J. Chosun Natural Sci.
Vol. 14, No. 4 (2021) pp. 183 – 188
https://doi.org/10.13160/ricns.2021.14.4.183

Modeling of Convolutional Neural Network-based Recommendation System

Tae-Yeun Kim†

National Program of Excellence in Software center, Chosun University, Gwangju, Korea

Abstract

Collaborative filtering is one of the commonly used methods in the web recommendation system. Numerous researches on the collaborative filtering proposed the numbers of measures for enhancing the accuracy. This study suggests the movie recommendation system applied with Word2Vec and ensemble convolutional neural networks. First, user sentences and movie sentences are made from the user, movie, and rating information. Then, the user sentences and movie sentences are input into Word2Vec to figure out the user vector and movie vector. The user vector is input on the user convolutional model while the movie vector is input on the movie convolutional model. These user and movie convolutional models are connected to the fully-connected neural network model. Ultimately, the output layer of the fully-connected neural network model outputs the forecasts for user, movie, and rating. The test result showed that the system proposed in this study showed higher accuracy than the conventional cooperative filtering system and Word2Vec and deep neural network-based system suggested in the similar researches. The Word2Vec and deep neural network-based recommendation system is expected to help in enhancing the satisfaction while considering about the characteristics of users.

Keywords: Modeling of Convolutional, Neural Network-based, Recommendation System

(Recceived December 3, 2021 Revised December 20, 2021 Accepted December 29, 2021)

1. Introduction

In response to increasing e-commerce, the need for a product recommendation is further increasing and enhancing the accuracy of the product recommendation is one of the main issues in the web-based product recommendation system[1,2]. The collaborative filtering is one of the commonly used methods in the web-based product recommendation system[2]. The general cooperative filtering system uses the user's product rating information to figure out the correlation similarity, find the neighbor users of high similarity, and recommend products by using the purchase information of the neighbor user.

Recently, Word2Vec is actively used in the text analysis field, which is one of the natural language processing fields[4,5]. Word2Vec converts a word into vector and it identifies the correlation between the words in the sentence, turns them into vector, and places the similar words in the close distance at the vector space[6]. Word2Vec is also applicable on fields other than the text analysis and number s of researchers are suggesting on utilizing Word2Vec on the cooperative filtering-based recommendation system[7]. According to these researches, the recommendation accuracy can be improved by figuring out the similarity between users and selecting the neighbor users based on the user vector information or product vector information derived from Word2Vec instead of directly using the user rating to figure out the user similarity. The similarity-based recommendation

†Corresponding author : tykim@chosun.ac.kr

system using the neighbor user, however, fundamentally involves an issue of difficulty in making recommendations for new users or new products.

These days, deep learning is showing a great performance in the image processing or natural language processing fields. In response, deep learning is applied on various fields and there are recent researches on applying the deep learning to the recommendation system[8]. For example, a research came up with the recommendation system that utilizes the recurrent neural network system in the environment where user rating information doesn't exist while another research pointed out that the Word2Vec and deep neural network-based recommendation system shows higher recommendation accuracy than the conventional user-based collaborative filtering method in the environment with the user rating information[9]. In addition, image processing field is utilizing the deep learning technology most actively and the deep learning technology is showing a remarkable performance in the image pattern recognition based on the Convolutional neural network (CNN) algorithm[10,11].

This study tries to suggest on applying the Word2Vec, which is actively used in the text analysis field for the movie recommendation with the user rating information, and CNN algorithm, which is the most actively used in the image processing field, on the movie recommendation system. The proposed system first uses Word2Vec to figure out the user vector and movie vector. Then, the user and movie vectors are made into the learning data. After that, movie is recommended by learning the CNN algorithm. For the evaluation on the performance of the proposed system, this study compares the user-based collaborative filtering system with the deep neural network-based recommendation showing the higher recommendation accuracy[12].

