International Journal of Advanced Research in Engineering and Technology (IJARET)
Volume 12, Issue 2, February 2021, pp.491-500 Article ID: IJARET_12_02_047
Available online at http://www.iaeme.com/IJARET/issues.asp?JType=IJARET&VType=12&IType=2
ISSN Print: 0976-6480 and ISSN Online: 0976-6499
DOI: 10.34218/IJARET.12.2.2020.047

EFFICIENT RECOMMENDATION SYSTEM USING BERT TECHNOLOGY

Anushree H
Student, Ramaiah Institute of Technology, Bangalore, India

Shashidhara H S
Associate Professor, Ramaiah Institute of Technology, Bangalore, India

ABSTRACT

Modelling the diverse desires of users through their economic models for recommendation systems is difficult and essential. Past Techniques utilize successive neural networks to convert the historical experiences of customers across left to right into encoded suggestion models. Recommendation systems in e-commerce are becoming an integral means of making consumers navigate the content accessible. Recommender systems are an important aspect of E-commerce platforms that assist consumers choose products of choice on a wide scale through huge investments. The metadata termed the contextual phrase, that incorporates the reference label, and the description of the quote proves have been used by several authors to locate the relevant research referenced. The lack of a well-established benchmarking dataset and no tool for recommendations that can achieve great efficiency has perhaps made the study challenging. The foundation of mutual marketplace stages in the rental property field is the personalization of recommendation systems. To support people, such a framework provides a valuable tool. The current technique aims to determine the context of the movie plot summary from the given movies using BERT as-a-service and to predict similar movie recommendations.

Key words: Recommendation, Recommender systems, E-commerce, Bidirectional Encoders Representation Transformers (BERT), Movie plot summary.

Cite this Article: Anushree H and Shashidhara H S, Efficient Recommendation System using Bert Technology, *International Journal of Advanced Research in Engineering and Technology (IJARET)*, 12(2), 2021, pp. 491-500.
http://www.iaeme.com/IJARET/issues.asp?JType=IJARET&VType=12&IType=2

1. INTRODUCTION

At the core of a successful recommendation framework is the precise interpretation of the customer's specifications. The actual desires of customers in several practical uses are inherently complex and changing, informed by their past activities. Recommendation systems are powerful text mining resources that are popular due to growing Internet connectivity, developments in customization, and evolving technology user behaviours while current recommender systems are effective in generating great suggestions, they also face difficulties like reliability, interoperability, and cold start. Neural networks, the machine learning approach used in many complicated projects, is being used in recommendation systems over the last few decades to increase the accuracy of suggestions. Recommendation systems are resources for knowledge processing that resolve the issue by offering customers with customised material that is feasibly enticing. Numerous ecommerce sites presently empower their platforms with recommendation systems, and in their regular tasks, several web users take full benefits of such facilities, like reading literature, listening to songs, and purchasing. The word item refers to the good or service that the framework suggests to its customers in a standard recommendation scheme. A suggested framework is required whether to examine past interests of like-minded customers or profit from the informative knowledge more about products by generating a list of suggested products for the customer or determining how far the customer would like a specific product.

Although automated recommender system are becoming reasonably effective in recommending material, while granted minimal details about the interests of the consumer, their efficiency benefits hugely. Those cases are not unusual, they arise once the consumers are unfamiliar to a recommendation system and once the interests of the consumers are not documented due to the nature of the system or even their own wishes for safety. In such instances, recommendations need to be made solely based on material which is being suggested. BERT is utilised for suggesting the responses based on the description of the movie plot in this proposed work. BERT is more efficient than collaborative-based information at preserving content-based information. The usefulness of BERT for quest and suggestion exposes decreases significantly as the number of teams in the exposes increases, particularly for collaborative-based information. By acquiring promising performance in various regression tasks, BERT has proven conclusively to be amazingly effective in computational linguistics. It has also demonstrated that post pre-training, they are likely to essentially retain facts in their specifications.

