

Enhanced Content-based Recommendation Method

Sampath Kini K¹ and Karthik Pai B H²

¹NMAM Institute of Technology (Nitte Deemed to be University)/CSE Department, Nitte, Karkala, India
Email: sampath@nitte.edu.in

²NMAM Institute of Technology (Nitte Deemed to be University)/ISE Department, Nitte, Karkala, India
Email: karthikpai@nitte.edu.in

Abstract— The intuitive and easily demonstrable benefits of content-based recommendation technology have made it widely adopted in education and e-commerce. However, traditional content-based recommendation methods face a challenge when it comes to analysis with regard to item modeling. These methods fail to consider the need for data parallelism technique upon large contents while producing structured data items. The objective of this research is to enhance the content analysis capability of traditional content-based recommendation algorithms by incorporating a multithreaded information vector space model for item modeling during content retrieval. Several experiments were carried out to assess the efficacy of this approach. The findings from these experiments indicate that the algorithm proposed in this paper exhibits satisfactory performance across different scenarios.

Index Terms— content-based recommendation, knowledge representation, Web Crawler, vector space and user preferences

I. INTRODUCTION

As information technology continues to advance, there is a growing focus on building the digital economy, digital society, and digital government. This progress is driving a transformation in production methods and overall lifestyle, emphasizing the importance of digital transformation in today's social development. Recommendation algorithms play a crucial role in unlocking the potential of data and uncovering valuable insights existing in web content. By analyzing user interactions such as browsing, collecting, purchasing, and considering item attributes, recommendation systems generate lists of recommended items which users are most likely to find very interesting. Content-based recommendation technology is widely applied in various fields, including e-commerce, education, and entertainment, due to its intuitive nature and ability to provide meaningful explanations on its advantages. The entire article is organized as follows. We have captured related work in this domain in section II. In section III, enhanced content based recommendation methodology is captured. In section IV, experimental procedure is captured. In section V, experimental results obtained during our study are discussed. In section VI, conclusion on our study on a new method of recommendation system is drawn.

II. RELATED WORK

The objective of the recommendation system is to analyze user preferences and provide suggestions for items that align with their interests, taking into account their past behavior and current application context. Various information service platforms have widely adopted this technology for extensive utilization[1, 2] like e-

commerce[3], social networks, and news websites[4]. In 2000, Mooney created a book recommendation system called Libra. This system used the content-based recommendation method and obtained book information from the Amazon e-commerce website. It constructed a model based on the book's content and generated recommendations using that model. Similarly, in 2004, Middleton developed Quickstep, an online academic research paper recommendation system. Quickstep also employed the content-based recommendation technology. It constructed a user interest model based on the classification information of papers viewed by the user. By calculating the similarity between the user's interests and the characteristics of papers, Quickstep generated paper recommendations. In 2009, Khribi et al. created a recommendation system that analyzed learners' browsing history to establish their preferences. It generated resource recommendations by including the contents of learning resources. In 2017, Soulef Benhamdi et al. developed NPR-EL, a personalized learning material recommendation system. It utilized collaborative filtering and content-based recommendation technology to aid learners using personalized learning materials within the learning environment. Sunandanad et al. designed a movie recommendation system in which they enhanced the content-based filtering algorithm by considering user demographic data[5]. Bagul et al. introduced another content-based recommendation method for literature recommendation[6]. Their method utilized LDA topic modeling. Tai et al. designed content-based recommendations using machine learning techniques[7]. After reviewing relevant literature, it has been noted that content-based recommendation technology is widely utilized in the field of education [8] due to its intuitive nature and ease of explanation. In this paper, the research on course recommendation also relies on the content-based recommendation method [9]. However, traditional content-based recommendation technology has a limitation related to speed and scalability in item modeling. The TF-IDF vector space models used in these approaches lack robustness. As a result, items in content are not efficiently discovered. Therefore, the objective of this research is to enhance the traditional content-based recommendation algorithm by incorporating data parallelism technique.

III. ENHANCED CONTENT RECOMMENDATION METHODOLOGY

A. Content-based recommendation model

The primary goal of the content-based recommendation algorithm is to suggest items to users based on their interests and preferences. It achieves this by identifying items that share similar characteristics to the ones the user has shown interest in or liked before. This algorithm relies on constructing feature models for both the items and users involved[10]. The recommendation process involves comparing the attributes of the items with the user's interests, identifying the items that are closely related to the user's preferences, and generating a list of recommendations accordingly. Fig.1 provides an illustration of the structure of the content-based recommendation approach.

