Enhance Document Contextual using Attention-LSTM to Handle Sparse Rating Matrix For E-Commerce Recommender System

Hanafi

Department of computer science, University of Amikom Yogyakarta, Jl. Ringroad Utara Condoncatur Depok, Sleman, 55283, Indonesia

Corresponding author: hanafi@amikom.ac.id.ac.id

***Abstract*—E-commerce is the most important service in last 2 decade. E-commerce service influnce growth of economy impact in world wide. Recommender system is essential mechanism to calculate product information for e-commerce user. The successfulness in recommender system adoption influence target revenue of e-commerce company. Collaborative filtering (CF) is the most popular algorithm to create recommender system. CF applied matrix factorization mechanism to calculate relathionship between user and product using rating variable as intersection value between user and product. However, number of rating very sparse where the number of rating only less than 4%. Product Document is the product side information representation. The document aims to advance matrix factorization work. This research consider to enhancement document context using LSTM with attention mechanism to capture contextual understanding of product review and incorporate with matrix factorization based on probabilistic matrix factorization (PMF) to produce rating prediction. This study employ real dataset using MovieLens dataset ML.1M, ML.10M and IMDB to observed our model called ATT-PMF. Movielens dataset represent of number sparse rating that only contains below 4% (ML.1M) and below 1% (ML.10M). Our experiment report show that ATT-PMF outperform than previous work morethan 2% in average. Moreover, our model also suitable to implement on huge datasets. For further research, enhancement of product document context will promising factor to handle sparse data problem in big data issue.**

***Keywords*— sparsity data, recommender system; matrix factorization; e-commerce; attention mechanism; PMF.**

Manuscript received 15 Oct. 2020; revised 29 Jan. 2021; accepted 2 Feb. 2021. Date of publication 17 Feb. 2021.

I. Introduction

E-commerce is the most popular application to provide online transaction in the internet. E-commerce influence high impact in growth of glabal economic enhancement. In everyday live, we cannot escape from online trasaction such as news paper to read, video to watch, food to delivery, game to play, number of friend to confirm. That means, in every live we need online transaction.

E-commerce service require a mechanism how to serve essential information about product to customer or buyer candidate. This engine responsible to compute information that very famaous called recommender system. Recommender system is automatic engine to calculate product fit information for customer. The succesful adoption of recommender system influence target marketing value. Recommender system is very important tools to achive busineess revenue for e-commerce company.

There are four classification of recommender system algorithm model to build e-commerce imcluded collaborative filtering, content based, knowledge based and demographic based filtering. According several literature [1][2][3], collaborative filtering is the most effective algorithm to produce product recommendation. Collaborative filtering work using user behavior activity record in the past, especially rating information. Rating is an expression of satisfied representation about service or product. This is why collaborative filtering more usefull over another algoritm such as content based that consider product characteristic calculation.

In early model of collaborative filtering depend on memory based algorithm also popular nearest neighbor. The memory based algorithm very simple model to calculate product recommendation. They no need data training and aslo no need develop complex algorithm. Memory based suitable for simple model, small dataset and no more number of transaction. However, memory based have shortcoming when applying in new additional data dan gigantic datasets. The another essential drawback of memory based is they cannot be integrated with another information such as user information of item information representation. Example of product recommendation presentation can bee seen on Fig. 1 below.

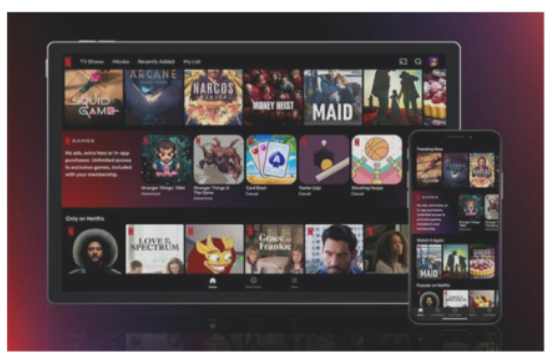

Fig. 1 E-commerce for movie online selling

The is a unique competition that belong to Netflix corporation. They will give the winner with 6 million dollar for who have succesful to achive morethan 10% over existing Nelflix engine performance. Majority academician, expert and researcher consider to applying another collaborative filtering model based on latent factor of famaous called matrix factorization. Eventhough, latent factor have been develop firtsly on early 2000 by Sarwar [4]. The latent factor model using singular value decomposition (SVD) that employ low rank dimentional. The advantage of matrix factorization model is they can be integrated with another information include side information of the product or user. The applying of latent factor of matrix factorization become popular in the Netflix competition aims to reach effectiveness level in producing rating prediction.

The essential problem have been rise when applying matrix factorization in sparse rating datasets. The sparse level rise when number of rating very minimum in the level below 4% or in extreme sparse level in below 1%. The effectiveness of rating prediction using latent factor degrade significantly when applying in sparse rating. Figure 1 below is ilustration unrating product cause sparse data. It have impact in product information for customer inacurate. It become major problem in collaborative filtering issue. Majority researcher involve product side information spesifically product document information such as product description, product information, and product testimony [5][6][7].

