

# **ВЕРОЯТНОСТНЫЙ ПОДХОД К ПРОГНОЗИРОВАНИЮ СПРОСА НА ТОВАРЫ ЭЛЕКТРОННОЙ КОММЕРЦИИ**

**И. Е. Кири**

*Московский финансово-промышленный университет «Синергия», Россия*  
E-mail: ivan-kirin@bk.ru

В статье рассмотрены два основных подхода, в основе которых лежит оценка вероятностей спроса на товары электронной коммерции: подход, основанный на детальной информации о клиенте и подход, основанный на детальной информации о товаре. Рассмотрены основные методы, модели оценки и прогнозирования спроса при использовании вероятностного подхода. Описаны возможности и особенности применения каждого подхода на практике. Выявлена ключевая разница в допустимых горизонтах прогнозирования каждого из подходов. Рассмотрены ключевые факторы, оказывающие влияние на эффективность прогноза. Проанализированы недостатки и трудности при реализации построения моделей.

## **PROBABILISTIC APPROACH TO FORECAST THE E-COMMERCE DEMAND**

**I. E. Kirin**

The article studies two main probabilistic approaches based on the estimation of the probabilities of demand for e-commerce: an approach based on detailed customer information and an approach based on detailed product information about. The main methods, estimation models and approaches of demand forecasting using the probabilistic approach are considered. The possibilities and peculiarities of the application of each approach in practice are described. A key difference in the allowable horizons for forecasting are identified. The possibilities of applying each approach are considered. The key forecasting factors are revealed. The difficulties in the model implementation are described.

За последнее десятилетие интернет изменил способы продажи и покупки товаров и услуг. Появление интернет-магазинов совершенствовало потребительский опыт клиентов. Развитие инструментов коммуникации, сетей 4G и 5G, распространение сетей Wi-Fi, технологий, позволяющих безопасно осуществлять платежи в сети, помогают рынку электронной коммерции наращивать обороты ежедневно.

Данный вид торговли характеризуется низкими транзакционными издержками, что снижает барьеры входа на рынок и позволяет, в том числе небольшим компаниям, успешно на нем конкурировать. Кроме того, данный вид торговли стирает географические границы между покупателем и продавцом и позволяет каждой стороне взаимодействовать друг с другом без временных и экономических затрат.

Сегодня крупные Consumer-to-Consumer (C2C) площадки электронной торговли перенасыщены товарами, которые простаивают месяцами и снимаются с продажи, не найдя своего покупателя (Ebay, Amazon, Авито и т.д.). Основной причиной длительного периода продажи является плохое качество пред-

ставленной информации о товаре, что не позволяет потребителю однозначно определиться с выбором [1]. Более того, среди большого количества товаров потребителю сложно найти свой товар. Поэтому оценка спроса на отдельный товар на площадке электронной коммерции сегодня выходит на первый план.

Оценка спроса на товары электронной коммерции получает сегодня все больше внимания среди зарубежных исследователей. Ввиду того, что оценка спроса на отдельный товар может принимать два исхода: осуществление покупки, либо ее отсутствие. Данная задача является стандартной задачей классификации и решается с помощью оценки вероятности продажи отдельного товара при заданных факторах. В современной литературе можно выделить два основных подхода к решению данной задачи: подход, основанный на детальной информации о товаре [2] и подход, основанный на детальной информации о клиенте [3].

В подходе, основанном на детальной информации о клиенте, используется предпосылка, согласно которой клиенты, осуществившие покупку однотипных товаров, обладают схожими интересами, предпочтениями и поведением. Задачей данного подхода является оценка вероятности продажи отдельного товара  $i$  для пользователя  $u$  на основе исторических данных о покупке товаров  $\{I\}$  потребителями  $\{U\}$ , схожими по предпочтениям с прогнозируемым пользователем.

Задача оценки вероятности покупки товара  $i$  пользователем  $u$  формулируется следующим образом [4]:

$$P(T_{ui} = C_j | S_{ui}, S_u, t) = \frac{P(S_{ui} | T_{ui} = C_j, S_u, t) P(T_{ui} = C_j | S_u, t)}{\sum_{m=1}^M P(S_{ui} | T_{ui} = C_m, S_u, t) P(T_{ui} = C_m | S_u, t)},$$

- $C_j$  – класс (1 – покупка товара, 0 отказ от покупки);
- $T_{ui}$  – исход для пользователя  $u$  при покупке товара  $i$ ;
- $S_{ui}$  – сессия пользователя  $u$  при выборе товара  $i$ ;
- $S_u$  – набор фактор пользователя  $u$ ;
- $t$  – момент времени.

Для оценки и прогнозирования вероятности в подходе, основанном на детальной информации о клиенте, чаще всего исследователи применяют дерево решений, логистическую регрессию [5], модель нейронной сети [6], случайный лес, модель Бернулли и модель Пуассона.

Ряд исследователей [7, 8] включили в свой анализ данные об активности и историю о кликах пользователя на торговых площадках, что повысило эффективность классификации на 5% по показателю F1 ( $2 * \frac{\text{precision} * \text{recall}}{\text{precision} + \text{recall}}$ ) относительно использования классических факторов. Авторы доказали, что информация о кликах может использоваться для прогнозирования предпочтений пользователей, однако, логирование данных пользователя содержит огромные массивы информации, поэтому необходимо учитывать специфику хранимой информации и извлекать только необходимую информацию. Исследователи использовали модели случайного леса и градиентный бустинг деревьев решений.

Основным недостатком подхода, основанном на анализе информации о клиенте, является сложность долгосрочного прогнозирования спроса ввиду необходимости дополнительного прогнозирования каждого из типов покупателей и их вероятности появления на электронной площадке, что значительно снижает качество прогноза.

