Wikification of Concept Mentions within Spoken Dialogues Using Domain Constraints from Wikipedia

Seokhwan Kim, Rafael E. Banchs, Haizhou Li
Human Language Technology Department
Institute for Infocomm Research
Singapore 138632
{kims,rembanchs,hli}@i2r.a-star.edu.sg

Abstract

While most previous work on Wikification has focused on written texts, this paper presents a Wikification approach for spoken dialogues. A set of analyzers are proposed to learn dialogue-specific properties along with domain knowledge of conversations from Wikipedia. Then, the analyzed properties are used as constraints for generating candidates, and the candidates are ranked to find the appropriate links. The experimental results show that our proposed approach can significantly improve the performances of the task in human-human dialogues.

1 Introduction

Linking mentions in natural language to the relevant concepts in knowledge-bases plays a key role in better understanding the meanings of expressions as well as further populating knowledge-bases with less human effort. Especially, Wikipedia has been widely used as a major target resource for linking. Most previous work on this Wikipedia-based linking task called Wikification (Mihalcea and Csomai, 2007) has focused on resolving ambiguities and variabilities of the expressions in written texts including newswire collections (McNamee and Dang, 2009; Ji et al., 2010; Ji et al., 2014) or microblog posts (Genc et al., 2011; Cassidy et al., 2012; Guo et al., 2013; Huang et al., 2014).

But writing and reading are not the only ways for exchange of information, since many communications between people in real life are performed through spoken dialogues also. Thus, we could expect to improve the understanding capabilities of applications based on Wikification and broaden the coverage of the contents in knowledge-bases, if Wikification is successfully performed also for human-human spoken conversations.

In this work, we focus on the following differences between spoken dialogues and written texts as sources for Wikification. Firstly, at least two speakers are engaged in a dialogue session, while the texts in newswire or microblogs are mostly written by a single author. Thus, the viewpoint of each speaker should be considered separately or jointly depending on the situation. Secondly, the correspondence between mentions and concepts in spoken dialogues tends to be dependent not only on the contexts explicitly mentioned in a given dialogue, but also on other information inferred by speakers based on their background knowledge. The other difference is that spoken utterances are more likely to be informal and noisy than written sentences, which makes expressions more ambiguous and variable.

To solve these issues, we propose a three step approach for Wikification on spoken dialogues. At the first step, a set of classifiers are used for analyzing the dialogue-specific aspects of a given mention. According to the analyzed results, the criteria in selecting concept candidates is determined, and then a ranking is performed on the filtered candidates to identify the concept that is the most relevant to the mention.

While many researchers have worked on linking named-entities (Bunescu and Pasca, 2006; Cucerzan, 2007; McNamee and Dang, 2009; Han and Sun, 2011; Han et al., 2011; Ji et al., 2014) or other types of concept mentions (Mihalcea and Csomai, 2007; Milne and Witten, 2008; Ferragina and Scaiella, 2010; Ratinov et al., 2011; Mendes et al., 2011; Cheng and Roth, 2013) to the relevant articles in Wikipedia, all the noun phrases including not only named entities or base noun phrases, but also complex or recursive noun phrases in a dialogue are considered as instances to be linked in this work. For the concept candidates, we divide every article into sub-sections and consider each section as a unit along with article-level concepts.

Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, pages 2225–2229,
Lisbon, Portugal, 17-21 September 2015.

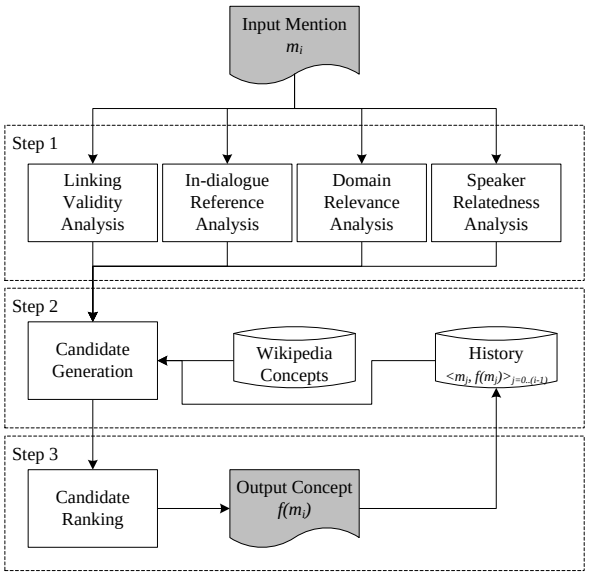

Figure 1: Overall architecture of three-step approach for Wikification on spoken dialogues

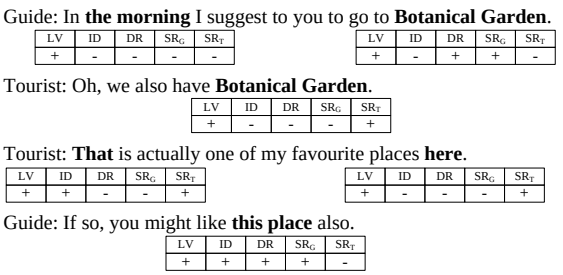

Figure 2: Examples of annotations for mention analysis: SR_G and SR_T denote guide and tourist relatedness, respectively.

