Extracting Relations in Social Networks from the Web Using Similarity Between Collective Contexts

Junichiro Mori[1,2], Takumi Tsujishita[1], Yutaka Matsuo[2], and Mitsuru Ishizuka[1]

[1] University of Tokyo, Japan
{jmori, tjstkm, ishizuka}@mi.ci.i.u-tokyo.ac.jp
[2] National Institute of Advanced Industrial Science and Technology, Japan
y.matsuo@aist.go.jp

Abstract. Social networks have recently garnered considerable interest. With the intention of utilizing social networks for the Semantic Web, several studies have examined automatic extraction of social networks. However, most methods have addressed extraction of the strength of relations. Our goal is extracting the underlying relations between entities that are embedded in social networks. To this end, we propose a method that automatically extracts labels that describe relations among entities. Fundamentally, the method clusters similar entity pairs according to their collective contexts in Web documents. The descriptive labels for relations are obtained from results of clustering. The proposed method is entirely unsupervised and is easily incorporated into existing social network extraction methods. Our method also contributes to ontology population by elucidating relations between instances in social networks. Our experiments conducted on entities in political social networks achieved clustering with high precision and recall. We extracted appropriate relation labels to represent the entities.

1 Introduction

Social networks have recently attracted considerable interest. For the Semantic Web, there is great potential to utilize social networks for myriad applications such as trust estimation [1], ontology construction [2], and end-user ontology [3].

Aiming at using social networks for the Semantic Web, several studies have addressed extraction of social networks automatically from various sources of information. Mika developed a system for extraction, aggregation, and visualization of online social networks for a Semantic Web community, called Flink [4]. In that system, social networks are obtained using Web pages, e-mail messages, and publications. Using a similar approach, Matsuo et al. developed a system called Polyphonet [5]. In line with those studies, numerous studies have explored automatic extraction of social networks from the Web [6,7,8,9].

Given social network extraction using the methods described above, the next step would be to explore underlying relations behind superficial connections in those networks. However, most automatic methods to extract social networks

merely provide a clue to the strength of relations. For example, a link in Flink [4] is only assigned the strength of its relation. A user might wonder what kind of underlying relation exists behind the link. In the field of social network analysis, it has been shown that rich information about underlying social relationships engenders more sophisticated analysis [10,11].

One reason for the lack of information about underlying relations is that most automatic extraction methods [6,4,8,9] use a superficial approach (e.g. co-occurrence analysis) instead of profound assessment to determine the type of relation. Matsuo et al. defines four kinds of relations in a research community and classifies the extracted relation [5]. They adopt a supervised machine learning method, which requires a large annotated corpus that requires a great deal of time and effort to construct and administer. In addition, it is necessary to gather domain-specific knowledge a priori to define the extracted relations.

Our goal is to extract underlying relations among entities (e.g., person, location, company) from social networks (e.g., person-person, person-location network). Thereby, we are aiming at extracting descriptive labels of relations automatically, such as affiliations, roles, locations, part-whole, and social relationships. In this paper, we propose a method that automatically extracts the labels that describe relations among entities in social networks. We obtain a local context in which two entities co-occur on the Web, and accumulate the context of the entity pair in different Web pages. Given the collective contexts of each entity pair, the key idea is clustering all entity pairs according to the similarity of their collective contexts. This clustering using collective contexts is based on our hypothesis that entity pairs in similar relations tend to occur in similar contexts. The representative terms in context can be regarded as representing a relationship. Therefore, the labels to describe the relations among entities are extracted from the clustering process result. As an exemplary scenario for our approach, we address a political and social network that is composed of two types of entities: politicians and geopolitical entities (GPEs).

Our method uses context information that is obtained during extraction of social networks. Consequently, the proposed method is easily incorporated into existing methods of social network extraction; it serves to enrich such networks by adding relation labels. In addition, the proposed method is entirely unsupervised. For that reason, our method requires neither a priori definition of relations nor preparation of large annotated corpora. It also requires no instances of relations as initial seeds for weakly supervised learning.

Identifying underlying relations is also important in ontology development. Recent studies have shown that social networks and collective knowledge contribute greatly to ontology extraction [2]. Because relation labels assigned to pairs of entities in social network can be regarded as non-taxonomic relations between instances, our work can be regarded as a specific case of ontology population in the context of social networks.