2. Recommendation Modeling

This study proposes a CNN algorithm-based movie recommendation system. The learning network for learning the user's movie rating is composed as Fig. 1.

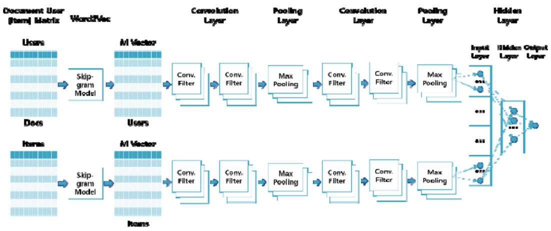

Fig. 1. Learning network configuration.

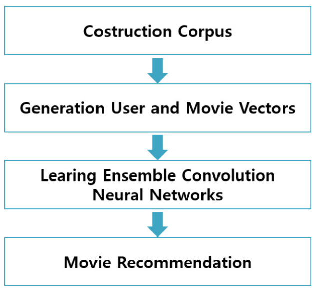

Fig. 2. Movie recommendation algorithm.

The Word2Vec and ensemble CNN algorithm-based movie recommendation stages are as Fig. 2.

In the text analysis, a document is divided into sentences and the words inside sentences become a corpus. As the movie data is composed in (User ID, Movie ID, and Movie Rating) structure, the sentence cannot be composed like the general text analysis. Instead, the intention sentences are needed for the analysis. In the corpus composition step, methods such as wDNN are used to generate user sentences and movie sentences by using the information on the movies that user watched and the user's movie ratings. This results in user corpus and movie corpus. Sentences are composed with words and this study regarded the user ID and movie ID as the words. The user sentences are composed only with user ID while movie sentences are composed only with movie ID.

The processes of generating the sentences composed of user ID are as follows. First, all user IDs that watched the same movie and their movie ratings are searched. Then, movie sentences are made by regarding the user ID as the word. For instance, a sentence is made by finding all user IDs that watched Movie A and using the lists of all users that gave the same points to the movie. Then, multiple sentences covering from the lists of users that gave the highest points to the lowest points are generated for Movie A. This process is repeated for all movies to generate the whole user sentence. Here, a corpus is made by combining the words (user ID) used in each sentence. In the movie sentence, movies are found and the sentences are made by using the lists of movies that got the same user ratings. After repeating this process for the all users, the whole movie sentences can be made. Then, a movie corpus is made by combining the words (movie ID) used in each sentence.

In the Word2Vec-based user vector generation, the user vector is generated by inputting the whole user sentences from the corpus composition step. Word2Vec uses the Gensim for a Python library. Word2Vec requires setting the window size to find the correlation between the words and word vector. This study set the word vector size to 200 and window size to 10. The 200 dimensional vectors figure out the correlation between words by taking 10 neighboring words (window size) in each sentence. The same process is repeated for the movie vector. As a result, each user ID and movie ID gets 200 dimensional vectors.

The raw data composed of user ID, movie ID, and user ratings is converted into the user vector, move vector, and user ratings to organize the learning data and use the CNN algorithm for learning. The whole model for the ensemble CNN algorithm is as Fig. 3. In this study, CNN algorithm-based learning uses Keras, a Python-based open source library, and Keras uses TensorFlow as a deep learning engine.

For the input data, the convolution layer uses the convolution kernels to extract the important characteristics of the designated area. The designated area is significant only when it is composed with correlated data. As 200 dimensional vectors indicating the user ID and movie ID are not correlated together, there should be a respective convolution layer for user ID and movie ID. In response, this study came up with an ensemble convolution model which is composed with the user convolution model (user model) that learns by input user vector and the movie convolution model (movie model) that learns by input movie vector separately and combines the two models in the input stage of the fully-connected stage.