Fig. 2 Unrating product cause sparse data, it have impact inaccurate product recommendation

The adoption of product document to increase effectiveness of latent factor model have been popular in recent decade [8]. Majority of them consider to reach contextual understanding of the document. They are several strategy to capture document understanding such as [9]. This model implement TF-IDF and statistical model to calculate and interpreted document representation that famaous called bag of word (BOW). This algoritm model sucessful in increasing latent factor model based on probabilistic matrix factorization (PMF). PMF is enhancement of SVD model that consider probabilistic mechanism based on gausian normal distribution [10].

The adoption of traditional natural language processing in latent factor (matrix factorization) proven to enhance of collaborative filtering. However, the traditional NLP faced the essential shortcoming in capturing understanding in contextual. The expert in NLP said where the contextual understanding of phrase can be reach by considering word order and subtle word, for example "the cat chase the mouse" is normal situation, while "the mouse chase the cat" is abnormal condition. According to bag of word mechanism, both of phrase are similar in meaning text [11].

The integrating product document and user information into PMF [12][13]. According to experiment report, their model that involves deep learning model reach better performance over traditional latent factor approach. The adoption of deep learning framework was proposed in last five years deep learning success to enhance performance in several computer science application such as image processing, voice recognizing, and natural language processing. The implementation of deep learning model in recommender system have been done by several researcher. Enhancement of matrix factorization based on PMF using auto encoder to calculate product document [14]. Similar model proposed to advance SVD model [15].

The advance of collaborative filtering model using CNN was proposed by Kim [16]. Their model that called ConvMF achive better performance than previous work based on auto encoder. The improvement of CNN on this model due to dimentional reduction mechanism to capture product document understanding. Different with previous work where they consider word order and subtle word to interpreting product document. Similar with Kim, another word order mechanism proposed by Hanafi [17] by sequential aspect mechanism where adoption of sequential mechanism achieve more useful over dimentional reduction based on CNN. Table 1 represent state-of-the-art of collaborative filtering algorithm by enhancement document of product understanding.

Table 1. Collaborative filtering algorithm by integrating product document and latent factor model

Ref.	Collaborative filtering algorithm model
[9]	**BOW**: Collaborative filtering model using PMF and TF-IDF to capture document understanding of product review document.
[12]	**LDA**: Collaborative filtering using PMF and integrated with statitical approach to interpreted product review document.
[13]	**CTR**: Collaborative filtering using PMF and integrated with topic modelling to interpreted product review document.
[14]	**CDL**: Collaborative filtering using PMF and integrated with auto encoder to intepreted document understanding of product review document.
[13]	**HCDR**: Collaborative filtering model using SVD and auto encoder to capture document understanding of product review document by considering dimentional reduction aspect of phrase.
[16]	**ConvMF**: Collaborative filtering model using PMF and CNN to capture document understanding of product review document by considering dimentional reduction aspect of phrase.
[17]	**LSTM-PMF**: Collaborative filtering model using PMF, word embedding and LSTM to capture document understanding of product review document by considering sequential aspect of phrase.
[18]	**SRCMF**: Collaborative filtering model using PMF and CNN aims interpreted social review document of the product.
[19]	**Att-ConvMF**: Collaborative filtering model using latent factor and optimizing product document understanding using CNN and attention mechanism.

Capturing document contextual understanding of product review become important aspect. In recent five year, there are several algorithm model that obtain some researcher such as word embedding based on word2vec, fastext, glove and so on. In recent year, there are several model using bidirectional word vector representation.

According to explanation on above, in early decade can be concluded that traditional NLP model become very famous model in integrating document representation into latent factor model such as BOW, LDA and CTR. However, in recent five years, some researcher involve deep learning model in enhancing document contextual understanding such as CDL, CNN, LSTM and Attention model. All of them employ latent factor model based on PMF aims to improve effectiveness of rating prediction. Different with previous work where Attention mechanism have no adoption in sequentil aspect in LSTM. Following to the literature on above, the main contribution on this study consist 2 contributon including:

1. Generating contextual understanding of product document using sequential to sequential aspect which integrated Attention and LSTM.

2. Combining between Attention-LSTM into traditional latent factor based on PMF to enhance effectiveness of rating prediction in large datasets.

