Exploring Challenges and Innovations in E-Commerce Recommendation Systems: A Comprehensive Review

Mashfiqur Rahman, Samiul Mushfik, Makhdum Ahsan Rupak, Muhammad Zubair Hasan, Minhaz Bin Farukee, and Shawrup Kumer Suter

Abstract Recommendation systems play a pivotal role in the digital age, with ongoing research focused on enhancing their effectiveness. This paper delves into the common challenges associated with developing these systems, including the cold-start problem, handling sparse datasets, and the use of matrix filling in hierarchical methods. We explore innovative approaches that include the integration of diverse algorithms and the application of alternative techniques, such as deep learning. Our research aims to establish an empirically based standard for various aspects of recommendation systems, thereby serving as a valuable reference for future studies.

Keywords Collaborative filtering · Content-based filtering · Association rule

1 Introduction

The e-commerce industry has seen significant growth in recent years, with online stores, streaming services, and meal delivery services rapidly expanding in both quantity and quality. With this rapid growth, the challenge of effectively and efficiently presenting a vast array of items to customers intensifies. Therefore, it's crucial

M. Rahman
Vanguard Australia, Melbourne, Australia

S. Mushfik
Stibo DX, Dhaka, Bangladesh

M. A. Rupak (✉)
Kona Software Lab Ltd, Dhaka, Bangladesh
e-mail: ahsanrupak97@gmail.com

M. Z. Hasan
Celloscope, Dhaka, Bangladesh

M. B. Farukee
ACI Ltd, Dhaka, Bangladesh

S. K. Suter
Delivery Hero, Berlin, Germany

S. Kumar et al. (eds.), *Fourth Congress on Intelligent Systems*, Lecture Notes in Networks and Systems 869, https://doi.org/10.1007/978-981-99-9040-5_8

to present users with items that align with their interests or purchasing tendencies. This subject has been the focus of numerous studies.

Recommendation systems traditionally highlight popular items. However, leveraging user-specific data can boost product consumption as these products are more likely to align with the individual preferences of the target consumer [1]. Collaborative filtering, based on user-item interactions, is prevalent.

Association rules are used to reveal correlations between variables. For instance, after using a baseline algorithm, keyword associations with coupon descriptions were used in [2]. The study in [3] employed association rules to identify product categories. The combination of the neighbourhood technique and the latent model-based approach is also a subject of study [4].

Collaborative filtering faces challenges like the 'cold-start' problem, where newly created users or items lack interaction data. Moreover, traditional algorithms do not consider the temporal aspect of user-item interactions, failing to distinguish between recent and older purchases. These issues can be addressed by modifying existing algorithms [5, 6].

Alternative strategies, like predicting future user purchases [7] or suggesting pricing for products [8], can greatly boost the performance of recommendation.

In the context of e-commerce, coupons are a popular tool to increase user engagement. The methods of coupon distribution are an active area of research [9]. However, simply distributing coupons randomly may not achieve the desired results. It is also crucial to ensure a satisfactory coupon usage ratio [10].

2 Background Knowledge

2.1 Collaborative Filtering (CF)

CF is a predictive technique that leverages the preferences of a large user base to estimate an individual user's interests [11]. CF primarily utilises two strategies: one is neighbourhood-based methods which can be divided into user-based and item-based CF. In the user-based approach, the algorithm identifies users with similar rating patterns to the active user, using their ratings to predict recommendations for the active user. Conversely, item-based CF generates recommendations for users based on item similarities. The other is latent factor model-based methods which predominantly employ matrix factorization, a popular technique that identifies the latent variables of users and items to formulate recommendations.

2.2 Content-Based Filtering (CBF)

CBF strategies, as outlined in [12], hinge on the utilisation of item descriptions and a user's preference profile to formulate recommendations. These strategies are particularly effective when there is a wealth of information about an item, such as its name and description, but limited data about the user. CBF recommenders perceive the recommendation process as a user-specific classification task. In this scenario, a classifier learns the user's preferences based on the attributes of items. Items are characterised by keywords, and a user profile is constructed to encapsulate the types of items the user is inclined towards.

3 Studies on Recommendation Systems

Various studies have been done on the topic of recommendation system discussing Improvement of algorithms, handling techniques for different problems faced in recommendation including coupon usage and disbursement.

