IJRRECS/October 2013/Volume-1/Issue-6/1283-1288 **ISSN 2321-5461**

INTERNATIONAL JOURNAL OF REVIEWS ON RECENT ELECTRONICS AND COMPUTER SCIENCE

ANALOGOUS ENTITY MINING FROM COMPARATIVE QUERY

B.Suvarna Laxmi[1], U.Aswani Kumar[2]

[1]M.Tech Student, Dept of CSE, TKR College of Engineering & Technology, Hyderabad, A.P, India

[2]Associate Professor, Dept of CSE, TKR College of Engineering & Technology, Hyderabad, A.P, India

ABSTRACT:

Matching up substitute options is one of the necessary steps in decision-making that we carry out on a daily basis however necessitating high knowledge expertise. A comparative query has to be a query with intention to contrast at least two entities. On the other hand, we scrutinize that a question is very probable to be a comparative query if it encloses at least two entities. We control this insight and expand a weakly supervised bootstrapping means to recognize comparative queries and take out comparators at the same time. Bootstrapping methods have been revealed to be very effectual in earlier information extraction study. Specially, the for the most part of applicable work is by Jindal and Liu happening on mining comparative sentences and relations. Their methods apply class sequential rules (CSR) and label sequential rules (LSR) cultured from annotated corpora to recognize relative sentences and take out relative relations correspondingly in the news and reassess domains. Our weakly supervised indicative extraction pattern mining method is a pattern-based approach comparable to Jindal and Liu method, but it is dissimilar in a lot of aspects. Our weakly supervised indicative extraction pattern mining method is based on two important suppositions: It is for the reason that the greatest indicative extraction pattern is probable to be the most exact and pertinent pattern for the given query. If a sequential prototype can be used to take out numerous dependable comparator pairs, it is very probable to be an indicative extraction pattern. If a comparator pair can be taken out by an indicative extraction pattern, the pair is dependable.

Keywords: Comparative query, Bootstrapping method, Sequential rules, Indicative extraction pattern.

1. INTRODUCTION:

Matching up substitute options is one of the necessary steps in decision-making that we carry out on a daily basis however necessitating high knowledge expertise. In the World Wide Web period, a comparison action on average involves: search for applicable web pages enclosing information regarding the targeted products, discovering challenging products, and recognize pros and cons [2]. To extract comparators from relative matter, we initially have to become aware of whether a question is relative or not. A comparative query has to be a query with intention to contrast at least two entities [13]. On the other hand, we scrutinize that a question is very probable to be a comparative query if it encloses at least two entities. We control this insight and expand a weakly supervised bootstrapping means to recognize comparative queries and take out comparators at the same time [5]. Bootstrapping methods have been revealed to be very effectual in earlier information extraction study. Our effort is comparable to them in provisions of methodology by means of bootstrapping method to take out entities by means of a specific relation. On the other hand, our job is dissimilar from theirs in that it necessitates not simply extracting unit but also make sure that the entities are taken out from relative questions which are usually not necessary in indicative extraction task [7]. Our effort on comparator mining is connected to the study on entity and relation removal in information extraction. Specially, the for the most part of applicable work is by Jindal and Liu happening on mining comparative sentences and relations. Their methods apply class sequential rules (CSR) and label sequential rules (LSR) cultured from annotated corpora to recognize relative sentences and take out relative relations correspondingly in the news and reassess domains [6] [12]. The similar methods can be applied to comparative query recognition and comparator mining from query. On the other hand, their methods classically can attain high exactitude but undergo from low recall. On the other hand, making sure high recall is vital in our intended application situation where users can subject arbitrary queries [3]. To concentrate on this difficulty, we develop a weakly-supervised bootstrapping pattern learning method by means of efficiently leveraging unlabeled query.

