symmetry

MDPI

Article

Feature Input Symmetry Algorithm of Multi-Modal Natural Language Library Based on BP Neural Network

Hao Lin

Institute of Linguistics, Shanghai International Studies University, Shanghai 200083, China; mmbg915@163.com

Received: 15 January 2019; Accepted: 22 February 2019; Published: 7 March 2019

Abstract: When using traditional knowledge retrieval algorithms to analyze whether the feature input of words in multi-modal natural language library is symmetrical, the symmetry of words cannot be analyzed, resulting in inaccurate analysis results. A feature input symmetric algorithm of multi-modal natural language library based on BP (back propagation) neural network is proposed in this paper. A Chinese abstract generation method based on multi-modal neural network is used to extract Chinese abstracts from images in multi-modal natural language library. The Word Sense Disambiguation (WSD) Model is constructed by the BP neural network. After the word or text disambiguation is performed on the Chinese abstract in the multi-modal natural language library, the feature input symmetry problem in the multi-modal natural language library is analyzed according to the sentence similarity. The experimental results show that the proposed algorithm can effectively analyze the eigenvalue symmetry problem of the multi-modal natural language library. The maximum error rate of the analysis algorithm is 7%, the growth rate of the analysis speed is up to 50%, and the average analysis time is 540.56 s. It has the advantages of small error and high efficiency.

Keywords: BP neural network; multi-modal; natural language; feature input; symmetric algorithm; WSD model

1. Introduction

Multi-modal natural language understanding is an important branch and frontier of AI (Artificial Intelligence). It is the culmination of various advances in AI and can be called the treetop of AI [1]. Its purpose is to understand the laws of language, analyze, and generate language [2]. Computational linguistics, as an important supporting discipline, has an in-depth theoretical framework and rich content. Natural language understanding is widely used. It is the core technology of web search technology. NLU (Natural language understanding) can be summarized as the trinity of human natural language itself, computational linguistics, and programming theory and technology. The linguistic theories and techniques realized by scientists and engineers have broken far beyond the scope and methods of linguists' conventional research and application of language. However, linguists and writers have difficulty in understanding natural language at present. The connection between the two is exactly where we are trying to integrate in the future. The discovery and construction of symmetry is a striking feature of Chinese language activities. When you start with your own words, what the characters see is a symmetry. Chinese is a language that starts to think. Starting from the couplets at home, Chinese people begin to symmetrically cultivate their ears. Chinese antithetical couplets are a kind of norm and training of thinking, which need to find symmetry, however they tend to be ordinary and word games [3].

At present, some researches have been done on the symmetry method of feature input of the multimodal natural language library. Literature [4] proposed a multi-modal data feature extraction

and selection method based on deep learning. In order to solve the huge structural difference between different language modes in a multi-modal environment, the feature extraction method of deep learning is introduced and the idea of deep learning is applied to multi-modal feature extraction. Then, a multi-modal neural network is proposed. For each mode, there is a multi-layer sub-neural network with an independent structure corresponding to it which is used to transform the features in different modes into the same modal features. At the same time, through a network layer common to all modes above these sub-neural networks, a connection is established between these different modes and, finally, a plurality of heterogeneous modes are converted into the same mode and a plurality of various modes are extracted therefrom. And the features of different data patterns are fused. However, this method cannot effectively remove redundant information or even noise information in advance, resulting in a large error rate. Literature [5] proposed a text language classification algorithm based on feature library projection. Firstly, the characteristics of all training samples are constructed according to a certain weighting strategy and all the sample feature information is retained by the feature library. Then, through the projection function, the feature database of each classification is mapped to the projection sample according to the feature set of the sample to be classified. Classification and input are completed by calculating the similarity between the new sample and each classified projection sample. However, this algorithm has a problem of high time complexity.

Therefore, in this paper, a feature input symmetry algorithm for a multi-modal natural language library based on BP (back propagation) neural network is proposed to extract Chinese abstracts from images in multi-modal natural language library and disambiguate words or texts in the abstracts. Finally, the similarity of words or texts after disambiguation is calculated and the feature input symmetry of the multi-modal natural language library is analyzed.

The rest of the paper is structured as follows: The second part mainly introduces the material method, including the introduction of the image Chinese abstract generation method based on multi-modal neural network, the process of constructing a WSD (Word Sense Disambiguation) model using BP neural network, and the characteristic input symmetry analysis in a multi-modal natural language library. The third part is the experimental part which mainly tests the performance and effectiveness of the method. The fourth part discusses the nonlinear mapping ability, self-learning and self-adaptive ability of the method, and discusses the generalization ability that is not fully reflected in the method. This also needs to be strengthened in the next research. Part five summarizes the full text.

2. Material Method

2.1. Method of Generating a Chinese Abstract of Images Based on Multi-Modal Neural Network

Image captioning is a cross-disciplinary subject that integrates computer vision, natural language processing, and machine learning. As the key technology of multi-modal processing, it has achieved remarkable results in recent years. At present, most of the research focuses on the generation of English abstracts from images, however few focuses on the generation of Chinese abstracts. This paper proposes a method of generating Chinese abstracts from images based on multi-modal neural network.

Most of the existing abstraction generation models of images are based on the encoder–decoder architecture. The encoder encodes the image to obtain visual features, and the decoder decodes the visual features to generate sentences, thus completing the multi-mode conversion from image to text [6]. This paper makes full use of multi-modal information, extracts image features and text features at the same time in the encoding process, and fuses multi-modal features in the decoding process to model the abstract.

2.1.1. Model Framework

Figure 1 shows the framework of a generation model of multi-modal neural network abstract.