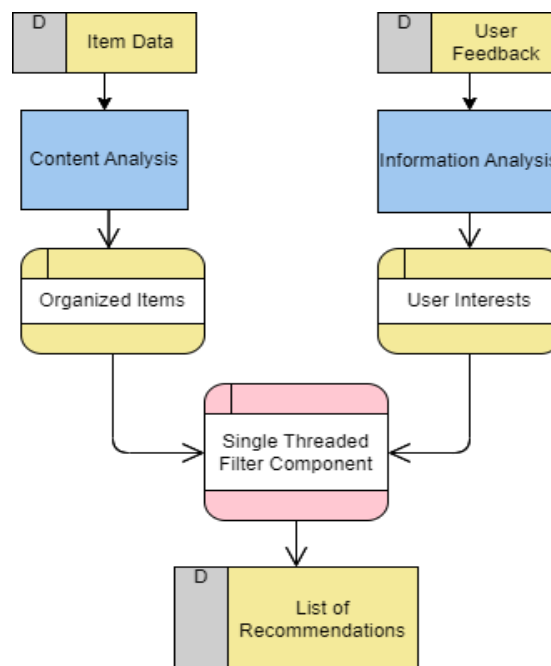


Figure 1. Content-based recommendation model.

The content analyzer is commonly employed to create models of object features. Within the content-based recommendation algorithm framework, this approach heavily relies on the descriptive attributes of items, which can be in the form of structured or unstructured data. To deal with unstructured data, it undergoes preprocessing to extract structured information and transform the original raw data into a specific format. This process is typically achieved through the utilization of a TF-IDF (Term Frequency-Inverse Document Frequency) dimensional space model, which helps organize and represent the data effectively [11].

The information learner acts as a tool for gathering user interaction data to build user models. In cases where the system directly asks users to check for the goods or services they receive, the collected feedback is treated as explicit feedback. At the same time, the system also observes and analyzes users' online behaviors to extract information about their interests, which is treated as implicit feedback. Implicit feedback is typically derived from users' online activities without requiring their active participation. In the context of this paper, the recommendation model focuses on modeling users' interests and generating recommendation lists based on the implicit feedback obtained from users' online course browsing behaviors.

The filtering component plays a crucial role in the recommendation system by comparing user interest information with item description information with the help of similarity measurement techniques like cosine similarity, Euclidean distance, and others. The matching results obtained are basically classified into two types: binary correlation and continuous correlation. Binary correlation captures user preferences as either liking or disliking an item. On the other hand, continuous correlation takes into account the degree of possible interest and generates a recommendation list in order of the level of potential interest. However, this traditional method of filtering suffers from performance issues with the larger data store containing organized data items built by content analyser. We will have to explore the opportunity to enhance this scenario to tackle the performance issues very well.

B. Enhanced Content-based recommendation model

The content-based recommendation algorithm utilizes a TF-IDF vector space model to extract item features, and the calculation of item similarity relies on cosine similarity. The TF-IDF vector space model operates by considering the term-frequency statistics within a corpus. It assumes that each feature item exists independently, with each dimension of the vector representing a specific feature item. The weight of feature items across all dimensions forms the item's feature vector representation. This approach allows for a comprehensive understanding of the item's characteristics, offering advantages such as straightforward interpretation and ease of use. One limitation of the current approach is the absence of considering the influence of data parallelism during the calculation of item content similarity. As a result, there is room for improvement in the execution time during generation of recommendation lists. To address this, multi-threading technology offers a valuable advantage by efficiently processing the data contents. By incorporating data parallelism techniques, we can enhance the recommendation engine speed by leveraging the underlying CPU cores. We have integrated data parallelism technology with the TF-IDF vector space model for item modeling and similarity calculation. This novel recommendation algorithm exploits the strengths of underlying processing cores. For a visual representation of the algorithm model, please refer to Fig. 2. In this approach, we have partitioned the organized items into equal sized sub partitions. Each of the individual sub partitions is processed by an separate threaded instance of filter component. This way all the sub partitions are processed simultaneously by the respective instance filter component in a separate threaded execution environment. This enables processing of organized items simultaneously on underlying multicore processing units.

