Journal of Engineering Science and Technology Review 14 (6) (2021) 137 - 144

JOURNAL OF
Engineering Science and Technology Review

www.jestr.org

Research Article

Dual Denoising Autoencoder Based on Neighbor-Attention Module for Implicit Feedback Recommendation

Qian Wan[1,*] and Chen Jiang [2]

[1]*Beijing Minghe Intelligent Technology Co. Ltd., Beijing, 100089, China*
[2]*Graduate School, Dongseo University, Busan, 47011, Republic of Korea*

Received 11 March 2021; Accepted 28 December 2021

Abstract

In recent years, the content-aware recommendation of implicit feedback data using denoising autoencoder has become the mainstream technology in the field of recommendation systems. However, the technology still faces the following main problems: the denoising autoencoder structure ignores the intrinsic structural duality property of the model, and the recommendation of traditional implicit feedback data based on the denoising autoencoder ignores the neighbor item information. To solve these problems, a dual denoising autoencoder based on neighbor-attention module (DDAENAM) was proposed. The encoder and the decoder were designed as a dual closed loop, and the dual attribute of the structure was used to train the encoder and the decoder at the same time, such that the feedback signal between the encoder and the decoder can be shared in the model. Then, a neighbor-attention module was proposed to extract word embedding and neighboring item information. Results show that in three standard implicit feedback datasets including CiteULike-a, MovieLens-20M, and Amazon-Books, DDAENAM achieves the best result compared with the state-of-the-art models, with an average improvement of 2.4%. Under the review of evaluation indices select recall rate, DDAENAM is slightly lower than the joint representation learning model (JRLM), except for the MovieLens-20M. The proposed DDAENAM in this study achieves the best result with the combination of three modules including attention module, term–attention module, and neighbor–attention module, indicating the effectiveness of the DDAENAM module.

Keywords: denoising autoencoder, neighbor-attention, feedback recommendation, dual learning

1. Introduction

Recently, recommendation systems have been widely applied to commodity recommendations of E-commerce, video recommendations of We-Media, and other fields. With a group of users, commodities as well as interaction information of users and commodities, the recommendation system can recommend goods according to the preferences of users. Personalized recommendation is one of the key applications of machine learning in E-commerce and other fields.

During the development of personalized recommendation system, two types of common data are usually needed, namely, user grading and commodity description, such as user grading or comments to a movie and commodity description abstract. At present, latent Dirichlet allocation [1], stacked denoising autoencoder (SDA) [2], and variational autoencoder (VAE) [3] are common modeling techniques for these two types of data. These models improve the performances of top-N recommendations based on user grading and commodity description. Collaborative deep learning and collaborative variational autoencoder (CVAE) are two representative learning methods that link learning of commodity contents to the recommendation task. These two methods generally use the SDA and VAE models. They learn expressions in the hidden layer from the bag-of-word model of commodity contents. Moreover, these two methods use the probability matrix decomposition technology.

Although existing personalization recommendation systems have achieved satisfying outcomes, they still have several limitations: (1) when traditional methods extract expressions of commodity information in the hidden layer by using the denoising autoencoder structure, they often ignore structural dual properties between the encoder and the decoder due to their dual properties in the model structure. The encoder and decoder structures are trained independently, thus resulting in no sharing of feedback signals between them. (2) Traditional methods ignore adjacent relation information among commodities, such as mutual citing information of movies of the same school or articles. Closely related commodities are very likely to share the same theme or have similar attributes. Therefore, exploring users' preference to similarity commodities of a given commodity is also beneficial to deducing the users' preference of the commodity. Traditional methods do not consider the importance of different terms of a commodity when learning expressions of word vectors in the hidden layer. Similarly, treatment of information words and other words might cause incomplete understanding of users on commodity information.

To solve the above problems, a content perception recommendation model is proposed via dual denoising autoencoder based on neighbor-attention module (DDAENAM). DDAENAM includes two modules: dual denoising autoencoder (DDAE), neighbor–attention module, and term–attention module. First, inspired by dual learning machine, a denoising autoencoder is designed as DDAE. Using the encoder and the decoder of the denoising autoencoder as the

*E-mail address: xjwjj06@126.com
ISSN: 1791-2377
doi:10.25103/jestr.146.16

dual closed-loop is suggested. The encoder and the decoder are trained simultaneously by using the strategy gradient approach, thus making them share feedback information. Next, the hidden layer characteristics among similar products are learned by designing a neighbor–attention module. Finally, a term–attention module is designed to learn the embedding of commodities from a series of terms. In the embedding layer, information terms could be selected in situations where complicated recurrent neural network or convolutional network is inapplicable. In this study, the proposed DDAENAM is assessed on three real-world datasets, and the validity of the proposed DDAE and attention mechanism is proposed.

The remainder of this study is organized as follows: Section 2 introduces the related work in recommendation system based on implicit feedback and dual learning. Section 3 describes the design of DDAENAM. Section 4 describes the experimental setup and analyzes the experimental results. Section 5 summarizes the conclusions.

