International Journal of Hybrid Information Technology
Vol. 9, No.7 (2016), pp. 237-248
http://dx.doi.org/10.14257/ijhit.2016.9.7.22

Clinical Relation Extraction with Deep Learning

Xinbo Lv[1],Yi Guan[*1], Jinfeng Yang[2] and Jiawei Wu[1]

[1]*School of Computer Science and Technology. Harbin Institute of Technology, Harbin, 150001, P.R. China*
[2]*School of Software. Harbin University of Science and Technology, Harbin, 150080, P.R. China*
lv_xinbo@hotmail.com, guanyi@hit.edu.cn, yangjinfeng@hrbust.edu.cn, 710167251@qq.com

Abstract

Relations between medical concepts convey meaningful medical knowledge and patients' health information. Relation extraction on Clinical texts is an important task of information extraction in clinical domain, and is the key step of building medical knowledge graph. In this research, the task of relation extraction is based on the task of concept recognition and is implemented as relation classification by the adoption of a CRF model. The proposed CRF-powered classification model depends on features of context of concepts. To remedy the problem of word sparsity, a deep learning model is applied for features optimization by the employment of auto encoder and sparsity limitation. The proposed model is validated on the data set of I2B2 2010. The experiments give the evidence that the proposed model is effective and the method of features optimization with the deep learning model shows the great potential.

Keywords*: relation extraction, clinical narrative, deep learning, auto encoder, sparsity limitation*

1. Introduction

An electronic medical record (EMR) is a repository for patient information within one health-care enterprise (e.g. within one hospital, author's note) that is supported by direct computer input and integrated with other information sources. EMR data is typically organized in tabs with elements such as problem, medication and allergy lists, and visits organized according to history, physical examination, assessment, and plan, and is stored as narratives, structural data, and images[1]. EMR data contains important medical information[2], and they should not only be used for the purpose for which they were collected, that is vander Lei's first law of medical informatics[3]. They can and should be used to generate new knowledge about the distribution and determinants of disease with technologies of information extraction or text mining, and also be used further for clinical decision supports[4] or evidence based medicine[5].

Clinical narrative is very important, concerns of facilitation, understandability and the improvement of presentation contribute the prevalence of narratives in EMR. Much clinical knowledge, including specifically clinical concepts and their relations, are buried in those semi-structural texts. Concept extraction, *i.e*. Named Entity Recognition (NER), is a sub-field of information extraction and refers to the task of recognizing expressions denoting entities, such as diseases, drugs, and tests, in free text documents[4]. Relation extraction builds on entity extraction and is the next step in the endeavor to create a structured representation of the contents of unstructured narratives. Concept extraction and relation extraction mainly apply statistics machine learning model, and there is no a priori reason to think that the

ISSN: 1738-9968 IJHIT

techniques tailored to either biomedical or to clinical text would not be useful (perhaps with modification) in the realm of clinical research narratives[6]. Therefore, it is important to study and develop models and systems for concept extraction and relation extraction in clinical domain. Clinical concept extraction and coding has been globally explored, however clinical relation extraction receives much less focuses. This research aims to investigate this task, specifically with deep learning. In this research, experiments are conducted on the data set of 2010 i2b2/VA relation challenge, and experimental results give the evidence of the effectiveness of the proposed method.