The remainder of this paper is structured as follows. Section 2 describes basic ideas of our approach and detailed steps of the proposed method. Section 3 describes our experiment. Section 4 describes results and evaluation. Section

5 compares our approach to other ongoing relevant research in social network extraction, relation extraction, and ontology population for the Semantic Web. We end our presentation with a discussion of future work, after which we provide concluding remarks in section 6.

2   Method

2.1   Problem Setting

In this paper, as an exemplary scenario for our approach, we use a political social network. Many studies of social network extraction from the Web have addressed researchers or students as entities [4,7,5]. Those individuals are easy for researchers to evaluate: they typically provide more than sufficient relational evidence (e.g., co-authors of a paper, co-members in a project, co-participants in a conference) through Web-based materials. Relations among political entities (e.g., politicians, geopolitical-entities) are also widely various; information for clues of relations is readily available from the Web (e.g., news sites, weblogs). In fact, political and social networks are one research target of social network analyses[1] Therefore, it is worthwhile to examine political social networks.

Figure 1 shows an automatically extracted social network from the Web using Mika and Matsuo's method [4,5]. The social network, including two types of political entities (a politician and a geopolitical location (GPE)), was extracted according to co-occurrence of two types of entities (politician-politician, politician-location) on the Web. In the network, a circular node represents a location entity and an elliptical node represents a political entity. Each edge in the network implies that there is a relation between entities. Given the social network, our task is to extract descriptive relation labels between entities in that social network. In particular, as an example of our approach, we address the relations between politicians and locations. Various relations exist among politicians and locations, for example "born in", "originally from", "elected in", "representing", and so on. These relations between politicians and locations have also been addressed in relation-extraction tasks of natural language processing and information extraction.

Given entity pairs in the social network (e.g., George W. Bush – United States, Junichiro Koizumi – Japan, etc.), our present goal is to extract labels to describe the relations of respective entity pairs (to discover relevant terms that relate a politician to a location). In the following section, we explain our basic idea to this purpose.

2.2   Concept

A simple approach to extract the labels that are useful for describing relations in social networks is to analyze the surrounding local context in which entities of

---

[1] Two focused sessions exist in the international social network conference (Sunbelt): politics and networks, politics and network structures.
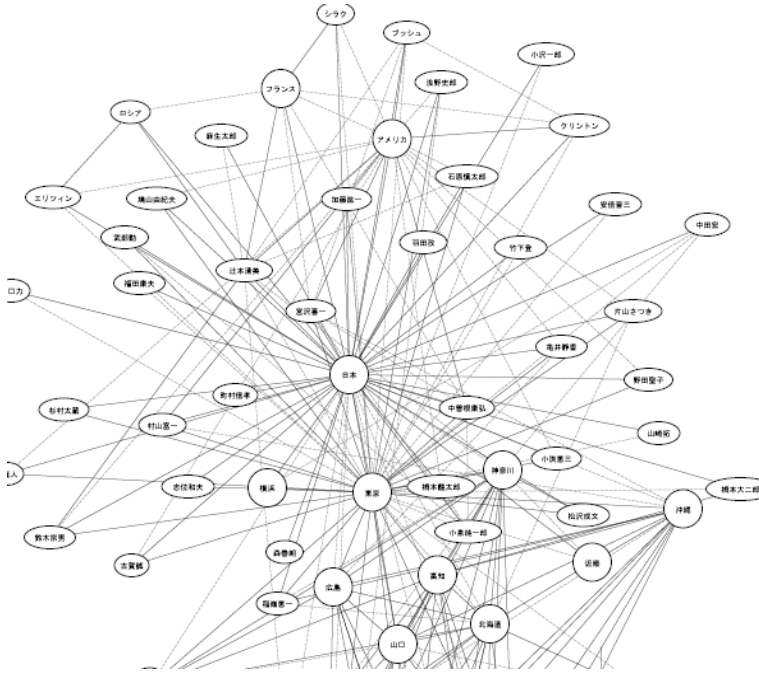
Fig. 1. Political social network extracted from the Web: a circular node represents a location entity and an elliptical node represents a political entity. Each edge in the network implies a relation between entities.

interest co-occur on the Web, and to seek clues to describe that relation. Local context is often used to identify entities or relations among entities in tasks of natural language processing or information extraction [12,13,14].