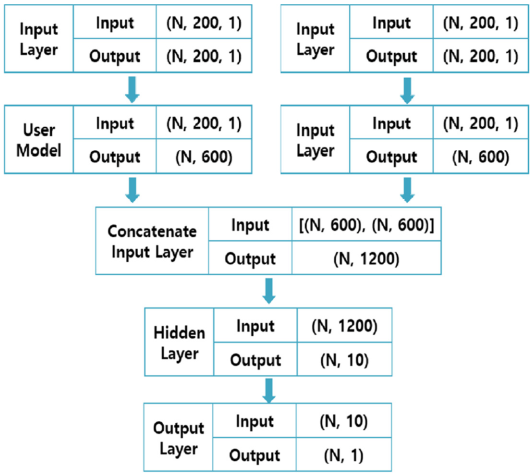

Fig. 3. The ensemble CNN algorithm model.

The input layer receives the random numbers of learning cases composed of 200 dimensional vectors and outputs without a conversion process. The user model and movie model process the input data through the convolution layer and pooling layer and ultimately output the random numbers of 600 dimensional vectors. The user model and movie model's output vectors are connected and integrated into 1200 dimensional vectors and inputted as the fully-connected layer (1200 input nodes for the fully-connected layer). The fully-connected layer's hidden layer is composed of 10 nodes and the data of the 1200 dimensional vectors is converted into 10 dimensional vectors for the output. The hidden layer's activation function is ReLU. The output layer composed of 1 node receives 10 dimensional vectors and outputs as 1 dimensional scalar. The activation function is not specified and the output layer's output is applied with the CNN algorithm to forecast the actual user rating from the cases of learning. "RMSprop" is

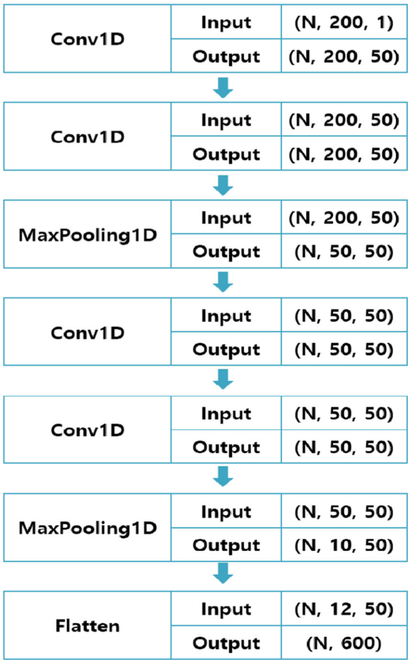

Fig. 4. User and movie CNN model.

used as the optimization method of the convolution model and “MSE” is used for the loss function.

The composition of the user model and movie model are the same in Fig. 6 and the details are as Fig. 4.

Generally, the convolution layer of the CNN algorithm mainly process images and it uses the two dimensional filter (kernel) to process the convolution. On the other hand, the user vector and movie vector in this study have the form of one-dimensional array of 200 dimensional vectors. Therefore, the convolution needs to be processed by using the one dimensional filter. In the one dimensional convolution processing, this study used Conv1D of Keras 2.1.3 Library and set the filter size to 5, filter number to 50, padding to “Same”, activation function to “ReLU, and stride to 1. Conv1D convolves the input data and one dimensional filter for the output and the output result becomes a feature map. As the filter number is set to 50, the convolution layer outputs 50 feature maps for each 200 dimensional individual user vector.

The pooling layer uses MaxPooling1D of Keras to compress the one-dimensional array data and the pooling size is set to 4. The input data is compressed in 1/4 size. MaxPooling1D receives the input vector data, sort out 4 elements in order, selects the biggest number among elements, and discards the remaining elements. The second pooling layer receives the input of 50 feature maps from the 50 dimensional vectors and outputs 50 feature maps compressed into the 12 dimensional vectors.

The flatten layer converts to a form of one-dimensional array for 600 (12*50=600) dimensional vectors.

By learning the CNN algorithm and inputting the user vector and movie vector for the recommended movie, the expected user rating is obtained from the output. As a result, the user is recommended a movie in the order of the highest user ratings.

