Graph-Based Relation Validation Method

Rashedur Rahman, Brigitte Grau, Sophie Rosset

HAL Id: hal-01571800
https://hal.science/hal-01571800

Submitted on 11 Sep 2019

Graph-Based Relation Validation Method

Rashedur Rahman[1], Brigitte Grau[2], and Sophie Rosset[3]

[1] IRT SystemX, LIMSI, CNRS, University Paris Saclay
rashedur.rahman@irt-systemx.fr
[2] LIMSI, CNRS, ENSIIE, University Paris Saclay
brigitte.grau@limsi.fr
[3] LIMSI, CNRS, University Paris Saclay
sophie.rosset@limsi.fr

Abstract. In this paper we present a relation validation method for KBP slot filling task by exploring some graph features to classify the candidate slot fillers as correct or incorrect. The proposed features with voting feature collectively performs better than the baseline voting feature.

1 Introduction

Relation extraction and validation plays an important role in information extraction task like slot filling (SF) for knowledge base population (KBP). SF defines the task of finding the filler-entity (or entity) from texts by justifying the relation (or slot) of a given entity (the query). It requires entity level relation extraction based on the mention level relations (MLR) and entity linking that are dependent on each other. Traditional MLR extractor does not achieve satisfying precision and recall for SF task [1] and it often results high confidence score for incorrect relations because of limited features and training data. We propose to work at the entity level for validating relations because we can use additional features that cannot be used at mention level.

In this paper we explore community-graph based features for validating SF relations that were not explored before. The community graph is made of neighbor entities. Han et al. [2] proposed *referent graph* for collective entity linking where they took into account the semantic relations among the neighbor entities. Friedl et al. [3] discussed the use of different centrality measurements to find the important and influential nodes in social networks. Solá et al. [4] explored the concepts of eigenvector centrality in multi-plex networks and showed the existence of such centrality which is unique. Information theoretic measurements have also been proposed for knowledge discovery in complex networks [5], or for validating answer in question-answering systems [6]. We propose to explore community network and information theoretic concepts for validating relations between query and candidate entities. We consider relation validation as a binary classification task where the candidate filler entities are generated by MLR extractor. Here we do not evaluate the KBP SF task but evaluate relation validation. We observe the classification performances of different feature sets and show that the model including all features increases the F-score by 2.6% compare to the baseline voting system.

2 Method Description

Let, a graph $G = (V, E)$, query relation (slot) R_q, query entity $v_q \epsilon V$, candidate filler-entities $V_c = \{v_{c1}, v_{c2}, \ldots, v_{cn}\} \epsilon V$ where $R_q = e(v_q, v_c) \epsilon E$. The candidate list is generated by relation extractor. Suppose other semantic relations $R_o \epsilon E$ where $R_o \neq R_q$. We define the task to classify whether a filler-entity c of C_v is correct or incorrect for a query relation (R_q) by analyzing the communities of query entity and candidate fillers. Fig. 1 shows an example of community based relation validation task where the query
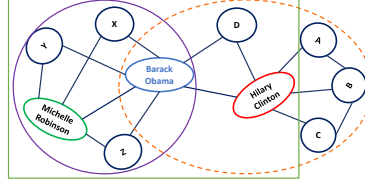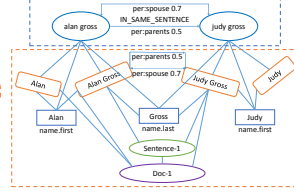