Table 1 shows keywords [2] that were extracted from local contexts of four entity pairs (Junichiro Koizumi – Japan, Yoshiro Mori – Japan, Junichiro Koizumi – Kanagawa, Yoshiro Mori – Ishizuka [3]). Keywords were extracted from the collective local contexts where co-occurrence of each entity pair was found. For each entity pair, the local contexts from 100 Web pages were collected. The keywords are ordered according to TF-IDF-based scoring, which is a widely used method in many keyword extraction methods to score individual words within text documents to select concepts that accurately represent the documents' contents. The keywords scored by TF-IDF can be considered as a bag-of-words model to represent the local context surrounding an entity pair.

---

[2] In our experiment, we mainly used Web pages in Japanese. Therefore, keywords in the table are translated from their original Japanese. The keyword beginning with a capital letter represents a Japanese proper noun.

[3] Junichiro Koizumi is the current Prime Minister of Japan and Yoshiro Mori is a former Prime Minister. Kanagawa is the prefecture where Koizumi was elected and Ishikawa is the prefecture where Mori was elected.

Table 1. Keywords obtained from each local context of four kinds of entity pairs: Junichiro Koizumi – Japan, Yoshiro Mori – Japan, Junichiro Mori – Kanagawa, and Yoshiro Mori – Ishikawa

(1) Junichiro Koizumi – Japan
pathology, Fujiwara, **prime minister**, Koizumi, Kobun-sha, politics, prime minister, visit, page,**prime minister**, products, **cabinet**, citizen, reform, **minister**, Warsaw, United States, Yasukuni, Yasukuni Shrine, revitalization, society
(2)Yoshiro Mori – Japan
rugby, **prime minister**, chairman, bid, **minister**, association, science, administration, **prime minister**, director, soccer, Africa, world, universe, competition, page, sport, gaffes, media, **cabinet**, director
(3) Junichiro Koizumi – Kanagawa
election, **prime minister**, Yokosuka, **candidate**, **congressional representative**, Saito, **Liberal Democratic Party**, Miura, Koizumi, Democratic Party, lower house, page, fair adversary, politics, endorsement, Liberal Democratic Party, house, president, running in a election, by-elections, constituent
(4) Yoshiro Mori – Ishikawa
Ichikawa, Yasuo, **prime minister**, **election**, Liberal Democratic Party, Okuda, **candidate**, Komatsu, **congressional representative**, **Liberal Democratic Party**, Yuji, Nomi, Kaga, Kanazawa, Nishimura, Page, Shinshin, answer, Matsutou, Komeito, winning in a election

We find that some keywords can serve as relevant labels to describe relations of an entity pair. However, other noise keywords that are irrelevant to describe the relations are also included because the keywords were extracted from collective local contexts of various kinds of Web pages. Using this simple approach, no additional information to decide relevant relation labels for entity pairs exists aside from the TF-IDF scoring. Therefore, we must find another clue to select relevant keywords for relation labels.

From a slightly different perspective, if we examine the common keywords (shown in bold typeface in the table) shared by (1) and (2), we note that the keywords that describe the relations of each entity pair, such as "prime minister" and "cabinet", are commonly shared. In fact, Koizumi and Mori are the current and former prime ministers of Japan. Similarly, if we look at common keywords of (3) and (4), we find that the keywords that describe the relations of each entity pair such as "election" and "candidate" are shared. In fact, Koizumi was elected in Kanagawa, a prefecture, and Mori was elected in Ishikawa. In contrast, if we compare Koizumi's keywords (1) with another of his keywords (3), we find that different keywords appear because of their respective links to different locations: Japan and Kanagawa (although both keywords are Koizumi's.).
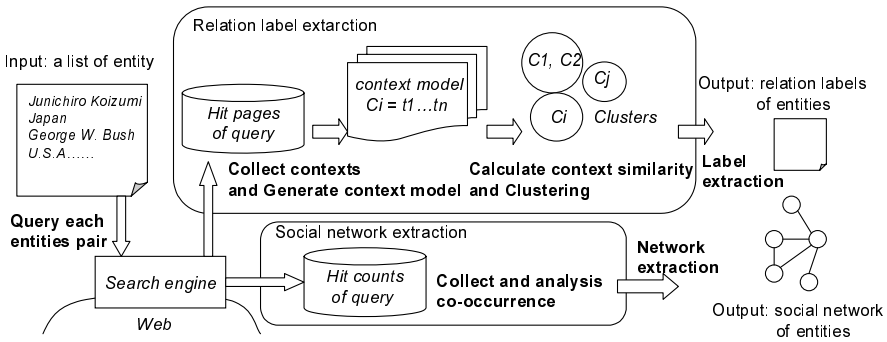
Fig. 2. Outline of the proposed method

