

---

# A Serendipity Recommendation Method for Book Categories Using BERT to Strengthen the Web Service of the Book

---

Youngmo Kim, Seok-Yoon Kim and Byeongchan Park\*

*Dept. of Computer Science and Engineering, Soongsil University, Republic of Korea  
E-mail: ymkim828@ssu.ac.kr; ksy@ssu.ac.kr; pbc866@gmail.com*

*\*Corresponding Author*

Received 29 May 2024; Accepted 08 November 2024

## Abstract

In the field of book search, research on a web service-based user-customized book recommendation system is being conducted to respond to increasingly diverse user requirements. The collaborative filtering algorithm, which is mainly used for book recommendation, has a problem in that it is difficult to reflect the user's recent interest without considering the changes in preference over time, and the user's satisfaction decreases because it repeatedly recommends only similar items.

In this paper, we propose a book recommendation method using category similarity based on deep learning. The proposed method is to predict books to be used next time by inputting users' past and current book usage history through BERT, a natural language processing model, and to recommend popular books in other categories with high similarity to the predicted book category in the BERT model to reflect serendipity. This method reflects serendipity, which can lead to users' recent interests and practical preferences, so that recommendation accuracy and user satisfaction can be satisfied at the same time.

**Keywords:** Web service, book category, BERT, serendipity, recommendation.

*Journal of Web Engineering, Vol. 24\_2, 199–216.*

doi: 10.13052/jwe1540-9589.2422

© 2025 River Publishers

## 1 Introduction

Due to the recent development of mobile devices and communication environments, various products and content can be consumed on platforms in the mobile-based web services environment. Accordingly, in order to increase sales and the number of users, the importance of technology that recommends products or content that may be of interest based on users' preferences and past behavior is growing [1, 2]. This recommendation technology has collaborative filtering algorithms and is used in various platforms in the mobile-based web services environment, such as OTT-based movie and drama recommendations, open-market-based product recommendations, and video sharing platform-based video recommendations. The collaborative filtering algorithm applied to existing book recommendations can be recommended by people with similar tastes based on the similarity of users and items [3]. However, as that does not take into account changes in preferences over time, it is not possible to know exactly what product the user is most interested in, and only recommends books of a type similar to those frequently used by users, so the users' satisfaction may decrease [4, 5]. Therefore, accurate recommendations that can be applied to real-world environments require recommendations that model contextual book preference patterns derived from users' past book use history and reflect serendipity, an indicator that provides new and satisfying items that they have not encountered before [6].

To address this problem, in this paper, we use the masked language model (MLM) of the deep learning model bidirectional encoder representations from transformers (BERT) [9]. In addition, we propose a serendipity recommendation method for book categories using BERT, which can recommend new books because it only provides consecutive recommendations from the past and recommend popular books in other categories with high similarity to the predicted book categories that reflect unexpectedness that can lead to users' practical preferences.

The composition of this paper is as follows. Following this introduction, Section 2 describes a sequential recommendation method based on a natural language processing deep learning model, collaborative filtering, which is an existing recommendation method, and the BERT model, a technique applied as a recommendation method in this paper. Section 3 describes the serendipity recommendation method for the book category using BERT proposed in this paper in the sequence of a preprocessing process, a sequential recommendation method, and an unexpected recommendation method. Section 4 reviews the experimental results, and Section 5 concludes with a conclusion.

