Cross-Lingual Question Answering Using Common Semantic Space

Amir Pouran Ben Veyseh
University of Tehran
Tehran, Iran
`veyseh@ut.ac.ir`

Abstract

With the advent of Big Data concept, a lot of attention has been paid to structuring and giving semantic to this data. Knowledge bases like DBPedia play an important role to achieve this goal. Question answering systems are common approach to address expressivity and usability of information extraction from knowledge bases. Recent researches focused only on monolingual QA systems while cross-lingual setting has still so many barriers. In this paper we introduce a new cross-lingual approach using a unified semantic space among languages. After keyword extraction, entity linking and answer type detection, we use cross lingual semantic similarity to extract the answer from knowledge base via relation selection and type matching. We have evaluated our approach on Persian and Spanish which are typologically different languages. Our experiments are on DBPedia. The results are promising for both languages.

1 Introduction

Large scale knowledge bases like DBPedia (Auer et al., 2007) and Freebase (Bollacker et al., 2008) provide structured information in diverse domains. Such resources are worthwhile to answer open-domain questions using structured query. In recent years, answering open-domain questions by querying knowledge bases has gained a lot of attentions (Yahya et al., 2012; Fader et al., 2013; Yih et al., 2014; Dong et al., 2015). These systems exploit many diverse methods like semantic parsing, information extraction (Yao et al., 2014) and deep learn-

ing (Yu et al., 2015; Bordes et al., 2015). While existing approaches focused only on English language, there are so many difficulties to cope with in cross lingual setting. On the one hand, lack of tools and resources, and on the other hand, vocabulary gap between source and target languages, frustrate any effort to adapt the existing approaches for languages other than English.

In this paper, we introduce a pipeline of stages for cross lingual question answering over knowledge bases. In the first stage, using a MaxEnt Markov Model, keywords are extracted. Syntactic and semantic features are utilized to do this job. In the second stage, using an SVM classifier, keywords that mention an entity are distinguished from ones that determine the answer type. In the next stage we try to find the most probable entity(s) in KB which can be linked to detected grounded entity(s). Several sources are used to find entities in KB, like abstract of entities in KB, cross lingual dictionaries like BabelNet (Navigli and Ponzetto, 2012) and the KB own cross lingual links (whenever such links exist). Also using extracted keywords we search in the ontology of the KB to predict type of entities that are answers. In the last stage, answers are extracted using two kinds of information: 1. Type of neighbours of found entities 2. Semantic similarity between relation labels of found entities and extracted keywords. Cross lingual semantic similarity are measured using the unified semantic space among languages proposed by Camacho-Collados (2015).

Our system doesn't rely on huge annotated data or any language specific resources except for a chunker. Thus our main contributions are:

15

- Introducing a staged cross lingual approach which can easily be adapted to any source language with an available chunker in that language.

- Reducing annotation effort and reliance on huge amount of training data which is a barrier for many resource scarce languages.

- Providing a new QA dataset for Persian and conducting experiences on two different languages.

2 Related Work

Early Question Answering systems like Baseball (Green Jr et al., 1961) were close domain. With the expansion of Linked Open Data, open domain knowledge bases like DBPedia and Freebase emerged. Several approaches have been proposed to provide natural language interface to these KBs. Some of them have utilized semantic parsing techniques (Fader et al., 2014; Cai and Yates, 2013). In these systems the question is converted to an intermediate logical form like lambda calculus and then using this interpretation of the question, the final query is constructed. Some systems have tried to use Information Extraction techniques to overcome QA task (Yao and Durme, 2014; Bordes et al., 2014). Yao and et. al (2014) showed that these two trends are not very different in their performances, but semantic parsing can target answering more compositional questions.

There are some attempts from deep learning researchers in this field (Yu et al., 2015; Yih et al., 2014; Dong et al., 2015). Sukhbaatar and et al. (2015) trained an end-to-end Memory Network contributed by Weston (2014). Their model is multi layered. In each layer any fact has two different input and output embedded form. After passing the embedded question through these layers, the predicted answer is obtained using a weight matrix.

All of aforementioned methods are mono lingual and can not be adapted for a resource scarce language mostly because of their reliance on huge training data or language specific resources both for question understanding or relation and entity extraction. Although cross lingual question answering over unstructured data is a well-known topic (Ahn

et al., 2004; de Pablo-Sánchez et al., 2005; Ligozat et al., 2006), but these approaches don't utilize all information available for QA over knowledge bases. Most of these systems deal with multilinguality using fully translation of given question or term-by-term translation of processed question in source language. Entity types and external links to cross lingual resources, existing in well-known open domain KBs, can be exploited to overcome many translation errors and our experiences corroborate this fact. Aggarwal and et al. (2013) proposed a cross lingual QA system over DBPedia. They have achieved 0.481 F_1 on QALD-2. Their system was unable to answer aggregation question. Moreover, unlike our method, they didn't utilize type information in the KB.

