African Journal of Advanced Pure and Applied Sciences (AJAPAS)
Online ISSN: 2957-644X
Volume 3, Issue 4, October - December 2024
Page No: 175-183
Website: https://aaasjournals.com/index.php/ajapas/index
ISI 2024: 0.877 SJIFactor 2024: 6.752 معامل التأثير العربي: 1.55

Enhancing Books Recommendation Systems Via External Information Based on Sentiment Analysis Factors

Mustafa M. Abuali [1*], Tarek M. Ghomeed [2]
[1,2] Department of Computer and Information Technology, College of Electronic Technology - Bani Walid, Libya

تحسين أنظمة التوصية بالكتب عن طريق استخدام معلومات خارجية مستخرجة من عوامل تحليل المشاعر

مصطفى مفتاح إبراهيم أبوعلي [*1]، طارق معتوق غميض [2]
[2،1] قسم الحاسوب وتقنية المعلومات، كلية التقنية الالكترونية، بني وليد، ليبيا

[*]Corresponding author: maboali2050@gmail.com

Received: August 30, 2024	Accepted: October 11, 2024	Published: October 17, 2024

Abstract:
Obtaining books or scientific research on a specific topic has become increasingly difficult due to the huge amount of material available on the Internet. In this age of information overload, traditional recommender systems alone are no longer enough. They often rely on the number of ratings and reviews a book has received, which may not accurately measure its actual scientific value, leading to recommendations that may inevitably waste the user's valuable time and effort.
To solve this critical issue, it is necessary to develop highly efficient recommendation systems that can provide satisfactory results and to achieve this, incorporating external information from sentiment analysis systems holds great promise. By harnessing the power of sentiment analysis, we can extract the true sentiment and value of a book.
By incorporating sentiment analysis into recommendation systems, we can improve their performance on multiple fronts. First, through sentiment analysis, we can determine not only the amount of feedback writers receive, but also the quality of the feedback. This comprehensive understanding helps filter out irrelevant or biased opinions, ensuring that recommendations are based on reliable and trustworthy sources but are also culturally relevant and contextually appropriate.
Our research aims to contribute a better understanding of the use of sentiment analysis for recommender systems in real-world scenarios. By exploring the potential of sentiment analysis and incorporating its insights, we strive to bridge the gap between traditional recommender systems and user expectations. Ultimately, our efforts seek to improve the user experience and practicality of book recommendations, enabling individuals to effortlessly discover valuable scholarly literature. Through this endeavour, we hope to provide users with reliable and targeted recommendations that enhance their quest for knowledge and expand the horizons of scientific exploration.

Keywords: E-book, book recommender, Sentimental analysis, Recommendation system, Natural language processing.

الملخص
أصبح الحصول على الكتب أو الأبحاث العلمية حول موضوع معين أكثر صعوبة بسبب الكم الهائل من المواد المتاحة على الإنترنت. في هذا العصر من فائض المعلومات، لم تعد أنظمة التوصيات التقليدية كافية بمفردها. فهي غالباً ما تعتمد على عدد التقييمات والمراجعات التي حصل عليها الكتاب، والتي قد لا تقيس بدقة قيمته العلمية الفعلية، مما يؤدي إلى توصيات قد تضيع وقت المستخدم وجهوده الثمينة. لحل هذه المشكلة الحرجة، من الضروري تطوير أنظمة توصيات عالية الكفاءة قادرة على تقديم نتائج مرضية. ولتحقيق ذلك، يحمل دمج المعلومات الخارجية من أنظمة تحليل المشاعر وعداً كبيراً. من خلال الاستفادة من قوة تحليل المشاعر، يمكننا استخراج الشعور والقيمة الحقيقية للكتاب.