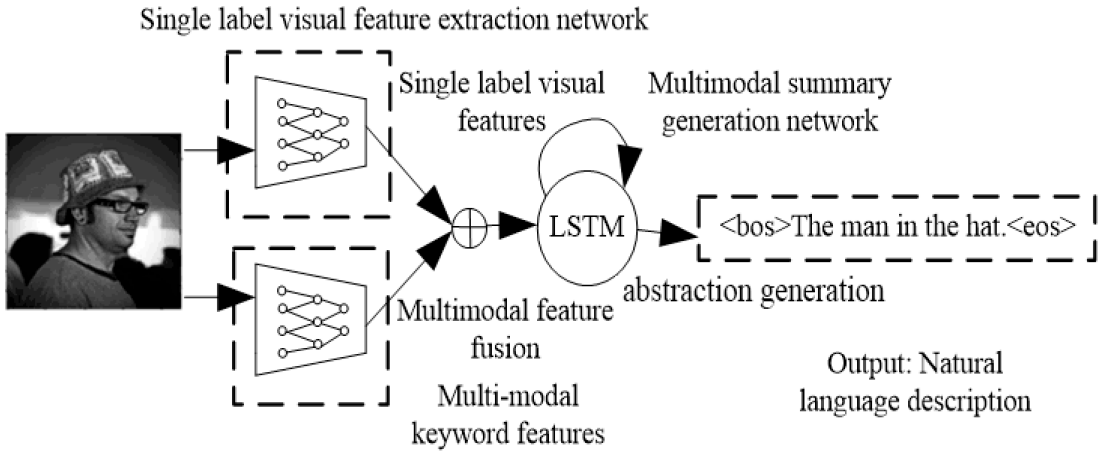

Figure 1. Framework for the summary generation model of multi-modal neural networks.

Encoder consists of two neural networks; one is the visual feature extraction network of a single label, the other is the keyword feature prediction network of a multi-label. Visual feature $V(I) \in R^n$ is the hidden layer output of a single label classification network which depicts the deep visual features of images, focuses on visual information, and uses real vector coding. Keyword features $W(I) = [w_1, w_2, ..., w_m]$ and $0 \leq w_i \leq 1$ are the output layer results of a multi-label classification network, reflecting the probability of keywords appearing in abstracts, focusing on text information, and using probability vector coding. The decoder consists of a multi-modal abstract generation network which integrates single-label visual features and multi-label keyword features, and the output is a Chinese sentence abstract of the image. Encoder is based on the convolutional neural network to model feature and is based on short-term and long-term memory network, decoder model sequence, and [7]. Because the data set of the Chinese abstract is limited, the generalization of neural network is reduced by using cooperative training method, so the three neural networks are trained separately on different data sets. For large enough abstract datasets, collaborative training is the ideal choice [8].

2.1.2. Visual Feature Extraction Network of the Single Label

For image I in multi-modal data set C, the visual feature extraction network $CNN_v(I)$ of the single label completes the feature mapping of $I \rightarrow V(I)R^n$, in which the network input is image I and the output is visual feature vector $V(I)$. Visual feature extraction network CNN_v adopts GoogleNet Inception V3 structure. The visual feature extraction network of the single label uses the idea of migration learning. The network is trained on ImageNet, a large-scale single label classification data set, and is tested on Flickr8k-CN, an image Chinese abstract data set. Different from the training process, the visual feature vector $V(I)$ extracted during the testing process is the hidden layer feature of the neural network, which can reflect the overall information of the image modality [9]. For the original image I in multi-modal data set C, firstly, the image is scaled and clipped to get three-channel RGB color image I' with the larger size of 299 × 299. Then, the image I' is processed by using the structure described in Table 1. The structure uses different Inception module groups to process the input matrix and splices the processing results of multiple module groups together. A highly structured feature is used to represent this. Finally, the whole feature of the image is obtained by aggregated features. Considering the difference between the single label classification task and the visual feature extraction task, we use Dropout regularization and normalization to improve the generalization ability of the model on the summary data [10]. Table 1 shows the setup of the single label's visual feature extraction network:

Table 1. Settings for a single label visual feature extraction network.

Type	Network Settings	Input Dimension
Convolution	3×3	$149 \times 149 \times 32$
Convolution	3×3	$147 \times 147 \times 32$
Hwa	3×3	$147 \times 147 \times 64$
Convolution	3×3	$73 \times 73 \times 64$
Convolution	3×3	$71 \times 71 \times 80$
Convolution	3×3	$35 \times 35 \times 192$
Incetion module group	3 Inception Modules	$35 \times 35 \times 288$
Incetion module group	5 Inception Modules	$17 \times 17 \times 768$
Incetion module group	3 Inception Modules	$8 \times 8 \times 1280$
Hwa	8×8	$8 \times 8 \times 2048$
Dropout	Keep ratio is 0.8	$1 \times 1 \times 2048$

2.1.3. Keyword Feature Prediction Network of the Multi-Label

For image I in multi-modal dataset C, the keyword feature prediction network $CNN_w(I)$ of the multi-label completes the feature mapping of $W(I)$, in which the network input is image I and the output is keyword feature vector $W(I)$. Keyword feature is encoded by probability vector, which reflects the probability of several keywords in the abstract. The keyword prediction network of the multi-label has six modules. The first five groups are consistent with VGGNet. Each group contains two to three convolution layers and one maximum pooling layer. The sixth module selects 1024 convolution cores to get 1024 sets of convolution outputs.

$x_k^5 \in R^{M_5 \times M_5}$ represents the K-th group features of the conv5-3 output of the group 5 module convolution layer, and Formula (1) represents group k feature mapping x_k^6 of the group 6 module convolution layer [11]:

$$x_k^6 = f\left(\sum_{p=1}^{n_5}\left(K_{kp}^6 \oplus x_p^5\right) + b_k^6\right) \tag{1}$$

where K_{kp}^6 represents the number of k-th convolution kernel connected to the p groups of characteristic x_p^5 in conv5-3, $\oplus$ represents the convolution operation, $b_k^6 \in R$ represents the offset of the k-th group, and $f(\cdot)$ is the ReLu activation function. The global average pooling process can be described as the form of Formula (2):

$$x_k = down\left(x_k^6\right) \tag{2}$$

In the formula, $down(\cdot)$ uses down sampling to calculate the average values of each group.

