World Scientific Research Journal Volume 5 Issue 7, 2019
ISSN: 2472-3703 DOI: 10.6911/WSRJ.201907_5(7).0006

Optimization of Conditional Random Field Model Based on Circular Neural Network

Yiming Liu

Liaoning University of Science and Technology, China

Abstract

The rapid development of Internet technology drives the exponential growth of online data, marking the arrival of the era of big data, and also means that information extraction technology will assume more important tasks. People need to accurately and quickly obtain target information from a large amount of data, and further improve the utilization rate of information. However, the existing extraction techniques often have problems such as limited application scope and large amount of work required for manual operation of corpus part-of-speech tagging. Analyzed in this paper using hidden markov model (HMM) training based information extraction method of the problems and the insufficiency, with the help of a conditional random field (CRF) principle on the processing characteristics of the knowledge representation, conditional random field model was put forward, at the same time, combining the cycle neural network (RNN) was carried out on the input variable length data is divided into N parts such as input to the neural network, so as to realize the neural network of variable length of input data processing. Through the experimental evaluation, under the condition of text classification, a better result is obtained by comparing the traditional hidden markov and bayesian network models.

Keywords

Conditional random field, circulatory neural network, informatio variable length text classification.

1. INTRODUCTION CLASSIFICATION

Chinese automatic word segmentation is a basic subject in the field of Chinese information processing and also the key to intelligent Chinese information processing. Chinese automatic word segmentation plays an important role in the automatic processing of Chinese information. Since the end of the 20th century, Chinese word segmentation has attracted much attention and a number of promising methods have emerged. There are two main methods of word segmentation technology, which are rule-based and statistics-based.

Formally, a word is a stable combination of characters, so in the context, the more times the adjacent characters appear simultaneously, the more likely it is to form a word. Therefore, the frequency or probability of co-occurrence between words can better reflect the credibility of words. The method based on probability statistics starts from probability statistics and computational science, through the statistical analysis of large-scale corpus, the computer can process natural language. Based on the method of probability statistics, its model has a rich mathematical foundation, which can be widely used. There are Hidden Markov Models (HMMs), Maximun Entropy Models (MEMs) and Condition Random Fields (CRFs).

ISSN: 2472-3703 DOI: 10.6911/WSRJ.201907_5(7).0006

2. RELATED WORK

2.1. Based on Hidden Markov Classification Model

Hidden markov model (HMM)It can be used to describe the statistical characteristics of the random process of serialization. As one of the statistical characteristics of the random process, how to better apply the HMM model to the text processing has become one of the research hotspots at present. In HMM, each document d has a category c attribute, and the objective is to establish a classification model based on training set to complete the classification of new text. In this paper, HMM model is used as a document generator. As shown, each class is mapped to an HMM model, and the model parameters are obtained through the training of data sets belonging to this class. When a document is to be classified, the probability distribution size belonging to the current category is calculated first, and then as the basis of belonging to the current category.

2.2. From Hidden Markov to Condition Random Airport

In order to facilitate model processing, a strict assumption of independence is given. For example, in the hidden markov model, we assume that the observation value at the moment only depends on the state at the moment, which ensures that all the observation values in the sequence are independent of each other. In fact, the data sequence is not completely represented as a set of independent units. When the data elements in the sequence have long distance dependence, it is more appropriate to allow such long distance dependence and make the observation sequence can be expressed as independent cross features. As an undirected graph model or markov random field, conditional random field is a statistical model used to mark and shard serialized data. The model is to calculate the joint probability of the whole marker sequence under the condition of given observation sequence requiring marking, rather than to define the distribution of the next state under the condition of given current state. Marking the distribution condition attribute of a sequence can make it possible to well match the real data, and in these data, the conditional probability of marking a sequence relies on the non-independent and interactive features of the observed sequence, and expresses the importance of the features by assigning different weights to the features.