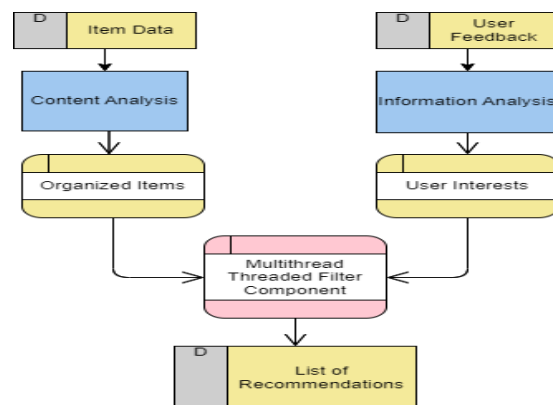


Figure 2. Enhanced Content-based recommendation model

IV. EXPERIMENTAL PROCEDURE

To validate the efficiency of the enhanced approach, we performed experiments using MOOC datasets. Given the nature of the content-based recommendation algorithm, we utilized different text data elements such as course titles, course overviews, and course teaching objectives. Through comparative experiments, we examined the variations in execution time trends when applying the enhanced recommendation algorithm. This enabled us to make speed comparisons between the old method and proposed enhanced approach on the datasets. Specifically, we used course titles, course overviews, and course teaching objectives as single data structure, with attributes as TF-IDF (course-title), TF-IDF (course-overview), and TF-IDF (teaching-objective) respectively. These experiments aimed to validate the impact and benefits of incorporating data parallelism into the recommendation process.

The experimental dataset was obtained using the scrapy crawler framework, which collected data from the icourse website. The dataset consists of various courses spanning a wide range of disciplines such as computer science, economic management, psychology, foreign languages, literature, history, arts, engineering, science, life science, philosophy, law, ideological and political education, pedagogy, innovation, entrepreneurship, and more. The crawled data primarily includes course titles, course overviews, and links to course pages, all in text format.

The design elements of the enhanced content-based recommendation technique is depicted in Fig. 3. The performance comparison of the enhanced recommendation method with the old approach is depicted in Fig. 4. We can note from the experiments that the speed of the recommendation process is related to the size of the data elements present in input content. In the experiment, we set the length range of the list to 10, which means recommending at least 1 similar course and at most 10 similar courses.

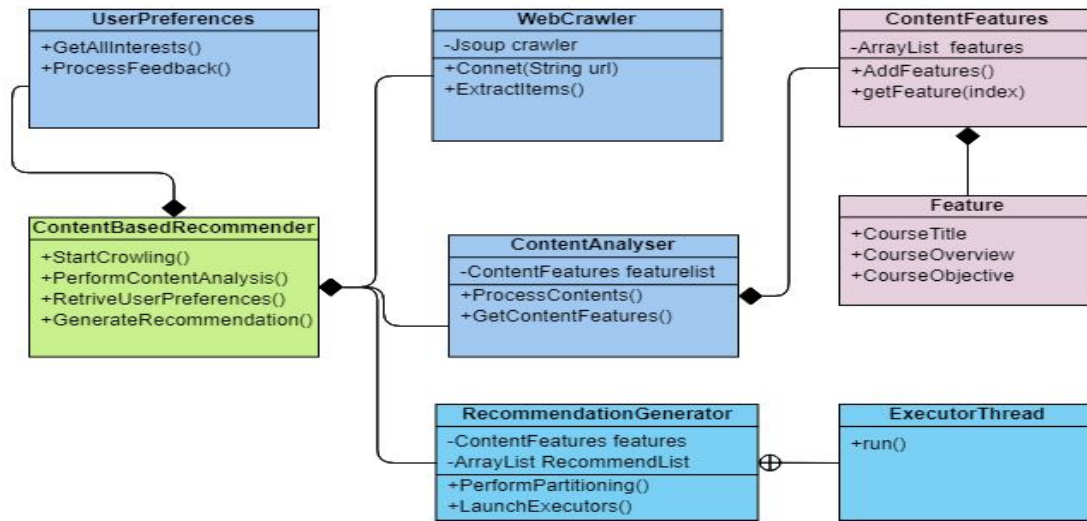


Figure 3. Enhanced Content-based recommendation model

V. RESULTS AND DISCUSSION

We performed experiments using MOOC datasets created using icourse website[12]. Total of 1447 courses are made available on this website. Given the nature of the content-based recommendation algorithm, we stored different text data elements such as course titles, course overviews, and course teaching objectives into an xls file. Through comparative experiments, we examined the variations in execution time trends when applying the enhanced recommendation algorithm. In the experiment, we have implemented the classic content-based recommendation algorithm and improved content recommendation algorithm through Java programming language. The experimental results are shown in Fig. 4. The data shows that the recommendation efficiency of the improved content recommendation algorithm is better compared with the classical content-based recommendation algorithm. The performance of the enhanced content recommendation algorithm rises by 27%. It is seen that it is efficient if we enhance the classic content-based recommendation algorithm by integrating data parallelism into item modeling through multithreading technology. Therefore experimental results also prove the rationality of the improvement.