The training process is carried out on the multi-modal image summary data set. The multi-classification labels are constructed with the results of the summary segmentation and the corresponding keywords are recorded as 1. For T keyword features and N training data (I_i, L_i), where $i = \{1, ..., N\}$, $L_i = (l_{i1}, l_{i2}, ..., l_{iT})$, the loss function of the neural network can be expressed in the form of Formula (3):

$$(\theta) = -\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{T}\left[l_{ij}\log p_{ij} + \left(1 - l_{ij}\right)\log\left(1 - p_{ij}\right)\right] + \lambda_0\|\theta\|_2^2 \tag{3}$$

In order to avoid over-fitting, the regularization term $\lambda_\theta\|\theta\|_2^2$ is added, where θ represents the model parameters and λ_θ is the trade-off coefficient. p represents the calculation parameters of the keyword characteristics. Different from the training process, the testing process predicts the key words that may appear in the summary sentences.

2.1.4. Multi-Modal Abstract Generation Network

For visual feature $V(I)$ and keyword feature $W(I)$, multi-modal abstract generates network $RNN(V(I))$, and $W(I)$ completes the mapping of $V(I)$ and $W(I) \rightarrow S(I)$, where $S(I)$ is the

Chinese abstract of image I. In this paper, long-term and short-term memory network is used to model the abstract generation process. The calculation process of network at time t is h_t and $c_t = LSTM(x_{t-1}, h_{t-1}, c_{t-1})$, where $x_t \in R^d$ is the input of time t, $c_t \in R_d$ is the state, $h_t \in R_d$ is the hidden cell state, and $LSTM(\cdot)$ function is expressed as follows:

$$\begin{aligned} i_t &= \sigma(W_i x_t + U_i h\partial_{t-1} + b_i) \\ c_t &= \tanh(W_c x_t + U_c h_{t-1} + b_c) \\ f_t &= \sigma\Big(W_f x_t + U_f h_{t-1} + b_f\Big) \\ c_t &= i_t \oplus c_t + f_t \oplus c_{t-1} \\ o_t &= \sigma(W_o x_t + U_o h_{t-1} + V_o c_t + b_o) \\ h_t &= o_t \oplus \tanh(c_t) \end{aligned} \tag{4}$$

where $i_t \in R_d$ is the input gate, $f_t \in R_d$ is the forgetting gate, $o_t \in R_d$ is the output gate, and b and ∂ represent the input state vector. According to the characteristics of long-term and short-term memory networks, four methods of abstract generation based on multi-modal neural networks are proposed, namely CNIC-X, CNIC-C, CNIC-H, and CNIC-HC. When $t > -1$, the word vector $s_t \in R_d$ in the abstract sentence is projected as the input of time t, namely $x_{t>-1} = W_S st$. At $t = -1$, the multi-modal abstract generation network fuses the features and calculates the hidden states h_0 and c_0 at $t = 0$. Four multi-modal abstract generation methods are as follows:

a. CNIC-X adds visual information and keyword information as input x_{-1} at time $t = -1$, where W_V and W_W are projection matrices:

$$\begin{aligned} x_{-1} &= W_V V(I) + W_W W(I) \\ h_0, c_0 &= LSTM(x_{-1}, 0, 0) \end{aligned} \tag{5}$$

b. CNIC-H takes visual information as input x_{-1} of time $t = -1$ and initializes the hidden unit state h_{-1} of time $t = -1$ with keyword information $W(I)$:

$$\begin{aligned} x_{-1} &= W_V V(I) \\ h_{-1} &= W_W W(I) \\ h_0, c_0 &= LSTM(x_{-1}, h_{-1}, 0) \end{aligned} \tag{6}$$

c. CNIC-C is similar to CNIC-H, keyword information $W(I)$ is used to initialize cell state c_{-1} at time $t = -1$:

$$\begin{aligned} x_{-1} &= W_V V(I) \\ c_{-1} &= W_W W(I) \\ h_0, c_0 &= LSTM(x_{-1}, 0, x_{-1}) \end{aligned} \tag{7}$$

d. CNIC-HC uses keyword information $W(I)$ to initialize the hidden state h_{-1} and c_{-1} at the time $t = -1$:

$$\begin{aligned} x_{-1} &= W_V V(I) \\ h_{-1} &= W_{W1} W(I) \\ c_{-1} &= W_{W2} W(I) \\ h_0, c_0 &= LSTM(x_{-1}, h_{-1}, c_{-1}) \end{aligned} \tag{8}$$

The output o_t at time t is used as the probability estimate of $p(s_t|s_o, ..., s_{t-1}, W(I), V(I))$ in the multi-modal summary generation network. The training objective is to maximize the likelihood function $J(\theta)$ and estimate the model parameter θ:

$$J(\theta) = \sum_{i=1}^{N} \sum_{t=0}^{n_i} \log(s_t|s_0, ..., s_{t-1}, W(I_i), V(I_i)) \tag{9}$$

In the training process, the real value s_t is used to estimate p_{t+1}. Considering the difference between the training process and the testing process, the column search algorithm is used to maintain TopK abstract candidates and the candidate with the greatest probability is used as the output of the optimal solution [12–14].

2.2. Building WSD Model Using BP Neural Network.

After extracting Chinese abstracts from images in multi-modal natural language libraries based on Section 2.1, it is necessary to disambiguate words in multi-modal natural language libraries before identifying the similarity between words or texts. Therefore, this section uses BP neural network to construct a WSD model, which is mainly used for word or text disambiguation in Chinese abstracts of images in multi-modal natural language libraries.

For the BP neural network model, how to determine the topological structure of the neural network is very important to the application effect of the neural network [15].