2. State of the art

2.1 Recommendation system based on implicit feedback

Research on early recommendation system mostly concentrates on recommendation based on the display feedbacks of users [4-5]. Recently, research concerning the recommendation system field turned to positive recommendations based on implicit feedback of users [6]. One type of major method of implicit recommendation system based on top-N is to use collaborative filtering technology to recommend to users a commodity list that they might be interested in. Item-based recommendation is more challenging than rate-based recommendation [7] and closer to recommendation scenes in practical life. Early works provide item-based recommendations by using the latent factor models. They use all missing data as negative sample through a uniform weighted mode or sample missing data from negative samples. Recently, Khan et al. [8] proposed a collaborative denoising auto-encoder (CDAE) based on implicit feedback top-N recommendation. Pádua et al. [9] proposed a collaborative filtering model based on neural network and uses the nonlinear user–commodity information interaction of multilayer sensor learning. Zhou et al. [10] proposed a recommendation algorithm based on representational learning method. The improved algorithm presented that the multiobjective node representation learning method based on bipartite graph network can solve the problem of dimensional disaster. Pan et al. [11] indicated that the proposal of a new measurement method for trust similarity can improve the sparsity of social data, deeply integrate trust data with users' implicit interactive information by using a denoising autoencoder, and, thus, profoundly incorporate the influence of social trust information.

2.2 Attention mechanisms in the recommendation system

The attention mechanism in the recommendation system means that attentions to different behaviors of a user are different during model prediction. Recently, attention mechanism has achieved great successes in many machine learning tasks, such as text classification in the natural language processing field [12] and machine translation [13]. In the recommendation system field, attention mechanism is also applicable. Cheng [14] proposed a model based on attention mechanism to measure the correlation dependence between users and commodity information. Liu et al. [15] proposed an autoencoder based on attention mechanism to judge user preferences.

The proposed neighbor–attention module and term–attention module are different from previous research. For the neighbor–attention module, the hidden layer characteristics of neighbor commodities are learned according to the importance score of the commodity and its neighbor commodity. In previous studies, information among commodities was hardly considered. For term–attention module, commodity information terms are selected by the multidimensional attention mechanism through comparison of scoring vectors.

Ni et al. [16] proposed a new recommendation model: comparative convolutional dynamic multiple attention. The model provides an accurate representation of user and item features and dynamically extracts potential feature vectors of user and items by using a multi-attention-based convolutional neural network. The multi-attention mechanism considers self-attention and cross-attention. Self-attention refers to the attention within users and items, while cross-attention refers to the mutual attention between users and items. A recommendation model of hierarchical attention cooperative neural network was also proposed. Du et al. [17] adopted a hierarchical attention mechanism to enrich the feature representation of users and items from the comment text. It uses two comment text-based parallel networks to conduct modeling for users and items separately, such that that the generated features are targeted.

2.3 Dual learning

Dual learning is a new machine learning paradigm. Dual learning solves the problem of insufficient training data in the practical application of machine learning. When sufficient artificially annotated large-scaled data for effective training are lacking, other signals to drive the training have to be found. The signal used in dual learning exists naturally in artificial intelligence (AI) tasks, but it is hardly used by people. Such signal is called structural dual property in AI tasks. Dual learning is proposed based on several observations, that is, many tasks of machine learning have dual forms. For example, English-to-Chinese translation and Chinese-to-English translation in machine translation are dual tasks. Image-based text generation and text-based image generation in the image understanding task are dual tasks. Voice recognition and voice synthesis in the voice recognition field are dual tasks.

Dual learning in machine translation is taken as an example. The machine translation task in Fig. 1 only has monolingual data that refer to document in English without annotations and a document in Chinese without annotations as well as two weak initial English-to-Chinese model and Chinese-to-English model. The dual learning task is to improve the translation abilities of the two initial models by continuous learning of monolingual data without annotations, finally obtaining two very strong English-to-Chinese model and Chinese-to-English model. To realize this goal, dual learning uses an English sentence without annotations and is then translated into Chinese through the initial English-to-Chinese translation model. Later, the Chinese sentence is translated again into English by the initial Chinese-to-English translation model. A series of feedback signals can be obtained to update the initial model by comparing the original sentence in English and the final sentence in English as well as grammar and terms in intermediate translation results. This process can be repeated continuously to improve models constantly. When mass monolingual data are obtained, dual learning can continuously improve the performances of translation models to very high standards.

The idea of dual learning is not only limited within the AI task that has dual tasks but also applicable to AI models with dual models, such as autoencoder (AE) and generative adversarial networks (GANs). The encoder and the decoder in AE all have structural dual properties, as shown in Fig. 2. The builder and the arbiter in GANs also have structural dual properties. However, these structural dual properties were often ignored in previous studies. The encoder and the decoder in AE are often trained separately in the recommendation system, and their feedback signals often are not shared. In this study, designing a denoising autoencoder as a dual closed-loop is suggested. The encoder and the decoder are trained simultaneously by using the structural dual properties between them, and feedback signals are used to update their training, thus improving the performances in extracting hidden layer characteristics of commodities.

Fang et al. [18] combined a dual learning algorithm with nonentangled representation theory and developed a new method called content-invariant dual learning to detect supervised or unsupervised change of remote sensing images. Yu et al. [19] proposed a new scheme to describe images from a visual and semantic perspective. Specifically, visual view was designed to capture appearance-level information in an image, including objects and their visual relationships. Meanwhile, semantic view enables the agent to understand high-level visual semantics from the entire image to local areas. On the basis of this dual-view image representation, we proposed a dual-encoded visual dialogue module that can adaptively select problem-related information from visual and semantic views in a hierarchy and focus on the visual content involved in various problems.

