Received: June 7, 2022. Revised: July 4, 2022.


*International Journal of*
Intelligent Engineering & Systems

http://www.inass.org/

# Enhance Rating Prediction for E-commerce Recommender System Using Hybridization of SDAE, Attention Mechanism and Probabilistic Matrix Factorization

**Hanafi[1]***

[1] *Department of Computer Science, Universitas Amikom Yogyakarta, Jl. Ringroad Utara, Condongcatur, Depok, Sleman, Yogyakarta, Indonesia*
* Corresponding author's Email: hanafi@amikom.ac.id

**Abstract:** E-commerce is essential application in world wide. In everyday live, we cannot escape from e-commerce transaction. E-commerce requires intelligent machine to deliver product information to customer. Intelligent machine popular called recommender system developed by matrix factorization. Rating is representation of customer expression for satisfied product or service. Unfortunately, number of ratings is too sparse due to majority customer lazy to give rating for e-commerce product. Number of sparse rating matrix have impact in matrix factorization in rating prediction. Moreover, extreme sparse rating has impact degrade performance significantly. Many researchers consider to enhance matrix factorization using customer and product information such as customer demographics information, customer testimony, product review, product description, and etc. In this research, we consider to incorporating Stack Denoising Auto Encoder (SDAE), attention mechanism aims to enhance product review document understanding representation and matrix factorization based on Probabilistic Matric Factorization (PMF) to produce rating prediction. According to experiment report, our model superior over previous work based on hybrid model between PMF, Convolutional Neural Network (CNN) and SDAE that popular called PHD in 1%, and superior over hybrid model between PMF, Long Short Term Memory (LSTM) and SDAE that popular called DDL-PMF in 0.9%, and achieved significantly in 8% over traditional PMF based on RMSE evaluation metrics.

**Keywords:** E-commerce, Recommender system, Deep learning, Matrix factorization, Sparse rating, Collaborative filtering, Attention mechanism.

## 1. Introduction

In two last decade, e-commerce become essential application in world wide. E-commerce famous to recognised with online shopping transaction. In everyday living, we cannot escape for online transaction based on e-commerce for what food to eat, what news to read, what movie to watch, what car to travel, what subject to learn, who people to friend, and etc [1].

Similar with traditional shopping transaction, customer require product information to choose suitable product. E-commerce or online shopping transaction require an engine to provide information for customer of customer candidates. The engine or machine responsible to producing automate information because impossible to produce product fit information manually. E-commerce automatic information engine popular called recommender system or recommendation system [2-4].

Majority researcher, expert, and academicians believed that successful of e-commerce recommender system influence marketing target. In the other hand, successful of recommender system implementation have impact in e-commerce business corporation revenue. Finally, Adoption of recommender system become important aspect to increase revenue growth of e-commerce company business.

Some research show that recommender system proved in increasing customer deal in many e-commerce company. For example, more than 40%

Received: June 7, 2022. Revised: July 4, 2022. 

Netflix customer deal with movie recommendation, 60% YouTube users watch video by recommendation, 20% of iTunes revenue obtained by recommendation. This is important reason of recommender system adoption in many large e-commerce company [5, 2, 6].

There are 4 basic recommender system algorithms included 1. Content based; product recommendation provided based on product behaviour, 2. Demographic based; product recommendation provided based on customer demographic information, 3. Knowledge based, product recommendation serve using knowledge based recommendation, 4. Collaborative filtering, product recommendation provide using customer behaviour in the past [6, 7].

According to several literature review, collaborative filtering become favourite algorithm that adopted in many e-commerce company. Collaborative filtering proved better performance in accuracy over content based. In early collaborative filtering algorithm, adoption of nearest neighbour model become majority approach to develop recommender system such as spearman rank, cosine similarity, adjust cosine similarity. The traditional collaborative filtering also popular called memory based collaborative filtering. However, memory based model face some shortcoming in high computation, require recalculation when any additional user and product information, incompatible to integrated with external information such as product and user information [8, 9].

According to drawback of memory based collaborative filtering mentioned above, in early 2006 rise a mechanism that popular matrix factorization and also popular latent factor model. The matrix factorization aims to handle disadvantage of memory based and increase effectiveness of rating prediction in sparse rating matrix. Matrix factorization has advantage that possible to integrating for side information [10]. Matrix factorization first introduced by Sarwar in early 2000 for rating prediction that popularized with Singular Value Decomposition (SVD) [11]. SVD applied low rank dimensional reduction based on matrix factorization to find correspondents between customer and product relationships. Rating become essential value to find intersection between them. Moreover, index of product for every consumer can be produced specifically following to user personalized. Example of sparse rating matrix show on Fig. 1.

A proposed model using Probabilistic Matrix Factorization (PMF) and topic modelling aim to increasing effectiveness of rating prediction. Topic modelling is a model to transform of product document using statistical approach to enhance document understanding by Wang [12]. Topic modelling responsible to transform document into word vector that integrated with PMF, while user information transforms into 2D latent space using probabilistic mechanism approach. Aims to enhance probabilistic approach, gaussian normal distribution applied in transforming user information. According to their experiment report, hybridization PMF and topic modelling outperform over traditional latent factor that only consider PMF based on RMSE and MAE evaluation metrics. PMF is enhancement version of SVD [13].

| | | | | |
|---|---|---|---|---|
| Hanafi | 3 | ? | 5 | ? |
| Mel | ? | 4 | ? | ? |
| Bert | 2 | ? | ? | 5 |
| Yana | ? | ? | 3 | ? |
| Siti | 2 | 4 | ? | 3 |

Figure. 1 Minimum rating matrix

A proposed model using PMF and AE aim to increasing effectiveness of rating prediction by Wang [14]. This model is enhancement of topic modelling where the document representation obtained by Auto Encoder (AE). According to deep learning point of view, AE is variant of deep learning algorithm that involve feature extraction mechanism. While, PMF responsible to bridging between user information and item document information. While, user information transformed into 2D latent space based on Gaussian normal distribution. An enhancement of document understanding using AE succeed to increase performance of topic modelling based on RMSE (root mean squared error) evaluation metrics.

