Exploit Multi Layer Deep Learning and Latent Factor to Handle Sparse Data for E-commerce Recommender System

Hanafi[1], Nanna Suryana Herman[2] and Abdul Samad Hasan Basari[2]
[1]*Departement of Computer Science University of Amikom Yogyakarta Yogyakarta, Indonesia*
[2]*Faculty of Information and Communication Technology University Teknikal Malaysia Malaka Malaka, Malaysia*

Keywords: E-commerce, Recommender system, Deep learning, LSTM, Sparse data, Collaborative filtering.

Abstract: E-commerce service have become popular way to shopping in recent decade. E-commerce machine requires a method to provide fit product information to customer called recommender system. The most of popular recommender system adopted for many large e-commerce companies named Collaborative filtering (CF) due to obtain relevant, fit and essential product information. Even though CF owned some benefit, it has shortcoming inaccurate recommendation when face minimum rating also popular named sparse data problem. Many researches have been conducted to proposed model how to generate rating prediction aim to handle sparse rating effectively. Most of them exploit latent factor model or matrix factorization (MF) to handle this problem, unfortunately, this problem fails to handle the problem when faced serious sparse data. Aims to improve the serious problem on above, several researchers involve auxiliary information in the form of product document or user demographic information respectively. Several researchers implemented Convolutional Neural Network (CNN) to extract product document review incorporate MF that responsible to produce rating prediction, another model exploited Stack Denoising Auto Encoder (SDAE) model as user demographic information extraction incorporate with MF. In this research, considered implementing dual information representation using deep learning model based on SDAE and Long Shorts Term Memory (LSTM) as product review document representation combined into PMF to generate rating prediction. According to experiment report, the proposed model called SLP (SDAE+LSTM+PMF) successful to obtained effectiveness rating prediction based on RMSE evaluation metrices over some current model based on traditional PMF more than 15% in average and superior over CNN more than 0.9% in average.

1 INTRODUCTION

Recommender system is one of the most important tools to build success e-commerce business company. Successful applied recommender system, it would influent the selling target achievement in online transaction. Many large world companies have been implementing recommender system to increase satisfied service to their company to make customer enjoyable looking for the product. It is a essential equipment to promote sales and services for many online websites and mobile applications. For instances, 80 percent of movies watched on Netflix came from recommendations (Gomez-Uribe and Hunt, 2015), 60 percent of video clicks came from home page recommendation in YouTube (Davidson et al., 2010). According to (Schafer et al., 2001) found that sales agents with recommendations by the NetPerceptions system achieved 60% higher average cross-sell value and 50% higher cross-sell success rate than agents using traditional cross-sell techniques based on experiments conducted at a U.K.-based retail and business group.

Based on algorithm approach (Hanafi et al., 2018), e-commerce recommender system divided into 4 types as follow: 1). Content based: the method to generate recommendation according to product classification approach. In the other hand, it is tending information retrieval to generate product recommendation 2). Knowledge based: this approach is to develops for specific necessary recommendation, the specific character is to provide product information rarely needed for individuals purpose for example house, loan, insurance, car. 3). Demographic based: product recommendation in which established to provide product recommendation according to demographic information. 4). Collaborative Filtering (CF): the mechanism to produce recommendation

Hanafi, . and Basari, A.
Exploit Multi Layer Deep Learning and Latent Factor to Handle Sparse Data for E-commerce Recommender System.
DOI: 10.5220/0009910603430351
In *Proceedings of the International Conferences on Information System and Technology (CONRIST 2019)*, pages 343-351
ISBN: 978-989-758-453-4

based on user attitude in the past such as rating, product review, comment, testimony, purchasing and etc. CF is the most successful technique that implemented in many large e-commerce company due to CF have ability to provide recommendation with special character as follow; provide product fit, serve relevant information, accurate, serendipity (Ricci et al., 2015). In common use, CF adopted rating as explicit feedback as basic compute to calculate similarity user for product uses rating matrix to generate product recommendation. The big challenge in collaborative filtering is just slight of user population who gave rating for product, totally about only less than 1 percent. The problem popular called sparse data also in extreme condition sparse data famous called cold start. When cold start happens, there is no recommendation possible to generated by recommendation system.