3 Method

In this section we introduce our proposed approach to deal with cross lingual QA over KBs. Any given question passes through four stages in a pipeline including 1. Keyword Extraction, 2. Keyword Type Detection, 3. Entity Linking & Ontology Type Extraction, and finally 4. Answer Extraction. We have employed QALD-5 as our dataset for training and testing, and DBPedia 2014 as our KB to extract answers from. We first briefly describe preparing dataset and then explain each of above stages in detail.

3.1 Preparing Dataset

QALD-5 is a multilingual QA dataset over DBPedia 2014 for QALD task at CLEF 2015. It contains 300 training questions in 7 languages with annotated keywords and queries to extract answers from DBPedia and 50 questions as test set. To add Persian translation to these dataset, the questions were translated to Persian by a language expert outside development team. To annotate keywords of each Persian questions we have used majority voting among 5 annotators. Each word has tagged as B, I or O. Also we have augmented this dataset with answer type tag. Each keyword has tagged as type detector or neutral. We have chosen these tags through majority voting among 5 annotators.

3.2 Keyword Extraction

In the first stage, the input question have to be analysed to extract content words which we call them

keywords. A MaxEnt Markov Model is used in order to extract these keywords through sequence labelling. The features used to train the model are: 1. Unigram, bigram and trigram of POS tags, 2. Chunk tag, 3. Position of the word in question, 4. IDF of word in corpus [1], 5. Exact match with entity labels in KB and 6. Babelfy tag(Moro et al., 2014).

3.3 Keyword Type Detection

In the second stage each keyword is classified as 1. Type Detector, 2. Grounded Entity or 3. Neutral. To do that we have utilized an SVM classifier with RBF kernel because it has the best performance in 10-fold cross validation in our experiments. The following features are used to train the SVM classifier: 1. Number of words in keyword, 2. POS tags of words in keyword, 3. Position of the first word of the keyword in question, 4. Average IDF of words of keywords in corpus, 5. Exact match with entity labels, 6. Babelfy tag, and 7. Match of translation with ontology types [2].

3.4 Entity Linking & Ontology Type Extraction

When the keywords that must be linked to some entities in the KB or refer to some types in the KB ontology have been found, we should link each of which to its appropriate entity or type. To do entity linking, results of queries over three different collections are merged with different weights and the first result is selected. Queries includes the entity mention augmented by other keywords with different weight and are over: 1. Entity abstracts in language specific chapters of the KB –Only entities which have $sameAs$ links to English version are indexed. For any query, the related English entity is returned. 2. Entity abstracts of English chapter of KB –Any keyword in query is translated to English using Google Translate. and 3. BabelNet –We use Babelfy for entity linking. Weights that cause minimum errors in training samples are chosen for augmenting keywords to queries and merging the results. To link keywords classified as type detector, we use Google Translate to translate the keyword to

English and then search in KB ontology. Only string similarity between translated keyword and ontology types is used.

3.5 Answer Extraction

In the last stage using extracted entities, ontology types and keywords classified as Neutral in subsection 3.3, we search in KB graph. All entities in 2-hop vicinity of the found entities in KB whose types are different from extracted ontology types are pruned. If there are entities of desired types with different path labels to the found entities, the cross lingual semantic similarity model contributed by Camacho-Collados (2015) is used to select the most similar relation with the keywords of tagged as Neutral. We have used unified vectors constructed according to BabelNet synsets.

To deal with aggregation questions, we have extracted all questions from the training data with atribute $aggregation = "true"$ and then grouped them in four types:

1. Sort 2. Count 3. Regular expresion 4. Time

For each of these groups, POS tags and words which are common among all members of that group have been extracted and their frequencies have been calculated and normalized. For a given question four scores related to each agregation type is caclulated using following formula:

$$Score_{type}(Q) = \sum_i N(POS_i) * S(POS_i)$$
$$+ \sum_j N(w_j) * S(w_j) \quad (1)$$

$N(*)$ is the number of occurance of $*$ in the question and $S(*)$ is the normalized score obtained for $*$ using agregation training questions. For each type which have some score greater than a threshold, agregation of that type is operated on the final result of the answer extracted. The threshold are calculated using all training questions.

4 Experiments

To evaluate our approach we have conducted experiments on Persian and Spanish. we have used DBPedia 2014 as the KB that answers must be extracted from it. We have tested our system on QALD-5 test set. It contains 49 questions in both languages.

---

[1]The language specific abstract of every entity in DBPedia was indexed by Lucene to obtain IDF of words.

