

TopC-CAMF: A Top Context Based Matrix Factorization Recommender System

Rosni Lumbantoruan¹, Paulus Simanjuntak², Ingrid Arionang³, Erika Simaremare⁴

Abstract—Online activities have been more and more vital as the digital business has expanded. Users can conduct most activities online such as online shops, hotel bookings, or online educations and courses. A large number of social users are drawn to the abundance of goods available on the Internet. The huge amount of information makes it impossible for social users to navigate it properly and efficiently. Many companies have offered a personalization to tackle this issue. It is proven that the personalized recommendation systems are able to suggest items to users based on their interests and needs that best suit them, which can be captured from user's contextual information. However, most of the studies capture this contextual information from the predefined contexts such as location and time. In this study, the personalized user context from the user's text review that they posted as they gave rating to an item was obtained. To this end, a new approach based on the matrix factorization recommendation model, TopC-CAMF, was proposed. TopC-CAMF investigates and finds the most important contexts or needs for each user by leveraging the deep learning model. First, all important contexts from user's text reviews were extracted. The next step was representing user preferences with the variations of most important contexts, namely top 5, top 10, top 15, top 20, and top 25 contexts. Then, the best top context variation was evaluated and the optimal one was used as the input for the matrix factorization method in providing better recommendations. Extensive experiments using three real datasets were conducted to prove the effectiveness of the TopC-CAMF in terms of root mean square error (RMSE), mean absolute error (MAE), mean squared error (MSE), normalized discounted cumulative gain (NDCG), and Recall.

Keywords—Context-Aware, Matrix Factorization, Personalization, Context Extraction.

I. INTRODUCTION

The rapid development of internet services and applications generates a large amount of information every day [1]-[3]. Based on *datareportal.com*, a website providing global reports on statistical data and industry growth, global internet users have increased up to 4.95 billion as of the beginning of 2022, which is 4% higher than the past year, and 79 zettabytes of data has been generated in 2022. This amount of data is so overwhelming that it takes users a long time to find information relevant to their interests [1]. This condition makes the existence of a recommendation system even more essential.

^{1,2,3,4} Fakultas Teknik Informatika dan Elektro, Pogram Studi Sarjana Sistem Informasi, Institut Teknologi Del, Jl. Sisingamangaraja, Sitoluama, Kecamatan Balige, Toba, Sumatera Utara 22381, INDONESIA (email: ¹rosni@del.ac.id, ²iss18068@students.del.ac.id; ³iss18034@students.del.ac.id; ⁴iss18054@students.del.ac.id)

[Received: 22 August 2022, Revised: 28 September 2022]

A recommendation system is a system that functions as an information filtering mechanism that is able to overcome the problem of information overload [4], [5] and help users find items or services that best suit their personal tastes or known as personalization [1]. Personalization is the act to customize an item or service based on behavior and preference knowledge of each user. It is one of the researchers' goals to provide recommendations according to user preferences using traditional recommendation systems.

In general, traditional recommendation systems can be classified into four categories based on the recommendation approach, namely content-based, collaborative filtering (CF), knowledge-based, and hybrid-based. Content-based models provide recommendations based on past user preferences. CF models provide recommendations by analyzing similar user behavior to identify candidate items. Knowledge-based models make recommendations by embedding domain-specific knowledge, while hybrid-based models make recommendations by combining the previous methods in a variety of ways [6]. The challenge faced is that traditional recommendation systems assume that user interests do not change and ignore the inherent relationship between preferences and user context, so that traditional recommendation systems only provide the same information for all users and cannot personalize results according to user interests [7].

To address these challenges, recommendation systems have been developed to provide personalized recommendation services and have been gaining increasing attention in both academic and industrial research [7]. One of the well-known recommendation techniques, namely collaborative filtering (CF), has the main concept of obtaining recommendations based on the similarity of user interests by analyzing the preferences of one user according to their history and then making recommendations based on the preferences of other people with the same interests [7]. The main approach used in CF is matrix factorization (MF). MF represents the relationship between users and items with a factor vector where the data are represented as multiplications of the user and item matrix. MF will decompose a large matrix into smaller matrices [8]. This technique is used to reduce the dimensions of the data for faster calculations without losing important information [8].

Another method used in this research was bidirectional encoder representations from transformers (BERT). BERT4Rec has been used to form item recommendations by employing BERT [9] and VCGN-BERT has been used to capture the user's main important contexts in text classification [10]. However, the aforementioned works ignore the influence and personalization of contexts on user preferences. Here, in this research, BERT was used as a deep learning method to

extract contexts or preferences for each user from user-contextual information and used them as input to the MF model to generate personalized recommendations. BERT's success in text comprehension specifically underlay the use of this method. In addition, deep learning's ability to process heterogeneous data sources brings more opportunities to recommend diverse items with unstructured data, such as review data [11].

Context-aware recommender systems (CARS) which incorporate contextual information have proven successful in enhancing the recommendation quality in returning the most relevant items to user preferences [12]-[14]. The previous approach mainly used topic modelling in retrieving the most important contexts for each user. The contribution in this study is to propose a new approach to the matrix factorization recommendation model by utilizing deep learning, namely BERT. BERT uses the contextual information review data as the main input of the model to provide an in-depth understanding of user preferences so that the system is able to provide the most relevant suggestions to users.

II. STATE OF THE ART OF RECOMMENDER SYSTEMS

A review of three existing studies closely related to this work, including CF, CARS, MF and deep learning was done.

