Journal of Information Hiding and Multimedia Signal Processing ©2017 ISSN 2073-4212
Ubiquitous International Volume **8**, Number **2**, March **2017**

Research on Entity Relation Extraction in TCM Acupuncture and Moxibustion Field

Shuihua Sun[1,2], Degen Huang[1], Jing Zhang[2,*]

1. School of Computer Science and Technology, Dalian University of Technology, Dalian 116024, China
2. College of Information Science and Engineering,
Fujian Provincial Key Laboratory of Big Data Mining and Applications,
Fujian University of Technology, Fuzhou, 350118, China
*corresponding author: jing165455@126.com

Received November 2016; revised January, 2017

Abstract. *For the context of entity relation instance in TCM acupuncture and moxibustion field, effective words, syntax and semantics features are chosen to combine into feature template, and the entity relation instances are vectorized. The classification models of entity relations in TCM acupuncture and moxibustion area are trained by the machine learning method which is based on support vector machine. The experimental results show that the feature template of entity relations in this thesis have excellent effects on the entity relation extraction in TCM acupuncture and moxibustion area. The F-measure of entity relation classification model of DM, HM, AM and DRM reached* $93.25\%, 87.19\%, 86.57\% and 84.57\%$ *respectively.*

Keywords: Entity in TCM acupuncture and moxibustion field; Relation extraction; Support vector machine(SVM); Feature selection

1. **Introduction.** Chinese acupuncture and moxibustion is a special health care technology based on Han culture. It includes acupuncture and moxibustion theory, acupoints, acupuncture techniques and related equipments. As a display of academic results and main carrier of acupuncture and moxibustion academic exchanges, TCM acupuncture and moxibustion literature has witnessed a rapid growth. Selecting effective knowledge from the abundance of acupuncture and moxibustion literature will be of great significance to establish the knowledge network, treat disease, stay healthy, develop traditional Chinese medicine, and speed up the process of standardization, modernization and internationalization of Chinese acupuncture and moxibustion.

Relation extraction[1-6] is the extraction of semantic relations among the entities from the massive text collection, and the transformation of original text information into structured information suitable for the database access. Relation extraction technology can be applied to automatic question answering, information retrieval and other fields. It has attracted widespread attention of the academia in recent years. At present, there are three main ways of entity semantic relation extraction: machine learning based on feature vector, machine learning based on kernel function and machine learning based on fusion of feature vector and kernel function. According to the pre-set feature templates, in Document [7-11] the instances of the entity relations in training set and test set are digitized to form multidimensional eigenvectors. Then, the support vector machine (SVM) and the maximum entropy (ME) are trained by the eigenvectors of the training set, and the obtained model is used to predict the entity relation category of the test set. The main research focus of this method is how to use the syntax, lexical and related features of the entity to form the most representative eigenvector of the relation. In Document [12-14], the syntactic tree and word sequence of sentences are treated as objects, and the similarity of the processed objects is calculated by defining the kernel functions of the tree kernel and the semantic sequence, thus avoiding the construction of high-latitude eigenvectors . In Document [15], eigenvector and tree kernel function are merged effectively. The kernel method can make up the defect that the eigenvector can not represent the structural feature. The eigenvector extends the tree kernel function to a large amount of data, and obtains a good relation extraction effect.

In the field of biomedicine, the task of relation extraction is to extract the semantic relations among different biomedical entities (diseases, drugs, genes, proteins, etc.) from the biomedical texts and express them in an understandable way. In Document [16], features including word, keyword, word distance between protein entities, and association path are combined into feature templates, and SVM statistical module is used to extract protein relations. In Document [17], the sentence information of the two entities and the local context information of the two entities are combined to extract the drug relations, and good results are obtained. In Document [18], six SVM classifiers were trained by using the order and distance of entities in the sentence, the lexical features, and the link syntax characteristics to identify the entity relations between diseases, symptoms, examinations and treatments. This research provides a reference for electronic medical record entity relation extraction.

This thesis aims to research on entity relation extraction in TCM acupuncture and moxibustion field. Scrap into the Chinese medical journals website (http://www.cqvip.com/) to get abstract information in the acupuncture and moxibustion literature, define the entity and entity relation type in the field, manually mark the entities and entity relations full of abstracts, so as to build Chinese acupuncture and moxibustion field corpus. Combine effective vocabulary, grammar and semantic features in the context of entity relation of TCM acupuncture and moxibustion into feature templates. The support vector machine statistical machine learning method is used to train the classification model of entity relationship in acupuncture and moxibustion field. The experimental results show that the suggested method has a good effect on the extraction of entity relations in TCM acupuncture and moxibustion field.