2. Related Works

Relation extraction is an important task of information extraction (IE) in the open domain, so as in the medical domain. Most research on relation extraction focus the binary relation, *i.e*. relation between two elements. This task normally is preceded by Named Entity Recognition (NER), and aims to classify the relation type between two entities or concepts. Concept relation can be formally described as $R = r(e_1, e_2)$, in which r is the relation type, e_1 is the first concept, and e_2 is the second concept. Relation extraction targets to recognize the pre-defined relation type between two concepts within one sentence, regardless of different sentences[7]–[9]. Relation extraction is implemented as to predict the most possible relation among some candidates as per context features of the two concepts. Therefore, relation extraction generally applies a certain classification model[8], such as CRF (Conditional Random Field), SVM (Support Vector Machine) and ME (Maximum Entropy). In clinical domain, Uzuner is the first to investigate the research of relation extraction between medical concepts, and defined six main categories of relation types in detail[10]. In the challenge of I2B2 2010, the relation scheme comprises three main types of relation, that is problem-test, problem-treatment, and problem-problem[7]. With Uzuner's relation scheme, concepts and their relation extracted from a medical record can be used as a skeleton of the record. Research on clinical relation extraction mainly employ classification model based on machine learning. OanaFrunza *etc*. [11]investigated three types of relations on the Medline[1] data set with three classification methods, and results showed that the all-for-one method outperformed others. In OanaFrunza's research, features include not only words of context but also semantic types of UMLS. Uzuner *etc*. [10]trained six classifiers based on SVM, and features for the classifiers comprise the order and distance between two concepts, words of concept context and link grammar[12] of sentences. Experiments showed that words of context played the key role in the classifiers. Uzuner's research inspired and motived many other works on the clinical relation extraction. Bryan Rink *etc*. [13] also applied SVM for the task of relation challenge of I2B2 2010. In their research, several tools and resource, such as GENIA[2], Wikipedia[3], WordNet[4], General Inquirer, were also used to help feature engineering. Results showed that words of context and medical knowledge were the great help to the relation classification model. However, a concept pair with few contexts may fail to be properly classified. To remedy this limitation, Demner-Fushman *etc*.[14] incorporated the semantic relations in UMLS into the classification model. Berry de Bruijn *etc*. [15]applied ME for their implementation of a semi-supervised classification model and used cTAKES (clinical Text Analysis and Knowledge Extraction System) [16]to parse sentences of clinical narratives and extract

[1] http://medline.cos.com/
[2] http://www.nactem.ac.uk/GENIA/tagger/
[3] http://www.wikipedia.org
[4] http://wordnet.princeton.edu/

dependency relation as features. Different types of relations may not be balance in clinical text, and machine learning based models may perform well on majority-class relations and perform poorly on minor-class relations. As per this problem, Ryan *etc*.[17] analyzed the bias of relation types and proposed a novel technique for weakly-supervised bootstrapping of a classifier for this task. According to the reviewed researched on the clinical relation classification task, a machine learning based model with features of concept context is an effective method and is generally focused by researchers. However, words in many machine learning models are represented as one-hot vectors, and the problem of sparsity has become an important bottleneck of models. This research aims to remedy the problem of sparsity of word features by applying a deep learning model for feature optimization.

3. Relation Scheme and Data Sets

The data sets adopted in this research were from 2010 i2b2/VA relation challenge. In the data sets, three types of concepts and their relations are defined in the task of this research, *i.e*. medical problems, tests and treatments. These defined concepts should be only complete noun phrases (NPs) and adjective phrases (APs) in clinic narratives, such as discharge summaries or progress notes. Medical Problems refer phrases that contain observations made by patients or clinicians about the patient's body or mind that are thought to be abnormal or caused by a disease. Treatments refer phrases that describe procedures, interventions, and substances given to a patient in an effort to resolve a medical problem. Tests stand for phrases that describe procedures, panels, and measures that are done to a patient or a body fluid or sample in order to discover, rule out, or find more information about a medical problem. Relation extraction aims to determine the type of relationship that exists between two concepts in the text from a sentence[10], therefor relations are bounded by sentences. Relations build on the medical problem, treatment, and test concepts that have already been identified. Definitely, this task is to further identify how problems relate to treatments, tests, and other medical problems in the text. Three major types of relations are defined, that is medical problems and treatments, medical problems and tests, medical problems and other medical problems. These defined concepts and relation can serve as a summary of a problem-oriented clinical record[10]. Relation scheme is summarized as follows.

I. Medical problems and treatment relations:
  a. Treatment improves medical problem (TrIP)
  b. Treatment worsens medical problem (TrWP).
  c. Treatment causes medical problem (TrCP).
  d. Treatment is administered for medical problem (TrAP).
  e. Treatment is not administered because of medical problem (TrNAP).
  f. Treatments and problems that are in the same sentence, but do not fit into one of the above defined relationships are not assigned a treatment-problem relationship.