## **2 Related Research**

### **2.1 Sequential Recommendation Method Based on a Natural Language Processing Deep Learning Model**

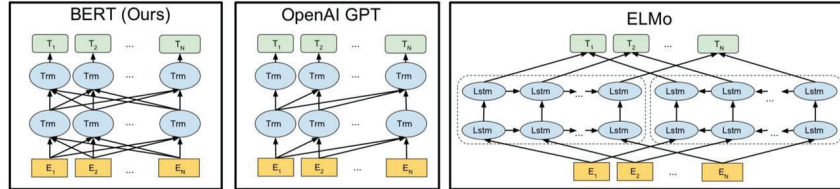
In a real environment, interaction between a user and an item appears in a consecutive sequence rather than independently, so the type of item preferred by the user may change over time [7]. Recently, a sequential recommendation method [8] that recommends the user's most preferred item at the current point in time by considering the order of items consumed by the user using a natural language processing deep learning model has been studied. Hidasi and others [9] have suggested that the sequence data of the items consumed by the user passes through the gated recurrent unit (GRU) layer, a recurrent neural network (RNN)-based algorithm, learns contextual relationships for all items in the sequence, and returns a preference score for the next item as the final output, showing better performance than the collaborative filtering recommendation method.

However, the RNN-based natural language processing recommendation system has the problem of gradient vanishing, in which the data of the past is lost as the input data becomes longer. To compensate for this, although session-level RNNs, which model item usage history within a session, and HRNN-based recommendation methods [10] consisting of two layers of user-level RNNs that model usage pattern changes between sessions and propagate those values to the next session, have been studied, the accuracy of recommendations is still low in fairly long data.

### **2.2 Collaborative Filtering**

Collaborative filtering [1, 11, 12] is most widely used to provide personalized recommendation services by calculating the similarity between users or items based on the assumption that groups with similar tastes for one item may have similar tastes for another. It is most widely used when nearest neighbor collaborative filtering, which is the most basic technique of collaborative filtering, can be divided into user-based collaborative filtering and item-based collaborative filtering according to the similarity measurement target.

User-based collaborative filtering is a method of recommending items that are commonly preferred by users with a high degree of similarity to customers by obtaining similarities between target customers and users. Item-based collaborative filtering is a method of predicting the rating a target customer will have for another item based on the rating of the item with high similarity by measuring the similarity between items previously evaluated by users.



**Figure 1** Structural comparison of the BERT language model and other language models.

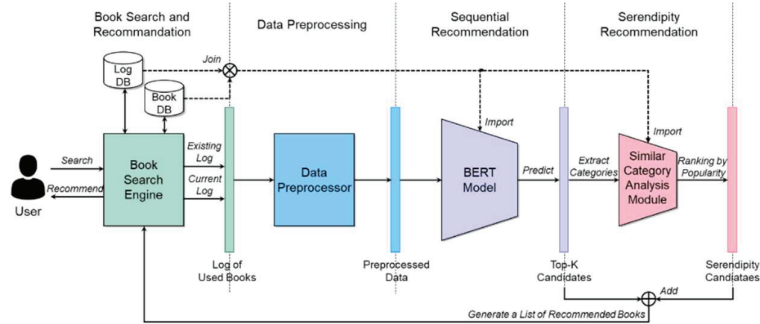
### 2.3 BERT

BERT [13] is a natural language processing deep learning model that performs fine tuning tasks that readjust hyperparameters together with additional training for tasks such as document classification, question answering, and translation after pre-training a language model using a large amount of unlabeled text data. Although it takes a lot of resources and time to directly train a large amount of data, the BERT model basically provides a pre-trained model for embedding a large number of words, so it performs many natural language processing tasks with relatively few resources. In addition, as shown in Figure 1, the entire context of a sentence can be used in both directions with a structure in which several transformer encoder blocks are overlapped, so it is more advantageous for modeling sequence data than a unidirectional language model.

## 3 Serendipity Book Recommendation Method Using Category Similarity

### 3.1 Overview of Serendipity Book Recommendation Using Category Similarity

In this paper, we use the masked language model of BERT, a deep learning model, and a method that reflects unexpectedness that can lead to user's actual preference by recommending new books that have not been used consecutively in the past. In addition, it uses a serendipity method that allows popular books in other categories that have a high similarity with the predicted book category to be recommended together. The proposed serendipity book recommendation method using deep learning-based category similarity consists of a book search and recommendation process, a preprocessing process, a sequential recommendation process using BERT, and a surprise recommendation process using serendipity, as shown in Figure 2.