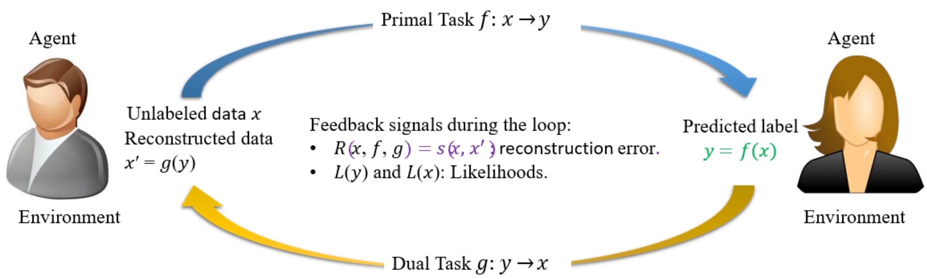

Fig. 1. Machine Translation based on Dual Learning

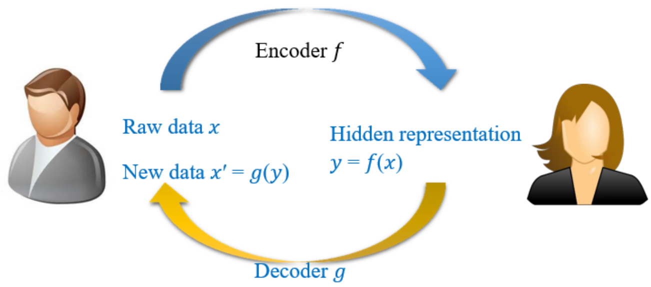

Fig. 2. Autoencoder based on Dual Learning

3. Methodology

The structure of the proposed DDAENAM is shown in Fig. 3. The main encoder is designed, and a dual decoder is used as the dual closed loop. The dual autoencoder learned in input binary user-item rating is introduced first. Next, the term-attention module is presented. Subsequently, the proposed neural portal structural module is discussed. Finally, the loss function and training process of the neighbor–attention module and model are introduced.

3.1 Dual denoising autoencoder

Constructing a model in user-item implicit data by using the stacking autoencoder to extract hidden layer characteristic information in implicit data is suggested. The traditional stacking autoencoder does not consider the structural dual information between the encoder and the decoder, and it trains the encoder and the decoder separately, with sharing feedback signals between them. A new dual stacking autoencoder structure is proposed here. First, the main encoder and a dual decoder structure are designed as a dual closed loop. Next, the main encoder and the dual decoder are trained commonly to make them share the feedback signals and thereby improve the performances of the dual stacking autoencoder in extracting hidden layer characteristics in implicit data.

The main encoder module is shown in Fig.3. First, it is used to encode the binary user-item rating $e_i \in R^n$, thus obtaining the hidden layer expression of item rating (e_i^r). The codes of the main encoder are expressed as follows:

$$\text{main encoder: } \begin{cases} e_i^1 = a_1(W_1 e_i + b_1) \\ e_i^r = a_2(W_2 e_i^1 + b_2) \end{cases} \tag{1}$$

The codes of dual decoder are expressed as follows:

$$\text{dual decoder: } \begin{cases} e_i^3 = a_3(W_3 e_i^r + b_3) \\ \hat{u} = a_4(W_4 e_i^3 + b_4) \end{cases} \tag{2}$$

where the subscript i in e_i^r refers to the specific item i information, and the superscript r is the hidden layer expression of codes in the binary item rating.

Moreover, $W_1 \in R^{v_1 \times m}$, $W_2 \in R^{v_1 \times m}$, $W_3 \in R^{v_1 \times m}$, and $W_4 \in R^{m \times v_1}$ represent the matrices of weight values. m is the number of users, v_1 is the dimension of the first hidden layer, and v denotes the dimension of the bottleneck layer.

3.2 Neighbor-attention module

After obtaining the integrating characteristic vector (e_i^j) of item rating and item content description, the item characteristic information similar with e_i^j, that is, neighbor–item characteristic information, is considered in the present study. In practical life, movies of the same school and papers with mutual citations have neighbor information. In the previous recommendation system, neighbor item information of a given item was often ignored.

In this study, a neighbor–attention module is proposed to extract characteristic information of an item and the neighbor item.

First, the neighbor item of a given item i is defined as ni. The information of n_i can be selected by setting item information similarity from the user binary rating. The hidden layer characteristics of ni are expressed by e_i^n :

$$e_i^n = \sum_{j \in n_i} soft\max(thanh(e_i^{gT} W_n e_i^g))e_i^g, \forall j \in n_i \tag{3}$$

where $W_n \in R^{v \times v}$ is the layer parameter of the neighbor–attention module. To extract information of i and n_i at the same time, the dual decoder is calculated as follows:

$$\hat{u}_i = a_4(W_4 a_3(W_3 e_i^g + b_3) + W_4 a_3(W_3 e_i^n + b_3) + b_4) \tag{4}$$

