

A Hybrid Method for Entity Hyponymy Acquisition in Chinese Complex Sentences¹

Yunru Cheng^a, Jianyi Guo^{a,b}, Yantuan Xian^{a,b}, Zhengtao Yu^{a,b}, Wei Chen^{a,b}, and Qiyue Yang^a

^a*School of Information Engineering and Automation, Kunming University of Science and Technology, Kunming, Yunnan, China*

^b*Key Laboratory of Pattern recognition And Intelligent computing of Yunnan College, Kunming, Yunnan, China*
e-mail: gjade86@hotmail.com

Received April 19, 2016; in final form, July 28, 2016

Abstract—Extracting entity hyponymy in Chinese complex sentences can be a highly difficult process. This paper proposes a novel hybrid approach that combines parsing with supervised learning and semi-supervised learning. First, conditional random fields (CRF) model is employed to obtain the candidate domain named entity. Pattern matching is then used to acquire candidate hyponymy. Next, predicate and symbol features, syntactic analysis, and semantic roles are introduced into the CRF features template to identify the hyponymy entity pairs. Finally, analysis of both the parallel relationship of entities among sentences and entity pairs in simple sentences is conducted to obtain the hyponymy entity pairs in Chinese complex sentences. The experimental results show that the proposed method reduces the manual work required for CRF markers and has an improved overall performance in comparison with the baseline methods.

Keywords: Chinese complex sentences, hyponymy, entity identification, CRF, pattern matching, syntactic analysis

DOI: 10.3103/S0146411616050035

1. INTRODUCTION

The automatic extraction of hyponymy between entities in a specific field is often a difficult step to accomplish when developing ontology knowledge bases or during construction of knowledge graphs, which is regarded as a crucial part for developing high-level natural language applications. Many algorithms have been developed to achieve hyponymy automatically, including dictionary-based, pattern-based, and statistical machine learning.

The dictionary-based method used to acquire hyponymy usually depends on field words, synonyms, or approximate synonyms in the man-made dictionary. For instance, Nakaya [1] obtained English domain concepts hyponymy with WordNet [2]. Moreover, Li [3] et al. used the geographical names dictionary for acquisition of toponym ontology concept relations in Chinese. However, the coverage of WordNet or HowNet [4] is still limited in the general field, and they are weak in the coverage of domain-specific terms and named entities.

Pattern-based methods mainly use linguistics and natural language processing technologies, which consider patterns as rules summarized according to some inherent regular language. For example, in this Chinese sentence “丽江的景点, 有丽江古城, 玉龙雪山和泸沽湖等” (“The scenic spots of Lijiang contains Lijiang old town, Yulong jokul, and Lugu lake, etc.”), here “Lijiang”, “Lijiang old town”, “Yulong jokul”, “Lugu lake” are entities, E_i is used to represent the entity i ($i = 1, 2, \dots$), and makes $E_1 = \text{Lijiang}$, $E_2 = \text{Lijiang old town}$, $E_3 = \text{Yulong jokul}$, $E_4 = \text{Lugu lake}$, the pattern “ E_1 contains E_2 , E_3 and, E_4 ” can be obtained, where E_1 is the hypernym of E_2 , E_3 , and E_4 in this context. Several researchers have studied the subject in regards to the English language. Hearst [5] raised 6 lexical-syntactic patterns and avoided the use of the pre-coding method of knowledge and automatic recognition hyponymy from large text corpus; Tuan [6] et al. put forward to build a term classification hierarchy model based on the combination of pattern matching and statistical methods; Bansal [7] et al. used an automatic acquisition pattern and a belief propagation algorithm to obtain the high precision hypernym of arbitrary noun phrases. In addition

¹ The article is published in the original.

to the English language, various studies have been conducted on Chinese. For example, Wu [8] et al. established a universal grammar pattern library for the hyponymy relationship according to the part-whole relationship of the six different kinds of forms; In view of the “ISA” pattern, Tang [9] et al. demonstrated hyponym concepts using syntactic parsing results; Tian [10] et al. proposed a combination of a bootstrapping method and a Chinese doubly anchored pattern of the hypernym. The pattern-based method is highly accurate for specific sentences, but compared to English, Chinese language has the more complicated sentence structure and semantic, so it is difficult to use this technique in generalization.

