A REVIEW ON APPLICATIONS OF MACHINE LEARNING IN E-COMMERCE

M. V. RAJESH and S. RAO CHINTALAPUDI

Computer Science and Engineering
Pragati Engineering College
India
E-mail: magavenkatarajesh@gmail.com
          srao.chintalapudi@gmail.com

Abstract

This paper is aimed to explore the applications of Machine Learning in the domain of E-commerce. Recently, E-commerce platforms have extensive impacts on our human life. The rapid growth in E-Commerce industry has lead to an exponential increase in the online purchases. Machine Learning (ML) is a multidisciplinary field which is a combination of statistics and computing techniques with the application of algorithms which is widely used in various domains. In this paper a review has been done on the usage of machine learning techniques in the E-commerce applications like Product Recommendations, Dynamic Price Adjustment, Supply and Demand Prediction, Fraud Detection and Segmentation, Personalization and Targeting.

I. Introduction

Machine Learning is a branch of Artificial Intelligence. Computing systems can be designed that can learn from data as being trained. Many different algorithms can be employed in machine learning based on the required output. There are many applications of machine learning like spam detection, voice recognition, stock trading, robotics, medicine and health care, advertising, retail and E-commerce, gaming analytics, internet of things, etc [1].

Operations from business to business (B2B) which means cutting costs in transactions between businesses and, business to consumer (B2C) which reflects sales of goods and services are included in E-commerce. E-Commerce

became the lifeline for the phenomenal growth of the Internet industry as it is carried on the Internet [18].

This paper is organized in five sections. To start with introduction of the topic of discussion which emphasizes on few important applications of machine learning in E-commerce, followed by the Related Work section, which highlights the research done on these applications. The next is the Data Set Collection section which enumerates various datasets used in these kinds of research activities. The following is the Research Challenges section, which describes the challenges involved in this research work. The paper is wrapped with the Conclusion section.

The exponential increase in the use of web and mobile applications for buying the products by consumers has given more options to choose. There have been many uses of machine learning in E-commerce, a few of them are listed below:

    i. Product Recommendations

    ii. Dynamic Price Adjustment

    iii. Supply and Demand Prediction

    iv. Fraud Detection

    v. Segmentation, Personalization and Targeting

i. Product Recommendations

The users of E-commerce platforms are flooded with lot of choices among the available set of products, which makes the searching a cumbersome and time consuming task. In any E-commerce platforms, there will be a huge collection of products available for purchase, which makes it very difficult for the users to find the product of their interest. Therefore customers expect the helping hand from technology to aid to search and find the products quickly which they are interested. So to automate and speed up the product selection thereby increasing the probability of purchase, product recommender systems are being developed and deployed for use.

Most of the e-commerce platforms have a recommender engine which analyses user search patterns so as to provide relevant recommendations. E-commerce, online auction, and books are few of the major areas where

recommender systems have contributed significantly [10]. Performance of any e-commerce company in terms of reliable products analysis is very crucial. In a broader sense, due to the global infrastructures of an e-commerce company, grouping of many similar items may happen in a different way. So the quality of the items gets analyzed by the customers based on the classification of the items [11].

The capability of the E-commerce platforms to easily search for an intended product will always enhance the purchasing experience of the customers. Products can be categorized using keywords which aid for the realization of the effective search capability [12]. In an E-Commerce (EC) environment, significantly important and a pivotal role is there for the reputation and trust, which enables multiple parties to establish relationships for the mutual benefit [16].

ii. Dynamic Price Adjustment

The E-commerce businesses are majorly dynamic with consistent competition among the market providers. Though the competition is highly stimulating for many merchants, it also creates abundant of new avenues for business improvement.