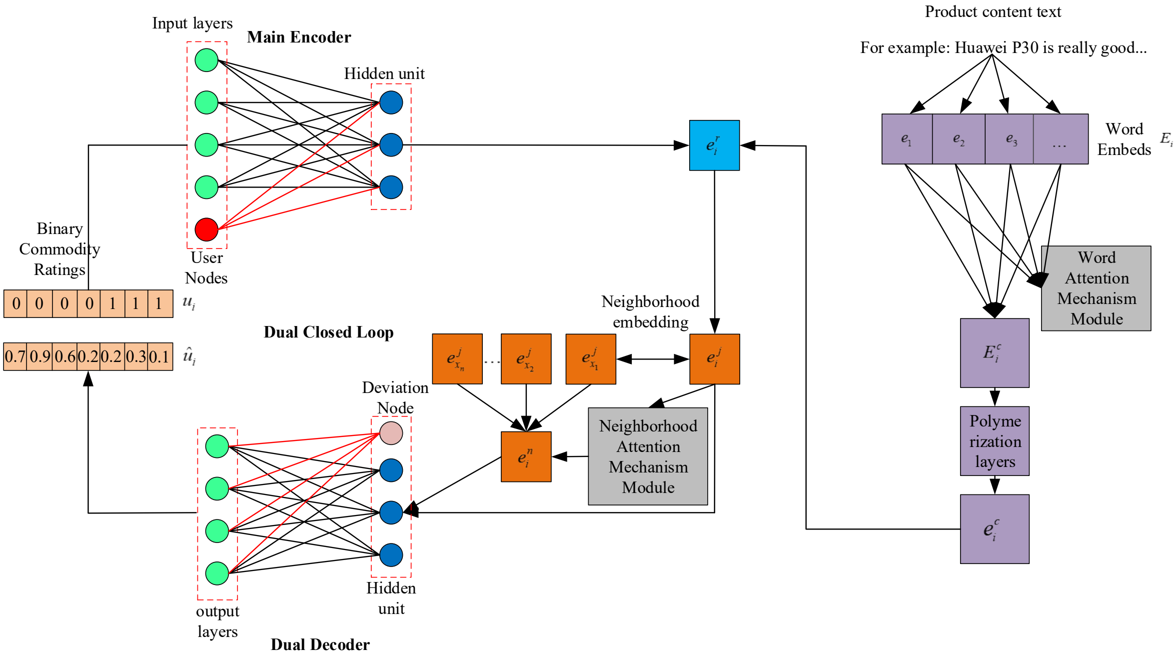

Fig. 3. Framework of Neighbor-Attention Dual Denoising Autoencoder

3.3 Term-attention module

The traditional item information embedding method ignores importance among terms of the same item. Here a term–attention module based on the descriptor sequence of the item is proposed. According to the purple module in Fig. 2, the proposed term–attention module selects information terms that have different importances to item description information adaptively.

First, item description text is input, such as “Huawei P30 is really good.” The text is embedded by using the embedding layer. The item is defined as i, and the description sentence of the item is defined as S_i. Each word is expressed by an independent thermal vector. The embedding layer encodes the independent thermal vector of the high-dimensional bag-of-word model of S_i and embeds it into a low-dimensional continuous real number vector. A word embedding matrix $E \in R^{d \times v}$ is encoded, where d refers to the embedding dimension of the word, and v denotes the number of words. The description text of an item is gained by inputting into the embedding layer.

$$E_i = \begin{bmatrix} & | & | & | & \\ \cdots & e_{m-1} & e_m & e_{m+1} & \cdots \\ & | & | & | & \end{bmatrix}, \text{ where } E_i \in R^{d \times v_i}, e_m \in R^d .$$

Inspired by the Transformer model in machine translation, the proposed term–attention module learns hidden layer information of description text of commodities by using multidimensional attention mechanism rather than complicated recurrent neural network and convolutional neural network because in practical applications, users prefer that item information can be summarized by several words rather than time-series of words. The term–attention module is used to distribute the embedding importance of different words from the word embedding E_i. It calculates the weight of the term–attention module from E_i by using the two-layer neural network of vanilla attention mechanism:

$$e_i = soft\max(w_{e_1}^{\mathrm{T}} \tanh(W_{e_2} E_i + b_{e_2})) , \tag{5}$$

where $w_{e1} \in R^{v}$, $w_{e2} \in R^{v\times v}$, and $b_{e2} \in R^{v}$. Here the SoftMax function assures that the sum of weight is 1. Next, the sum of weights of E_i is calculated:

$$E_i^c = \sum_{e_j \in E_i} e_{i,j} e_j \tag{6}$$

Here term–attention module is used to calculate the importance of single word embedding, which leads to a single item content recommended by the recommendation system model. If e_i is the item information containing a single dimension, the performance of the recommendation system decrease. Therefore, a term-attention module based on multidimensional e_i is proposed to calculate the weights of word embedding. Multidimensional e_i can contain different aspects of the contents of an item. Suppose c_e is the attention features extracted from word embedding and here $w_{e1} \in R^{c_e \times l_i}$ is expanded. The weights of the term-attention module are calculated by using the multidimensional attention mechanism:

$$e_i = soft\max(w_{e_1} \tanh(W_{e_2} E_i + b_{e_2}) + b_{e_1}) \tag{7}$$

where $e_i \in R^{c_e \times l_i}$ is the attention weight matrix, and $b_{e1} \in R^{c_e}$ is the error term. According to word embedding extracted by the multidimensional attention mechanism, the characteristic matrix of commodity content can be obtained:

$$E_i^c = E_i E_i^{\mathrm{T}} \tag{8}$$

Finally, the hidden layer characteristics of the content feature matrix of an item can be gained by using the neuropolymeric layer:

$$e_i^c = e_t(E_i^{c\mathrm{T}} w_t) \tag{9}$$

where w_t is the parameter of neuropolymeric layer, and $e_t()$ is an active function.