**Figure 2** Serendipity book recommendation method using category similarity based on deep learning.

**Table 1** Metadata items of books

Item	Explanation	
User information	User code	Identification code of book user
	UCI	UCI code to identify used books
	Keyword	Categories of books used
	Use date	Year, month, day of use of books
Used book feedback information	Use form	Type of use of books (purchase/inquiry)
	Grade	Ratings given to books used

### 3.2 Data Preprocessing Process for Learning

The data preprocessing process proposed in this paper preprocesses the data to be input to the sequential recommendation process based on the book user’s previous book use history and the information of the book currently being used in the book search and recommendation process. The metadata items of books used by the user are metadata items used to extract preprocessing data for application to the BERT learning model, and are log information when the book user used the book in the past, as shown in Table 1.

The data preprocessing process for learning enables book users to classify and analyze according to user code criteria based on previously used book information, select and learn only preferred books for each user, and convert them into sequences based on the date and time of use. This process is shown in Figure 3.

First, after confirming the usage pattern of the book, whether the user prefers the book is normalized in a binary format of 0 (non-preference) and 1 (preference). Users can give ratings between 0 and 5 for purchased books, and the distribution of these ratings is different for each user. Therefore, if

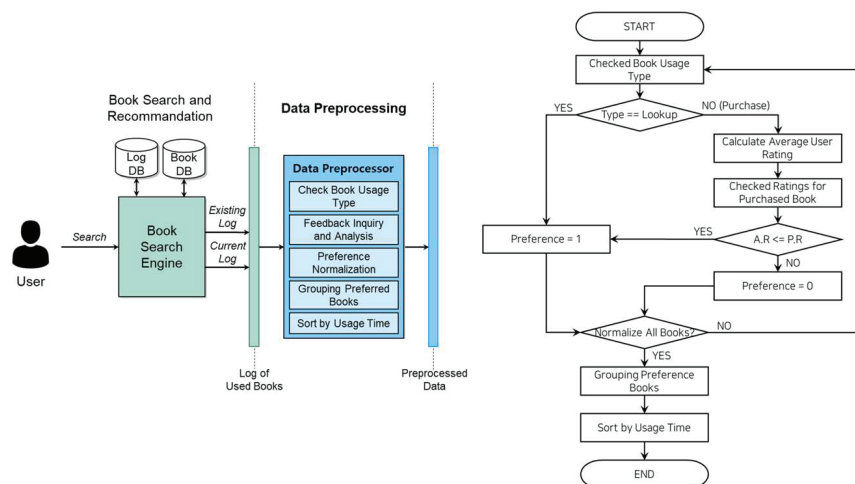


Figure 3 Flowchart of data preprocessing process for training.

the book usage type is “purchase”, books that have given each user a rating above the average rating are judged as preference books. After normalizing the preference of all books, only the preferred books among the books used by each user are grouped based on the user code, and, finally, the list of preference books grouped by user code is sorted in order of the date of use to complete the preprocessing process.

### 3.3 Sequential Book Recommendation Process Using BERT

The sequential book recommendation process using BERT is the process of generating K book lists so that users can recommend books that are more likely to be used next time, as shown in Figure 4.

This process is a method to reflect only the recently used books among the books used by the book user from the past to the present, and the data input to the BERT model is entered with a fixed maximum length. In this paper, a fixed maximum length of 200 is set, and the preprocessed book user’s preference list is searched through a preprocessing process. If the length is different from the set length, the length is adjusted either leaving only the most recent 200 favorite books or by the zero padding technique. Figure 5 shows the adjustment process when the maximum length is 6.

