Domain Specific Cross-Lingual Knowledge Linking Based on Similarity Flooding

Liangming Pan(✉), Zhigang Wang, Juanzi Li, and Jie Tang

Tsinghua National Laboratory for Information Science and Technology,
Department of Electronic Engineering, Tsinghua University, Beijing 100084, China
{plm,wzhigang,ljz,tangjie}@keg.cs.tsinghua.edu.cn

Abstract. The global knowledge sharing makes large-scale multi-lingual knowledge bases an extremely valuable resource in the Big Data era. However, current mainstream multi-lingual ontologies based on online wikis still face the limited coverage of cross-lingual knowledge links. Linking the knowledge entries distributed in different online wikis will immensely enrich the information in the online knowledge bases and benefit many applications. In this paper, we propose an unsupervised framework for cross-lingual knowledge linking. Different from traditional methods, we target the cross-lingual knowledge linking task on specific domains. We evaluate the proposed method on two knowledge linking tasks to find English-Chinese knowledge links. Experiments on English Wikipedia and Baidu Baike show that the precision improvement of cross-lingual link prediction achieve the highest 6.12 % compared with the state-of-art methods.

Keywords: Knowledge linking · Cross-lingual · Similarity flooding

1 Introduction

In the era of information globalization, sharing knowledge across different languages becomes an important and challenging task. Online encyclopedias, which have already become an indispensable part in people's life for knowledge acquisition, are the primary focus of globalized knowledge sharing. One fundamental research, namely *cross-lingual knowledge linking*, aims at automatically discovering *cross-lingual links* (CLs), i.e., links between articles describing the same subjects in different languages. Inter-wiki cross-lingual links can largely enrich the cross-lingual knowledge and facilitate knowledge sharing across different languages. These CLs also serve as a valuable resource for many applications, including machine translation [16], cross-lingual information retrieval [10,15], and multilingual semantic data extraction [1,3], etc.

However, the problem of cross-lingual knowledge linking is non-trivial and poses a set of challenges. One of the most serious challenges is that lexical similarities, such as the edit distance between articles, are impracticable to be utilized because of the language gap. Recognizing this problem, designing language-independent features (e.g. the link structure of articles) forms the basis of recent

F. Lehner and N. Fteimi (Eds.): KSEM 2016, LNAI 9983, pp. 426–438, 2016.
DOI: 10.1007/978-3-319-47650-6_34

related works for cross-lingual knowledge linking [9,11,13,14]. However, several problems for cross-lingual knowledge linking are still in need of further investigation. First, large number of known CLs are required in all the aforementioned approaches for serving as either training data or seed set. However, in most cases, there are less existing CLs or none at all between different wikis. Clearly, it is time consuming and tedious to annotate all these training CLs. Thus, unsupervised methods need to be developed for cross-lingual knowledge linking. Second, the methods to date have been entirely focus on the general framework for knowledge linking, with less focus on finding CLs in specific domains. In fact, if we focus the knowledge linking task on one specific domain, many domain-specific features (e.g. the properties in the infoboxes of wiki) can be utilized to discover new CLs more accurately.

Based on these considerations, we propose an unsupervised domain-specific framework for cross-lingual knowledge linking. Our model takes the articles of two cross-lingual wikis K and K' as input, and the articles are all from one specific domain D. Each wiki article is a single web page and describes a realworld entity in domain D. The semantic relations between these entities are mostly contained in the wiki infoboxes and hyperlinks between articles. After extracting these semantic relations, we first build up a domain-specific knowledge graph (KG) for each wiki, and then we match the entities between the two KGs via an adaptive variation of the similarity flooding algorithm (SF) [8]. Specifically, we make the following contributions:

1. We propose a novel unsupervised knowledge linking framework for cross-lingual wikis in specific domains. The method does not dependent on any pre-given CLs, which is more practically applicable than previous methods.
2. We propose a variation of the similarity flooding algorithm for entity matching between cross-lingual wikis. The SF is commonly used for ontology alignment and can hardly be applied to entity matching because of the computational challenge. We tackle with this problem via reducing the number of nodes in the pairwise connectivity graph (PCG).
3. We conduct experiments on different wiki data sets. Experimental results show that our method outperforms the state-of-the-art framework for cross-lingual knowledge linking.