[2]Keywords are translated to English using Google Translate.

Language(Method)	Precision	Recall	F_1
Persian(ours)	53.3	51.7	52.4
Spanish(ours)	55.2	53.4	54.2
Persian(Fully MT)	30.7	28.3	29.5
Spanish(Fully MT)	33.4	31.1	32.2

Table 1: Performance of our approach for Persian and Spanish questions compared with baseline.

Method	F_1
ours(English)	65.2
Xser(Xu et al., 2014)	63.0

Table 2: Performance of our approach for English questions compared with best participant of QALD-5.

As a baseline we translate each question to English using Google Translate. Table 1 shows the result of our approach for both Persian and Spanish questions compared with results of the baseline. Errors in translating of named entities in fully translating a question is one of main sources of errors in baseline with proportion of 64%.

We have compared the performance of the monolingual version of our approach with the best participant of QALD-5 challenge. Table 2 shows the results. Despite less annotation cost for training the model compared with Xser, our system improved F_1 by 2.2%.

Since our proposed approach consist of a pipeline of pre-processing, we have evaluated internal stages of our system. Table 3 shows the results for each stage. The reported accuracies are average accuracy over 10-fold cross validation.

We have also evaluated the influence of calculating semantic similarity using unified vectors on accuracy of our method. Semantic similarity of the

Stage(Language)	Macro Accuracy
Keyword Extraction(Persian)	92.3
Keyword Extraction(Spanish)	92.8
Keyword Classification(Persian)	88.7
Keyword Classification(Spanish)	89.3
Ontology Type Selection(Persian)	84.4
Ontology Type Selection(Spanish)	85.6
Entity Linking(Persian)	78.6
Entity Linking(Spanish)	79.2
Relation Selection(Persian)	61.8
Relation Selection(Spanish)	63.8

Table 3: Performance of each stage.

Method	F_1
Unified Vectors	52.4
Translation & Word2Vec	46.7

Table 4: Performance of our method with unified vectors and Word2vec semantic similarity

translation of a Persian or Spanish phrase and an English phrase using Word2Vec was used as a baseline. The result for Persian are reported in Table 4.

5 Error Analysis

Wrong Relation Selection is the main source of error (32% of errors in test set). Table 3 shows a great accuracy lost between this stage and its previous one. The diversity in paraphrasing different relations in KB is one of most important hindrance to remedy vocabulary gap in this stage even in monolingual setting . Moreover, for some questions the answer are connected to entities grounded in questions by more than one relation. Our approach doesn't select more than one relation.

Other sources of errors include: wrong entity linking (27%), questions with no grounded entity (22%), wrong ontology type selection (15%) and others (4%)

6 Conclusion & Future Works

Question answering over knowledge bases is one of the popular topics for researchers in semantic parsing, information extraction, deep learning and so on. However various systems in recent years have been proposed but cross-linguality is rarely studied so far. In this paper we proposed a cross lingual system using unified semantic representation of concepts in different stages for keyword extraction, keyword classification, entity linking, type extraction and relation selection. Although our experiments show usefulness of proposed approach but there are still a lot of rooms for future works. More investigation on relation extraction is needed. Deep learning approaches like Memory Networks show promising results and we plan to adapt our system for these approaches. Extending our method for other KBs that don't have versions in other languages like Freebase and also other datasets like WebQuestions (Berant et al., 2013) is another room for future work. Recently there was some research on dialect-level differences between Persian and Dari (Malmasi and Dras, 2015). Adapting and evaluating our method in cross-dialect setting has been left for future work.

18

References

Nitish Aggarwal, Tamara Polajnar, and Paul Buitelaar. 2013. Cross-lingual natural language querying over the web of data. In *NLDB 2013, June 19-21, 2013. Proceedings*, pages 152–163.

Kisuh Ahn, Beatrix Alex, Johan Bos, Tiphaine Dalmas, Jochen L. Leidner, and Matthew Smillie. 2004. Cross-lingual question answering with QED. In *Working Notes for CLEF 2004 Workshop, September 15-17, 2004.*

Sören Auer, Christian Bizer, Georgi Kobilarov, Jens Lehmann, Richard Cyganiak, and Zachary G. Ives. 2007. Dbpedia: A nucleus for a web of open data. In *The Semantic Web,ISWC 2007 + ASWC 2007, November 11-15, 2007.*, pages 722–735.

Jonathan Berant, Andrew Chou, Roy Frostig, and Percy Liang. 2013. Semantic Parsing on Freebase from Question-Answer Pairs. *Proceedings of EMNLP*, pages 1533–1544.

Kurt D. Bollacker, Colin Evans, Praveen Paritosh, Tim Sturge, and Jamie Taylor. 2008. Freebase: a collaboratively created graph database for structuring human knowledge. In *Proceedings of the SIGMOD 2008, June 10-12, 2008*, pages 1247–1250.