2.3. Circular Neural Network Based on Conditional Random Field

Circulatory neural network can remember the relationship of input information through time series. However, in practical applications, the front and end of time series may be far away, and problems will arise when two distant data in time series need to be integrated. In theory, circular neural network can not be restricted by the information distance in the sequence, but in practical application, it is found that when the information distance is too far, the calculation may not converge, and the correct result cannot be obtained. In order to solve the problem of dealing with the long-distance information relationship in sequences, the long and short term memory network was proposed. The long and short term memory network is improved on the basis of circular neural network, and it has achieved good results in the processing of time relation of many long series input, and has been widely used.

ISSN: 2472-3703 DOI: 10.6911/WSRJ.201907_5(7).0006

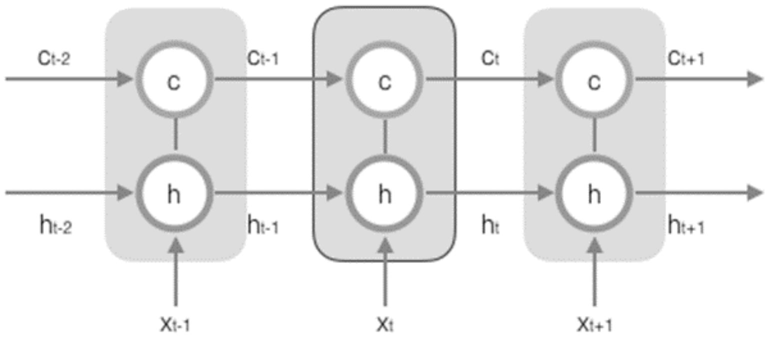


Fig 1. LSTM hidden layer expanded by time dimension

3. THE CRF - RNN METHOD

The traditional method of automatic summarization USES the sentence-level features and word-level features of the document to calculate the weight of the sentence as the abstract sentence, so as to generate the document abstract. Later, automatic abstracting based on circular neural networks was a breakthrough, which took advantage of the shallow semantic features.

3.1. Probability Graph Model Based on RNN

GRU(Gated Recurrent Unit) is a variation of LSTM, which simplifies the LSTM structure, but it does not affect the model effect. A specific GRU unit USES an update gate to replace the forgetting gate and the input gate of the LSTM unit, and does not distinguish between the hidden state and the memory unit state. That is, the linear self-update is not built on the additional memory unit, but the linear accumulation is built on the hidden state directly, and the control gate is used to regulate.

"Forget gate" : determines how much of the previous unit state ct-1 remains in the current ct.

"Output" (the output gate) : "forget the door" and "input" can control unit state the content of c, "output" long-term memory can control the current output, the influence of decision unit state ct how much output to the current output value of the ht both short-term and long-term memory network, which can determine the next set of neurons from the neurons of access to information. Finally, the output gate and cell state are used to obtain the final output of the long and short term memory network.

$$\tilde{c}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (1)$$

$$c_t = f_t \otimes c_{t-1} + i_t \otimes \tilde{c}_t \quad (2)$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (3)$$

$$h_t = o_t \otimes \tanh(c_t) \quad (4)$$

3.2. The Specific Implementation

The pseudo-code of the algorithm based on conditional random field and neural network fusion is described as follows:

The pseudo-code of the algorithm based on conditional random field and neural network fusion is described as follows:

Step 1. Initialize the parameter vector to zero

ISSN: 2472-3703
DOI: 10.6911/WSRJ.201907_5(7).0006

Step 2. Read the Chinese word segmentation training sentence in the corpus

Step 3. The characteristic string is generated according to the characteristic template and the characters of each current position and stored in the characteristic dictionary to get the characteristic index of the corresponding characteristic string of the character in the dictionary. If the training corpus is not finished, go to step 2, otherwise go to step 4

Step 4. The probability of node and state transition is calculated according to parameter vector and character index of words in each sentence