The rest of this paper is organized as follows. Section 2 presents some related works. Section 3 presents some basic concepts and the problem formulation. In Sect. 4 we present our detailed approaches. The experimental results are reported in Sect. 5. Finally we conclude our work in Sect. 6.

2 Related Work

Our work is relevant to cross-lingual knowledge linking, which concerns the discover of missing cross-lingual links across online wikis. There exist several related works. Sorg and Cimiano [11] proposed a classification-based approach to infer new CLs between German Wikipedia and English Wikipedia.

Erdmann et al. [4] extracted a dictionary from Wikipedia by analyzing the link structure of Wikipedia. Hassan et al. [6] address the task of cross-lingual semantic relatedness by exploiting the cross-lingual links available between Wikipedia versions in multiple languages. Wang et al. [14] employed a factor graph model which leverages link-based features to find CLs between English Wikipedia and Chinese Wikipedia. All the aforementioned works intend to propose a general framework for cross-lingual knowledge linking. To our best knowledge, our work is the first to focus on discovering CLs in specific domains.

Ontology and instance matching is another related problem. The goal of ontology and instance matching is to find equivalent elements between two heterogeneous semantic data sources. Currently, there exist several systems for ontology matching, such as Silk [12], idMesh [2], SOCOM [5] and RiMOM [7]. The Silk and idMesh focus on monolingual matching tasks, while SOCOM and RiMOM can deal with ontology matching across languages.

3 Preliminaries

In this section, we introduce some basic concepts, and formally define the key problem of domain-specific cross-lingual knowledge linking.

Definition 1. ***Knowledge Graph.*** Let E be a set of entities and R be a set of binary relations. A knowledge graph G is defined as a directed graph whose nodes correspond to entities in E and edges of the form (s, r, t), where $s, t \in E$ and $r \in R$. Each edge (s, r, t) indicates that there exists a relationship r from the entity s to entity t.

Definition 2. An ***Online Wiki*** can be represented as $K = \{a_i\}_{i=1}^{p}$, where a_i is a disambiguated article in K and p is the size of K. A wiki article $a \in K$ is formally defined as a 4-tuple $a = (title, text, info, link)$, where $title$ denotes the title of the article a, $text$ denotes the unstructured text description of a, $info$ is the infobox associated with a and $link$ is the set of hyperlinks in article a (hyperlinks in infoboxes are not count). Specifically, $info = \{(attr_i, value_i)\}_{i=1}^{q}$ represents the list of attribute-value pairs for the article a.

Figure 1 gives an example of these four important elements concerning the article named "Steve Jobs". Given two online wikis, K and K', a *correspondence* between entities $e \in K$ and $e' \in K'$, denoted as $\langle e, e' \rangle$, signifies that e and e' are equivalent. Cross-lingual knowledge linking is the task of finding correspondences between multi-language online wikis, which is formally defined as follows.

Definition 3. **Cross-lingual knowledge linking.** Given two online wikis, K and K', *knowledge linking* is the process of finding correspondences between K and K'. If K and K' are in different languages, we call it the problem of *cross-lingual knowledge linking.*

In our problem, we further choose a domain D and extract all articles of domain D from K and K' to form two new online wikis, denoted as K_D and K'_D. We then define the problem of cross-lingual knowledge linking between K_D and K'_D as *domain-specific cross-lingual knowledge linking.*