3.1 Importance of Personalised Recommendation

Email and Internet advertising with offers are prevalent which were meticulously selected by professionals. But it is expensive and time-intensive, because it disregards consumer preferences and purchasing history. A CBF system streamlines coupon selection and personalises suggestions in [1] to increase click-through rates and conversions. The suggested method utilised only positive input and depicted shopping trip prediction as a one-class classification (OCC) problem using only positive samples and false negative samples from the user's purchase history. RFC and XGBoost algorithms were used for classification since they are capable of handling noisy data, outliers, and missing features. To validate the model, two baseline methodologies were explored. One baseline picks the most popular discounts based on click data from the past, while the other utilises click data from the future. The RFC model anticipated coupon click-through rates of 96.80%, while the XGBoost model predicted click-through rates of 95.85%, both outperforming the benchmarks.

3.2 Combination of Collaborative Filtering and Other Approaches

In [2], author suggested a novel coupon offer recommendation algorithm that combines the baseline algorithm and keyword association. Each user's suggested offer weights are calculated using keyword association criteria. Since offers expire quickly, the suggested technique cannot advocate expired offers using association rules. Thus, keywords were evaluated instead of offers. This approach offers consumers fresh discounts depending on their purchase history and word pair association criteria. The approach outperforms the quantity algorithm in percentage of hits, average rank,

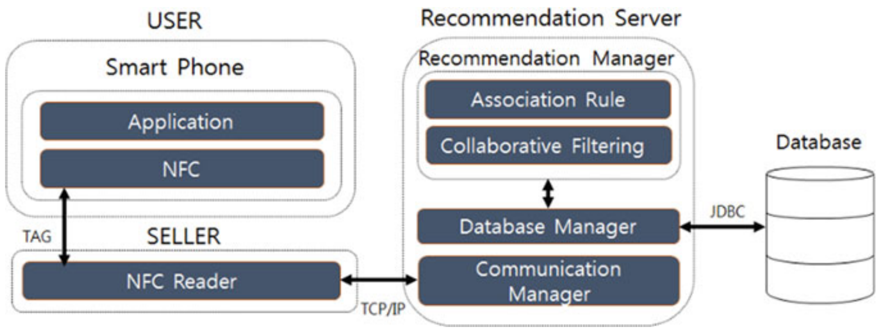

Fig. 1 System architecture

and mean reciprocal rank. Top 5 suggestions had a 7.8% increase in hits, top 10 a 2.4% rise, and top 20 a 3.4% increase. This technique produced lower average rank and greater mean reciprocal rank than the baseline for top 5, top 10, and top 20 recommendations.

The article [3] uses NFC to implicitly collect customer data and forecast companies with a high likelihood of a customer visit using the association rule and collaborative filtering. The recommendation system has two main parts: Recommendation Manager and Communication Manager (Fig. 1). The Communication Manager responds to consumer phone recommendations. The recommendation manager analyses usage patterns using association rule-based k-means algorithm and recommends things for comparable consumers using collaborative filtering. The recommendation server returns the list to the consumer.

Association rules find product categories that clients are most likely to buy, while CF suggests underused things from highly comparable ones. Apriori was used to find association rules. Customers could not directly review things because this study collected data indirectly. The review matrix included the customer's visit history. Collaborative filtering employed the Pearson correlation coefficient and K-NN algorithm for the nearest neighbour.

Tan et al. [13] provide a recommendation engine architecture for e-learning platforms that uses collaborative filtering and association rules to collect data, engage with users, and provide recommendations. The ratio of actual study hours to course duration was used to construct rating ratings. The Pearson correlation coefficient calculated user proximity to identify neighbourhood. Finally, the top-N algorithm in user-specific regions created suggestion lists. If data was scarce, a non-personalised baseline popular items method was used.

Due to the continual evolution of fashion content and the long tail, purchasing behaviours of users material-based algorithms fail. The authors of [4] recommend identifying similar products and applying user-level personalisation to them. The proposed algorithm assigns a value to each item based on the affinity between the target product and the user's previously used products and taste. First, they gathered implicit user data which were weighted to generate ratings. Item-based CF developed candidate goods. Using Alternate Least Square Matrix Factorization and Bayesian Personalised Ranking Matrix Factorization, they then generated user and item latent vectors to measure user-product similarity and added these two scores linearly. MAP@15 was the statistic in question. The combination of user score and item-item similarity increased the MAP value from 0.04 to around 0.045. Similarity between users and items that is weighted increased the score. Most sparse user-item matrix recommendation system challenges may apply this strategy.

