Feature-Based Relation Classification Using Quantified Relatedness Information

Jin-Xia Huang, Key-Sun Choi, Chang-Hyun Kim, and Young-Kil Kim

Feature selection is very important for feature-based relation classification tasks. While most of the existing works on feature selection rely on linguistic information acquired using parsers, this letter proposes new features, including probabilistic and semantic relatedness features, to manifest the relatedness between patterns and certain relation types in an explicit way. The impact of each feature set is evaluated using both a chi-square estimator and a performance evaluation. The experiments show that the impact of relatedness features is superior to existing well-known linguistic features, and the contribution of relatedness features cannot be substituted using other normally used linguistic feature sets.

Keywords: Relation classification, feature selection, semantic relatedness, probabilistic relatedness, feature-based.

I. Introduction

Relation extraction has gained increasing interest in recent years based on its application in information extraction, knowledge building, information retrieval, and machine translations. The task of relation extraction is identifying the relationships between two or more entities in a given context. Most of the existing works have focused more on the relations between named entities [1]-[3]. Meanwhile, for the purpose of knowledge building, there is also increasing need toward relation extraction for general or domain specific terms [4]-[6]. The latter task is more challenging for several reasons. The semantic categories of terms are more varied than those of named entities, which means that the sense ambiguities of terms are relatively high, and the relation types between terms can be manifold.

Relation extraction can be separated into relation detection and relation classification tasks. Relation candidates are detected by patterns in the detection phase and classified into certain relation types in the classification phase. In relation classification tasks, feature-based machine learning approaches have been broadly used in recent years. The features generally employed in existing researches include the lexicon, part-of-speech (POS), syntactic features, and semantic categories of the entities [1]-[4].

These surface features are employed under an implicit assumption that certain words (POS tags, syntactic roles, and semantic types of entities) have stronger relatedness with certain relation types than other words. For example, an entity in the semantic category of 'Person' is likely to have a 'general-staff' or an 'owner' relation with an entity belonging to an 'Organization' category, but unlikely to have a 'part-of' or 'located' relation.

Under this assumption, this letter proposes relatedness features to represent the relatedness between words and relation types in an explicit way. The proposed relatedness features include a semantic relatedness feature, which represents the semantic relatedness between patterns and relation types by using the semantic similarity, and a probabilistic relatedness feature, which represents the probabilistic relatedness between patterns and relation types using a corpus. The experiments showed that the proposed features contributed to the classification performance in a significant way, and that the proposed features were competitive with deep knowledge features including syntactic ones.

Manuscript received Aug. 25, 2009; revised Jan. 1, 2010; accepted Feb. 23, 2010.

Jin-Xia Huang (phone: +82 42 860 6512, email: hgh@etri.re.kr), Chang-Hyun Kim (email: chkim@etri.re.kr), and Young-Kil Kim (email: kimyk@etri.re.kr) are with the Software Research Division, ETRI, Daejeon, Rep. of Korea.

Key-Sun Choi (email: kschoi@kaist.ac.kr) is with the Division of Computer Science, KAIST, Daejeon, Rep. of Korea.

doi:10.4218/etrij.10.0209.0353

II. Problem Description

The problem described in this letter is the classification of explicit relationships between two entities occurring in a sentence, with an assumption that the two entities are already detected as relation candidates, using a simple pattern matching approach. The entities can be general or domain-specific terms and named entities.

Assume a relation candidate with entities e_1 and e_2, in which context W matches pattern p. We want to predict the relation type r:

$$f:(e_1, e_2, p, W) \rightarrow r.$$

The relation type r can be either *isa*, *usedFor*, *produces*, *provides*, or no-relation. No-relation means that the relation candidate does not hold any relation type. Considering that each relation type already has its own predefined patterns, the multi-classification task can be transferred to a binary classification task, in which the predicted result is either positive or negative for certain relation type r. For example, a relation candidate that is detected by an *isa* pattern will be classified as either *isa* or *non-isa*.

III. Feature Set for Relation Classification

1. Relatedness Features

A. Semantic Relatedness Feature

The semantic relatedness $PatSim(p_i, r)$ between pattern p_i and relation type r is decided by the semantic similarity between w_i and r, where w_i is the main word of p_i (for example, w_i ='use' for p_i = 'be used for', and w_i ='available' for p_i = 'be available for') decided by a human developer, and the similarity $sim(w_i, r)$ between w_i and r is directly proportional to the semantic similarity between w_i and $W=\{w_1,\cdots, w_i,\cdots, w_n\}$, which is the main word set of the pattern set P for relation type r.

$$PatSim(p_i,r) \cong sim(w_i,r) \cong sim(w_i,W)/\max_j\{sim(w_j,W)\}, \quad (1)$$

$$sim(w_i,W)=\sum_{j=1}^{n} 1/dis(w_i,w_j)=\sum_{j=1}^{n} 1/\min(n_{ij}+1), \quad (2)$$

where $dis(w_i, w_j)$ indicates the shortest distance between the synsets of w_i and w_j in WordNet,[1] and n_{ij} is the minimum node number between the two synsets.