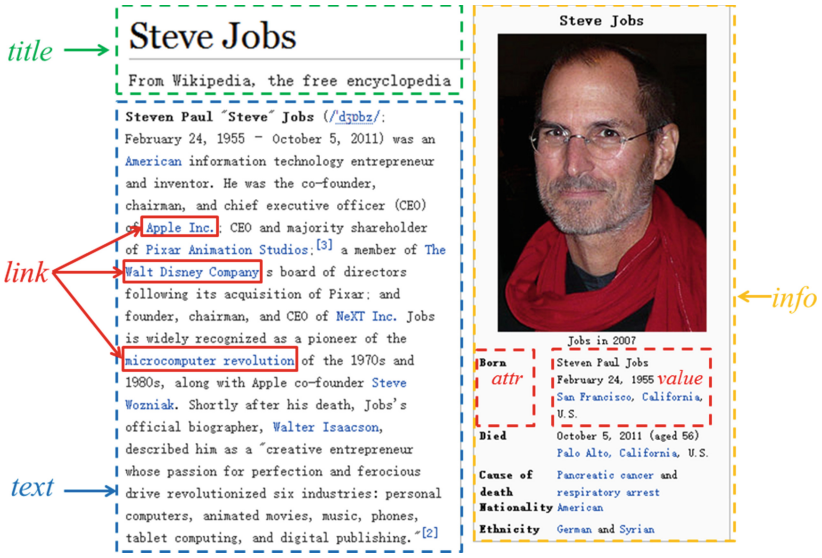


Fig. 1. An example of online wiki articles

4 The Proposed Approach

Figure 2 shows the framework of our proposed approach. There are two major components: *Knowledge Graph Construction* and *Graph-based Knowledge Linking*. *KG Construction* aims to build two knowledge graphs, denoted as G_D and G'_D, for K_D and K'_D. Specifically, based on a common relation set, we extract semantic relations in a structured form of subject-predicate-object triples from infoboxes and hyperlinks of K_D and K'_D. Then, G_D and G'_D are constructed by these triples. The goal of *Graph-based Knowledge Linking* is to discover CLs between G_D and G'_D based on the variation of SF algorithm. Two main processes in this algorithm are PCG construction and similarity propagation. In the following subsections, the *KG Construction* and *Graph-based Knowledge Linking* are described in detail.

4.1 Knowledge Graph Construction

Online wiki's infoboxes contain rich structured information of various entities. Among all the infobox attributes, those attributes having hyperlinks in its values identify semantic relations between entities, which are important for creating domain knowledge graph. Because attribute names are usually annotated by human editors, an attribute often have many surface names in infoboxes of online wiki. For example, in the movie domain, the attributes "Starring" and "Actor List" both refer to the actors of a movie. Furthermore, attributes across wikis may also have same semantic meanings (e.g. "Starring" and "演员表"). Therefore, we need to unify all synonymous attribute names of K_D and K'_D to a

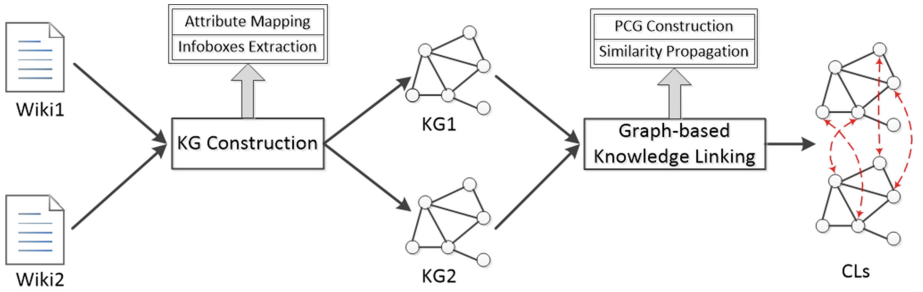


Fig. 2. The framework of the proposed method

disambiguated attribute. We call it the construction of *Attribute Mapping*, which is formally defined as follows:

Definition 4. **Attribute Mapping.** Given two online wikis K_D and K'_D in domain D, let us denote A_D and A'_D as the set of infobox attribute names of K_D and K'_D. An *attribute mapping* is a set of disambiguated attributes, denoted as $AM = \{r_i\}_{i=1}^{q}$. Each $r_i \in AM$ can be represented as a set $r_i = s_i \cup s'_i$, where $s_i \subset A_D$ and $s'_i \subset A'_D$. Attribute names in r_i have identical semantic meanings.

Figure 3 shows an example of attribute mapping of movie domain between English Wikipedia and Baidu Baike[1]. Based on the attribute mapping AM, we can define a function Map to map an attribute name to its corresponding disambiguated attribute in AM.