2.2.1. How to Quantify Parameters.

WSD relies on the context content of the disambiguated word in the multi-modal natural language library to determine its meaning. Therefore, the input vector in the model should be the disambiguated word and its context content [16]. In this experiment, the mutual information of each sense item and its context word is calculated as the input vector. The MI (Mutual Information) calculation formula is as follows:

$$MI(w_1, w_2) = \log \frac{P(w_1, w_2)}{P(w_1)P(w_2)}. \tag{10}$$

In the formula, $P(w_1)$ and $P(w_2)$ denote the probability of w_1 and w_1 appearing alone in multi-modal natural language libraries, and $P(w_1, w_2)$ denotes the probability of the co-occurrence of w_1 and w_2.

2.2.2. Pretreatment Before the Experiment.

Before the experiment, multiple meanings of the words to be disambiguated in the multi-modal natural language library are used to replace the words.

$$setW_p = W_1/W_2/.../W_i \tag{11}$$

where W_p represents the disambiguation word in the multi-modal natural language library, and W_i is the i-th meaning of the word to be disambiguated. Therefore, it is represented by the i-th meaning of the disambiguation words in the multi-modal natural language library. In experiments, the meaning of polysemous words in each sentence is required in a multi-modal natural language library.

2.2.3. Determining the Word Input Vector.

According to statistics, it is found that most polysemous words in Chinese have two to four meanings, while in polysemous words, the most polysemous words are verbs, with an average of 2.56 meanings per verb.

Therefore, according to the statistical results, in the construction model, three meanings of common polysemous words are selected to construct their vector features [17,18]. In repeated experiments, the number of words in context is determined. By calculating the contextual location information gain (i.e., the influence of the context words to be disambiguated on the polysemous words), five words before and after the disambiguation words, namely $(-M + N)M = N = 5$, are selected.

In the model, the determined word input vector is:

$$\begin{aligned} V_{\text{input}} = \quad & \{MI11, MI12, ..., \text{MI1i}, \text{M.11}', \text{MI12}', \\ & ..., \text{MI1j}', \text{MI121}, \text{MI122}, ...\text{MI12}i, MI21', \\ & MI22', ..., MI2j', \text{MI31}, \text{MI32}, ...MI3i, \\ & MI31', MI32', ..., MI3j' \end{aligned} \tag{12}$$

In the Formula (12), $i = j = 5$. $M\text{Im}i$ denotes the mutual information between the mth sense item with disambiguation and the ith word above in the multi-modal natural language library; $M\text{Imy}'$ denotes the mutual information between the mth sense item with disambiguation and the jth word below in the multi-modal natural language library.

2.2.4. Determining the Word Output Vector.

For each word to be disambiguated in a multi-modal natural language library, a model is trained. The average meanings of polysemous words are two to three. Therefore, the average meanings of each word to be disambiguated in the experiment are two to three. In the experiment, three meanings of each word to be disambiguated are selected to form an output vector, namely, the three nodes defined in the output are respectively:

$$S1 = -1; \quad S2 = 0; \quad S3 = 1 \tag{13}$$

2.2.5. Determining the Number of Nodes in the Hidden Layer.

The number of nodes in the hidden layer is relatively flexible. In the experiment, the enumeration method is used to determine the optimal experimental results of 15 nodes [19]. Generally speaking, the number of nodes in the hidden layer is larger, which can reduce the number of iterations and improve the accuracy of the model. However, it is not the more nodes the better, but rather, it is the more complex the reasons are [20].

2.2.6. The Process of Training Models.

In the process of training, the multi-modal natural language database is used for training. Because this method is directed disambiguation, the selected corpus needs to be labeled manually [21–24]. At the same time, for each meaning of each polysemous word, the same number of examples should be chosen for learning. In the neural network, there is a balance problem. If there are more examples of one sense than others, it is likely that the model will focus on this sense in recognition. For example, for a polysemous word $W(w_1, w_2, w_3)$, there are 20 example sentences about w_1, 30 example sentences about w_2, and 50 example sentences about w_3. In the process of training corpus, we must choose the same example sentences to train, that is, 20 sentences of each choice to train, in order to prevent the inconsistency between meanings leading to the instability of the model [25].

2.3. Feature Input Symmetry Analysis in the Multi-Modal Natural Language Library

The symmetry of sentences is mainly judged by similarity. After disambiguating words or texts in multi-modal natural language libraries in Section 2.2, this paper analyses the feature input symmetry in multi-modal natural language libraries based on similarity.

According to the idea of blending semantic features into other features, this paper puts forward the calculation of similar words and puts emphasis on the calculation of similarity of component relations. Considering the morphology, word order, length, semantics, and component relations of sentences, it is hoped that the similarity of sentences in multi-modal natural language libraries can be measured more accurately [26].

2.3.1. Similar Words.

This refers to a pair of words of which their similarity reaches a certain threshold, for example, the threshold is set to 0.9 (the similarity range of words is (0, 1), including the same words and synonyms. In this paper, the word similarity calculation is based on the word similarity calculation method of CNKI (China National Knowledge Infrastructure) [27–29]. By putting forward the concept of similar words, we can avoid adding synonyms manually, avoid adding additional synonym dictionaries, and recognize synonyms automatically, so as to improve the similarity between sentences.

2.3.2. Improving Word Similarity.

Based on the concept of similar words, we can improve the calculation method of morphological similarity, replacing the number of the same words with the number of similar words, so as to improve the morphological similarity of sentences in line with the actual needs. The formula for calculating the similarity of the improved sentence A, B is as follows:

$$WordSim(A, B) = 2 * \frac{SimWord(A, B)}{Len(A) + Len(B)} \tag{14}$$

where $SimWord(A, B)$ is the number of similar words contained in A, B, and $Len(A)$ and $Len(B)$ are the number of words in sentence A, B.

2.3.3. Improving Word Order Similarity.

Similarly, when calculating word order similarity, this paper uses word order similarity of similar words to replace word order similarity of the same words, so as to improve the word order similarity of sentences [30].

2.3.4. Similarity of Component Relations.

This paper uses the open Chinese natural language processing system developed by Harbin Institute of Technology Social Computing and Information Retrieval Research Center to analyze language data.