2. **Definition of types of entities and entity relations in TCM acupuncture and moxibustion.** Before semantic entity relation extraction, it is necessary to define the type of semantic relations to be extracted, and then predict more probable semantic relation type according to the context features of the two entities. The definitions of entity types in different areas are different, then the definitions of semantic relation types of entities are also different. In this thesis, five types of TCM acupuncture and moxibustion entities are predefined, so are four kinds of entity relation types. Named entity types and its examples of Chinese acupuncture and moxibustion field are illustrated in Table 1. Entity relation types and its examples of Chinese acupuncture and moxibustion are shown in Table 2.

3. **Entity relation extraction.** Entity relation extraction is usually transformed into classification. The key of the research is how to express the relation instance effectively and calculate the similarity among the relation instances. This thesis studies the extraction of entity relations in the field of TCM acupuncture and moxibustion based on eigenvectors. The specific steps are as follows: 1) Select the appropriate feature set and

Table 1. Named entity types and examples of TCM acupuncture and moxibustion field

Entity Type	Description	Annotation Example
Named Entity of Disease	Restrictedly indicate the specific TCM disease names, such as scapulohumeral periarthritis, coronary heart disease, cervical spondylosis.	<Disease>coronary heart disease< /Disease>
Named Entity of Health Care	Restrictedly indicate specific Chinese Medical terms relating to health care, for example, immune function, constitution, microcirculation, spleen and stomach function, blood stasis, erythrocyte sedimentation rate, etc.	<Health>immune function< /Health>
Named Entity of Disease Treatment and Health Care Method	Restrictedly indicate the name of a specific disease treatment or method of care, generally referred to as moxibustion or acupuncture terms, for example, traction, acupuncture, abdominal acupuncture, ear laser acupuncture, acupoint injection, electromagnetic waves, filiform needle acupuncture, electro-acupuncture.	<Method>scalp acupuncture< /Method>
Named Entity of Meridians	Restrictedly indicate specific human meridians and acupuncture points, for example, Shenmen point, lumbar Jiaji point, Gate of Life, Yao Yangguan, Shenshu, lower limb gallbladder meridian, bladder meridian and so on.	<Acupoint>Jiaji Points (Lumbar Segments)< /Acupoint>
Named Entity of Medication	Restrictedly indicate the name of specific drugs, for example, angelica injection, white mustard seed powder, citicoline, bee venom, compound salvia miltiorrhiza, coenzyme A and so on.	<Drug>Vitamin B12< /Drug>

express the entity relation instances as the eigenvector; 2) Use SVM to train the classification model of entity relation in the TCM acupuncture and moxibustion field, which can make the classification model obtain the discriminable effect with the given data; 3) Test and evaluate the classification models of the entity relations. The process of entity relation extraction in Chinese acupuncture and moxibustion field is shown in Figure 1.

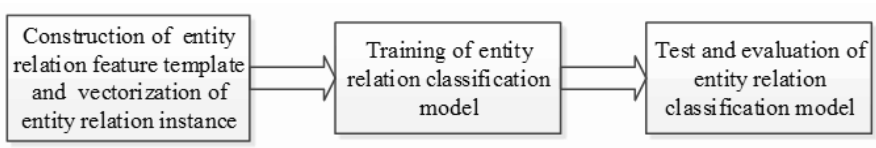

Figure 1. Entity relation extraction in TCM acupuncture and moxibustion

3.1. **Options of entity relation features.** This thesis only considers the relations between two entities within the scope of a sentence in the field of TCM acupuncture and moxibustion, and do not recognize the relations between the entities across sentences. The input of entity relation extraction system in TCM acupuncture and moxibustion is one sentence and two entities which have been marked in the sentence. The output is the semantic relation between the two entities. In order to capture the local and global features of the entity relations, to construct the eigenvector and improve the classification

Table 2. Entity relation types and its examples of Chinese acupuncture and moxibustion