Step 5. Calculate the forward-backward vector, and then calculate the value of the gradient vector and the logarithmic likelihood function. If there are still sentences without the gradient vector, go to step 4; otherwise, go to step 6

Step 6. Add the penalty factor, and then use l-bfgs algorithm to make the optimal estimation of the parameter vector, and the obtained parameter vector is used to calculate the next node, the transfer probability and the penalty factor. If the algorithm converges or reaches the maximum number of training times, it is finished, otherwise go to step 4

```
Node n = root;
        String gbk = GbkMap.INSTANCE.getGbk(words);
        for (int i = 0; i < gbk.length(); i++) {
            int pos = (gbk.charAt(i) - 48);
            if (pos > 22) {
                return false;
            }
            if (n.getNext().elementAt(pos) == null) {
                return false;
            } else {
                n = n.getNext().elementAt(pos);
            }
        }
        return n.isWord();
}
```

4. EXPERIMENTAL EVALUATION

The Chinese coding order in conditional random field has an influence on the space sparsity of even number groups, and the frequency of Chinese characters in Chinese character set is used as the basis for recoding. This method puts forward a research direction to solve the spatial sparse problem of even number groups. This method is applied in this paper and combined with the method of constructing the node with the largest number of children first. The experimental results show the spatial sparse problem of even number groups.

That's a good problem to solve. To further reduce the space sparse problem, based on the above mentioned coding method, put forward the serial number of features in the template information preference code, and make the serial number coding method of continuous, the purpose of this method is to reduce the number set is constructed for the first time, the minimum characteristics of serial number coding of the double array subscript before waste of space. This method can effectively reduce the space sparsity problem of the double-byte query structure.