The preference list that has gone through the length adjustment process is converted into an embedding vector that can learn the BERT model and predict the result. At this time, in order to apply the existing BERT model

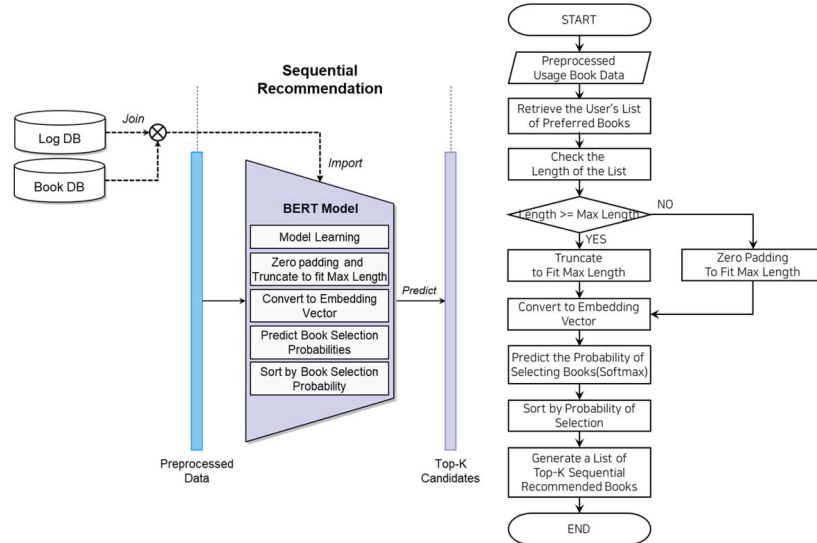


Figure 4 Flowchart of the sequential book recommendation process using BERT.

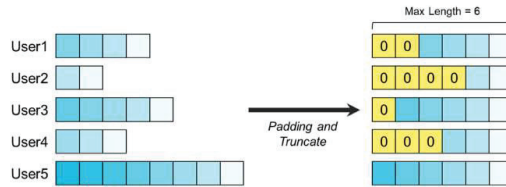


Figure 5 Length adjustment of the BERT model input data.

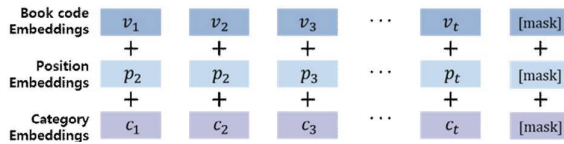
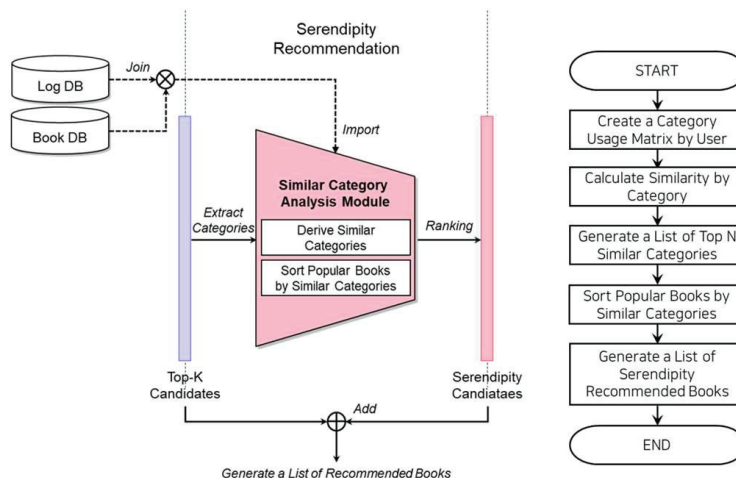


Figure 6 BERT input embedding configuration for book recommendation.

specialized for natural language processing tasks to book recommendation, token embedding, which is a word embedding vector, and segment embedding that plays a role in distinguishing sentences are first removed, and then it is configured to add book code embedding and position embedding indicating the order of book use, and book category embedding, and the last token is replaced with [MASK] to predict the book to be selected at the next time, as shown in Figure 6.



**Figure 7** Flowchart of serendipity book recommendation using category similarity.

If a user inputs the list of preference books converted into embedding vectors into the learned BERT model, the BERT model predicts a probability value between 0 and 1 to select the book at the next time for all other books that do not exist in the list of preference books, and creates a list of  $K$  sequential recommended books in the order of the highest predicted choice probability.