II. Test relations and medical problems:
  a. Test reveals medical problem (TeRP).
  b. Test conducted to investigate medical problem (TeCP).
  c. Tests and problems that are in the same sentence, but do not fit into one of the above defined relationships are not assigned a test-problem relationship.

III. Medical problem and other medical problems:
  a. Medical problem indicates medical problem (PIP).

b. Pairs of medical problems that are in the same sentence, but do not fit into one of the above defined relationships are not assigned a problem-problem relationship.

Clinical narratives in the data sets comprise discharge summaries contributed by Partners Health care, Beth Israel Deaconess Medical Center, and the University of Pittsburgh Medical Center, and progress notes contributed by the University of Pittsburgh Medical Center. A total of 394 training reports, 477 test reports, and 877 unannotated reports were de-identified and released to challenge participants with data use agreements. Details about 2010 i2b2/VA relation challenge can be referred by [7], [10], and the relation scheme and annotation specification can be accessed at the site of I2B2.

4. Methodologies

Relation extraction aims to determine the type of relationship that exists between two concepts in the text from a sentence. Therefore, relation extraction models can be implemented as classifiers, usually based on supervised machine learning. These models view the type of relationship as a predefined tag, and predict the tag of the relationship according to its context features, including the type of concepts, word and POS of the context, *etc*. Broadly adopted classification models include CRF, SVM and ME. Many models implemented with CRF have achieved excellent performance on multiple NLP tasks; therefore the proposed model is designed based on CRF. Specifically, CRF++ is adopted, which is the mostly applied implementation of CRF model.

Features are key importance of a statistic machine learning model. Feature engineering including feature design, feature selection, feature optimization and learning has obtained more and more focuses. Deep learning has achieved state-of-the-art performance in many NLP tasks. In Most NLP tasks, deep learning is applied for feature learning, especially learning word embeddings, which are dense and low dimension vector representations of words. In this research, deep learning is also applied for learning vector representation of words in the context of concept relations. In order to extract features, clinical text should be preprocessed by normalization, removing stop words, and POS tagging.

4.1. Preprocessing and Features Extraction

Clinical text is very different from open domain text, which poses a special challenge to NLP in clinical text. Those unique characteristics includes: (1) many sentences are ungrammatical and composed of short, telegraphic phrases; (2) shorthand of words are widely used, such as abbreviations, acronyms, and local dialectal shorthand phrases; (3) misspellings abound in clinical texts exist; (4) clinical narratives can contain any characters, *i.e*. numbers, symbols, *etc*. Meystre has surveyed the main features of clinical[6]. Therefore, preprocessing on clinical text plays an important role in the task of information extraction. With respect to Upper/lower case and different morphologies of words, a normalization technology is adopted. In this research, luiNorm[5] is applied for word normalization, which is based on UMLS (Unified Medical Language System)[18]. For shorthand of words, metamap[6] is applied for mapping complete medical concepts from the large medical data base of UMLS. Furthermore, POS (part of speech) of words are very useful

[5] http://lexsrv3.nlm.nih.gov/LexSysGroup/Projects/lvg/2014/web/index.html
[6] https://metamap.nlm.nih.gov/

information for the method of information extraction. GENIA[7] is used to tag POS of words automatically.

Context features of relationships are selected and extracted after preprocessing on clinical text. Features are categorized to four types: words of context, POS of words, concept type, and distance between two concepts. These features are listed in table 1.

Table 1. Feature Template for the CRF Model

Type of feature	Description of feature
Word	Words of a concept
Word	Two words before a concept
Word	Two words after a concept
POS	POS of words of a concept
POS	POS of two words before a concept
POS	POS of two words after a concept
Distance	The gap of two concepts
Concept type	Type of a concept

.

4.2. Features Optimization with Deep Learning

Words indicate important context information, and express somewhat pattern of relations. Words are used as features, also called BOW (bag of words). However, words are represented one-hot vector, and lead to sparsity problem in classification model. In order to counter this problem, a deep learning model is adopted to learn and optimize new word feature.

