

International Journal of Computer Science and Mobile Computing



A Monthly Journal of Computer Science and Information Technology

ISSN 2320-088X

IJCSMC, Vol. 3, Issue. 5, May 2014, pg.1173 – 1177

RESEARCH ARTICLE

AN APPROACH TOWARDS REMOVAL OF COMPARABLE ASPECTS BY ACCUMULATING QUERIES

Dr. M.V.Siva Prasad, P.Guru Lingam, N.Mahesh Babu

ABSTRACT:

The additional direction towards using the World Wide Web as the probable source of answer as well as the usage of the generic search engines to get back the information concerning the question and additionally post-processing to take out the answers for the queries. Weakly supervised indicative extraction pattern mining method is a pattern-based approach. It is dissimilar from other approaches such as an alternative of using various class sequential rules and label sequential rules, and aims to become skilled at sequential prototype which can be able to be used to recognize comparative enquiry and take out comparators. Jindal and Liu process is reasonable to system of weakly supervised in terms of accuracy in comparative question recognition and recall is considerably lower than weakly supervised indicative extraction pattern mining. Attempt on comparator mining is connected to the study on entity and relation removal in information extraction and make sure the entities are taken out from relative questions which are usually not necessary in indicative extraction task.

Keywords: Comparator mining, Pattern mining, weakly supervised system, Query

1. INTRODUCTION

Most of the work in text mining does not make use of any form of natural-language processing, treating documents as an unsystematic bag of words as is distinctive in information retrieval [4]. For the most part of the knowledge that may be mined from text cannot be exposed by means of an easy bag-of-words illustration. The entities referenced within a document and associations asserted concerning between these entities cannot be indomitable by means of a standard vector-space depiction. In spite of the modern progress, there still remain difficulties in the present day's systems of information extraction. Initially the accurateness and the robustness of the systems of machine extraction can still be extremely enhanced. In recent times, methods of pattern

learning were applied to take out comparable entities from query logs in addition to web documents and their experiments explained that query logs are better to web documents as resources from which to take out comparable entities [8]. Increasing numbers of internet applications that making use of technologies of information technology that construct knowledge bases from the web pages that generate databases of job listings; that put up query systems of newsgroup; and moreover generate predict databases from web pages [10]. Human error during throughout the information extraction is moreover caused by means of lapse of concentration while the errors of systems of automated extraction are due to its moderately shallow understanding of the input text [1]. Consequently the errors of machine generated are extremely difficult to track down and to precise. Subsequently, generating a system of information extraction in a novel domain is tricky and time consuming frequently necessitates huge efforts by means of domain specialists in addition to computational linguists [5]. The additional direction towards using the World Wide Web as the probable source of answer as well as the usage of the generic search engines to get back the information concerning the question and additionally post-processing to take out the answers for the queries [2]. Maximum efforts were done by Jindal and Liu on mining comparative sentences and relations which apply class sequential rules and label sequential rules cultured from annotated corpora to recognize relative sentences and take out relative relations correspondingly in the news and reassess domains and their methods classically can attain high exactitude but undergo from low recall. The analogous technique can be functional to comparative query appreciation and comparator mining from query [11]. To take out comparators from relative matter, should become aware of whether a question is relative or not and comparative query has to be a query with purpose to compare at least two entities [6]. A question is very probable to be a comparative query when it encloses at least two entities and it was controlled and expand a weakly supervised bootstrapping means to recognize comparative queries and take out comparators at the same time [3]. Making sure high recall is vital in our intended application situation where users can subject arbitrary queries and hence weakly-supervised bootstrapping pattern learning was implemented by means of efficiently leveraging unlabeled query.