3.4 Model loss function and model training

The proposed model is constructed based on implicit data. For better modeling of users' preference information to items and implicit data information, the confidence matrix is inserted into the square loss function. The model loss function of the gate–attention dual autoencoder is designed as follows:

$$L_{GADAE} = \sum_{i=1}^{n}\sum_{j=1}^{m}\left\| C_{j,i}(D_{j,i} - \hat{D}_{j,i}) \right\|_2^2 = \left\| C_T \odot (D^T - \hat{D}^T) \right\|_F^2 \tag{10}$$

where $\odot$ is the product of elements, and $|\cdot\|_F$ is the F-paradigm of matrix. The confidence matrix is defined as follows:

$$C_{i,j} = \begin{cases} \rho & if\ the \quad D_{i,j} = 1 \\ 1 & if\ not \end{cases} , \tag{11}$$

where $C \in R^{m\times n}$.

The objective function of the gate–attention dual autoencoder is as follows: $C_{i,j} = \begin{cases} \rho & if\ the \quad D_{i,j} = 1 \\ 1 & if\ not \end{cases}$

$$L = L_{GADAE} + \lambda(\| W_* \|_F^2 + \| w_t \|_2^2) \tag{12}$$

Algorithm 1: Training of DDAENAM

Input: user binary rating vector (u_i), item content embedding vector (e_i^c), integrating vector (e_i^j), and neighbor vector (e_i^n);

Output: parameterized (θ_E) main encoder (*E*), parameterized (θ_D) dual encoder (D);

Initialize θ_E and θ_D randomly

repeat

Get a minibatch of m pairs $\{e_i, d_i\}_{j=1}^m$;

Calculate the gradient:

$G_E = \nabla_{\theta_E}(\frac{1}{m})\sum_{j=1}^{m}[L_1(f(e_j;\theta_E), y_j]$;

$G_D = \nabla_{\theta_D}(\frac{1}{m})\sum_{j=1}^{m}[L_2(f(d_j;\theta_D), y_j]$;

Update θ_E and θ_D

$\theta_E \leftarrow Opt_1(\theta_E, G_E), \theta_D \leftarrow Opt_2(\theta_D, G_D)$

until model convergence

The optimizers used are Opt_1 and Opt_2 . The main encoder and the dual decoder in DDAENAM are trained at the same time to share feedback signals between the two models.

4. Experimental analysis and discussion

4.1 Experimental data

The experimental dataset selects three standard implicit feedback datasets in the real world, including CiteULike-a, movielens-20M [20], and Amazon-Books. The three datasets choose different fields and have different sparsities. CiteULike-a provides user preferences for the title and content of articles. Movielens-20M is the movie description dataset collected from the online movie website IMDb. Amazon-Books is a comment dataset on the Amazon website. Statistics of experimental datasets are listed in Table 1.

Table 1. Experimental data statistics

Dataset	User	Item	Rate	Word	Density
CiteULike-a	5511	16980	204986	8000	0.217%
movielens-20M	138493	18307	1997049	12397	0.788%
Amazon-Books	65476	41264	1947765	27584	0.072%

4.2 Experimental evaluation indices

In the experiment, evaluation indices select recall rate (Recall@k) and normalized discount cumulative gain expressed in ranking index (NDCG@k). Recall@k is equal to a multi-interest inquiry in the recommendation system, that is, each user is an inquiry word, and then Top k items related with reach inquiry word are returned; this means that the percentage of Top k items that each user is interested in is returned. NDCG@k expresses that the accumulations of scores of relevance of each recommendation results are used as the score of the whole recommendation list.

4.3 Comparison experimental model

To verify the validity of the proposed DDAENAM on the content perception recommendation system, it is compared with three types of recommendation system model.

The first type is the traditional recommendation systems based on implicit feedback. 1. The weighted regularized matrix factorization (WRMF) uses the implicit binary feedback as a living binary example and then considers all items in the user-item interaction matrix (including unobserved items). WRMF has a confidence value to control weights of the positive and negative items. 2. CDAE uses DAE to learn hidden layer features from the implicit feedback information.

The second type uses the recommendation system models based on the bag-of-word model. 3. CVAE is a generative model with hidden variables and uses the VAE model to generate item content and user rating. 4. The collaborative metric learning model (CMLM) learns item features by using the metric learning strategy.

The third type uses recommendation system models based on word sequences. 5. Convolutional matrix factorization (ConvMF) extracts content information and features of items by the convolutional neural network and probability matrix decomposition technology. 6. Joint representation learning model (JRLM) [21] is a framework model that learns characteristics of the Top-N recommendation system jointly.

4.4 Experimental environment and experimental setting

The software and hardware environments for the operation of the experimental program are listed in Table 2.

Table 2. Experimental environment

Item	Details
Processor	Intel(R) Core i9-9900k
Graphics card	Nvidia GeForce GTX Titan X
Development language	Python 3.6
Development framework	Pytorch

First, the proposed DDAENAM is presented. The maximum length of word embedding for the content description of items is 300, and the grid searching hyperparameter of four datasets is [m,100,50,100,m]. The grid searching thresholds for CiteULike-a, movielens-20M, Amazon-Books, and Amazon-CDs are $\rho=5$, $\rho=20$, $\rho=15$, and $\rho=20$, respectively. The network learning rate is 0.01, and the batching size is 1024.