عند دمج تحليل المشاعر في أنظمة التوصيات، يمكننا تحسين أدائها على عدة جوانب. أولاً، من خلال تحليل المشاعر، يمكننا تحديد ليس فقط كمية الملاحظات التي يتلقاها الكتاب، ولكن أيضاً جودة هذه الملاحظات. هذا الفهم الشامل يساعد في تصفية الآراء غير ذات الصلة أو المتحيزة، مما يضمن أن التوصيات تعتمد على مصادر موثوقة وجديرة بالثقة، بالإضافة إلى كونها ملائمة ثقافياً وملائمة للسياق.
تهدف أبحاثنا إلى تقديم فهم أفضل لاستخدام تحليل المشاعر في أنظمة التوصيات في سيناريوهات العالم الحقيقي. من خلال استكشاف إمكانيات تحليل المشاعر ودمج رؤاها، نسعى إلى سد الفجوة بين أنظمة التوصيات التقليدية وتوقعات المستخدمين. في النهاية، تهدف جهودنا إلى تحسين تجربة المستخدم وعملية تقديم توصيات الكتب، مما يمكّن الأفراد من اكتشاف الأدب العلمي القيم بسهولة. ومن خلال هذا المسعى، نأمل في تقديم توصيات موثوقة ومستهدفة تساعد المستخدمين في سعيهم للمعرفة وتوسيع آفاق الاستكشاف العلمي.

الكلمات المفتاحية: الكتب الإلكترونية، نظام توصية الكتب، تحليل المشاعر، نظام التوصية، معالجة اللغة الطبيعية.

Introduction

The use of book recommendation systems has become increasingly important in helping users find books that suit their interests. These systems traditionally use a person's reading history, book ratings, and profiles to create recommendations. More recently, there has been a shift towards incorporating external information to enhance the performance of recommendations. Social media and e-commerce platforms have become valuable source for mining textual data, which includes user reviews and comments. Sentiment analysis, a branch of recommender systems, focuses on automatically detecting sentiment such as positivity or negativity from user reviews. It has been applied in various systems, including movie recommendations and public platforms [6].

To improve book recommendation systems, we propose to add sentiment analysis factors to identify hidden sentiments and analyze user sentiment. Using data from Goodreads, a public online platform for book lovers, we aim to enhance recommendation systems by incorporating sentiment analysis, sentiment concordance, and sentiment discrepancy between readers and rating users. This approach allows us to delve into the emotional side of readers' experiences with books.

The results showed that using sentiment analysis features significantly improves book recommendation systems. By considering the sentiments and sentiments expressed in user reviews, the system can better understand the preferences and needs of individual users. This makes it possible to create more accurate and personalized book recommendations, ensuring a higher level of user satisfaction.

Furthermore, emotional concordance measures the degree of agreement between the reader's emotional response and the feelings expressed in the book. This aspect helps in identifying books that match well with the user's emotional preferences. On the other hand, sentiment discrepancy analyzes the gaps between the emotions felt by readers and the emotions expressed by assessment users. This analysis provides valuable insights into potential contradictions that may exist in interpreting the emotional impact of the book.

By merging sentiment analysis results such as sentiment agreement or sentiment discrepancy with Goodreads dataset by merging the field extracted from sentiment analysis and adding it to the books entity i.e. book database Recommendation System, we create a more comprehensive and accurate understanding of user preferences. This not only enhances the accuracy of book recommendations, but also allows for a deeper exploration of the emotional connections between readers and the books they love. The inclusion of sentiment analysis factors paves the way for a more personalized and satisfying reading experience for users.

To utilize from this integration and make it more useful for the user, saving effort and time. There are many challenges facing this research, namely how to approximate the results of recommendations similar to user behavior and interests, in addition to challenging the “cold start” problem [11], since the user is new to the system and we do not have information about his behavior. In this research, we build a tool that leverages recommendations that are close to what the user is searching for.

The purpose of this proposal is to improve the quality of the recommender system using external information-based factors of sentiment analysis systems, by incorporating these factors as a feature of the dataset on which the recommendation algorithm is trained to ensure accuracy.

Related Works

Chnabel, T.; Bennett, P.N Joachims:[1] proposed enhancing recommendation systems has become an important matter and must go beyond algorithms, by using the process of generating information and feedback, which increases the chances of reducing the cost of accessing information and facilitating the recommendation process. This can contribute significantly to the accuracy of learning and the effectiveness of recommendation systems.