After parsing, we can get the information of each component of a sentence. In this paper, we only use the five major component relations in a sentence: subject-predicate, verb-object, fixed-middle, adverbial-middle, and verb-complement. The number of occurrences of these five component relations in a sentence constitutes a vector, which is called component relation vector. Then, the cosine formula is used to calculate the cosine values of the component relation vectors of the two sentences and the similarity of the component relation between the two sentences is obtained [31].

For example: sentence A: "mobile bank supports those bank accounts?" Sentence B: "which bank accounts can use mobile banking services?" After parsing of A, B, we can get A4 component relational vectors of A, B: [1, 1, 3, 1, 0] and [1, 1, 4, 1, 0], respectively. The cosine similarity of A and B component relational vectors is calculated according to the cosine formula, as follows:

$$compSim(A, B) = \frac{\sum_{i=1}^{n} A * B}{\sqrt{\sum_{i=1}^{n} (A)^2} + \sqrt{\sum_{i=1}^{n} (B)^2}} \tag{15}$$

The similarity of the component relationship between A and B is 0.993 when vector A, B is replaced by the upper form.

2.3.5. Improving Sentence Similarity.

The calculation formula of sentence similarity based on a multi-feature mixture is as follows:

$$SentSim(A,B) = \alpha * WordSim(A,B) + \beta * SentLenSim(A, B) + \gamma * WordOrderSim(A,B) + \lambda * SematicSim(A,B) + \eta * CompSim(A,B) \quad (16)$$

Among them, $\alpha, \beta, \gamma, \lambda, \eta$ is the weight coefficient of each feature similarity and it satisfies $\alpha + \beta + \gamma + \lambda + \eta = 1$.

2.3.6. Algorithm Description.

Input: sentence A, B to be calculated similarity.

Output: similarity between sentences A and B;

Step 1: separate the sentences A and B separately.

Step 2: remove the stop words from the participle results and extract the keywords of the sentences.

Step 3: calculate the pairs of similar words in sentences A and B.

Step 4: syntactic analysis of sentence A and B to get component relation vector;

Step 5: The morphological similarity, sentence length similarity, word order similarity, semantic similarity, and component relationship similarity of sentences A and B are calculated respectively by using the obtained similarity word pairs and component relation vectors.

Step 6: Compute the sentence similarity of sentences A and B according to the improved sentence similarity formula and output the results.

3. Results

3.1. The Effectiveness of the Algorithm in this Paper

(1) Initialization, find out M sentences containing W in a large multi-modal natural language library;

(2) For each sentence containing disambiguated words, mutual information is calculated and input into the BP model, error between the output value and expected value is calculated, and weight is adjusted.

(3) When the error is close to 0, the model is completed; otherwise, it will return to (2).

(4) Test the unknown text, input the mutual information in its context, and export its meaning.

Using the proposed algorithm, the training data of the word meaning is predicted, as shown in Table 2. Hownet is used to train polysemous words by using people's daily corpus.

From Table 2, it can be found that for polysemous words with less than three meanings, such as "gu", there are only two meanings S1 and S2 in Hownet; three meanings are still selected, and only the third meaning is complemented by 0, so as to achieve the consistency of the model; for polysemous words with more than three meanings, such as "performance" in Hownet, we have four meanings. We choose the three most frequent items which are related to the specific corpus. It can be seen from the table that the input is actually a sparse matrix and for a context, it only biases towards a certain meaning. Figure 2 depicts the comparison between the predicted and actual values of the algorithm proposed in this paper.

As shown in Figure 2, the predicted value of the proposed algorithm is basically consistent with the actual value. The maximum predicted value of the proposed algorithm is 0.9, the actual value is 0.99, and the error is only $0.99 - 0.9 = 0.09$. This shows that the proposed algorithm can effectively analyze the feature input symmetry of multi-modal natural language libraries.

Table 2. Raw data training table.

	Discretion Words		Material	Guk	Performance	Lottery
The similarity between terms and context words	S1	*M*11	−0.9501	0	−0.4451	−0.6979
		*M*12	−0.8132	0	−0.1988	0
		*M*13	−0.8936	0	−0.0153	0
		*M*14	−1.4447	−0.5252	−0.7468	0
		*M*15	−0.8318	0	−1.2028	0
		*M*11′	−0.6721	0	−1.1389	0
		*M*12′	−0.9797	−0.5028	−1.2722	0
		*M*13′	−0.8801	0	−0.1988	−0.2523
		*M*14′	−0.2026	0	−1.8462	0
		*M*15′	−1.3784	0	−1.8381	0
	S2	*M*21	−1.3093	−1.2722	−1.8318	0
		*M*22	−1.6658	−0.2685	0	0
		*M*23	0	−1.1987	0	−0.7948
		*M*24	0	−0.6038	0	0
		*M*25	0	−0.8462	−1.9318	0
		*M*21′	0	−0.9318	0	0
		*M*22′	−1.9871	−0.466	−1.0196	0
		*M*23′	0	−1.4186	0	0
		*M*24′	0	0.5028	0	0
		*M*25′	0	−0.5678	0	0
	S3	*M*31	0	0	0	−0.8385
		*M*32	−1.5028	0	0	−0.5466
		*M*33	0	0	0	−0.7948
		*M*34	0	0	0	−0.173
		*M*35	0	0	0	−0.8757
		*M*31′	0	0	0	−0.8939
		*M*32′	−1.7095	0	−0.9883	−0.5836
		*M*33′	0	0	0	−1.5681
		*M*34′	0	0	0	−0.4449
		*M*35′	0	0	0	−0.9568