A. Collaborative Filtering (CF)

The CF method is a method in which the collection and examination of information is based on user behavior, activity, or preferences. This method anticipates the interests of certain users by using their similarities with other users [15]. A CF-based approach and a three-segment bio-inspired clustering ensemble-based clustering technique has been used for user clustering, predictions, and recommendation results, to improve the process of making recommendations and grouping users on Yelp and TripAdvisor datasets [16]. Consistent prediction results indicate that this approach has outperformed other recommendation system approaches. Reference [17] proposed a CF method using a deep learning method, namely the restricted Boltzmann machine (RBM)-convolutional neural network (CNN), to provide insight to the role of big data analytics in the implementation of an effective recommendation system. This method achieved accuracy by resolving data privacy and security issues found in CF-based recommendation systems. A modification to user-based CF was made by calculating similarity between users using user-supplied text reviews for a single item reviewed together [18]. Then, the scores of this similarity were used as weights in the ranking prediction. In this study, the rating prediction was generated in addition to CF by learning the users' preferences from the users' review texts or user/item metadata when users' review texts were unavailable.

B. Context-Aware Recommender System (CARS)

Context-aware is an evolution of the traditional recommendation system that applies context information to improve the quality of recommendations. CARS has proven to improve the accuracy of recommendations by adapting to user preferences for different contextual situations. With the help of

available contextual information, CARS takes over modeling and predicts user tastes and preferences. This information is usually in the form of ratings and is modeled as a function of not only items and users, but also context. The rating function can be defined as following.

$$R: user \times item \times context \rightarrow rating$$

where *user* and *item* are the domain of user and item, *rating* is the domain of rating, and *context* determines contextual information related to the application.

Research [19] proposed a CARS for promoting online learning for individual learners. By using the social context between students and learning objects, then classifying them using the K-nearest neighbor and decision tree methods, this system was able to classify appropriate types of learners and provide personalized learning. Dynamic contexts have been addressed for mobile users such as location, time, and weather conditions [4]. Specifically, this research was AI-based mobile so that the suggestions and information offered by the system were appropriate and relevant to user preferences. Meanwhile, [5] presented a CARS using an MF approach to improve system performance on large datasets and looking at the effect of combining contextual information with traditional recommendation systems on the accuracy of the recommendation results. Personalized user contexts have been proposed in [12], [13]. Specifically, [12] proposed a declarative context-aware recommender system (D-CARS) to extract personalized user context using topic modelling, namely the user window non-negative matrix factorization (UWNMF) and a new declarative context-aware recommendation algorithm based on the subspace ensemble tree model (SETM) for building personalized user profiles and identifying candidate items. Meanwhile, [13] proposed personalized context in an interactive manner by retrieving feedback from users during the interaction and learning the users' preferences based on this feedback. In contrast to recent studies, this study aims to capture user-important contexts and find the optimal number of contexts to represent user preferences without degrading the model's ability to return the most relevant items to the users.

C. Matrix Factorization (MF)

In a situation where the recommendation system is connected to the item ranking matrix, MF is proposed to break down a large matrix into a smaller one. The reality is that there are many users and items in this approach, many of which have not yet been rated or purchased by the users. Indirectly, this matrix will consequently grow and become sparse, hence, a solution to handle the sparsity in this technique needs to be proposed so that the dimensions of the data become smaller [8]. Some of the research has addressed this issue by incorporating social network information [20] and a preference refinement [21]. Specifically, [20] proposed a recommendation method based on social networks using the MF technique. Preference factors and social relationships between users were considered and users who achieved superior performance results were classified. In addition, based on this research, MF was able to be a solution in overcoming data sparsity and has a significant impact in its use with multi-dimensional user information. An

MF-based clustering, namely, a cluster refinement on preference embedded MF (CREPE MF), has been proposed [21]. This algorithm was based on two-stage clustering, namely the preference network integration and the inter-item similarity network using a social network subgraph on the Yelp dataset. They mentioned that preference networks could capture user preferences better in the user latent feature matrix compared to social networks.

In this study, user and item metadata in addition to users' rating to items were taken into account. Additionally, context-specific data from user-posted reviews written by users after they gave an item a rating were gathered.

D. Deep Learning

A deep learning recommender system is proposed to train artificial neural networks by automatically learning representations of data, such as text, video, and images. The representation used in this study was a text in the form of a user review of an item. A deep learning-based model has been used to study item properties and user behavior based on a review text called the deep cooperative neural networks (DeepCoNN) model [22]. This model was able to extract semantic information from the review by using pretrained word insertion techniques to study latent factors between users and items. It is similar to MF but adds the use of neural networks so that the prediction accuracy and overall system performance are maximized. Reference [23] proposed a graph convolutional matrix completion (GC-MC) approach using a graph auto-encoder framework for matrix completion and could be generalized to include secondary information from users and items. Meanwhile, [24] proposed a large pretraining language model approach such as BERT, which had been applied to text ranking problems. The results showed superior performance on several datasets, namely the MS MAR-CO and TREC-COVID datasets. On the other hand, in this study, the deep learning approach, namely BERT, was used to filter word context into a rating review that could describe users' reviews by studying the context of the word based on the surrounding words (context words).

III. METHODOLOGY

This section describes the framework of the proposed approach, namely TopC-context-aware matrix factorization (CAMF), as depicted in Fig. 1. There are three main components of the TopC-CAMF: 1) data and text preprocessing; 2) context extraction with BERT; and 3) recommendation generation using TopC-CAMF.