2. LITERATURE SURVEY

Yuyangzi Fu et. al., [1] proposed an approach that could fix problems affecting conventional recommendation systems like cold start and the suggested material. They performed experiments with a complete cold-start paradigm on a large-scale real-life dataset. By utilizing the BERT framework for understanding product details and design significance among different products, they presented a new methodology to product-based collaborative filtering. From the context of natural language to the e-commerce background, they apply the masked language simulation and next phrase determination tasks. They utilize the name codes of the product as material, together with symbol descriptors, rather than specifically describing a product with a unique number, to resolve the cold start issue. In interpreting product names and understanding associations among products through extensive inventory, they prove the effectiveness of their framework over a conventional neural network method.

Xingjie Feng et. al., [2] research the customised rating determination task based on analysis texts for suggestion. Much relevant studies have currently used convolution neural networks to collect local contextual data, but term intensity and global contextual data have been lost. In conjunction, they try comparing the user or product embedding to the embedding analysis that carries certain unrelated analysis of textual data into the client requirement or product data.

The pre-trained BERT framework and the neural factorization machines (FM) are combined to accomplish the task of rating estimation. In order to explore the value of data analysis, the presented approach will successfully reduce the complexity of the latest CNN-based approach that can collect term frequency and general background details. Moreover, latest research has shown that FM performs better than internal products.

Chanwoo Jeong et. al., [3] presented a deep learning-based model and well organised dataset for context sensible paper reference recommendation. Their model includes an information encoder and a background encoder that utilizes a surface of Graph Convolutional Networks and BERT that is a pre-trained text database system. They suggested a new dataset comprising background phrases for cited sources and text documentation by changing the associated dataset.

Current databases are not up-to-date and there is no consistent background identification with respect to the knowledge reference suggestion research. They invented and launched the dataset to resolve these issues. The suggested dataset contains revised articles, covering most of the papers, which offers a tool for extracting information metadata easily and reliably has a very well viewpoint.

PavlosMitsoulis-Ntompos et. al., [4] proposed a simple Deep Personalized Recommendation framework to measure dependent embeddings of travellers. In order to create brief historical context to tailor suggestions for travellers, their approach incorporates listing embeddings in a regulated framework. This technique, which is implemented in the manufacturing environment, is computationally powerful and flexible and enables us to collect non-linear correlations. For a broad online vacation rental platform site, they proposed a technique which incorporates shallow and deep machine learning to understand travellers and list embedding. In the production setting, we implemented this device. In this sense, their findings indicate Deep Average Networks could perform quite complicated neural networks.

Zeynep Batmaz et. al., [5] include a detailed overview of suggestive strategies depended on deep learning to empower and direct new authors concerned in the topic. They review collected four-dimensional research, that are machine learning frameworks used in recommender system, solutions for recommender system problems, understanding and prominence of suggestion contexts, and purposeful properties. They also provide a detailed quantitative review of sector articles and summarize by addressing learned perspectives and potential future studies on the topic.

Because of the rising computing capacity and big data processing capabilities, artificial neural networks have started to gain considerable interest in recent years. As an evolving computer science field, authors effectively create and train profound design structures that facilitate machine learning. Many techniques in computer vision, object detection, processing of natural language, and recognition of voice presently use convolutional neural networks as a primary method. Convincing deep learning technology abilities also enable the researcher to apply deep architectures in recommendation jobs.

Dawen Liang et. al., [6] presented a collective music content-aware recommendation framework which combines an information retrieval framework with a multi-layer artificial neural modular system. In music suggestion, the approach yields better output provided information and recommender systems data. They produce a collective music recommender system that is sensitive of quality. The framework has, as the name implies, two aspects: the information model and the model of collaborative filtering. Collaborative filtering, where products are suggested to a person generated on certain members with specific habits of product purchase, is a commonly utilised strategy to suggestion.