### 3.4 Serendipity Book Recommendation Process Using Category Similarity

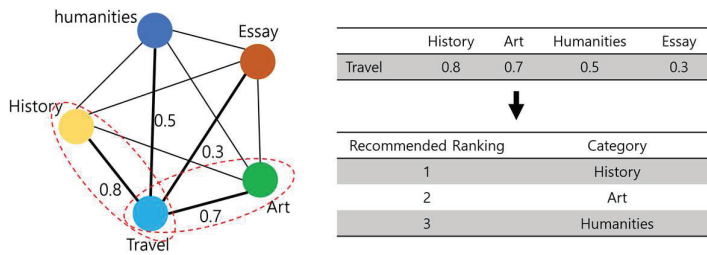
The serendipity recommendation process using category similarity is the process of generating a list of popular books in the category similar to the category of top- $K$  candidate books generated in the sequential recommendation process to provide book recommendations reflecting serendipity that can lead to users' practical preferences. Figure 7 illustrates the flowchart of this process.

First, in order to derive the top  $N$  categories similar to the categories of sequential recommended candidate books, users can calculate the similarity between each category vector in the matrix by organizing the number of books used by category, as shown in Figure 8.

Second, among the methods of measuring vector similarity, the value between 0 and 1 is extracted and the similarity between all category vectors is applied by applying the cosine similarity function that does not require

	User 1	User 2	...	User N	Calculate similarity between two vectors
Travel	15	0	...	1	
History	7	1	...	1	
Art	6	4	...	0	
...	...	....	...	...	
Essay	4	2	...	3	

**Figure 8** An example of a matrix of book usage in categories by user.



**Figure 9** Example of creating a similar category list.

a normalization process and can be accurately measured even in multidimensional vectors. In Equation (1),  $\text{CosSim}(A, B)$  represents the similarity obtained by measuring the cosine distance of category vector  $A$  and category vector  $B$ , and  $A_i$  and  $B_i$  mean the number of times user  $i$  used categories  $A$  and  $B$ , respectively.

$$\text{CosSim}(A, B) = \cos(\theta) = \frac{A \cdot B}{\|A\| \|B\|} = \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n A_i^2} \sqrt{\sum_{i=1}^n B_i^2}} \quad (1)$$

Third, among the list of top-K candidate books generated through the sequential recommendation process, N similar categories with the highest similarity to categories with high frequency are obtained. For example, if the category of a book recommended the most in the sequential recommendation top-K list by a specific user is “travel”, then “history”, “art”, and “humanities” categories with the highest similarity to travel are derived into similar categories, as shown in Figure 9.

Lastly, after sorting the books according to the degree of popularity for each of the derived similar categories, a list of serendipity recommendation books arranged in the order of highest user popularity is created with the most popular books.

**Table 2** Experiment environment

Experiment Environment	Specification
CPU	Intel Core i9 12900K
RAM	32 GB
GPU	NVIDIA Geforce RTX 3090
OS	Linux Ubuntu 20.40 LTS
Programming language	Python 3.9.13
Algorithm framework	Tensorflow 2.7.0

**Table 3** Example of user book log data

User-code	Book-code	Category	Timestamp	Type	Rating
A0459829SI5B5WIOZZHX	6545793	Literature & Fiction	1386547200	P	5
A0459829SI5B5WIOZZHX	7181701	Literature & Fiction	1373500800	P	5
A0459829SI5B5WIOZZHX	7230206	Literature & Fiction	1367452800	P	2
A0459829SI5B5WIOZZHX	7441282	Literature & Fiction	1364169600	P	1

## 4 Experimental Results

### 4.1 Experiment Environment and Method

In order to implement and verify the serendipity book recommendation method using category similarity based on deep learning proposed in this paper, experiments were conducted in the environment shown in Table 2.

The dataset to be used in the experiment was collected from May 1996 to October 2018, and “Amazon Books 2018” was used, which consisted of a total of 2,935,525 book products and 51,311,621 reviews. For book review data, data from January 2017 to October 2018 were selected from the entire dataset, users who wrote less than 5 reviews were excluded, and rows with missing values were removed. It was converted into user log data in the same format as Table 3.