E-Commerce is one of those industries where there will be very tough competition, especially when it comes to niche consumer segments such as electronic products and beauty products. So it is very crucial to leverage as many advantages as possible for attraction and retention of customers. Offering more competitive prices for the products of interest is one of the most effective ways of doing that, i.e., by permitting the merchants to automatically adapt the product prices in order to respond and take benefit of the altered market conditions. As similar to predicting the financial stock prices, the possibility to adjust the price by forecasting competitors' product prices is also possible [13].

iii. Supply and Demand Prediction

One of the fundamental aspects of retail and ecommerce is the ability to understand the past events and predict the future sales. Demand Forecasting is the process of anticipating the future sales of the products by quantitatively analyzing the effects of product price, promotion activities,

seasonal dependency, and working holidays in the society. The volume of products that consumers will be willing to pay during a specified time period is called demand. On the other hand, forecasting is the process of calculating the future result of customer's acts in a specified time period [7].

To provide an efficient E-commerce platform, how to balance the demand and supply in market segments is one of the most fundamental problems. Always there will be a review section under their products in the e-commerce websites for the consumers to express their opinions which aid for the evaluation of items by prospective consumers, which in turn plays a major role in e-commerce analysis and classification of positive and negative opinions is Sentiment Analysis [14].

iv. Fraud Detection

In principle, fraud in an E-commerce platform occurs when a deceiver visits an online market place and undertakes an unauthorized transaction by employing the particulars of a stolen or fake bank card. Now the merchant will not receive any legal remittance for the products; thus, the E-commerce platform will charge back to the owner of the stolen or fake bank card for the products for which order is placed [9].

As the frequency of online transactions increase, the curve for E-commerce fraud also rises. This situation is especially obvious now, when the society is dragging away from in-store off-line sales. Customers are required to make online purchases to stay safe, due to the COVID-19 prevailing in society everywhere in the world. Therefore, every vendor and financial organizations like banks should pay attention to the recent patterns in e-commerce fraud, so as to detect the same and to ensure the protection of their business. In order to decrease the E-commerce frauds, it is required to develop and deploy a procedure for the risk management and mitigation by all the participating E-commerce major stakeholders like banks and vendors. The environment of transaction processing should be more controlled as openness might be prevalent in all channels of the transactions like mobile, online networks and staff.

Fraudulent E-commerce websites makes many of Internet users as victims leading to the loss for the customers. As there will be an on-going large number of E-commerce transactions, their datasets are also large.

Therefore to identify fraudulent transactions, fast and efficient algorithms are required. Rule-based or re-training based methods are used for fraud detection. It is not easy to detect fraud as it is happening or within a short time span and also required advanced techniques [21].

v. Segmentation, Personalization and Targeting

In an E-commerce platform, quality product descriptions are critical for providing competitive customer experience. To improve the likelihood of purchase of the product by the customers, an accurate and attractive description certainly helps. Ecommerce platforms are separated from their customers due to lack of in person and physical interfacing. The salesperson interacting with a customer in an offline sales mode segments and targets, and personalizes the customer's experience to get them to buy as he/she can quickly takes in what they are saying, their body language, behaviour, and many other factors in order to help the customer.

To meet the diverse needs of different users, personalization plays an important role. There will be a collection of different types of search and recommendation tasks, in many large portals such as Taobao and Amazon, operating simultaneously for personalization [19].

II. Related Work

Liaoliang Jiang et al. [2] proposed a slope one algorithm based on the fusion of trusted data and user similarity, which can be deployed in various recommender systems. In the field of collaborative filtering recommendations, the slope one algorithm is easy to implement and also efficiently and effectively works. However, the prediction accuracy of the slope one algorithm is not very high. Another drawback of the slope one algorithm is that it does not perform so well with personalized recommendation tasks. They have performed the verification of the algorithm through execution using the Amazon dataset and which showcased that their algorithm for recommendation system produced the outcome more precisely than the conventional slope one algorithm.

C. Chavaltada et al. [12] compared different machine learning techniques and their performance on product categorization. P. Hamsagayathri et al.

[15] classified the customer reviews on women clothing *e*-commerce using various classifiers and their performances are evaluated using accuracy, precision and recall parameters.

S. Serth et al. [3] developed Price Wars, an open continuous time framework to replicate the dynamic pricing competition, to solve the problems encountered by both E-commerce professionals and researchers. It is time-consuming to test the automated pricing procedures and also very risky when tried in production, for e-commerce practitioners. On the other end, the researchers also have trouble to understand and analyze how strategies for pricing gets their real time implementation under huge competition in the dynamic E-commerce businesses.