Michael Tsang et. al., [7] focused on a popular use of the suggestion for machine learning: estimation website. They find that their concepts of communication are both insightful and reliable, for instance predominantly outpacing current models of recommendation. W ell beyond suggestion, the similar approach to understanding experiences could provide fresh perspectives into contexts, like image and text classification. They proposed a methodology both to perceive the observations of black-box recommendation systems and increase them. In specific, in a purpose recommender system, in which both input and output variants are black boxes, they intend to perceive function interacts from an origin recommendation system and specifically encrypt these interacts.

Michael Fleischman et. al., [8] analysed the issues with automated recommender systems while knowledge regarding user needs is minimal. They compare the issue with one of the calculations of probably available and adapt Natural Language Processing strategies to the movie recommender field. Utilizing concept identifiers, they identify two techniques, a naive word-space method and a more advanced method, and analyse their efficiency equated to reference, global market, and industrial strategies.

They address this issue of making suggestions as one of material similarity calculation lacking interest knowledge. Employed in the film sector, for every movie, they have exposure both to organised details and text-based scenario explanations. Their parameters of resemblance are linear functions of metrics of resemblance for individual information type.

3. PROPOSED SYSTEM

The proposed architecture provides the complete overview of the system. Primarily, the dataset must be collected and stored. The dataset is loaded into the NLP module. After loading the dataset, data pre-processing is carried out to eliminate the redundant data, null values and stored the essential columns into the data frame using pandas. The NLP module will perform the tasks like segmentation, tokenization, stemming, POS tagging and other lemmatization tasks. The yield from the attention layers of the embedding layers are achieved from the BERT pre-trained check points that are downloaded from the official git hub link, that are utilized as word-piece tokenization for the plot summary data, for instance, the token “Standing” will be tokenized as ‘stand’,’##ing’ to obtain a better context of the sentence obtained. The model is loaded from the pbtxt file obtained from the earlier step, which is also accessible in the BERT pre-trained checkpoints. BERT feature vectors are now acquired using tf.Estimator by graph parsing. These are the vector embeddings corresponding to which are the output of the 11th attention layer of BERT model. These vectors are context aware encodings, which can be used to find the similarities of sentences using classification algorithms such as KNN and using the Euclidean Distance module. The below figure 1 shows the architecture of the proposed system.

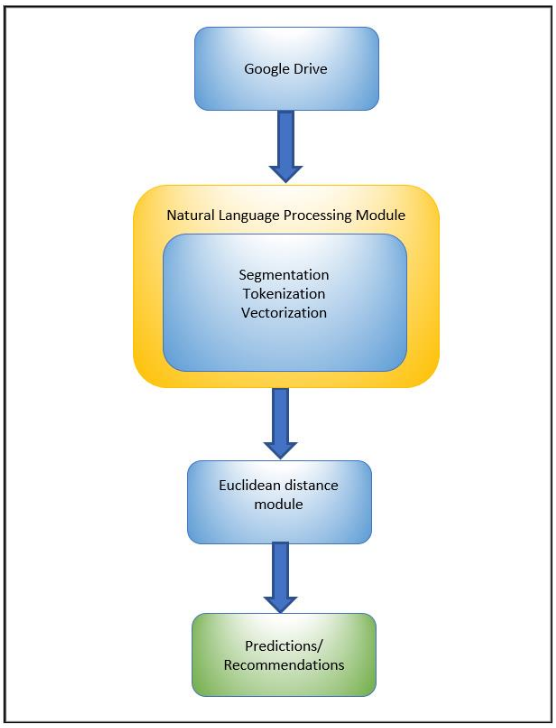

Figure 1 System architecture

Collection of datasets is obtained from the CMU Movie Summary Corpus Website and stored it in Google drive after cleaning the data set. Pre-processing of data is performed in the NLP module, where we obtain the vectorized form of plot summary from bert_vectorizer which is used as a feature. Euclidean distance module and KNN algorithm is used to obtain the top-n similar movies from the dataset.

