

Extraction of Comparable Entity Using Decision Support System

¹S.Rajasekhar Reddy, ²M.Siva Lakshmi M.Tech, ³S.G.Nawaz M.Tech

¹PG Schalar, Dept. of CSE, S.K.D Engineering College, Anantapur, *Rajsreerama@Gmail.Com*

²Assistant Professor, M. Tech (CSE), S.K.D. Engineering College, *Shivacse06@Gmail.Com*

³Associate Professor, Dept. Of CSE, S.K.D Engineering College, Anantapur, *Sngnawaz@Gmail.Com*

Abstract: Any persons decision will be based on comparison of one thing with the other. But it is not easy to know the alternatives. For this purpose, a new approach is used automatically to extract comparative units from comparative questions. A method called bootstrapping is followed for the identification of comparable units. The experimental results show that the method achieves 82.5% in comparative question identification and 83.3% in extraction comparable unit.

The comparison activity is very common in the daily life. this requires high knowledge skill. Magazines such as Consumer Reports and PC Magazine and online media such as CNet.com strive. For example, if we consider "alto" and "swift" are comparable as "car manufacturers", but it is rarely seen that people comparing "alto Focus" and "swift". The results will be very useful in helping the users" exploration of alternative choices by suggesting similar units based on other users" previous requests. First it is detected whether a question is comparative or not.

Keywords: Comparative units, comparative questions, bootstrapping, decision support system

1. INTRODUCTION

The comparison activity involves: search for relevant web pages find competing products, containing information about the targeted products, identify pros and cons read reviews. The definition of comparative questions and comparators are as follows:

Comparator: An entity which is a target of comparison in a comparative question.

Comparative Question: A question that intends to compare two or more entities and it has to mention these entities explicitly in the question. For example, it is sensible to propose "iPod speaker" or "iPod batteries" if the user is interested in "iPod", but it is not compared with "iPod". The items that are comparable with "iPod" such as "iPhone" or "PSP" which were found in comparative questions predict simply based on item similarity between them. though they are all music players, "iPhone" is mainly a mobile phone, and "PSP" is mainly a portable game device.

2. RELATED WORK

Bootstrapping methods have been shown to be very effective in previous information extraction research CSR is a classification rule. It maps a sequence pattern $S(s_1s_2 \dots s_n)$ to a class C. In our problem, C is either comparative or non-comparative. A sequential pattern is called an indicative extraction pattern (IEP) if it can be used to identify comparative questions and extract comparators in them with high reliability. Once a question matches an IEP, it is classified as a comparative question and the token sequences corresponding to the comparator slots in the IEP are extracted as comparators. There are two key steps in our method: (1) pattern generation and (2) pattern evaluation three kinds of sequential patterns are generated from sequences of questions:

Generalized Patterns: A lexical pattern can be too specific. Thus, we generalize lexical patterns by replacing one or more words

Lexical Patterns: Lexical patterns indicate sequential patterns consisting of only words and symbols

Specialized Patterns: In some cases, a pattern can be too general.

The bootstrapping process starts with a single IEP. An error analysis for the cases where our method fails to extract correct comparator pairs:

- 23.75% of errors on comparator extraction are due to wrong pattern selection.

- The remaining 67.63% of errors come from comparative questions which cannot be covered.

To investigate the robustness of our bootstrapping algorithm for different seed configurations, it compares the performances between two different seed IEPs. The same techniques can be applied to comparative question identification and comparator mining from questions.