A proposed model using PMF and CNN aim to increasing effectiveness of rating prediction by Kim [15]. CNN tried to advance product document understanding. In this research, Kim also applied word embedding application using GloVe (Global Vector for Word Vector Representation) before process and integrate into CNN machine. While, CNN responsible to transform product document into 2D latent space before compute and hybrid PMF and user information representation that calculate into 2D latent space using Gaussian normal distribution. The enhancement of PMF for rating prediction using GloVe and CNN success to increase effectiveness of rating prediction. Moreover, CNN outperform over

Received: June 7, 2022. Revised: July 4, 2022. 

hybridization between PMF and AE. The PMF and CNN model implemented in ML-1M, ML-10M, and Amazon datasets.

Similar with previous work, Hanafi proposed a hybrid model using PMF and deep learning approach based on LSTM [16-18]. The objective of this study to advance product document understanding using LSTM. In this model also applied GloVe framework to transform product document latent space. The involvement of LSTM aims to calculate of product document with sequential technique. Different with CNN model that consider dimensional reduction technique using convolutional process. LSTM has benefit to capture product document understanding over CNN model due to word order consideration. According to experiment report, hybridization between LSTM and PMF superior over previous work based on topic modelling, AE, CNN with significant performance. Similar with previous work, the PMF and LSTM also applied Gaussian normal distribution to transform user information latent space.

A proposed model based on PMF and dual deep learning model using SDAE and CNN [19]. It was contrast with some previous work, an enhancement user information using SDAE was applied. SDAE responsible to change Gaussian normal distribution for transforming user information latent space. While, PMF and LSTM have similar task with previous work. The hybridization technique is this research popular called PHD (Probabilistic Hybrid Deep Learning). PHD also applied in similar datasets following to previous work using ML-1M, ML-10M, and Amazon datasets. According to experiment report, PHD model success to increase effectiveness of previous achievement based on PMF and CNN. In the other hand, SDAE to transform user information plays important role to enhance effectiveness of rating prediction. Table 1 show some state of the art in collaborative filtering recommender system

A study involves PMF and dual deep learning based on SDAE and LSTM aim to enhance effectiveness of sparse rating matrix by Hanafi [20]. This hybrid model named Dual Deep Learning and PMF (DDL-PMF). Almost similar with previous work, they tried to enhancement PHD model using GloVe and LSTM and implemented the model in ML-1M, ML-10M and Amazon datasets. According to experiment report, DDL-PMF success to increase performance over PHD model as best previous work. This model was evaluated by root mean squared error (RMSE) and Mean Absolut Error (MAE). Author believes that role of SDAE and LSTM play essential aspect in increasing of effectiveness level in rating prediction task.

A new model proposed in recent year using PMF and Attention mechanism [21, 22]. Attention is essential finding in recent year. It has ability in optimizing deep learning task in computer science field such as text processing, voice recognizing, image processing. In this study, attention approach used to advance document context of product document representation. In several network security research, involvement of deep learning achieved effectiveness in attack detection [23]. The adoption of attention mechanism in the study success to increase product document understanding. It was believed the performance of attention mechanism and PMF outperform over best previous work using LSTM and PMF based on RMSE evaluation metrics.

Table 1. State of the art of collaborative filtering recommender system

| Ref. | Method | Model based / latent factor | | Explicit feedback | Auxiliary information | | Deep learning | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | SVD | PMF | Rating | Item side document | User side information | AE | CNN | LSTM | Attention |
| [13] | PMF | - | √ | √ | - | - | - | - | - | - |
| [24] | CTR | - | √ | √ | √ | - | - | - | - | |
| [14] | CDL | - | √ | √ | √ | - | √ | - | - | - |
| [21] | ATT-PMF | - | √ | √ | √ | - | - | - | √ | |
| [25] | HCDR | √ | - | √ | √ | - | √ | - | - | - |
| [26] | CNN-PMF | - | √ | √ | √ | - | - | √ | - | |
| [16]–[18] | LSTM-PMF | - | √ | √ | √ | - | - | - | √ | - |
| [19] | PHD-PMF | - | √ | √ | √ | √ | √ | √ | - | |
| [20] | DDL-PMF | - | - | - | √ | √ | √ | - | √ | - |
| SDAE, Attention, and PMF | | | √ | √ | √ | √ | √ | - | - | √ |

Received: June 7, 2022. Revised: July 4, 2022. 

Collaborative filtering still faces in several challenge in the future research. The enhancement of side information also popular named auxiliary information required improvement in the term information resources and algorithm enhancement model. Social media information and internet of thing can be explored. According to description on above, the contribution of this study can be concluded as follow:

- Novel model in hybridization between SDAE to enhance user information representation, Attention mechanism to enhance contextual product document understanding and PMF responsible to generate rating prediction for sparse rating matrix.
- Enhancement of product document understanding using attention mechanism and word embedding using GloVe.
- Enhance user demographic information representation using feature extraction model based on SDAE.