Based on the observations described above, we hypothesize that if the local contexts of entity pairs in the Web are similar, then the entity pairs share a similar relation. Our hypothesis resembles previously tested hypotheses related to context [15,14]: words are similar to the extent that their contextual representations are similar. According to that hypothesis, our method clusters entity pairs according to the similarity of their collective contexts. Then, the representative terms in a cluster are extracted as labels to describe the relations of each entity pair in the cluster, assuming that each cluster represents different relations and that the entity pair in a cluster is an instance of a certain relation. The key point of our method is that we determine the relation labels not by examining the local context of one single entity pair, but by the collective local contexts of all entity pairs of interest. In the following section, we explain the precise steps of our proposed method.

2.3   Procedure

Our method for extraction of relation labels in social networks includes the following steps.

1. Collect co-occurrence information and local context of an entity pair
2. Extract a social network that is composed of entity pairs.
3. Generate a context model of each entity pair.
4. Calculate context similarity between entity pairs.
5. Cluster entity pairs.
6. Select representative labels to describe relations from each cluster.

Figure 2 depicts the outline of our method. Our method requires a list of entities (e.g., personal name, location name) to form a social network as the input; it then outputs the social network and a list of relation labels for each entity pair. Although collection of a list of entities is beyond the scope of this paper, one might use named entity recognition to identify entities and thereby generate a list of entities of interest.

The first step is to collect co-occurrence and local contexts of each entity pair from the Web. Many existing methods of social network extraction use a search engine and its resultant query hit counts to obtain co-occurrence information of entities from the Web [Matsuo, Mika]. In line with such methods, we use Google [4] to collect co-occurrence information and generate a social network, as shown in Fig. 1.

Using co-occurrence information, we also collect local contexts in which elements of an entity pair of interest co-occur within a certain contextual distance of one another within the text of a Web page. For this, we downloaded the top 100 web pages included in the search result of corresponding search query to each entity pair (in our example of a politician and location name, the query is "Junichiro Koizumi AND Japan"). This can be accomplished in the process of collecting co-occurrence information, which uses search query hit counts.

2.4 Context Model and Similarity Calculation

For each entity pair, we accumulate the context terms surrounding it; thereby, we obtain the contexts of all entity pairs. As the next step, to calculate the similarity between collective contexts of each entity pair, we require a certain model that represents the collected context. In our method, we propose a context model that represents the context using a bag-of-words and a word vector [16]. We define the context model as a vector of terms that are likely to be used to describe the context of an entity pair (e.g., the keywords list shown in Table 1 can be considered as an example of the context model.). A context model $C_{i,j}$ of an entity pair (e_i, e_j) is defined as the set of N terms $t_1, ..., t_N$ that are extracted from the context of an entity pair as $C_{i,j}(n, m) = t_1, ..., t_N$, where both n and m are parameters of the context window size, which defines the number of terms to be included in the context. In addition, m is the number of intervening terms between e_i and e_j; n is the number of words to the left and right of either entity.

Each term t_i in the context model $C_{i,j}(n, m)$ of an entity pair (e_i, e_j) is assigned a feature weight according to TF-IDF-based scoring defined as $tf(t_i) \cdot idf(t_i)$. Therein, $tf(t_i)$ is defined by the term frequency of term t_i in all the contexts of the entity pair (e_i, e_j). Furthermore, $idf(t_i)$ is defined as $log(|C|/df(t_i))+1$, where $|C|$ is the number of all context models and $df(t_i)$ is the number of context models including term t_i. With the weighted context model, we calculate the similarity $sim(C_{i,j}, C'_{i,j})$ between context models according to the cosine similarity as follows: $sim(C_{i,j}, C'_{i,j}) = C_{i,j}C'_{i,j}/(|C_{i,j}||C'_{i,j}|)$.

In our exploratory experiment, we tried probability distribution-based scoring and several similarities such as L1 norm, Jensen-Shannon and Skew divergence [13]. According to those results, TFIDF-based cosine similarity performs well.

2.5 Clustering and Label Selection

Calculating the similarity between the context models of entity pairs, we cluster all entity pairs according to their similarity. This is based on our hypothesis

[4] http://www.google.com

described in Sect. 2.2: the local contexts of entity pairs in the Web are similar, and the entity pairs share a similar relation.