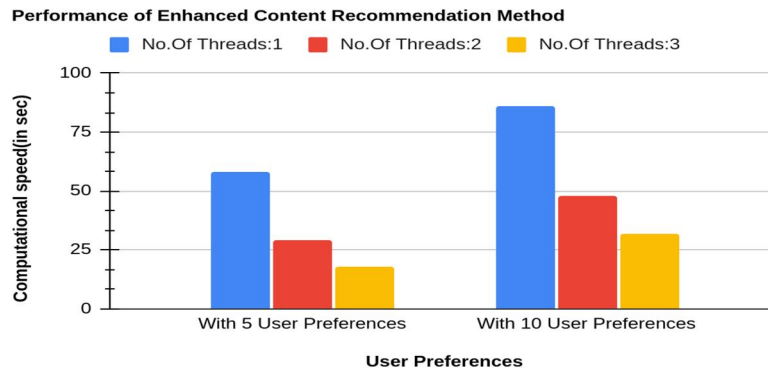


Figure 4. Performance Comparison Of Enhanced Content-based recommendation model

VI. CONCLUSIONS

Addressing the limitations of the classic content-based recommendation algorithm in terms of processing, this paper introduces an enhanced content recommendation algorithm that incorporates data parallelism. The data items are processed with TF-IDF technique to generate item features. The algorithm calculates item similarity reasonably faster, resulting in an improved item recommendation method that enhances both speed and scalability of the proposed method. In order to verify the effectiveness of this improved recommendation method, a series of comparative experiments were conducted using a specifically selected dataset. These experiments offer empirical evidence of the algorithm's scalability and its capability to address the shortcomings of conventional approaches.

REFERENCES

- [1] Chang, jia-wei, et al, "music recommender using deep embedding-based features and behavior-based reinforcement learning", multimedia tools and applications (2021): 1-28.
- [2] Wu, zonghan, et al, "a comprehensive survey on graph neural networks.", iee transactions on neural networks and learning systems, 32.1 (2020): 4-24.
- [3] Felfernig, alexander, et al, "an integrated environment for the development of knowledge-based recommender applications", international journal of electronic commerce, 11.2 (2006): 11-34.
- [4] Girsang, abba suganda, sani muhamad isa, and ikrar harvy, "recommendation system journalist for getting top news based on twitter data.", journal of physics: conference series, vol. 1807. No. 1. Iop publishing, 2021.
- [5] Sunandana, g., et al, "movie recommendation system using enhanced content-based filtering algorithm based on user demographic data.", 2021 6th international conference on communication and electronics systems (icces). Ieee, 2021.
- [6] Bagul, dhiraj vaibhav, and sunita barve, "a novel content-based recommendation approach based on lda topic modeling for literature recommendation.", 2021 6th international conference on inventive computation technologies (icict). Ieee, 2021.
- [7] Tai, yifan, zhenyu sun, and zixuan yao, "content-based recommendation using machine learning.", 2021 ieee 31st international workshop on machine learning for signal processing (mlsp). Ieee, 2021.
- [8] Dascdascalu, maria-iuliana, et al, "educational recommender systems and their application in lifelong learning.", behaviour & information technology 35.4 (2016): 290-297.alu mi, bodea cn, mihailescu mn. Educational recommender systems and their application in lifelong learning. Behav inf technol. 2016;35(4):290–
- [9] Bagher, rahimpour cami, hamid hassanpour, and hoda mashayekhi, "user trends modeling for a content-based recommender system.", expert systems with applications 87 (2017): 209-219.
- [10] Huang, ran, and ran lu, "research on content-based mooc recommender model." 2018 5th international conference on systems and informatics (icsai), pp. 676-681,ieee, 2018.
- [11] Rani, ujjwal, and karambir bidhan, "comparative assessment of extractive summarization: textrank tf-idf and lda.", journal of scientific research 65.1 (2021): 304-311.
- [12] <https://www.icourse163.com/courses/all/> last accessed 20/june/2023