3.1 Natural Language Processing

NLTK (Natural Language Toolkit) is a built-in python library collection. It is primarily utilized in the collection and classification of texts. The package for the NLTK library also symbolizes multiple qualifying classifiers. The NLTK library is used primarily to build a model for terms. In this one, we need to calculate the number of instances that are observed for every word. The entire tweet sentiment can be determined by assigning a sentiment compound polarity rating to every word using a sentiment analysis.

3.2 BERT as a Service

The main technological advancement of BERT is to implement transformer's bidirectional teaching, a common focus model, to machine translation. This is in contrast with previous attempts that focused at a textual series from left to right or mixed training from leftmost and from rightmost. BERT utilizes transformer, a system of concentration that in a textual learns conceptual relationships among words. Transformer contains two different frameworks in its regular context: an encoder which absorbs the input text and a decoder that generates a task estimation. Although the objective of BERT is to create a classification models, just the encoder technique is employed.

There is a concern while training language models to identify a predictive target. The next term in a series is predicted by several models, a lateral strategy that ultimately restricts contextual learning. BERT utilizes two preparation methods to resolve this problem:

a. Masked LM

15 percent of the words in every series are substituted with a [MASK] symbol prior to ingesting individual words into BERT. The system then tries to forecast the initial value of the masked terms, considering the context given in the series from the other, non-masked, terms. The estimation of the output terms in technical terms needs:

- Including a classification top layer on the output of the encoder.
- The embedding matrix multiplies the output vectors, converting themselves into the language element.
- Using softmax to measure the likelihood of every term in the vocabulary.

b. Next Sentence Prediction

Until approaching the model, the data is processed in the appropriate way to enable the system to differentiate among the two sentences in training.

- At the start of the initial sentence, a [CLS] token is placed, and at the completion of every sentence, a [SEP] token is placed.
- To every token, an embedded sentence implying sentence X or sentence Y is connected. Sentence embeddings with such a vocabulary of 2 are identical in principle to token embeddings.
- To signify its location in the series, a positional integration is applied to every token.

The basic methods are carried out to predict whether the second sentence really is related to the initial sentence:

- Through the transformer model, the whole input data enters.
- Utilizing a standard classification layer, the output of the [CLS] token is translated into a vector.

3.3. Nearest Neighbour

One of several best machine learning algorithms dependent on the ensemble classification methodology is Nearest Neighbour. This algorithm assumes that the latest scenario is identical to the available cases and places the new scenario into the classification that is most relevant to the cases available. The Nearest Neighbour algorithm preserves all relevant information and categorizes, based on certain characteristics, a new observation. This means that it could be conveniently categorized into a well-suite group by utilizing an algorithm as new information emerges. It is possible to use the Nearest Neighbour algorithm for Regression and Classification, but it is mainly used for challenges with classification. It is a non-parametric technique that means the raw data doesn't really make any predictions. It is often referred to as a slow learner method as it does not automatically learn from the training dataset, but rather preserves the dataset and executes an operation on the dataset at the point of classification. At the training point, the algorithm only preserves the dataset and afterwards categorizes the data into a group that is very close to the new dataset when it receives new data.

The main steps of Nearest Neighbour algorithm working:

- Choose the neighbours' number K
- Determine the Euclidean distance from the number of neighbours of K
- According to the determined distance measure, consider the K nearest neighbours
- Count the number of observations in group between these k neighbours

- Allocate new data to the group which has neighbour's maximum number

3.3 Euclidean Distance Module

Euclidean Distance Module is a class used with nearest neighbour algorithm to detect movies with plot vectors most identical to the movie in query and send this back to the customer.

4. RESULTS

The validations carried out on the actual results from recommendation of movies using BERT as a service obtained by using plots of the movie.