Ideally, the clustering process terminates when it generates a relevant number of clusters that correspond to the number of relations that entity pairs can hold. However, we do not know what kinds of relation pertain. Therefore, we do not know in advance how many clusters we should make. For that reason, we employ hierarchical agglomerative clustering, which is similarity-based and which uses a bottom-up clustering method.

Several clustering methods exist for hierarchical clustering: single linkage, average linkage and complete linkage. We used those different methods in our exploratory experiment. According to those results, complete linkage performs well because it is conservative in producing clusters and does not tend to generate a biased large cluster. In complete linkage, the similarity between the clusters CL_1, CL_2 is evaluated by considering the two most dissimilar elements as follows: $min_{C_{i,j} \in CL_1, C'_{i,j} \in CL_2} sim(C_{i,j}, C'_{i,j})$.

Initially, each entity pair forms its own cluster. Then the clustering algorithm repeats the step that merges the two most similar clusters still available until the cluster quality drops below a predefined threshold. The cluster quality is evaluated according to two measures [17]: the respective degrees of similarity of entity pairs within clusters and among clusters.

After the clustering process terminates and creates a certain number of clusters, we extract the terms from a cluster as labels to describe the relations of each entity pair in the cluster. This is based on the assumption that each cluster represents a different relation and each entity pair in a cluster is an instance of similar relation. The term relevancy, as a cluster label, is evaluated according to a TFIDF-based measure in the same manner as weighting the terms in a context model. However, in this process, the term frequency is determined for all contexts of a cluster. The underlying idea is to extract terms that appear in the cluster, but which do not appear in other clusters. With a cluster CL's labels $l_1, ..., ln$ scored according to the term relevancy, an entity pair, e_i and e_j, that belongs to the CL can be regarded as holding the relations described by $l_1, ..., ln$.

3   Experiment

Using our proposed method, we extracted labels to describe relations of each entity pair in a social network. We chose 143 distinct entity pairs (a politician and a GPE) that comprise the social network shown in Fig. 1. The politicians mainly include chiefs of state of Japan and other countries. The GPE includes locations such as country, prefectural district, and city. Examples of entity pairs are "Junichiro Koizumi – Japan", "George W. Bush – United States of America", and "Shintaro Ishihara – Tokyo".

We created a context model of each entity pair using nouns and noun phrases from parts-of-speech (POS) surrounding entity pairs in a Web page. We exclude stop words, symbols, and highly frequent words. For each entity pair, we download

Table 2. Manually assigned relation labels of entity pairs of "Junichiro Koizumi – Japan", "George W. Bush – United States of America"

Junichiro Koizumi-Japan	prime minister
George W. Bush – United States of America	president, chief of state

the top 100 web pages in the process of collecting co-occurrence information for extraction of social network. For the context size, we used two parameters, m and n, as explained in Sect. 2.4. As a baseline of the context size, we assigned 10 and 5, respectively, to m and n.

We used complete-linkage agglomerative clustering to cluster all entity pairs. Thereby, we created five distinct clusters according to the predefined thresholds of two quality measures within the clusters and among the clusters, as explained in Sect. 2.5. To evaluate the clustering results and the extracted labels, two human subjects analyzed the context terms of each entity pair and manually assigned the relation labels (three or fewer possible labels for each). Examples of manually assigned relation labels of the entity pair of "Junichiro Koizumi – Japan", "George W. Bush – United States of America" are shown in Table 2. Then, a cluster label was chosen as the most frequent term among the manually assigned relation labels of entity pairs in the cluster. The manually assigned relation labels are used as ground truth in the subsequent evaluation stage.

In Table 3 [5], the left column shows the label of each cluster. The right column shows the highly scored terms that are extracted automatically from each cluster. They can be considered as the labels that describe relations of each entity pair in the cluster. The terms are sorted by relevancy score.

4   Evaluation

We first evaluated the clustering results. For each cluster cl, we counted the number of entity pairs $EP_{cl,correct}$ whose manually assigned relation labels included the label of cluster cl. We also counted the entity pairs $EP_{cl,total}$ in the cluster cl. Next, for each relation label l, we counted the number of entity pairs $EP_{l,correct}$ that have the relation label l whose cluster label is l. We also counted the entity pairs $EP_{l,total}$ that have the relation label l. Then, precision and recall of the cluster were calculated as:

$$precision = \Sigma_{cl \in CL} \frac{EP_{cl,correct}}{EP_{cl,total}}, \ recall = \Sigma_{l \in L} \frac{EP_{l,correct}}{EP_{l,total}}.$$

According to *precision* and *recall*, we evaluated clusters based on the F measure as $F = 2 * precision * reall/(precision + recall)$.