Hassanpour, N. Abdolvand, and S.RajaeeH: [2] proposed People's ideas are considered one of the important issues that enhance users' decision to use a particular product or benefit from it. Most platforms also use a recommendation system to increase customer satisfaction, and this will increase the opportunity for profit. Generating information by paying attention to the psychological aspects and opinions of users is considered one of the things that generates collective wisdom that has a positive impact. On customer decision-making.

Sundermann, Dominguez:[4] proposed They surveyed existing recommender systems, studied contextual information, extracted user opinions and recommended integrating the data with recommender systems.

Sallam et al [5] In this study, to improve recommendation systems, two methods were used: collaborative filtering based on individual value analysis and item-based, where the degree of similarity between the target item and the other item is determined. Thus, the filtering process is completed and its results are calculated based on the greatest degree of similarity between the elements, as well as greatest rating.

Y. Park, and A. Choi:[7] In the last application fields, many social relationships between humans, which were performed by face-to-face meetings in the last century, have shifted into online meetings within the last decade. It is so identical in this situation that book reading habits have shifted into online interactions. Thus, online book recommendation and evaluation methods have gained significant interest within the last ten years.

Gong, S.; Cheng, G:[9] User-based techniques go through two main stages to forecast items' ratings for a specific user. The first stage locates similar users to the target user. The second stage obtains rates from similar users to the active user, then using them to produce recommendations. There have been many collaborative filtering algorithm measures that calculate the similarities among users. The commonly used similarity measures in the literature include mean-squared difference, Pearson correlation, cosine similarity, Spearman correlation, and adjusted cosine similarity.

Schafer, J.B.; Frankowski:[10] Content-based methods attempt to create a user profile to predict ratings on unseen items. Successful content-based methods use tags and keywords. A measure of the usefulness of content-based filtering is typically calculated using heuristic functions, such as a cosine similarity measure. Content-based filtering can be used in many cases, as feature values can be easily extracted. Content-based filtering is not typically used in cases where feature values must be entered manually. This is manageable for small data sets, but when thousands of new products are added daily, this task becomes impossible. Content-based filtering does not require other users' data, as the predicted recommendations are user-specific. Thus, these techniques scale the system to handle multiple users. Content-based filtering is user-independent since this system only requires item and user profile analysis to obtain recommendations.

Research Hypotheses

Hypothesis 1: Integrating external information from sentiment analysis systems will improve the performance of book recommendation systems.

Hypothesis 2: The enhanced recommendation system will provide more accurate and diverse book recommendations to users.

Hypothesis 3: By addressing the "cold start" problem, the proposed system will be able to provide relevant recommendations even for new users with no previous behavior data.

Hypothesis 4: The integration of sentiment analysis factors will result in higher user satisfaction and improved decision-making when selecting books and reducing the data size facilitates easier handling and processing. All tables should be inserted in the main text article at its appropriate place

Research Methodology

The proposed recommendation system to enhance the quality of the recommendation system by incorporating sentiment analysis factors as features into the dataset, which will be used to train and improve the accuracy of the recommendation algorithm in Figure 1 illustrates the methodology and general scheme of this research.

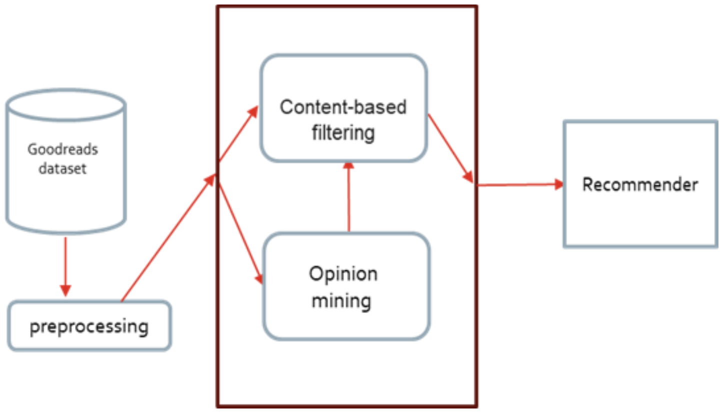

Figure 1: proposed recommendation system.