CF having several advantages over another approach such Content base, Demographic base, Knowledge base. However, CF have essential limitation caused lack in rating collection from customers. CF rely on rating as basic calculation to generate product recommendation. Several attempts have been made to reduce sparse data, so they hope the product recommendations method having more accurate. The use of auxiliary information has proven to improve the accuracy of a product recommendation. Some of the auxiliary information that have been explored to handle sparse data have been done for instance audio features in music recommendations (Van den Oord et al., 2013) (Wang and Wang, 2014), color features on online fashion shop (Jaradat, 2017), documents recommendation in online news (Park et al., 2017).

With the final objective to beat the issue of information flooding, utilizations of recommendation system have pulled in a lot of consideration in last decade. Numerous business companies have included recommender system into their e-commerce intelligent system, for example, Amazon1, JD2, Taobao3, and, etc. One viable technique for recommendation is to estimate new rating of various users and items. Presently, the most famous strategies in this field are Content-based Filtering (CBF) and Collaborative Filtering (CF). CBF utilizes the setting of users or items to estimate new rating. For instance, we can create user inclinations as indicated by their age, sexual orientation, or diagram of their colleague, and etc. Likewise, genres and items review can be exploited to generate product preference for various users. On the opposite side, CF utilizes the rating of users for items in their review to estimate new ratings. For example, a table of N users $\{u_1, u_2, u_3, u_4, u_5, \ldots u_N\}$ and table of M items $\{v_1, v_2, v_3, v_4, v_5, \ldots v_N\}$. Concurrently, a table of items v_{11} , given rated by users u_i. These ratings can either be explicit feedback on rating star 1-5, or implicit feedback on a scale of 0-1. Our objective is to estimate new feedback from users without records. Moreover, CF often more powerful over CBF due CF (F. Ricci and Saphira, 2011). Fig. 1 is example of movie review from users that very popular to integrated into latent factor model in the term of matrix factorization.

Figure 1: example review product for movie

One of the most outstanding methodologies in CF are Probabilistic Matrix Factorization (PMF) (Mnih and Salakhutdinov, 2008) and Singular Value Decomposition (SVD) (Sarwar et al., 2001). However, they are usually against with sparse data (extreme sparse in cold start issue), so they stall to squeeze effective description from users and items. To address the trouble related with this issue, a functioning line of research since the previous decade and several of method have been proposed (Zhou et al., 2011) (Yi et al., 2016) (Trevisiol et al., 2014). All of these methods consider utilizing the side information of users or items into CF to produce more effective characteristic. In addition, the most normally utilized side information text documents, therefore several methods use text document approach, for example, Latent Dirichlet Allocation and Collaborative Topic Modelling have been put forward by (Ling et al., 2014), (Wang and Blei, 2011)hile earlier research has inspected text document as a bag-of-words, it might be desirable over ponder the effect of sequences of words in text documents; consequently, past techniques are constrained to several expand.

Figure 2: Rating given by users.

According fig. 2, majority problem of sparse data problem caused minimum rating.

2 PREVIOUS WORK

DDeep learning is one of the machine learning approaches that used to solve data sparse problems and overcome some of the weaknesses in several approaches described above. Deep learning is a derivative of a resurgent neural network because it can produces outperform result in image processing, natural language processing. Some experts attempt to empower deep learning in dealing with problems that exist in the recommender system. Several study was use deep learning to extract content feature aims to handle sparse data in rating product, such as (Van den Oord et al., 2013) enhance Deep Convolutional Neural Network to create audio music classifier aim to develop auto music recommendation, due no data product record, the recommendation difficult to generate.

According (HANAFI et al., 2018) where author conducting approach to eliminate sparse data by combining between multi-layer neural network and non-negative matrix factorization, even the result of hybridization could be work, however the results are less accurate than some studies that use similar technique.

