Future Generation Computer Systems 37 (2014) 468–477

Contents lists available at ScienceDirect

Future Generation Computer Systems

journal homepage: www.elsevier.com/locate/fgcs

Mining temporal explicit and implicit semantic relations between entities using web search engines

Zheng Xu [a,*], Xiangfeng Luo [b], Shunxiang Zhang [c], Xiao Wei [b], Lin Mei [a], Chuanping Hu [a]

[a] The Third Research Institute of Ministry of Public Security, Shanghai, China
[b] Shanghai University, Shanghai, China
[c] Anhui University of Science and Technology, China

HIGHLIGHTS

- Defining the problem of generating temporal semantic relation between entities.
- Automatically generating structured temporal semantic relation between entities.
- Helping users understand and explore the semantic relation between entities.

ARTICLE INFO

Article history:
Received 20 December 2012
Received in revised form
17 June 2013
Accepted 26 September 2013
Available online 10 October 2013

Keywords:
Temporal semantic relation
Semantic annotation
Content analysis
Web mining

ABSTRACT

In this paper, we study the problem of mining temporal semantic relations between entities. The goal of the studied problem is to mine and annotate a semantic relation with temporal, concise, and structured information, which can release the explicit, implicit, and diversity semantic relations between entities. The temporal semantic annotations can help users to learn and understand the unfamiliar or new emerged semantic relations between entities. The proposed temporal semantic annotation structure integrates the features from IEEE and Renlifang. We propose a general method to generate temporal semantic annotation of a semantic relation between entities by constructing its connection entities, lexical syntactic patterns, context sentences, context graph, and context communities. Empirical experiments on two different datasets including a LinkedIn dataset and movie star dataset show that the proposed method is effective and accurate. Different from the manually generated annotation repository such as Wikipedia and LinkedIn, the proposed method can automatically mine the semantic relation between entities and does not need any prior knowledge such as ontology or the hierarchical knowledge base. The proposed method can be used on some applications, which proves the effectiveness of the proposed temporal semantic relations on many web mining tasks.

1. Introduction

Mining semantic relations between entities plays an important role in many tasks including web mining [1], information retrieval [2], semantic web [3–5], learning technologies [6,7], and web search [8]. In fact, the huge number of real-world entities is organized by numerous semantic relations in the World Wide Web.

Mining semantic relations between entities has received much attention lately. In open information extraction systems such as TextRunner [9], the goal is to mine a large set of relations without the need for any human knowledge and operations. In the object level mining system such as Renlifang [10], the relations between people are given to the users. An entity relation graph can be generated automatically to link the related persons. TREC retrieval conference [11] performs an open relation extraction task every year. The extracted relations can be used in drug reaction [12], bio medical [13], and policy analysis [14]. Given the dynamics, huge scale, and unstructured feature of web, mining semantic relations between entities is still a challenge [15]. The reasons are as follows.

(1) *Explicit semantic relation*. The explicit semantic relations mean the entities co-occurrence in the same repository or corpus. For example, "apple" and "iPhone" may appear in the same web page about the apple computer company. But because of the different description of natural language, a semantic relation can be expressed in different patterns. For example, consider the *Birthplace* relation between a person and a place.

* Corresponding author. Tel.: +86 13817917970.
E-mail addresses: xuzheng@shu.edu.cn (Z. Xu), luoxf@shu.edu.cn (X. Luo), sxzhang@shu.edu.cn (S. Zhang), xwei@shu.edu.cn (X. Wei), l_mei72@hotmail.com (L. Mei), cphu@vip.sina.com (C. Hu).

0167-739X/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.future.2013.09.027

One pattern for this relation is *Person is born in Place*, and other pattern is *Person's birthplace is Place*. Moreover, two entities might have more than one semantic relation. For example, after a *Birthplace* is established between a person and a place, a *Live-in* relation may exist between them. The different lexical syntactic patterns and the different relations between the same pair of entities bring the parsing and mining challenge to extract semantic relations.

(2) *Implicit semantic relation*. The implicit semantic relations mean the entities do not appear in the same repository or corpus. But some entities can be the bridge for mining an implicit relation. In other words, a pair of entities cannot be linked by a direct semantic relation. For example, John is the head of IBM. David is the manager of Microsoft. Suppose John will attend a conference and he notices that David will attend the same conference. Assume that John does not know David and would like to chat with him. John searches the page of David in LinkedIn.[1] He finds David worked for IBM before joining Microsoft. Thus, John can easily find out the implicit relation between him and David: they both worked at IBM. The implicit relation adds the difficulty for the mining procedure since the existing co-occurrence mining method cannot perform.

(3) *Temporal semantic relation*. The temporal semantic relations mean the entities contain different relations in different time intervals. For example, "Barack Obama" is the president of United States. He has four experiences including the president, US Senator, State Senator, and Senior lecturer in law. Thus, the entity "Barack Obama" might be related to different entities in different time intervals. The temporal feature of semantic relation should be considered in the mining procedure.