ISSN: 2472-3703 DOI: 10.6911/WSRJ.201907_5(7).0006

4.1. Structure Diagram of the Experimental System

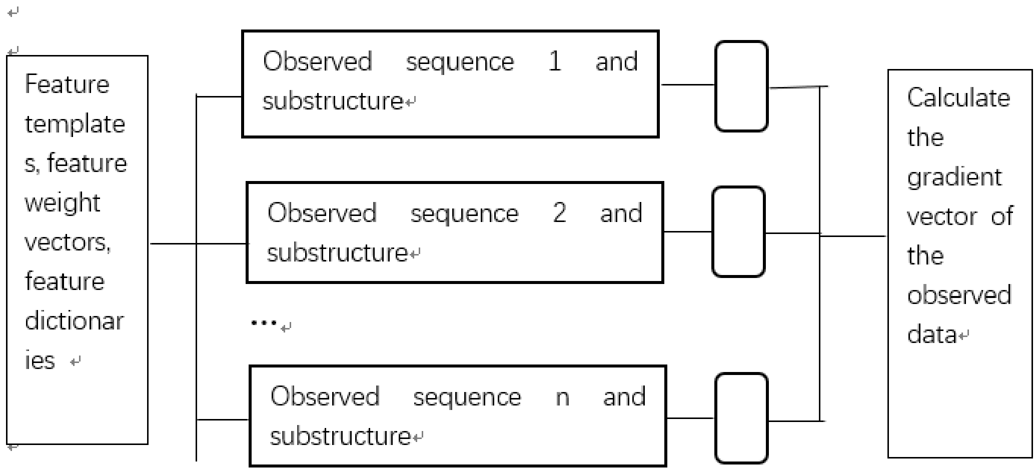


Fig 2. Flow chart of conditional random field experiment

4.2. Comparison of Cross Models

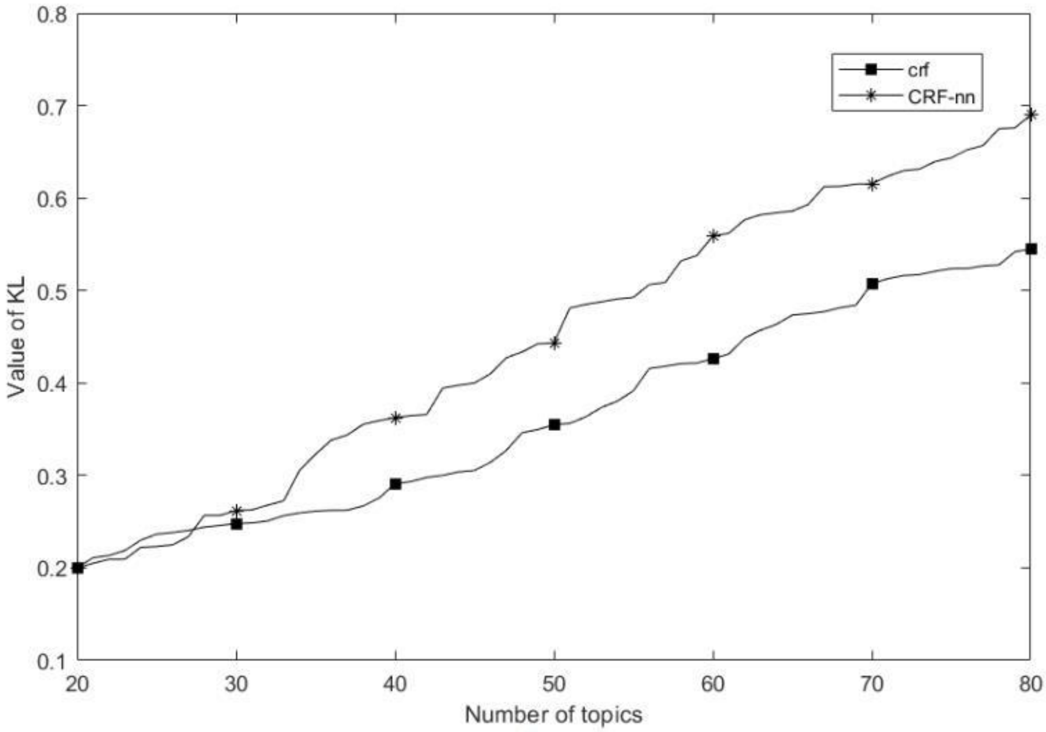
In the experiment, the prior distribution parameter of the theme model was set to 50K, K +, and K was the theme number. Beta takes an empirical value of 1.01. The value of K was determined by pretreatment experiments. Shows the su-4 values for different topics. This paper finally determines that K value is 50 as the number of topics. In the experiment, the Greedy Algorithm is used to adjust the parameters on [0,1] by using 0.1 as the stepping length, and the optimal parameters are eventually obtained . As there are few characteristic parameters and the final cost value is constrained by KL value, which is calculated by the parameters of unsupervised learning method, the dependence of parameters on data is not obvious. In addition, because the named entity features depend on the performance of the named entity recognition system, due to the limitation of the recognition accuracy, we initially give it a small value, later will do in-depth research on the named entity features.

Test corpus k value	HMM	CRF	CRF-RNN
K=5	89.78%	90.88%	91.12%
K=10	88.28%	89.16%	89.92%
K=20	88.64%	88.75%	89.72%

Fig 3. Comparison of experimental results

4.3. Experimental Results and Analysis

In this paper, the Chinese word segmentation algorithm based on crf neural network is implemented, and the Chinese word segmentation algorithm based on lstm neural network model is compared with the traditional Chinese word segmentation algorithm based on conditional random field (crf).



5. CONCLUSION

This paper proposes a participle model based on CRF-RNN. Chinese word formation and configuration are linguistic features realized by constantly affixing different endings after the word stem, and according to the statistics of the length and frequency of Chinese words,The characteristic template suitable for Chinese part of speech tagging is designed. This method can make full use of Chinese context information, avoid the problem of mark bias, get the global optimal result of part of speech tagging, and effectively improve the performance of part of speech tagging.

REFERENCES

[1] Ketkar N. Introduction to PyTorch[J]. 2017.