A. Text and Data Preprocessing

Prior to generating the list of recommendations, data with text and data preprocessing were prepared. Then, features given to the preprocessed data, specifically the users' review texts for an item, were extracted. It was believed that the most crucial factors for users to consider when selecting an item were included in these review texts. Subsequently, the user preferences in a range of significant contexts were simulated using the retrieved contexts. Selecting relevant contexts is crucial when modeling user preferences since irrelevant

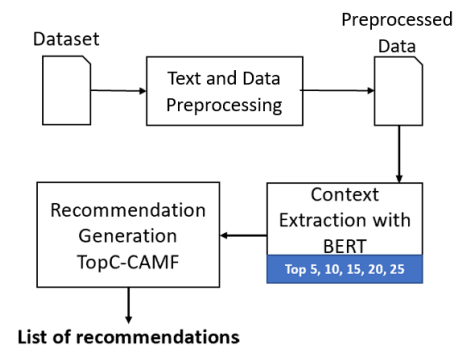


Fig. 1 Framework of TopC-CAMF.

circumstances can reduce the effectiveness and efficacy of recommendation [12], [25]. The details of components of the proposed TopC-CAMF are explained as follows.

B. Matrix Factorization (MF)

Gathering information from other users to make prediction recommendations in CF makes user and item relationships form a rating matrix. This rating matrix is usually sparse or incomplete, because generally users only give ratings to certain products, so there are missing data entries in the matrix. The missing value should be sought so that the recommendations can be personalized appropriately using the latent variable model. The recommended method for finding these hidden variables is MF since it provides solutions for scalability and better predictive accuracy in finding hidden structures in big data. Prediction with MF on the value of user review or r_{ui} is described in (1) below.

$$\hat{r}_{ui} = \mu + b_u + b_i + q_i^T p_u \quad (1)$$

where μ represents the overall average rating, parameters b_u and b_i indicate the observed deviations from user u and item i . For a particular item i , the q_i element measures the extent to which the item has these factors, positive or negative. For a given u user, the p_u element measures the user's level of interest in the item that is high on the related factor, either positive or negative. The resulting point product, $q_i^T p_u$, captures the interaction between user u and item i of the user's overall interest in the characteristics of the item.

C. Context-Aware Matrix Factorization (CAMF)

In general, context awareness can be defined as a process for analyzing, extracting, and utilizing contextual data to provide specific functions that are appropriate to the current context of a particular task. In this study, contextual information was combined with the aim of producing more relevant suggestions by adapting them to the contextual situation of the user. There are three approaches to the CAMF model that deal with context differently [26]. For CAMF-C, it is assumed that each context value has a global influence on the rating independently of the item. CAMF-C uses one parameter for each value of the contextual factor. The second model is CAMF-CI, which will use one parameter for each pair of contextual values and items. In this manner, capturing contextual factors will have different effects on ratings depending on the item. The expression for rating prediction with CAMF is shown in (2) [26].

$$r_{ui c_1 \dots c_k} = \vec{v}_u \times \vec{q}_i + \underline{i} + b_u + \sum_{j=1}^k B_{ij}c_j. \tag{2}$$

where $r_{ui c_1 \dots c_k}$ is the predicted rating for user u for item i with contextual value $c_1 \dots c_k$. Variables \vec{v}_u and \vec{q}_i are vectors of user u and item i being multiplied. The recommendation models experimented with in this study were where the AC-CAMF was the variant of the proposed TopC-CAMF, with details as follows.

1) *Context-Aware Matrix Factorization (CAMF)*: The CAMF model is a comparison model (baseline) that uses contextual information when providing rating predictions. Contextual information used is user-reviewed data that does not go through any process (original). This method is analyzed to see the effect of the original data review in providing recommendations to users.

2) *Top Dominant Context-Aware Matrix Factorization (TopC-CAMF)*: The TopC-CAMF model is a CAMF model that uses variations in a number of contexts in making recommendations. In user review data, user preferences are extracted with five context variations, namely top 5, top 10, top 15, top 20, and top 25. Then, based on the five variations of the context, one context that has the best value is selected (dominant context) in performing user preferences based on evaluation metrics

3) *All Context-Context-Aware Matrix Factorization (AC-CAMF)*: The AC-CAMF model is a CAMF model that uses all context, which is the entire context representing user reviews. This method is analyzed to see the effect of the entire context in providing recommendations to users.

D. *Bidirectional Encoder Representations from Transformer (BERT)*

BERT is a two-way transformer model that studies the context of words based on the surrounding words (context words) so that this model has a deeper understanding of context than the one-way language model. Steps taken to extract context from user review data using BERT are detailed as follows.

1) *Extract Feature*: This stage is the first step in the process of implementing the BERT model. The preprocessed data will be processed with the aim of extracting a list of candidate keywords from the review document.

2) *Word Embedding*: After the keyword candidate list is obtained, the review data and the candidate list will be converted into vector form, respectively. BERT was used for this purpose as it showed good results for similarity and paraphrasing tasks.

3) *Extract Keyword*: This is a process carried out to extract words and phrases that are most relevant to the reviewed data. In this study, the KeyBERT method was used to extract keywords by utilizing the BERT language model. Researchers

TABLE I
DATASET

Beauty	Office	IncarMusic
167,725	110,472	42
29,004	14,224	139
252,056	138,084	4,012
review, keyword	review, keyword	title, drivingStyle, landscape, mood, naturalphenomena, RoadType, sleepiness, trafficConditions, weather
336 MB	238 MB	93 KB

have used BERT embedding and cosine similarity to determine which subphrases in the document are most similar to the document itself. The results of the embedding in the previous stage are extracted for words, and then the most similar words are selected according to cosine similarity, as shown in (3).

$$similarity = \cos \cos (\theta) = \frac{A.B}{||A||.||B||} \tag{3}$$

where A represents vector A , B represents vector B to compare the similarities, $A.B$ is the dot product of vectors A and B , and $||A||/||B||$ is the cross product of $|A|$ and $|B|$. In contrast to the previous framework, [9] treated the item's name as the only context that was significant for item recommendations. Thus, they did not incorporate other contexts than item's name in the recommendation generation. Meanwhile, [10] generated text classification by extracting the main context from each sentence and combining it with contexts extracted using a graph neural network. In this study, the most crucial contexts from user reviews were studied to define the most optimal number of contexts to generate item recommendations.