Type of Entity Relationship	Description	Annotation Example
Disease-treatment $< Relation - DM >$	Relation$< Relation - DM >$obtains therapeutic relationship between two entities. It is used to indicate what disease could be treated in what way or what drugs or what acupuncture points.	This paper introduces the recent 10-years’ progress in the study of $< Relation - DM >< Method >$ acupuncture $< /Method >$ in the treatment of $< Disease >$ scapulohumeral periarthritis$< /Disease >< /Relation - DM >$.
Health-care $< Relation - HM >$	Relation$< Relation - HM >$obtains health care relations between two entities. It is used to indicate what role health care plays in what way or what drugs or what acupuncture points.	$< Relation - HM >$ Inject$< Drug >$compound danshen injection$< /Drug >$and$< Drug >$dushen injection$< /Drug >$ into acupoints to boost$< Health >$blood circulation$< /Health >$improve$< Health >$blood stasis$< /Health >$, nourish$< Health >$heart$< /Health >$invigorate$< Health >$pulse$< /Health >$and better$< Health >$qi$< /Health >$advance$< Health >$yang$< /Health >$function$< /Relation - HM >$.
Meridian points-treatment and health-care method$< Relation - AM >$	Relation $< Relation - AM >$decides the effective relationships of meridians and the treatment and health care methods between two entities. It is used to indicate what method in what acupuncture points and meridians have therapeutic or healthcare effects.	$< Relation - AM >< Acupoint >$ Governor vessels$< /Acupoint >< Method >$electro-acupuncture$< /Method >< /Relation - AM >$electric field treatment of $< Disease >$semi-transected spinal injury $< /Disease >$ in rats, is a simple, safe, effective treatment.
Medication-treatment and health-care method$< Relation - DRM >$	Relation $< Relation - DRM >$ obtains effective relations of medications and the therapeutic care approach between two entities. It is used to indicate the manner in which drugs are used for therapeutic or healthcare purposes.	In this paper, $< Relation - DRM >< Drug >$compound danshen injection$< /Drug >$and$< Drug >$dushen injection$< /Drug >< Method >$are injected$< /Method >< /Relation - DRM >$ to treat$< Disease >$Coronary heart disease$< /Disease >$ of 102 patients.

performance, we summed up the effective description of the lexical, syntactic and semantic features of entity relations in the field of TCM acupuncture and moxibustion based on the synthetical analysis of the context of the entity relation. The features include:

1) Features of entity types. The respective types of the two entities, which can be obtained from the annotated corpora in the TCM acupuncture and moxibustion field.

2) Features of all words in entities. The whole vocabulary of each entity.

3) Context features of entities. The first three words in the front and in back of each two entities, and parts of speech of these words.

4) Features of verbs. Take the verb closest to Entity 2. If there are two verbs meeting the condition, take the verb behind Entity 2.

5) Features of word distance. Refer to the number of words between two entity pairs which forms entity relation. With word distance ≥ 0.

6) Features of clauses. Whether the two entities are in a same clause, if in a same one they are 1, otherwise 0.

7) Features of interval entities. Whether there are other entities between the two entity pairs that constitute the entity relation, if there are any, the feature value is 1; if there is nothing, the feature value is 0.

In conclusion, as illustrated in Figure 2, we construct an entity relation combination feature template for the entity pairs (E1, E2) composed of any two entities in a sentence in the field of TCM acupuncture and moxibustion.

E1.TYPE, E2.TYPE, E1.TEXT, E2.TEXT,
Wi-3, Wi-2, Wi-1, Wi+1, Wi+2, Wi+3, Ti-3, Ti-2, Ti-1, Ti+1, Ti+2, Ti+3,
Wj-3, Wj-2, Wj-1,Wj+1, Wj+2, Wj+3, Tj-3, Tj-2, Tj-1, Tj+1, Tj+2,Tj+3,
Verb, Word_Distance, Clause, Interval

FIGURE 2. The Template of entity relation features

3.2. **Vectorization of Entity Relation Instance.** The vector space model is a model that transforms text features into numerical features. When the vector space model is used to extract the entity relations, the entity relation instance is given the specific eigenvalue according to the pre-set feature to form the multidimensional eigenvector. In this paper, according to the entity relation feature template described in Fig. 2, the entity relation instances in the field of Chinese acupuncture and moxibustion are mapped into eigenvector. The input of the quantization module of entity relation instances is a set of abstracts of Chinese acupuncture and moxibustion which are manually labeled with entities and entity relations. The output is the instance vector set in Chinese acupuncture and moxibustion field. The program flow is shown in Fig. 3.