Deep learning is built on neural network method. Firstly, deep learning is a feature learning method with the purpose of representing data using multiple neural networks, which makes it learning method without surveillance by selecting a suitable feature according to the input data. Compared with local description, deep learning uses the data more effective by distribution of the data, which also shows that less parameters can divide more subspaces. As shown in Figure 1, the left one is locally described, where the numbers of features and subspaces are the same. However, distribution feature can separate the same subspaces with only several features. The discriminated feature shows more effective ability to divide data space.

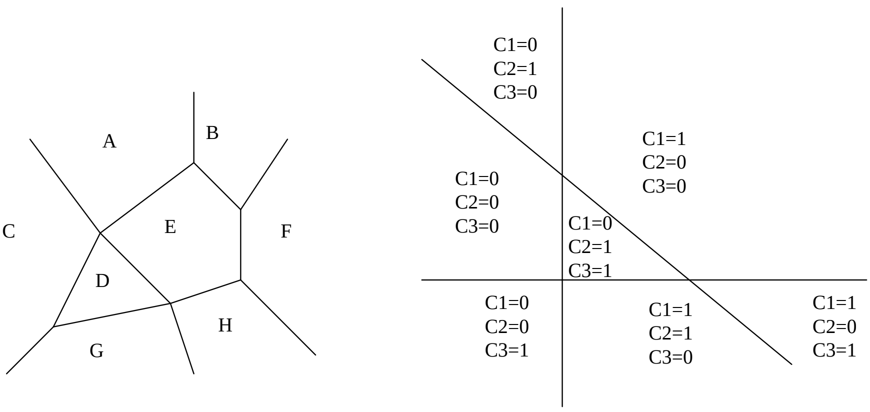

Figure 1. Local Representation and Distributed Representation of Data

[7] http://www.nactem.ac.uk/GENIA/tagger/

In surveillance neural network, the nodes in each hidden layer is determined by nodes activated in last layer, which shows part of the complicated results and preceding information of nodes in last layer and even earlier layers. Such structure can accomplish a sophisticated function approximation. From a different view, the nodes among multiple layers can be considered as a change of description. The denotation of nodes is obtained with the combination of various nodes in last layer, which is referred to as a multiple nonlinear mapping structure. Thus, another expression of the data is obtained and the data can be denoted by multiple layers.

Compared with neural network algorithm, the features of deep learning are concluded in two aspects. Deep learning uses neural network to denote the learning process, and each node represents an element of a new description, while in neural network nodes are considered as calculation nodes. On the other hand, neural network algorithm is based on the adjustment of error from predicted results in forward computation and standard results. Contrarily, deep learning is based on data pre-training process without surveillance and shows data regeneration in layer-by-layer, which can used to avoid local minimum and overfitting to some extent.

4.2.1. Deep Auto Encoder

Auto encoder is a basic structure used for deep learning. Deep learning uses continuous layer-wise learning to obtain data expression in distinct layers. As for untagged data, another expression is obtained in the data. This method can use original data as input and auto encoded in single layer, then compare with original data and output data to finish backward propagation process of error. The basic structure is shown in Figure 2.

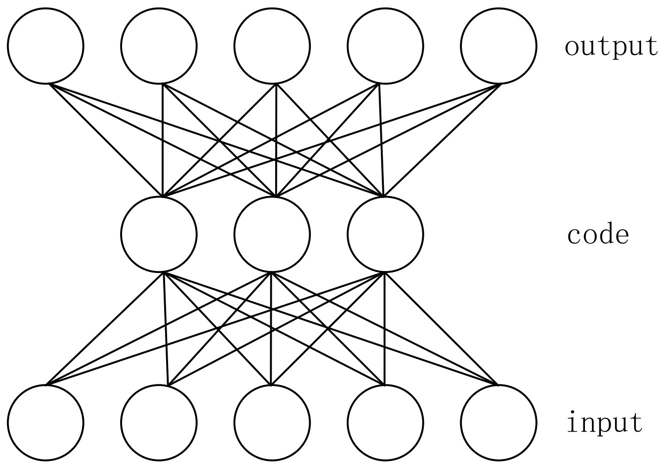

Figure 2. Architecture of Auto Encoder

From the structure in Figure 2, hidden nodes are calculated by networks from input data. The hidden nodes can be represented as the original input through calculation in the next layer. Then the hidden layer is considered as a description of the original input data. As for the input vector $V=(v_1,v_3,...,v_m)$, and the randomly initialized neural network weight matrix W . The second hidden nodes are calculated by the first layer, formulated as Equation (1).

