Proceedings of CONF-SEML 2025 Symposium: Machine Learning Theory and Applications
DOI: 10.54254/2755-2721/2025.TJ23599

Research on the Design of Information Resource Automatic Classification and Recommendation System Based on Deep Learning

Xiao Han

Shungeng Campus, University of Jinan, Jinan, China
19153103170@163.com

Abstract: This article proposes a framework for automatic classification and recommendation of information resources based on deep learning. Through neural networks and natural language processing techniques, automatic classification of information resources can be effectively achieved, and personalized suggestions can be provided based on user behavior and semantic features of items. The classification adopts a combination of convolutional neural network and long short-term memory network for more accurate localization of text labels; The combination of deep learning collaboration and content recommendation algorithm is recommended to improve the recommendation effect. The experiment shows that this design method has improved accuracy and recommendation effectiveness compared to traditional classification and recommendation methods, and has the advantages of high accuracy and high real-time performance, which can meet the needs of information processing in large-scale data processing.

Keywords: Deep learning, Information resources, Automatic classification, Recommendation system, Data mining

1. Introduction

With the rapid development of information technology, big data classification and precise personalized recommendation have become research hotspots. However, traditional classification and recommendation methods cannot complete large-scale data analysis and suffer from problems such as large errors and low efficiency. Deep learning methods, as an excellent batch processing technology for large-scale data, have achieved excellent performance in natural language processing and recommendation systems. Based on the multi-layer perceptual neural network (MLP) model in deep learning, feature extraction and automatic classification of information resources can be achieved. Combined with user behavior data and item semantic features, personalized recommendations can be made to obtain more accurate information classification and higher recommendation results. It has good real-time and adaptability and can better address the difficulties brought by big data.

2. Neural networks and backpropagation algorithms

A neural network is a collection of multi-layered neurons that mimic the pattern of neural cells connecting to each other in the human brain, forming a deep digital structure. Nodes are connected by weights, and each neuron receives an input signal. The input information is weighted to obtain the output result, which is then processed through an activation function. The learning method of the network is completed through backpropagation algorithm, which calculates the error between the network output and the actual label, and backpropagates the error to each layer of the network, updating the weight of each layer.Let the neural network have L layers, and the output of each layer is$a^{(l)}$,wherel represents the l -th layer and $a^{(0)}$ is the input layer. The output of each layer is calculated using the following formula, as shown in formula (1).

$$a^{(l)} = \sigma(W^{(l)}a^{(l-1)} + b^{(l)}) \tag{1}$$

Among them,$a^{(l)}$represents the output of the l layer,$W^{(l)}$ is the weight matrix of the l -th layer, $b^{(l)}$ is the bias of the l -th layer,and σ is the activation function.The backpropagation algorithm uses the chain rule to calculate gradients and minimizes the error function through gradient descent. The steps include: first, calculating the output through forward propagation, and then calculating the error of the output layer based on the error function; Then calculate the error layer by layer and backpropagate it, reducing the error by adjusting the network weights.The calculation of the error in the output layer is shown in formula (2), which describes the gradient propagation mechanism of the loss function on the output.

$$\delta^{(L)} = \frac{\partial J}{\partial a^{(L)}} \cdot \sigma'(z^{(L)}) \tag{2}$$

Among them, $\frac{\partial J}{\partial a^{(L)}}$ is the gradient of the activation of the output layer by the loss function,$\sigma'(Z^{(L)})$ is the derivative of the activation function with respect to the weighted input. This formula describes the calculation method of output error and is a key step in the backpropagation algorithm.Through multiple iterations of training, neural networks can learn complex features and patterns in data. The backpropagation algorithm effectively solves the gradient calculation problem in neural network training and is the core technology of deep neural network training.

3. Architecture design of automatic classification and recommendation system

3.1. Overall system architecture and core module function design

This system is composed of modules, including data collection, text preprocessing, classification, recommendation, and user feedback. The main function is to automatically classify data information and provide personalized recommendations between these components. The data collection module collects consumers' activity history and item information from multiple channels. The text preprocessing module includes data cleaning, noise removal, and word segmentation to ensure the quality of the data. The classification module provides accurate classification results. Personalized recommendations based on user's past history and content features for joint learning and semantic feature extraction. The user feedback module provides feedback on recommendation results based on user interaction behavior, and can self learn and self adjust the system.