The graph depicted in Fig. 3 shows that the clustering results vary depending on the context size. Consequently, to find the optimal context size, we calculate

---

[5] In our experiment, we mainly used Web pages in Japanese. Therefore, keywords in the table are translated from their original Japanese.

Table 3. Cluster label and automatically extracted relation labels from a cluster

1 mayor	mayor, citizen, hosting, president, affairs, officer, matter, answer, city, conference
2 president	president, administration, world, Japan, economics, policy, war, principle, politics, Iraq
3 prime minister	prime minister, administration, politics, article, election, prime minister, government, peace
4 governor	prefectural governor, governor, president, prefectural-government, committee, Heisei, prefectural administration, mayor, comment, prefectural assembly
5 congressional representative	congressional representative, election, Liberal Democratic Party, candidate, lower house, Democratic Party, proportional representation

the F-measure by changing two size parameters: m and n. Expanding the context size from the minimum, the F-measure takes an optimal value when m is around 30 and n is around 10 (Fig. 3 and Table 4) . We employed this optimal context size to extract the relation labels in our experiment. After reaching the peak, the value of the F-measure decreases as the context size increases. The wider context window tends to include noise terms that are not appropriate to represent the context, thus rendering the similarity calculation between the contexts irrelevant. The optimal context size depends on the structural nature of language. Consequently, we must choose the context size carefully when applying our methods to a different language.

To evaluate the automatically extracted relation labels, we compared the cluster label (left column of Table 3) with the automatically extracted relation labels (right column of Table 3). We found that the relation label that has the highest score is equal to the corresponding cluster's relation label. Precision of the clustering results in our experiment is quite high, as shown above. Therefore, we can say that each entity pair in a cluster is represented properly by the highest-scored relation label from the cluster. In addition, if we examine other automatically extracted relation labels, we find that various terms that represent the relations are extracted.

5   Related Work

Aiming at extracting underlying relations in social networks from the Web, our method is related closely to existing extraction methods of social networks. Several studies have addressed extraction of social networks automatically from various sources of information such as the Web, e-mail, and contacts [6,7,8,9,4]. While most approaches for social network extraction have focused on the strength of the relation, few studies have addressed automatic identification of underlying relations. Matsuo et al. employed a supervised machine learning method to classify four types of relations in a research community [5]. There have also been several important works that have examined supervised learning of relation extraction
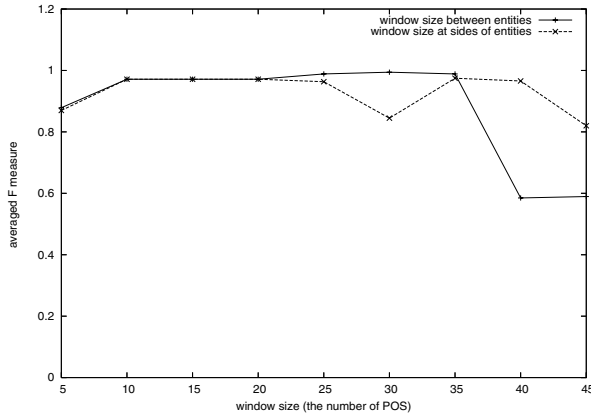
Fig. 3. F measure of clustering results vs. Context window size with two parameters: one is the number of intervening terms between entities and another is the number of words to the left and right of either entity

Table 4. Clustering performance in parameters of context window size with two parameters: m and n

Context window size n, m	Precision	Recall	F-measure
$n = 10$, $m = 30$	0.992	0.995	0.994
$n = 5$, $m = 10$	0.88	0.85	0.86
All terms in a Web page	0.76	0.677	0.716

in the field of natural language processing and information extraction [18,19,20]. However, a supervised method requires large annotated corpora, which cost a great deal of time and effort. In addition, it is necessary to know the domain specific knowledge to define extracted relations a priori. Our method is fully unsupervised and requires no annotated corpora. Furthermore, our method works domain independently and requires no pre-defined relations. For further improvement of our method, it might be worth considering exploitation of weakly supervised and bootstrapping methods [21,22] that rely on a small set of pre-defined initial seeds instead of a large annotated corpus.