$$h_{1,j}=\sigma(\sum_i W_{ij}^1 V_i + b^{(1)}) \qquad (1)$$

where, $\sigma(z)=\frac{1}{1+e^{-z}}$, as the activation function of nodes, mapping the data to the internal (0,1), is also another description of neural nodes.

A data transformation is obtained in each calculation process. When output data fits the input data, the hidden nodes are called a new description of data. Then, delete the output layer, and use the obtained hidden layer as a new input layer for next training. The process is the deep auto encode layer-wise training process, which can be obtained new description of original data and each layer is more advanced than last layers.

After layer-wise training, one basic search start point is obtained for assumed space. Then, a minor adjustment is preceded by output tagged data. As for the entire network weight matrix, the error is sent back to the network using backward propagation algorithm. Since the distribution expression of data is obtained for the hidden layers training, the so-called adjustment is minor and determines the major direction.

Auto encode is calculated using neural network, and attempt to study a function $h_{w,b}(x)\approx x$, which is also an equation function. The expression can constrain the entire network structure in many aspects, such as controlling the number of hidden layer nodes, limiting the activation proportion of hidden nodes and so on. Some improvement is made in the description of the data using such data transformation. For example, the process of generating hidden layer nodes shows each node in hidden layer is decided by linear combination of all nodes in last layer through weight matrix and then realizes the nonlinear transformation of activation function. From the view of feature, this process can be also considered as the combination of the feature in last layer. On the other hand, if the number of nodes limiting the hidden layers is small, a reasonable compression is preceded in the description of the data. Using this method, the noise of the data is to some extent reduced. Finally, the major directions in the data are remained after the entire process, which amounts to the projective result in data space with the largest variance. Similar to the Principal Component Analysis (PCA), auto encoder can add or limit the hidden nodes in nonlinear transformation.

4.2.2. Sparsity Limitation

In the process of auto encoder study, the objective is to keep the description of the output and original input unchanged, as shown in fig2. Thus, the loss function is defined as Equation (2).

$$J(w,b)=\frac{1}{2}\left\|h_{w,b}(x)-x\right\|^2 \qquad (2)$$

Using the backward propagation algorithm with loss function, the adjustment value of error is sent to the preceding weight matrix. As for the feature expression, the recognition of the partial information is more easily identified with sparser features. When the node value is set to 1, the node is considered as being activated, or depressed. That’s the feature of the node, and the limitation in one layer with most nodes being depressed is called sparse limitation. The sparse feature means that only a little feature are activated. Due to the limitation of sparsity, the feature to discriminate is maintained, while the classification of feature is enhanced greatly. Thus, the loss function is redefined as Equation (3)

$$J(w,b)=\frac{1}{2}\left\|h(x)-x\right\|^2+\frac{\lambda}{2}\sum_{i,j}W_{ij}^2 \qquad (3)$$

The rest of the training method is the same as in the basic deep auto encoder learning method. Specifically, using feature learning layer-wise, hidden layers become another description of the original input, and after determining the entire multiple network layers, the final model is obtained with tagged data by back propagation method.

4.2.3. Features Optimization with Deep Learning

Features extracted following the template in table 1 comprise two groups, one is a group of word features, and the other is a group of features including others. Features of context words suffer the problem of sparsity for their one-hot representations in classical machine learning models. To remedy the sparsity of words' representations, a deep learning model is applied for learning somewhat abstract or general words' representations. In the data set, high frequent words may be less discriminative, which called stop words and to be removed. In the experiments, the highest 1.5% frequent words are removed.