II. MATERIAL AND METHOD

A. Probabilistic Matric Factorization (PMF)

PMF is very popular latent factor model to produce rating prediction. Many researcher consider to adoption PMF in many research and application for recommendation. PMF is advance version of SVD. The essensial work mechanism can be explained as follow, for instance M is represent of movie and N is represent of user. While integer is represent of rating value began from *1* to *k* where R_{ij} is represent of user i with movie j. While $U \in R^{D\times N}$, $V \in R^{D\times M}$. Then, U and V become represent of the user and movie latent factor. Aim to calculate rating prediction can be computed by $R_{ij} = U_i^T, V_j$. PMF is SVD version with Gausian normal distribution approach. The vector representation of a user and movie resulted from distribution with rating correspondent, where distribution mechanism can be computed with equation 1 as follows:

$$p(R_{ij}|U_i, V_j, \sigma^2) = N(R_{ij}|U_i^T V_j, \sigma^2) \tag{1}$$

A mechanism to transform latent vector of the item, the PMF model consider applying a zero-mean spherical gausion prior that can be calculated with equation 2 as follows:

$$p(V|\sigma_V^2) = \prod_{j=1}^{M} N\,(V_i|0, \sigma_V^2 I) \tag{2}$$

While, a mechanism to transform latent vector of the item, the PMF model consider applying a zero-mean spherical gausion prior that can be calculated with equation 3 as follows:

$$p(U|\sigma_U^2) = \prod_{i=1}^{N} N\,(U_i|0, \sigma_U^2 I) \tag{3}$$

B. Capturing document contextual with LSTM and Attention Mechanism

1) LSTM

RNNs, or recurrent neural networks, are a type of artificial neural network in which the data from the previous step is used as input for the data from the following step. The main problem with RNNs, on the other hand, is the occurrence of gradient vanishing and exploding problems during back propagation. For this reason, Hochreiter and Schmidhuber [20] developed the Long Short-Term Memory (LSTM) in 1997 as a solution to the problem. A modified version of recurrent neural networks, Long Short-Term Memory (LSTM) networks are able to learn information from earlier time steps and apply it to subsequent time steps. Cell states, as opposed to the conventional feedforward neural networks, are used to channel data as it flows through the LSTM network. LSTMs are capable of selectively remembering and forgetting information in this manner. As a result, it was its gating mechanism that allowed RNNs to overcome the "short-term memory" problem. LSTM units are composed of a cell, an input gate, an output gate, and a forget gate, which is the last gate in the chain. The cell retains data over a random period of time, and the three gates regulate the flow of information into and out of the cell on a continuous basis (Fig. 3).

LSTM is sub-class of neural network and categorical feed forward neural network. One of advantage of LSTM approach is considering to connection among past information secsion and current secsion process. According to context of NLP, this is essential aspect to capture contextual understanding of phrase in a document. The LSTM consist several hidden stage related to the input layer, output layer the hidden state and prior process. The advantage of LSTM to initiate sequential aspect due to some critical calculation stage in the hidden stage.

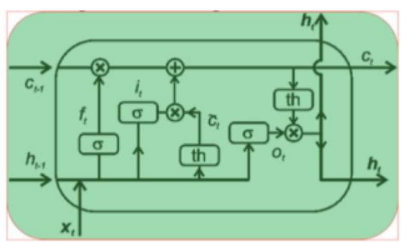

Fig 3. Basic mechanism work of LSTM

$$\begin{aligned}
i &= \sigma(x_i U^i + S_{t-1} W^i) \\
f &= \sigma(x_i U^f + S_{t-1} W^f) \\
o &= \sigma(x_t U^o + S_{t-1} W^o) \\
g &= tanh(x_i U^g + S_{t-1} W^g) \\
c_t &= c_t - Of + gOi \\
s_t &= tanh(c_t) - O_o gOi
\end{aligned} \tag{4}$$

i, f, o	It is the symbol of (*i*) input, (*f*) forget, and (*o*) output gate. All of them own similar equations, and they just have different parameter matrices. It is known as gate due to the sigmoid role that determines the value either *0* or *1*.
g	It is the symbol of the hidden state where it is computed based on the existing input and the past hidden state.
c_t	It is the symbol of the internal memory of the hidden state. It has become a combination of the previous memory c_{t-1} that is multiplied with a forget gate and the new hidden state g multiplied by the input gate.
Ct	It is the symbol of the memory of the hidden state. The computed output of the hidden state on s_t is multiplied using the output gate.

2) Attention Mechanism

One of the most important ideas in deep learning research in the last decade is the attention mechanism. Furthermore, it is a method that mimics the focus of the human mind. However, a wide range of artificial intelligence models, including natural language processing [21] and computer vision [22], now make use of this strategy. Seq2Seq models from the domain of Neural Machine Translation were used to create it initially. When using the seq2seq approach, the encoder analyzes the input data and compresses it into a context vector of a fixed length (sentence embedding), and the decoder uses the context vector in its computations to produce an output that has been transformed. Seq2Seq challenges have shown the enormous strength of this architecture, but it is hampered by a critical flaw. Sentence embedding is generated in a single vector that becomes more difficult for a machine to process as the length of the input data increases. As a result, it is unable to store longer input data because it tends to forget portions of it.

To help neural machine translation remember long source sentences, Bahdanau [23] introduced an attention mechanism. This mechanism creates shortcuts between a single context vector and the entire source input, rather than building a single context vector from the start. For each output feature, the weights of these shortcut connections can be adjusted (Fig. 3). Data that isn't critical is emphasized while the rest fades into the background.