IV. EXPERIMENTAL EVALUATION

In this section, the experimental stages of the research are described.

A. *Experimental Setup*

Experiments were carried out on three datasets, namely the Beauty, Office, and InCarMusic product datasets as depicted in Table I. Beauty and Office product datasets were taken from Amazon product data, which contains product reviews and metadata. The InCarMusic dataset was taken from Github and was a dataset collecting suggestions about car drivers' contextual music. The context used in this research was selected. In the Beauty and Office dataset, the review contexts and keywords were used. Meanwhile, since InCarMusic does not have a review, any contextual information the dataset has, such as driving style, mood, and weather, were used.

The dataset was divided into 90% of training set and 10% of testing set. The training set was used by the model to learn the users' preferences, given that their contexts and test sets were used to assess the performance of the model.

B. *Evaluation Methodology*

Evaluation was used as a basis or benchmark to find out how well the performance of the model has been built in making

predictions and providing recommendations. These experiments aim to address three research questions (RQs).

1. [RQ-1]: How to identify contexts that affect user preferences from user text reviews on items and/or metadata from items and users?
2. [RQ-2]: How does the number of retrieved contexts affect the users' preferences?
3. [RQ-3]: How effective is the proposed TopC-CAMF in returning the most relevant items to users?

These RQs were evaluated using three categories of metrics. The first category is predictive accuracy metrics, which include mean average error (MAE), mean squared error (MSE), root mean square error (RMSE). The second category is ranking accuracy metrics, namely normalized discounted cumulative gain (NDCG). The last category is classification accuracy metrics (recall). Predictive accuracy metrics look at the effectiveness of the recommendation model in predicting ratings and providing recommendations to users. In (4)-(6), N is the number of data samples, Y_i is the actual data value and Y'_i is the predicted data value.

$$MAE = \frac{1}{N} \sum_{i=1}^N |Y_i - Y'_i|. \quad (4)$$

$$MSE = \frac{1}{N} \sum_{i=1}^N (Y_i - Y'_i)^2 \quad (5)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (Y_i - Y'_i)^2} \quad (6)$$

Ranking accuracy metrics measure the ability of the recommender system to place the items most relevant to users at the top of the recommendation list rather than at the end. In (7), Z is the normalization that makes the perfect NDCG ranking worth 1, $r(j)$ is the relevance of the j document position from the ranking results.

$$NDCG = \frac{1}{Z} \sum_{j=1}^n \frac{2^{r(j)}}{\log 2(1+j)} = \sum_{j=1}^n \frac{G_i}{D_i}. \quad (7)$$

Classification accuracy metrics measure the frequency with which the recommendation system makes the right or wrong decision about whether an item is good. In (8), TP is true positive and FN is false negative.

$$Recall = \frac{TP}{TP+FN}. \quad (8)$$

V. RESULTS AND DISCUSSION

This section describes the results of the implementation and a discussion of the results obtained.

A. Identifying Users' Preferences in Terms of Contexts

To answer RQ1, the users' preferences in terms of context from stated users' text reviews for an item, item and user metadata were identified. It was believed that not all contexts were important to users, thus, rigorously selecting the most important contexts for users were deemed necessary. Here, top user contexts (5, 10, 15, 20, 25) were assessed and the model's performance in returning recommendations when utilizing these various contexts were compared. In the next subsection, steps in order to capture these contexts are presented.

1) *Text Preprocessing*: User contextual information, such as review data, is data that has an unstructured and inconsistent form. As a result, a process is required to clean and prepare the data before it can be used properly. The text preprocessing carried out included lower case, punctuation removal, stop word removal, and lemmatization. Lower case is used to convert all text data to lowercase so that the use of capital letters is equivalent in review data. Remove punctuation is used to remove characters or noise that are irrelevant and have no effect on processing, such as punctuation and other characters that are not recognized in the dataset. Stop word removal is used to filter out important words, while lemmatization is used to reduce or change words into meaningful basic forms. For example, given a user review "I BOUGHT THIS Avalon Organics: Biotin B Complex Thickening Conditioner, 14 oz AND THE SHAMPOO AND USED IT DURING THE WHOLE 2012 AND NOT ONLY DID NOT THICKENED MY HAIR BUT IT MADE IT VERY DRIED". The result of the preprocessed review will be "buy avalon organics biotin b complex thicken conditioner oz shampoo use whole thicken hair make dry". Here, it can be seen that in this user review, some frequently appearing words such as "I" and "the" are excluded. The preprocessed text was used to extract the context candidates at the next step. This context candidates then were used as the presentation of user preferences.

2) *Extracting Context Candidates*: Here, context-extraction of user's preferences using BERT was proposed in three steps: Firstly, a list of candidate contexts was produced by retrieving important words from the preprocessed data, which represent the user's preferences regarding an item. Here, BERT was used to extract the context candidates using the n-gram range of (0, 1), which determined the length of the candidate contexts to be generated, i.e., unigram or one context only. Given a preprocessed user review "love moisturizer would recommend someone dry skin fine line wrinkle use brand day night serum", the extracted context candidates were "[brand, day, dry, fine, line, love, moisturizer, night, recommend, serum, skin, use, wrinkle]". It can be clearly seen that the retrieval of context candidates returns important words without deteriorating the meaning of the review itself. Secondly, the similarity amongst contexts were evaluated and then the most important context as top 5, top 10, top 15, top 20, and top 25 contexts was returned. Top 5 can be interpreted as the most important 5 contexts representing a user's preferences. This information was retrieved by first presenting the user review in vector form using BERT. For example, given user review in Table II, the retrieved contexts could be moisturizer, skin, and serum. Finally, after getting the vector form from the list of candidate keywords and the review document, context extraction was performed by looking for context candidates that were similar and represented the review document. The similarity of the context candidates was assessed using the cosine similarity.