## 2. Methodology

The proposed framework that involves SDAE, PMF, Attention, and some pre-processing approach. Deep learning is the novel achievement in machine learning algorithm in 2 last decade. It has success to reach improvement in several computer science research and industry such as image processing field [27], text processing for recommender system [16, 17, 28-31]. The detail explanation of experiment scenario explained in this section as follow:

### 2.1 PMF

PMF is representation of collaborative filtering recommender system based on model based. It also popular called model based collaborative filtering. PMF is enhancement of SVD model that responsible to produce rating prediction. The basic work of PMF is suppose we have *M* represent of item and variable *N* as user representation. While, rating value representation start from *1* to *k*. Then, the representation of user *i* for movie *j* can be obtained by $R_{ij}$. *U* and *V* represents of users and movies latent factor that produced by $U \in R^{D \times N}$ and $V \in R^{D \times M}$. The processing to observed rating value as follow:

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

Aim to transform 2D latent vector consider to apply Zero mean spherical Gaussian Prior of user as;

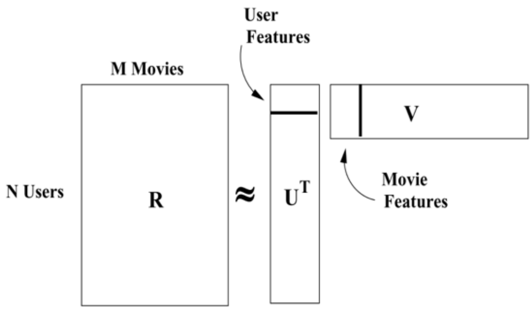


Figure. 2 PMF task to produce rating prediction

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

While, transforming item representation 2D latent space can be obtain by;

$$p(V|\sigma_V^2) = \prod_{j=1}^{N} N\left(V_j\middle|0, \sigma_V^2 I\right) \quad (3)$$

For the further process, PMF responsible to produce rating prediction will integrated with user and item information with SDAE dan attention mechanism aim to increase effectiveness to generate rating prediction.

### 2.2 SDAE

SDAE was introduced by Vincent to enhance feature extraction of auto encoder algorithm [32]. It used multiple combination of stacked layer to generate weight representation in hidden layer. In this research adopted PHD model by Liu [19]. They combined two essential parts of users side information. Detail framework of SDAE model can be seen on Fig. 3 below.

SDAE work by employ matrix *U* for generating latent factor model. While, process to produce user latent factor can be calculate with equation as follows:

$$h_l = g\left(C_l h_l + Q_l \tilde{X} + b_l\right) \quad (4)$$

Variable $Q_l$ and $C_l$ represent of weight parameter to each session. While $b_1$ responsible to solving the bias vector for each network. $R_i$ represent the corrupt value of *h0*, variable *g*() responsible to calculate non linier activation function. Symbol *X* represent the corrupt value of $\widetilde{X}$. While, the output value *L* can be computed with equation as follow:

$$\hat{R}_i = f\left(C_L h_L + b_{\hat{R}_i}\right) \; and \; \hat{X} = f\left(Q_L h_L + b_{\hat{X}}\right) \quad (6)$$

In the second layer of activation function represent as *f*(). While, $\frac{L}{2}$ as represent 2 layers of

Received: June 7, 2022. Revised: July 4, 2022. 

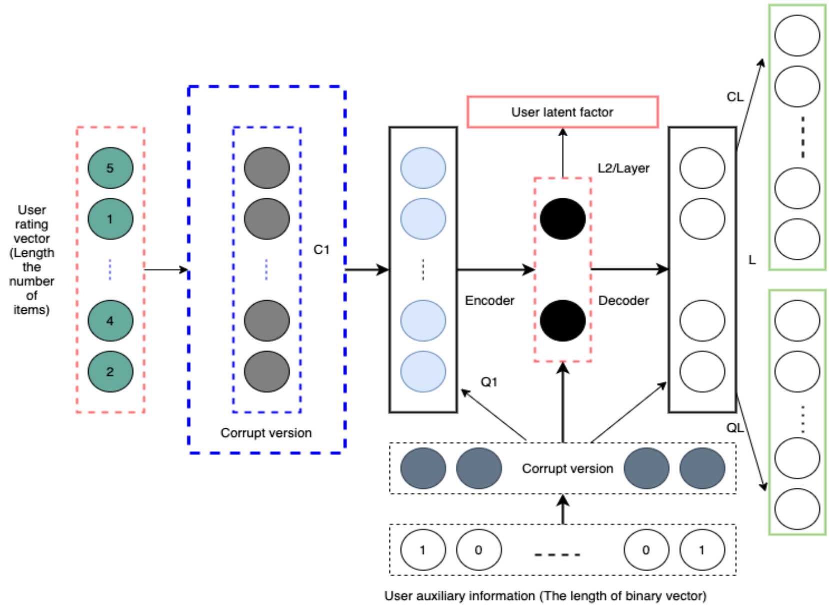

Figure. 3 Architecture of SDAE

SDAE where first of $\frac{L}{2}$ responsible for encoder task and second of $\frac{L}{2}$ responsible for decoder task.

### 2.3 Attention

Attention is essential finding is recent year. This method plays important role in enhancement of deep learning work performance. Attentions mechanism adopt of human mind attention in mimics. In older approach, application in machine learning model including computer vision, language processing involves traditional neural network. However, a new strategy implement sequence to sequence (seq2seq) model that was used in majority translation machine. According to basic seq2seq mechanism, the encoder responsible to calculate the incoming data and compresses the data to context vector in the form of fix length (common process named sentence embedding) [33]. The second process is decoding step where decoder applied the context vector computation to generate an output that was converted. The shortcoming of seq2seq model solved with tremendous achievement with this architecture. Sentence embedding was produced by one vector that very difficult for an algorithm to compute a length of input data.

Attention mechanism suitable neural machine improvement with remembering technique for long sentence category. Attention mechanism first introduces in recent year [33]. Attention makes a data input, develop one context vector in the beginning process. In every output data, they adjusted the weight of shortcut relation. The process put data that it is not important into the background.