There are multiple attention weights that are calculated for each of the inputs because not all of them will be used in generating the corresponding output, so the attention mechanism calculates multiple attention weights. The weighted sum of the annotations is used to create the context vector C_i for the output result y_i.

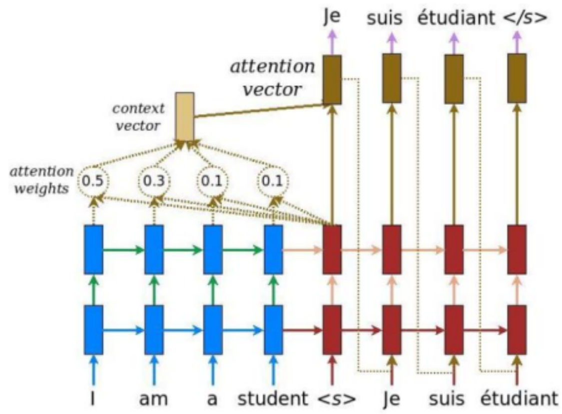

Fig 4. Basic work of Attention mechanism

Since not all the inputs would be used in generating the corresponding output, The attention mechanism calculates multiple attention weights marked by $\alpha(t,1)$, $\alpha(t,2)$, .., $\alpha(t, t)$. The context vector C_i for the output result y_i is produced applying the weighted sum of the annotations:

$$C_i = \sum_{j=1}^{T_x} \alpha_{ij} h_j \tag{7}$$

By normalizing the output of a feed-forward neural network described by the function that captures alignment between input at *j* and output at *i*, the attention weights are computed." In the following equation, a softmax function is used to compute the weights α_{ij}.

$$\alpha_{ij} = \frac{exp(e_{ij})}{\sum_{k=1}^{T_x} exp(e_{ik})} \tag{5}$$

$$e_{ij} = a(s_{i-1}, h_j) \tag{6}$$

e_{ij} is the output score of a feed forward neural network described by the function a that attempts to capture the alignment between input at *j* and output at *i*.

C. Hybridization Attention-LSTM and PMF

According to the attention-LSTM, using regression applications such as rating prediction in a collaborative filtering recommender system is not appropriate. The output of attention-LSTM in the form of a 2D vector representation cannot be used to predict the rating directly. To address the aforementioned issue, it must be combined with matrix factorization, such as PMF.

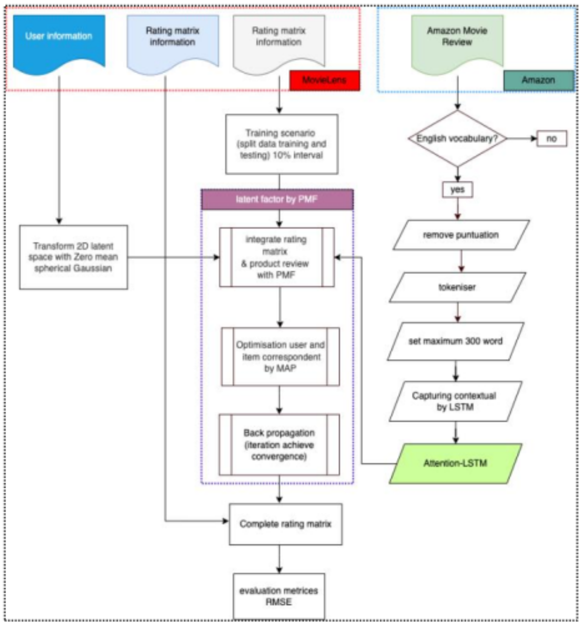

Fig. 5 Atention-LSTM and PMF framework

PMF is in charge of calculating the relationship between the user's latent model and the product's latent space, which strengthens user and item correlation. For instance, N as symbol of the user and M as symbol of the the item, the strategy to compute rating score is $R \in R^{NxM}$, while the formula to compute user representation is $U \in R^{KxN}$ and item representation can be compute by $V \in R^{KxN}$, then the table of product can be calculated by $U^T V$. According to probabilistic point of view, the normal distribution can be computed by:

$$p(R|U,V,\sigma^2) = \prod_{i=1}^{N}\prod_{j-1}^{M}[N(R_{ij}|U_i^T V_j,\sigma^2)]^{I_{ij}} \tag{8}$$

Where:

μ : mean of total population
σ^2 : variance score
I_{ij} : an indicator function as a generative model for user

1) User latent vector representation

MovieLens collects user information representations that only include user and rating information. The user latent model territory employs a zero mean spherical Gaussian prior by incorporating the user data variance value σ^2 with the formula is as follows:

$$p(U|\sigma_U^2) = \prod_{i-1}^{N} N(U_i|0,\sigma_U^2 I) \tag{9}$$

2) Item latent vector representation

Item information representation is obtained from AIV in the form of AIV item documents. A product document 2D vector 50 is obtained after a series of processes based on the LSTM mechanism. From a probabilistic standpoint, the item latent model is given by:

$$p(V|W,X,\sigma_W^2) = \prod_{j}^{M} N(v_j|attention\_lstm(W,X_j)\sigma_W^2 I) \tag{10}$$