NiHao - the Chinese word segmentation tool developed by our lab is used in word segmentation and POS tagging. In order to make entities intact when they are divided from the original corpus, such as Jingming, Abdominal acupuncture, Acupoint application and so on, we first use the entity sets extracted from the corpus to train word segmentation dictionary, and then use the word segmentation tool to segment the original corpus and tag the part of speech.

3.3. **The Generation of Entity Relation Classification Model.** After the entity relation instance in the corpus of Chinese acupuncture and moxibustion field is mapped into the feature vector set, the next step is to select the appropriate classifier to classify the vectors with similar features into one class. At present, mature machine learning classification algorithms, such as K-nearest neighbor (KNN), support vector machine (SVM), maximum entropy (MaxEnt), etc., are applied to relation extraction. SVM was first proposed by Vapnik at reference [19]. The basic idea of SVM is to get the maximum edge of segmentation training data for hyperplane constructed by weight vector. Because SVM classifier has many advantages such as good universality, high classification precision, fast classification speed and independence of the number of training samples, SVM algorithm is selected to construct the entity relation classifier in TCM acupuncture and moxibustion field. First, separate entity relation instance feature vector set into train set and test set; then train the SVM classifier by the train set to get the classification model which can distinguish the entity relation type in the field of Chinese acupuncture and moxibustion; finally, evaluate the predictability of the classification model of the entity relation by using the test set in TCM acupuncture and moxibustion field.