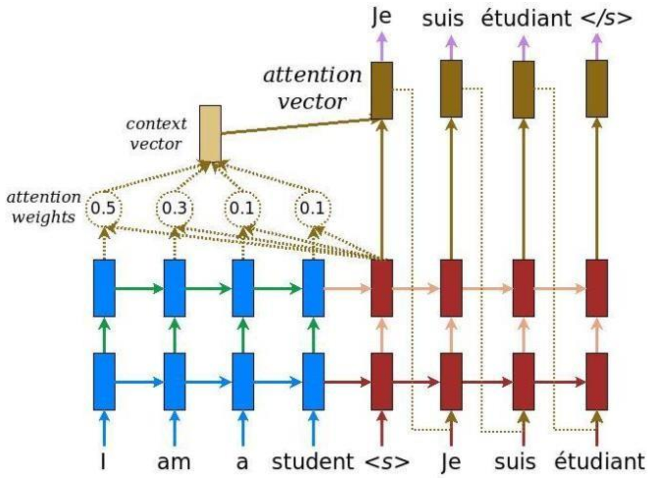

Figure. 4 Basic work of attention mechanism [33]

As a result, there are several attention values that was computed for every input. Not for all input are computed to produce the output correspondent. According to the reason, attention machine computes overall attention weight. The weigh sum is developed with objective to employ the context vector ($C_i$) to produce the output ($y_i$). The graphic illustration of Attention mechanism work can be seen on Figure 4 below.

While not every input would be applied in producing the output correspondent, the attention algorithm compute every weight signed by $\alpha(t,1)$, $\alpha(t,2)$, .., $\alpha(t, t)$. Aims to generate the weight sum using the context vector with $C_1$ and the output result $y_i$ with the equation as follow:

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

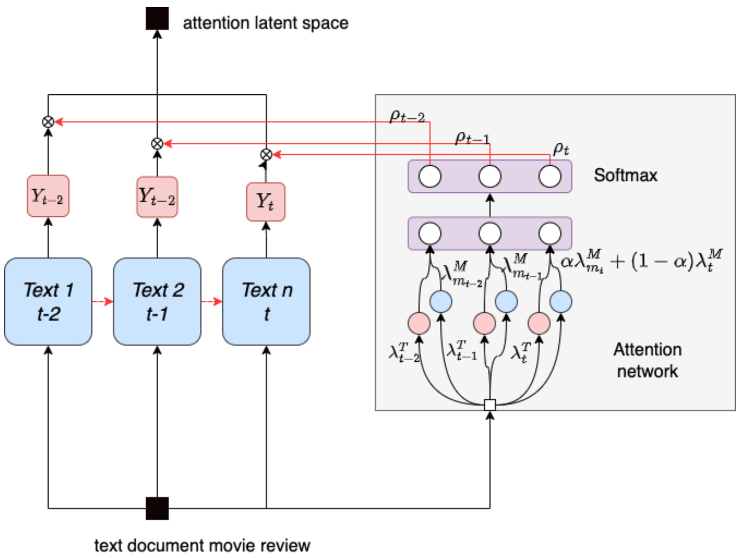


Figure. 5 Attention mechanism to transform product document latent space

Table 2. Text pre-processing step

| No | Text pre-processing |
|---|---|
| 1 | Setting up maximum length of sentences up to 300 words, this step aims to capture of the contextual meaning of the document. The limiting role taken according to previous study decided with 300 maximum words. |
| 2 | Delete stop words in the sentences, this step aims delete unused characters and symbol which it does not influent the meaning. |
| 3 | Compute TF-IDF value for every document, the goal of deleting the score of words that too much dominant in every document review on pre-processing datasets. |
| 4 | Delete overall non-vocabulary words from catalogue sentences, this step is not influent the meaning context and it can reduce the computation cost. |

Aim to catch correspondent between an input j and an output *i* using normalizing feed forward neural network, attention weight was calculated by using SoftMax function to enumerate the weight $\alpha_{ij}$ with the equation as follow:

$$\alpha_{ij} = \frac{\exp(e_{ij})}{\sum_{k=1}^{Tx} \exp(e_{ik})} \quad (8)$$

$$e_{ij} = a\left(h_j, S_{i-1}\right) \quad (9)$$

$e_{ij}$ represent the value of the output from feed forward of neural network that acquired from the function aims to capture correspondent among input of *j* and output of *i*.

The product document in the term of text information. It requires some process to transform from text into 2d dimensional data representation. Table 3 show some pre-processing stage of product document. This research applied NLTK (Natural Language Tool Kits) library aims to handle text processing step.

According to matrix factorization technique, it needs text dimensional latent space in the term of product representation (*W*) to support PMF task. The detail of attention mechanism to compute text document using sequential to sequential (seq2seq) aspect can be seen on Fig. 5 below. Every product review set maximum in 300 words. Every text document that was transform into vector space would input into attention mechanism input.

### 2.4 Integrating SDAE, attention and PMF

The basic concept of proposed model consists of 3 essential algorithms including SDAE, Attention mechanism and PMF. Adoption of SDAE and attention aims to handling user and item side information, while PMF responsible to bridging SDAE and attention in generating rating prediction. The basic work explanation task on above can be calculated with equation:

$$p = (R|U, V, \sigma^2) = \prod_i^N \prod_j^M N(R_{ij}|u_i^T v_j, \sigma^2)^{I_{ij}} \quad (10)$$

In this proposed model, it considers to apply normal distribution $N(x|\mu, \sigma^2)$ as represent pdf (probability density function). SDAE is user representation point of view which is contain 3 important factors. *Xi* as user side information representation for user *i*, *W+* as internal weight value representation, and varepsilon of Gaussian noise. User latent factor representation can be computed as follow:

Received: June 7, 2022. Revised: July 4, 2022. 