3. Experiments and Results Analysis

To assess the performance of the proposed system, this study used the FilmTrust data provided by LibRec. The FilmTrust data covers the ratings of 1,000 users on the 1500 movies from 0.5 points to 4 points. The FilmTrust data is composed of 3 parts, {user ID, movie ID, movie Rating}, and 24,385 rating cases with the data density of 1.05%.

For the criteria on assessing the recommendation accuracy of the proposed system, this study used MAE (Mean Absolute Error), the mean of the absolute value and forecast ratings, commonly used the assessment of the accuracy. The whole data is divided into learning data (90%) and test data (10) and the learning data was used to build a movie recommendation model and forecast the user ratings in the test data. To build experiment models and set parameters, this study used the part of the learning data as the validation data and found the learning model. After confirming the learning model, this study used the whole learning data for learning the model. To assess the performance of the test data, this study carried out the 10-fold cross validation. For the experiment, this study used Python 3.9, Keras 2.1.3 (Open library of Python), TensorFlow, and Word2Vec of Gensim library under the hardware environment of i7 processor, 16G memory, and GTX 1060. The hyper parameter, output vector size M and window size W, of Word2Vec, M was set to 200 and W was set to 100. In composing the CNN algorithm, this study tested the numbers of forms in the pre-experiments and ultimately chose the models in Fig. 2 and Fig. 3 with relative higher performance. For the experiment, user and movie convolution models were finally selected. In the user and movie convolution models, this study set the convolution layer's numbers of kernel to 50, kernel size to 5, padding to "Same", activation function to "ReLU", and stride to 1 while setting the max pooling with the pooling of 4 in the pooling layer. In the fully-connected layer, hidden layer 1 was adopted while "ReLU" was adopted as the activation function for 10 hidden layers in the hidden nodes. Additionally, dropout was set to 0 based on the pre-experiments on the dropout. To assess the recommendation accuracy of the proposed system (wCNN), this study compared it with a system that utilizes the deep neural network (wDNN).

wDNN is a deep neural network consisting of hidden layer 2 for the 400 dimensional data obtained by combining the user vector and movie vector from Word2Vec. CF stands for the conventional user-based cooperative filtering. The results for 10-fold cross validation on the experiment data are as Table 1. The MAE (Mean Absolute Error) for wDNN was 0.6672 while MAE for wCNN was 0.6585 implying that the wCNN proposed in this study shows higher accuracy than wDNN.

Additionally, in the 10-fold cross validation including the new user or new product from substituting the new user with user vector's mean vector and substituting the new product with movie vector's mean vector, the wDNN had MAE of 0.6813 while wCNN had MAE of 0.6509. Such results proved that the accuracy of the movie recommendation can be improved by utilizing Word2Vec and ensemble CNN algorithm.

Table 1. Periodic table of elements

No	Collaborative Filtering	wDNN	wCNN
1	1.0074	0.6753	0.6413
2	0.9865	0.6653	0.6721
3	1.0356	0.6529	0.6318
4	1.0234	0.6750	0.6510
5	1.0947	0.6884	0.6398
6	1.0304	0.6477	0.6972
7	0.9783	0.6713	0.6554
8	1.0076	0.6814	0.6552
9	1.0216	0.6608	0.6763
10	0.9511	0.6534	0.6651
Mean	1.0137	0.6672	0.6585

4. Conclusion

In response to increasing numbers of products traded in the e-commerce, it is getting more important to increase the recommendation accuracy in the web-based product recommendation system. While the cooperative filtering system was widely used in the past, there are Word2Vec or deep learning-based systems these days. This study proposed Word2Vec and ensemble CNN algorithm-based system to enhance the recommendation accuracy in the movie recommendation. The proposed method used the Word2Vec to figure out the user vector and movie vector, came up with convolution model and movie convolution model that forecast ratings from the input of user vector and movie vector respectively, and built up the ensemble CNN algorithm model for combining the two models together. This study composed the convolution model with the convolution layer that processes the convolution with the one-dimensional filter, pooling layer for the 1D max pooling of the one-dimensional input, and flatten layer for connecting with the fully-connected layer's input layer.