4. **Experiments and Result Analysis.**

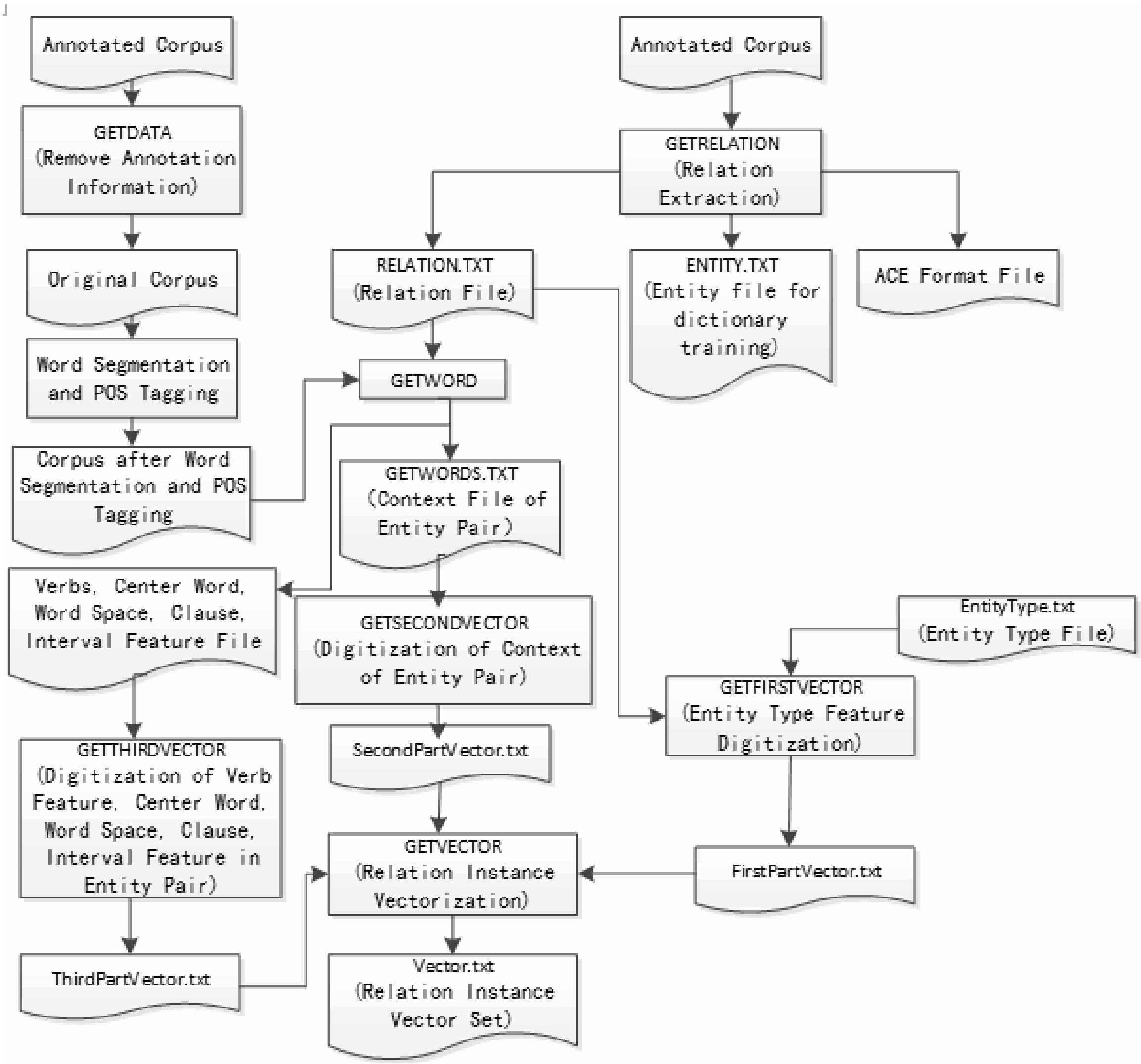

FIGURE 3. Entity relation instance vectorization program flow chart

4.1. **Test corpus and evaluation indicator.** We scrap 4.2M literature information of "Acupuncture Clinical Journal" published between 2009 and 2013 from the website (http://www.cqvip.com/). Randomly select 350 abstracts from the literature, and tag them manually in accordance with pre-defined five categories of entity types and four categories of entity relation types in Chinese acupuncture and moxibustion field. According to the feature template of entity relations in Figure 2, we extract the instances of entity relations sentence by sentence from the annotated corpus, and map them into eigenvectors.The design of Taiwan University Professor Lin Zhiren - libSVM tool is used to construct entity relation classification model in Chinese acupuncture and moxibustion field. The entity relation classification model is evaluated by three common evaluation indicators - precision rate(P), recall rate (R) and F-value. Assumeing that the number of samples correctly classified is TP, the number of samples that are misclassified is FN, the number of samples incorrectly classified into this group is FP, the specific definition of the evaluation index is as follows:

$$P = \frac{TP}{(TP + FP)} \times 100\%. \tag{1}$$

$$R = \frac{TP}{(TP + FN)} \times 100\%. \quad (2)$$

$$F - value = \frac{2 \times P \times R}{(P + R)} \times 100\%. \quad (3)$$

4.2. **Analysis of experiment results.** A total of 3016 relation instances are extracted from 350 manually-tagged corpora, including four types of relation instances of the pre-defined types and one class of relation instances which indicates there are no pre-defined types between two entities. In this paper, a relation instance between two entities that has no pre-defined type is tagged as a relation instance of type <Relation-Null>. The instance of <Relation-Null> relation is introduced to increase the number of negative cases to solve the problem of imbalance of positive and negative data in test and training data. In order to verify the effectiveness, we test the method of this paper and that of [8] comparatively. Two sets of different feature sets are used to vectorize entity relations in the field of Chinese acupuncture and moxibustion; the 5-fold cross-validation experiments are carried out at the instance level; four entity relation classification models are constructed by using the libSVM tool; we take the average of precision rate, recall rate and F-value of 5 groups of experiments on entity relation classifier. The results of the comparative experiments are shown in Table 3.

Table 3. The Results of Experimental Comparison between Suggested and Baseline Method

Number of Article	Number of Entity Relation Instance	Experi-mental Method	Type of Entity Relationship	Number of Positive Instance	Number of Negative Instance	$P_{Average}$	$R_{Average}$	$F_{Average}$
350	3016	Baseline Experiment	DM- Entity Relationship	1084	1932	92.954%	92.268%	92.610%
350	3016	Baseline Experiment	HM- Entity Relationship	249	2767	98.462%	76.760%	86.267%
350	3016	Baseline Experiment	AM-Entity Relationship	138	2878	96.036%	81.080%	87.927%
350	3016	Baseline Experiment	DRM-Entity Relationship	60	2956	100.000%	62.346%	76.806%
350	3016	Suggested Method	DM- Entity Relationship	1084	1932	93.606%	92.902%	93.253%
350	3016	Suggested Method	HM- Entity Relationship	249	2767	98.462%	78.230%	87.188%
350	3016	Suggested Method	AM-Entity Relationship	138	2878	94.964%	79.546%	86.574%
350	3016	Suggested Method	DRM-Entity Relationship	60	2956	98.000%	74.370%	84.565%

In the feature template of the baseline experiment, the entity type and the entity context feature are used. In addition to the features used in the baseline experiment, the feature templates of the method including verb feature, entity all word feature, word distance feature, clause feature and interval entity features are also added into the feature template of method used in this thesis. The data in Table 3 shows that the entity type and the entity context feature contribute greatly to the correct classification of the entity relations in the field of Chinese acupuncture and moxibustion, but the addition of the new feature set further enhances the F-measure of the entity relation classifier of DM, HM and DRM types one more step.