1. **Dataset**

The dataset will be obtained from the "kaggle" website, which contains comprehensive data on books, users, and ratings. This dataset called "Goodreads" will provide a rich source of information for training and improving the recommendation system and it included multiple relations as:

Books Table: Contains information about books, including book ID, title, author, genre, publication date, and description.

Ratings Table: Stores user ratings for books, including rating ID, user ID, book ID, rating date, and rating value.

Reviews Table: Captures user reviews, including review ID, user ID, book ID, review text, and review date.

Users Table: Holds user-related data, such as user ID, name, email, demographic information, and any other relevant attributes.

2. **Research Methodology**

Data Processing: Handle missing values, duplicate entries, and inconsistencies. Standardize formats and remove redundant information.

Content Filtering System Implementation: Current Suggestions Analysis: Implementing the content filtering system using TF-IDF technology and global similarity.

Primary Selection: Utilize book files and data from the Goodreads database. Database Search Enhancement: Incorporate sentiment analysis results into the database search.

3. **Feature Engineering**

Identify and select relevant features, such as book attributes, user interactions, and sentiment analysis results, for accurate recommendations. Interpretation of Book Feelings:

Comments Collection: Utilize available book comments from the database. Sentiment Analysis Algorithm Utilization: Implement algorithms like BERT or VADER to analyze the sentiment of each comment (positive, negative).

New Search Creation: Medium Effect Integration: Incorporate medium reactions and mitigate negative effects. Results Storage: Add results to a new search in the database, such as “average ratings”.

Filter books with more than 1000 reviews: Let's denote the set of all books as $\boldsymbol{B}$ and each book $\boldsymbol{bi}$ has $\boldsymbol{Ri}$ reviews. We define the set of books with more than 1000 reviews as:

$$B' = \{bi \in B \mid Ri > 1000\}$$

Filter positive reviews: For each book $\boldsymbol{bi}$ in the set $\boldsymbol{B'}$, we define the set of reviews $\boldsymbol{Ti}$ and filter the positive reviews based on sentiment analysis:

$$Ti = \ = \{\, tij \in Ti \mid S(tij) = 1\,\}$$

Where $\boldsymbol{S(tij)}$ represents the sentiment analysis result for review $\boldsymbol{j}$.

Calculate the ratio of positive reviews for each book: the number of positive reviews for each book $\boldsymbol{bi}$ in the set $\boldsymbol{B'}$ is , we calculate the ratio for each book as follows:

$$Ai = \mid Ti \mid \backslash Ri$$

Where $\boldsymbol{Ai}$ is the ratio of positive reviews to the total number of reviews for book $\boldsymbol{bi}$.

Calculate the average ratio across all books: to find the average ratio of positive reviews across all books in the set $\boldsymbol{B'}$ we sum the ratios and divide by the n number of books in the set:

$$A' = \frac{1}{\mid B' \mid^1} \sum_{bi \in B'}^{n} (Ai)$$

By combining these equations, we obtain the average number of positive reviews for books that have received more than 1000 reviews, given that the sentiment analysis result is positive.

4. Model Selection

Choose appropriate machine learning algorithms, such as collaborative filtering or content-based filtering, considering their strengths and weaknesses.

Model Training and Evaluation: Train selected models using the Goodreads dataset and assess their performance using metrics like precision, recall, and F1-score.

5. System Testing and Deployment

Test the enhanced system on a separate dataset and through user feedback to ensure effectiveness. Deploy the system on a user-accessible platform.

Continuous Improvement: Regularly collect user feedback, monitor system performance, and update the recommendation system based on user interactions and emerging trends.

Proposed System Model

Hybrid Model:
The proposed hybrid model combines Content-Based Filtering with Sentiment Analysis techniques to improve recommendation accuracy and relevance.