Rainer Schlosser et al. [4] experimented with the stochastic dynamic pricing models in dynamic business environments by employing various product dimensions, like cost, performance, and customer liking. It is very tough to obtain the pricing strategies that will be efficient because of the complicated nature of E-commerce markets. In a first step, they analysed how the probability of sales is impacted by particular consumer attitude and the step wise interaction of pricing strategies with the help of a simulated test market. In a second step, they formulated an effective pricing strategy which includes the curse of dimensionality using a dynamic programming model.

Jiatu Shi, Huaxiu Yao, et al. [5] proposed a novel algorithm, RMLDP, to incorporate a multi-pattern fusion network (MPFN) with a meta learning paradigm. This algorithm considers both local and global temporal patterns for segment demand prediction. In this paradigm, the transferable knowledge is regarded as the model parameter initializations of MPFN, which are learned from diverse source segments. The conventional machine learning models were greatly successful on data-sufficient segments, but they may fail in a large-portion of segments in E-commerce platforms, because to learn well-trained models there are no sufficient records. In this paper, this problem in the context of market segment demand prediction was tackled.

Adrian Micu., et al. [20] proposed applications of machine learning in digital marketing and E-commerce by examining qualitative determinant factors on brand logos and correlations on companies' income, with profit, the

number of employees, images and number of product images on E-commerce homepage.

J. Nanduri, Y. W. Liu, et al. [6] bring up two inventive methods, to effectively handle the confrontation of identifying various fraud styles as fraud islands (link analysis) and multi-layer machine learning model. The dynamic and diverse nature of E-commerce fraud patterns is the main challenge for the prevention of E-commerce operational frauds. The correlation between different fraudulent entities is verified and thereby the hidden complex fraud patterns are identified using Fraud Islands, which are formed with link analysis. Their experimentation proved that the efficiency of fraud decisions can be predominantly enhanced by employing different machine learning algorithms which were trained based on various types of fraud labels.

E. Khoo, et al., [17] investigates a website that might contain some specific styles that imply it is fraudulent based on aspects like HTML tags, text content and pictorial content. The efficacy of the detection of fraudulent e-commerce websites is analyzed using machine learning techniques.

Qibin Chen, Junyang Lin, et al., [7] combined the power of neural networks and knowledge base to explore a new way to generate personalized product descriptions. Specifically, in the context of E-commerce proposed a product description generation model called KnOwledge Based pErsonalized (or KOBE). KOBE considers a variety of important factors during text generation, to make the description both informative and personalized; they include product aspects, user categories, and knowledge base.

A. Suvarna, et al. [22] explore the existence of gender bias in product descriptions described on E-commerce websites and also proposes a framework to provide gender-neutral product descriptions.

III. Dataset Collection

Datasets are very much required for the research activity to be carried in the area of machine learning. To carry out the analysis and further research in this topic, the required datasets can be collected from various sources, few of which are listed below.

- https://data.world/
- https://archive.ics.uci.edu/ml/datasets.php
- https://datarade.ai/
- https://www.kaggle.com/
- https://datastock.shop/
- https://github.com/

IV. Research Challenges

There are significant challenges to take up the research in this topic of machine learning in e-commerce, due to the dynamism and diversity involved in the business aspects of the online business platforms. The opportunities and challenges of marketplaces are continuously changing due to steady rise of e-commerce. Application of predominantly real time based approaches based on the competitor's market decisions, demand projections, or certain consumer purchase patterns that alter the prices of products can be called dynamic pricing. One of the challenges is that vendor's don't have the efficient environment to test their strategies of pricing accurately before deploying them in real time markets. Nonexistence of replicating platforms for voluminous pricing competitions with challenging pricing approaches is another challenge from the research point of view [3].