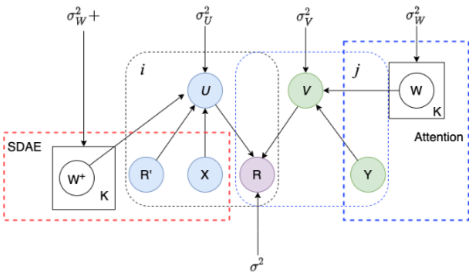


Figure. 6 Graphic illustration of hybridization strategy and data transformation

$$u_i = sdae(W^+, X_i) + \varepsilon_i \quad (11)$$

We implemented zero mean spherical Gaussian prior for graphic representation of Gaussian normal distribution.

$$p\left(w^+ \middle| \sigma_{W^+}^2\right) = \prod_k N\left(w_k^+ \middle| 0, \sigma_{W^+}^2\right) \quad (12)$$

The user latent factor can be calculated with the equation on below, following to probabilistic density function mechanism.

$$p\left(U \middle| W^+, X, \sigma_U^2\right) = \prod_i^N N\left(u_i \middle| sdae(W^+, X_i), \sigma_U^2\right) \quad (13)$$

This model considers to applied some probabilistic mechanism of data dimensional transformation. Similar with user's latent factor, item's latent factor can be developed with following the equation in below.

$$p\left(V \middle| W, Y, \sigma_Y^2\right) = \prod_j^M N\left(v_j \middle| attention(W, Y_j), \sigma_V^2 I\right) \quad (14)$$

Where, variable $v_j$ can be calculated with equation as follow:

$$v_j = attention(W, Y_j) + \epsilon_j \quad (15)$$

The detail explanation of data distribution and integration between SDAE, Attention and PMF according to probabilistic point of view can be explained on Fig. 6.

Table 3. Notation of equation

| notation | description |
|---|---|
| $U$ | user latent factor representation |
| $V$ | item latent factor representation |
| $\sigma^2$ | variance value (in this research can be form of the user or item representation) |
| $\varepsilon_i$ | epsilon variable of the item |
| $\sigma_U^2$ | variance value of user information representation in the term of demographic information or tags # information |
| $\sigma_V^2$ | variance value of item representation in the term of product document |
| $W^+$ | internal weight from product document representation |
| $\sigma_V^2$ | variance value of the item |
| $R_{ij}$ | actual value of rating from the user *i* and item *j* |
| $M$ | matrix of item representation from MovieLens datasets |
| $N$ | matrix of users representation from MovieLens datasets |
| $I_i$ | diagonal matrix |
| $I_{ij}$ | indicator function of the matrix |
| $\mu$ | mean value |
| $v_j$ | product representation of item *j* |
| $\sigma$ | standard deviation |
| $R$ | symbol of actual rating matrix |
| $R'$ | result of rating matrix prediction from PMF |
| $X$ | user auxiliary information in the term of demographic information of the user |
| $Y$ | item auxiliary information in the term of product document information |

The development of model considers to adopt mathematic and statistic formula. Table 3 show the complete notation in this study.

### 2.5 Recommender system datasets

In this research consider to applied in real datasets and most popular research in e-commerce recommender system. The datasets consist of two categories, MovieLens ML.1M and ML.10M as rating representation and Amazon as product review document representation [34]. ML.1M contain 1

Table 4. Datasets characteristic

| Dataset category | User side information | Item side information | Number of users | Number of items | Number of ratings | Sparse level (%) |
|---|---|---|---|---|---|---|
| **ML-1M** | Gender / age / occupation / zip code | Movie descriptions | 6,040 | 3,706 | 1,000,209 | 4.47% |
| **ML-10M** | Tags application | Movie descriptions | 69,878 | 10,677 | 10,000,054 | 1.34% |
| **Amazon** | Demographic characteristics | Movie review | 81,339 | 18,203 | 238,352 | 0.016% |

Received: June 7, 2022. Revised: July 4, 2022. 

millions of ratings and 10 millions number of rating. The complete characteristic and feature of the datasets can be seen on Table 4 below. The objective of datasets implementation in this model is to observe the ability in handling sparse rating matrix problem.

### 2.6 Evaluation metrics used RMSE

Aims to observe the effectiveness of propose model, in this research consider to evaluate using RMSE evaluation metrics. This research considers to split the dataset into data training and data testing with 10% interval sparseness level of rating, where 10% data training and 90% data testing are categorised into highest sparseness level and vice versa. The formula of RMSE to evaluated of rating prediction result can be calculated with equation as follows:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i,j} Z_{i,j}^{P}(R_{ij} - \hat{R}_{ij})^2} \qquad (16)$$

Where, N variable represents the total number of ratings, while $Z_{i,j}^{P}$ represents the test rating. It means the actual value of rating from datasets compared with rating value from prediction result.

## 3. Result and discussion

Sparsity is serious problem in many computer science territory. In e-commerce recommender system, sparsity problem due to minimum number of ratings also become major issue. This study considers to enhance sparse problem with enhance user and item information with SDAE, Attention, and PMF. The experiment aims to observed the effectiveness of proposed model that was described on below.

First experiment session divided ML.1M into 10% data training and 90% data testing. It categorical highly sparse level due to only 10% data for learning mechanism, while 90% data for testing mechanism. In the second process conduct similar schema with 10% interval of additional data training and reducing data testing with 10% interval also until last experiment scenario in the eight sessions.