2 Method

2.1 Mention Analysis

The first step in our proposed approach (Figure 1) is analyzing the following four types of binary properties of a given mention: linking validity (LV), in-dialogue reference (ID), domain relevance (DR), and speaker relatedness (SR).

Linking validity of the mention is determined by the decision whether it is matched with any Wikipedia concept or not. Since only the mentions assigned with positive validity values are proceeded to the further processes, this classification can be considered as a joint task for target mention identification and NIL detection.

Another type of analysis focuses on the references between the mention and the linking history. If the mention is matched with one in the set of concepts for the previous mentions in the same session, it has a positive value for the in-dialogue reference property.

The other two types of properties are defined for indicating the relevances of the mention to the contents that are specific to the target domain or the profiles of each speaker in the conversation. For these analyses, the whole Wikipedia collection is partitioned into subsets according to the domain or speaker-relevances. In this work, the concepts in these subsets are automatically collected with no manual effort by utilizing the domain knowledge also from Wikipedia. First, we retrieve the 'List' or 'Index' pages in Wikipedia that are related to the topic or the profile of a speaker. Then, all the articles listed on these seed pages are collected and considered as the related concepts in the corresponding sets.

Since every property has a positive or a negative value as a result, each analysis can be considered as a binary classification problem. In this work, we train support vector machines (SVM) (Cortes and Vapnik, 1995) from the dialogues annotated with the corresponding labels as shown in Figure 2 based on the features listed in Table 1.

2.2 Candidate Generation

After analyzing the above property values of a given mention, a set of concepts to be disambiguated are selected from Wikipedia. These candidates are retrieved from a Lucene [1] index on the whole Wikipedia collection with the fields of article title, section title, redirection, category, and body texts. Each query to the search engine is prepared with the combination of the mention phrase and its analyzed properties as constraints for filtering. If the value for in-dialogue reference is positive, the searching is restricted to the set of concepts linked with the previous mentions in the same session. Similarly, the domain relevance and speaker relatedness values provide the filtering condition within the corresponding subsets introduced in Section 2.1.

One practical issue on this candidate generation step is how to combine the multiple constraints when we have more than one positive properties for a given mention. The simplest way is taking the intersection of the corresponding constraints. However, we should consider the fact that the properties assigned automatically can be erroneous, since none of the analyzer is perfect.

[1] http://lucene.apache.org/

Level	Name	Description
Mention	SP	the speaker who spoke that mention
	WM	word n-grams within the mention
	LM	lemma n-grams within the mention
	PM	POS n-grams within the mention
	NE	named entities within the mention
	NP	base noun phrases within the mention
Utterance	BW	the words before the mention
	AW	the words after the mention
	BL	the lemmas before the mention
	AL	the lemmas after the mention
	BP	the POS tags before the mention
	AP	the POS tags after the mention
	IU	whether the mention previously occurs in the same utterance
Dialogue	EO	whether the phrase is previously mentioned in the dialogue history
	EOS	whether the phrase is previously mentioned by the same speaker in the history
Wikipedia	IW	whether the phrase is a title of an entry in Wikipedia
	IWD	whether the phrase is a title of an entry in the set of domain-relevant concepts
	IWS_k	whether the phrase is a title of an entry in the set of k-th speaker-relevant concepts

Table 1: List of features for training the models for mention analysis

For the noisy cases, the intersection-based filtering could be risky, because the errors are also jointly accumulated. To circumvent the impact of errors from the previous step, we also try to use the union of the constraints and compare it with the intersection case later in Section 3.

2.3 Candidate Ranking

In this work, linking a given mention to its most relevant concept is determined by ranking SVM (Joachims, 2002) which is a pairwise ranking algorithm learned from the ranked lists. For each pair of a mention m in the training data and its candidate concept c, the ranking score $s(m, c)$ is assigned as follows:

$$s(m,c) = \begin{cases} 4 & \text{if } c \text{ is the exactly same as } f(m), \\ 3 & \text{if } c \text{ is the parent article of } f(m), \\ 2 & \text{if } c \text{ belongs to the same article} \\ & \text{but different section of } f(m), \\ 1 & \text{otherwise.} \end{cases}$$

where $f(m)$ is the annotation of m in the training dataset. The list of candidates assigned with their scores provides the relative orders for a given mention, and it can be converted into a set of pairwise constraints which are trained by ranking SVM with the features in Table 2.