3.3 Handling Problems in Recommendation

Cold start and data sparsity still plague many real-world recommendation settings, even if collaborative filtering generally produces good suggestions. A research presents a hybrid recommendation system that uses collaborative filtering (coupon usage history) and other data sources to improve suggestions for new users and new offers. It integrates demographics, discounts, and usage history. The proposed system design is shown in Fig. 2. The suggested system matches new and current commodities using user and offer data instead of preference data. This extra data covers demographics, location, and product category metadata. NLP extracted discount description tags from unstructured text. Thus, a dictionary of term-category associations was created to match new things with the categories most similar to the description terms. The study used few months of user profiles, offer descriptions, and coupon redemption data.

Traditional CF recommendation systems in e-commerce online platforms have two significant issues: data sparsity and lack of timeliness. CF uses a sparse user-item rating matrix. The authors of this study [6] strongly suggest considering time and putting more weight to data close to evaluation period and gradually decreasing weight to data distant from evaluation period as consumer choice and preferences dynamically change over time and more recent transactions are more likely to exhibit relevant state of consumer preferences. This research proposes a realistic solution technique and investigates the accuracy effects of closest neighbours. Hierarchical filling reduces CF rating matrix sparsity. Calculating RF * IRF for each category of items in the initial rating matrix and filling in the average item rating of RF * IRF larger than threshold is the fundamental notion. Time and Pearson's correlation coefficient are used to calculate user-user similarity [14]. The model beats classical CF in MSE and RMSE.

3.4 Different Approaches for Recommendations

Predicting user purchases may advise or discount things. Sequential pattern analysis (SPA) predicts user purchases. This method assigns support values to common sequence patterns. These patterns mirror the user's buying sequence. Candidate: later

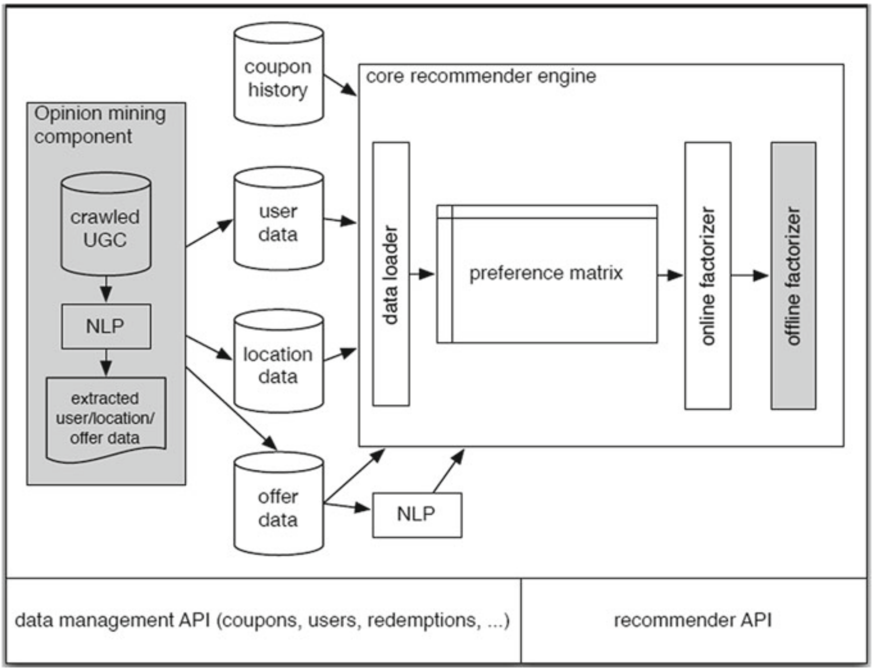

Fig. 2 Recommendation system architecture

matching pattern part. Candidate support values are added to determine scores. SPA recommends top-scoring products. Due to the vast number of users with varied purchase sequences, support is limited. Long-distance transactions are patterns; thus, recency is disregarded. Authors proposed an alternative. Analysing purchase patterns in subsequences which ensures recency.