No	Method	Description	Ref.
1	PMF	Probabilistic Matrix Factorization (PMF) is standard rating prediction approach that only involve rating for collaborative filtering	[11]
2	NMF	Recommender system based on collaborative filtering applied use non-negative matrix factorization (NMF) to generate recommendation.	[19]
3	SVD	Recommender system based on collaborative filtering applied use Singular value decomposition (SVD) as low rank dimensional factorization.	[20]
4	CTR	Collaborative Topic Regression (CTR) is a state-of the art recommendation model, which combines collaborative filtering (PMF) and topic modeling (LDA) to use both ratings and documents.	[17]
5	CDL	Collaborative Deep Learning (CDL) is another state-of-the-art recommendation model, which enhances rating pre- diction accuracy by analyzing documents using Deep Learning approach.	[21]
6	CONVMF	Novel model Collaborative Filtering that includes item document representation using CNN incorporate with traditional latent factor based on PMF.	[22]
7	PHDMF	Novel model Collaborative Filtering that incorporate user information using SDAE dan item document information using CNN mixed into latent factor based on PMF.	[23]

Figure 3: comparison of state-of-the-art approach.

Source of Figure 3 : (Mnih and Salakhutdinov, 2008), (Wang and Blei, 2011), (Zhang et al., 2006), (Wang et al., 2015b), (Sarwar et al., 2002), (Park et al., 2017), (Liu et al., 2017).

3 OUR CONTRIBUTION

Since last decade deep learning have become trending to solving in several computer science research field because of outstanding improvement. Feature extraction of deep learning has solid characteristic, the increase in the number of studies has adopted deep learning technology by implementing side information to produce effective description. For example, Deep Recurrent Neural Network (RNN) for News Recommendation (Park et al., 2017), another Author proposed a model to dig user rating latent factor using Stacked Denoising Autoencoder (SDAE) (Wang et al., 2015a) (Wang et al., 2015b) and Convolutional Neural Network (CNN) (Wang et al., 2015a) a.k.a.

ConvMF, which difference angle the inquiry of words in text documents. In any case, ConvMF just thinks about item side information (e.g., text documents, review, synopsis, abstract, etc.), so users' latent factors still have no effective description. Another approach by Author (Kim et al., 2016) shows that SDAE is proficient at extricating remarkable feature in users' latent factor without involving text documents, this causes items latent factor stay with equal classical methodologies. Hence, according to handle the problem on above, we incorporate SDAE-NN and CDNN into a probabilistic model to develop more effectively extract users' latent factor and items latent factor.

1. We propose hybrid model involve users latent factor layer by using SDAE-NN and items latent factor by using DC-NN, both of them incorporating by probabilistic model. Our proposed model called MultiLayer Deep Learning (MLDL) recommender system. To the best our insight, MLDL is the pioneer model to incorporating two deep learning layers (SDAE-NN and DC-NN) into probabilistic point of view.

2. We widely show that MLDL is a mixing of a several best performance techniques but with a more effective representation.

3. We establish unique strategy and conduct examinations which demonstrate that MLDL successful to eliminate CF sparse data issue.

4 PROBLEM DEFINITION

Similar with some existing state of the art, this experiment adopts document of product review to integrated into probabilistic matrix factorization to produce rating prediction (PMF). Following to famous pre-processing procedure, we have n users, m items, and an extreme sparse in rating matrix $R \in R^{nxm}$. Every input R_{ij} of R corresponds to user's i rating on item j. Involvement document product review explain in below, the auxiliary information of users and items are denoted by $X \in R^{nxe}$ and $Y \in R^{mxf}$, respectively. Let u_i, $v_j \in R$ be user i latent factor vector and item j latent factor respectively, where k is the dimensionality of latent space. In this process, the corresponding matrix form of latent factor for users and items are U = *u[1:n]* and V = *v[1:m]*, separately. Given the sparse rating matrix R and the side information matrix X and also Y, our objective is to learn effective users latent factor U and Item latent factor V, and after that to estimate the missing rating in R.

5 OUR APPROACH

In this section, we consider explaining three essential method to develop SLP (SDAE-LSTM-PMF) to produce rating prediction to handle sparse data in recommender system based on collaborative filtering.

5.1 A. Probabilistic Matrix Factorization (PMF)

PMF is traditional latent factor model using matrix factorization to produce rating prediction. This model considers Gaussian Normal Distribution to create user and item latent factor representation. First model proposed by Salakhutdinov (Mnih and Salakhutdinov, 2008). Unfortunately, majority collaborative filtering based on traditional latent factor whether PMF, SVD, SVD involve temporal effect, Non-Negative Matrix Factorization (NNMF) obtain inaccurate rating prediction when faced with sparse data. Aim to handle this problem, considering item or user side information are needed. In this experiment, considered PMF required to obtain rating prediction integrated with SDAE and LSTM.