The statistical machine learning method is mainly based on corpus, linguistic knowledge, and the statistical language model, and it uses the machine-learning algorithm to obtain the concept hyponymy relationship. In the field of English, Fan [11] et al. considered the task as a matrix factorization problem to extract the relation. For Chinese, Xia [12] et al. proposed a method of graph clustering to extract the hyponymy; Fu [13] et al. joined a variety of features into SVM (support vector machine) based on the RBF (radial basis function) kernel. Statistical machine learning is considered the main method for hyponymy extraction due to its advantages of reducing manual annotation and extracting information automatically.

The three methods described above generally extract entity hyponymy relations in a simple sentence. For complex sentences, it is necessary to co-operate syntactic patterns with entity recognition as well as other approaches to improve the overall performances, especially for the Chinese. In addition, patterns should be extracted using the bootstrapping method.

The current paper proceeds as follows: Section 2 describes the methods and the framework; Section 3 discusses the process of hyponymy acquisition; Section 4 reports the results; Section 5 presents the conclusions.

2. PROPOSED METHOD

In the majority of cases, Chinese sentences are complex. For example, consider the following sentence: “唐三彩器物类别, 如生活用品, 主要有瓶、壶、罐、钵、杯、盘、盂、烛台、砚、枕等, 模型有亭台楼榭、仿木箱柜、住房、仓库、厕所、牛车、马车等, 俑类有各种人物与动物, 如贵妇、达官、男女侍、武士、天王、胡人及马、骆驼、猪、羊、鸡、狗等” (“The tri-colored glaze pottery of Tang Dynasty has three main categories, including living goods, models, and pottery figurines. The living goods mainly include bottles, kettles, pots, alms bowls, cups, plates, containers, candlesticks, inkslabs, and pillows. The models include pavilions, imitation wood cabinets, houses, warehouses, toilets, cow carts, and carriages. Pottery figurines include all kinds of personage and animals, such as dowagers, high ranking officials, attendants, warriors, heavenly kings, the northern barbarian tribes in ancient China, horses, camels, pigs, sheep, chicken, dogs, etc.”). This sentence contains two layers of hyponymy. Using pattern matching or the CRF method may not accurately identify these layers. Thus, a combined method must be employed. Sang [14] et al. extracted the lexical patterns in an English simple sentence that might contain hyponymy and analyzed the dependency relation between the elements in that sentence. Liu [15] et al. extracted Chinese entity relations by making use of the syntactic characteristics and semantic characteristics in combination with the SVM algorithm to ensure that the F-Measure value increased significantly; Chen [16] et al. used not only the result of parsing trees, but also assembled six candidate feature sets using constraint conditions to obtain Chinese hyponymy.

Our work proposes a hybrid method for entity hyponymy acquisition. The following steps were expressed as four stages (detailed in Section 3). Figure 1 shows the overall framework of the proposed method.

3. HYPONYMY ENTITY PAIRS EXTRACTION

3.1. Extracting Candidate Domain Entities using the CRF Model

In the field of tourism, word segmentation tools cannot accurately identify entities, owing to segmentation tools are not perfect, and there no scenic spots dictionary exists for this field. The current work extract candidate domain entities by fusing domain entity feature and CRFs, which ensured that the word was not separated by segmentation tools.

The process includes 4 steps: (1) Preprocessing. The corpus crawled from the Web is first processed by deleting the duplicated corpus, de-noising, separating sentences, word segmentation and part-of-speech tagging [17] et al. (2) Feature selection and rules marking. By selecting the features to identify the entities and marking the entity features for these sentences as training data of domain entity recognition. (3) Obtaining domain entity recognition model. Domain entity recognition model is generated by learn-