B. Probabilistic Relatedness Features

Probabilistic relatedness is acquired from labeled data by calculating the percentage of positive cases out of the relation candidates detected by the pattern (or patterns with the same main word) for certain relation type r, which equals the accuracy of the pattern(s).

$$PatProb(p_i,r)=|r(p_i)|/(|r(p_i)|+|non-r(p_i)|), \quad (3)$$

$$PatMainProb(p_i,r)=\text{average}\{PatProb(p_m,r)\}, \forall w_m=w_i, \quad (4)$$

where $|r(p_i)|$ indicates the number of the positive cases of pattern p_i, $|non–r(p_i)|$ indicates the number of the negative cases of p_i, and p_m indicates the pattern which shares the same main word with pattern p_i.

Table 1 shows some examples of relatedness scores. The relatedness features of pattern 'be a form of' can be described as '*PatSim*: 1.0 *PatProb*: 0.7 *PatMainProb*: 0.5' in feature vector according to Table 1.

Table 1. Examples of relatedness scores.

Relation	Pattern (main word)	*PatSim*	*PatProb*	*PatMain Prob*
isa	be a form of (be)	1.0	0.7	0.5
	refer to (refer)	0.3	0.4	0.4
usedFor	be used for (use)	1.0	0.6	0.5
	be available for (available)	0.1	0.4	0.4
produces	develop (develop)	0.4	0.5	0.5
	create (create)	0.4	0.4	0.4
provides	offer (offer)	0.4	0.5	0.5
	release (release)	0.3	0.6	0.6

2. Existing Linguistics Features

This letter uses most of the linguistic features employed in existing works [1]-[4], which include the following:

Word features: These include the string that matches the pattern, the main word of the matched pattern, the domain and range entities, headwords, and all words of the entities;

Context features (at word level): The words before and after the domain and range entities (the window size is 2 in this letter);

POS, syntactic, and dependency features: These represent the POS, syntactic, and dependency tags of the words included in the context window, respectively.

IV. Experiments

In our experiments, the relation candidates are already extracted by matching predefined patterns on parsed texts. A

1) http://wordnet.princeton.edu/

Connexor parser[2] is used for parsing, and a Bayesian classifier, which is provided by WEKA,[3] is adopted for classification.

1. Dataset and Classification Performance

The data set is collected from Wikipedia pages in an IT domain and labeled by human developers. The inter-rater agreement is evaluated between two developers who are assigned to label the test set and computed using Cohen's Kappa statistic (κ) [7].

In Table 2, the second column shows the number of patterns. The third and fourth columns show the numbers of candidates in each set. The percentage of candidates holding the relation type *r*, which equals the relation detection accuracy, is in parentheses. The fifth column shows the Kappa scores, which are either moderate (*isa*) or fair agreement on different relation types.

Table 2. Data set and relation detection accuracy.

Relation	No. of patterns	Training set	Test set	Kappa
isa	89	35,389 (54.7%)	1,158 (50.2%)	0.58
usedFor	22	720 (43.2%)	126 (42.9%)	0.21
produces	46	1,038 (51.4%)	155 (38.1%)	0.24
provides	17	1,803 (48.2%)	317 (47.3%)	0.22

2. Evaluation on Feature Selection

First, the impact of each feature set is evaluated for *isa* relation type using a chi-square[4] estimator in WEKA. According to our evaluation, the impact of each feature set in its average ranking is as follows: word level features are on the top, then the relatedness feature set is next, followed by POS, dependency, and syntactic feature sets.

Then, to evaluate the contribution of each feature set, the features are added gradually in their order of impact. As a result, we can see that both the accuracy and F-measure are increased gradually, and the increasing ratio of relatedness feature set is higher or comparable with existing linguistic feature sets, which is shown in Fig. 1.

To check the contribution of the relatedness feature set, we eliminated the relatedness features from full feature sets (−relatedness). As the figure shows, both F-measure and accuracy suffer, which means that relatedness features cannot be substituted by other feature sets.