This experiment scenario aims to observe the effectiveness of the model with attention mechanism to capture product document context from review, where document context with $W$ expect to increase share weigh of product document representation. Finally, according to experiment report on Table 4, the propose model outperform over previous work that involve CNN [15] and LSTM model in capturing of document context [18]. While, according to user information representation in PHD [19] and DDL-PMF [20], they used similar algorithm based on SDAE. In the other hand, it can be concluded that role of Attention mechanism very important to increase share weight ($W$) for product document representation. ML.1M MovieLens is categorised sparse datasets where the number of ratings only 1.41%. However, the performance of SDAE, Attention and PMF achieved better performance over best previous work using PMF [13] and SDAE-LSTM-PMF [20].

The second experiment applied the model in ML.10M datasets which is categorical huge datasets. The result of experiment and comparison show that role of SDAE and attention to support PMF in rating prediction is very essential. SDAE responsible to extract user information representation in the term of tags application to enhance of internal user share weight $W^{+}$ representation, and item share weight

Table 5. RMSE comparison result on ML.1M

| Ratio (%) | Hybrid Collaborative Filtering Model | | | |
|---|---|---|---|---|
| | PMF [13] | PHD-PMF [19] | DDL-PMF [20] | SDAE-ATT-PMF |
| 10% | 1.64697 | 0.98684 | 0.96298 | **0.94787** |
| 20% | 1.26577 | 0.94889 | 0.93392 | **0.91881** |
| 30% | 1.11180 | 0.93053 | 0.90986 | **0.89475** |
| 40% | 1.03992 | 0.91326 | 0.89842 | **0.88331** |
| 50% | 0.99064 | 0.89819 | 0.89371 | **0.87859** |
| 60% | 0.95897 | 0.88936 | 0.88095 | **0.86584** |
| 70% | 0.93369 | 0.88146 | 0.87272 | **0.85761** |
| 80% | 0.91134 | 0.87237 | 0.86605 | **0.85139** |
| 90% | 0.90452 | 0.86919 | 0.85837 | **0.84315** |
| $\sum$ | 9.76335 | 8.19009 | 8.07698 | **7.94132** |
| $\bar{X}$ | 1.08481 | 0.91001 | 0.89744 | **0.88236** |
| Improvement 8% over PMF, 1% over PHD-MF, 0.9 over DDL-PMF, whole in average. | | | | |

Table 6. RMSE comparison result on ML.10M

| Ratio (%) | Hybrid Collaborative Filtering Model | | | |
|---|---|---|---|---|
| | PMF [13] | PHD-PMF [19] | DDL-PMF [20] | SDAE-LSTM-PMF |
| 10% | 1.27539 | 1.17821 | **1.32981** | 1.37942 |
| 20% | 1.05233 | 0.83530 | 0.90216 | **0.90068** |
| 30% | 0.96513 | 0.81901 | 0.80812 | **0.798412** |
| 40% | 0.91827 | 0.80651 | 0.79945 | **0.785681** |
| 50% | 0.88834 | 0.79962 | 0.78819 | **0.779172** |
| 60% | 0.86673 | 0.79220 | 0.78134 | **0.770198** |
| 70% | 0.85071 | 0.78252 | 0.77681 | **0.764091** |
| 80% | 0.84055 | 0.77991 | 0.76145 | **0.755189** |
| 90% | 0.82796 | 0.76186 | 0.75998 | **0.749038** |
| ∑ | 8.48541 | 7.55283 | 7.70731 | **7.681881** |
| $\bar{X}$ | 0.94282 | 0.83920 | 0.85636 | **0.853542** |
| Improvement 8% over PMF, 1% over DDL-PMF, whole in average | | | | |

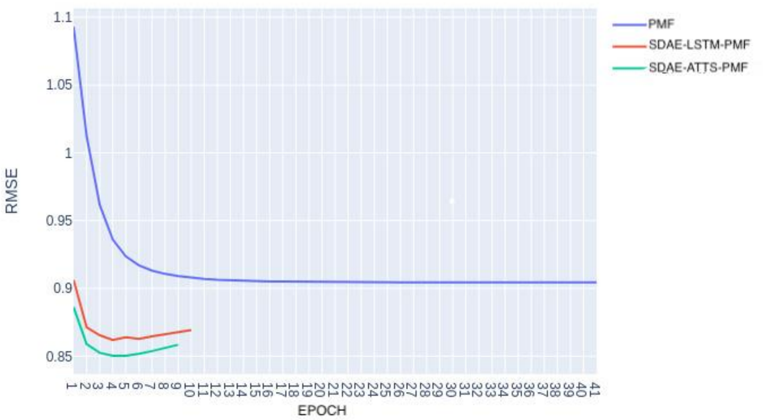


Figure. 7 Graphic RMSE evaluation result and impact of attention mechanism

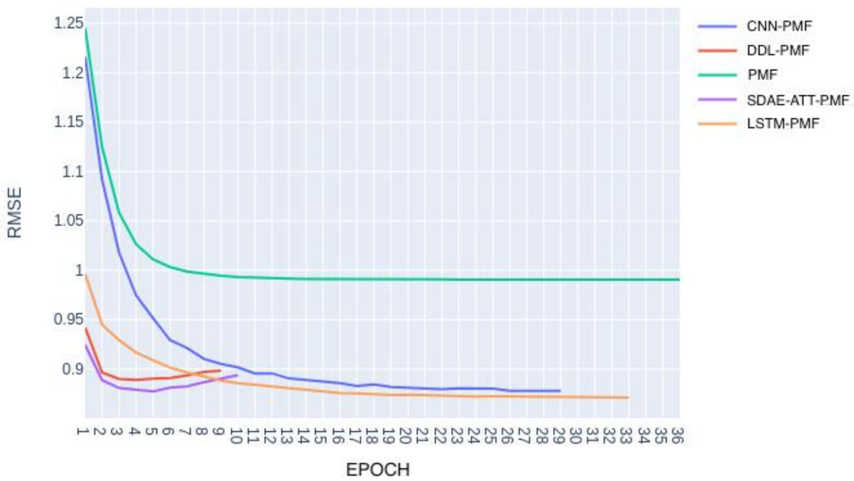


Figure. 8 Graphic RMSE evaluation result and effect of SDAE and attention

*W* representation. The share weight value obtains by attention mechanism more effective with seq2seq aspect over previous work based on CNN and LTSM. Table 6 show that proposed model achieves better performance in effectiveness significantly.