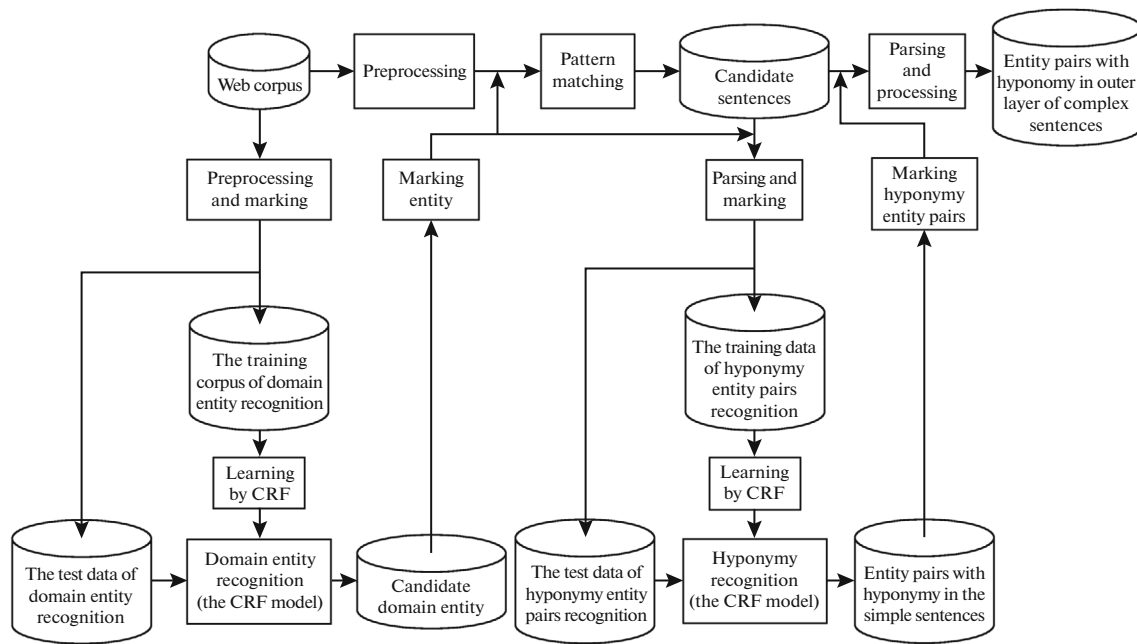


Fig. 1. Overview of the proposed approach.

ing the training corpus of domain entity recognition based on CRFs. (4) Generating candidate entity. The test data of domain entity recognition was input into the domain entity recognition model to identify the domain entity, and make the identified entity as the candidate entity. Finally, these domain entities are joined into the domain dictionary.

3.1.1. Principle of CRFs. CRFs, represented by an undirected probability graph model, are used to markup and serialize data segmentation. They can be used to express the long distance dependence and overlap of elements. When using CRFs, the specific task is to choose the appropriate feature-set for the model, and the simple feature is often used to reflect the complex linguistic phenomenon.

The principle of CRFs can be demonstrated through the following example. Suppose X indicates the observed sequence to be marked, and C indicates the respective joint probability distribution of the observed sequence. Given an effective observation sequence X , which has a length m , $X = x_1, x_2, \dots, x_m$, the marked sequence C , $C = c_1, c_2, c_3, \dots, c_m$ can be obtained, and the set can be defined as follows:

$$P(C/X) = \frac{1}{Z(x)} \exp \left\{ \sum_{k=1}^k \lambda_k f_k(c_t, c_{t-1}, x_t) \right\}, \quad (1)$$

where $Z(x)$ indicates a normalization factor constant, which makes all probability values fall into the range $[0,1]$ of the result state sequence. $Z(x)$ can be expressed as follows:

$$Z(x) = \sum_c \exp \left\{ \sum_{k=1}^k \lambda_k f_k(c_t, c_{t-1}, x_t) \right\}, \quad (2)$$

where $f_k(c_t, c_{t-1}, x_t)$ is a transfer function and represents the transition probability of the marking sequence in the observing sequence X when the position in t and $t-1$; λ_k indicates the corresponding feature weight vector, which is calculated in the training sample.

3.1.2. Feature selection and rules marking. In the process of named entity extraction, it can be found that some nouns may form a domain named entity if they have an attribute (ATT) relationship after syntactic parsing. For example, the “唐三彩” (“Tangsancai”) and “器物” (“utensils”) in the above example sentence have an attribute relationship, it can be inference that “唐三彩器物” (“Tangsancai utensils”) may be a named entity. Moreover, the nouns grouped between the Chinese punctuation “、” and the punctuation “、” itself may be domain entities. An example in this sentence is “仿木” (“imitating wood”).

