Deep Metric Learning for text data based on Triplet Network

Xin Deng 1,a, Wei Wu 2,b, Fei Wang 3,c

1 College of Computer Science and Software Engineering, Shenzhen University, Shenzhen, 518060 P.R. China
2 Software Development Department, Yunce Company Shenzhen, 518060 P.R. China
3 College of Computer Science and Software Engineering, Shenzhen University, Shenzhen, 518060 P.R. China

a 1062388964@qq.com
b wuwei0955@163.com
c wangfei2017@email.szu.edu.cn

Abstract—Text data has not only the characteristics of high dimension and serialization, but also most of the data are not labeled. This makes it difficult for the traditional method to achieve good results only relying on the training class mark to solve the problem of text classification. Later, some people put forward the method of metric learning to solve the problem of text classification. Its principle is to learn the measurement distance function for a specific task according to different tasks, that is to say, samples are classified by "distance measurement". The triplet network used in this paper is an algorithm based on "distance measurement". By optimizing the original triplet network, we can learn a non-linear measurement space with stronger data representation ability, so as to achieve better results than the traditional metric learning. Finally, we analyze the performance and efficiency of the method on IMDB movie review data set, and prove the superiority of the method for text classification.

1. INTRODUCTION

Triplet network is a method of deep metric learning, which is mainly used in image field at first. The principle is to learn an embedding function, which maps samples from feature space to measure space[1]. Then through the training parameters, the distance between the same kind of samples is close, and the distance between different kinds of samples is far away, so as to complete the classification[2]. Inspired by this method, this paper improves triplet network according to the characteristics of high dimension and serialization of text data. The model structure, loss function, selection of triplet and distance function which are more suitable for text classification scene are used. Finally, the experimental analysis of the method is carried out on IMDB movie review data set, and the experimental results are better than the traditional metric learning method.

2. RELATED WORK

Metric learning is an algorithm based on distance to realize classification[3]. As early as 2006, some people proposed the research method of text classification based on metric learning. The paper metric learning for text documents by Lebanon et al. has good reference significance and research value[4].
Later, as deep learning becomes more and more popular, there have been many researches on the classification of composition texts by deep metric learning. For example, the paper “Survey on distance metric learning and dimensionality reduction in data mining”[5] and “Low rank similarity metric learning in high dimension” published around 2015[6]. These researches make the deep metric learning method more and more popular.

Triplet network is a popular deep metric learning algorithm in recent years, which has been widely concerned by researchers all over the world. As early as 2014, Elad Hoffer first proposed the concept of triple network depth measurement learning in the paper "Deep metric learning using triplet network", defined the network structure of triple network and proved the advantages of triplet network method compared with traditional methods[7].

Later, during 2015-2017, Florian Schroff and Jiawei Liu et al. applied the triplet network method to face recognition and computer vision image classification[8][9]. They put forward some varieties of triple network and completed the end-to-end training combined with convolutional neural network. Finally, the accuracy of face recognition is higher than the traditional method, so triplet network is mostly used to study image problems.

Compared with triplet network in the field of image research, there are relatively few achievements in the field of text. In recent years, only Stefan Thaler have proved the feasibility of triplet network for text classification research[10]. Based on the previous research, this paper explores the general idea, network structure and application value of triplet network to solve the problem of text classification.

3. TRIPL ET NETWORK
Here we present our all experimental detail, including the model structure. We want to show that our optimized triplet network model has stronger representation ability and is more suitable for text type data.

3.1. Data source
Our dataset is IMDB movie review dataset. IMDB data set is a well-known data set, which is mostly used in the field of text classification. The dataset is a binary text dataset containing 50000 movie reviews from the Internet. The data set is divided into 25000 reviews for training and 25000 reviews for testing. Both the training set and the test set contain 50% positive and 50% negative reviews.

3.2. Data Pre-processing
In order to make the data adapt to the model better, we need to preprocess the data. First of all, we should remove the stop words that contain mood particles, conjunctions and special characters, etc. Next, we need to use word2vector to process sentences into the following formats:

\[
\text{Sent}_1 = \{W_{11}, W_{12}, \ldots, W_{1n}\}
\]

\[
\text{Sent}_2 = \{W_{21}, W_{22}, \ldots, W_{2n}\}
\]

\[
\ldots
\]

\[
\text{Sent}_{n-1} = \{W_{n1-1}, W_{n2-1}, \ldots, W_{m-1}\}
\]

\[
\text{Sent}_n = \{W_{n1}, W_{n2}, \ldots, W_{nn}\}
\]

Where \(\text{sent}_n\) stands for sentences, \(W_{nn}\) stands for words. And when the vocabulary is built, every word has a unique id, so the sentences can be transferred to vectors. We define word’s id as \(k_i\) in the upper form.

\[
\text{Sent}_1 = \{k_{11}, k_{12}, \ldots, k_{1n}\}
\]

\[
\text{Sent}_2 = \{k_{21}, k_{22}, \ldots, k_{2n}\}
\]

\[
\ldots
\]

\[
\text{Sent}_{n-1} = \{k_{n1-1}, k_{n2-1}, \ldots, k_{nn-1}\}
\]
Sentₙ={kₙ1,kₙ2,...,kₙn}

In addition, it should be noted to prevent the length inconsistency of the sequence, we also need to truncate the sentence to a uniform length.