While variable v_j as item representation that produced by attention and LSTM frame work. It can be obtained by:

$$v_j = attention\_lstm(W,X_j) + \varepsilon_j \tag{11}$$

3) Optimizing model and produce rating prediction mechanism

This is the last processing consist 3 essential step. When trying to calculate an unknown quantity, the MAP statistic comes in handy. It is similar to the posterior distribution in terms of optimise and learning. To be more specific, it seeks to optimize the learning variable while taking the MAP application into consideration. This method employed by a posteriori through the use of hyperparameters to analyze user and movie features. The complete of computing scenario can be seen on equation belwo:

$$\begin{aligned}&\max_{U,V,W} p(U,V,W|R,X,\sigma^2,\sigma_U^2,\sigma_V^2,\sigma_W^2)\\ &= \max_{U,V,W}[p(R|U,V,\sigma^2)p(U|\sigma_U^2)p(V|W,X,\sigma_V^2)p(W|\sigma_W^2)]\end{aligned} \tag{12}$$

The training scenario to learn correspondent between user and item aims to minimazing loss function, the detail formula can be seen in equation below.

$$\begin{aligned}\mathcal{L}(U,V,W) = &\sum_{i}^{N}\sum_{j}^{M}\frac{I_{ij}}{2}(r_{ij} - u_i^T v_j)_2 + \frac{\lambda U}{2}u_{i2}\\ &+ \frac{\lambda V}{2}\sum_{j}^{M} v_j - attetnion\_lstm(W,X_j)_2\\ &+ \frac{\lambda W}{2}\sum_{k}^{W_k} W_{k2}\end{aligned} \tag{12}$$

This step is critical because it involves optimizing W, which represents the weight variable and bias variable for each layer in the back-propagation algorithm, which is an important step in the process. The update mechanism is designed to optimize every layer, including V, U, and W, until convergence is required. It is optimized until convergence is required. The following is the formula that was used to predict the unknown rating:

$$\begin{aligned}r_{ij} \approx \mathbb{E}[r_{ij}|u_i^T v_j,\sigma^2] &= u_i^T v_j\\ &= u_i^T(attention\_lstm(W,X_j + \epsilon_j)\end{aligned} \tag{13}$$

D. Dataset

One of the most widely used datasets for e-commerce experiments is Movielens. In 1997, the University of Minnesota's School of Computing developed it. The MovieLens datasets [31, 11] were used in the majority of recommender system experiments. We were looking for information to help with our own personal recommendations.

These datasets have some categories that are dependent on how many ratings, how many users, how many products, and how dense the sparse ratings are. AIV's product review document was used in this experiment. This is a well-known Amazon dataset [32][33][36]. According to Table 3, here are some of the dataset's characteristics.

Table 2. Dataset characteristic

Dataset category	number of users	number of movie	number of rating	Sparse level
1M	6.040	3.544	993.482	4.64%
AIV	29.757	15.14	135.188	0.03%

This experiment utilizes two MovieLens categories: ML-1M, which contains one million ratings at a sparse level of 4.64 percent, and ML-10M, which contains ten million ratings at a sparse level of 1.41 percent. This is a critical factor in determining the performance of Attention-LSTM-PMF under certain sparsity level conditions.

E. Evaluation metrics

It is necessary to evaluate the performance of Attention-LSTM-PMF. The RMSE evaluation matrix is the most frequently used technique for determining the effectiveness of rating prediction [34][35]. The experiment's scenario is divided into nine parts, each of which divides the dataset by a 10% interval ratio, as follows: sparseness level 80% (20:80), sparseness level 60% (40:60), sparseness level 40% (60:40), sparseness level 20 (80:20).

The training process's output was evaluated using the root mean square error (RMSE) evaluation matrix. The evaluation matrices have the following formula:

$$RMSE = \sqrt{\frac{\sum_{i,j}^{N,M}(r_{ij} - \hat{r}_{ij})^2}{(total\ \#\ rating)}}$$

Finally, the Attention-LSTM-PMF prediction result is compared to the actual rating based on dataset resources from MovieLens. This is a mechanism to observe the effectiveness of attention-LSTM model to support latent factor based on PMF.

III. RESULT AND DISCUSSION

Sparse data in recommender system teritory is very important issue in long time. The growth of huge of data also increase the issue more popular. The combining between Attention-LSTM and PMF is become a solution to handle sparse data. This experiment attempt to observing the effectiveness of the algorithm. Figure 6 demonstrated the training result involved 80% sparseness level with ratio 20% data training and 80% data testing. The training scenario compared LSTM and PMF with Attention and without Attention. The impact of Attention show that success to increase effectiveness on rating prediction even adopted in extreme sparse level. Attention-LSTM and PMF represented in red colour and LSTM-PMF wthout PMF represented in blue colour.