Let $D=\{w_1,w_2,...,w_i,...,w_n\}$ represents the set of all words of contexts, and each word is represented as a one-hot vector v_i. For a given concept E_k, P_k denotes the set of words in the context of E_k. To represent the word context of E_k, each vector of a word in is added, *i.e.* $c_k = \sum_{w_j \in P_k} v_j$. With the input of c_k, the deep learning model is expected to product more abstract representations of the context of concepts. Finally, the outcomes, the optimized features of word context of concepts are applied for the relation classification model. For comparisons, the original word features of context of concepts are also used to train a baseline model.

5. Results and Analysis

In the experimental stage, 177 training documents, comprising 3120 pairs of relations, and 259 testing documents, comprising 6293 pairs of relations, are used to train and validate classification models. The evaluation metrics are precision, recall and F-measure, calculated as Equation (4), (5) and (6)

$$precision(P)=\frac{TP}{TP+FP} \qquad (4)$$

$$recall(R)=\frac{TP}{TP+FN} \qquad (5)$$

$$F-measure(F)=\frac{2\times P\times R}{P+R} \qquad (6)$$

TP denotes the count of properly classified relations, FP denotes the count of relations which are classified to the target type but actually not, FN denotes the count of relations which are not classified to the target type but actually true.

For comparison, three models are trained and evaluated, one is based on the original word features and is named baseline, and the other two are based on the optimized word features by a deep learning model and is named deepAE (deepAE means the deep learning model is implemented with auto encoder) and deepSAE (deepSAE means the deep learning model is implemented with both auto encoder and sparsity limitation). Three groups of experimental results are listed in the following table 2 to table 4, in which the maximum values of corresponding metrics are specified in bold style.

As per comparisons between the results of the baseline and of the deep learning based model, the proposed model deepSAE perform better than the baseline, especially at the minor types, *i.e.* TrIP, TrNAP, *etc*. These achievements give the evidence that deep learning based model can remedy the problem of sparsity effectively. Furthermore, sparsity limitation is an effective constraint condition during the process of feature optimization. However, it is also can be seen that there is not big difference between the model and the baseline. The superiority with the deep learning model is subtle on this research largely because the data set is too small to feed the deep learning.

Table 2. Results of Baseline

Relation types	P(%)	R(%)	F(%)
TrAP	76.2	**97.2**	85.4
TeRP	89.4	**98.2**	**93.6**
TrIP	**58.3**	9.2	15.9
TrCP	**67.3**	39.2	49.5
TrWP	0	0	0
TrNAP	**64.3**	8.0	14.3
PIP	**100.0**	**100.0**	**100.0**
TeCP	72.8	29.3	41.8

Table 3. Results of Deep Auto Encoder Based Model

Relation types	P(%)	R(%)	F(%)
TrAP	77.2	90.1	83.1
TeRP	89.2	94.1	91.6
TrIP	50.9	17.8	**26.3**
TrCP	46.8	**47.4**	47.1
TrWP	42.9	**2.7**	**5.2**
TrNAP	48.4	13.4	21.0
PIP	99.1	98.1	98.6
TeCP	56.8	38.2	45.6

Table 4. Results of Sparse Deep Auto Encoder Based Model

Relation types	P(%)	R(%)	F(%)
TrAP	**78.2**	95.6	**86.0**
TeRP	**90.6**	96.2	93.3
TrIP	44.3	**17.8**	25.4
TrCP	67.0	44.4	**53.4**
TrWP	**66.7**	1.8	3.6
TrNAP	43.6	**15.2**	**22.5**
PIP	**100.0**	**100.0**	**100.0**
TeCP	63.0	**39.3**	**48.5**