Table 1. Marking rules of domain entity

Tag	Description
ATTB	The prelude noun of the nouns groups which have ATT relationship
ATTE	The afterbody noun of the nouns groups which have ATT relationship
PB	The prelude noun of the nouns grouped between the Chinese punctuation “、” and “、”
PE	The afterbody noun of the nouns grouped between the Chinese punctuation “、” and “、”

Table 2. Bootstrapping method for pattern extraction

Input: the seeds set with hyponymy (Seeds_sets); the corpus of scenic spots of Sougou Baike (Corpus)

Output: the patterns set with hyponymy (Pattern_sets); the number of patterns (N)

Step:

- 01: Searching the Seeds_sets in the Corpus and get some instances (S), and then, employing S to produce pattern instances.
- 02: Generalizing pattern instances to get the patterns (Candidate_patterns), and joining into the Pattern_sets. The number of Candidate_patterns is n , $N = N + n$.
- 03: Matching the patterns of Candidate_patterns in the Corpus, obtaining entity pairs of S as NewSeeds_sets. Join the NewSeeds_sets into Seeds_sets.
- 04: Making use of NewSeeds_sets, repeating the step 01, 02 and 03 until the N no longer increases. The iteration is over, and return the Pattern_sets.

Table 3. Examples of hyponymy patterns

Pattern	Note
E (包括 /包含 /主要有 /:) E_1, E_2, \dots, E_n (In English: E (include/contain/mainly have/:) E_1, E_2, \dots, E_n)	E is the domain concept in the pattern; E_1, E_2, \dots, E_n are concept instances; E_1, E_2, \dots, E_n is appositive and constitutes a category.
E_1, E_2, \dots, E_n (组成 /等) E (In English: E_1, E_2, \dots, E_n (compose of/and so on) E)	

and “箱柜” (“bin”). Thus, the ATT relationship and the punctuation “、”-were taken as characteristics to be joined in the CRF model to extract a named entity. The examples of marking format are “唐三彩/ATTB 器物/ATTE” and “仿木/PB 箱柜/PE”. Table 1 describes the marking rules of domain entity.

It is difficult to extract the hyponymy relation sentences in the tourism domain by pattern matching, due to the fact that there are no unified and fixed patterns, which contain hyponymy in the Chinese tourism field. Therefore, the bootstrapping method was applied.

3.2. Extracting Candidate Sentences by Pattern Matching

3.2.1. Extracting pattern by bootstrapping method. Table 2 describes the bootstrapping method algorithm [18] for pattern extraction semi-automatically. Whose basic idea is that given a small set of seed instances for a hyponymy relation, the system learns lexical patterns, applies them to extract new instances, the required pattern scale can be obtained finally by progressive learning.

Table 3 shows some typical examples of the patterns studied. Latter experiments proved that these patterns might cover the majority of hyponymy relation sentences.

3.2.2. Extracting candidate sentences. The corpus crawled from the Web was first preprocessed by deleting the duplicated web pages, de-noising, separating sentences. Then marking candidate entity (obtained in section 3.1) of these sentences and E_i replace the entity with label. At Last, extracting the candidate sentences by pattern matching.

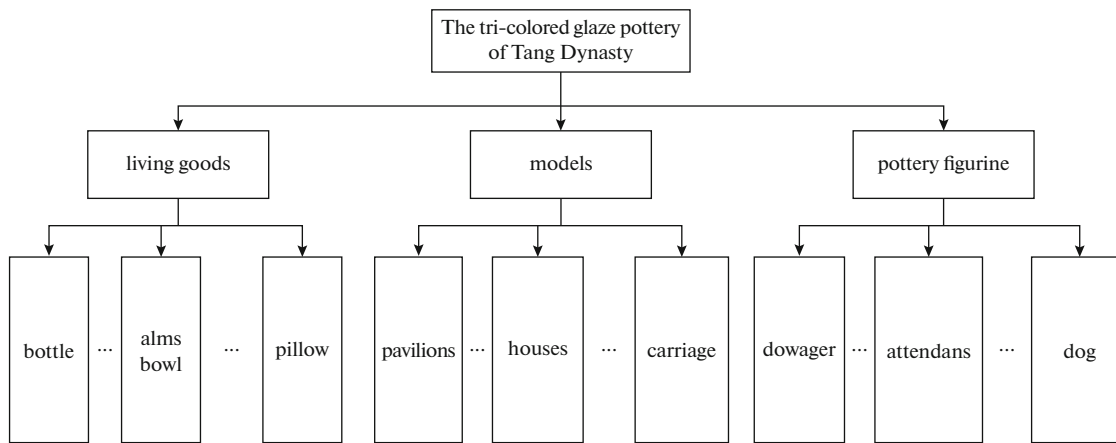


Fig. 2. Hyponymy model of sentences.