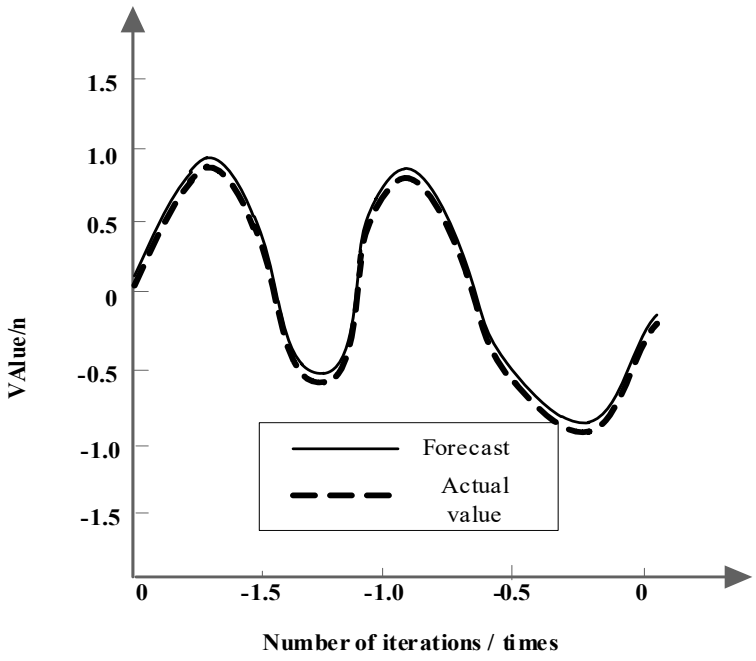

Figure 2. Comparison of the predicted and actual values of this algorithm.

3.2. Algorithm Performance Analysis

It can be seen from experiment 3.1 that the proposed algorithm can effectively analyze the symmetry of words in multi-modal natural language libraries. In order to verify the performance advantages of the proposed algorithm, the simulation experiments are carried out by using the proposed algorithm, the knowledge retrieval algorithm based on natural language processing, and a speech-to-word conversion efficient decoding algorithm based on the language model.

3.2.1. Comparison of Error Rates

Figure 3 shows the error rate of the input symmetry analysis results of the multi-modal natural language library of three different algorithms when the number of words analyzed is 40.

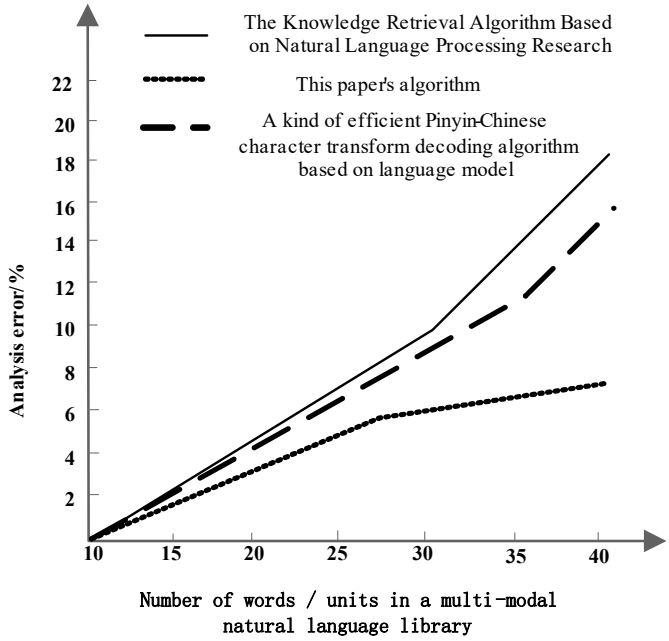

Figure 3. Comparison of error rates for symmetric analysis of three algorithms.

Analysis of Figure 3 shows that with the increase of the number of words to be analyzed in the multi-modal natural language library, when the number of words is 40, the maximum error rate of this algorithm is 7% for feature input symmetry analysis of the multi-modal natural language library; the knowledge retrieval algorithm based on natural language processing is used to analyze feature input symmetry of the multi-modal natural language library, and the maximum error rate is 20%. An efficient speech-to-word decoding algorithm based on the language model has a maximum error rate of 17% for feature input symmetry analysis of multi-modal natural language libraries. From this, we can see that the proposed algorithm has the least analysis error and the advantage of high accuracy.

3.2.2. Efficiency Comparison Results

Figure 4 shows the speed growth rate comparison results of the characteristic input symmetry analysis of the multi-modal natural language library of three different algorithms when the number of words analyzed is eight.

Analyzing Figure 4, we can see that the speed growth rate of the proposed algorithms is higher than that of the other two. With the increase of the number of experiments, the speed growth rate of the feature input symmetry analysis of the multi-modal natural language library is as high as 50% after the fifth experiment. The speed of feature input symmetry analysis for multi-modal natural language libraries based on knowledge retrieval of natural language processing increases by 20%, and the speed of feature input symmetry analysis for multi-modal natural language libraries by an efficient decoding

algorithm based on the language model increases by 20%. From this, we can see that the speed of this algorithm analysis is increasing rapidly.

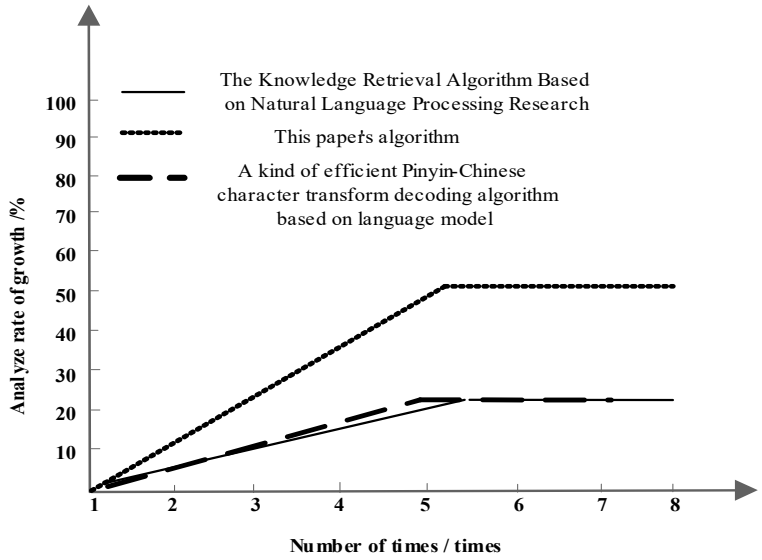

Figure 4. Comparison results of symmetric analysis speed growth rate of three algorithms.

