Finding Better Options by Comparator Extraction for Decision Making With Comparative Questions

Priyanka R. Shirsath
PG Student
Department of Computer Engineering
SES's R.C. Patel Institute of Technology, Shirpur, India.
shirsathpiya@gmail.com

Ujwala M. Patil
Associate Professor
Department of Computer Engineering
SES's R. C. Patel Institute of Technology, Shirpur, India.
patil_ujwala2003@rediffmail.com

Abstract:-Nowadays internet has become very easy to access, people seem to use it all the time, which provide them a better guideline. Therefore people can make better decision even when they want to shop, do study, make online friends, want to use some services or wants to go out for vacation on some places. However searching for relevant options become time consuming and irritating process as there are several links available even for the single thing. Thus it becomes more difficult for people to choose one best option among so many of them as search engine shows many result. For this reason, this paper focuses precisely on showing the better result and better alternatives. For this, a weakly supervised bootstrapping approach is used which aims to identify comparative questions and evaluate comparable entities concurrently. By using this method people can have better option among the given choices and appropriate alternatives for the targeted entity in the entered query. The weakly supervised bootstrapping approach is very useful for knowing the better alternatives in people, products, places, any applications or any services. The experiment result shows the better performance of this method by measuring the Recall, Precision and F-Score.

Keywords:-*Comparable Entity Mining, Information Extraction, Bootstrapping, Sequential Pattern Mining, Pattern Generation, Pattern Evaluation, Entity Extraction, Entity Raking.*

*****

I. INTRODUCTION

Comparison is one of the best way to know the advantages and disadvantages of the things. For example, when someone is interested in particular place like lakes, National Parks, Cities for vacations or in products like mobile phones, cars, bikes for purchasing or in any services such as medical treatments, Hotel services, he/she would aspire to know the another possibility before they make the final decision. In this world people easily compare two or more things but only when they know everything about these entities. For example, choosing between car and bike going out for college is easy as we know about these transport so we choose the relevant one. That means comparison involves a high knowledge. Consumer Reports, PC Magazines and Online Media such as CNet.com are the sources that provide such a high knowledge of comparison and try to satisfy the need of comparison.

As most of the people use search engines to know about their interested things, it becomes time consuming and time eating process. Because there are so many of links available showing the result for people, history, places, product etc. Thus, this paper focuses only on showing the best option and best alternative which is mostly required to make final decision. For example, if someone is interested in purchasing laptop of better standard, he may posts a question about it such as "Which is better HCL or Acer?" then the best among the two of this query will be given as result and also Dell, Apple and HP etc. will be shown as an alternative option for the laptop depending upon ranking method [1] explain in this paper in methodology section.

The comparison becomes difficult for the entities having different functionalities which may create another problem. For example, one might compare "iPhone" with "Samsung Galaxy Note 3" as a mobile phone or compare "iPhone" with "PSP" as a portable game device. The comparison becomes complicated when an entity has different identities. For Example, people do compare "Paris versus London" as a location and "Paris versus Nicole" as a celebrity.

As comparison is very crucial to make decision, plenty of questions are posted online. For Example, "Itautec or Lanix? Which laptop brand is better?" this is the comparative question. There can be a question having two or more entities but may not have comparison intention. For example, "Does Shri Lanka have more caves than Malashiya?"

Therefore, the comparative question and comparators are defined as follows:

- **Comparative Question.** It is a question having two or more entities with comparison intention.
- **Comparator.** It is posted entity/items which is a target of comparison in comparative question.

According to the above definition a comparative question must compare two or more things or comparing at least two entities/items. For example,

Q1. "Which one is best?"
Q2. "Is Suzuki Hayabusa coolest bike?"
Q3. "Which antivirus is better Avast or Avira?"

From above definition, Q3. Is the comparative question and the "Avast", "Avira" are the comparators. Accordingly, as stated in the above definition the questions Q1. And Q2. Are not comparative question.

The goal of this work is to detect the comparative question and mine comparators concurrently. For this a weakly supervised bootstrapping approach is used. The subsequent comparators are shown as result by ranking them. The result will be more useful for the people who do not know the alternative options for the entities in which they are interested. This method is also useful for recommender system and for any company launching the new product so that they can know their competitors.

The remaining paper is organized as follows: section II is about previous works. Section III represents the weakly supervised bootstrapping method. It enlighten that how comparative question is identified and how comparative question is used to detect comparators. Section IV presents the assessment of Existing Method and Improved Method.

II. RELATED WORK

The work in past few years is done on sentiment and opinion extraction and classification. But specifically detecting comparative question and mining comparators at the same time was not done yet.

Z. Kozareva et al. presented an approach to weakly supervised semantic class learning from web, which captures two properties associated with pattern-based extractions: popularity and productivity. There were errors created by their algorithm which were initiated by incorrect proper name extraction [2].

The work on comparator mining is associated to the research on entity and relation extraction in information mining. Precisely, the most applicable work is done by Nitin Jindal and Bing Liu on extracting comparative sentences and mining the relation of comparative sentence. Their methods applied Class Sequential Rules (CSRs) and Label Sequential Rules (CSRs) learned from annotated corpora to ascertain comparative sentences and excerpt comparative relations respectively in the news and review domains. In their experiment their method can typically attain high precision but struggle from low recall [3, 4].