Name	Description
SP	the speaker who spoke that mention
WM	word n-grams within the surface of m
WT	word n-grams within the title of c
EMT	whether the surface of m is same as the title of c
EMR	whether the surface of m is same as one of re-directions to c
MIT	whether the surface of m is a sub-string of the title of c
TIM	whether the title of c is a sub-string of the m's surface form
MIR	whether the surface of m is a sub-string of a re-directed title to c
RIM	whether a re-directed title to c is a sub-string of the m's surface form
PMT	similarity score based on edit distance between the surface of m and the title of c
PMR	maximum similarity score between the surface of m and the redirected titles to c
OC	whether c previously occurred in the full dialogue history
OC_w	whether c occurred within w previous turns with $w \in \{1, 3, 5, 10\}$

Table 2: List of features for training the ranking SVM model

3 Evaluation

3.1 Data

To demonstrate the effectiveness of our approach to Wikification on spoken dialogues, we performed experiments on a dialogue corpus which consists of 35 sessions collected from human-human conversations in English about tourism in Singapore between actual tour guides in Singapore and tourists from the Philippines. All the recorded dialogues with the total length of 21 hours were manually transcribed, then the 31,034 utterance were pre-processed by Stanford CoreNLP toolkit [2]. Each noun phrase in the constituent trees provided by the parser is considered as an instance for Wikification and manually annotated with the corresponding concept in Wikipedia. 34,949 mentions have been linked to the concepts in Wikipedia.

As a pool for candidate generation, we built a Lucene index based on Wikipedia database dump as of January 2015 which has 4,797,927 articles and 25,577,464 sections in total. From this collection, 11,128 and 27,186 articles have been considered as Singapore-related and Philippines-related concepts, respectively, for the filtering based on domain and speaker relevances.

[2] http://nlp.stanford.edu/software/corenlp.shtml

Features	LV	ID	SR_G	SR_T
M	86.29	69.15	**71.10**	**72.94**
M+U	**86.90**	70.43	70.43	68.85
M+D	86.17	71.09	70.56	71.52
M+W	86.21	68.96	70.66	71.86
M+U+D	86.82	**72.37**	70.12	68.30
M+U+W	86.84	70.13	70.19	68.78
M+U+D+W	86.77	72.20	69.94	68.10

Table 3: Comparisons of the performances in F-measure of mention analyzers with different combinations of features: M,U,D,W denotes mention-level, utterance-level, dialogue-level, and Wikipedia features, respectively

3.2 Mention Analysis

Based on the annotated dialogues, we built four mention analyzers for LV, ID, SR_G, and SR_T, where SR_G is for the guides and SR_T is for the tourists in the conversations. In this work, only the information where each speaker is from was considered as a profile to analyze the speaker-related properties. Since all the guides participated in the data collection are from Singapore and the main topic of the conversations is also about Singapore, we omitted DR which should have the same results as SR_G in the experiments.

For each analyzer, we trained the SVM models using SVM^{light} [3] with the features in Table 1. All the evaluations were performed in five-fold cross validation to the manual annotations with precision, recall, and F-measure.

Table 3 compares the performances of the seven combinations of feature sets for each analyzer. Based on these results, we selected the model that achieved the best performance for each analyzer to process the mentions for the further steps.

3.3 Candidate Generation

For each mention in the corpus, we prepared four sets of candidates with different filtering constraints. While the first baseline set was retrieved with no filtering, the others were generated according to the procedure described in Section 2.2. When more than one positive values were provided from mention analyzers, intersection and union operators were applied for combining multiple constraints. In the last set, the property values manually annotated in the training data were considered as the correct constraints, which is intended for comparing with the others to investigate the influence of errors in mention analysis. For every set, we retrieved top 100 candidates satisfying the given constraints from the Lucene index with Wikipedia collection and added one more special candidate for NIL detection.

Method	P	R	F
No filtering	26.85	22.52	21.24
Intersection	44.37	27.35	33.84
Union	38.04	31.97	34.74
Manual (Oracle)	39.90	34.72	37.13

Table 4: Comparisons of the performances of Wikification on spoken dialogues

3.4 Candidate Ranking

For each set of candidates, we trained a ranking function using SVM^{rank}[4] with the features in Table 2. Both training and testing the ranking models were performed also based on five-fold cross validation with the same divisions as the former evaluation. After getting the ranking results, we took the top-ranked candidate for each list and considered it as a result of Wikification for the corresponding mention.