4.1 Output Validation

The recommendations produced here uses movie plot and assuming the genres of the movie are obtainable and are standard in nature, the genres of the recommended movie are compared to the particular movie, assuming the recommendations are good in nature we should expect similar genres in all of the recommended movies. Two examples are shown below.

Movie Name: The Crypt

Genre: ['Horror']

Figure 2 plots the number of times the genre has occurred in 10 recommendations by the system for the movie 'The Crypt'. As observed form the plot, the genre of the source movie 'Horror' has been tagged to 9 out of 10 recommendations meaning we can conclude that there are no false positives in the result.

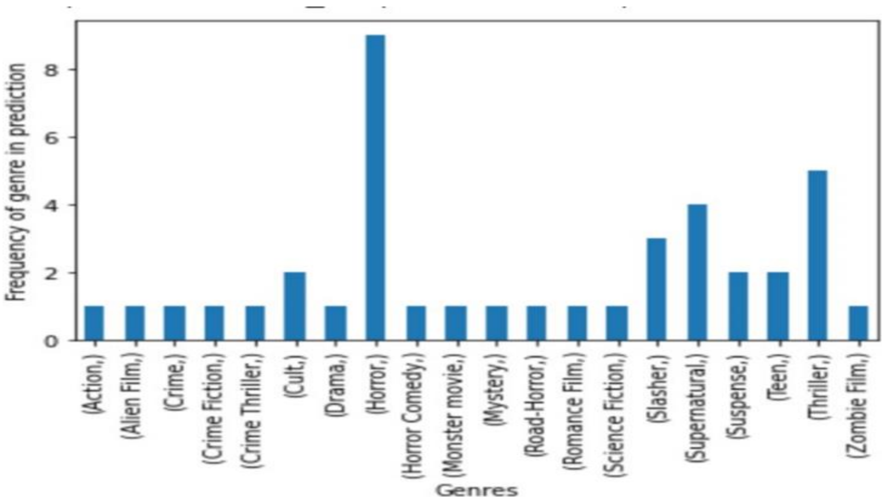

Figure 2: Genre Plot of recommendations for movie 'The Crypt'

Movie Name: The Black Shield of Falworth

Genre: ['Costume drama', 'Period piece', 'Costume Adventure', 'Adventure',]

Figure 3 plots the number of times the genre has occurred in 10 recommendations by the system for the movie 'The Black Shield of Falworth'. As observed form the plot, the genre of the source movie 'Adventure' and 'drama' has been tagged to 8 out of 10 recommendations and it is also interesting to observe that even though source movie doesn't have 'drama' as a genre value there is a very high co-relation between 'Costume drama' and 'drama' which we know from the to the domain knowledge.

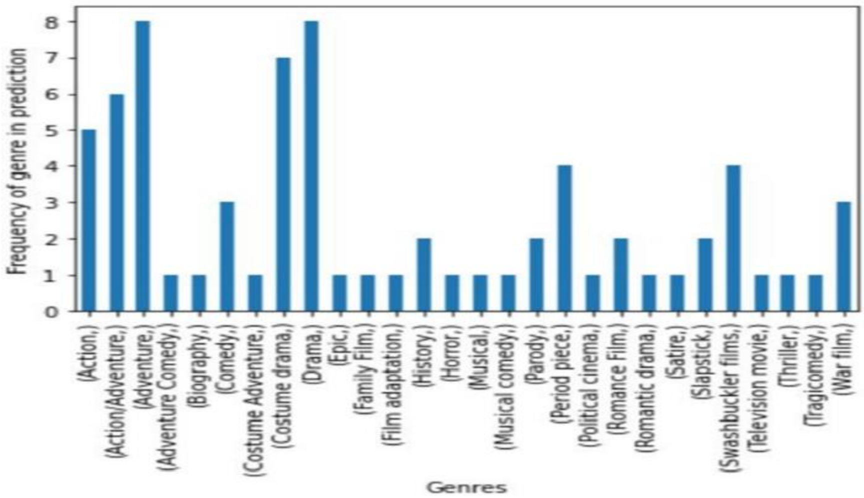

Figure 3: Genre Plot of recommendations for movie 'The Black Shield of Falworth'