Content-Based Filtering:
Content-Based Filtering focuses on analyzing the content and attributes of books to make recommendations. In this model, book attributes such as title, author, genre, keywords, and other metadata will be processed to understand the context and theme of each book. This approach enables the system to recommend books similar to those that the user has shown interest in the past or books that align with their preferences based on the content.

Sentiment Analysis:
Sentiment Analysis is integrated into the model to incorporate user feedback and opinions. The system will analyze user-generated content, such as reviews, ratings, or comments, to extract sentiment scores or polarity (positive, negative, neutral) associated with each book. This additional layer of information helps the system understand the emotional response and satisfaction level of users towards specific books

Data Fields and Integration:
The hybrid model introduces new data fields derived from sentiment analysis. These fields can include sentiment scores, polarity, or any other relevant indicators extracted from user-generated content. These fields will be integrated into the existing dataset as additional attributes, enhancing the richness and in formativeness of the data.

Recommendation Filtering and Accuracy Measurement:
The hybrid model employs a filtering mechanism that combines content-based filtering with sentiment analysis. The system will generate recommendations by considering both the similarity in book content and the sentiment associated with user interactions. The accuracy of the recommendations will be measured again using appropriate evaluation metrics, such as precision, recall, F1-score, or Mean Average Precision (MAP).

The proposed hybrid model aims to provide more nuanced and contextually relevant book recommendations by leveraging the strengths of both content-based filtering and sentiment analysis. The integration of sentiment analysis adds an extra dimension to the recommendation process, potentially improving user satisfaction and the overall effectiveness of the system.

Scopes and Limitation
The constraints of this study are that it will be carried out utilizing a content filtering and sentiment analysis system, which incorporates elements from sentiment analysis into the dataset, while avoiding the use of collaborative filtering based on user similarity. The recommendation system in this study is based on aggregation of user behaviors and book usage, rather than aggregation based on user emotions.

This study's recommendation system is primarily founded on the process of aggregating user behaviors and book usage, rather than solely relying on the aggregation rooted in user emotions and feelings. The emphasis lies on analyzing and synthesizing the patterns of actions, preferences, and interactions exhibited by the users, in conjunction with their utilization of books, as the fundamental basis for generating recommendations. By focusing on objective data regarding user behavior and book utilization, rather than solely relying on subjective emotional factors, this recommendation system strives to provide personalized and accurate recommendations that align with the specific interests and preferences of the users.

Results
The experiment was conducted in different settings: no external information was used, some data were merged during preprocessing, data cleaning and deduplication were used, content-based filtering was used for analysis, and recall and accuracy were checked in Figure 4. There is a large amount of data that was processed to obtain real data for all authors and not to rely on book names or book numbers only, as the book ID and book name fields were adopted together as the primary key for books and all authors were displayed with the number of their publications as shown in Figure 2. As for the average and number of reviews conducted on books, it was observed that there were a large number of reviews for books that received a poor evaluation, which makes them unwanted data in our experiment as shown in Figure 3.

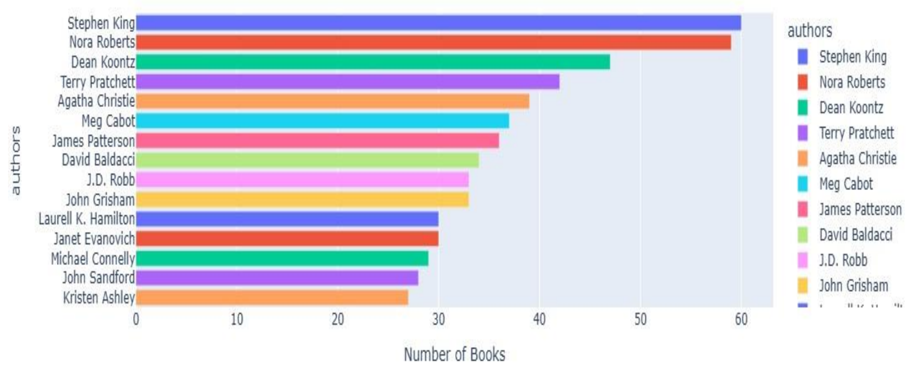

Figure 2: authors and N of books.