3.2. Classification module design: label recognition model based on neural network

Classification is one of the core modules of the entire system, responsible for accurately identifying various types of information resources as corresponding labels. This module uses neural networks for text feature extraction and multi label prediction, and adopts multi-layer perceptron (MLP) for text modeling, which is one of the mainstream methods of deep learning classification currently. [1]The input text is first converted into a sequence of word vectors and sent to the network for feature extraction through multiple hidden layers. Finally, in the output layer, the probability distribution of various labels is obtained through the softmax function, achieving automatic classification of multiple types of text.

In order to further improve classification performance, the model introduces a hybrid structure combining Convolutional Neural Network (CNN) and Long Short Term Memory Network (LSTM). CNN excels at extracting local important features, such as keyword phrases or fixed phrase structures, while LSTM can model long-range contextual relationships in text [2]. Combining the two can fully leverage the local perception ability of CNN and the temporal modeling ability of LSTM, thereby improving the semantic understanding depth and accuracy of classification.Assuming the input text consists of n words and their word vectors form a matrix$X \in \mathbb{R}^{n\times d}$, where d is the dimension of the word vectors. The CNN part uses convolution kernels$W_c \in \mathbb{R}^{k\times d}$ to perform convolution operations on continuous word vectors with a window size of k, obtain the feature representation as shown in formula (3).

$$c_i = f(W_c \cdot X_{i:i+k-1} + b_c) \tag{3}$$

Among them, f is the ReLU activation function, $b_c$is the bias term, $c_i$represents the activation output of the i-th window. All composed feature maps are max pooled to obtain fixed length vectors.The pooled feature sequence is input into the LSTM layer. LSTM preserves long-term dependency information through its gating mechanism, and the core calculation is shown in formula (4).

$$\begin{gathered} i_t = \sigma(W_i \cdot h_{t-1} + U_i \cdot x_t + b_i) \\ f_t = \sigma(W_f \cdot h_{t-1} + U_f \cdot x_t + b_f) \\ o_t = \sigma(W_o \cdot h_{t-1} + U_o \cdot x_t + b_o) \\ c_t = f_t \odot c_{t-1} + i_t \odot \tanh(W_c \cdot h_{t-1} + U_c \cdot x_t + b_c) \\ h_t = o_t \odot \tanh(c_t) \end{gathered} \tag{4}$$

The final output vector of LSTM is passed through a fully connected layer and a softmax classifier to output the probability distribution of the corresponding category. During the model training process, cross entropy is used as the loss function and backpropagation algorithm is employed for parameter updates.The overall model structure is shown in Figure 1.

Figure 1: Structure diagram of CNN-LSTM hybrid classification model

3.3. Recommendation module design: personalized algorithm combining semantic features

The recommendation module follows the core concept of the system to meet personalized needs. This module is mainly based on deep learning technology, combining behavioral data and item semantic features to establish a recommendation model. Firstly, analyze and study the user's past behavior data (such as browsing, purchasing, selection records, etc.) to establish the user's preference patterns; Further analyze and study the semantic features of the item, such as title, description, category, label, etc., to obtain deeper information about the item. By understanding the user's preferences and product features through the above methods, we can infer the likelihood of

DOI: 10.54254/2755-2721/2025.TJ23599

the user liking the product. In order to achieve better recommendation results, a mixed use approach is adopted: on the one hand, collaborative filtering is used to select other objects based on user similarity, and on the other hand, item attributes and consumers' past preference tendencies are used as references to find items.

4. Experimental analysis

4.1. Experimental dataset construction and model training parameter settings

This study mainly uses the Amazon Product Reviews Dataset, which contains approximately 100000 English text samples covering five major product categories including books, electronic products, and home furnishings. It has broad application value in recommendation systems and text classification, and has good representativeness and diversity. Data preprocessing includes noise text cleaning, duplicate item removal, English word segmentation (implemented using the NLTK library), lowercase unification, removal of special symbols, and stop word filtering based on the NLTK English stop word list. The text corpus is ultimately embedded using GloVe (Global Vectors for Word Representation) pre trained word vector models, which are converted into fixed length word vector sequences for efficient processing by neural network models. The dataset is divided into training set, validation set, and testing set in a ratio of 7:1:2, which are used for model training, hyperparameter adjustment, and performance evaluation, respectively. [3] This study adopts a hybrid model combining Convolutional Neural Network (CNN) and Long Short Term Memory Network (LSTM), which combines CNN's ability to extract local text structures with LSTM's ability to model contextual semantic relationships, to achieve multi-level mining of text features. In terms of hyperparameter settings, the learning rate is 0.001, the batch size is 32, the number of training rounds is 50, the loss function uses Cross Entropy Loss, and the optimizer uses Adam to improve the convergence speed of the model. To enhance the generalization ability of the model, cross validation techniques are used to avoid overfitting problems. To comprehensively evaluate the effectiveness of this model, multiple control experiments were conducted. In the classification task, Naive Bayes model based on TF-IDF features and K-Nearest Neighbors (KNN) classification algorithm are selected as the traditional benchmark methods; In the recommendation task, a comparison is made between the collaborative filtering recommendation algorithm based on Singular Value Decomposition (SVD) and the content-based recommendation algorithm. In terms of experimental evaluation dimensions, the classification model measures its performance through accuracy, precision, recall, and F1 score; The recommendation model uses metrics such as Root Mean Square Error (RMSE), accuracy @ k, recall @ k, coverage and diversity to comprehensively evaluate the accuracy and user satisfaction of the recommendation system. All models are trained and tested under the same dataset partitioning and parameter settings to ensure experimental fairness and result reliability.