3.3. Hyponymy Entity Pairs Extraction in Simple Sentences

Consider the Chinese sentence mentioned in Section 2, “唐三彩器物类别，如生活用品，主要有瓶、壶、罐、钵、杯、盘、盂、烛台、砚、枕等，模型有亭台楼榭、仿木箱柜、住房、仓库、厕所、牛车、马车等，俑类有各种人物与动物，如贵妇、达官、男女侍、武士、天王、胡人及马、骆驼、猪、羊、鸡、狗等。” Based on the definition of hyponymy relationship, it can be seen that this example is a complex sentence containing 2 layers of hyponymy (Fig. 2), and 4 typical simple sentences containing hyponymy. For example, it can be seen that “唐三彩器物，生活用品” (“the tri-colored glaze pottery of Tang Dynasty, living goods”), “生活用品，瓶” (“living goods, bottle”), and “模型，亭台楼榭” (“model, pavilions”) are hyponymy entity pairs, but this is difficult for automatic recognition.

By deeply observation, we found that the sentence is complex in most cases. Complex sentences are composed of clauses, which are typically simple sentences. Thus, hyponymy entity pairs must first be extracted from the clauses (i.e. simple sentences) in a complex sentence.

3.3.1. The process of extract hyponymy entity pairs in simple sentences. The procedure contains 4 sections: (1) Marking entity. For these candidate sentences, marking entity is first performed. (2) Generating the training data. Dependency parsing is used to parse the candidate sentences and the hyponymy features are obtained. And then, the candidate sentences with hyponymy features labeling as the training data of hyponymy entity pairs recognition. (3) Obtaining hyponymy recognition model. Hyponymy recognition model is generated by learning the training corpus of hyponymy entity pairs recognition based on CRFs. (4) Identifying entity with hyponymy in simple sentences. The test data of hyponymy entity pair recognition is injected into the hyponymy recognition model to identify the entity with hyponymy in simple sentences.

3.3.2. Feature selection and rules marking. In dependency parsing [19], several features can be obtained in addition to the dependency relation (clauses structure feature), such as semantic role labeling (shallow semantic analysis technology, labeling certain phrases in the sentence for a given predicate argument), agent, patient, time, and place.

Semantic role labeling can identify the predicate verbs, A0 (the agency of action) and A1 (the influence of the action). Figure 3 shows that A0 generally contains hypernym entities, and A1 contains hyponym entities in simple sentences. Moreover, those sentences might contain the hyponymy if they contained the coordinate relationship (COO) during syntactic parsing. Table 4 describes the tags of relation types present in Figs. 3 and 4.

In addition, there are other characters that can help identify the hyponymy, such as words like “包括” (“include”), “包含” (“contain”), and “主要有” (“mainly have”), and symbols like “:” and “、”. These combinations of features might better express entity hyponymy. If the sentences have the deep semantic relations, such as COO, then it will be labeled as COO. Table 5 shows the marking rules of hyponymy.

The current work is combining these features with the CRFs to obtain the hyponymy recognition model. Selecting the combination feature involves linear extraction and combination. And making t_i represent the i -th word, c_i is the part of speech of the i -th word, d_i denote the feature of “include”, “contain” and “:” of the i -th word, e_i express the feature of “、” of the i -th word, f_i represent the feature of COO

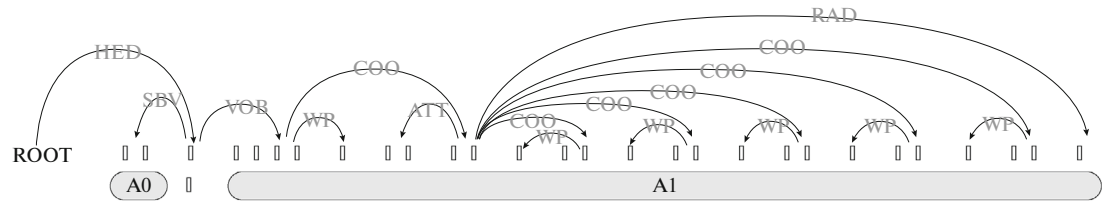


Fig. 3. Syntactic parsing and semantic role labeling

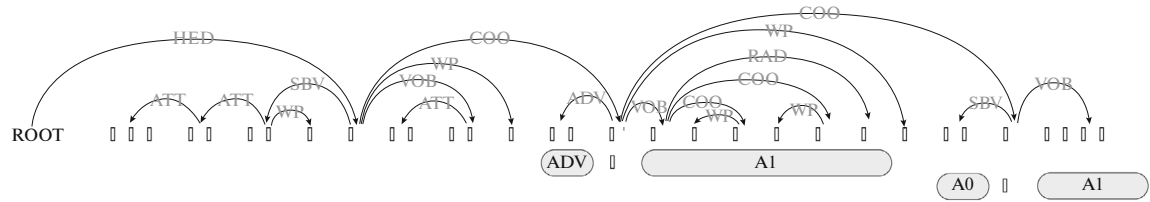


Fig. 4. Syntactic parsing and semantic role labeling.