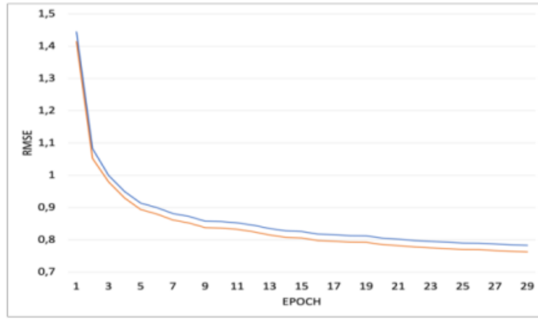

Fig 6. RMSE result on sparseness level 80%

The next section scenario attempt to decrease sparseness level into 40% (Fig. 7) 60% (Fig. 8) and finally on 80% (Fig. 8). In every section scenario, Attention-LTSM proven better perform over LSTM without Attention.

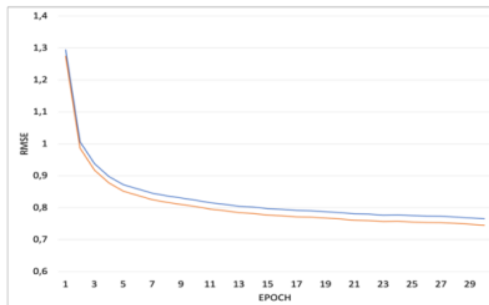

Fig 7. RMSE result on sparseness level 60%

Reffer to RMSE evaluation report, every training scenario reach below 0.80 where lower RMSE achievement is better. This result show that Attention-LSTM proven better effectiveness over previous work. The detail ecperiment result can be seen on Fig. 10.

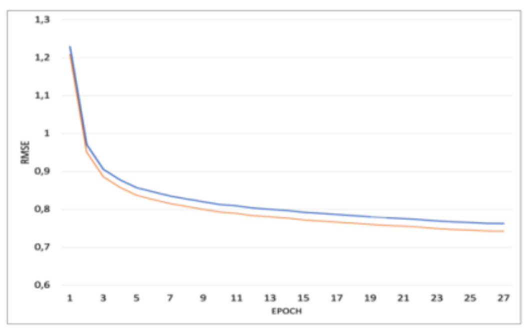

Fig 8. RMSE result on sparseness level 40%

Our best achievement demonstrated on Fig. 9 where the RMSE result reach amost 0.731 and Attention mechanism success to support LSTM work to increase effectiveness of PMF in generating rating prediction.

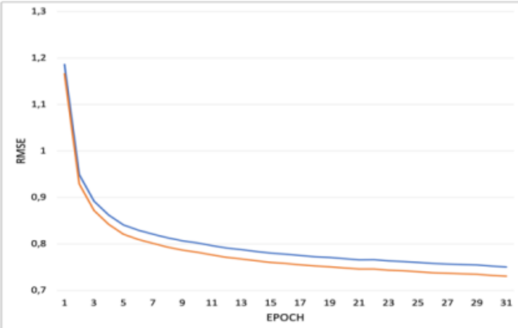

Fig 9. RMSE result on sparseness level 20%

Fig. 10 show the effectiveness of Attention mechanism over LSTM in every training scenario. Orange colour is represent 80% sparseness level, green represented 40% sparseness level, grey represented 60%, and the last is blue colour that represented 20% spparseness level .The key success factor of Attention is increasing share weigh of latent factor document of the product that represented in *W* where this variable play essential rule in inreasing the value of variable *V*. In every case specifically for natural language processing, seq2seq aspect success in increasing document understanding point of view.

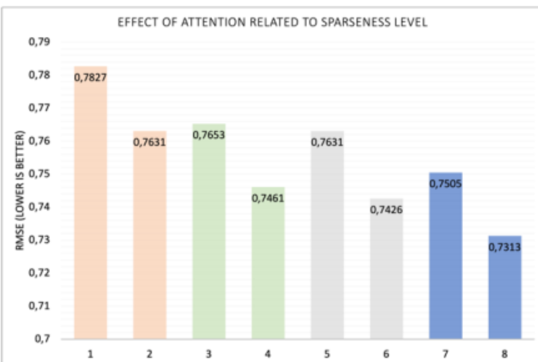

Fig 10. RMSE comparison result on various sparseness level

The effectiveness of every PMF with document understanding model presented on Fig. 11. The schematic experiment comparison in this research only applied 50% (50:50) sparseness level. The competitor in this experiment are included auto encoder and PMF (blue color), CNN and PMF (red colour), LSTM and PMF (green colour), and our model Attention-LSTM and PMF (yellow colour). Overall models are state-of-the-art in recommender system in lasr five year. Our model achieve significant performance over another deep learning model.