The similarity calculation was performed by calculating the similarity measure between all candidates of keyword embedding and the embedding document. Finally, the results were sorted in descending order to obtain the top relevant contexts. There were five variations of the extracted contexts,

TABLE II
TOP CONTEXTS

User Review	I have a very dry, frizzy hair. I have tried many hair products and cannot make it look sleek. I bought KMS product through my friend recommendation. It was not overnight change, but after a month, my hair looks much sleek. The conditioner is very rich and I do like the light smell a lot. The product arrives on time although the package looks different from what is on the picture.
Top 5	dry, conditioner, sleek, hair, frizzy.
Top 10	change, package, recommendation, product, products, dry, conditioner, sleek, hair, frizzy.
Top 15	smell, try, buy, light, look, change, package, recommendation, product, products, dry, conditioner, sleek, hair, frizzy.
Top 20	month, make, arrive, overnight, different, smell, try, buy, light, look, change, package, recommendation, product, products, dry, conditioner, sleek, hair, frizzy.
Top 25	kms, like, picture, friend, time, month, make, arrive, overnight, different, smell, try, buy, light, look, change, package, recommendation, product, products, dry, conditioner, sleek, hair, frizzy.

TABLE III
COMPARISON OF NUMBER OF CONTEXTS FOR CAMF IN TERMS OF MSE, RMSE, AND MAE

Dataset	Top	Metrics		
		MSE	RMSE	MAE
Beauty	5	0.2825	0.5315	0.3841
	10	0.3196	0.5653	0.4137
	15	0.3196	0.5653	0.4137
	20	0.3924	0.6264	0.4553
	25	0.3924	0.6264	0.4553
Office	5	0.2643	0.5141	0.3806
	10	0.3093	0.5561	0.4182
	15	0.3093	0.5561	0.4182
	20	0.3093	0.5561	0.4182
	25	0.3093	0.5561	0.4182
InCarMusic	5	0.4149	0.6441	0.5084
	10	0.4185	0.6440	0.5074
	15	0.4185	0.6440	0.5074
	20	NA	NA	NA
	25	NA	NA	NA

namely top 5, top 10, top 15, top 20, and top 25 as depicted in Table II. From Table II, it can be seen that utilizing a user review, important contexts of the user can be retrieved. For example, top 5 contexts

include “dry”, “conditioner”, “sleek”, “hair”, and “frizzy”. Supposed they are expanded to ten contexts (top 10), additional contexts include “product” which can be seen as the more general context from previous top 5 context “conditioner”. Therefore, the more the contexts, the more general the preferences will be and the less capable it will be of distinguishing one user from others.

B. Effect of Context Numbers in Modeling User Preferences

To answer RQ2, the model’s performance in returning the recommendation list with varied numbers of contexts for each

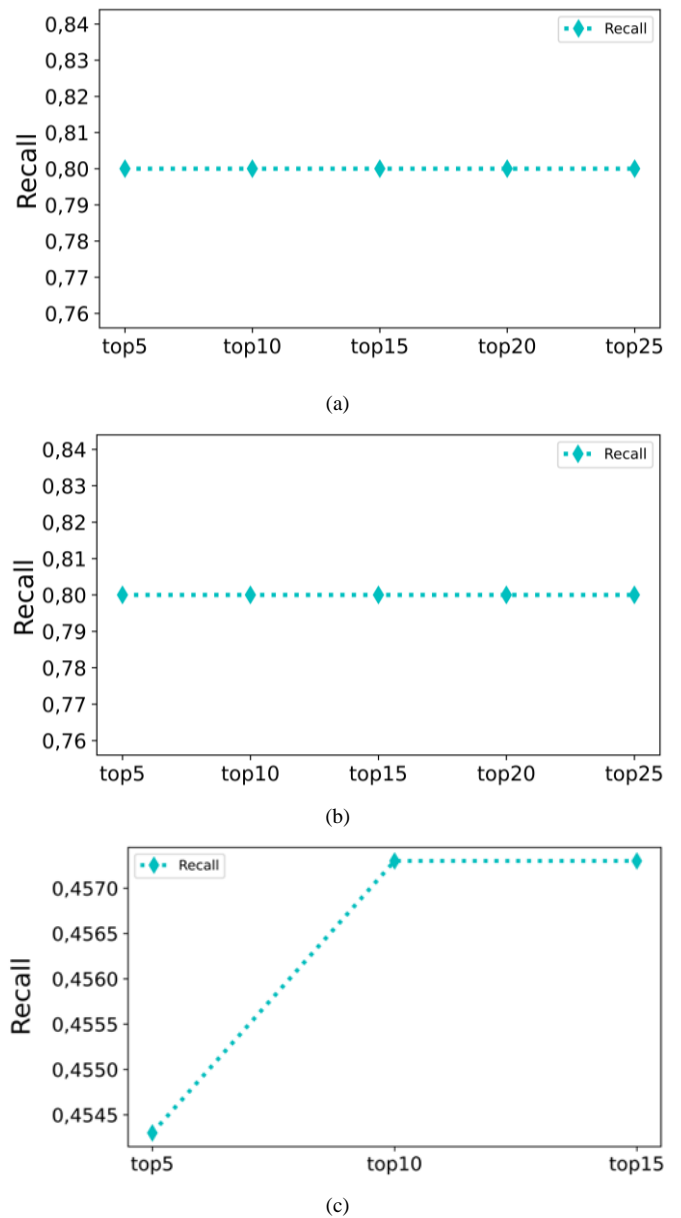


Fig. 2 Effect of different number of contexts for recommendation performance in terms of recall over three datasets, namely (a) Office, (b) Beauty, and (c) InCarMusic datasets.

user was assessed. Specifically, each user preference using top 5, top 10, top 15, top 20, and top 25 contexts were modeled. Considering that InCarMusic dataset does not have user review text as the contexts but rather predefined contexts, this dataset only contains top 5, top 10, top 15 contexts.