2. METHODOLOGY:

Our weakly supervised indicative extraction pattern mining method is a pattern-based approach comparable to Jindal and Liu method, but it is dissimilar in a lot of aspects: as an alternative of using various class sequential rules and label sequential rules, our process aims to become skilled at sequential prototype which can be able to be used to recognize comparative enquiry and take out comparators concurrently [8] [14]. A sequential prototype is called an indicative extraction pattern if it can be used to recognize comparative query and take out comparators in them with high consistency. Our weakly supervised indicative extraction pattern mining method is based on two important suppositions: It is for the reason that the greatest indicative extraction pattern is probable to be the most exact and pertinent pattern for the given query [1] [4]. If a sequential prototype can be used to take out numerous dependable comparator pairs, it is very probable to be an indicative extraction pattern. If a comparator pair can be taken out by an indicative extraction pattern, the pair is dependable [9]. Based on these two suppositions, we plan our bootstrapping algorithm as revealed in fig1. The bootstrapping procedure begins with a single indicative extraction pattern. From it, we take out a set of early seed comparator pairs [15]. For every comparator pair, all queries holding the pair are getting back from a question gathering and observed as comparative query. From the comparative query and comparator pairs, all potential sequential prototypes are produced and assessed by means of calculating their dependability score. Patterns assessed as dependable ones are indicative extraction patterns and are added into indicative extraction pattern ordnance [11]. Then, novel comparator pairs are taken out from the question assortment by means of the newest indicative extraction patterns. The new comparators are additional to a consistent comparator repository and used as novel seeds for pattern learning in the subsequently iteration. All queries from which consistent comparators are taken out are disconnected from the assortment to permit finding novel patterns economically in later on iterations [10]. The procedure iterates in anticipation of no more novel patterns can be originated from the question assortment. There are two steps in our process: such as generation of patterns and pattern assessment.

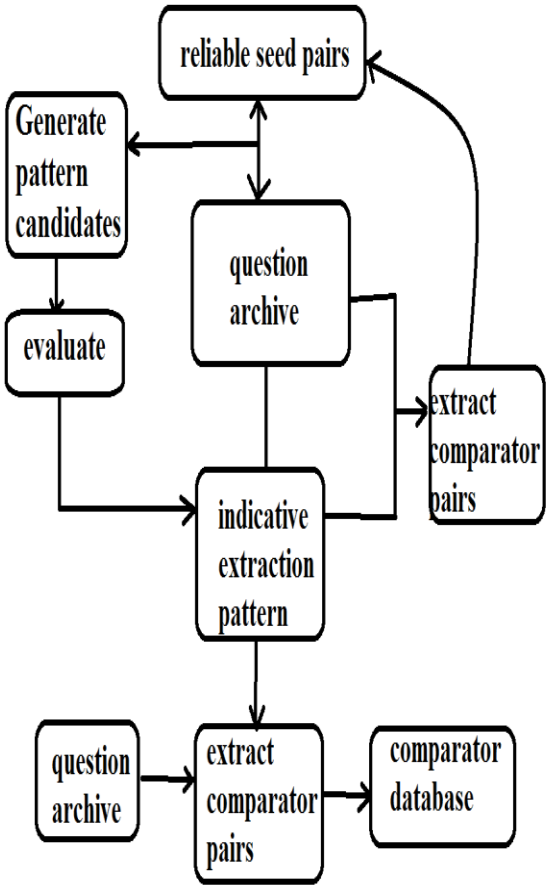

Fig1: An outline of the bootstrapping algorithm

3. RESULTS:

In terms of accuracy the Jindal and Liu process is reasonable to our method in comparative question recognition. On the other hand, the recall is considerably lower than ours. In terms of recall, our scheme outperforms Jindal and Liu method by means of 35% and 22% in comparative question recognition and comparator extraction correspondingly. In our examination the low recall of Jindal and Liu method is mostly caused by short coverage of learned class sequential rules prototypes over the test set. The end-to-end experimentation, our weakly supervised means performs considerably better than Jindal and Liu means. The consequence also emphasizes another benefit of our method that recognize comparative query and takes out comparators concurrently by means of one single pattern. Jindal and Liu method uses class sequential rules (CSR) and label sequential rules (LSR). Its performance goes down considerably due to error propagations. The performance of our bootstrapping algorithm is steady in spite of considerably different number of seed pairs generated through the two indicative extraction patterns. This consequence implies that our bootstrapping algorithm is not responsive to the selection of indicative extraction pattern.