5.2 SDAE to Extract User Demographic Information

SDAE is sub class of neural network model where it is based on feed forward approach. This model very popular to adopt in deep learning mechanism. Basic concept of this approach by follow the term of autoencoder mechanism by using feedforward neural network that it has an input layer, one hidden layer and an output layer as follow figure 5. The output layer has same number of neurons as the input layer for the purpose of reconstructing its own inputs. This makes an autoencoders a form of unsupervised learning, which means no labelled data are necessary just a set of input data instead of input-output pairs. In SLP model (SDAE-LSTM-PMF, SDAE consider extracting user demographic information representation to transform within 2D latent space. User demographic information were used as input of SDAE for feature learning representation. The result of SDAE output in the term of 2D latent space would be integrated with PMF. Before SDAE processes be conducted, user demographic information would be pre-processing using vectorized to transform user information into vector space.

5.3 Long Shorts Term Memory (LSTM) to Extract Product Review Document

TItem document information has become popular information to support latent factor model to improve performance in producing recommendation in last decade. Majority of previous work adopted LDA model to interpreted product review document. Some of researcher using CNN to make more deeper understanding of product document interpretation. However, several of them fail to capture contextual understanding of product review document due to ignore sequential word aspect to develop latent semantic representation. Aims to improve the problem in contextual understanding, this research considers to adopted LSTM model to detect contextual understating of product review document.

LSTM method adopted to transform item document product review. The detail explanation of our model shows on Fig 3 below where item document obtains from IMDB, then pre-processing process using NLTK (Natural Language Tool Kits) module conduct some process such as lemmatization, stop word removal, remove punctuation and vectorized into one hot encoding. After pre-processing stage success to implemented, exploiting LSTM using TensorFlow would be utilities in the next processes. The final process resulted 2D latent space in the form 50 dimensional. So, the document product review as product representation would integrate with SDAE as user information representation into latent factor model based on PMF.

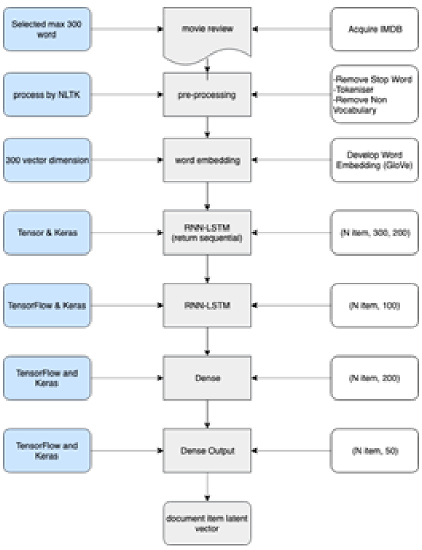

Figure 4: LSTM to extract item document information

5.3.1 Hybrid SDAE, LSTM and PMF

The success story of model based using latent factor based on matrix factorization popularized several researchers and academicians in Netflix competition event on early 2006. Majority of matrix factorization success to improve memory-based model significantly. Began in this competition, model based using matrix factorization become favorited model to increase performance of collaborative filtering recommendation.

There are several researches considered to integrated deep learning model with matrix factorization to increase performance in handling sparse data due to deep learning has tremendous achievement for several computer sciences field such Audio, image processing, sentiment analysis, text mining. A kind of deep learning class is convolutional neural network, feedforward neural system that has effectively been employed to Computer Vision (Krizhevsky et al., 2012), Natural Language Processing Kim (Kim, 2014) and Audio Signal Processing (Piczak, 2015). Essentially, there are a few examinations have utilized CNN in recommender system, for example, content-based music recommendation (Van den Oord et al., 2013), fashion shop based colour feature (Jaradat, 2017), which opposed with conventional method utilizing a bag of words as description of the audio signal with deep CNN, and Dynamic Convolutional Neural Network (DCNN), which utilizes CNN to produce items' latent factors however this model lack in users latent factor due the users security issue. In other hand, their effort just thinks one sided latent factor (i.e., item auxiliary information), which uses estimated rating not similar to SVD via gradient descent. Hence, we utilize items' latent factor developed by LSTM and user' latent factor by a family of SDAE. We have prepared SDAE model which consideration imputing missing value to increase effective representation of user latent factor. In the previous work, consideration integrated user and item information representation into matrix factorization has been made that called PHDMF. This model exploit CNN to capture document of product review understanding. Different with PHDMF, our approach exploits LSTM to capture contextual understanding of product review document. The detail of SLM model demonstrated on Figure 5 below.

