021;3:716–728. DOI: 10.3390/forecast3040044

8. Uria B., Côté M.-A., Gregor K., Murray I., Larochelle H. Neural auto-regressive distribution estimation. *JMLR*. 2016;17(1):7184–7220.

9. Krasheninnikov V.R., Kuvayskova Yu.E., Malenova O.E., Subbotin A.Yu. Models of systems of quasiperiodic processes based on cylindrical and circular images. *Izvestiya Samarskogo nauchnogo tsentra Rossiyskoy akademii nauk*. 2021;23(1):103–110. (In Russ.) DOI: 10.37313/1990-5378-2021-23-1-103-110

10. Andriyanov N.A., Vasiliev K.K. The forces in the lever mechanism of the gripper by contact of the slave link with the solid. *Vestnik UIGTU*. 2019;1(85). Available at: <https://cyberleninka.ru/article/n/svoystva-avtoregressiy-s-kratnymi-kornyami-harakteristicheskikh-uravneniy> (accessed 03.08.2022). (In Russ.)

11. Vasiliev K.K., Popov O.V. [Autoregressive models of random fields with multiple roots]. In: *Trudy 4-y konferentsii “ROAI: novyye informatsionnyye tekhnologii”* [Proceedings of the 4th conference “ROAI: new information technologies”]. 1998;4(1):258–260. (In Russ.)

12. Krasheninnikov V.R., Subbotin A.Yu. Doubly stochastic model of a quasi-periodic process as an image on a cylinder. In: *Proceedings of the International Scientific and Technical Conference “Advanced Information Technologies”*; 2018. P. 1017–1021.

13. Vasiliev K.K., Dementiev V.E., Andriyanov N.A. [Estimating the parameters of doubly stochastic random fields]. *Radiotekhnika*. 2014;(7):103–106. (In Russ.)

14. Vasil'ev K.K., Dement'ev V.E., Andriyanov N.A. Doubly stochastic models of images. *Pattern Recognition and Image Analysis*. 2015;25(1):105–110. DOI: 10.1134/S1054661815010204

15. Dementyev V.E., Andriyanov N.A., Vasilyev K.K. Use of Images Augmentation and Implementation of Doubly Stochastic Models for Improving Accuracy of Recognition Algorithms Based on Convolutional Neural Networks. In: *2020 Systems of Signal Synchronization, Generating and Processing in Telecommunications (SYNCHROINFO)*; 2020. P. 1–4. DOI: 10.1109/SYNCHROINFO49631.2020.9166000

16. Alzen J.L., Langdon L.S., Otero V.K. A logistic regression investigation of the relationship between the Learning Assistant model and failure rates in introductory STEM courses. *IJ STEM*. 2018;5:56–63. DOI: 10.1186/s40594-018-0152-1

17. Coates A. and Ng A.Y. *Learning Feature Representations with K-means*. Stanford University Press; 2012. 318 p.

18. Sevast'yanova M.D., Zhelyabin D.V., Andriyanov N.A. [Application of applied data processing methods in the problem of clustering multidimensional data in the field of education]. In: *Sovremennyye problemy proyektirovaniya, proizvodstva i ekspluatatsii radiotekhnicheskikh sistem: sb. nauch. tr.* [Modern problems of design, production and operation of radio engineering systems: collection of scientific papers]; 2021. P. 172–177.

Информация об авторах

Андрянов Никита Андреевич, канд. техн. наук, доцент Департамента анализа данных и машинного обучения, Финансовый университет при Правительстве Российской Федерации, Москва, Россия; naandriyanov@fa.ru.

Атаходжаева Мадина-Бону Рустамовна, магистрант Департамента анализа данных и машинного обучения, Финансовый университет при Правительстве Российской Федерации, Москва, Россия; atakhodzhaeva01@gmail.com.

Бородин Евгений Игоревич, аспирант Департамента анализа данных и машинного обучения, Финансовый университет при Правительстве Российской Федерации, Москва, Россия; eiborodin2021@edu.fa.ru.

Information about the authors

Nikita A. Andriyanov, Cand. Sci. (Eng.), Ass. Prof. of the Department of Data Analysis and Machine Learning, Financial University under the Government of the Russian Federation, Moscow, Russia; naandriyanov@fa.ru.

Madina-Bonu R. Atakhodzhaeva, Master's student of the Department of Data Analysis and Machine Learning, Financial University under the Government of the Russian Federation, Moscow, Russia; atakhodzhaeva01@gmail.com.

Evgeny I. Borodin, Postgraduate Student of the Department of Data Analysis and Machine Learning, Financial University under the Government of the Russian Federation, Moscow, Russia; eiborodin2021@edu.fa.ru.

Статья поступила в редакцию 14.03.2022

The article was submitted 14.03.2022