4.2. Comparison experiment and practical application case verification of effectiveness

Take the item recommendation on e-commerce platforms as an example to verify the accuracy of system functions. By using web crawling technology to obtain consumer browsing, clicking, and shopping behavior data, text preprocessing is performed to clean, denoise, and segment the data. In the classification module, machine learning neural network algorithms are used to automatically classify products and extract product features. The recommendation module combines user behavior characteristics with product features to provide personalized recommendations for users through a combination of collaborative and content-based recommendation methods. The user feedback module makes corresponding adjustments to the user's recommendation results based on user

DOI: 10.54254/2755-2721/2025.TJ23599

interaction information feedback. The performance comparison of the classification and recommendation system before and after optimization is shown in Figure 2.

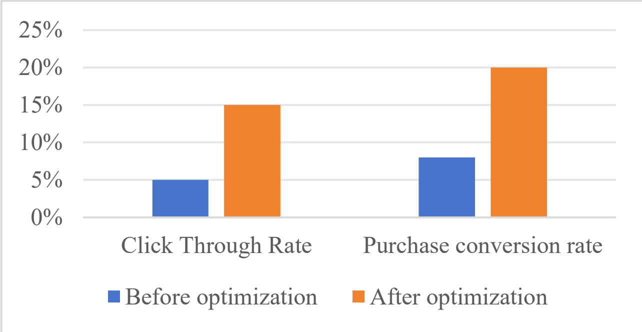

Figure 2: Performance comparison of classification and recommendation system before and after optimization

Figure 4 shows the comparison results of the two key indicators of "click through rate" and "purchase conversion rate" before and after system optimization. It can be seen that after optimization, the click through rate increased from 5% to 15%, a growth of 200%; The purchase conversion rate increased from 8% to 20%, a growth of 150%. This indicates that the introduction of deep learning CNN-LSTM model structure and personalized recommendation strategy has significantly improved the system's performance in user behavior prediction and conversion.

Conclusion: In the process of designing an automatic classification and recommendation system for information resources, this study proposes multiple innovative technological paths: (1) integrating convolutional neural networks (CNN) and long short-term memory networks (LSTM) in the classification model, fully combining CNN's ability to extract local features of text and LSTM's ability to model semantic sequence dependencies, effectively improving the accuracy of text label recognition; (2) In the recommendation module, a dynamic fusion strategy based on user behavior and item semantic features was introduced for the first time, and a personalized deep recommendation model with adaptive capability was constructed; (3) In terms of system architecture design, the modular integration of data collection, text preprocessing, deep learning modeling, and user feedback mechanisms enhances the robustness and scalability of the system in complex scenarios. Overall, the system combines neural networks and natural language processing techniques, and outperforms traditional methods in terms of information classification accuracy and personalized recommendation performance. The experimental results show that the proposed method has high accuracy, real-time performance, and system stability, and can effectively meet the needs of large-scale data processing. In the future, research can further enhance the diversity recommendation and real-time update capabilities of the model, promoting the evolution of information recommendation systems towards a more intelligent and adaptive direction.

References

[1] *Krishnamoorthi S , Shyam G K .DESIGN OF RECOMMENDENDATION SYSTEMS USING DEEP REINF ORCEMENT LEARNING – RECENT ADVANCEMENTS AND APPLICATIONS[J].journal of theoretical an d applied information technology, 2024, 102(7):2908-2923.*

[2] *Zhao Y , Zhao H .RESEARCH ON DATA MINING AND REINFORCEMENT LEARNING IN RECOMME NDATION SYSTEMS[J].Scalable Computing: Practice & Experience, 2024, 25(3).*

[3] *Gündoan, Esra, Kaya M , Daud A .Deep learning for journal recommendation system of research papers [J].Scientometrics, 2023, 128(1):461-481.*