Fig. 1: Community graph   Fig. 2: Knowledge graph

entity, type and slot name are *Barack Obama*, *person* and *spouse* accordingly. The slot filler candidates are *Michelle Robinson* and *Hilary Clinton* that are linked to *Barack Obama* by *spouse* relation hypothesis. The communities of *Barack Obama* (green rectangle), *Michelle Robinson* (purple circle) and *Hilary Clinton* (orange ellipse) are constructed by *in_same_sentence* relation which means the pair of entities are mentioned in the same sentences in the texts. We want to classify *Michelle Robinson* as the correct slot filler based on community analysis. We create the community graph of entities from the knowledge graph as illustrated in Fig. 2. The knowledge graph is generated by applying some processing on the texts that includes named entity recognition (NER), sentence splitting, relation extraction (RE). The RE extracts the semantic relation between a pair of entities and gives a confidence score. The knowledge graph represents the documents, sentences, mentions as nodes and their relations. Entity mentions are connected to entities in the community graph and an edge connects a pair of entities in the community graph based on the associated relation hypothesis. Since the community graph is constructed based on the knowledge graph, the semantics is maintained in the community graph. We include *person, location and organization* typed entities as the community members in our community-graph-based analysis. A community is built with a set of entities which are mostly inter-related.

We assume that a correct filler-entity of a SF query should be a strong member in the community of the query entity and such community can be extracted from the texts by extracting semantic relations and/or based on their existences in the same sentences. We hypothesize that the *network density* (eq. 1) of a community of a correct filler-entity with the query entity should be higher than a community of an incorrect filler-entity with the query entity. In Fig. 1 the community of *Michelle Robinson* with *Barack Obama* is more dense than the community of *Hilary Clinton*. Eigenvector centrality [7] measures the influence of a neighbor node to measure the centrality of a node in a community. We quantify the influence of the candidate fillers in the community of a query entity by calculating the absolute difference between the eigenvector centrality scores of the query entity and a filler entity. We hypothesize that the difference should be

smaller for a correct filler than an incorrect filler. We also hypothesize that the *mutual information* (eq. 3) and *similarity* (eq. 2) between the community of a correct filler and the community of the query entity should be higher than an incorrect filler. The community of an entity (query entity or a candidate filler-entity) is expanded up to level 3 for measuring the eigenvector centrality and mutual information.

Additionally, we defined 6 ratios at the collection level as features based on the equations 4 to 9. For example, we calculate the ratio of occurrences of a filler-entity mention that are associated by the query relation with the query entity to the total number of mentions of that filler (eq. 4). We also include the confidence score (given by a relation extractor) as feature.

$$\rho_{network} = \frac{number\ of\ existing\ edges}{number\ of\ possible\ edges} \tag{1}$$

$$cosine\ similarity = \frac{|X \cap Y|}{\sqrt{|X||Y|}} \tag{2}$$

where, X and Y are the set of community-members of query and filler entity accordingly

$$MI(X, Y) = H(X) + H(Y) - H(X, Y) \tag{3}$$

where, $H(X) = -\sum_{i=1}^{n} p(x_i)\ log_2(p(x_i))$, X and Y are the communities of query and filler entity accordingly

and p(x) refers to the probability of centrality degree of a community-member

$$r_{mention}(e_c) = \frac{\#\ of\ mentions\ associated\ with\ e_q}{total\ number\ of\ mentions} \tag{4}$$

$$r_{hyp1}(e_q, e_c) = \frac{\#\ of\ documents\ with\ r_q(e_q, e_c)\ hyp.}{total\ number\ of\ r_q(e_q, e_c)\ hyp.} \tag{5}$$

$$r_{hyp2}(e_q, e_c) = \frac{total\ number\ of\ r_q(e_q, e_c)\ hyp.}{\#\ of\ sentences\ containing\ e_q\ and\ e_c} \tag{6}$$

$$r_{doc}(e_q, e_c) = \frac{\#\ of\ documents\ with\ r_q(e_q, e_c)\ hyp.}{total\ document\ count\ with\ all\ candidates} \tag{7}$$

$$relFreq_{hyp}(e_c) = \frac{number\ of\ r_q(e_q, e_c)\ hyp.}{total\ r_q\ count\ with\ all\ candidates} \tag{8}$$

$$r_{entity}(r_q, e_q, e_c) = \frac{\#\ of\ different\ entities\ in\ all\ r_q(e_q, e_c)\ sentences}{total\ entity\ count\ in\ all\ r_q(e_q, e_c)\ sentences} \tag{9}$$