Aim to observe the performance of the model in achieving convergence value for training process, Author tried to evaluated the model using curve illustration, where blue colour line represents of PMF training result, orange colour linr represent of DDL-PMF and green colour line represent of SDAE,

Received: June 7, 2022. Revised: July 4, 2022. 

attention and PMF (the proposed model). The detail of training process can be seen on Fig. 7.

PMF as traditional latent factor of collaborative filtering reach low accuracy and high computation to learn process because of only consider rating matrix factor to generate rating prediction. While, Orange colour line that represent SDAE-LSTM-PMF (DDL-PMF) achieve better performance significantly over PMF. Certainly, it was influenced by side information impact for user information representation in $U$ that support by share weight $W^{+}$ and also item information representation share weight in $W$. In this research, internal item share weight was enhancement using seq2seq aspect by attention mechanism. It was influence in effectiveness and reduce the computation cost in rating prediction learning process.

The second experiment involves some hybridization of collaborative filtering algorithm aim to observe seq2seq aspect of attention in capturing product review and SDAE as user information where green colour as PMF model, blue colour as CNN and PMF model, orange colour as LSTM and PMF model, red colour as hybridization between SDAE, LSTM and PMF model, and purple colour as hybridization between SDAE, Attention and PMF.

According to RMSE evaluation result in the term of curve illustration Fig. 8 can be observed that the model involving item information representation only (blue and orange) achieved better performance over PMF. In the other hand, they only consider $W$ as internal item share weight. However, the training process to reach convergence required high computation and repeat much more over the model that involve both user and item information representation (red and purple).

The different approached by DDL-PMF (SDAE, LSTM and PMF) and SDAE, Attention and PMF where they consider to adopt both of information to support PMF task. According to experiment result show on figure below, they achieved low computation. Both of them reach convergence quickly. Moreover, they reach better performance in effectiveness over previous work. We believed that the effectiveness of both model due to enhancement of user and item information by SDAE and attention mechanism that applies seq2seq aspect to produce internal share weight $W$ and $W^{+}$ to support $U$ as user representation and $V$ as item representation.

## 4. Conclusion

In this study, Author adopts attention mechanism to enhance collaborative filtering based by enhancement product document information. Attention mechanism responsible to enhance product document representation in previous work that majority model applied CNN and LSTM. Attention mechanism consider to implement seq2seq aspect. In the other hand, seq2seq aspect responsible to enhance product document understanding in the contextual point of view to support PMF in generating rating prediction.

The attention mechanism model that combined with SDAE and PMF applied in ML.1M. According to the experiment report and comparison, attention mechanism succeeds to generate rating prediction with tremendous result. Attention achieved better performance in 1.6% in average over DDL-PMF, 3% in average over PHD-MF, and 8% in average over traditional PMF. The impact of involvement product document enhancement using attention mechanism play important role in effectiveness of this model.

The second experiment demonstrated the involvement of attention mechanism suitable to adopt in huge datasets (ML.10M) that contain 10 Millions rating and success to increase effectiveness of rating prediction 2.5% in average over previous best perform using DDL-PMF, and achieved 8% in average over PMF model. Moreover, attention mechanism model also achieves in low repetition to achieve training convergence. Author believes that enhancement of item document representation based on attention and user information representation become essential factor in performance result.

For the future work, involvement user and item side information will become essential factor to handle sparse data problem. Social media information is promising information that will be integrated into collaborative filtering based on model based. Social media information can be generated easily. The challenge in the future is how to integrated and selected relevant information that adequate to adopt in recommender system.

## Conflicts of Interest

The authors declare no conflict of interest on this research.

## Acknowledgments

This work was supported by Universitas Amikom Yogyakarta, Indonesia.

## References

[1] Hanafi, R. Widyawati, and A. S. Widowati, “Effect of service quality and online servicescape toward customer satisfaction and loyalty mediated by perceived value”, In: *Proc. of IOP Conference Series: Earth and*

*Environmental Science*, Vol. 704, No. 1, 2021, doi: 10.1088/1755-1315/704/1/012011.

[2] J. A. Konstan and J. Riedl, "Recommender systems : from algorithms to user experience", *User Model User Adaptive*, pp. 101-123, 2012, doi: 10.1007/s11257-011-9112-x.

[3] 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.

[4] Hanafi, N. Suryana, and A. S. B. H. Basari, "Paper Survey and Example of Collaborative Filtering Implementation in Recommender System", *Journal of Theoretical and Applied Information Technology*, Vol. 95, No. 16, 2017.

[5] 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.

[6] J. Bobadilla, F. Ortega, A. Hernando, and A. Gutiérrez, "Recommender systems survey", *Knowledge-Based Systems*, Vol. 46, pp. 109-132, 2013, doi: 10.1016/j.knosys.2013.03.012.

[7] W. C. Neto, M. L. H. Alcaraz, R. V. García, and F. G. Sánchez, "Social knowledge-based recommender system. Application to the movies domain", *Expert Systems with Applications*, Vol. 39, No. 12, pp. 10990-11000, 2012, doi: 10.1016/j.eswa.2012.03.025.

[8] F. Ricci, L. Rokach, and B. Shapira, "Recommender Systems Handbook", *2nd ed. London: Springer New York Heidelberg Dordrecht London*, 2015, doi: 10.1007/978-0-387-85820-3.