Zhao et al. [8] offer a breakthrough e-commerce recommendation system that emphasises product price and personalised suggestions to boost actual company profits. The paucity of e-commerce recommender research and the requirement for particular tailoring drove the authors' findings. Before then, most research neglected product pricing's dynamic nature, unlike other recommender features. The authors think altering product pricing for each user and utilising them in recommenders might revolutionise recommender design. This demands customers' Willingness to Pay (WTP) for linked things. WTP enables customer-specific pricing for marketing. Through a new auction-based method, taste-based paid participants choose basic products from 120k skincare products. The pre-developed e-commerce website outperforms Amazon's price using the provided method. The method cannot be generalised since it was tested on pre-selected commodities. Topic dependence, product replacement, and customer responsiveness to targeted discounts are further difficulties.

Table 1 Summary

Method	Overall (%)	Daily average (%)	Daily std. (%)
popularity	48.9	48.6	4.0
user_segment	36.4	36.7	3.6
interest sim	10.9	10.7	1.9
item CF	13.0	13.0	1.6

3.5 Coupon Usage in Recommendation

This study [9] suggests different discount coupon recommendation methods which predict user coupon consumption based on coupon availability. Customers received various coupon recommendations from authors. Different methods performance are summarised in Table 1.

Popularity-based strategies perform better. No suggestion techniques were used, and customers printed coupons upfront. Discount coupons entice people, but they don't necessarily buy. Predicting coupon utilisation likelihood improves coupon usage ratio. Machine learning analyses coupon use behaviour in [10]. This paper contributes fourfold. Coupon use, feature extraction and impact, 'old user effect' and other overfitting features, and comparison with other approaches.

4 Conclusion

Today's e-commerce uses recommendation algorithms. Individualization trumps popularity. Researchers integrated various algorithms for multiple purposes to improve recommendations. Integration of several data sources and hierarchical filling techniques alleviate cold-start and sparse data concerns. It is possible to investigate collaborative and content-based filtering concerns.

Conventional algorithms market items based on user or product similarity. Beyond these methods, the system may be enhanced through creative thought. Predicting consumer purchases and adjusting pricing can be useful in indicating.

References

1. Xia Y, Di Fabbrizio G, Vaibhav S, Datta A (2017) A content-based recommender system for e-commerce offers and coupons. ECOM@ SIGIR
2. Huang Y (2016) Personalized recommendation of coupon deals by keywords association rules. J Ind Intell Inf 4
3. Jooa J, Bangb S, Parka G (2016) Implementation of a recommendation system using association rules and collaborative filtering. Procedia Comput Sci 91:944–952

4. Agarwal P, Vempati S, Borar S (2018) Personalizing similar product recommendations in fashion E-commerce. ArXiv Preprint ArXiv:1806.11371
5. Grivolla J, Campo D, Sonsona M, Pulido J, Badia T (2014) A hybrid recommender combining user, item and interaction data. In: 2014 international conference on computational science and computational intelligence, vol 1, pp 297–301
6. Zhang Z, Xu G, Zhang P (2016) Research on e-commerce platform-based personalized recommendation algorithm. Appl Comput Intell Soft Comput
7. Lu H (2014) Recommendations based on purchase patterns. Int J Mach Learn Comput 4:501
8. Zhao Q, Zhang Y, Friedman D, Tan F (2015) E-commerce recommendation with personalized promotion. In: Proceedings of the 9th ACM conference on recommender systems, pp 219–226
9. Mutanen T, Nousiainen S, Liang H (2010) Personalized recommendation on discount coupons. Annu Int Conf Comput Sci Educ Innov Technol 127–132
10. He J, Jiang W (2017) Understanding users' coupon usage behaviors in e-commerce environments. In: 2017 IEEE international symposium on parallel and distributed processing with applications and 2017 IEEE international conference on ubiquitous computing and communications (ISPA/IUCC), pp 1047–1053
11. Ricci F, Rokach L, Shapira B (2011) Introduction to recommender systems handbook. In: Recommender systems handbook, pp 1–35
12. Aggarwal C et al (2016) Recommender systems. Springer, Berlin
13. Tan H, Guo J, Li Y (2008) E-learning recommendation system. In: 2008 international conference on computer science and software engineering, vol 5, pp 430–433
14. Benesty J, Chen J, Huang Y, Cohen I (2009) Pearson correlation coefficient. In: Noise reduction in speech processing, pp 1–4