This study also compared the proposed system (wCNN) with wDNN to assess the recommendation accuracy. The recommended system remarkably improved the accuracy of the movie recommendation system compared to the conventional user-based cooperative filtering. In addition, wCNN showed higher accuracy than the deep neural network-based system (wDNN).

In addition, this study suggested a system that utilizes both Word2Vec and ensemble CNN algorithm to enhance the recommendation accuracy in the movie recommendation and verified the system by using the FilmTrust data. In order to generalize the proposed system, it is necessary to apply the system on the different data additionally. Furthermore, deep learning technologies are developing quickly and there needs to be further studies on improving the accuracy by building the ensemble model including such new models.

References

[1] Lee, Y., Won, H., Shim, J., and Ahn, H., "A Hybrid Collaborative Filtering-based Product Recommender System using Search Keywords", *Journal of Intelligence and Information Systems*, vol. 26, no. 1, pp. 151-166, 2020.

[2] Cho, S. E., and Lim, H. S., "A Study on Product Recommendation System Based on User Search keyword", *Journal of Digital Contents Society*, vol. 20, no. 2, pp. 315-320, 2019.

[3] Park, S. J., Kim, Y. M., and Ahn, J. J., "Development of product recommender system using collaborative filtering and stacking model," *Journal of Convergence for Information Technology*, vol. 9, no. 6, pp. 83-90, 2019.

[4] Turner, C. A., Jacobs, A. D., Marques, C. K., Oates, J. C., Kamen, D. L., Anderson, P. E., and Obeid, J. S., "Word2Vec inversion and traditional text classifiers for phenotyping lupus," *BMC medical informatics and decision making,* vol. 17, no. 1, pp. 1-11. 2017.

[5] Hu, K., Wu, H., Qi, K., Yu, J., Yang, S., Yu, T., Zheng, J., and Liu, B., "A domain keyword analysis approach extending Term Frequency-Keyword Active Index with Google Word2Vec model", *Scientometrics*, vol. 114, no. 3, pp. 1031-1068, 2017.

[6] Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., and Dean, J., "Distributed representations of words and phrases and their compositionality", *In Advances in neural information processing systems*, pp. 3111-3119, 2013.

[7] Park, S., J., and Byun, Y., C., "Improving Recommendation Accuracy based on Machine Learning using Multi-Dimensional Features of Word2Vec", *Journal of Advanced Information Technology and Convergence*, vol. 19, no. 3, pp. 9-14, 2021.

[8] Zhao, Z., and Chen, X., "Deep Reinforcement Learning based Recommend System using stratified sampling", *In IOP Conference Series: Materials Science and Engineering*, vol. 466, no. 1, pp. 012110, 2018.

[9] Devooght, R., and Bersini, H., "Collaborative filtering with recurrent neural networks", arXiv preprint arXiv:1608.07400, 2016.

[10] Liu, S., Ji, H., and Wang, M. C., "Nonpooling convolutional neural network forecasting for seasonal time series with trends", *IEEE transactions on neural networks and learning systems*, vol. 31, no. 8, pp. 2879-2888, 2019.

[11] Xu, W., Huang, H., Zhang, J., Gu, H., Yang, J., and Gui, G., "CNN-based skip-gram method for improving classification accuracy of chinese text", *KSII Transactions on Internet and Information Systems (TIIS)*, vol. 13, no. 12, pp. 6080-6096, 2019.

[12] Fong, C. M., Wang, H. W., Kuo, C. H., and Hsieh, P. C., "Image quality assessment for advertising applications based on neural network", *Journal of Visual Communication and Image Representation*, vol. 63, pp. 102593, 2019.