[9] C. C. Aggarwal, "Recommender systems: The Textbook", *London: Springer International Publishing Switzerland*, 2016, doi: 10.1007/978-3-319-29659-3.

[10] Y. Koren, R. Bell, and C. Volinsky, "Matrix Factorization Techniques for Recommender Systems", *IEEE Publisher*, Vol. 40, No. 8, pp. 42-49, 2009.

[11] 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, doi: 10.1.1.38.744.

[12] S. Purushotham, Y. Liu, and C. C. J. Kuo, "Collaborative Topic Regression with Social Matrix Factorization for Recommendation Systems", In: *Proc. of International Conference on Machine Learning*, *Eidinburg*, 2012, doi: 10.1016/j.nimb.2008.03.150.

[13] R. Salakhutdinov and A. Mnih, "Probabilistic Matrix Factorization.", In: *Proc. of Advances in Neural Information Processing Systems 20th*, pp. 1257-1264, 2007, doi: 10.1145/1390156. 1390267.

[14] H. Wang, N. Wang, and D. Y. Yeung, "Collaborative Deep Learning for Recommender Systems", In: *Proc. of ACM Knowledge Discovery and Data Mining*, 2015. doi: 10.1145/2783258.2783273.

[15] D. Kim, C. Park, J. Oh, S. Lee, and H. Yu, "ConVolutional Matrix Factorization for Document Context-Aware Recommendation", *10th ACM Conference on Recommender System*, pp. 233-240, 2016, doi: 10.1145/2959100.2959165.

[16] 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, doi: 10.1088/1742-6596/1577/1/012006.

[17] 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, doi: 10.1088/1742-6596/1577/1/012002.

[18] 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, doi: https://doi.org/10.1155/2021/ 8751173

[19] J. Liu and D. Wang, "PHD : A Probabilistic Model of Hybrid Deep Collaborative Filtering for Recommender Systems", In: *Proc. of 9th Asian Conference on Machine Learning*, pp. 224-239, 2017.

[20] Hanafi, E. Pujastuti, A. Laksito, R. Hardi, R. Perwira, A. Arfriandi, Asroni, "Handling Sparse Rating Matrix for E-commerce Recommender System Using Hybrid Deep Learning Based on LSTM, SDAE and Latent Factor", *International Journal of Intelligent Engineering and Systems*, Vol. 15, No. 2, pp. 379-393, 2022, doi: 10.22266/ijies2022.0430.35.

[21] S. Chen and M. Wu, "Attention collaborative autoencoder for explicit recommender systems", *Electronics (Switzerland)*, Vol. 9, No. 10, pp. 1-12, 2020, doi: 10.3390/electronics9101716.

Received: June 7, 2022. Revised: July 4, 2022. 


[22] 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, doi: 10.1109/ACCESS.2018.2887100.

[23] A. Sunyoto and Hanafi, "Enhance Intrusion Detection (IDS) System Using Deep SDAE to Increase Effectiveness of Dimensional Reduction in Machine Learning and Deep Learning", *International Journal of Intelligent Engineering and Systems*, Vol. 15, No. 4, pp. 125-141, 2022, 10.22266/ijies2022.0831.13.

[24] C. Wang and D. M. Blei, "Collaborative Topic Modeling for Recommending Scientific Articles", In: *Proc. of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '11*, p. 448, 2011, doi: 10.1145/2020408.2020480.

[25] J. Wei, J. He, K. Chen, Y. Zhou, and Z. Tang, "Collaborative Filtering and Deep Learning Based Hybrid Recommendation for Cold Start Problem", In: *Proc. of 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, doi: 10.1109/DASC-PICom-DataCom-CyberSciTec.2016.149.

[26] Hanafi, N. Suryana, and A. S. H. Basari, "InVolve ConVolutional-NN to Generate Item Latent Factor Consider Product Genre to Increase Robustness in Product Sparse Data for E-commerce Recommendation", In: *Proc. of Journal of Physics: Conference Series*, Vol. 1201, No. 1, 2019. doi: 10.1088/1742-6596/1201/1/012004.

[27] 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, doi: 10.26555/ijain.v7i1.577.

[28] Hanafi, N. Suryana, and A. S. H. Basari, "InVolve ConVolutional-NN to Generate Item Latent Factor Consider Product Genre to Increase Robustness in Product Sparse Data for E-commerce Recommendation", *Journal of Physics: Conference Series*, Vol. 1201, No. 1, 2019, doi: 10.1088/1742-6596/1201/1/012004.

[29] 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, doi: 10.18517/ijaseit.9.3.8843.

[30] Hanafi, N. Suryana, and A. S. B. H. Basari, "Recommender System Based Tensor Candecomp Parafact Algorithm-ALS to Handle Sparse Data In Food Commerce Information Services", *International Journal of Simulation System Science and Technology*, pp. 1-9, 2019, doi: 10.5013/IJSSST.a.19.06.60.

[31] Hanafi, N. Suryana, and A. Samad, "Dynamic conVolutional neural network for eliminating item sparse data on recommender system", *International Journal of Advances in Intelligent Informatics*, Vol. 4, No. 3, pp. 226-237, 2018.

[32] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, and P. A. Manzagol, "Stacked Denoising Autoencoders: Learning Useful Representations in a Deep Network with a Local Denoising Criterion Pierre-Antoine Manzagol", *Journal of Machine Learning Research*, Vol. 11, pp. 3371-3408, 2010, doi: 10.1111/1467-8535.00290.

[33] 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.

[34] J. McAuley, "Amazon Product Data", 2021, https://jmcauley.ucsd.edu/data/amazon/ (accessed Aug. 23, 2021).