Figure 5: Integrate SLM model include SDAE, LSTM and PMF

Figure 5 on above show overview of the probabilistic matrix factorization approach to support SLP, which combine SDAE and LSTM model into PMF. Based on a probabilistic point of view, the formula given by:

$$p(R|U,V,\sigma^2) = \prod_i^N \prod_j^M N(R_{ij}|u_i^T v_j, \sigma^2)^{I_{ij}} \quad (1)$$

Where $N(x|\mu,\sigma^2)$ is the probability density function of the Gaussian normal distribution with mean μ and variance σ^2. Thus, to finalize users' latent factor can be produce by reference equation on below.

$$u_i = sdae(W^+, X_i) + \varepsilon_i$$

$$\varepsilon_i = N(0, \sigma_U^2 I) \quad (2)$$

To calculate $w_k^+ in W^+$ we consider using zero-mean spherical Gaussian prior, equation according on below.

$$P(w^+|\sigma_{w+}^2) = \prod_k N(w_k^+|0, \sigma_{w+}^2) \quad (3)$$

Finally, the distribution over user latent factor given by:

$$p(U|W^+, X, \sigma_U^2) = \prod_i^n N(u_i|sdae(W^+, X_1), \sigma_u^2) \quad (4)$$

Similarly, the user's latent factor, an item latent factor given by formulation structure:

$$v_j = lstm(W, Y_j) + \varepsilon_j$$

$$\sigma_j = N(0, \varepsilon_v^2 I) \quad (5)$$

dcnn representation of the output of dcnn approach, also use zero-mean spherical Gaussian prior.

$$P(w\sigma_v^2 = \prod_k N(N_k, 0, \sigma_w^2)) \quad (6)$$

$$p(V|W,Y,\sigma_v^2) = \prod_j^m N(v_j|dcnn(W,Y_j), \sigma_v^2 I) \quad (7)$$

5.4 Evaluation Measure Metric

RMSE is frequently used measure of the differences between values (sample and population values) predicted by a model or an estimator and the values actually observed. Root Mean Squared Error (RMSE) is might the most famous metric used in evaluating accuracy of predicting rating. The system generates predicted ratings $\bar{r}_{ui}$for test set t of user-item pairs (u,i) for which the true rating r_{ui} are known. Typically, r_{ui} are known because they are hidden in an offline experiment. The RMSE between the predicted and actual rating is given by:

$$RMSE = \sqrt{\frac{1}{|t|} \Sigma_{(u,i)\in t} (\bar{r}_{ui} - r_{ui})^2} \quad (8)$$

6 EXPERIMENT SETTING

This research focus to develop a method to handle user rating sparse data extremely in case collaborative filtering recommender system involve deep learning machine to find relationship latent factor between user rating from a product. According our best knowledge, in state of the art by use matrix factorization has failure to addressing user rating sparse data extremely. So, the result of recommendation sometime getting the mistake (Koren et al., 2009). This research is categorical lab scale research involve public datasets. We consider MovieLens dataset as convince datasets were applied in so many study in recommender system research field, detail specification refer to (Harper and Konstan, 2015).

6.1 Device and Library Tools

Our experiment involves several tools include software and hardware. There are several tools and software that involve includes Python with several libraries such as tensor flow for deep learning implementation and GeForce GTX 1001 for running convolutional neural networks supported by processors that we use Xeon 2.4 Ghz.