The experimental comparison models use WRMF, CDAE, CVAE, CMLM, ConvMF, and JRLM. First, the hidden variable dimensions of all models are set to 50. The experimental settings of WRMF and CDAE are consistent with those in the original papers. To realize better performances, the experimental parameters of the CVAE model are set as $\lambda_u=0.1$, $\lambda_v=10$, $\lambda_r=0.01$. For CMLM, the marginal parameter and the experimental parameters are m=2 and $\lambda_f=0.1$, $\lambda_c=1$. For ConvMF, the experimental settings are consistent with those in the original paper. For JRLM, the batching learning dimension is 64.

4.5 Comparison of model performances

The results of Recall@k and NDCG@k of the proposed DDAENAM as well as experimental comparison models WRMF, CDAE, CVAE, CMLM, ConvMF, and JRLM are proposed in this study. The results of Recall@k and NDCG@k of these models under CiteULike-a are shown in Fig. 4 and 5, respectively.

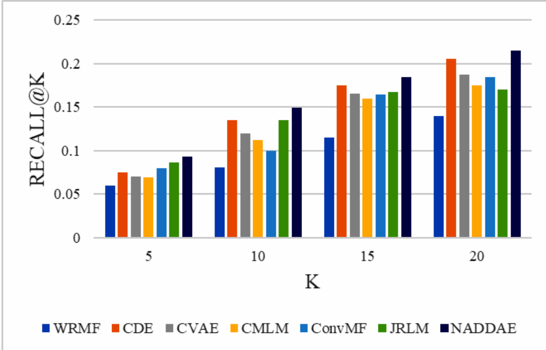

Fig. 4. Recall@k on CiteULike-a datasets

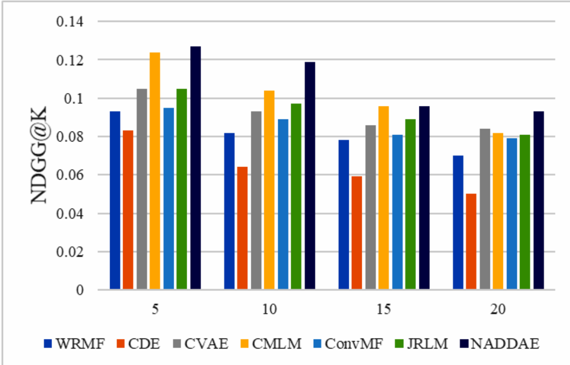

Fig. 5. NDCG@k on CiteULike-a datasets

The results of Recall@k and NDCG@k of these models under movielens-20M are shown in Fig. 6 and 7, respectively.

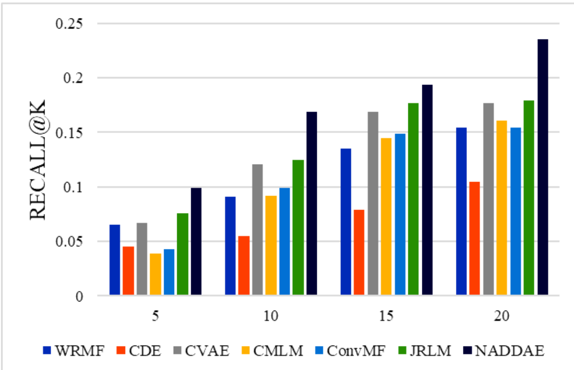

Fig. 6. Recall@k on movielens-20M datasets

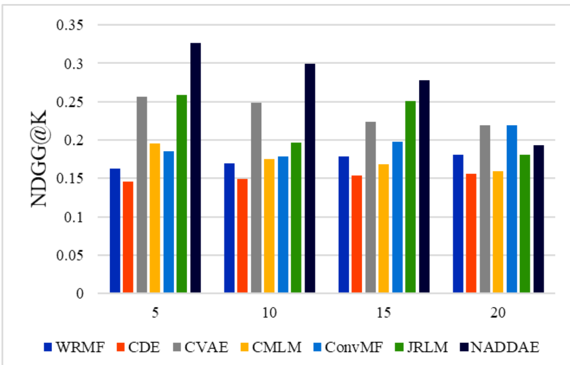

Fig. 7. NDCG@k on movielens-20M datasets

The results of Recall@k and NDCG@k of these models under Amazon-Books are shown in Fig. 8 and 9, respectively.