of the i -th word, g_i express the feature of semantic roles of the i -th word. Table 6 shows the part of feature template of the hyponymy entity pair recognition.

In the Table 6, t_n denotes the single word feature; $t_n t_{n+1}$ is the combination feature of two words; $t_0 c_0 f_0$ represents the combination feature of word, part of speech and COO; $t_0 e_1 e_{+1}$ expresses the combination feature of word, the former “、” and the next “、”; $t_0 e_1 e_{+1} f_n$ is the combination feature of word, the former “、”, the next “、” and the COO; $t_0 e_1 e_{+1} f_n g_n$ denotes the combination feature of word, the former “、”, the next “、”, the COO and the semantic roles.

3.4. Hyponymy Entity Pairs Extraction in Outer Layer of Complex Sentences

When extracting the outer layer hyponymy relation entity pairs in complex sentences, the hyponymy entity pairs in the clauses obtained from Section 3.3 must first be labeled within the candidate sentences. Then, the coordinate relation of hypernyms among clauses can be obtained by syntactic analysis. Finally, the entity pairs with hyponymy of the outer layer in the complex sentences can be obtained by analytical processing.

Figure 4 shows the COO relation among “生活用品”, “模型” and “佣类”. The COO entity relationship is directly marked out in simple sentences, but the hyponymy relation of the outer layer is marked out by the predicate verbs among them in complex sentences. Moreover, these annotations require a certain degree of processing, because the COO relationship of a multilayer is defined as a layer in the process of syntax analysis. We first extracted the hyponymy entity pairs in the simple sentences. Based on this, we can identify the hyponymy entity pairs of the multilayer structure of the complex sentences, the results can be seen in Fig. 2. In front of “生活用品”, there is the “如” (“for example”), as shown in Figure 4(In order to show the COO relation of the predicate verbs “如” and “有” more clearly, the examples of

Table 4. Tags of relation type

Tag	Description
SBV	Subject-verb
VOB	Verb-object
ATT	Attribute
ADV	Adverbial
RAD	Right adjunct
HED	Head
WP	Punctuation

Table 5. Marking rules

Tag	Semantic relationships
A0	The agency of action
A1	The influence of the action
0	No specific feature
1	Have specific features
COO	Coordinative relation

Table 6. Feature templates

Feature
$t_n, n = \{-2, -1, 0, 1, 2\}$
$t_n t_{n+1}, n = \{-2, -1, 0, 1\}$
$t_0 c_0 f_0$
$t_0 e_{-1} e_{+1}$
$t_0 e_{-1} e_{+1} f_n, n = \{-2, -1, 0, 1, 2\}$
$t_0 e_{-1} e_{+1} f_n g_n, n = \{-2, -1, 0, 1, 2\}$

“钵、杯、盘、盂、烛台、砚、枕” of “生活用品” were omitted in Fig. 4). Here, “如” and “有” (“have”) are the COO relationship, but “生活用品” is a hypernym of “瓶”. Thus, the hyponymy must first be recognized in the simple sentences. When combined with the hyponymy entity pairs of the outer layer in complex sentences, there could be interference (Fig. 5).

4. METHODS AND RESULTS

4.1. Data Preparation

Data was crawled by this paper using WebCrawler coding from Sogou Baike tourism field entries. A total of 568 entries and 29,045 sentences were analyzed. The number of sentences that contained hyponymy was 1,362. The corpus is divided into two parts, of which 4/5 as the training corpus, and the rest of 1/5 as the test corpus. The experiment applies the 5-fold cross-validation, and uses the mean value as the result.