No	Device and Library	Specification
1	Processor	Intel Xeon Quad core, 2.4 Ghz
2	Memory	16 GB
3	GPU	GeForce GTX 1001
4	Tensor Flow	Deep learning tools
5	Keras	Deep learning tools
6	Anaconda	Web interface
7	Python	Tool programming
8	Scikit-learn	TF-IDF, Bag of Word
9	Pylearn	
10	Surface	Data analytics
11	NLTK	NLP module
12	Matplotlib	Data analytics visualization
13	NumPy	Matrix Factorization

Figure 6: Detail specification device and library.

6.2 Datasets

In this research to understand the performance of SLP model, we considered to implemented on three real datasets obtain from Movielens and AIV. The detail characteristic of datasets shows on Figure 7 below.

Dataset Table	Item Side Information	Users	Items	Items Ratings
100K	movie descriptions	943	1,546	94,808
ML-1M	movie descriptions	6,040	3,544	993,482
ML-10M	movie descriptions	69,878	10,073	9,945,875
Amazon	Movie reviews	81,339	18,203	238,352

Figure 7: Characteristic of Datasets.

To show the effectiveness of our approach in terms of rating prediction. Firstly, we applied real-world datasets acquired from MovieLens (Harper and Konstan, 2015) and Amazon 3. These datasets contain of consumer' explicit ratings for products on rating scale of 1 to 5. Amazon dataset contains opinion for products as item description documents. Because of MovieLens does not include item description documents, we generate the documents use corresponding items from IMDB server. Equal with (Liu et al., 2017), we conduct to preprocessed description documents for all of them datasets as follows: 1) setting the data with maximum length of raw documents to 300, 2) removing stop words, 3) calculated tf-idf score for each word, 4) removing corpus specific stop words that have the document frequency higher than 0.5, 5) Choose top 8000 distinct words as a vocabulary, 6) removing all non-vocabulary words from raw documents. As a result, average numbers of words per document are 97.09 on MovieLens-1M (ML-1M), 92.05 and Amazon Instant Video (AIV), We consider to deleted items that have no their description documents in dataset table, and specially for the case of Amazon dataset, we consider to removed user's data that have just only less than 3 ratings. The result, statistics of each data show that three datasets have different characteristics on table 2. Finally, even though some users have removed by preprocessing, Amazon dataset is still extremely sparse compared to the others.

7 RESULT

The result of experiment report shows on figure below include some scenario training process where we divided dataset into training and testing categories with interval 10%. So, our scenario experiment included 10/90, 20/80, 30/70, 40/60, 50/50, 60/40, 70/30, 80/20, 90/10. The complete experiment result demonstrated figure below.