Table III shows the model’s performance in terms of MSE, RMSE, and MAE giving different number of contexts. In Table III, the best performance is denoted by a bold. From this table, it is clear that the top 5 contexts demonstrate the best performance by returning the lowest error values of 0.2825, 0.5315 and 0.3841 for MSE, RMSE, and MAE for Beauty datasets, respectively; for Office dataset, error values of 0.2643, 0.5141, and 0.3806 were returned. Top 10 displayed the ideal number of contexts for the InCarMusic dataset due to

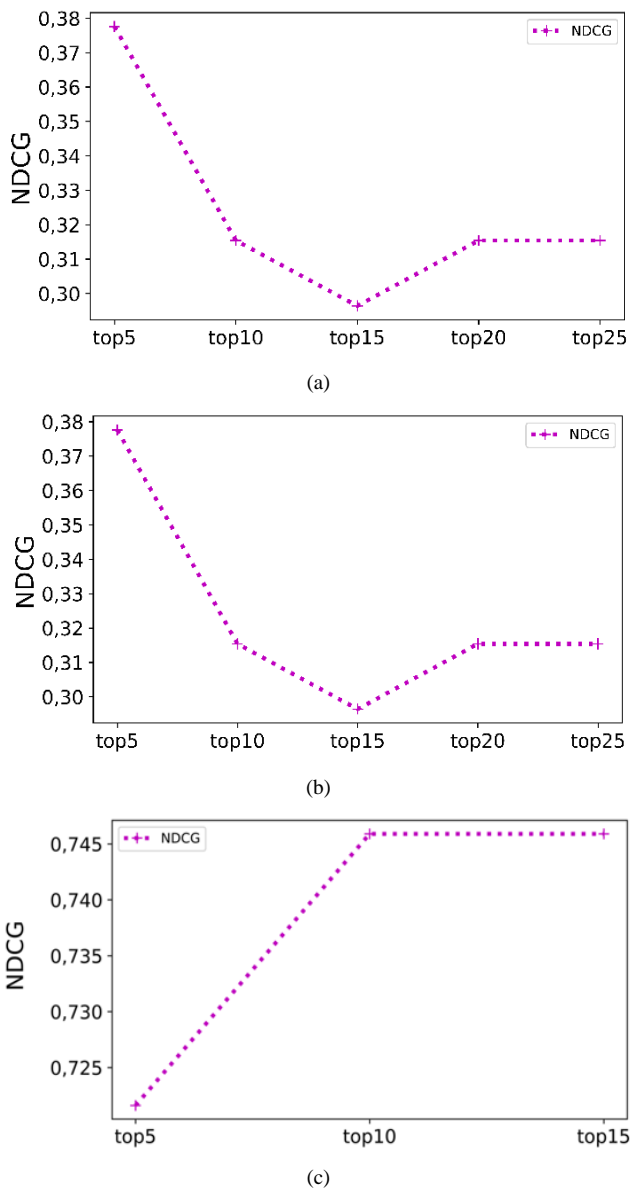


Fig. 3 Effect of different number of contexts for recommendation performance in terms of NDCG over 3 datasets, namely (a) Office, (b) Beauty, and (c) InCarMusic datasets.

InCarMusic's inadequate contexts and the absence of the user review text.

However, it is clear that there were no significant differences between the top 5 and top 10 performances for InCarMusic. Thus, it can be concluded that in terms of MSE, RMSE, and MAE, the top 5 contexts outperform the strategies with an alternative number of contexts.

The effect of different numbers of contexts in users' preferences for recommendation performance are depicted in Fig. 2 and Fig. 3. The results in Fig. 2 shows that in terms of recall for Office and Beauty dataset, the model was able to return the best performance for top 5 contexts. As for InCarMusic which does not have user review as contexts, the addition of contexts from 5 to 10 contexts could enhance the performance of the model. It proves that identifying the

TABLE IV
COMPARISON OF TOPC-CAMF WITH COMPETITORS IN TERMS OF MSE

Dataset	TopC-CAMF	CAMF	AC-CAMF
Beauty	0.2825	0.3199	0.5066
Office	0.2643	0.3114	0.5375
IncarMusic	0.4149	0.4148	0.6151

TABLE V
COMPARISON OF TOPC-CAMF WITH COMPETITORS IN TERMS OF MAE

Dataset	TopC-CAMF	CAMF	AC-CAMF
Beauty	0.3841	0.5329	0.4138
Office	0.3806	0.5573	0.4187
IncarMusic	0.5084	0.6054	0.5074

TABLE VI
COMPARISON OF TOPC-CAMF WITH COMPETITORS IN TERMS OF NDCG

Dataset	TopC-CAMF	CAMF	AC-CAMF
Beauty	0.3776	0.3389	0.3216
Office	0.3528	0.3286	0.3490
IncarMusic	0.7216	0.7074	0.7459

important contexts are essential in understanding the user's preferences.