In order to mine the accuracy semantic relations between entities, in this paper, we study the problem of automatically generating temporal semantic relations (TSR) for entities. We want to provide temporal semantic relations between entities, which indicate the semantics and hidden association of entities. Besides considering the temporal feature of semantic relation, we also divide the semantic relation into two aspects: explicit and implicit. The explicit semantic relations mean the entities co-occurrence in the same repository or corpus. The implicit semantic relations can be related by some other entities though they do not appear together.

Despite the mining method of semantic relations, an understandable and clear annotation of relations is also important. What is a good annotation of the semantic relations between entities? Let us see two examples from IEEE[2] and Renlifang,[3] which are shown in Figs. 1 and 2. In Fig. 1, we search for "Zheng Xu" and "Xiangfeng Luo" in IEEE. The returned result shows that "Zheng Xu" and "Xiangfeng Luo" have a *co-author* relation. From Fig. 1, we can see that the annotation for a semantic relation is structured as follows. First, the co-occurrence context is presented. For example, "Zheng Xu" and "Xiangfeng Luo" are both the authors of the paper "Building association link network for semantic link on web resources". Second, the faceted information is given to show the different aspects of the relation. For example, "Zheng Xu" and "Xiangfeng Luo" are both the authors of two different papers. Besides, the publication time of the paper is given.

Similarly, if we can provide the structured and semantic related information of a pair of entities for a user, it will be very helpful for her/him to understand and further explore it. Thus, inspired by the annotations from IEEE, the appropriate semantic annotation of a semantic relation should include.

Building Association Link Network for Semantic Link on Web Resources

Xiangfeng Luo; Zheng Xu; Jie Yu; Xue Chen
Automation Science and Engineering, IEEE Transactions on
Volume: 8 , Issue: 3
Topic(s): Robotics & Control Systems
Digital Object Identifier: 10.1109/TASE.2010.2094608
Publication Year: 2011 , Page(s): 482 - 494

Association Link Network: An Incremental Semantic Data Model on Organizing Web Resources

Zheng Xu; Xiangfeng Luo; Wenjun Lu
Parallel and Distributed Systems (ICPADS), 2009 15th International Conference on
Topic(s): Communication, Networking & Broadcasting ; Computing & Processing (Hardware/Software)
Digital Object Identifier: 10.1109/ICPADS.2009.14
Publication Year: 2009 , Page(s): 793 - 798

Fig. 1. The annotation from IEEE.

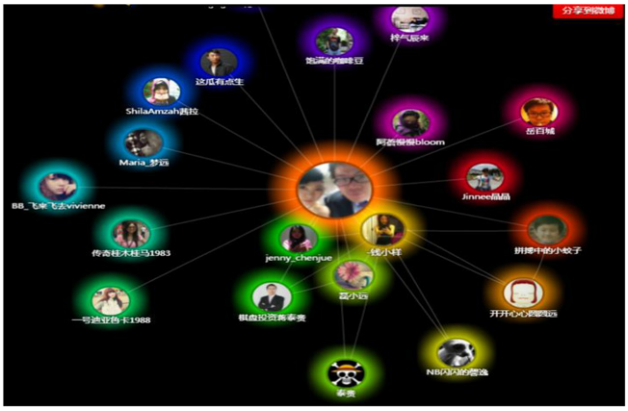

Fig. 2. The annotation from the dictionary.

(1) *Semantic relation context*. Given a semantic relation, the example context can help the users to understand it. Moreover, the example context can help users to use the relations in a real task.

(2) *Faceted information*. Given a pair of entities, we should give the different semantic relations of it, which can show the whole aspects to the user.

(3) *Temporal information*. In the different time interval, a pair of entities might have different semantic relations. The appropriate annotation in a different time interval should be mined.

In Fig. 2, we issue the weibo[4] (a Chinese social website similar to twitter[5]) account of "Zheng Xu" as the query in Renlifang. The nodes in Fig. 2 mean a person and the link between two nodes means these two weibo accounts refer the same information. From Fig. 2, we can see that the annotation for a semantic relation in Renlifang is a graph based structure. Moreover, some implicit semantic relations can be mined from Fig. 2. For example, "Shila Amzah" and "Jinnee" are both have the *reference* relation with "Zheng Xu". So the implicit relation between them is they are both friends of "Zheng Xu". We add two components to the annotations of a semantic relation.

(4) *Connection entity*. Given a pair of entities, we can give the connection entity to reflect their implicit relation.

(5) *Visual thesaurus graph*. Given a semantic relation, we can give a graph based entity network.

[1] www.linkedin.com.
[2] www.ieee.org.
[3] www.renlifang.msra.cn.
[4] www.weibo.com.
[5] www.twitter.com.