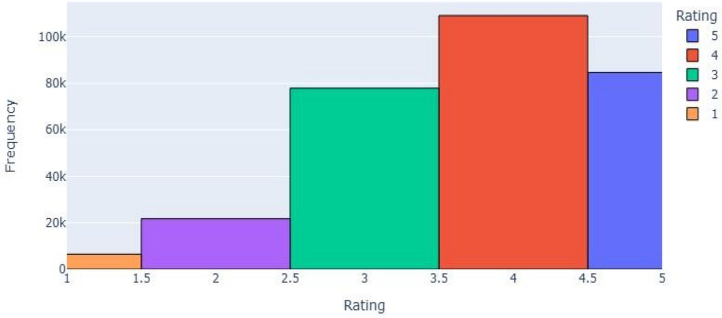

Figure 3: poor and high ratings.

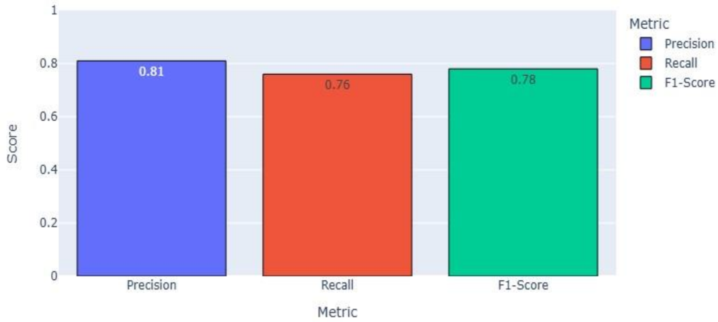

Figure 4: precision, recall, F1-score.

During the experiment, to further enhance the evaluation process, we introduced an additional step: integrating a

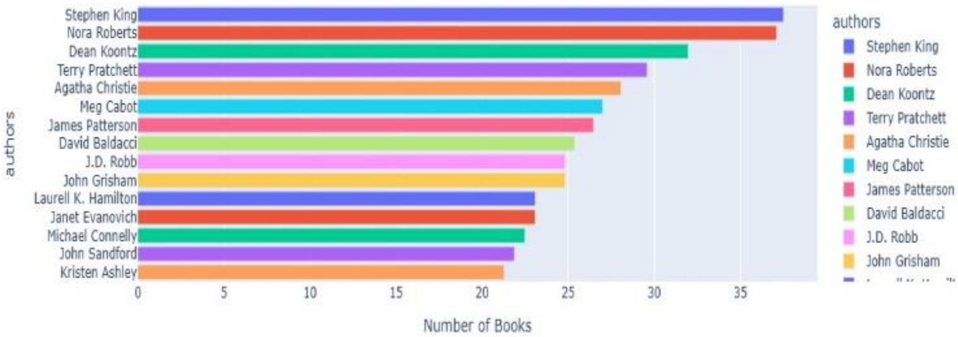

Figure 5: precision, recall, F1-score.

sentiment analysis column into the existing database. This column assigned a binary value of zero for negative sentiment and one for positive sentiment. By leveraging this sentiment analysis, we were able to effectively filter reviews using a zero-based filter, and the results showed that the number of books to be included in the evaluation was reduced, as shown in Figure 5. This fine-tuning process aimed to improve the dataset and eliminate any potentially misleading data points, ensuring that a book with a large number of negative reviews would not be retained even if it had a high rating, which resulted in a boost in the average rating, as shown in Figure 6. Furthermore, the filtered data was reintegrated into the recommendation system, paving the way for new queries and enabling us to calculate precision and recall with greater accuracy, as shown in Figure 7, which shows an improvement in the values of the metrics by which the model's performance was measured. This iterative approach allowed us to continually improve our recommendation system. By implementing these comprehensive measures, we ensured a robust and reliable experimental process, leading to accurate and meaningful results. Each step has been carefully designed to address specific aspects of our research objectives, ensuring a comprehensive assessment of the effectiveness, accuracy, and recall of the recommendation system.