Another challenge can be pointed here is that, in a large-portion of segments in E-commerce platforms, due to unavailability of sufficient records to learn well-trained models, the conventional machine learning models may fail to perform well [5]. One more important challenge is *e*-commerce transaction fraud prevention due to the dynamic and diverse nature of fraud patterns. Serious threats are imposed onto *e*-commerce because of transaction frauds [8]. The users being surfaced the right content at the right time plays a vital role in *e*-commerce in providing a superior shopping experience to customers which is called product recommendation [9]. To understand customers' intentions, timely and provide valuable assistance to find what they are looking for and during their entire shopping process is one of the challenges. The product recommendation algorithm to filter the interested items from vast collection of products needs to be accurate, which always makes it complicated and tough for implementation [2].

V. Conclusion

Hence, a review has been carried out on previous works covered on the usage of machine learning used in the various aspects of e-commerce operations like Product Recommendations, Dynamic Price Adjustment, Supply and Demand Prediction, Fraud Detection and Segmentation, Personalization and Targeting. Collaborative filtering recommendations, market segment demand prediction, detecting diverse fraud patterns, neural networks and knowledge base has been focused in this review. Lastly, Data Mining techniques, Machine Learning approaches and several other algorithms has proven their efficiency in the area of E-commerce.

References

[1]   Machine Learning for Big Data, Jason Bell, Wiley Big Data Series, Wiley India Pvt. Ltd., Reprint (2017).

[2]   L. Jiang, Y. Cheng, L. Yang, et al. A trust-based collaborative filtering algorithm for E-commerce recommendation system, J Ambient Intell Human Comput 10 (2019), 3023-3034. https://doi.org/10.1007/s12652-018-0928-7.

[3]   S. Serth et al., An Interactive Platform to Simulate Dynamic Pricing Competition on Online Marketplaces, 2017 IEEE 21st International Enterprise Distributed Object Computing Conference (EDOC), Quebec City, QC, 2017, pp. 61-66, doi: 10.1109/EDOC.2017.17.

[4]   Rainer Schlosser and Martin Boissier, Dynamic Pricing under Competition on Online Marketplaces: A Data-Driven Approach. In Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '18). Association for Computing Machinery, New York, NY, USA, 705–714. DOI:https://doi.org/10.1145/3219819.3219833. (2018).

[5]   Jiatu Shi, Huaxiu Yao, et al., Relation-aware Meta-learning for Market Segment Demand Prediction with Limited Records. Conference'17, July 2017, Washington, DC, USA arXiv:2008.00181v1 [cs. LG] 1 Aug (2020).

[6]   J. Nanduri, Y. W. Liu, K. Yang and Y. Jia Ecommerce Fraud Detection Through Fraud Islands and Multi-layer Machine Learning Model, In: Arai K., Kapoor S., Bhatia R. (eds) Advances in Information and Communication, FICC 2020. Advances in Intelligent Systems and Computing, vol 1130. (2020) Springer, Cham. https://doi.org/10.1007/978-3-030-39442-4_41.

[7]   Qibin Chen, Junyang Lin and et al., Towards Knowledge-Based Personalized Product Description Generation in *E*-commerce, KDD 19: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining July 2019 Pages 3040-3050https://doi.org/10.1145/3292500.3330725.

[8]   S. Wang, C. Liu, X. Gao, H. Qu and W. Xu, Session-based fraud detection in online *e*-commerce transactions using recurrent neural networks. In: Joint European Conference on Machine Learning and Knowledge Discovery in Databases, Springer, Cham, September (2017), 241-252.

[9]   Konstantina Christakopoulou, Alex Beutel, Rui Li, Sagar Jain and Ed H. Chi. 2018. Q and R: A Two-Stage Approach Toward Interactive Recommendation. In KDD'18. 139-148.