### 4.2 Experiment

In the preprocessing process, a specific book user was selected from the dataset, and the previously used books were classified into preference books and non-preference books, and only preference books were extracted. As a result of sorting based on the preferred books used, 27 preferred books could be selected as “Literature & Fiction” for each category, 11 for “Humor & Entertainment”, 10 for “Romance”, and 3 for “History”.

In the sequential recommendation process, preference books selected through a preprocessing process were input into the BERT model so that users can receive recommendations for books to use next, and the books

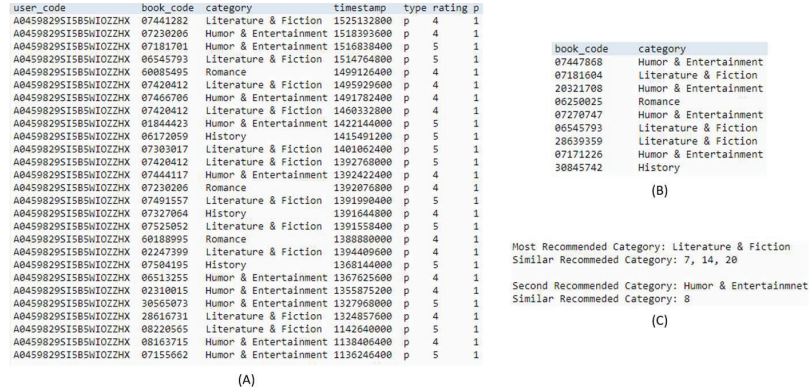


Figure 10 Experimental results.

in the “Humor & Entertainment” category were mainly recommended. In addition, although the number of uses was the lowest at 3, it was confirmed that books in the “Romance” category, which were used relatively recently, were included in the list of recommended books. Through this result, it can be confirmed that the BERT model makes a recommendation that reflects the user’s recent interests.

In the serendipity recommendation process, the categories of recommended books are checked through the sequential recommendation process to identify the most recommended categories and the categories recommended in the next order. First, the similarity for each category was calculated by applying the cosine similarity based on the number of books used by all users in each category.

Within the sequential recommendation list, popular books within similar categories are recommended by checking the similarity between “Humor & Entertainment”, the category of books recommended by the recommended user the most, and “Literature & Fiction”, the category recommended in the second place, and other categories. The category with the highest similarity to “Humor & Entertainment” is “Computers & Technology”, and the categories with the highest similarity to the next recommended category, “Literature & Fiction” are confirmed to be “History”, and “Mystery and Thriller & Suspense”, as shown in Figure 10.

### 4.3 Analysis of Experimental Results

To verify the recommendation performance of the serendipity book recommendation method using category similarity based on deep learning proposed

in this paper, hit rate, normalized discounted cumulative gain (NDCG) [14, 15], and serendipity indicators [16] are used as performance evaluation indicators. The hit rate represents the ratio of the number of users whose recommendation was hit to the total number of users, as shown in Equation (2). When the recommendation model, which has completed learning by excluding some of the correct answer books preferred by the user, recommends  $K$  items for each user, if the correct answer book excluded from the learning process is included among all the user's preferred books, the user is set as the user to whom the recommendation was successful. NDCG is a method of measuring how good the current model's recommendation list is compared to the most ideal recommendation combination, and when recommending  $K$  items to a user, the higher the ranking of items with relevance to the recommendation results, as shown in Equation (3). Serendipity measures the average similarity between a book newly recommended to a user and a book previously used. The lower the similarity, the higher the serendipity, as shown in Equation (4).

$$\text{Hit Rate} = \frac{\# \text{ of Hit User}}{\# \text{ of User}} \quad (2)$$

$$\text{DCG}_K = \sum_{i=1}^K \frac{rel_i^{opt}}{\log_2(i+1)} \quad (3)$$

$$\text{Unserendipity}_U = \sum_{u \in U} \frac{1}{|U||H_u|} \sum_{h \in H_u} \sum_{j \in R_{u,K}} \frac{\text{CosSim}(i, h)}{K} \quad (4)$$

$U$ : Set of all users

$H_u$ : User  $u$ 's favorite books in the past

$R_{u,K}$ : Top  $K$  items recommended to user  $u$ .