The analysis time of the three algorithms in the above experiments is described in Tables 3–5.

Table 3. Time consuming analysis of the knowledge retrieval algorithm based on natural language processing/s.

Number of Experiments/Order	Number of Analytical Words/Number				
	10	20	30	40	Mean Value
1	168	168	167.5	167.4	167.725
2	336	336	335	334.8	335.45
3	504	504	502.5	502.2	503.175
4	672	672	670	669.6	670.9
5	840	840	837.5	837	838.625
6	1008	1008	1005	1004.4	1006.35
7	1176	1176	1172.5	1171.8	1174.075
8	1344	1344	1340	1339.2	1341.8
Mean value	756	756	753.75	753.3	754.76

Table 4. Time consuming analysis of this paper/s.

Number of Experiments/Times	Number of Analytical Words/Number				
	10	20	30	40	Mean Value
1	120	120	120	120.5	120.13
2	240	240	240	241	240.25
3	360	360	360	361.5	360.13
4	480	480	480	482	480.5
5	600	600	600	602.5	600.63
6	720	720	720	723	720.75
7	840	840	840	843.5	840.88
8	960	960	960	964	961
Mean value	540	540	540	542.25	540.56

Table 5. Time consuming analysis of a speech conversion algorithm based on language model/s.

Number of Experiments/Times	Number of Analytical Words/Number				
	10	**20**	**30**	**40**	**Mean Value**
1	144	144.5	144.5	1144.5	144.38
2	288	289	289	289	288.75
3	432	433.5	433.5	433.5	433.13
4	576	578	578	578	577.5
5	720	722.5	722.5	722.5	721.88
6	864	867	867	867	866.25
7	1008	1011.5	1011.5	1011.5	1010.63
8	1152	1156	1156	1156	1155
Mean value	648	650.25	650.25	775.25	680.94

From Table 3, Table 4, and Table 5, it can be seen that the average analysis time of the knowledge retrieval algorithm based on natural language processing is 754.76 s, the average analysis time of the proposed algorithm is 540.56 s, and the efficient decoding of a speech-to-character conversion based on the language model is 754.76 s. The average analysis time of an efficient speech-to-word conversion decoding algorithm based on the language model is 680.94 s. Compared with the other two algorithms, the average analysis time of the proposed algorithm is the lowest when analyzing the symmetry of feature input in the multi-modal natural language library. This algorithm has the advantage of high analysis efficiency.

4. Discussion

(1) Nonlinear mapping capability: The BP neural network used in this paper essentially implements the mapping function from multi-modal natural language library features from input to output. Mathematical theory proves that the three-layer neural network can approximate any nonlinear continuous function with arbitrary precision, which makes it especially suitable for solving complex internal mechanism problems. That is to say that the BP neural network is more suitable for analyzing this paper than other models. and can meet the requirements of input symmetry of natural language database.

(2) Self-learning and adaptive ability: the BP neural network can automatically extract the "reasonable rules" between output data by learning and adaptively memorize the learning content in network weights. That is to say that the BP neural network has strong self-learning and self-adaptive ability, which makes the proposed method more than the multi-modal data feature extraction and selection method based on deep learning and the text classification algorithm based on feature library projection. It can effectively cope with the symmetry of different words, making the final analysis result more accurate.

(3) Generalization ability: When designing the pattern classifier, the so-called generalization ability refers to whether the system can accurately classify and output the unknown mode or the noisy pollution mode after training. The goal of this part is this paper. This is also the focus of the next study.

5. Conclusions

Targeting the problem that the eigenvalue symmetry analysis of the existing methods have high error rates and long analysis times, this paper proposes a multi-modal natural language library feature input symmetric analysis algorithm based on the BP neural network. The image Chinese abstract generation method based on the multi-mode neural network is used to make full use of multi-mode information, extract image and text features in the encoding process, and integrates multi-mode features to model the abstract in the decoding process. The WSD model constructed by the neural network is mainly used for the disambiguation of the Chinese abstract of the image in the multi-modal natural language library. Finally, based on the similarity between words and texts, the eigenvalue

symmetry in the multimodal natural language libraries is analyzed. Through the analysis of the experimental data, it can be seen that the algorithm can effectively analyze the input symmetry of the multi-modal natural language library. After many experiments, the maximum error rate of the multimodal natural language library input symmetry analysis is 7% and the average analysis time is 540.56 s, which indicates that the algorithm not only has small analysis error, however it also has fast analysis efficiency and satisfactory results. The experimental results show that the method has more practical application value.