In order to statistically analyze the influence of each feature in the feature set defined in this paper on the correct classification of entity relations in the field of Chinese acupuncture and moxibustion, we experiment on adding features into new feature sets on the basis of the baseline experiment, after which the P, R, F values of each classifier, and the difference in F-measure with the baseline experiment are recorded. The results are shown in Table 4.

Table 4. The influence of each partial feature on the classification result of entity relationship

Type of Entity Relationship	Feature	$P_{Average}$	$R_{Average}$	$F_{Average}$	$F_{Average}$-$F(Baseline)_{Average}$
DM	Verb	93.42%	92.63%	93.023%	0.413%
DM	All Words of Entities	93.606%	92.994%	93.299%	0.689%
DM	Word Distance	92.956%	92.444%	92.699%	0.089%
DM	Clause	92.936%	92.084%	92.490%	−0.120%
DM	Separation of Entity	92.926%	91.996%	92.459%	−0.151%
HM	Verb	98.462%	77.098%	86.480%	0.213%
HM	All Words of Entities	98.436%	75.952%	85.744%	−0.523%
HM	Word Distance	98.448%	75.656%	85.560%	−0.707%
HM	Clause	98.406%	74.848%	85.025%	−1.242%
HM	Separation of Entity	98.436%	75.274%	85.311%	−0.956%
AM	Verb	96.868%	78.804%	86.907%	−1.020%
AM	All Words of Entities	96.868%	78.114%	86.486%	−1.441%
AM	Word Distance	95.084%	81.08%	87.525%	−0.402%
AM	Clause	96.988%	81.080%	88.323%	0.396%
AM	Separation of Entity	95.996%	80.210%	87.396%	−0.531%
DRM	Verb	100.000%	62.346%	76.806%	0.000%
DRM	All Words of Entities	100.000%	74.756%	85.555%	8.749%
DRM	Word Distance	100.000%	73.218%	84.539%	7.733%
DRM	Clause	100.000%	60.528%	75.411%	−1.395%
DRM	Separation of Entity	100.000%	56.528%	72.227%	−4.579%

From the data in Table 4, it can be seen that different features have different effects on the classification results of different types of entity relationships. Features of verbs, all word features of entities, features of word distance can improve the correctness of DM type classification; all word features of entities, features of word distance can greatly improve the correctness of DRM classification; verb features have positive effect on the classification of entity relations of HM, so do features of clause on the classification of entity relations of AM. Therefore, different combination of features should be chosen for different types of entity relation classification. The separation of entity features has negative effect on the classification of four types of entity relations, which could be considered to be removed from the feature set. Verb features, all word features of entities do not contribute to the correct classification or produces negative effects on the correct classification of HM and AM types due to segmentation errors, for example: “Fire needle / medicine has/YOU-VERB clear/ADJ lower blood pressure/NVERB-N effect/NVERB-N” . “Pressure” is a disease named entity, is all word features of entities. “Lower” is the verb, the word segmentation tool does not identify them correctly, which directly leads to the decrease of the classification ability of the entity relation or the interference to the entity relation classification.

5. **Conclusion.** This thesis defines 5 kinds of Chinese acupuncture and moxibustion entity types and 4 types of entity relations. Aiming at the text features of Chinese acupuncture and moxibustion field, this paper discusses the effective feature set of entity relationship classification. In the corpora of Chinese acupuncture and moxibustion field, we use SVM machine learning algorithm based on eigenvector to extract entity relations. The experimental results show that the suggested method can effectively improve the performance of DM, HM and DRM entity relation extraction tasks, and the contributions of verb feature, all word features of entities, word distance feature and clause feature are different for different types of entity relation classification. In the following studies, we will discuss the best feature combination of different types of entity relationship classification in Chinese acupuncture and moxibustion field. At the same time, we will further collect the vocabulary of Chinese acupuncture and moxibustion and adding the training of word segmentation dictionary in order to improve the accuracy of text segmentation in Chinese acupuncture and moxibustion field, so as to better classify the entity relationship through verb features and all the word features of entities.