4.2 Graphical Evaluation Depictions

Statistical analysis is achieved on the recommended movie plots to advance more vision into the recommendation quality. The below figure 4 and figure 5 represents the line plot presenting the co-sine similarity among the plot of the selected movie and the recommended movie. As we can see from the two plots, the similarity score is always above 8 which means the plots of the recommended movies when exposed to documented similarity yields a very good score.

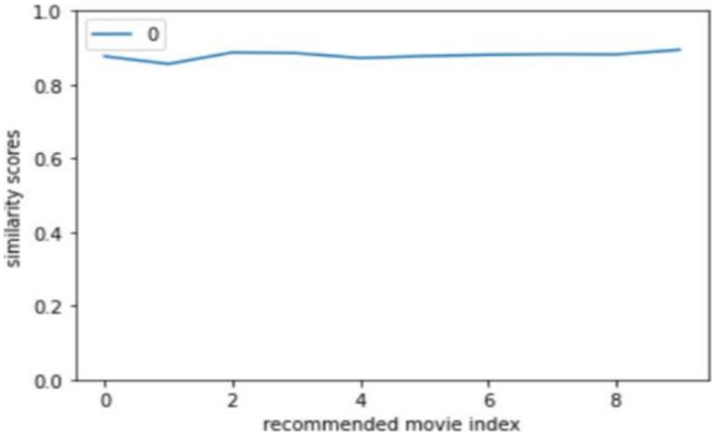

Figure 4: Similarity score Plot of recommendations for movie 'The Crypt'

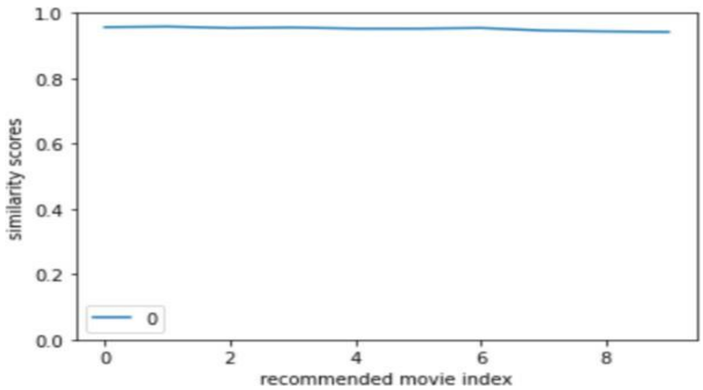

Figure 5: Similarity score Plot of recommendations for movie 'The Black Shield of Falworth'

Table 1 provides an overview of the five movie graphs.

Table 1: Average Similarity scores of Recommendations

Movie Name	Average Similarity Scores for top 10 recommendations	Standard Deviation
The Crypt	0.8794657	0.009727768
The Black Shield of Falworth	0.9516918	0.0053467895
Mr. Romeo	0.87620133	0.007351816
Saptapadi telugu film	0.8166755	0.0112663
Wackiki Wabbit	0.8513583	0.012052061

4.3 Result Screenshots

Fig. 6 shows the home page which is represented on a load of website. One must login with proper credentials to open the recommendation system.

Figure 6: Login Page

One movie must be selected from the dropdown for which recommendation should be done which is shown in Fig. 7.