4. CONCLUSION:

We expand a weakly supervised bootstrapping means to recognize comparative queries and take out comparators at the same time. Our weakly supervised indicative extraction pattern mining method is a pattern-based approach comparable to Jindal and Liu method, but it is dissimilar in a lot of aspects such as an

alternative of using various class sequential rules and label sequential rules, our process aims to become skilled at sequential prototype which can be able to be used to recognize comparative enquiry and take out comparators concurrently. Our weakly supervised indicative extraction pattern mining method is based on two important suppositions: It is for the reason that the greatest indicative extraction pattern is probable to be the most exact and pertinent pattern for the given query. If a sequential prototype can be used to take out numerous dependable comparator pairs, it is very probable to be an indicative extraction pattern. If a comparator pair can be taken out by an indicative extraction pattern, the pair is dependable. Our comparator mining outcomes can be used for a business search or product reference system. The investigational results show that our method is effectual in both comparative query recognition and comparator removal. It considerably progress recall in both responsibilities whereas uphold high accuracy.

REFERENCES:

[1] Mary Elaine Califf and Raymond J. Mooney. 1999. Relational learning of pattern-match rules for information extraction. In *Proceedings of AAAI'99 /IAAI'99*.

[2] Claire Cardie. 1997. Empirical methods in information extraction. *AI magazine*, 18:65–79.

[3]Dan Gusfield. 1997. *Algorithms on strings, trees, and sequences: computer science and computational biology*. Cambridge University Press, New York, NY, USA

[4] Taher H. Haveliwala. 2002. Topic-sensitive pagerank. In *Proceedings of WWW '02*, pages 517–526.

[5] Glen Jeh and Jennifer Widom. 2003. Scaling personalized web search. In *Proceedings of WWW '03*, pages 271–279.

[6] Nitin Jindal and Bing Liu. 2006a. Identifying comparative sentences in text documents. In *Proceedings of SIGIR '06*, pages 244–251.

[7] Nitin Jindal and Bing Liu. 2006b. Mining comparative sentences and relations. In *Proceedings of AAAI '06*.

[8] Zornitsa Kozareva, Ellen Riloff, and Eduard Hovy. 2008. Semantic class learning from the web with hyponym pattern linkage graphs. In *Proceedings of ACL-08: HLT*, pages 1048–1056.

[9] Greg Linden, Brent Smith and Jeremy York. 2003. Amazon.com Recommendations: Item-to-Item Collaborative Filtering. *IEEE Internet Computing*, pages 76-80.

[10] Raymond J. Mooney and Razvan Bunescu. 2005. Mining knowledge from text using information extraction. *ACM SIGKDD Exploration Newsletter*, 7(1):3–10.

[11] Dragomir Radev, Weiguo Fan, Hong Qi, and Harris Wu and Amardeep Grewal. 2002. Probabilistic question answering on the web. *Journal of the American Society for Information Science and Technology*, pages 408–419.

[12]Deepak Ravichandran and Eduard Hovy. 2002. Learning surface text patterns for a question answering system. In *Proceedings of ACL '02*, pages 41–47.

[13]Ellen Riloff and Rosie Jones. 1999. Learning dictionaries for information extraction by multi-level bootstrapping. In *Proceedings of AAAI '99 /IAAI '99*, pages 474–479.

[14]Ellen Riloff. 1996. Automatically generating extraction patterns from untagged text. In *Proceedings of the 13th National Conference on Artificial Intelligence*, pages 1044–1049.

[15]Stephen Soderland. 1999. Learning information extraction rules for semi-structured and free text. *Machine Learning*, 34(1-3):233–272.