Acknowledgment. The authors wish to thank Natural Science Foundation of Fujian Province of China(2014J01218), the Project in Fujian Provincial Education Bureau (JAT160328 JZ160461), the scientific research project in Fujian University of Technology (GY-Z160066).

REFERENCES

[1] K. Lee,S. Lee, S. Park, et al. BRONCO: Biomedical entity Relation ONcology COrpus for extracting gene-variant-disease-drug relations, *[J]. Database*, 2016: baw043.

[2] H. Wan, M. F. Moens, W. Luyten, et al. Extracting relations from traditional Chinese medicine literature via heterogeneous entity networks,*[J]. Journal of the American Medical Informatics Association*,vol. 23, no. 2, pp. 356-365, 2016.

[3] Bravo, J. Pinero, et al. Extraction of relations between genes and diseases from text and large-scale data analysis: implications for translational research, *[J]. BMC bioinformatics,* vol. 16, no. 1, pp. 1, 2015.

[4] S. Riedel, L. Yao, A. McCallum, et al. Relation extraction with matrix factorization and universal schemas[J]. 2013.

[5] J. Weston, A. Bordes, O. Yakhnenko, et al. Connecting language and knowledge bases with embedding models for relation extraction, *[J]. arXiv preprint arXiv:*, pp. 1307–7973, 2013.

[6] Min B, Grishman R, Wan L, et al. Distant Supervision for Relation Extraction with an Incomplete Knowledge Base, *[C]//HLT-NAACL.* pp. 777-782, 2013.

[7] N. Kambhatla, Combining lexical, syntactic, and semantic features with maximum entropy models for extracting relations, *[C]//Proceedings of the ACL 2004 on Interactive poster and demonstration sessions. Association for Computational Linguistics*, 22, 2004.

[8] W. Che, T. Liu, S. Li, Automatic Entity Relation Extraction, *[J]. Journal of Chinese Information Processing,*,vol. 19, no. 2, pp. 1-6.

[9] D. Jing, L. Sun, Y. Feng, et al. Chinese Automatic Entity Relation Extraction[J]. Journal of Chinese Information Processing, vol. 21, no.4, pp. 80-85, 2007.

[10] W. Li, P. Zhang, F. Wei, et al. A Novel Feature-based Approach to Chinese Entity Relation Extraction[C]. Proceedings of ACL-08, Columbus, Ohio, USA, pp. 89-92, 2008.

[11] Y. Chen,Q. Zheng, W. Zhang, Omni-word Feature and Soft Constraint for Chinese Relation Extraction[C].Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics,Baltimore,Maryland,USA, pp.572-581, 2014.

[12] D. Zelenko,C. Aone, A. Richardella, Kernel methods for relation extraction[J]. The Journal of Machine Learning Research, vol. 3, pp. 1083-1106, 2003.

[13] M. Zhang, J. Zhang, J. Su,et al., A composite kernel to extract relations between entities with both flat and structured features*[C]//Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the Association for Computational Linguistics,Sydney,Australia*, pp.825-832, 2006.

[14] B. Liu, F. Li, L.Liu, et al. Implementation of a Kernel-Based Chinese Relation Extraction System, *[C]//Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the Association for Computational Linguistics,Sydney,Australia*, vol. 44, no. 8, pp. 1406-1411, 2007.

[15] R. Lin, J. Chen, X. Yang, et al. Research on Multi-information Fusion Chinese Relation Extraction Technology*[J]. Journal of Xiamen University(Natural Science)*, vol. 50, no. 3, pp. 540-545, 2011.

[16] Z. Yang, H. Lin, Y. Li, BioPPISVM Extractor: A Protein-protein Interaction Extractor for Biomedical Literature Using SVM and Rich Feature Sets*[J]. Journal of Biomedical Informatics*,2010, vol. 43(1):88-96.

[17] I. Segura-Bedmar, P. Martinez, C. de Pablo-Sá nchez, Using a shallow linguistic kernel for drug - drug interaction extraction, *[J]. Journal of biomedical informatics*, vol. 44(5): 789-804, 2011.

[18] O. Uzuner, J. Mailoa, R. Ryan, T. Sibanda, Semantic relations for problem-oriented medical records[J]. *Artificial Intelligence in Medicine*,vol.50, no. 2, pp. 63-73, 2010.

[19] V. N. Vapnik, Statistical Learning Theory, *[M]. New York:John Wiley and Son*,1998.