3 Dataset and Experiments

We use the assessments of 2014 Cold Start Slot Filling (CSSF) evaluation task that contains 100 queries (50 for PERSON and 50 for ORGANIZATION) for building our corpus. Each PERSON query includes 16 slots (*per:spouse, per:children* etc) and an ORGANIZATION query contains 25 slots. The assessment files provide the correct and incorrect responses of the queries with the document reference that support the relation. We compile a subset of KBP-2014 evaluation corpus and we select 1942 documents for 7 slots (*org:top_members_employees, org:founded_by, per:statesorprovinces_of_residence, per:cities_of_residence, per:member_employees_of, per:children, per:spouse*) from the source document ids given in the assessment files and the queries because our current system limits to extract these relations. The documents (of correct and incorrect responses) are taken for a query slot if at least one responded slot filler string is justified as correct by NIST. The dataset includes 168 correct and 289 incorrect fillers from 97 queries. There exist multiple correct fillers for some queries because all the slots that we discuss here are multi-valued slots.

We trained several binary classifiers in Weka3.8 evaluated with 10 fold cross validation and the best F-score was achieved by SMO classifier. We group the relation valida-

Feature Set	Precision	Recall	F-score
Voting (baseline)	76.0	72.9	72.6
Graph + voting	75.0	73.5	73.5
Graph + collective(all) + voting	75.4	74.5	74.6
Graph + collective(eq. 5 & 6) + voting	**76.0**	**75.2**	**75.2**

Table 1: Classification performances (in %)

tion features into 3 sets: (i) *graph features*: equations 1, 2 and 3. (ii) *collective*: equations 4 to 9 and confidence score (iii) *voting (baseline)*: counts the maximum vote of a filler. Table 1 depicts the classification performances of different feature sets. The voting baseline obtains an F-score of 72.6%. The graph feature set and voting collectively obtain an F-score of 73.5%. We achieve the highest F-score of 75.2 which is 2.6% higher than the voting baseline by using all the graph features, two of the collective features (eq. 4 and eq. 5) and the voting features. The best scoring features have been selected by using ranker attribute selection (Relief method) in Weka. We evaluate the proposed features for validating relation between a pair of entities instead of evaluating its impact on the SF task because it has some requirements that we do not consider here. However, the experimental results strongly support the proposed community-graph based features for validating relations that could be very effective for KBP SF tasks.

4   Conclusion

In this paper we explored some community-graph based and corpus level collective features for validating relations that obtained promising results to classify correct and incorrect relations. The proposed graph features increased the F-score by 2.6% that strongly argues to continue graph based analysis for validating relation hypothesis.

References

1. Surdeanu, M., Ji, H.: Overview of the english slot filling track at the tac2014 knowledge base population evaluation. In: Proc. Text Analysis Conference (TAC2014). (2014)
2. Han, X., Sun, L., Zhao, J.: 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, ACM (2011) 765–774
3. Friedl, D.M.B., Heidemann, J., et al.: A critical review of centrality measures in social networks. Business & Information Systems Engineering **2**(6) (2010) 371–385
4. Solá, L., Romance, M., Criado, R., Flores, J., del Amo, A.G., Boccaletti, S.: Eigenvector centrality of nodes in multiplex networks. Chaos: An Interdisciplinary Journal of Nonlinear Science **23**(3) (2013) 033131
5. Holzinger, A., Ofner, B., Stocker, C., Valdez, A.C., Schaar, A.K., Ziefle, M., Dehmer, M.: On graph entropy measures for knowledge discovery from publication network data. In: Availability, reliability, and security in information systems and HCI. Springer (2013) 354–362
6. Magnini, B., Negri, M., Prevete, R., Tanev, H.: Is it the right answer?: exploiting web redundancy for answer validation. In: Proceedings of the 40th Annual Meeting on Association for Computational Linguistics, Association for Computational Linguistics (2002) 425–432

7. Bonacich, P., Lloyd, P.: Eigenvector-like measures of centrality for asymmetric relations. Social networks **23**(3) (2001) 191–201