According to Fig. 3, the top 5 is the best number of contexts for both Office and Beauty dataset by returning the highest NDCG value. Meanwhile, for InCarMusic dataset with a poor context, the model needed to enhance the contexts in order to get better user's preferences. Thus, for InCarMusic, better recall was displayed in the top 10 contexts.

C. Effectiveness Comparison with Competitors

To answer RQ3, the TopC-CAMF was compared with CAMF and other alternatives that were named as AC-CAMF in terms of error metrics (RMSE, MAE, and MSE) and ranking metrics (NDCG and recall). The comparison of the competitors for all three datasets can be seen in Table IV and Table V. The details of the methods are as follows:

1) *TopC-CAMF*: proposed CAMF algorithm which incorporates top 5 contexts in modelling user's preferences.

2) *CAMF*: this method does incorporate contexts in generating the recommendation but does not take into account the user's text review as contexts.

3) *AC-CAMF*: the alternative CAMF which does not filter importance contexts but uses all retrieved terms from the text reviews as context.

The TopC-CAMF model predicted ratings more accurately than the other two approaches for the entire dataset examined, according to a comparison of the three models' MSE, RMSE, and MAE values. Due to other approaches, CAMF and AC-CAMF did not support the TopC-CAMF strategy for selecting the most crucial contexts to express user preferences.

Table VI shows the comparison of methods in terms of NDCG values. Based on the comparison of the three models in terms of NDCG values, the TopC-CAMF model was superior to the other two CAMF and AC-CAMF across all the datasets. There were no significant differences among the strategies in

terms of Recall. As for Beauty, Office, and InCarMusic, respectively, each TopC-CAMF yielded 0.8000, 0.7900, and 0.4543. In contrast, CAMF provided the same recall for the Beauty dataset, with a higher recall for Office (0.8250), but a lower recall for Office (0.3655). AC-CAMF, on the other hand, produced better results for the Beauty and Office datasets, with values of 0.8200 and 0.8500, respectively, and 0.4573 for the InCarMusic dataset. Despite the fact that there were fewer users and items in the InCarMusic dataset, it appeared that AC-CAMF performed better. As a result, almost all the items may be appropriately retrieved for the target users when the model gives results in terms of recall.

VI. CONCLUSION

In this paper, a new approach on top of context-aware matrix factorization, namely TopC-CAMF, was proposed. In this research, three research questions were addressed, namely RQ-1, RQ-2, and R-Q3. RQ-1 is on how to identify the users' important contexts. In RQ-2, after retrieving the candidate contexts, the optimal number of contexts in representing the user's profile was examined. Finally, in RQ-3, the proposed TopC-CAMF was compared with CAMF and the alternative approach AC-CAMF.

TopC-CAMF identifies the user preferences through contextual information in terms of users' review texts and user/item metadata. The user context was then extracted, with the top 5 contexts being the best context in the Beauty and Office datasets, and the top 10 contexts in the InCarMusic dataset. This difference was found due to differences in contextual information in each dataset. With a slightly different performance for top 10 and top 5 for InCarMusic, the rest of the experiment of top 5 for all datasets were employed. Finally, the effectiveness of the proposed method in various metrics was evaluated. Experimental results have proved that the proposed new approach has outperformed the competitors, namely CAMF and the proposed alternative method, AC-CAMF, in terms of effectiveness in generating recommendations to target users.

Future works are expected to improve TopC-CAMF to more comprehensively captured contexts that can be inherited from the identified user dominant contexts. These comprehensive contexts will be personalized to each user to represent their preferences.

CONFLICT OF INTEREST

The author, whose name is listed in the article entitled "TopC-CAMF: A Top Context Based Matrix Factorization Recommender System" states that there is no conflict of interest with any parties that may influence the interpretation or the representation of this paper.

AUTHOR CONTRIBUTION

Conceptualization, Rosni Lumbantoran, Ingrid Arintonang, Erika Simaremare, Paulus Simanjuntak; methodology, Rosni Lumbantoran, Ingrid Arintonang, Erika Simaremare, Paulus Simanjuntak; software, Paulus Simanjuntak, Rosni Lumbantoran; validation, Rosni Lumbantoran, Ingrid

Aritonang, Erika Simaremare, Paulus Simanjuntak; formal analysis, Rosni Lumbantoran; investigation, Paulus Simanjuntak, Rosni Lumbantoran; resources, Paulus Simanjuntak, Rosni Lumbantoran, Ingrid Arintonang, Erika Simaremare; data curation, Rosni Lumbantoran, Erika Simaremare; writing—original draft preparation, Rosni Lumbantoran, Erika Simaremare, Ingrid Arintonang, Paulus Simanjuntak; writing—review and editing, Rosni Lumbantoran, Paulus Simanjuntak, Erika Simaremare, Ingrid Arintonang; visualization, Paulus Simanjuntak; supervision, Rosni Lumbantoran; project administration, Rosni Lumbantoran.

ACKNOWLEDGMENT

The funding of this work is fully supported by the Institution of Research and Community Services (LPPM) of Institut Teknologi Del.