Table 4 compares the final performances of Wikification obtained by ranking on the candidates generated with different sets of constraints. Both approaches, intersection and union, outperformed the baseline by 12.60 and 13.50 in F-measure, respectively. While the intersection strategy contributed to produce more precise outputs than the others even including the case with manual filtering, the other proposed approach with union achieved more gain in recall with slightly better F-measure than the former one.

4 Conclusions

This paper presented a Wikification approach for spoken dialogues. In this approach, a set of dialogue-specific properties were analyzed for generating concept candidates. Then, supervised ranking was performed on these candidates to identify the relevant concepts. Experimental results show that the proposed constraints help to improve the performances of the task on spoken dialogues.

[3] http://svmlight.joachims.org/

[4] http://www.cs.cornell.edu/people/tj/svm_light/svm_rank.html

References

Razvan C. Bunescu and Marius Pasca. 2006. Using Encyclopedic Knowledge for Named entity Disambiguation. In *EACL*, volume 6, pages 9–16.

Taylor Cassidy, Heng Ji, Lev-Arie Ratinov, Arkaitz Zubiaga, and Hongzhao Huang. 2012. Analysis and Enhancement of Wikification for Microblogs with Context Expansion. In *COLING*, volume 12, pages 441–456.

Xiao Cheng and Dan Roth. 2013. Relational inference for wikification. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pages 1787–1796.

Corinna Cortes and Vladimir Vapnik. 1995. Support-vector networks. *Machine learning*, 20(3):273–297.

Silviu Cucerzan. 2007. Large-Scale Named Entity Disambiguation Based on Wikipedia Data. In *EMNLP-CoNLL*, volume 7, pages 708–716.

Paolo Ferragina and Ugo Scaiella. 2010. Tagme: on-the-fly annotation of short text fragments (by wikipedia entities). In *Proceedings of the 19th ACM international conference on Information and knowledge management*, pages 1625–1628.

Yegin Genc, Yasuaki Sakamoto, and Jeffrey V Nickerson. 2011. Discovering context: Classifying tweets through a semantic transform based on wikipedia. *Foundations of Augmented Cognition*, page 484.

Stephen Guo, Ming-Wei Chang, and Emre Kiciman. 2013. To Link or Not to Link? A Study on End-to-End Tweet Entity Linking. In *HLT-NAACL*, pages 1020–1030.

Xianpei Han and Le Sun. 2011. A generative entity-mention model for linking entities with knowledge base. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1*, pages 945–954. Association for Computational Linguistics.

Xianpei Han, Le Sun, and Jun Zhao. 2011. Collective entity linking in web text: a graph-based method. In *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval*, pages 765–774.

Hongzhao Huang, Yunbo Cao, Xiaojiang Huang, Heng Ji, and Chin-Yew Lin. 2014. Collective tweet wikification based on semi-supervised graph regularization. *Proceedings of the ACL, Baltimore, Maryland*.

Heng Ji, Ralph Grishman, Hoa Trang Dang, Kira Griffitt, and Joe Ellis. 2010. Overview of the tac 2010 knowledge base population track. In *Third Text Analysis Conference (TAC 2010)*.

Heng Ji, H. T. Dang, J. Nothman, and B. Hachey. 2014. Overview of tac-kbp2014 entity discovery and linking tasks. In *Proc. Text Analysis Conference (TAC2014)*.

Thorsten Joachims. 2002. Optimizing search engines using clickthrough data. In *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 133–142.

Paul McNamee and Hoa Trang Dang. 2009. Overview of the tac 2009 knowledge base population track. In *Text Analysis Conference (TAC)*, volume 17, pages 111–113.

Pablo N Mendes, Max Jakob, Andrés García-Silva, and Christian Bizer. 2011. Dbpedia spotlight: shedding light on the web of documents. In *Proceedings of the 7th International Conference on Semantic Systems*, pages 1–8.

Rada Mihalcea and Andras Csomai. 2007. Wikify!: linking documents to encyclopedic knowledge. In *Proceedings of the sixteenth ACM conference on Conference on information and knowledge management*, pages 233–242.

David Milne and Ian H. Witten. 2008. Learning to link with wikipedia. In *Proceedings of the 17th ACM conference on Information and knowledge management*, pages 509–518.

Lev Ratinov, Dan Roth, Doug Downey, and Mike Anderson. 2011. Local and global algorithms for disambiguation to wikipedia. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1*, pages 1375–1384. Association for Computational Linguistics.