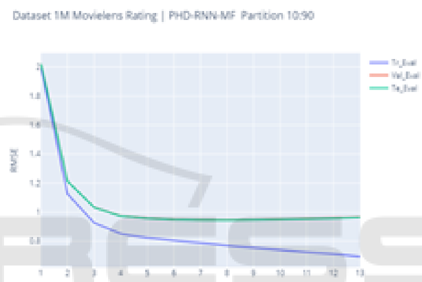

Figure 8: Evaluation metrices result ratio 10:90

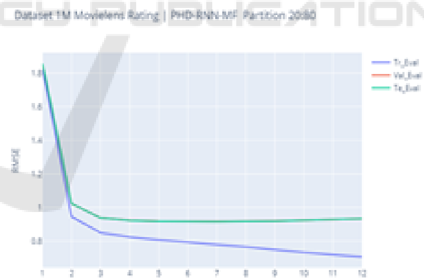

Figure 9: Evaluation metrices result ratio 20:80

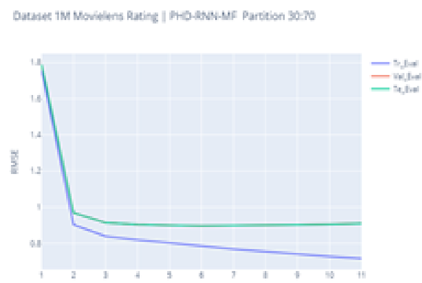

Figure 10: Evaluation metrices result ratio 30:70

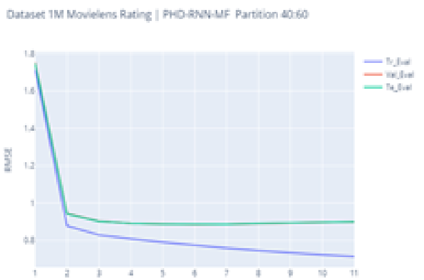

Figure 11: Evaluation metrices result ratio 40:60

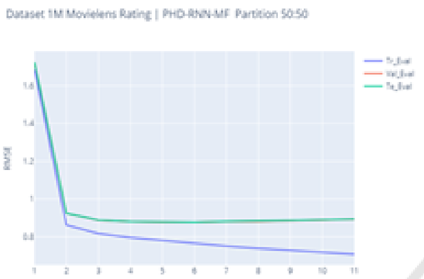

Figure 12: Evaluation metrices result ratio 50:50

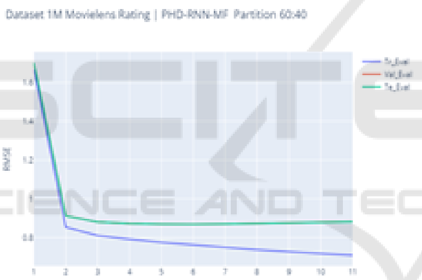

Figure 13: Evaluation metrices result ratio 60:40

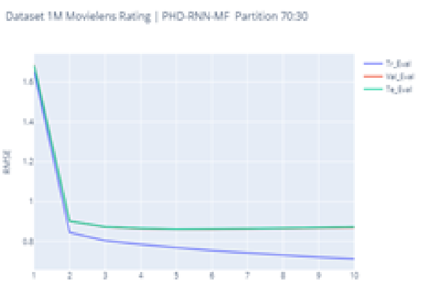

Figure 14: Evaluation metrices result ratio 70:30

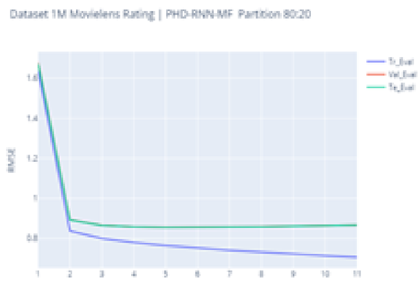

Figure 15: Evaluation metrices result ratio 80:20

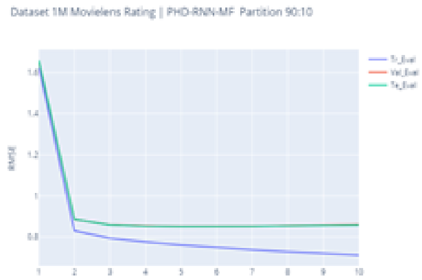

Figure 16: Evaluation metrices result ratio 90:10

8 PRELIMINARY FINDING

Based on this experiment it can be shown our proposed hybrid model called SLP success to produce rating prediction using ninth scenario training and testing process. According to the experiment report on above, the best performs on ratio 90:10 composition training and testing compare to another competition. User information latent factor and item document latent factor play important role to support PMF in producing rating prediction. Our plan for future experiment considers to implemented with several parameter and comparison with previous state of the art to understand effectiveness rating prediction level.

REFERENCES

Davidson, J., Liebald, B., Liu, J., Nandy, P., Van Vleet, T., Gargi, U., Gupta, S., He, Y., Lambert, M., Livingston, B., et al. (2010). The youtube video recommendation system. In *Proceedings of the fourth ACM conference on Recommender systems*, pages 293–296.

F. Ricci, L. R. and Saphira, B. (2011). Recommender systems handbook.

Gomez-Uribe, C. A. and Hunt, N. (2015). The netflix recommender system: Algorithms, business value, and innovation. *ACM Transactions on Management Information Systems (TMIS)*, 6(4):1–19.

HANAFI, SURYANA, N., BASARI, A. S. B. H., et al. (2018). Hybridization approach to eliminate sparse data based on nonnegative matrix factorization & deep learning. *Journal of Theoretical & Applied Information Technology*, 96(14).

Hanafi, SURYANA, N., BASARI, H., BIN, A. S., et al. (2018). An understanding and approach solution for cold start problem associated with recommender system: A literature review. *Journal of Theoretical & Applied Information Technology*, 96(9).

Harper, F. M. and Konstan, J. A. (2015). The movielens datasets: History and context. *Acm transactions on interactive intelligent systems (tiis)*, 5(4):1–19.