4.2. Evaluation Standards

Certain parameters, including accuracy (P), recall (R), and F-Measure value (F), were employed to evaluate the extraction results of domain hyponymy. The parameters are defined as follows:

$$P = \frac{A}{B} \times 100\%, \quad (3)$$

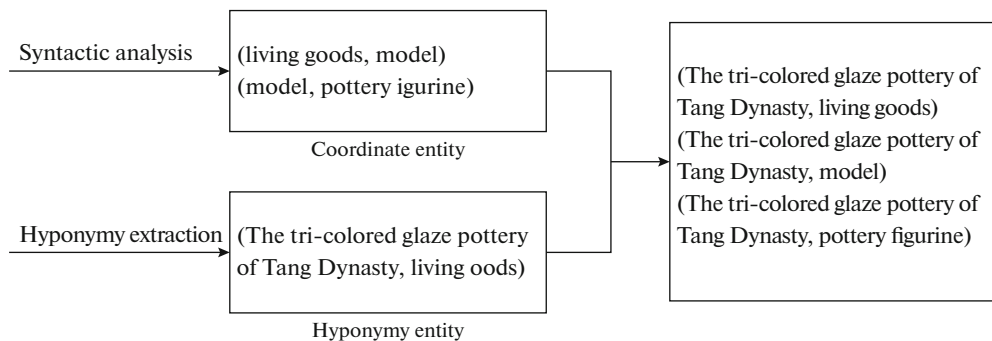
**Fig. 5.** Extracting the entity pairs with hyponymy in outer layer of complex sentences.

Table 7. Experimental results

Algorithm	Accuracy	Recall	<i>F</i>
Pattern matching	63.15%	80.60%	70.82%

Table 8. Experimental results from proposed method and traditional methods

Algorithm	Test Type	Accuracy	Recall	<i>F</i>
CCRFs (traditional characteristic)	Closed test	73.45%	68.82%	71.06%
	Open test	64.43%	59.83%	62.04%
CCRFs + SVM (headword path)	Closed test	80.11%	79.86%	79.98%
	Open test	74.51%	75.89%	75.19%
Pattern + CCRFs (multi-feature)	Closed test	82.54%	78.64%	80.54%
	Open test	77.83%	77.91%	77.87%
Proposed Method	Closed test	83.27%	80.68%	81.95%
	Open test	80.01%	78.37%	79.18%

$$R = \frac{A}{C} \times 100\%, \quad (4)$$

$$F = \frac{2 \times P \times R}{P + R} \times 100\%, \quad (5)$$

where A represents the correct identification number of domain entities hyponymy; B represents the total number of domain entities hyponymy; and C represents the total number of domain entities hyponymy in the corpus.

4.3. Results

The candidate sentences with hyponymy were first extracted using the defined patterns. Here, the top 10 high-frequency patterns were used. Table 7 shows the experimental results.

In the test process of extracting hyponymy, 4 groups of experiments were made, including the current method and 3 traditional methods for comparison. The first method was CCRFs (cascaded conditional random fields) with traditional characteristics proposed in the literature [20]; the second was the method presented in literature [21], which applied the part of speech and the suffix of characteristic combined with CCRFs to extract entities, and combined the feature of headword path with SVM [22] to extract entity hyponymy; the third was CCRFs with multi-features plus the pattern-based method, and the CRF++ toolkit [23] was used. Table 8 shows the comparison of the results from the various methods.

Although the proposed method resulted in some errors due to pattern matching to extract hyponymy sentences, it does not require a great deal of manual annotation. Thus, it saves a large amount of manpower and time. Moreover, it has a better performance than the other methods in terms of accuracy and recall rate.

Table 7 indicates that the factors affecting the experimental data are as follows:

- (1) A small number of domain entities failed to be recognized.
- (2) Some sentences that might contain hyponymy were filtered out by pattern matching, which resulted in part of the entity pairs being unable to be extracted.
- (3) There was an error associated with the syntactic parser tool. Though syntactic parsers are evolving and improving, an error rate still exists. This paper used the NLP (natural language processing) parser from MIT (Massachusetts Institute of Technology) [24].

5. CONCLUSIONS

This article proposed a hybrid method for extracting hyponymy in a complex sentence by combining syntactic analysis features and other features with CRFs model. Chinese language contains many complicated characteristics that are worth studying. Manual annotation is one of the greatest limitations in the

proposed method, along with the use of the syntactic. Future studies will involve improving the current method and finding deep semantic combination features for hyponymy acquisition.

ACKNOWLEDGMENTS

This research was supported in part by the National Natural Science Foundation of China (Grant nos. 61262041, 61472168, and 61562052) and the key project of National Natural Science Foundation of Yunnan province (Grant no. 2013FA030).