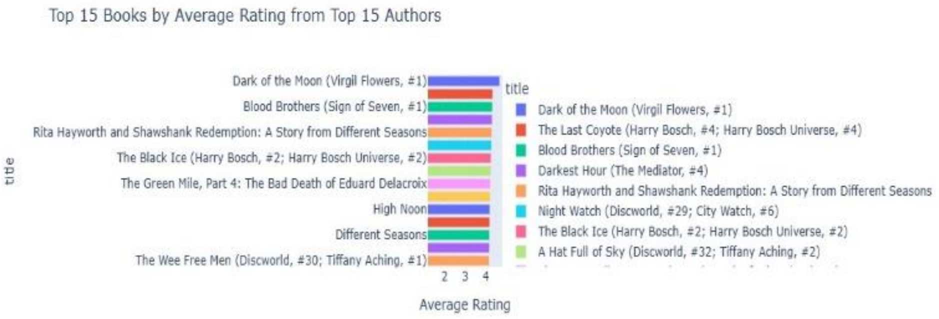

Figure 6: Average rating.

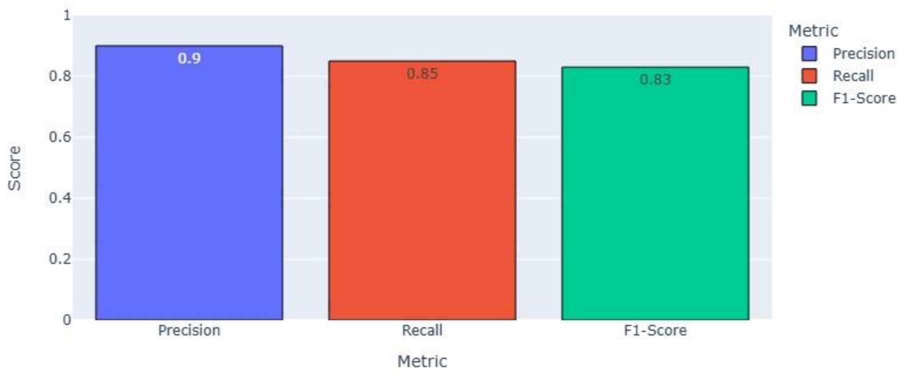

Figure 7: Accuracy after adding sentiment attribute.

Discussions the result

In this paper, we demonstrate how book recommendations can be improved by using such external information. We extract sentiment analysis factors from user comments using a predefined set of English and book-specific terms and a detailed analysis of the Goodreads dataset. We use these factors to extend two classical recommendation approaches: user-based and item-based. Via an extensive set of experiments, we demonstrate that both simple extensions are indeed able to significantly enhance the recommendation accuracy. Furthermore, our experiments enable us to uncover the conditions under which our proposed approach leads to the greatest improvements.

The hybrid model was tested in five different configurations: without any external information, with metadata, with textual review data, with both textual review data and metadata, and using only external information. The results of the evaluation showed that, in identical conditions, the implementation utilizing external information, which included sentiment analysis, was able to provide more accurate book recommendations to the user. When considering potential practical applications, the recommendation system that used an equal combination of external information demonstrated higher accuracy in recommending new books. The experiment involved a comprehensive recommendation system that initially analyzed which recommendation source best reflected a user's personal preference. The findings indicated that when all recommendation factors were combined, the accuracy and popularity of the sentiment model's top-three recommendations were comparable to those of the rating model. Sentiment analysis expanded the historical knowledge of user-item ratings and behaviors, revealing that a user who frequently posted negative comments might still rate some books positively. Additionally, it was observed that a book with a very high rating was not necessarily widely liked, as the skewed positive rating distribution suggested that the total number of users who favored the book was relatively small.