Finally, we need to divide the training data into 80% training set and 20% verification set. We also need to scramble the data and repeat the experiment many times, which is also to reduce the influence of data sampling imbalance on the experimental results.

3.3. Model

As we know, the model is the core of an algorithm and plays an important role. Our model is an improved triplet network. The structure of our triplet network model includes the following parts: the triplet selection, loss function, measure space and final classification model. The model structure is shown in Figure 1 below:

![Figure 1. The model structure diagram of our triplet network](image)

In the above picture, <X', X, X> represents the triplet sequence of the input. Then, we define an LSTM network structure and output an 8-dimensional vector space. Then in the metric space, we use the Mahalanobis distance to calculate the distance before different sample points, so that the semantically similar sequences are close to each other and the dissimilar sequences are far away from each other. Then we define the triple loss function and use Adam as the optimizer to minimize the loss function. At this time, the samples of different categories can be separated. Finally, we use SVM model to complete the final classification.

Compared with the traditional triplet network, in order to better adapt to the text type data, our model uses the LSTM model instead of the traditional CNN model, uses the Mahalanobis distance instead of the traditional European distance and finally uses SVM as the classifier to complete the text classification. In addition, we can try some unsupervised methods to optimize the selection of triplet. Some outliers are excluded by outlier detection methods such as Isolated Forest and OneClassSvm, and the rest normal samples are used as anchor points. This method can effectively avoid the impact of outlier samples on the results, so as to achieve better classification results.

3.4. Triplet selection

Before training the model, we need to choose different triplet pairs (A, P, N). Where A represents the anchor point, P refers to the same sample as class A, and N refers to a different sample from class A.
Generally, we have two methods to traverse triples. One method is called hard_triplet_loss, which means that after randomly selecting an anchor point, the nearest heterogeneous sample and the farthest similar sample are selected to form a triplet. Another one is called all_triplet_loss, which means that after randomly selecting an anchor point, it traverses all samples of the same type and different types of anchor point to form n triples.

Obviously, the two ways have their own characteristics. Hard_triplet_loss can form the optimal triplet, but it has higher time and space complexity. Although the time complexity of all_triplet_loss is relatively low, but because there are many easy triplets that are not the best choice, the average loss is very small. Therefore, we should choose to use the triple loss according to the actual problems.

3.5. Triplet Loss

The calculation method of triple loss is different from that of traditional loss. It is not to adjust the loss by predicting the difference between the real class and the real class, but to minimize the loss function by adjusting the distance between the anchor point and the positive and negative samples. The calculation method is shown in Formula 1.

\[ L = \max(d(a, p) - d(a, n) + m \text{ margin}, 0) \]  

(1)

Where a is the anchor point, p is the same sample, n is the different sample, d is the distance between sample points, margin is the manually set distance.

From the above formula, we find that the traditional triplet loss only considers the relative distance between positive and negative sample pairs, but not the absolute distance between positive sample pairs. Therefore, we propose an improved triplet loss and add a driving term d(a, p). In this way, not only the distance between positive and negative samples in the feature space is far enough, but also the distance between positive samples is close enough. Experiments prove that our improved loss method has better results than the traditional triplet loss method. The specific formula is as follows:

\[ L = \max(d(a, p) + d(a, p) - d(a, n) + m \text{ margin}, 0) \]  

(2)

3.6. Distance metric

The traditional triplet network uses the Euclidean distance to measure the similarity between two unknown samples, but the biggest problem of the Euclidean distance is to treat the differences between different attributes of the samples equally, which can only be applied to the case where the measurement standards of each component are unified, and the text does not fit the characteristics of the text data.

In this paper, we use a more suitable Mahalanobis distance. Mahalanobis distance represents the covariance distance of data. Compared with Euclidean distance, it takes into account the relationship between various characteristics of samples and is not affected by dimensions, so it can more truly reflect the distance between different samples. The specific calculation method is shown in the formula:

\[ D_n(x) = \sqrt{(x - \mu)^T \sum^{-1}(x - \mu)} \]  

(3)

3.7. Algorithm pseudo code

The following algorithm pseudo code shows the process of our triplet model training.