In order to verify the performance of the proposed method, we have compared the recommendation methodology using the existing RNN model, the recommendation methodology using only the BERT model, and the method proposed in this paper. The recommendation hit rate, NDCG, and serendipity index have been measured when the size  $K$  of the recommended book list to the user was set from 1 to 20. The detailed experimental results are shown in Figure 11.

Table 4 shows the average performance comparison results according to the size of the recommendation list for each model.

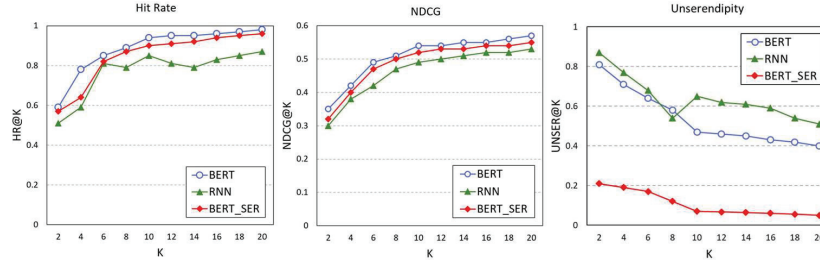


Figure 11 Experimental results.

Table 4 Average performance comparison for each model

Model	Hit Rate	NDCG	UNSER
RNN	0.7712	0.4647	0.6385
BERT	0.8869	0.5087	0.5379
BERT_SER	0.8482	0.4936	0.1056

## 5 Conclusion

Based on the similarity between the user and the item, existing book recommendation methods recommend items preferred by people with similar tastes, and recommend only similar types of books without considering changes in user preference over time, which reduces satisfaction with use. In order to improve this limitation, we have proposed a serendipity book recommendation method using category similarity based on natural language processing model BERT. Using BERT’s masked language model, a natural language processing model trained to predict masked words within a sentence and introducing sequential book recommendation techniques that view user-used books as sentences and predict books to be used at the next point in time, the hit rate of the proposed recommendation method has been improved by 9.98% and the NDCG by 6.22% compared to the recommendation method using the existing RNN model. In addition, by introducing a serendipity recommendation technique that recommends popular books in other categories with high similarity to the predicted book categories in the BERT model, serendipity was reduced by about 83.46% compared to the existing method without compromising recommendation accuracy.

Through the experimental results, it is confirmed that the proposed recommendation method shows high accuracy and can make recommendations that reflect serendipity that improves users’ satisfaction. If the proposed recommendation method is used, it is expected that it will be possible to

recommend personalized books that reflect serendipity, which can lead to practical preferences while reflecting users' recent interests in the mobile web service environment. However, the proposed recommendation method has a limitation in that it does not utilize information other than rating data and category data. As a future study, it is necessary to apply various attribute information such as local information and review text to recommendation for more sophisticated personalized recommendation.

### **Acknowledgement**

This research project supported by Ministry of Trade, Industry and Energy(MOTIE) and Korea Evaluation Institute of Industrial Technology(KEIT) in 2023 (No. 20016990).

We also appreciate Ms Youngsun Kwon for her sincere contribution to this research.