Fig. 3. An example of attribute mapping in movie domain

Given the Map function, we can build a domain knowledge graph for an online wiki using semantic relations contained in infoboxes. The knowledge graph construction algorithm is presented in Algorithm 1, where K is the input online wiki and G is the output knowledge graph. V and E are the vertex set and edge set of G, respectively. Notice that hyperlinks also represent relations between

[1] http://baike.baidu.com/.

entities. If article a has a hyperlink to article b, there exists some kind of relationship between a and b, thus we also add an edge $(a, <relatedTo>, b)$ to the knowledge graph.

Algorithm 1. Knowledge Graph Construction

Input: K, AM
Output: $G = (V, E)$
$V = \varnothing, E = \varnothing$;
foreach *article* a ***in*** K **do**
  V.add(a);
  foreach $(attr, value)$ ***in*** $a.info$ **do**
    $r = Map(attr)$;
    if $r \neq None$ ***and*** $value \in K$ **then**
      $E.add(< a, r, value >)$;
    end
  end
end
foreach *hyperlink* h ***in*** $a.link$ **do**
  $E.add(< a, relatedTo, h >)$;
end

4.2 Graph-Based Knowledge Linking

After the construction of knowledge graphs for two input wikis, we match equivalent entities between the two graphs. Because semantic relations between entities are language-independent features [14], we assume that the two knowledge graphs share similar structures. Similarity flooding is an efficient algorithm for alignment between two similar ontologies. Two main processes in SF are pairwise connectivity graph (PCG) construction and similarity propagation. If the number of nodes for two input graphs are m and n, the scale of nodes number for PCG will be $O(m \times n)$. Thus, the SF algorithm is possible for relatively small m and n (e.g. in schema matching tasks), but clearly infeasible when m and n are too large. The knowledge graph of one input wiki may contain thousands of nodes, which makes the nodes number of the PCG become billions. Thus, SF can not be directly applied to the knowledge linking because of the computational challenge. To tackle with this problem, we propose a variation of SF algorithm which reduce the scale of the propagation graph. The algorithm are described in detail in the following subsections.

Initial Similarity Computation. The two KGs are construed from link-based structural information such as infoboxes and innerlinks of wikis. The text descriptions of entities in KG are not considered. In our method, we calculate the initial similarities of entities based on their text description. In order to

calculate texture similarities across languages, some priori knowledge about the two languages should be given. The priori knowledge is defined as the *Domain Dictionary* in our model. Specifically, the domain dictionary is a set of word pairs between two languages. More specifically, given two different languages A and B and the domain D, the domain dictionary Dic_D is defined as:

$$Dic_D = \{<w_i^A, w_i^B>\}_{i=1}^{L} \tag{1}$$

where $<w_i^A, w_i^B>$ is a translation equivalent pair between language A and B (e.g. <"China","中国">). To provide sufficient priori knowledge for initial similarity computation, word pairs in Dic_D should be relevant to the domain D. Given the domain dictionary Dic_D, we represent each entity in KGs as a L dimensional vector. Without loss of generality, for an entity e from a wiki written in language A, the *i-th* dimension of e's vector is the frequency of w_i^A appeared in its corresponding wiki article. Finally, the initial similarity between entity e and e' is the cosine similarity of their corresponding vectors.

Propagation Graph Construction. Based on the assumption that a part of the similarity of two elements should propagate to their respective neighbors [8], we convert the two knowledge graphs to a *Similarity Propagation Graph* (SPG) as follows:

Definition 5. **Similarity Propagation Graph.** Given two knowledge graphs G and G', $((e,e'),r,(o,o')) \in SPG(G,G')$ if and only if: (1) $(e,r,o) \in G$, (2) $(e',r,o') \in G'$, (3) $Sim(e,e') > \theta$, and (4) $Sim(o,o') > \theta$. The $Sim(e,e')$ indicates the initial similarity between entity e and e'. θ is a pre-given threshold.

Each node in the SPG represents a candidate alignment pair between the two KGs. To reduce the scale of SPG, the conditions (3) and (4) in Definition 5 remove the entities pairs which have low initial similarities.