**Algorithm 1** Triplet network training

1: Load data to X,Y
2: Update X,Y by word2vector
3: Split TrainX, TestX from X
4: Split TrainY, TestY from Y
5: Build LSTMNetwork
6: Update Y, Y' by input TrainX, TrainY into LSTM Network
7: Generate A, P, N by all_triplet_loss or hard_triplet_loss
8: Update AP by d(A-P)
9: Update AN by d(A-N)
10: Update L by max[AP-AN+margin, 0]
11: Train LSTMNetwork by L
12: Test LSTMNetwork use TestX, TestY

4. EXPERIMENT

4.1. Data distribution map
Figures 2 and 3 below reflect the distribution of test data before and after using the triplet network model. Because of the large sample size of IMDB full data set, we randomly selected 1000 samples to replace the full data set.

Figure 2. The Distribution of data before and after using triple network

Figure 3. The Distribution of data before and after using triple network
From the figure above, we can see that after the training of triple network, the data samples of different categories can be more distributed to achieve the effect of classification.

4.2. **Confusion matrix**

Figure 3 and 4 below shows the results of confusion matrix and normalized confusion matrix after using our improved triplet network model for 25000 training datasets. From the figure, We can see that triplet network can distinguish two types of text clearly.

![Figure 4. The confusion matrix of experimental results](image1)

![Figure 5. The normalized confusion matrix of experimental results](image2)

4.3. **Comparative experiment results**

Table 1 shows the experimental results of triplet network compared with other traditional metric learning algorithms and deep metric learning algorithms in IMDB dataset. From the results, it can be seen that triplet network is superior to the traditional metric learning algorithms in terms of the performance of several key indicators, such as accuracy, precision, recall and F1_score.
In the experiment in the table below, we choose two traditional metric learning methods and two deep metric learning methods to compare with our triplet network. From the experimental results, we can see that our improved triple network algorithm has obvious advantages in dealing with text classification.

| Evaluation | Metric Learning Method |
|------------|------------------------|
|            | LMNN | ITML | DDML | Siamese | Triplet |
| Accuracy   | 0.538 | 0.582 | 0.698 | 0.783 | 0.833 |
| Precision  | 0.549 | 0.580 | 0.704 | 0.792 | 0.839 |
| Recall     | 0.582 | 0.578 | 0.712 | 0.785 | 0.832 |
| F1_score   | 0.596 | 0.604 | 0.705 | 0.776 | 0.824 |

In addition, in the experiment, there are some key variables that affect the experimental results, such as: the number of anchor points selected, the number of layers of neural network, the setting of positive and negative sample margin, etc. We can optimize these indicators for better training results.

5. CONCLUSIONS
This paper explores a new method of text classification using deep metric learning, and proves that deep metric learning method based on triplet network can also achieve good results in text classification. Based on the traditional triplet network, this paper optimizes and innovates the triplet selection, triplet loss function, triplet network structure and distance measurement function, which makes the classification effect of triplet network better than that of traditional metric learning method in text field. We provide important academic value and application value for triplet network research in the field of text in the future.

In the future, there is still a large optimization space for triplet network applied in text field. We can optimize the triplet anchor selection and the distance function to achieve better experimental results. In addition, the research methods of this paper can also be extended to other fields.

REFERENCES
[1] Aggarwal C C, Zhai C X. A survey of text classification algorithms[M] Mining text data. Springer, Boston, MA, 2012: 163-222.
[2] Kulis B. Metric Learning: A Survey[J]. Foundations & Trends® in Machine Learning, 2013, 5(4):287-364.
[3] Nguyen B, Morell C, De Baets B. Large-scale distance metric learning for k-nearest neighbors regression[J]. Neurocomputing, 2016, 214: 805-814.
[4] Lebanon G. Metric learning for text documents[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2006, 28(4): 497-508.
[5] Weinberger K Q, Saul L K. Distance metric learning for large margin nearest neighbor classification[J]. Journal of Machine Learning Research, 2009, 10(Feb): 207-244.
[6] Davis J V, Dhillon I S. Structured metric learning for high dimensional problems[C] Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2008: 195-203.
[7] Hoffer E, Ailon N. Deep metric learning using triplet network[C] International Workshop on Similarity-Based Pattern Recognition. Springer, Cham, 2015: 84-92.
[8] Schroff F, Kalenichenko D, Philbin J. Facenet: A unified embedding for face recognition and clustering[C] Proceedings of the IEEE conference on computer vision and pattern recognition. 2015: 815-823.
[9] Liu J, Zha Z J, Tian Q I, et al. Multi-scale triplet cnn for person re-identification[C] Proceedings of the 24th ACM international conference on Multimedia. ACM, 2016: 192-196.
[10] Thaler S, Menkovski V, Petkovic M. Deep metric learning for sequential data using approximate information [C] International Conference on Machine Learning and Data Mining in Pattern Recognition. Springer, Cham, 2018: 269-282.