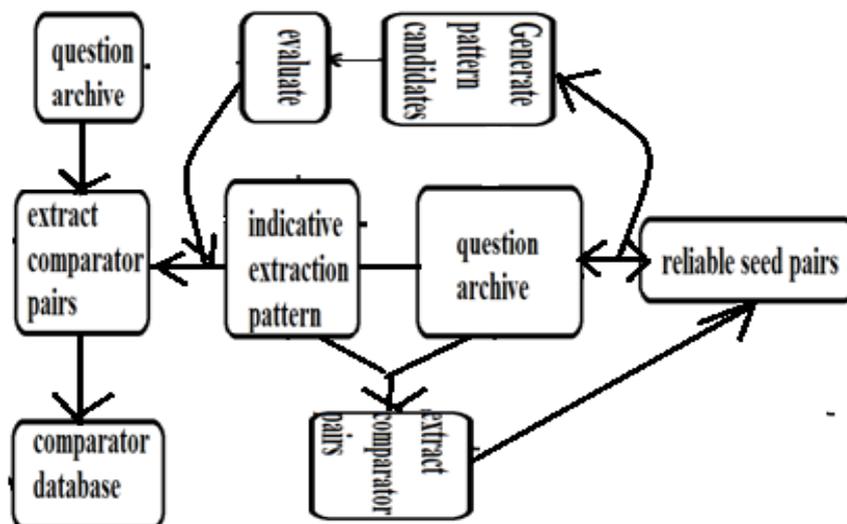


Fig1: An outline of the bootstrapping algorithm

2. METHODOLOGY

In decision-making, matching up substitute options is one of the necessary steps that were performed on a daily basis however necessitating high knowledge expertise [14]. In earlier information extraction bootstrapping methods shown in fig1 have been revealed to be very effectual and attempting to comparable to provisions of methodology by means of bootstrapping method take out entities by means of a specific relation [9]. Attempt on comparator mining is connected to the study on entity and relation removal in information extraction and make sure the entities are taken out from relative questions which are usually not necessary in indicative extraction task. The bootstrapping procedure begins with a single indicative extraction pattern and takes out a set of early seed comparator pairs. Each and every one query holding the pair are getting back from a question gathering and observed as comparative query for every comparator pair [13]. Entire potential sequential prototypes are produced and assessed by means of calculating their dependability score from the comparative query and comparator pairs. Patterns assessed as dependable ones are indicative extraction patterns and are added into indicative extraction pattern ordnance [15]. Weakly supervised indicative extraction pattern mining method is a pattern-based approach, 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 [7]. New comparators are additional to a consistent comparator repository and used as novel seeds for pattern learning in the subsequently iteration and the comparator pairs are taken out from the question assortment by means of the newest indicative extraction patterns. There are two steps in our process: such as generation of patterns and pattern assessment. All queries from which consistent comparators are taken out are disconnected from the assortment to permit finding novel patterns economically in later on iterations [16]. The procedure iterates in anticipation of no more novel patterns can be originated from the question assortment. 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. 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 [12]. 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.

3. RESULTS

Making sure high recall is vital in intended application situation where users can subject arbitrary queries and hence weakly-supervised bootstrapping pattern learning was implemented by means of efficiently leveraging unlabeled query. System of weakly supervised performs considerably better than Jindal and Liu means. Bootstrapping algorithm performance is steady in spite of considerably different number of seed pairs generated through the two indicative extraction patterns and hence bootstrapping algorithm is not responsive to the selection of indicative extraction pattern. Jindal and Liu process is reasonable to system of weakly supervised in terms of accuracy in comparative question recognition and recall is considerably lower than ours which is mostly caused by short coverage of learned class sequential rules prototypes over the test set. 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 and label

sequential rules and its performance goes down considerably due to error propagations. It recognizes relative sentences and takes out relative relations correspondingly in the news and reassesses domains and their methods classically can attain high exactitude but undergo from low recall.

4. CONCLUSION

The entities referenced within a document and associations asserted concerning between these entities cannot be indomitable by means of a standard vector-space depiction. Methods of pattern learning were applied to take out comparable entities from query logs in addition to web documents and their experiments explained that query logs are better to web documents as resources from which to take out comparable entities. Maximum efforts were done by Jindal and Liu on mining comparative sentences and relations which apply class sequential rules and label sequential rules cultured from annotated corpora to recognize relative sentences and take out relative relations. Weakly supervised indicative extraction pattern mining method is based on important suppositions such as the greatest indicative extraction pattern is probable to be the most exact and pertinent pattern for the given query; when a sequential prototype can be used to take out numerous dependable comparator pairs, it is very probable to be an indicative extraction pattern.