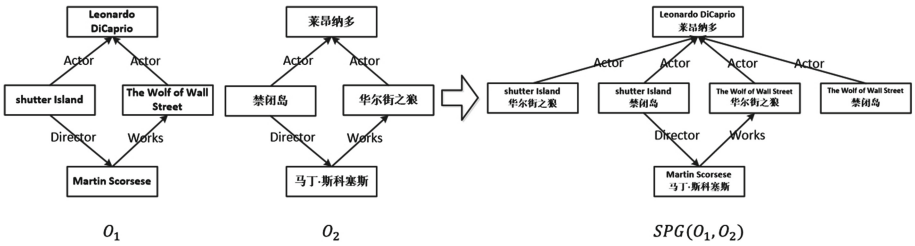


Fig. 4. An example of the construction of SPG

Figure 4 shows an example of SPG construction. The left side of the figure are two small cross-lingual KGs of movie domain, denoted as O_1 and O_2. The right side shows the SPG of O_1 and O_2. In $SPG(O_1, O_2)$, nodes are entity pairs

from two KGs that have some structural relationship in common. For example, "shutter Island" and "禁闭岛" are two entities in O_1 and O_2. They are constructed into a node in $SPG(O_1, O_2)$ because they share the same relationship "Director".

Similarity Propagation. The similarity propagation starts from initial similarities between nodes of two KGs and runs an iterative propagation in the SPG. In each iteration, the similarity of a given matching pair would be propagated to the neighborhood matching pairs. The iteration stops when no similarity changes or after a predefined number of steps. Formally, let us denote $\sigma^i(e, e')$ as the similarity between e and e' after the i-th iteration. The iteration equation to perform similarity propagation is defined as follows:

$$\sigma^{i+1}(e, e') = \frac{1}{Z}\left(\sigma^0(e, e') + \sigma^i(e, e') + \varphi^i(e, e')\right) \tag{2}$$

$$\varphi^i(e, e') = \sum_{(o,o') \in IN(e,e')} \omega(o, o') \cdot \sigma^i(e, e') \tag{3}$$

$$Z = \max_{(e,e') \in SPG(G,G')} (\sigma^{i+1}(e, e')) \tag{4}$$

As defined in Eq. 2, for any entity pair (e, e') in SPG, the similarity of (e, e') in the $(i + 1)$-th iteration is dependent on its similarity of the i-th iteration $(\sigma^i(e, e'))$, its initial similarity $(\sigma^0(e, e'))$ and the similarity gain from its neighbors $(\varphi^i(e, e'))$. In Eq. 3, $\varphi^i(e, e')$ is defined as the weighted sum of similarities of (e, e')'s adjacent nodes. $IN(e, e')$ is the set of incoming neighbors of the node (e, e') in SPG. $\omega(o, o')$ is the propagation weight which is simply defined as the inverse of the number of out-linking relationships for the node (o, o'). Z is a normalization factor defined in Eq. 4.

5 Experiments

5.1 Datasets

The proposed method can be used to find cross-lingual links between any online wikis in different languages. To evaluate the performance of our method, we extract the articles of the movie domain from three different online wikis. Specifically, one English language online wiki—English Wikipedia—and two Chinese language online wikis—Baidu Baike and Chinese Wikipedia—are chosen for our experiments. For English Wikipedia and Chinese Wikipedia, we employ the latest publicly available Wikipedia dump[2], which includes **9,834,664** articles for English Wikipedia and **886,437** articles for Chinese Wikipedia. As for the Baidu Baike, which is the largest web-based encyclopedia in China, we crawled **6,223,649** web pages from the latest Baidu encyclopedia.

[2] https://dumps.wikimedia.org/enwiki/20160113/.

To further create datasets of movie domain, we utilize the category system of online wikis. Specifically, we select several typical categories for the movie domain such as "film", "actor" and "director", and we extract articles with these categories from the aforementioned three online wikis. As a result, a total of **222,022** movie domain articles are extracted from English Wikipedia. As for the Baidu Baike and Chinese Wikipedia, we extract **112,164** and **58,638** articles for the movie domain, respectively. We refer to these three datasets as **EWM** (English Wikipedia of movie domain), **ZWM** (Chinese Wikipedia of movie domain) and **BBM** (Baidu Baike of movie domain).