REFERENCES

- [1] G. Zhang, Y. Liu, and X. Jin, "A Survey of Autoencoder-Based Recommender Systems," *Front. Comput. Sci.*, Vol. 14, No. 2, pp. 430–450, Apr. 2020.
- [2] A.H. Ali and M.Z. Abdullah, "Recent Trends in Distributed Online Stream Processing Platform for Big Data: Survey," *2018 1st Annu. Int. Conf. Inf., Sci. (AiCIS)*, 2018, pp. 140–145.
- [3] S. Raza and C. Ding, "Progress in Context-Aware Recommender Systems-An Overview," *Comput. Sci. Rev.*, Vol. 31, pp. 84–97, Feb. 2019.
- [4] M.C. Rodríguez-Hernández and S. Ilarri, "AI-Based Mobile Context-Aware Recommender Systems from an Information Management Perspective: Progress and Directions," *Knowl.-Based Syst.*, Vol. 215, pp. 1–29, Mar. 2021.
- [5] M.H. Abdi, G.O. Okeyo, and R.W. Mwangi, "Matrix Factorization Techniques for Context-Aware Collaborative Filtering Recommender Systems: A Survey," *Comput. Inf. Sci.*, Vol. 11, No. 2, pp. 1–10, Mar. 2018.
- [6] D. Monti, G. Rizzo, and M. Morisio, "A Systematic Literature Review of Multicriteria Recommender Systems," *Artif. Intell. Rev.*, Vol. 54, No. 1, Jan. 2021.
- [7] Z. Cui *et al.*, "Personalized Recommendation System Based on Collaborative Filtering for IoT Scenarios," *IEEE Trans. Serv. Comput.*, Vol. 13, No. 4, pp. 685–695, Jul.-Aug. 2020.
- [8] S. Kulkarni and S.F. Rodd, "Context Aware Recommendation Systems: A Review of the State of the Art Techniques," *Comput. Sci. Rev.*, Vol. 37, pp. 1–33, Aug. 2020.
- [9] F. Sun *et al.*, "Bert4rec: Sequential Recommendation with Bidirectional Encoder Representations from Transformer," *Int. Conf. Inf., Knowl. Manag.*, 2019, pp. 1441–145.
- [10] Z. Lu, P. Du, and J.-Y. Nie, "VGCN-BERT: Augmenting BERT with Graph Embedding for Text Classification," *Adv. Inf. Retr.*, 2020, pp. 369–382.
- [11] S. Zhang, L. Yao, A. Sun, and Y. Tay, "Deep Learning Based Recommender System: A Survey and New Perspectives," *ACM Comput. Surv.*, Vol. 52, No. 1, pp. 1–38, Jan. 2020.
- [12] R. Lumbantoran, X.M. Zhou, Y.L. Ren, and Z.F. Bao, "D-CARS: A Declarative Context-Aware Recommender System," *IEEE Int. Conf. Data Mining (ICDM)*, 2018, pp. 1152–1157.
- [13] R. Lumbantoran, X. Zhou, Y. Ren, and L. Chen, "I-CARS: An Interactive Context-Aware Recommender System," *IEEE Int. Conf. Data Mining (ICDM)*, 2019, pp. 1240–1245.
- [14] R. Lumbantoran, X. Zhou, and Y. Ren, "Declarative User-Item Profiling Based Context-Aware Recommendation," *Int. Conf. Adv. Data Mining, Appl.*, 2020, pp. 413–427.

- [15] A.S. Tewari, J.P. Singh, and A.G. Barman, "Generating Top-N Items Recommendation Set Using Collaborative, Content Based Filtering and Rating Variance," *Procedia Comput. Sci.*, Vol. 132, pp. 1678–1684, 2018.
- [16] R. Logesh *et al.*, "Enhancing Recommendation Stability of Collaborative Filtering Recommender System Through Bio-Inspired Clustering Ensemble Method," *Neural Comput., Appl.*, Vol. 32, No. 7, pp. 2141–2164, 2020.
- [17] A.K. Sahoo, C. Pradhan, R.K. Barik, and H. Dubey, "DeepReco: Deep Learning Based Health Recommender System Using Collaborative Filtering," *Comput.*, Vol. 7, No. 2, Jun. 2019.
- [18] M. Terzi, M. Rowe, M.A. Ferrario, and J. Whittle, "Text-Based User-kNN: Measuring User Similarity Based on Text Reviews," in *User Modeling, Adaptation, and Personalization. Lecture Notes in Computer Science*, Vol. 8538, V. Dimitrova *et al.*, Eds., Cham, Switzerland: Springer, 2014, pp. 195–206.
- [19] W. Intayoad, T. Becker, and P. Temdee, "Social Context-Aware Recommendation for Personalized Online Learning," *Wirel. Pers. Commun.*, Vol. 97, No. 1, pp. 163–179, Nov. 2017.
- [20] C. Xu, "A Novel Recommendation Method Based on Social Network Using Matrix Factorization Technique," *Inf. Process., Manag.*, Vol. 54, No. 3, pp. 463–474, May 2018.
- [21] L.R. Divyaa, A. Tamhane, and N. Pervin, "A Clustering Based Social Matrix Factorization Technique for Personalized Recommender Systems," *AMCIS 2018 Proc.*, 2018, pp. 1–10.
- [22] L. Zheng, V. Noroozi, and P.S. Yu, "Joint Deep Modeling of Users and Items Using Reviews for Recommendation," *Proc. 10th ACM Int. Conf. Web Search, Data Min. (WSDM 2017)*, 2017, pp. 425–433.
- [23] R. van den Berg, T.N. Kipf, and M. Welling, "Graph Convolutional Matrix Completion," 2017, *arXiv:1706.02263v2*.
- [24] H. Zhuang *et al.*, "Ensemble Distillation for BERT-Based Ranking Models," *Proc. 2021 ACM SIGIR Int. Conf. Theory Inf. Retr. (ICTIR 2021)*, 2021, pp. 131–136.
- [25] R. Lumbantoruan, "Declarative Context-Aware Recommendation," Ph.D. dissertation, RMIT University, Melbourne, Australia, 2021.
- [26] L. Baltrunas, B. Ludwig, and F. Ricci, "Matrix Factorization Techniques for Context Aware Recommendation," *Proc. Fifth ACM Conf. Recommender Syst.*, 2011, pp. 301–304.