[2] Hochreiter S, Schmidhuber J. Long short-term memory. Neural Computation,1997,9(8): 1735-1780.

[3] Hu B, Lu Z, Li H, et al. Convolutional neural network architectures for matching natural language sentences[C]. Advances in Neural Information Processing Systems, 2014: 2042-2050.

[4] ZHANG S, ZHANG S, CHEN X, et al. Analysis and research of cloud computing system instance[C]// International Conference on Future Networks, Washington, 2010:88-92.

[5] HINTON G E, SALAKHUTDINOV R R. Reducing the dimensionality of data with neural networks[J]. Science,2006,7(28):504-507.

[6] BENGIO Y, SCHWENK H, SENÉCAL J, et al. Neural probabilistic language models[J]. Journal of Machine Learning Research, 2003, 3(6):1137-1155.

[7] WANG H, CAN D, KAZEMZADEH A, et al. A system for real-time Twitter sentiment analysis of 2012 U.S. Presidential election cycle[C]// ACL 2012 System Demonstrations, Jeju Island, Korea,2012:115-120.

[8] MIKOLOV T, SUTSKEVER I, CHEN K, et al. Distributed representations of words and phrases and their compositionality[C]// Proc of the NIPS, Lake Tahoe, 2013:3111-3119.

ISSN: 2472-3703 DOI: 10.6911/WSRJ.201907_5(7).0006

[9] KIM Y. Convolutional neural networks for sentence classification[C]//Proc of the EMNL, Belgium, 2014:87-93.

[10] BAHDANAU D, CHO K, BENGIO Y. Neural machine translation by jointly learning to align and translate [J]. Computer Science, 2014

[11] Hochreiter S, Schmidhuber J. Long short-term memory[J]. Neural computation, 1997, 9(8): 1735-1780.

[12] Ward R, Wu X, Bottou L. AdaGrad stepsizes: Sharp convergence over nonconvex landscapes, from any initialization[J]. 2018.

[13] Mukkamala M C, Hein M. Variants of RMSProp and Adagrad with Logarithmic Regret Bounds[J]. 2017.

[14] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks[C]. International Conference on Neural Information Processing Systems. Curran Associates Inc. 2012:1097-1105.

[15] Yang J, Zhang D, Frangi A F, et al. Two-dimensional PCA: a new approach to appearance-based face representation and recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2004, 26(1):131-137.

[16] Graves A, Schmidhuber J. Framewise phoneme classification with bidirectional LSTM and other neural network architectures[J]. Neural Networks, 2005, 18(5-6):602-610.

[17] Zhang X, Zhao J, Lecun Y. Character-level Convolutional Networks for Text Classification[J]. 2015,27(6):649-657.

[18] ZHANG S, ZHANG S, CHEN X, et al. Analysis and research of cloud computing system instance[C]// International Conference on Future Networks, Washington, 2010:88-92.

[19] HINTON G E, SALAKHUTDINOV R R. Reducing the dimensionality of data with neural networks[J]. Science,2006,7(28):504-507.

[20] BENGIO Y, SCHWENK H, SENÉCAL J, et al. Neural probabilistic language models[J]. Journal of Machine Learning Research, 2003, 3(6):1137-1155.

[21] WANG H, CAN D, KAZEMZADEH A, et al. A system for real-time Twitter sentiment analysis of 2012 U.S. Presidential election cycle[C]// ACL 2012 System Demonstrations, Jeju Island, Korea,2012:115-120.

[22] MIKOLOV T, SUTSKEVER I, CHEN K, et al. Distributed representations of words and phrases and their compositionality[C]// Proc of the NIPS, Lake Tahoe, 2013:3111-3119.

[23] KIM Y. Convolutional neural networks for sentence classification[C]//Proc of the EMNL, Belgium, 2014:87-93.

[24] BAHDANAU D, CHO K, BENGIO Y. Neural machine translation by jointly learning to align and translate [J]. Computer Science, 2014.