To create gold standard for evaluation, we construct a evaluation dataset that contains equivalent article pairs between EWM and ZWM. These article pairs are acquired from the existing Chinese-English cross-lingual links within Wikipedia. As a result, we obtain **2,678** CLs between EWM and ZWM. We also create a dataset of CLs between EWM and BBM. Firstly, a total of 10,000 articles are randomly selected from EWM. Then, each selected article is sent to a human annotator to find its corresponding article in BBM. However, not all articles in English Wikipedia have CLs to Baidu Baike. Finally, we obtain a total of **4,022** CLs from the 10,000 sampled articles as our evaluation dataset.

Knowledge Graph Construction. As for the knowledge graph construction, we use the attribute mapping shown in Fig. 3 for all the three datasets. In other words, we consider four kinds of relations—<Actor>,<Director>,<Writer> and <Works>—, which are frequently occurred among entities in the movie domain. Then, we apply the Algorithm 1 defined in Sect. 4.1 to construct a KG for each wiki dataset. In addition with the <relatedTo> relation defined in Algorithm 1, there exist a total of five kind of edges in the constructed KGs. Detailed statistics of the three constructed KGs are presented in Table 1. For example, in the KG for the EWM dataset, there exist 185,453 edges with the relation <Actor>.

Table 1. Statistics of knowledge graphs for three datasets

Dataset	#Nodes	#Edges				
		<Actor>	<Director>	<Writer>	<Works>	<relatedTo>
EWM	220,989	185,453	73,705	48,500	373,550	681,208
ZWM	57,842	81,717	23,544	11,730	93,257	151,299
BBM	111,768	154,112	18,921	9,603	180,370	363,006

5.2 Methods for Comparison

We define four cross-lingual linking methods as the comparison methods.

- **Title Edit Distance (TED).** This method first translates the titles of Chinese articles into English by Google Translation API[3], then we calculate the similarity between all article pairs using the edit distance of their titles.
- **Initial Similarity (IS).** This method directly regards the initial similarities defined in Sect. 4.2 as the final result. Due to the removal of the similarity propagation step, the essence of this method is a translation-based method which calculates the similarities between article texts.
- **Simple Similarity Propagation (SSP).** This method conducts the similarity propagation process without the influence of initial similarity. We initialize all nodes in the SPG with a unified initial similarity (set to 0.5 in our experiments). Accordingly, the Eq. 2 is rewritten as:

$$\sigma^{i+1}(e, e') = \frac{1}{Z}\left(\sigma^{i}(e, e') + \varphi^{i}(e, e')\right) \tag{5}$$

- **Linkage Factor Graph (LFG).** The LFG model [14] is a state-of-art method for cross-lingual knowledge linking. The method first calculates several language-independent features from input wikis and then proposes a factor graph model to discover cross-lingual links.

Implementation Details. For the construction of domain dictionary for the IS method and the proposed method, we first rank all the words in the EWM dataset by their TF-IDF, then we select the top 1000 ranked words to form a set of keywords. Finally, we translate each of the keyword into Chinese by Google Translation API to get 1000 English-Chinese word pairs. These word pairs are employed as our domain dictionary. For the LFG, we use 0.001 learning rate and run 2500 iterations in all the experiments to get its best performance.

Evaluation Metrics. We evaluate our approach on the Chinese-English cross-lingual links constructed in Sect. 5.1, i.e. the 2,678 EWM-ZWM cross-lingual links and the 4,022 CLs between the EWM dataset and the BBM dataset. For an arbitrary candidate matching pair, all comparison methods are able to predict its similarity, indicating its confidence level of being equivalent. Intuitively, for an article e in the source wiki, the article in the target wiki which has the highest similarity with e is regarded as its predicted CL. Thus, we use the prediction accuracy on the evaluation CLs to evaluate different knowledge linking methods, denoted as $P@1$. In addition, we also evaluate the methods by $P@5$, which is defined as the percentage of articles that have correct equivalent articles in its Top-5 candidates.