[10]   Khalid Anwar, Jamshed Siddiqui and Shahab Saquib Sohail, Machine learning-based book recommender system: a survey and new perspectives, International Journal of Intelligent Information and Database Systems, Volume 13, Issue 2-4, August 2020. doi.org/10.1504/IJIIDS.2020.109457

[11]   B. M. Cheekati and S. V. Padala, Evaluating the Progressive Performance of Machine Learning Techniques on E-commerce Data, In: Satapathy S., Bhateja V., Das S. (eds) Smart Computing and Informatics. Smart Innovation, Systems and Technologies, vol 78. Springer, Singapore. https://doi.org/10.1007/978-981-10-5547-8_10

[12]   C. Chavaltada, K. Pasupa and D. R. Hardoon, A Comparative Study of Machine Learning Techniques for Automatic Product Categorisation, In: Cong F., Leung A., Wei Q. (eds) Advances in Neural Networks - ISNN 2017. ISNN 2017. Lecture Notes in Computer Science, vol 10261. Springer, Cham. https://doi.org/10.1007/978-3-319-59072-1_2

[13]   Houda Bakir, Ghassen Chniti and Hédi Zaher, E-Commerce Price Forecasting Using LSTM Neural Networks International Journal of Machine Learning and Computing, Vol. 8, No. 2, April 2018 doi: 10.18178/ijmlc.2018.8.2.682

[14]   A. Noor and M. Islam, Sentiment Analysis for Womens E-commerce Reviews using Machine Learning Algorithms, 2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT), Kanpur, India, 2019, pp. 1-6, doi: 10.1109/ICCCNT45670.2019.8944436.

[15]   P. Hamsagayathri and K. Rajakumari, Machine learning algorithms to empower Indian women entrepreneur in *E*-commerce clothing, 2020 International Conference on Computer Communication and Informatics (ICCCI), Coimbatore, India, 2020, pp. 1-5, doi: 10.1109/ICCCI48352.2020.9104111.

[16]   Elshrif Ibrahim Elmurngi and Abdelouahed Gherbi, Unfair Reviews Detection on Amazon Reviews using Sentiment Analysis with Supervised Learning Techniques. Journal of Computer Science 2018, 14 (5): 714.726 DOI: 10.3844/jcssp.2018.714.726

[17]   E. Khoo, A. Zainal, N. Ariffin, M. N. Kassim, M. A. Maarof and M. Bakhtiari, Fraudulent *e*-Commerce Website Detection Model Using HTML, Text and Image Features, In: A. Abraham, M. Jabbar, S. Tiwari, I. Jesus (eds) Proceedings of the 11th International Conference on Soft Computing and Pattern Recognition (SoCPaR 2019). SoCPaR 2019. Advances in Intelligent Systems and Computing, vol 1182. Springer, Cham. https://doi.org/10.1007/978-3-030-49345-5_19 (2021)

[18]   Vinay Kumar Jain and Shishir Kumar, Improving Customer Experience Using Sentiment Analysis in E-Commerce, Handbook of Research on Intelligent Techniques and Modeling Applications in Marketing Analytics 2017, 10.4018/978-1-5225-0997-4.ch012

[19] Yabo Ni, Dan Ou, Shichen Liu, Xiang Li, Wenwu Ou, Anxiang Zeng, and Luo Si, Perceive Your Users in Depth: Learning Universal User Representations from Multiple E-commerce Tasks. In Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & amp; Data Mining (KDD 18). Association for Computing Machinery, New York, NY, USA (2018), 596-605.
DOI:https://doi.org/10.1145/3219819.3219828

[20] Adrian Micu, Marius Geru, Alexandru Capatina, Constantin Avram, Robert Rusu and Andrei Alexandru Panait, Leveraging e-Commerce Performance through Machine Learning Algorithms, Annals of Dunarea de Jos" University of Galati Fascicle I. Economics and Applied Informatics Years XXV – no2/2019.
DOI https://doi.org/10.35219/eai1584040947

[21] J. Shaji and D. Panchal, Improved fraud detection in *e*-commerce transactions, 2017 2nd International Conference on Communication Systems, Computing and IT Applications (CSCITA), Mumbai, 2017, pp. 121-126, doi: 10.1109/CSCITA.2017.8066537.

[22] A. Suvarna, K. Dey, S. Nagar, N. Madaan and S. Mehta, Handling Gender Biases in E-Commerce Product Specifications, 2019 Grace Hopper Celebration India (GHCI), Bangalore, India, 2019, pp. 1-5, doi: 10.1109/GHCI47972.2019.9071916.