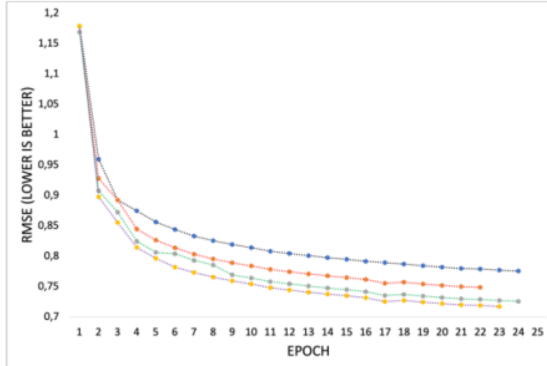

Fig 11. RMSE comparison result over AE, CNN, LSTM and Attention-LSTM

IV. CONCLUSION

Sparse data issues, which are caused by a minimum rating, continue to be a significant problem in the recommendation system. In this study, we proposed a latent factor model incorporating Attention, LSTM, and PMF while taking word sequential to calculate in word order to interpret document understanding and capture the contextual insight contained within the product review documents. According to the results of our experiment described above, our model outperformed previous work. Adoption Attention into LSTM-PMF performed well, according to the researchers, because of the impact of contextual insight representation of the document to support latent factors based on PMF in increasing the effectiveness level of generating ratings, which was believed to be the case. Furthermore, the involvement product documents developed with the help of Attention and LSTM improve the efficiency of the training process, allowing for greater convergence in an overall training scenario. In order to learn contextual insight interpretability, there are several methods available, such as bidirectional word encoder representation. Consideration of the bidirectional model to improve contextual understanding of the document may improve the performance of matrix factorization in predicting the rating matrix. Future research work will be made more difficult as a result of this. PMF is a matrix factorization method that is a subset of the matrix factorization method. This can be further enhanced by incorporating other matrix factorization methods, such as SVD, SVD++ and Non-Negative Matrix Factorization (NNMF), which only takes into rating information. With the help of some of the technique previously mentioned there is a possibility to improve the effectiveness of rating prediction in sparse data in large datasets.

ACKNOWLEDGMENT

We would to thank for Universitas Amikom Yogyakarta to support our research and help us to provide some hardware tools and application.

REFERENCES

[1] Hanafi, N. Suryana, and A. Sammad, "An Understanding and Approach Solution for Cold Start Problem Associated with