Задачей подхода, основанном на детальной информации о товаре, является оценка вероятности продажи товара на основе исторических данных о продажах аналогичных продуктов. В данном подходе используется классическая информация о товаре: категория, цена, количество, признак распродажи, дата появления, дата реализации на площадке электронной коммерции [9]. Проведя анализ в различных категориях товаров на электронной площадке Ебай, исследователи выяснили, что больше всего покупателей стимулируют к покупке ценовые факторы: цена и признак распродажи. Ряд авторов выявили [10], что помимо классических факторов товара на поведение потребителя также оказывает влияние качество и количество отзывов о товаре.

С развитием сентиментального анализа стал распространен анализ описания к товару. Например, в своей работе [11] авторы описали подход к выделению и классификации ключевых слов из текстового описания. В своем исследовании авторы проанализировали влияние различных слов на конверсию предложения товара на площадке электронной торговли. Более того, исследователи в своей работе [12] проанализировали влияние текстового описания и выявили, что включение данного фактора в регрессионную модель снижает ошибку прогноза на 30%. Авторы использовали модель нейронных сетей [6].

Критически важным элементом объявления о продаже товара на площадках электронной коммерции является его изображение. Визуальная аналитика требует значительных компьютерных мощностей. Ряд исследователей [13, 14] описали подход к выделению факторов из изображений, оказывающих влияние на выбор потребителя. Авторы выяснили [13], что потребителей привлекают яркие картинки с хорошим качеством, небольшим количеством различных цветов и минимальным количеством объектов на изображении. В своем анализе авторы использовали модель нейронных сетей.

Основной сложностью вероятностного подхода при прогнозировании спроса на товары электронной коммерции на основе данных о товаре – является техническое ограничение при анализе всех описанных факторов одновременно. В связи с этим совокупное влияние факторов на вероятность совершения сделки на площадках С2С практически не исследовано.

Таким образом, существует два основных подхода к прогнозированию спроса на отдельный товар на площадке электронной коммерции. Каждый из подходов обладает рядом недостатков и трудностей при его реализации на практике. При этом, ввиду специфики данных, подход, основанный на детальной информации о клиенте, может быть использован с целью краткосрочного прогнозирования, а вероятностный подход, основанный на детальной информации о товаре для среднесрочного и долгосрочного прогнозирования.

## СПИСОК ЛИТЕРАТУРЫ

1. *Xue W., Xiao B., et al.* Intelligent mining on purchase information and recommendation system for e-commerce // International Conference on Industrial Engineering and Engineering Management (IEEM). 2015.
2. *Rogers S. K.* Item-to-item Recommendations at Pinterest // Proceedings of the 10th ACM Conference on Recommender Systems. Boston. Massachusetts, USA. ACM: 393-393. 2016.
3. *Zheng Y., Liu C., et al.* Neural Autoregressive Collaborative Filtering for Implicit Feedback // Proceedings of the 1st Workshop on Deep Learning for Recommender Systems. Boston. MA. USA. 2016. P. 2-6.
4. *Chadha A., Kaur P.* Comparative analysis of recommendation system // Emerging Trends and Technologies in Libraries and Information Services (ETTLIS). 2015 4th International Symposium on. 2015.
5. *Сорокин А. С.* Построение скоринговых карт с использованием модели логистической регрессии // Интернет-журнал Науковедение. 2014. № 2 (21). С. 82.
6. *Сорокин А. С.* Использование нейросетевых моделей в поведенческом скоринге // Прикладная информатика. 2015. Т. 10. № 2 (56). С. 92-109.
7. *Li, D., Zhao G., et al.* A Method of Purchase Prediction Based on User Behavior Log // 2015 IEEE International Conference on Data Mining Workshop (ICDMW). 2015.
8. *Oyvind H., Thorstein K., Hai Thanh N., Helge L, Kofod-Petersen A.*: Probability-based Approach for Predicting E-commerce Consumer Behaviour Using Sparse Session Data. 2016.
9. *Wu X, Bolivar A.* Predicting the conversion probability for items on C2C ecommerce sites // CIKM '09 Proceedings of the 18th ACM conference on Information and knowledge management Pages. 2010. P. 1377-1386.
10. *Zhu F., Zhang X.* Impact of online consumer reviews on sales: The moderating role of product and consumer characteristics // Journal of Marketing. 2010. Vol. 74 (2). P. 133–148.
11. *Hui Yuan, Wei Xu.* Topic sentiment mining for sales performance prediction in e-commerce // Springer Science+Business Media New York. 2017.
12. *Pryzant R, Chung Y, Jurafsky D.* Predicting Sales from the Language of Product Descriptions. 2017.
13. *Haibin C, Roelof van Zwol, Azimi J., Manavoglu E., Zhang R., Zhou Y., Navalpakkam V.* Multimedia Features for Click Prediction of New Ads in Display Advertising. 2012.
14. *Абламейко С. В., Краснопрошин В. В. Образцов В. А.* Распознавание образов и анализ изображений: теория и опыт решения практических задач. Минск. БГУ. 2013.

## АНАЛИЗ ЭФФЕКТИВНОСТИ ПРОГНОЗИРОВАНИЯ УСПЕВАЕМОСТИ НА ОСНОВЕ FUZZY LOGIC

**Е. В. Кондратьева, О. Ю. Кондратьева**

*Саратовский государственный университет, Россия*  
E-mail: elka@sgu.ru

Статья посвящена анализу эффективности прогнозирования успеваемости обучающихся с использованием математического пакета FUZZY LOGIC.