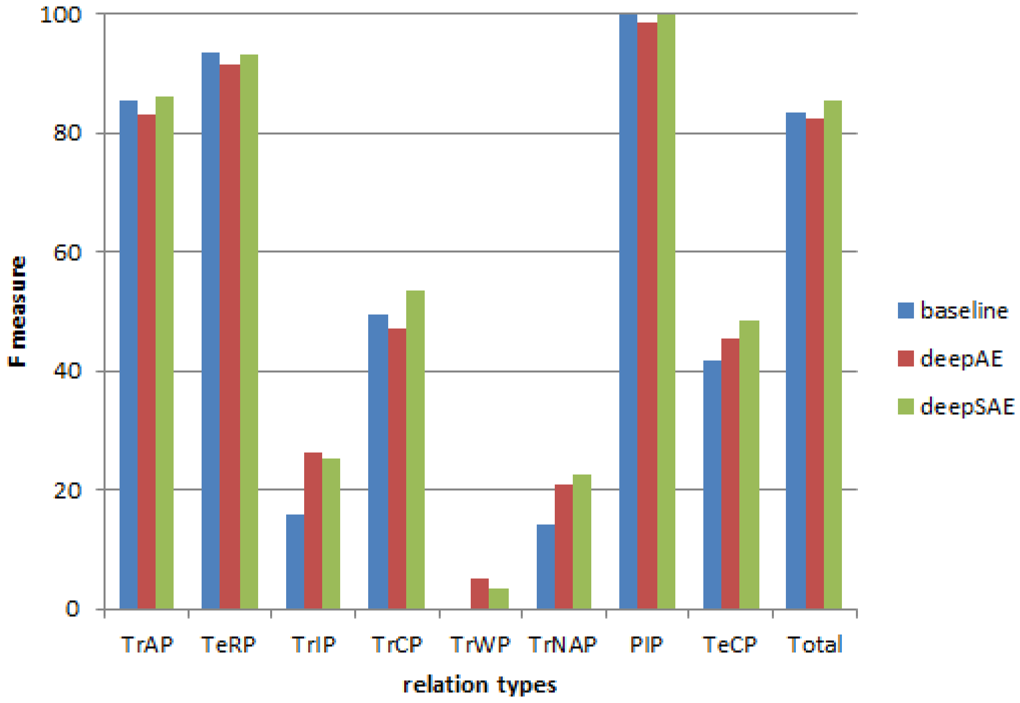

Figure 3. Compare Between Baseline and Improved Method

For the purpose of visual comparisons, F-measures of the three groups of results are illustrated in histograms shown in Figure 3, including the total evaluated results of all types of relations. The total F-measures of total relations have climbed over 80%. The state-of-art system in i2b2 2010 achieved the F-measure 0.737[7], and the proposed model has greatly outperformed the best system in i2b2 2010. Therefore, the adoption of CRF and context features optimization is suitable for the task of relation classification on clinical text.

6. Conclusions and Future Work

Narrative is an important type of data in electronic medical records, in which plenty of medical knowledge is buried. Therefore information extraction on clinical narratives, such as medical concept recognition and concept relation classification, is an important task, and has been focused by many researchers and companies. This research investigated the task of relationships classification on clinical narratives. In this research, a relation classification model is proposed by adopting a CRF model. Meanwhile, context features of concepts are extracted and are optimized by a deep learning model. Experiments were conducted, and results were compared with a baseline model and the state-of-art system in i2b2 2010. Those comparative experiments showed the effectiveness of the proposed model.

Although the model has obtained some effectiveness, it is subtle and has large space to get improved in the minor types of relationships. The first of all future work is to apply the deep learning model on a large scale data set for learning better word representations. Secondly, a medical concept may have a hypernym in a knowledge base, such as UMLS, MeSH, then a relation between hypernyms may be helpful for relation classification. Therefore the second future work is to incorporate UMLS or MeSH into the classification model.