Antoine Bordes, Sumit Chopra, and Jason Weston. 2014. Question answering with subgraph embeddings. In *Proceedings of EMNLP 2014, October 25-29, 2014*, pages 615–620.

Antoine Bordes, Nicolas Usunier, Sumit Chopra, and Jason Weston. 2015. Large-scale simple question answering with memory networks. *CoRR*, abs/1506.02075.

Qingqing Cai and Alexander Yates. 2013. Large-scale semantic parsing via schema matching and lexicon extension. In *Proceedings of ACL 2013, 4-9 August 2013, Volume 1: Long Papers*, pages 423–433.

José Camacho-Collados, Mohammad Taher Pilehvar, and Roberto Navigli. 2015. A unified multilingual semantic representation of concepts. In *Proceedings of ACL 2015, July 26-31, 2015, Volume 1: Long Papers*, pages 741–751.

César de Pablo-Sánchez, Ana González-Ledesma, José Luis Martınez-Fernández, José Maria Guirao, Paloma Martinez, and Antonio Moreno. 2005. Miracles 2005 approach to cross-lingual question answering. In *Working Notes of CLEF*.

Li Dong, Furu Wei, Ming Zhou, and Ke Xu. 2015. Question answering over freebase with multi-column convolutional neural networks. In *Proceedings of ACL 2015, July 26-31, 2015, Volume 1: Long Papers*, pages 260–269.

Anthony Fader, Luke S. Zettlemoyer, and Oren Etzioni. 2013. Paraphrase-driven learning for open question answering. In *Proceedings of ACL 2013, 4-9 August 2013, Volume 1: Long Papers*, pages 1608–1618.

Anthony Fader, Luke Zettlemoyer, and Oren Etzioni. 2014. Open question answering over curated and extracted knowledge bases. In *The 20th KDD '14, August 24-27, 2014*, pages 1156–1165.

Bert F Green Jr, Alice K Wolf, Carol Chomsky, and Kenneth Laughery. 1961. Baseball: an automatic question-answerer. In *western joint IRE-AIEE-ACM computer conference*, pages 219–224. ACM.

Anne-Laure Ligozat, Brigitte Grau, Isabelle Robba, and Anne Vilnat. 2006. Evaluation and improvement of cross-lingual question answering strategies. In *Proceedings of the Workshop on Multilingual Question Answering*, pages 23–30. ACL.

Shervin Malmasi and Mark Dras. 2015. Automatic language identification for persian and dari texts. In *Proceedings of PACLING*, pages 59–64.

Andrea Moro, Alessandro Raganato, and Roberto Navigli. 2014. Entity Linking meets Word Sense Disambiguation: a Unified Approach. *TACL*, 2:231–244.

Roberto Navigli and Simone Paolo Ponzetto. 2012. Babelnet: The automatic construction, evaluation and application of a wide-coverage multilingual semantic network. *Artif. Intell.*, 193:217–250.

Sainbayar Sukhbaatar, Jason Weston, Rob Fergus, et al. 2015. End-to-end memory networks. In *NIPS*, pages 2431–2439.

Jason Weston, Sumit Chopra, and Antoine Bordes. 2014. Memory networks. *CoRR*, abs/1410.3916.

Kun Xu, Yansong Feng, and Dongyan Zhao. 2014. Answering natural language questions via phrasal semantic parsing. In *Working Notes for CLEF 2014 Conference, September 15-18, 2014.*, pages 1260–1274.

Mohamed Yahya, Klaus Berberich, Shady Elbassuoni, Maya Ramanath, Volker Tresp, and Gerhard Weikum. 2012. Natural language questions for the web of data. In *Proceedings of EMNLP-CoNLL July 12-14, 2012*, pages 379–390.

Xuchen Yao and Benjamin Van Durme. 2014. Information extraction over structured data: Question answering with freebase. In *Proceedings of ACL 2014, June 22-27, 2014, Volume 1: Long Papers*, pages 956–966.

Xuchen Yao, Jonathan Berant, and Benjamin Van Durme. 2014. Freebase QA: Information Extraction or Semantic Parsing? *Proceedings of the ACL 2014 Workshop on Semantic Parsing*, pages 82–86.

Wen-tau Yih, Xiaodong He, and Christopher Meek. 2014. Semantic parsing for single-relation question answering. In *Proceedings of ACL 2014, June 22-27, 2014, Volume 2: Short Papers*, pages 643–648.

Yang Yu, Wei Zhang, Chung-Wei Hang, and Bowen Zhou. 2015. Empirical study on deep learning models for question answering. *CoRR*, abs/1510.07526.