REFERENCES

- [1] S. Li, C.-Y. Lin, Y.-I. Song, and Z. Li, "Comparable Entity Mining from Comparative Questions," Proc. 48th Ann. Meeting of the Assoc. for Computational Linguistics (ACL '10), 2010.
- [2] D. Radev, W. Fan, H. Qi, H. Wu, and A. Grewal, "Probabilistic Question Answering on the Web," J. Am. Soc. for Information Science and Technology, pp. 408-419, 2002.
- [3] M.E. Califf and R.J. Mooney, "Relational Learning of Pattern- Match Rules for Information Extraction," Proc. 16th Nat'l Conf. Artificial Intelligence and the 11th Innovative Applications of Artificial Intelligence (AAAI '99/IAAI '99), 1999.
- [4] R.J. Mooney and R. Bunescu, "Mining Knowledge from Text Using Information Extraction," ACM SIGKDD Exploration Newsletter, vol. 7, no. 1, pp. 3-10, 2005.
- [5] C. Cardie, "Empirical Methods in Information Extraction," Artificial Intelligence Magazine, vol. 18, pp. 65-79, 1997.
- [6] N. Jindal and B. Liu, "Mining Comparative Sentences and Relations," Proc. 21st Nat'l Conf. Artificial Intelligence (AAAI '06), 2006.
- [7] Z. Kozareva, E. Riloff, and E. Hovy, "Semantic Class Learning from the Web with Hyponym Pattern Linkage Graphs," Proc. Ann. Meeting of the Assoc. for Computational Linguistics: Human Language Technologies (ACL-08: HLT), pp. 1048-1056, 2008.
- [8] E. Riloff, "Automatically Generating Extraction Patterns from Untagged Text," Proc. 13th Nat'l Conf. Artificial Intelligence, pp. 1044-1049, 1996.
- [9] D. Ravichandran and E. Hovy, "Learning Surface Text Patterns for a Question Answering System," Proc. 40th Ann. Meeting on Assoc. for Computational Linguistics (ACL '02), pp. 41-47, 2002.
- [10] E. Riloff and R. Jones, "Learning Dictionaries for Information Extraction by Multi-Level Bootstrapping," Proc. 16th Nat'l Conf. Artificial Intelligence and the 11th Innovative Applications of Artificial Intelligence Conf. (AAAI '99/IAAI '99), pp. 474-479, 1999.
- [11] L. Page, S. Brin, R. Motwani, and T. Winograd, "The PageRank Citation Ranking: Bringing Order to the Web," Stanford Digital Libraries Working Paper, 1998.
- [12] N. Jindal and B. Liu, "Identifying Comparative Sentences in Text Documents," Proc. 29th Ann. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval (SIGIR '06), pp. 244-251, 2006

- [13] Greg Linden, Brent Smith and Jeremy York. 2003. Amazon.com Recommendations: Item-to-Item Collaborative Filtering. *IEEE Internet Computing*, pages 76-80.
- [14] Dan Gusfield. 1997. *Algorithms on strings, trees, and sequences: computer science and computational biology*. Cambridge University Press, New York, NY, USA
- [15] Mary Elaine Califf and Raymond J. Mooney. 1999. Relational learning of pattern-match rules for information extraction. In *Proceedings of AAAI'99/IAAI'99*.
- [16] Stephen Soderland. 1999. Learning information extraction rules for semi-structured and free text. *Machine Learning*, 34(1-3):233–272.

Authors' Profile



Dr.M.V.Siva Prasad awarded PhD from Nagarjuna University,Guntur,received M.Tech. [SE] from VTU, Belgaum and B.E. [CSE] from Gulbarga University, presently working as principal in Anurag Engineering College (AEC),Ananthagiri(V),Kodad(M), Nalgonda(Dt.), Andhra Pradesh, India.



Mr.P.Gurulingam received Master of Technology (Computer Science & Engineering) from Jawaharlal Nehru Technological University (JNTUH). My research interests include Information Security, and Mobile Computing. Presently working as Associate Professor in CSE Department in Anurag Engineering College (AEC), Ananthagiri (V), Kodad (M), Nalgonda (Dt.), Andhra Pradesh, India.



N.Mahesh babu Pursuing Master of Technology (Computer Science)from Jawaharlal Nehru Technological University (JNTUH). My research interests include Information Security, Web Services, Cloud Computing, Data Mining and Mobile Computing. Presently working as PG Assistant in the department of CSE in Anurag Engineering College (AEC), Ananthagiri (V), Kodad(M), Nalgonda(Dt.), Andhra Pradesh, India.