### **References**

- [1] J. Son, S. B. Kim, H. Kim, S. Cho, "Review and Analysis of Recommender Systems," *Journal of the Korean Institute of Industrial Engineers*, Vol. 41, No. 2, pp. 185–208, 2015.
- [2] K. Song., I. Moon, "Introduction to recent recommendation system research and future works," *Communications of the Korean Institute of Information Scientists and Engineer*, Vol. 39, No. 3, pp. 16–23, 2021.
- [3] S. Kim, Y. J. Roh, M. R. Kim, "A Narrative Study on User Satisfaction of Book Recommendation Service based on Association Analysis," *Journal of Korean Library and Information Science Society*, Vol. 52, No. 3, pp. 287–311, 2021.
- [4] S. Wang, L. Hu, Y. Wang, L. Cao, Q. Z. Sheng, M. Orgun, "Sequential recommender systems: challenges, progress and prospects," *arXiv preprint arXiv:2001.04830*, 2019.
- [5] P. Adamopoulos, A. Tuzhilin, "On unexpectedness in recommender systems: Or how to better expect the unexpected," *ACM Transactions on Intelligent Systems and Technology (TIST)*, Vol. 5, No. 4, pp. 1–32, 2014.
- [6] D. Kotkov, J. Veijalainen, S. Wang, "Challenges of serendipity in recommender systems," In *International conference on web information systems and technologies*, SCITEPRESS, 2016.

- [7] M. Quadrana, P. Cremonesi, D. Jannach, “Sequence-aware recommender systems,” *ACM Computing Surveys (CSUR)*, Vol. 51, No. 4, pp. 1–36, 2018.
- [8] H. Fang, D. Zhang, Y. Shu, G. Guo, “Deep learning for sequential recommendation: Algorithms, influential factors, and evaluations,” *ACM Transactions on Information Systems (TOIS)*, Vol. 39, No. 1, pp. 1–42, 2020.
- [9] B. Hidasi, A. Karatzoglou, L. Baltrunas, D. Tikk, “Session-based recommendations with recurrent neural networks,” *arXiv preprint arXiv:1511.06939*, 2015.
- [10] M. Quadrana, A. Karatzoglou, B. Hidasi, P. Cremonesi, “Personalizing session-based recommendations with hierarchical recurrent neural networks,” In *proceedings of the Eleventh ACM Conference on Recommender Systems*, pp. 130–137, 2017.
- [11] C. C. Aggarwal, “*Recommender systems*,” Cham: Springer International Publishing, Vol. 1, 2016
- [12] Y. Hu, Y. Koren, C. Volinsky, “Collaborative filtering for implicit feedback datasets,” In *2008 Eighth IEEE international conference on data mining*, pp. 263–272, IEEE, 2008.
- [13] J. Devlin, M. W. Chang, K. Lee, K. Toutanova, “Bert:Pre-training of deep bidirectional transformers for language understanding,” *arXiv preprint arXiv:1810.04805*, 2018.
- [14] X. He, L. Liao, H. Zhang, L. Nie, X. Hu, T. S. Chua, “Neural collaborative filtering,” In *Proceedings of the 26th international conference on world wide web*, pp. 173–182, 2017.
- [15] G. Shani, A. Gunawardana, “Evaluating recommendation systems,” In *Recommender systems handbook* (pp. 257–297). Springer, Boston, MA, 2011.
- [16] Y. C. Zhang, D. Ó. Séaghdha, D. Quercia, T. Jambor, “Auralist:introducing serendipity into music recommendation,” In *Proceedings of the fifth ACM international conference on Web search and data mining*, pp. 13–22, 2012.

## Biographies



**Youngmo Kim** received his Ph.D. degree in Computer Engineering from Deajeon University, Daejeon, Korea in 2011. He is currently adjunct professor in Soongsil University. He is also working on several standardization and national project.



**Seok-Yoon Kim** received his B.Sc. degree in electrical engineering from Seoul National University in 1980. He received his M.Sc. and Ph.D. degrees in ECE from University of Texas at Austin, in 1990 and 1993, respectively. He is currently with the School of Computing, Soongsil University.



**Byeongchan Park** received his B.Sc., M.Sc. and Ph.D. degrees in Computer Science and Engineering from Soongsil University, Korea, in 2015, 2018 and 2023, respectively. He is currently with the Dept. of Computer Science and Engineering, Soongsil University. He is also working on several national R&D projects.