REFERENCES

1. Nakaya, N., Kurematsu, M., and Yamaguchi, T., A domain ontology development environment using a MRD and text corpus, *Proc. of the Joint Conf. on Knowledge Based Software Engineering*, 2002, pp. 242–253.
2. *WordNet: A Lexical Database for English*, Princeton University. <http://wordnet.princeton.edu/wordnet/>.
3. Li, H., Li, W., Liang, R., et al., Toponym ontology concept semantic relation research based on place name dictionary and thesaurus, *China Place Name*, 2010, vol. 10, pp. 71–74.
4. Dong, Z. and Dong, Q., HowNet. http://www.keenage.com/html/c_index.html.
5. Hearst, M.A., Automatic acquisition of hyponyms from large text corpora, *Proceedings of the 14th Conference on Computational Linguistics*, 1992, vol. 2, pp. 539–545.
6. Tuan, L.A., Kim, J., and Kiong, N.S., Taxonomy construction using syntactic contextual evidence, *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2014, pp. 810–819.
7. Bansal, M., Burkett, D., Melo, D.G., et al., Structured learning for taxonomy induction with belief propagation, *ACL*, 2014, no. 1, pp. 1041–1051.
8. Wu, J., Luo, B., and Cao, C., Acquisition and verification of mereological knowledge from Web page texts, *J. East China Univ. Sci. Technol.*, 2006, vol. 32, no. 11, p. 1310.
9. Tang, Q., Lv, X.Q., and Li, Z., Research on domain ontology concept hyponymy relation extraction, *Microelectron. Comput.*, 2014, vol. 6, pp. 68–71.
10. Tian, F., Yuan, C., and Ren, F., Hyponym extraction from the web by bootstrapping, *IEEE Trans. Electr. Electron. Eng.*, 2012, vol. 7, no. 1, pp. 62–68.
11. Fan, M., Zhao, D., Zhou, Q., et al., Distant supervision for relation extraction with matrix completion, *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*, 2014, vol. 1, pp. 839–849.
12. Xia, F., Cao, X., Fu, J., et al., Extracting part-whole relations based on coordinate structure, *J. Chin. Inf. Process.*, 2015, vol. 29, no. 1, pp. 88–96.
13. Fu, R., Qin, B., and Liu, T., *Exploiting multiple sources for open-domain hypernym discovery*, *EMNLP*, 2013, pp. 1224–1234.
14. Sang, E.T.K. and Hofmann, K., Lexical patterns or dependency patterns: Which is better for hypernym extraction?, *Proceedings of the Thirteenth Conference on Computational Natural Language Learning*, 2009, pp. 174–182.
15. Liu, H., Che, W., and Liu, T., Feature engineering for Chinese semantic role labeling, *J. Chin. Inf. Process.*, 2007, vol. 21, no. 1, pp. 79–84.
16. Chen, Y., Zheng, Q., and Chen, P., Feature assembly method for extracting relations in Chinese, *Artif. Intell.*, 2015, vol. 228, pp. 179–184.
17. Zhang, H., NLPPIR: Chinese word segmentation system. <http://ictclas.nlpir.org/>.
18. Pennacchiotti, M. and Pantel, P., A bootstrapping algorithm for automatically harvesting semantic relations, *Proceedings of Inference in Computational Semantics (ICoS-06)*, 2006, pp. 87–96.
19. The Research Center for Social Computing and Information Retrieval at Harbin Institute of Technology (HIT-SCIR): Language Technology Platform. <http://www.ltp-cloud.com/>.
20. Mo, Y., Guo, J., Yu, Z., et al., Hyponymy extraction of domain ontology concept based on CCRF, *Comput. Eng.*, 2014, vol. 40, no. 6, pp. 138–141.
21. Wang, C. and Yang, Z., An acquisition method of domain-specific terminological hyponym based on structure features of sentence, *J. Chongqing Univ. Posts Telecommun. (Nat. Sci. Ed.)*, 2014, vol. 3, p. 19.
22. Chang, C. and Lin, C., LIBSVM – A library for support vector machines. <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>.
23. Kudo, T., CRF++: Yet another CRF toolkit. <https://taku910.github.io/crfpp/>.
24. Che, W., Li, Z., and Liu, T., LTP: A Chinese language technology platform, *Proceedings of the 23rd International Conference on Computational Linguistics: Demonstrations*, 2010, pp. 13–16.