Because recent studies have shown that social networks greatly contribute to ontology extraction [2], identifying underlying relations is important for ontology development. Currently, several studies are examining the use of relation extraction for ontology learning and population [23]. Although ontology learning and population share the common goal of facilitating ontology construction, they differ slightly. Whereas ontology learning mainly addresses extraction of taxonomic relations among concepts, the goal of ontology population is extraction of non-taxonomic relations among instances of concepts [24]. In our case, because the labels (non-taxonomic relations) of relations are assigned to pairs of entities of social networks (relation instances), our work can be regarded as a specific case of ontology population in the context of social networks.

Relation extraction for ontology population is typically an unsupervised approach. Because ontology population is usually intended to extract information about instances from large and heterogeneous sources such as the Web, a fully supervised approach that assumes numerous training instances is not feasible for large-scale exploitation, as pointed out in some precedent studies [25]. Therefore, several studies have exploited unsupervised or semi-supervised approaches. Particularly, the current approaches for relation extraction in ontology population are classifiable into two types: those that exploit certain patterns or structures, and those that rely on contextual features.

Pattern-based approaches [26,27,28] seek phrases or sentence structures that explicitly show relations between instances. However, most Web documents have a very heterogeneous structure, even within individual web pages. Therefore, the effectiveness of the pattern-based approach depends on the domain to which it is applied. Rather than exploiting patterns or structures, context-based approaches [29,30,31] assess contextual syntactic, semantic, and co-occurrence features. Several studies have employed contextual verb arguments to identify relations in text [29,31], assuming that verbs express a relation between two ontology classes that specify a domain and range. Although verbs are relevant features to identify relations, we assume that syntactic and dependency analyses are applicable to text collections. Because the Web is highly heterogeneous and often unstructured, syntactic and dependency structures are not always available. For that reason, we employed a contextual model that uses a bag-of-words to assess context. Therefore, the method is applicable to any unstructured documents in the Web. As shown in our experiment, the simple context model performed well to extract descriptive relation labels without depending on any syntactic features in text.

Aiming at extraction of the relation labels in automatically extracted social network from the Web, our method is a Web mining method. Recent approaches of Web mining toward the Semantic Web use the Web as a huge language corpus and combine it with a search engine. This trend is observed not only in recent social network extraction [4,5] but also in ontology population for entities [32,33] and relations [27,34]. The underlying concept of these methods is that it uses globally available Web data and structures to annotate local resources semantically to bootstrap the Semantic Web. In line with this, our approach utilizes the Web to obtain the collective contexts that engender extracting representative relations in social network. As pointed in [35], we claim that relations should be defined not by local information, but rather by a global viewpoint of a network composed of individual relations.

6   Conclusions and Future Work

We propose a method that automatically extracts labels that describe relations between entities in social networks. The proposed method is entirely unsupervised and domain-independent; it is easily incorporated into existing extraction methods of social networks.

Future studies will explore the possibilities of extending the proposed method to relations in other types of social networks. Enriching social networks by adding relation labels, our method might contribute to several social network applications such as finding experts and authorities, trust calculation, community-based ontology extraction, and end-user ontology.

References

1. Golbeck, J., Hendler, J.: Accuracy of metrics for inferring trust and reputation in semantic web-based social networks. In: Proceedings of the 14th International Conference on Knowledge Engineering and Knowledge Management (EKAW). (2004)
2. Mika, P.: Ontologies are us: A unified model of social networks and semantics. In: Proceedings of the 4th International Semantic Web Conference (ISWC). (2005)
3. Brickley, D., Miller, L.: Foaf vocabulary specification. namespace document. (2005)
4. Mika, P.: Flink:semantic web technology for the extraction and analysis of social networks. Journal of Web Semantics **3**(2) (2005)
5. Matsuo, Y., Mori, J., Hamasaki, M., Ishida, K., Nishimura, T., Takeda, H., Hashida, K., Ishizuka, M.: Polyphonet: An advanced social network extraction system. In: Proceednings of the 15th International Word Wide Web Conference (WWW). (2006)
6. Kautz, H., Selman, B., Shah, M.: The hidden web. AI Magazine **18**(2) (1997) 27–36
7. Adamic, L.A., Adar, E.: Friends and neighbors on the web. Social Networks **23**(3) (2003)
8. Harada, M., Sato, S., Kazama, K.: Finding authoritative people from the web. In: Proceedings of the Joint Conference on Digital Libraries (JCDL). (2004)
9. Culotta, A., Bekkerman, R., McCallum, A.: Extracting social networks and contact information from email and the web. In: Proceedings of the 1st Conference on Email and Anti-Spam (CEAS). (2004)
10. Scott, J.: Social Network Analysis: A Handbook. Sage Publications, London (2000)
11. Wasserman, S., Faust, K.: Social network analysis. Methods and Applications. Cambridge University Press, Cambridge (1994)
12. Grefenstette, G.: Explorations in Automatic Thesaurus Construction. Kluwer (1994)
13. Lin, D.: Automatic retrieval and clustering of similar words. In: Proceedings of COLING-ACL98. (1998)
14. Schutze, H.: Automatic word sense dicrimination. Computational Linguistics **24**(1) (1998)
15. Harris, Z.: Mathematical Structures of Language. Wiley (1968)
16. Raghavan, V., Wong, S.: A critical analysis of vector space model for information retrieval. Journal of the American Society for Information Retrieval **35**(5) (1998)
17. Kannan, R., Vempala, S., Vetta, A.: On clustering: Good, bad and spectral. Computer Science (2000)
18. Zelenko, D., Aone, C., Richardella, A.: Kernel methods for relation extraction. Machine Learning Research **2003**(2) (2003)
19. Culotta, A., Sorensen, J.: Dependency tree kernel for relation extraction. In: Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics (ACL). (2004)