Future Work

In future work, in order to gain a deeper understanding of review sentiment factors, it is necessary to explore and analyze not only the datasets derived from Goodreads, but also integrate other different review datasets. Moreover, it is very important to delve deeper into the obtained results and analyze the sentiment of the social aspect comprehensively. Moreover, leveraging the size of the dataset to support the book recommendation system will prove to be an indispensable step forward. In a broader sense, the concerted efforts made in these areas will undoubtedly play a pivotal role in solving the irrelevance problem and greatly improving the effectiveness of the book recommendation system. Also, the development of a hybrid system that includes all book recommendation techniques and sentiment analysis factors, especially those with bias, will be added and studied to extract the values and factors useful for recommendation systems.

Conclusion

This study investigates the popular method of book recommendation, content-based filtering, and proposes an improved approach to personalized book recommendation. The optimization combines emotion scores and emotional items based on part-of-speech tags with an item-based collaborative filtering method. Five distinct features based on the emotional part-of-speech tag were identified and used for recommendations. Experiments were conducted using the Goodreads dataset, demonstrating the highly successful performance of the proposed method compared to baseline methods. In addition, influential emotional part-of-speech tags are analyzed within the full recommendation process, which has not been explored previously. The proposed approach demonstrates that sentiment features not only enhance recommendation performance but also improve the model by incorporating external features. The method outperforms and competes with basic methods across different evaluation metrics. As future work, in addition to factors based on emotional content, other context-aware features such as position on the page and the time at which the item is evaluated will be evaluated.

References

[1] Schnabel, T.; Bennett, P.N.; Joachims, T. Improving recommender systems beyond the algorithm. arXiv 2018, arXiv:1802.07578.

[2] Hassanpour, N. Abdolvand, and S. RajaeeHarandi, “Improving Accuracy of Recommender Systems using Social Network Information and Longitudinal Data,” Journal of AI and Data Mining Vol. 8, No. 3, pp. 379-389, 2020.

[3] Kun Zhou, Wayne Xin Zhao*†, Shuqing Bian, Yuanhang Zhou, Ji-RongWen†, and Jingsong Yu. 2020. Improving Conversational Recommender Systems via Knowledge Graph-based Semantic Fusion. In26th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD ’20), August 23–27, 2020, Virtual Event, USA.ACM, New York, NY, USA.

[4] Sundermann,C.;Domingues,M.;Sinoara,R.;Marcacini,R.;Rezende,S. Using Opinion Mining in Context-Aware Recommender Systems: A Systematic Review.Information2019,10,42.https://doi.org/10.3390/info10020042Aware Recommender Systems: A Systematic Review.Information 2019,10,42.

[5] Sallam,R.;Hussein,M.;Mousa,H.An Enhanced Collaborative Filtering-Based Approach for Recommender Systems.Int. J. Comput. Appl. 2020,176, 9–15.

[6] M. Ashok, S. Rajanna, P. V. Joshi and Sowmya Kamath S, "A personalized recommender system using Machine Learning based Sentiment Analysis over social data," 2016 IEEE Students' Conference on Electrical, Electronics and Computer Science (SCEECS), Bhopal, India, 2016, pp. 1-6, doi: 10.1109/SCEECS.2016.7509354.

[7] H. Ko, S. Lee, Y. Park, and A. Choi, "A survey of recommendation systems: recommendation models, techniques, and application fields," Electronics, 2022.

[8] Z. Fayyaz, M. Ebrahimian, D. Nawara, A. Ibrahim, et al., "Recommendation systems: Algorithms, challenges, metrics, and business opportunities," applied sciences, vol. 2020. mdpi.com, 2020.

[9] Gong, S.; Cheng, G. Mining user interest change for improving collaborative filtering. In Proceedings of the 2008 Second International Symposium on Intelligent Information Technology Application, Shanghai, China,21–22 December 2008; Volume 3, pp. 24–27.

[10] Schafer, J.B.; Frankowski, D.; Herlocker, J.; Sen, S. Collaborative filtering recommender systems. In The Adaptive Web; Springer: Berlin/Heidelberg, Germany, 2007; pp. 291–324add alternative text to a picture in Word 2010.

[11] P. J. Feng, P. Pan, T. Zhou, H. Chen, C. Luo, "Zero shot on the cold-start problem: Model-agnostic interest learning for recommender systems," in Proceedings of the 30th ACM, 2021.