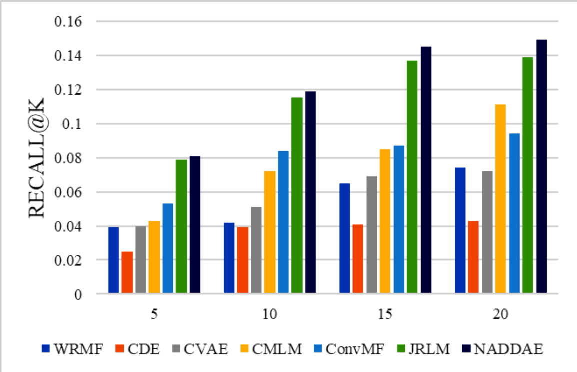

Fig .8. Recall@k on Amazon-Books datasets

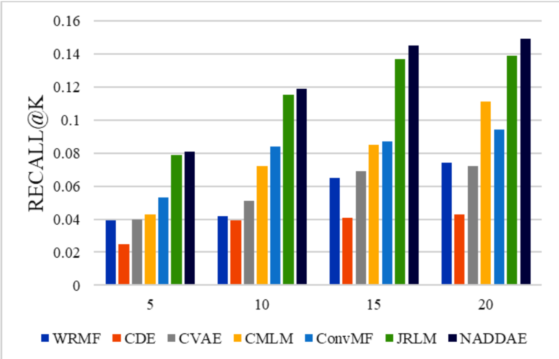

Fig. 9. NDCG@k on Amazon-Books datasets

According to the comparison results under the three datasets, the proposed DDAENAM is superior to WRMF, CDAE, CVAE, CMLM, ConvMF, and JRLM in terms of Recall@ (k=5,10,15,20). With respect to NDCG@k, DDAENAM is slightly inferior to JRLM in terms of NDCG@20 under movielens-20M, but it is superior to all six comparison models with respect to NDCG@ (k=5,10,15,20) under the two remaining datasets. The Recall@k and NDCG@k results of DDAENAM are better than those of ConvMF and JRLM, although they are constructed based on doc2vec[] model of sentence vectors and use convolutional neural network structure because ConvMF and JRLM ignore the neighbor–attention characteristic information of user rating and item contents. The proposed DDAENAM combines the neighbor item information of a given item and highlights the information words. The Recall@k and NDCG@k results of DDAENAM are better than those of CVAE and CMLM because they use the bag-of-word vector rather than the term–attention module during content embedding of items, thus resulting in information independence among characteristic words of an item. Differently, the proposed DDAENAM uses the word embedding technique and the term–attention module and considers the importance among description words of the same item. DDAENAM is substantially better than WRMF and CDAE. Finally, the Recall@k and NDCG@k results of DDAENAM are better than those of the frameworks based on autoencoder, CDAE and CVAE, because this study suggests designing the encoder and decoder structures in the autoencoder as a dual closed-loop, training them simultaneously, and considering structural dual information and feedback signals between them.

4.6 Validity analysis of modules

The proposed DDAENAM is composed of three modules, namely, DDAE module, neighbor–attention module, and term–attention module. To verify the validity of these three modules, an ablation analysis of the model is carried out. The Recall@k and NDCG@k results of the combinations of the above three modules are verified on CiteULike-a, movielens-20M, and Amazon-Books, where k=10.

The following models are applied for implicit feedback data recommendation: 1. DAE with the simplest denoising autoencoder structure, 2. DDAE with dual structural properties, 3. denoising autoencoder with neighbor–attention module (DAE+NA), 4. denoising autoencoder with term-attention module (DAE+WA), 5. denoising autoencoder with neighbor–attention module and term–attention module (DAE+NA+WA), 6. DDAE with neighbor–attention module (DDAE+NA), 7. DDAE with term-attention module (DDAE+WA), and the proposed DDAENAM integrating all three modules.

The experimental results are shown in Table 3. The results of DDAE are better than those of DAE, which proves the validity of DDAE. Recall@10 and NDCG@10 of DAE+WA+NA are better than those of DAE on the three datasets, which proves the validity of the attention module, the term–attention module, and neighbor–attention module. Recall@10 and NDCG@10 of the proposed DDAENAM under the combination of three modules are optimal under the three datasets.

Table 3. Model module analysis on three datasets

Models	CiteULike-a		Movielens-20M		Amazon-Books	
	Recall@10	NDCG@10	Recall@10	NDCG@10	Recall@10	NDCG@10
DAE	0.0347	0.0441	0.0377	0.0381	0.0741	0.0414
DDAE	0.0941	0.0677	0.0784	0.0978	0.0784	0.0357
DAE+WA	0.098	0.1079	0.0855	0.1574	0.0745	0.0441
DAE+NA	0.1158	0.0947	0.0957	0.1941	0.0957	0.0478
DAE+WA+NA	0.1253	0.1258	0.1244	0.2248	0.0991	0.0478
DDAE+WA	0.1142	0.0953	0.1027	0.1007	0.0944	0.0447
DDAE+NA	0.1341	0.1573	0.1477	0.2247	0.1027	0.0688
DDAENAM	0.1494	0.1973	0.1651	0.2970	0.1125	0.0746

5. Conclusions

In this study, a DDAENAM for implicit data recommendation was proposed. A DDAE of dual closed loop was designed. Specifically, the encoder and decoder were designed as dual closed loops and trained simultaneously by using the dual property of structure. Therefore, the feedback signal between the encoder and the decoder could be shared in this model. Combining neighbor attention and term attention, DDAE learned the expressions of hidden layer characteristics of item information. The item information words and neighbor item features were learned via neighbor attention and term attention, respectively. The following conclusions were obtained:

(1) The DDAE based on neighbor-attention module was compared with experimental comparison models, such as

WRMF, CDAE, CVAE, CMLM, ConvMF, and JRLM at Recall@k (k=5,10,15,20) and NDCG@k. In terms of CiteULike-a, MovieLens-20M, and Amazon-Books, DDAENAM achieves the best result compared with the six comparison models, with an average improvement of 2.4%. Under the NDCG@k review, DDAENAM is slightly lower than the JRLM model under NDCG@20, except for the MovieLens-20M. Nevertheless, DDAENAM in CiteULike-a and Amazon-Books under NDCG@k(k=5,10,15,20) is better than the six comparison models.