Jaradat, S. (2017). Deep cross-domain fashion recommendation. In *Proceedings of the Eleventh ACM Conference on Recommender Systems*, pages 407–410.

Jolly, P. (1982). Nickel catalyzed coupling of organic halides and related reactions.

Kim, D., Park, C., Oh, J., Lee, S., and Yu, H. (2016). Convolutional matrix factorization for document context-aware recommendation. In *Proceedings of the 10th ACM Conference on Recommender Systems*, pages 233–240.

Kim, Y. (2014). Convolutional neural networks for sentence classification. *arXiv preprint arXiv:1408.5882*.

Koren, Y., Bell, R., and Volinsky, C. (2009). Matrix factorization techniques for recommender systems. *Computer*, 42(8):30–37.

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105.

Ling, G., Lyu, M. R., and King, I. (2014). Ratings meet reviews, a combined approach to recommend. In *Proceedings of the 8th ACM Conference on Recommender systems*, pages 105–112.

Liu, J., Wang, D., and Ding, Y. (2017). Phd: a probabilistic model of hybrid deep collaborative filtering for recommender systems. In *Asian Conference on machine learning*, pages 224–239.

Mnih, A. and Salakhutdinov, R. R. (2008). Probabilistic matrix factorization. In *Advances in neural information processing systems*, pages 1257–1264.

Park, K., Lee, J., and Choi, J. (2017). Deep neural networks for news recommendations. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, pages 2255–2258.

Piczak, K. J. (2015). Environmental sound classification with convolutional neural networks. In *2015 IEEE 25th International Workshop on Machine Learning for Signal Processing (MLSP)*, pages 1–6. IEEE.

Ricci, F., Rokach, L., and Shapira, B. (2015). Recommender systems: introduction and challenges. In *Recommender systems handbook*, pages 1–34. Springer.

Sarwar, B., Karypis, G., Konstan, J., and Riedl, J. (2001). Item-based collaborative filtering recommendation algorithms. In *Proceedings of the 10th international conference on World Wide Web*, pages 285–295.

Sarwar, B., Karypis, G., Konstan, J., and Riedl, J. (2002). Incremental singular value decomposition algorithms for highly scalable recommender systems. In *Fifth international conference on computer and information science*, volume 1. Citeseer.

Schafer, J. B., Konstan, J. A., and Riedl, J. (2001). E-commerce recommendation applications. *Data mining and knowledge discovery*, 5(1-2):115–153.

Trevisiol, M., Aiello, L. M., Schifanella, R., and Jaimes, A. (2014). Cold-start news recommendation with domain-dependent browse graph. In *Proceedings of the 8th ACM Conference on Recommender systems*, pages 81–88.

Van den Oord, A., Dieleman, S., and Schrauwen, B. (2013). Deep content-based music recommendation. In *Advances in neural information processing systems*, pages 2643–2651.

Wang, C. and Blei, D. M. (2011). Collaborative topic modeling for recommending scientific articles. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 448–456.

Wang, H., Shi, X., and Yeung, D.-Y. (2015a). Relational stacked denoising autoencoder for tag recommendation. In *Twenty-ninth AAAI conference on artificial intelligence*.

Wang, H., Wang, N., and Yeung, D.-Y. (2015b). Collaborative deep learning for recommender systems. In *Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining*, pages 1235–1244.

Wang, X. and Wang, Y. (2014). Improving content-based and hybrid music recommendation using deep learning. In *Proceedings of the 22nd ACM international conference on Multimedia*, pages 627–636.

Yi, P., Yang, C., Zhou, X., and Li, C. (2016). A movie cold-start recommendation method optimized similarity measure. In *2016 16th International Symposium on Communications and Information Technologies (ISCIT)*, pages 231–234. IEEE.

Zhang, S., Wang, W., Ford, J., and Makedon, F. (2006). Learning from incomplete ratings using non-negative matrix factorization. In *Proceedings of the 2006 SIAM international conference on data mining*, pages 549–553. SIAM.

Zhou, K., Yang, S.-H., and Zha, H. (2011). Functional matrix factorizations for cold-start recommendation. In *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval*, pages 315–324.