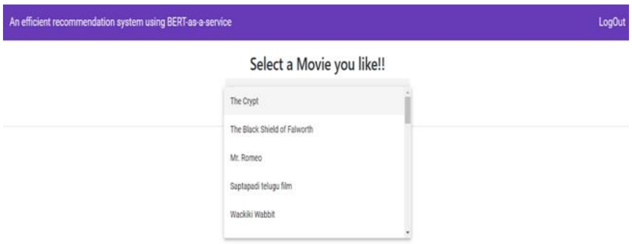

Figure 7: Selection Page with available movie list

Fig. 8 shows the page where the recommendations are shown. Upon selecting a movie, the user interface loads top 5 recommendations on the screen. User can click on right or left to scroll other movies. For each movie a poster, movie name and its genres are displayed.

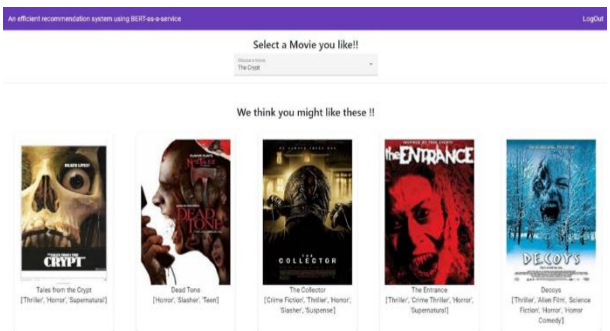

Figure 8: Movie Recommendations

5. CONCLUSION

In this work, it is demonstrated that the heavily pre-trained language models are capable of handling huge data and can be used for obtaining better recommendations for movies, books or any data that is in text sequence manner. In this work, it is tested to decide how much stronger a recommendation can be. It can be found by determining a similar plot description of movies with the movie under consideration. It gives an experience which has been preserved with the information of the BERT model in the form of pre-trained model parameters assists to create the functional vectors with probabilistic values obtained from the parameters of 11th layer of encoding module in the BERT module. These language vectors are mapped against the plot summaries. It is depicted that in the recommendation, 70-80 percent of the predicted movies consists similar genres as the genre of the movie under consideration. The design is tested against the recommendation framework using the cosine similarity measure, which produces the similar recommended top 10 films, provided the similarity score which ranges around 0.82-0.95 for a scale of 0 to 1 for the example movies evaluated for recommendations. Literature reviews are relevant for future work to examine the degree to which the recommendation model based on BERT operates together with user feedback and ratings.

REFERENCES

[1] Yuyangzi Fu and Tian Wang, Item-based Collaborative Filtering with BERT, Proceedings of the 3rd Workshop on e-Commerce and NLP, 2020, pp. 54–58.

[2] Xingjie Feng and Yunze Zeng, Neural Collaborative Embedding from Reviews for Recommendation, IEEE, Vol. 20, 2017, pp. 1-9.

[3] ChanwooJeong, Sion Jang, Hyuna Shin, Eunjeong Park and Sungchul Choi, A Context-Aware Citation Recommendation Model with BERT and Graph Convolutional Networks, Test Engineering and Management, 2019.

[4] PavlosMitsoulis-Ntompos, MeisamHejazinia, Serena Zhang and Travis Brady, A Simple Deep Personalized Recommendation System, 2019, pp. 22-26.

[5] Zeynep Batmaz, Ali Yurekli, Alper Bilge and CihanKaleli, A review on deep learning for recommender systems: challenges and remedies, Artificial Intelligence Review, 2018.

[6] Dawen Liang, Minshu Zhan, and Daniel P. W. Ellis, Content-aware collaborative music recommendation using pre-trained neural networks. Semantic Scholar.

[7] Michael Tsang, Dehua Cheng, Hanpeng Liu, Xue Feng, Eric Zhou and Yan Liu, Feature interaction interpretability: A case for explaining ad-recommendation systems via neural interaction detection, ICLR, 2020, pp. 1-19.

[8] Michael Fleischman and Eduard Hovy, Recommendations Without User Preferences: A Natural Language Processing Approach, Proceedings of the 8th International Conference on Intelligent User Interfaces, Miami, Florida, USA, 2003.