(2) In the analysis of the effectiveness of the DDAENAM module, model ablation analysis is performed to verify the Recall@k and NDCG@k of each module combination in the three datasets of CiteUlike-a, MovieLens-20M, and Amazon-Books. The proposed DDAENAM in this study achieves the best result at Recall@10 and NDCG@10 with the combination of three modules including attention module, term–attention module, and neighbor–attention module, indicating the effectiveness of the DDAENAM module.

DDAENAM improves performances by using the structural dual property of the autoencoder, but the input preprocessing text does not use the preprocessing linguistic model for text embedding. Future studies can adopt preprocessing and training of binary item rating and text to improve the performances of the whole model.

References

1. Xu, G., Wang, H. F., "The Development of Topic Models in Natural Language Processing". *Chinese Journal of Computers*, 34(08), 2011, pp.1423-1436.
2. Han, J., Zhang, D., Hu, X., Guo, L., Ren, J., Wu, F., "Background prior-based salient object detection via deep reconstruction residual". *IEEE Transactions on Circuits and Systems for Video Technology*, 25(8), 2014, pp. 1309-1321.
3. Bai, Q., Wu, Y., Zhou, J., He, L., "Aligned variational autoencoder for matching danmaku and video storylines". *Neurocomputing*, 454, 2021, pp. 228-237.
4. Wang, D., Chen, Z., Yue, W. J., Gao, X., Wang, F., "Probabilistic matrix factorization recommendation with explicit and implicit feedback". *Journal of Computer Applications*, 35(9), 2015, pp. 2574-2578.
5. Wei, X. U., Fu, D., "Combining clustering and collaborative filtering for implicit recommender system". *Computer Engineering and Design*, 35(12):2014, pp. 4181-4185.
6. Li, K. L., Su, H. D., Rong, J. Y., "Personalized recommendation system incorporated with the implicit feedback of projects and users". *Journal of Chinese Computer Systems*, 41(3), 2020, pp. 519-525.
7. Lim, H., Poleksic, A., Yao, Y., Tong, H., He, D., Zhuang, L., Meng, P., Xie, L., "Large-scale off-target identification using fast and accurate dual regularized one-class collaborative filtering and its application to drug repurposing". *PLoS Computational Biology*, 12(10), 2016, e1005135.
8. Khan, Z. A., Zubair, S., Imran, K., Ahmad, R., Butt, S. A., Chaudhary, N. I., "A new users rating-trend based collaborative denoising auto-encoder for top-n recommender systems". *IEEE Access*, 7, 2019, pp.141287-141310.
9. Pádua, F. L., Lacerda, A., Machado, A. C., Dalip, D. H., "Multimodal data fusion framework based on autoencoders for top-N recommender systems". *Applied Intelligence,* 49(9), 2019, pp.3267-3282.
10. Zhou, W., Yang, Y., Du, Y., Haq, A. U. Pairwise deep learning to rank for top-N recommendation. *Journal of Intelligent and Fuzzy Systems*, (4), 2021, pp. 1-12.
11. Pan, Y., He. F., Yu, H. A correlative denoising autoencoder to model social influence for top-N recommender system. *Frontiers of Computer Science*, 14(3), 2020, pp, 1-13.
12. Manshu, T., Bing, W., "Adding prior knowledge in hierarchical attention neural network for cross domain sentiment classification". *IEEE Access*, 7, 2019, pp.32578-32588.
13. Qiu, J., Liu, Y., Chai, Y., Si, Y., Wu, Y., "Dependency-based local attention approach to neural machine translation". *Computers, Materials and Continua*, 58(2), 2019, pp.547-562.
14. Cheng, L., "Research on the attention mechanism-based bidirectional LSTM model for the sentiment classiifcation of Chinese product reviews". *Software Engineer*, 20(11), 2017, pp.4-6.
15. Liu, H. H., Ren, H. R., He, H. C., "Group recommendation method based on self-attention mechanism". *Application Research of Computers*, 37(12), 2020, pp.3572-3577.
16. Ni, J., Huang, Z., Yu, C., Lv, D., Wang, C. Comparative convolutional dynamic multi-attention recommendation model. *IEEE Transactions on Neural Networks and Learning Systems*, 2021. doi: 10.1109/TNNLS.2021.3053245.
17. Du, Y., Wang, L., Peng, Z., Guo, W. Review-based Hierarchical Attention Cooperative Neural Networks for Recommendation. *Neurocomputing,* 447(8), 2021, pp.38-47.
18. Fang, B., Chen, G., Ouyang, G., Chen, J., Kou, R., Wang, L. Content-invariant dual learning for change detection in remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing,* doi: 10.1109/TGRS.2021.3064501.
19. Yu, J., Jiang, X., Qin, Z., Zhang, W., Hu, Y., Wu, Q. Learning dual encoding model for adaptive visual understanding in visual dialogue. *IEEE Transactions on Image Processing*, 30, 2021, pp.220-233.
20. Harper, F. M., Konstan, J. A., "The movielens datasets: History and context". *Acm Transactions on Interactive Intelligent Systems*, 5(4), 2015, pp.19.
21. Wang, Z., Xia, H., Chen, S., Chun, G., "Joint representation learning with ratings and reviews for recommendation". *Neurocomputing*, 425, 2021, pp.181-190.