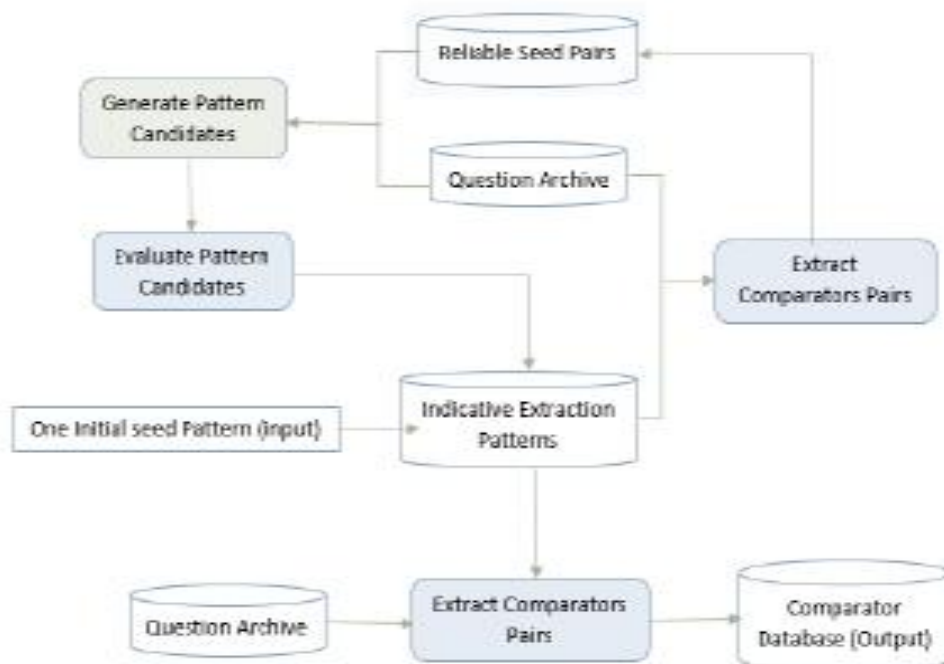


Fig1. Overview of the bootstrapping algorithm

CSR is a classification rule, which maps a sequence pattern S to a class C . Every CSR is associated to two parameters: support and confidence. **Confidence** is the proportion of sequences labeled as C in the sequences containing the S . **Support** is the proportion of sequences in the collection containing S . The question once matches an IEP then it is classified as comparative question and the token sequences corresponding to the comparator slots in the IEP are extracted as comparators. For generating sequential patterns two symbols, are used, they are #start and #end which are attached to the beginning and the end of a sentence in the question

3. EXPERIMENTS

Source Data: All the experiments are conducted on about questions of 60M.

Evaluation Data: Two data sets were created separately for estimation. First 5,200 questions are collected by sampling 200 questions.

Table1. Can show the difference between our mining and query recommendation comparison / article.

Channel	Gap	iPod	Kobe	Canon
Channel handbag	Gap Copuons	iPod nano	Kobe Brayant stats	Canon t2i
Channel sunglass	Gap outlet	iPod touch	Lakers Kobe	Canon printers
Channel earrings	Gap card	iPod best buy	Kobe espn	Canon printer drivers
Channel Watches	Gap careers	iTunes	Kobe Dallas Mavericks	Canon downloads
Channel shoes	Gap Casting call	Apple	Kobe NBA	Canon copiers
Channel jewelry	Gap adventures	iPod shuffle	Kobe 2009	Canon scanner
Channel clothing	Gap navy	iPod support	Kobe san Antonio	Canon jenses
Door	Banana republic	iPod classic	Kobe Brabant 24	Nikon

As shown in the table, usually suggests a mixed set of two types of queries related target entity "Google related search results": (1)specified in sub-queries to the original query (e.g., "Chanel Handbag "" Chanel ") and (2) its comparable entities (e.g., "Dior "" Chanel "). It confirms our claims that mining and query recommendation comparator / item are related but not the same.

4. RESULTS

The end-to-end experiments, weakly supervised method performs significantly better. It also analyzed the effect of pattern generalization and specialization. The performance of the bootstrapping algorithm is stable in spite of of significantly different number of seed pairs that are generated by the two IEPs. For example, it is compared to „HP’,,, Lexmark’, or „Xerox’, the printer manufacturers, and also compared to „Nikon’, „Sony’, or „Kodak’, the digital camera manufactures.

5. CONCLUSION

It presents a new supervised method for identifying comparative question sand extraction of comparator pairs simultaneously. The experimental results show that the method is effective in both comparative question identification and comparator extraction. The comparator mining results can be used for a commerce search or product recommendation system.

ACKNOWLEDGEMENT

The work was done when the author worked as an intern at Microsoft Research Asia.

REFERENCES

- [1] Mary Elaine Califf and Raymond J. Mooney. Relational learning of pattern-match rules for information extraction. In Proceedings of AAAI’99
- [2] Claire Cardie. 1997. Empirical methods in information extraction. AI magazine
- [3] Dan Gusfield. 1997. Algorithms on strings, trees, and sequences: computer science and computational biology. Cambridge University Press