References

[1] H. J. Tange, A. Hasman, P. F. De Vries Robbé and H. C. Schouten, “Medical narratives in electronic medical records”, Int. J. Med. Inform., vol. 46, no. 1, **(1997)**, pp. 7–29.
[2] R. C. Wasserman, “Electronic medical records (EMRs), epidemiology and epistemology: reflections on EMRs and future pediatric clinical research”, Acad. Pediatr., vol. 11, no. 4, **(2011)**, pp. 280–287.
[3] J. van der Lei, “Use and abuse of computer-stored medical records”, Methods Inf. Med., vol. 30, no. 2, **(1991)**, pp. 79–80.
[4] D. Demner-Fushman, W. W. Chapman and C. J. McDonald, “What can natural language processing do for clinical decision support?”, J. Biomed. Inform., vol. 42, no. 5, **(2009)**, pp. 760–772.
[5] J. Frankovich, C. A. Longhurst and S. M. Sutherland, “Evidence-based medicine in the EMR era”, N. Engl. J. Med., vol. 365, no. 19, **(2011)**, pp. 1758–9.
[6] S. M. Meystre, G. K. Savova, K. C. Kipper-Schuler and J. F. Hurdle, “Extracting information from textual documents in the electronic health record: a review of recent research”, Yearb. Med. Inform., vol. 47, no. Suppl 1, **(2008)**, pp. 128–144.
[7] Ö. Uzuner, B. R. South, S. Shen and S. L. DuVall, “2010 i2b2/VA challenge on concepts, assertions, and relations in clinical text”, J. Am. Med. Inform. Assoc., vol. 18, no. 5, **(2011)**, pp. 552–6.
[8] D. Nadeau and S. Sekine, “A survey of named entity recognition and classification”, Lingvisticae Investig., vol. 30, no. 1, **(2007)**, pp. 3–26.
[9] J. Chen, “Automatic relation extraction among named entities from text contents”, National University of Singapore, **(2006)**.
[10] O. Uzuner, J. Mailoa, R. Ryan and T. Sibanda, “Semantic relations for problem-oriented medical records.,” Artif. Intell. Med., vol. 50, no. 2, **(2010)**, pp. 63–73.
[11] O. Frunza and D. Inkpen, “Extraction of Disease-Treatment Semantic Relations from Biomedical Sentences”, Proc. 2010 Work. Biomed. Nat. Lang. Process., no. July, **(2010)**, pp. 91–98.
[12] D. D. K. Sleator and D. Temperley, “Parsing English with a Link Grammar”, Third Int. Work. Parsing Technol., vol. 64, no. October, **(1995)**, pp. 91.
[13] B. Rink, S. Harabagiu and K. Roberts, “Automatic extraction of relations between medical concepts in clinical texts”, J. Am. Med. Informatics Assoc., vol. 18, no. 5, **(2011)**, pp. 594–600.
[14] A. A. Demner-Fushman D, E Apostolova, R Islamaj Dogan, FM Lang, J Mork, A Neveol, S Shooshan, M Simpson, “NLM’s system description for the fourth i2b2/VA challenge”, in Proceedings of the 2010 i2b2/VA Workshop on Challenges in Natural Language Processing for Clinical Data, **(2010)**.
[15] B. De Bruijn, C. Cherry, S. Kiritchenko, J. Martin and X. Zhu, “Machine-learned solutions for three stages of clinical information extraction: the state of the art at i2b2 2010”, J. Am. Med. Informatics Assoc., vol. 18, no. 5, **(2011)**, pp. 557–562.
[16] G. K. Savova, J. J. Masanz, P. V Ogren, J. Zheng, S. Sohn, K. C. Kipper-Schuler, and C. G. Chute, “Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES): architecture, component evaluation and applications.,” J. Am. Med. Informatics Assoc., vol. 17, no. 5, pp. 507–513, **(2010)**.
[17] R. J. Ryan, “Groundtruth budgeting: a novel approach to semi-supervised relation extrac-tion of medical language”, MIT Master’s Thesis, **(2011)**.
[18] O. Bodenreider, “The Unified Medical Language System (UMLS): integrating biomedical terminology”, Nucleic Acids Res., vol. 32, no. Database issue, **(2004)**, pp. 267–270.

Authors

Xinbo Lv, he received the B.S. degree in Computer Science and Technology from Harbin Institute of Technology in 2006 and the M.S. degree in Computer Science and Technology from Harbin Institute of Technology in 2008. Now he is a Ph.D. candidate at School of computer science and technology in Harbin Institute of Technology, and his research interests cover natural language processing, machine learning and information extraction on electronic medical records.