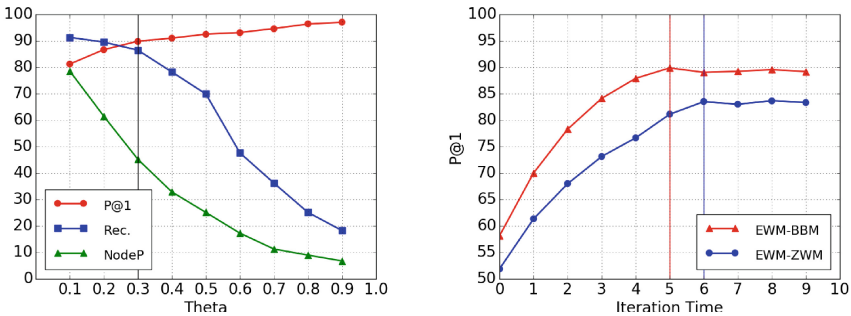
5.3 Influences of Parameters

There are two parameters in our method that may influence the performance including: (1) pruning threshold θ, (2) iteration time T. In this section, we look into the influences of these parameters.

[3] http://code.google.com/intl/zhcn/apis/language/translate/overview.html.

The Parameter θ. The parameter θ defined in Definition 5 is used for pruning the SPG. The entity pair having a lower initial similarity than θ is not considered as a candidate CL, and will be removed from the SPG. Figure 5(a) demonstrates the influence of θ on the knowledge linking between EWM and BBM. In the figure, the *Rec.* denotes the coverage rate of the PCG for the entity pairs in the evaluation set, indicating the recall of our method. *NodeP* is defined as the number of nodes in PCG divided by the nodes number of PCG when $\theta = 0$. From the figure, we observe that the *Rec.* declines rapidly with the increasing of θ. The reason is: if θ is set too high, we may prune the SPG too much. As a result, many correct entity pairs may be excluded from the SPG, which leads to a poor recall. However, if θ is too low, the number of nodes in PCG will increase rapidly, which makes the algorithm computationally challenging. We observe that the method reaches its best F1 score when $\theta = 0.3$. The similar observations are also got on the experiment between EWM and ZWM.

The Iteration Time T. Figure 5(b) shows the P@1 of the proposed method with different iteration time T on the two knowledge linking tasks. The parameter θ is set to 0.3 for both tasks. From the figure, we observe that the proposed method reaches its best performance through 4 to 6 iterations. In the knowledge linking task between EWM and BBM, our method converges to the best performance (89.89 % in terms of P@1) after the 5-th iteration. Similarly, for the datasets of EWM and ZWM, the method converges to the P@1 of around 83 % after 6 iterations.



(a) The influence of the parameter θ (b) The influence of iteration time T

Fig. 5. The study of parameter influence on the proposed method (%).

5.4 Results Analysis

After we explore the influences of parameters, we further employ baseline methods to compare with the proposed method. Table 2 summarizes the performance of 5 different methods on different knowledge linking tasks. From the table, we find that the proposed method outperforms all baselines on both two tasks.

According to the result, the TED method gets the lowest $P@1$ of 55.32 % and 54.79 %, because only the entity titles are used in this method. The IS method utilizes the article texts of entities, while the SSP method take advantage of the semantic information contained in the infoboxes. The two methods have better performances than TED because of utilizing more information of the wiki article. The proposed method can be regraded as the combination of the IS and the SSP. By combining the texture information and the structural information in a synergistic way, our method outperforms the IS and the SSP by 21.75 % and 12.28 % with regards to $P@1$, respectively. Compared with the LFG method, our method focuses on the knowledge linking task in specific domains. The experimental results show that our method outperforms the LFG (+4.68 % in terms of $P@1$ in average) with regards to both $P@1$ and $P@5$.

Table 2. Performance of knowledge linking with different methods (%).

Tasks	Metrics	Methods				
		TED	IS	SSP	LFG	Proposed
EWM-BBM	$P@1$	55.32	68.14	77.61	83.73	**89.89**
	$P@5$	62.91	75.53	86.56	88.21	**93.28**
EWM-ZWM	$P@1$	54.79	61.88	70.11	80.26	**83.51**
	$P@5$	61.53	67.03	80.35	82.29	**87.33**

6 Conclusion and Future Work

In this paper, we propose a cross-lingual knowledge linking approach for discovering domain-specific cross-lingual links across online wikis. Our approach combines both domain semantic relations and texture features of the wiki article, and employs a variation of similarity flooding to predict new cross-lingual links. Evaluations on two cross-lingual linking tasks show that our approach can outperform the state-of-the-art method by an average precision of 4.68 %. Our future work is to studying the automatical construction method for the attribute mapping, to extend our approach to a more general one.