Recommender System : A Literature Review," *Journal of Theoretical and Applied Information Technology*, vol. 96, no. 09, pp. 2677–2695, 2018.
[2] Y. Koren, R. Bell, and C. Volinsky, "Matrix Factorization Techniques for Recommender Systems," *IEEE*, vol. 40, no. 8, pp. 42–49, 2009.
[3] Hanafi, N. Suryana, and A. Samad, "Deep Learning for Recommender System Based on Application Domain Classification Perspective : a Review," *Journal of Theoretical and Applied Information Technology*, vol. 96, no. 14, pp. 4513–4529, 2018.
[4] B. M. Sarwar, G. Karypis, J. a Konstan, and J. T. Riedl, "Application of Dimensionality Reduction in Recommender System - A Case Study," *Architecture*, vol. 1625, pp. 264–8, 2000.
[5] Hanafi, N. Suryana, and A. S. H. Basari, "Generate Contextual Insight of Product Review Using Deep LSTM and Word Embedding," *Journal of Physics: Conference Series*, vol. 1577, no. 1, 2020.
[6] Hanafi, N. Suryana, and A. S. B. H. Basari, "Convolutional-NN and word embedding for making an effective product recommendation based on enhanced contextual understanding of a product review," *International Journal on Advanced Science, Engineering and Information Technology*, vol. 9, no. 3, 2019.
[7] Hanafi, N. Suryana, and A. S. H. Basari, "Deep Contextual of Document Using Deep LSTM Meet Matrix Factorization to Handle Sparse Data: Proposed Model," *Journal of Physics: Conference Series*, vol. 1577, no. 1, 2020.
[8] Hanafi, N. Suryana, and A. S. B. H. Basari, "Convolutional-NN and word embedding for making an effective product recommendation based on enhanced contextual understanding of a product review," *International Journal on Advanced Science, Engineering and Information Technology*, vol. 9, no. 3, pp. 1063–1070, 2019.
[9] G. Ling, M. R. Lyu, and I. King, "Ratings meet reviews, a combined approach to recommend," *Proceedings of the 8th ACM Conference on Recommender systems - RecSys '14*, pp. 105–112, 2014.
[10] R. Salakhutdinov and A. Mnih, "Probabilistic Matrix Factorization.," *Proc. Advances in Neural Information Processing Systems 20 (NIPS 07)*, pp. 1257–1264, 2007.
[11] C. C. Aggarwal, *Machine Learning for Text*. NEWYORK: Springer International Publishing AG, 2018.
[12] J. Liu and D. Wang, "PHD : A Probabilistic Model of Hybrid Deep Collaborative Filtering for Recommender Systems," in *ACML*, 2017, pp. 224–239.
[13] C. Wang and D. M. Blei, "Collaborative Topic Modeling for Recommending Scientific Articles," *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '11*, p. 448, 2011.
[14] H. Wang, N. Wang, and D.-Y. Yeung, "Collaborative Deep Learning for Recommender Systems," in *KDD conference*, 2015, pp. 1235–1244.
[15] J. Wei, J. He, K. Chen, Y. Zhou, and Z. Tang, "Collaborative Filtering and Deep Learning Based Hybrid Recommendation for Cold Start Problem," *2016 IEEE 14th Intl Conf on Dependable, Autonomic and Secure Computing, 14th Intl Conf on Pervasive Intelligence and Computing, 2nd Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress(DASC/PiCom/DataCom/CyberSciTech)*, pp. 874–877, 2016.
[16] D. Kim, C. Park, J. Oh, S. Lee, and H. Yu, "Convolutional Matrix Factorization for Document Context-Aware Recommendation," in *Proceedings of the 10th ACM Conference on Recommender Systems - RecSys '16*, 2016, pp. 233–240.
[17] Hanafi and B. M. Aboobaider, "Word Sequential Using Deep LSTM and Matrix Factorization to Handle Rating Sparse Data for E-Commerce Recommender System," *Computational Intelligence and Neuroscience*, vol. 2021, no. 1, 2021.
[18] X. Wang, X. Yang, L. Guo, Y. Han, F. Liu, and B. Gao, "Exploiting Social Review-Enhanced Convolutional Matrix Factorization for Social Recommendation," *IEEE Access*, vol. 7, pp. 82826–82837, 2019.
[19] B. Zhang, H. Zhang, X. Sun, G. Feng, and C. He, "Integrating an attention mechanism and convolution collaborative filtering for document context-aware rating prediction," *IEEE Access*, vol. 7, pp. 3826–3835, 2019.
[20] S. Hochreiter and J. Urgen Schmidhuber, "Lstm," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
[21] Y. Kim, "Convolutional Neural Networks for Sentence Classification," in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2014, pp. 1746–1751.
[22] Hanafi, A. Pranolo, and Y. Mao, "Cae-covidx: Automatic covid-19 disease detection based on x-ray images using enhanced deep convolutional and autoencoder," *International Journal of Advances in Intelligent Informatics*, vol. 7, no. 1, pp. 49–62, 2021.
[23] J. Chorowski, D. Bahdanau, D. Serdyuk, K. Cho, and Y. Bengio, "Attention-based models for speech recognition," *Advances in Neural Information Processing Systems*, vol. 2015-Janua, pp. 577–585, 2015.

ICAITI_2022_submissions_135-169_2022-02-16_1644978762/ICAITI_2022_paper_145.pdf

ORIGINALITY REPORT

22%
SIMILARITY INDEX

PRIMARY SOURCES

1	www.ncbi.nlm.nih.gov Internet	438 words —	10%
2	journalofbigdata.springeropen.com Internet	190 words —	4%
3	www.hindawi.com Internet	83 words —	2%
4	www.joiv.org Internet	39 words —	1%
5	medium.com Internet	36 words —	1%
6	downloads.hindawi.com Internet	33 words —	1%
7	sci2s.ugr.es Internet	17 words —	< 1%
8	www.aclweb.org Internet	17 words —	< 1%
9	Udit Singhania, B. K. Tripathy. "chapter 7 Text-Based Image Retrieval Using Deep Learning", IGI Global, 2021	16 words —	< 1%

Crossref

10	www.ijain.org Internet	16 words — < 1%
11	coek.info Internet	10 words — < 1%
12	assets.researchsquare.com Internet	9 words — < 1%
13	ijain.org Internet	9 words — < 1%
14	mafiadoc.com Internet	9 words — < 1%
15	www.kdnuggets.com Internet	9 words — < 1%
16	Wei, Jian, Jianhua He, Kai Chen, Yi Zhou, and Zuoyin Tang. "Collaborative filtering and deep learning based recommendation system for cold start items", Expert Systems with Applications, 2017. Crossref	8 words — < 1%
17	aclweb.org Internet	8 words — < 1%
18	arxiv.org Internet	8 words — < 1%
19	mail.ijain.org Internet	8 words — < 1%
20	pdfs.semanticscholar.org Internet	8 words — < 1%

21	thesai.org Internet	8 words — < 1%
22	www.insightsociety.org Internet	8 words — < 1%
23	www.jatit.org Internet	8 words — < 1%
24	Hanafi, Burhanuddin Mohd Aboobaider. "Word Sequential Using Deep LSTM and Matrix Factorization to Handle Rating Sparse Data for E-Commerce Recommender System", Computational Intelligence and Neuroscience, 2021 Crossref	7 words — < 1%
25	insightsociety.org Internet	4 words — < 1%