References

1. Jia, R.L. The Knowledge Retrieval Algorithm Based on Natural Language Processing Research. *Microelectron. Comput.* **2016**, *33*, 130–133.
2. Zhang, Z.Q.; Zhang, T.H.; Wu, Q. A Kind of Efficient Pinyin-Chinese Character Transform Decoding Algorithm Based on Language Model. *Intell. Comput. Appl.* **2016**, *6*, 38–41.
3. Li, H. A Review on the Research of Word Similarity Algorithms. *Modern Inf.* **2015**, *35*, 172–177.
4. Zhao, L. Research on Feature Extraction and Selection Method of Multimodal Data Based on Deep Learning. *Tianjin University.* **2016**, *13*, 56–60.
5. Yin, S.F.; Zheng, H.; Xu, S.H.; Rong, H.; Zhang, N. A Text Classification Algorithm Based on Feature Library Projection. *J. Cent. South Univ. (Sci. Technol.)* **2017**, *7*, 100–107.
6. Zhao, H.; Shi, S.; Jiang, H. Calibration of AOTF-Based 3D Measurement System Using Multiplane Model Based on Phase Fringe and BP Neural Network. *Opt. Express* **2017**, *25*, 10413. [CrossRef] [PubMed]
7. Jia, W.; Zhao, D.; Shen, T. An Optimized Classification Algorithm by BP Neural Network Based on PLS and HCA. *Appl. Intell.* **2015**, *43*, 1–16. [CrossRef]
8. Kolaei, A.; Rakheja, S.; Richard, M.J. A Coupled Multi-Modal and Boundary-Element Method for Analysis of Anti-Slosh Effectiveness of Partial Baffles in a Partly-Filled Container. *Comput. Fluids* **2015**, *107*, 43–58. [CrossRef]
9. Guo, J.G.; Huang, S. Study on Architecture of Big Data Based on Military Intelligence Analysis and Service System. *J. China Acad. Electron. Inf. Technol.* **2017**, *1*, 3–5182.
10. Zoccali, M.; Gonzalez, O.A.; Vasquez, S. The GIRAFFE Inner Bulge Survey (GIBS). I. Survey Description and a Kinematical Map of the Milky Way Bulge. *Astron. Astrophys.* **2017**, *562*, 118–130.
11. He, Y.; Zhuang, S. Research on Direct Current Control of Active Power Filter Based on Deadbeat Control. *J. Power Supply* **2015**, *99*, 652–658.
12. Gao, W.; Farahani, M.R.; Aslam, A.; Hosamani, S. Distance Learning Techniques for Ontology Similarity Measuring and Ontology Mapping. *Clust. Comput. J. Netw. Softw. Tools Appl.* **2017**, *20*, 959–968. [CrossRef]
13. Desai, R.; Patil, B.P. Adaptive Routing Based on Predictive Reinforcement Learning. *Int. J. Comput. Appl.* **2018**, *40*, 73–81. [CrossRef]
14. Schultz, D.; Spiegel, S.; Marwan, N. Approximation of Diagonal Line Based Measures in Recurrence Quantification Analysis. *Phys. Lett. A* **2015**, *379*, 997–1011. [CrossRef]
15. Zhang, W.H.; Pei, F.; Liu, P. Electrochemical Impedance Analysis of LiFePO_4/C Batteries in Cycling Process. *Chin. J. Power Sources* **2015**, *39*, 54–57.
16. Blandin, R.; Arnela, M.; Laboissière, R. Effects of Higher Order Propagation Modes in Vocal Tract like Geometries. *J. Acoust. Soc. Am.* **2015**, *137*, 832. [CrossRef] [PubMed]
17. Zhang, C.T.; Wang, L.K. Common Architecture Research Public Institutions Saving Optimization Control System. *Autom. Instrum.* **2015**, *98*, 50–52.
18. Hosamani, S.M.; Kulkarni, B.B.; Boli, R.G.; Gadag, V.M. QSPR Analysis of Certain Graph Theocratical Matrices and their Corresponding Energy. *Appl. Math. Nonlinear Sci.* **2017**, *2*, 131–150. [CrossRef]
19. Fornasiero, E.F.; Opazo, F. Super-Resolution Imaging for Cell Biologists: Concepts, Applications, Current Challenges and Developments. *Bioessays* **2015**, *37*, 436–451. [CrossRef] [PubMed]
20. Cui, G.M.; Zhang, D.D.; Liu, P.L. Research on New Gas Analytic Instrument. *Comput. Simul.* **2018**, *45*, 4–90.
21. Bernard, D.E. Multi-Modal Natural Language Query System for Processing and Analyzing Voice and Proximity-Based Queries. *J. Acoust. Soc. Am.* **2015**, *130*, 640. [CrossRef]

22. Mi, C.; Wang, J.; Mi, W. An Aimms-Based Decision-Making Model for Optimizing the Intelligent Stowage of Export Containers in a Single Bay. *Discret. Contin. Dyn. Syst. Ser S* **2019**, *12*, 1117–1133.
23. Ismaeel, S.; Karim, R.; Miri, A. Proactive Dynamic Virtual-Machine Consolidation for Energy Conservation in Cloud Data Centres. *J. Cloud Comput.* **2018**, *7*, 10. [CrossRef]
24. Hosny, M.; Raafat, M. On Generalization of Rough Multiset via Multiset Ideals. *J. Intell. Fuzzy Syst.* **2017**, *33*, 1249–1261. [CrossRef]
25. Wu, X.; Chen, H.; Wang, Y. BP Neural Network Based Continuous Objects Distribution Detection in WSNs. *Wirel. Netw.* **2016**, *22*, 1–13. [CrossRef]
26. Duncan, A.B.; Lelièvre, T.; Pavliotis, G.A. Variance Reduction Using Nonreversible Langevin Samplers. *J. Stat. Phys.* **2016**, *163*, 457–491. [CrossRef] [PubMed]
27. Qu, W.; Wang, D.; Feng, S.; Zhang, Y.; Yu, G. A Novel Cross-Modal Hashing Algorithm Based on Multi-Modal Deep Learning. *Sci. China (Inf. Sci.)* **2017**, *60*, 092104. [CrossRef]
28. Zdeněk, H.; Přemysl, S. Time Symmetry of Resource Constrained Project Scheduling with General Temporal Constraints and Take-Give Resources. *Ann. Oper. Res.* **2018**, *248*, 1–29.
29. Nekouie, N.; Yaghoobi, M. A New Method in Multi-Modal Optimization Based on Firefly Algorithm. *Artif. Intell. Rev.* **2016**, *46*, 1–21. [CrossRef]
30. Han, B. Algorithm for Constructing Symmetric Dual Framelet Filter Banks. *Math. Comput.* **2015**, *84*, 1–34. [CrossRef]
31. Yang, P.; Tang, K.; Lu, X. Improving Estimation of Distribution Algorithm on Multi-modal Problems by Detecting Promising Areas. *IEEE Trans. Cybern.* **2015**, *45*, 1438–1449. [CrossRef] [PubMed]