References

1. Bizer, C., Lehmann, J., Kobilarov, G., Auer, S., Becker, C., Cyganiak, R., Hellmann, S.: Dbpedia - a crystallization point for the web of data. Web Semant. Sci. Serv. Agents World Wide Web **7**(3), 154–165 (2009)
2. Cudré-Mauroux, P., Haghani, P., Jost, M., Aberer, K., De Meer, H.: idMesh: graph-based disambiguation of linked data. In: Proceedings of WWW, pp. 591–600 (2009)
3. De Melo, G., Weikum, G.: MENTA: inducing multilingual taxonomies from wikipedia. In: Proceedings of CIKM, pp. 1099–1108 (2010)

4. Erdmann, M., Nakayama, K., Hara, T., Nishio, S.: Improving the extraction of bilingual terminology from wikipedia. Int. J. TOMM **5**(4), 31 (2009)
5. Fu, B., Brennan, R., O'Sullivan, D.: Cross-lingual ontology mapping – an investigation of the impact of machine translation. In: Gómez-Pérez, A., Yu, Y., Ding, Y. (eds.) ASWC 2009. LNCS, vol. 5926, pp. 1–15. Springer, Heidelberg (2009). doi:10.1007/978-3-642-10871-6_1
6. Hassan, S., Mihalcea, R.: Cross-lingual semantic relatedness using encyclopedic knowledge. In: Proceedings of EMNLP, pp. 1192–1201 (2009)
7. Li, J., Tang, J., Li, Y., Luo, Q.: RiMOM: a dynamic multistrategy ontology alignment framework. Int. J. of TKDE **21**(8), 1218–1232 (2009)
8. Melnik, S., Garcia-Molina, H., Rahm, E.: Similarity flooding: a versatile graph matching algorithm and its application to schema matching. In: Proceedings of ICDE, pp. 117–128 (2015)
9. Oh, J.H., Kawahara, D., Uchimoto, K., Kazama, J., Torisawa, K.: Enriching multilingual language resources by discovering missing cross-language links in wikipedia. In: Proceedings of WI-IAT, pp. 322–328 (2008)
10. Potthast, M., Stein, B., Anderka, M.: A wikipedia-based multilingual retrieval model. In: Macdonald, C., Ounis, I., Plachouras, V., Ruthven, I., White, R.W. (eds.) ECIR 2008. LNCS, vol. 4956, pp. 522–530. Springer, Heidelberg (2008). doi:10.1007/978-3-540-78646-7_51
11. Sorg, P., Cimiano, P.: Enriching the crosslingual link structure of wikipedia - a classification-based approach. In: Proceedings of the AAAI Workshop on Wikipedia and Artificial Intelligence (2008)
12. Volz, J., Bizer, C., Gaedke, M., Kobilarov, G.: Discovering and maintaining links on the web of data. In: Bernstein, A., Karger, D.R., Heath, T., Feigenbaum, L., Maynard, D., Motta, E., Thirunarayan, K. (eds.) ISWC 2009. LNCS, vol. 5823, pp. 650–665. Springer, Heidelberg (2009). doi:10.1007/978-3-642-04930-9_41
13. Wang, Z., Li, J., Tang, J.: Boosting cross-lingual knowledge linking via concept annotation. In: Proceedings of IJCAI, pp. 2733–2739 (2013)
14. Wang, Z., Li, J., Wang, Z., Tang, J.: Cross-lingual knowledge linking across wiki knowledge bases. In: Proceedings of WWW, pp. 459–468 (2012)
15. Wang, Z., Li, Z., Li, J., Tang, J., Pan, J.Z.: Transfer learning based cross-lingual knowledge extraction for wikipedia. In: Proceedings of ACL, pp. 641–650 (2013)
16. Wentland, W., Knopp, J., Silberer, C., Hartung, M.: Building a multilingual lexical resource for named entity disambiguation, translation and transliteration. In: Proceedings of ICLRE, pp. 3230–3237 (2008)