2) http://www.connexor.eu/
3) http://www.cs.waikato.ac.nz/ml/weka/
4) A chi-square is a popular feature selection method, which evaluates features individually with respect to class.

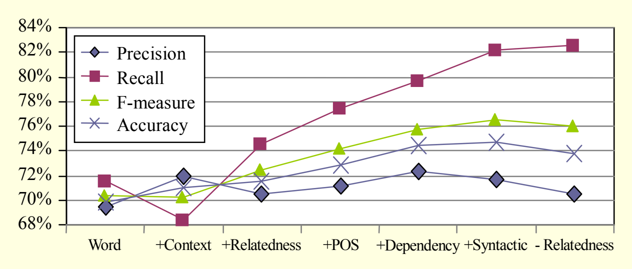

Fig. 1. Contribution of each feature set for *isa* relation type.

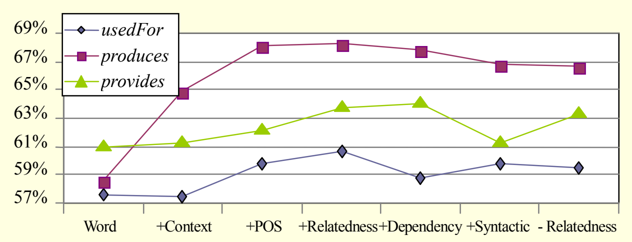

Fig. 2. Feature evaluation on other three relation types with F-measure.

Similar evaluation is also performed on the other three relation types. According to our evaluation, the impact ranking using a chi-square is changed: POS shows higher impact in relation types *provides* and *produces*. However, in whatever case, the impact of the relatedness feature set is higher than dependency and syntactic feature sets; and the highest F-measure is achieved when the relatedness feature set is employed, which is shown in Fig. 2. It means the relatedness feature set shows similar tendency in all relation types.

The contribution of the relatedness features is relatively low in *produces*. It is because its patterns have relatively more diverse senses than other relation types. The patterns use the main words not only like 'produce, make, develop' which are semantically close to 'produce,' but also like 'begin, carry, pioneer, start.' How to fine-tune the pattern set to get the reliable semantic relatedness will be one of our future projects.

V. Conclusion

Feature selection is important for feature-based relation classification. This letter proposes probabilistic and semantic relatedness features to represent the relatedness between patterns and relation types in an explicated way. The experiments showed that the relatedness features have a big impact on the classification performance and cannot be substituted by existing linguistic features.

As future work, we intend to explore the relatedness information between the semantic categories of terms and relation types, while this letter focuses on pattern-related relatedness. Using extracted relations in question answering [8] and applying this research in bilingual relation extraction [9] might be included as part of future projects.

References

[1] N. Kambhatla, "Combining Lexical, Syntactic, and Semantic Features with Maximum Entropy Models for Extracting Relations," *Proc. 42nd ACL*, 2004, article no. 22.

[2] G. Zhou and M. Zhang, "Extracting Relation Information from Text Documents by Exploring Various Types of Knowledge," *Inf. Process Manage.*, vol. 43, no. 4, 2007, pp. 969-982.

[3] Z. Zhang, "Mining Relational Data from Text: From Strictly Supervised to Weakly Supervised Learning," *Inf. Syst.*, vol. 33, no. 3, 2008, pp. 300-314.

[4] R. Girju, A. Badulescu, and D. Moldovan, "Automatic Discovery of Part-Whole Relations," *Computational Linguistics*, vol. 32, no. 1, 2006, pp. 83-135.

[5] H. Liu and P. Singh, "ConceptNet: A Practical Commonsense Reasoning Tool-Kit," *BT Technol. J.*, vol. 22, no. 4, 2004, pp. 211-226.

[6] J.X. Huang, J.A. Shin, and K.S. Choi, "Enriching Core Ontology with Domain Thesaurus through Concept and Relation Classification," *Proc. OntoLex*, *ISWC*, 2007.

[7] J. Cohen, "A Coefficient of Agreement for Nominal Scales," *Educational and Psychological Measurement*, vol. 20, no. 1, 1960, p. 37.

[8] H.J. Oh, S.H. Myaeng, and M.G. Jang, "Enhancing Performance with a Learnable Strategy for Multiple Question Answering Modules," *ETRI J.*, vol. 31, no. 4, Aug. 2009, pp. 419-428.

[9] Y.S. Hwang and Y.K. Kim, "Automatic Acquisition of Paraphrases Using Bilingual Dependency Relations," *ETRI J.*, vol. 30, no. 1, Feb. 2008, pp. 155-157.