20. Kambhatla, N.: Combining lexical, syntactic, and semantic features with maximum entropy models for extracting relations. In: Proceedings of ACL. (2004)
21. Brin, S.: Extracting patterns and relations from the world wide web. In: Proceedings of the WebDB Workshop at 6th International Conference on Extending Database Technology (EDBT). (1998)
22. Agichtein, E., Gravano, L.: Extracting relations from large plain-text collections. In: Proc. of the 5th ACM International Conference on Digital Libraries (ACMDL00). (2000) 85–94
23. Buitelaar, P., Cimiano, P., Magnini, B.: Ontology Learning from Text: Methods, Evaluation and Applications. IOS Press, Amsterdam (2005)
24. Maedche, A.: Ontology Learning for the Semantic Web. Kluwer (2002)
25. Cimiano, P.: Ontology learning and populations. In: Proceedings of the Dagstuhl Seminar Machne Learning for the Semantic Web. (2005)
26. Velardi, P., Navigli, R., Cuchiarelli, A., Neri, F.: Evaluation of ontolearn, a methodology for automatic population of domain ontologies. In: P. Cimiano, and B. Magnini, editors, Ontology Learning from Text: Methods, Applications and Evaluation. IOS Press. (2005)
27. Geleijnse, G., Korst, J.: Automatic ontology population by googling. In: Proceedings of the 17th Belgium-Netherlands Conference on Artificial Intelligence (BNAIC). (2005)
28. Ciravegna, F., Chapman, S., Dingli, A., Wilks, Y.: Learning to harvest information for the semantic web. In: Proceednings of the 1st European Semantic Web Symposium. (2004)
29. Kavalec, M., Maedche, A., Svatek, V.: Discovery of lexical entries for non-taxonomic relations in ontology learning. In: Van Emde Boas, P., Pokorny, J.,Bielikova, M.,Stuller, J. (eds.). SOFSEM 2004. (2004)
30. Cimiano, P., Volker, J.: Towards large-scale open-domain and ontology-based named entity classification. In: Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP). (2005)
31. Schutz, A., P.Buitelaar: Relext: A tool for relation extraction from text in ontology extension. In: Proceedings of the 4th International Semantic Web Conference (ISWC). (2005)
32. Cimiano, P., Ladwig, G., Staab, S.: Gimme' the context: Context-driven automatic. semantic annotation with c-pankow. In: Proceednings of the 14th International Word Wide Web Conference (WWW). (2005)
33. Etzioni, O., Cafarella, M., Downey, D., Kok, S., Popescu, A., Shaked, T., Soderland, S., Weld, D., Yates, A.: Web-scale information extraction in knowitall(preliminary results). In: Proceednings of the 13th International Word Wide Web Conference (WWW). (2004)
34. Boer, V., Someren, M., Wielinga, B.: Extracting instances of relations from web documents using redundancy. In: Proceedings of the 3rd European Semantic Web Conference (ESWC). (2006)
35. Matsuo, Y., Hamasaki, M., Takeda, H., Mori, J., Danushka, B., Nakamura, H., Nishimura, T., Hashida, K., Ishizuka, M.: Spinning multiple social network for semantic web. In: Proceedings of the 21st National Conference on Artificial Intelligence (AAAI). (2006)