Kennedy, C. studies the gradability of comparatives and measure of gradability. The logic, on which the semantic analysis is constructed, is not directly applicable to identifying comparative question and mines comparators [5].

Hu and Liu, presented some methods to abstract opinions from customer reviews, so that to identify the product features mentioned on by customers and to determine whether the opinions are positive or negative [7].

D. Ravichandram et al. developed a method for learning surface text patterns automatically. In the experiment only those questions were used which do not have a long answers because it was affecting precision of patterns [8].

Turney offered a method constructed on mutual information between document phrases and the words “excellent” and “poor” to discovery suggestive words for sentiment classification [9].

Ellen Rilof et al. presented a multilevel bootstrapping algorithm that generates both the semantic lexicon and extraction patterns simultaneously. As input, the technique requires unannotated training texts and a handful of seed words for a category. To make the method more vigorous, a second level of bootstrapping is added (metabootstrapping) that retains only the most reliable lexicon entries produced by mutual bootstrapping and then restarts the process. The algorithm constructs high-quality dictionaries for several semantic categories [10].

III. METHODOLOGY

The bootstrapping methods has been shown a great effect in previous information extraction research [2], [6], [8], [10], [11]. The weakly supervised bootstrapping approach is similar to [3, 4] but different in many aspect. The weakly supervised bootstrapping method for comparator mining is a pattern-based approach which goals to learn sequential pattern further which can be used to identify comparative question and mine comparators instantaneously.

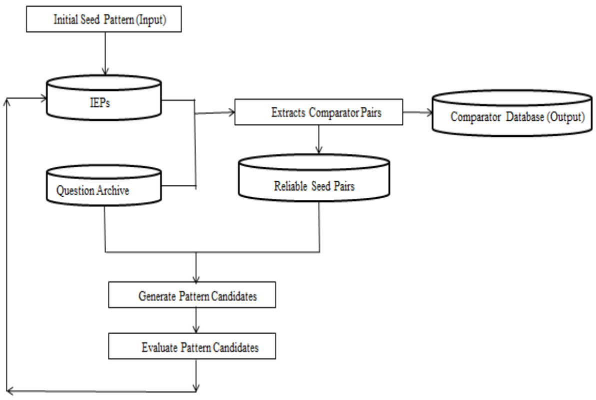

Figure 1. Flow of Weakly Supervised Bootstrapping Method

1.1 Steps of Bootstrapping Method

- The bootstrapping process starts with a single a sequential pattern which is an IEP used for detecting the comparative questions.
- From that sequential pattern a set of initial seed comparator pairs is mined.
- For each comparator pair, all questions containing the same comparator pair are retrieved from a question collection and regarded as comparative questions.
- By using recently identified comparative questions and comparator pairs, all possible new sequential patterns are generated and evaluated by measuring their reliability score.
- Patterns evaluated as reliable ones are IEPs and are also added into an IEP repository.
- From the question collection, using the latest IEPs the new comparator pairs are extracted.

- The new comparators are added into a reliable comparator repository as they are mined from IEPs and used as new seeds for pattern learning in the next iteration.
- The process repeats until no more new patterns are found from the question collection.

Input: *CP*, *G*

Initialize solution: *Q*←*{}*, *P*←*{}*, *P*new←*{}*, *CP*new←*CP*

```
Repeat
P← P + Pnew
Qnew← ComprativeQuestionIdentify( CPnew )
Q← Q+ Qnew
for qiϵ G do
if IsMatchExistingPatterns(P, qi) then
Q← Q - qi
end if
end for
Pnew← MineGoodPatterns( Q )
CPnew← {}
for qiϵ G do
cp← ExtractComprableComparators(P, qi)
If cp≠ NULL and cp ∉ CP then
CPnew← CPnew + {cp}
end if
end for
until Pnew = {}
return P
```

Figure 2. Bootstrapping Algorithm [1].

Bootstrapping method has two parts:

1. Mining Indicative Extraction Pattern

An Indicative Extraction Pattern [IEP] is a sequence of the pattern which is used to identify comparative question and mine comparators. An IEP is a combination of a POS tag, a symbol denoting either a comparator ($C), the beginning (#start) and the end of a question (#end).

Table 1:Example of candidate IEP of the question "Which place is better Hawai or Miami?"

	Sequential Pattern
1.	<#start Which place is better $C or $C? #end>
2.	<#start Which place is better Hawai/NNP or Miami /NNP? #end>
3.	<#start $C/NNP or $C/NNP #end>
4.	<#start Which/WDT place is/VBZ better/JJS $C or/CC $C? #end>
5.	<#start Which/WDT place/NN is/VBZ better/JJS $C or/CC $C? #end>

IEP mining method is depend upon the following assumption:

- If a sequential pattern is used to mine many comparators then it is reliable IEP.
- If any comparator pair is used to mine new IEPs then it is said to be reliable comparator pair.

Once any question hit by user matches with an IEP, it is classified as comparative question. Therefore in above steps, the initial sequential pattern is considered as IEP. To generate a comparative question as sequential pattern and evaluate new IEPs two key steps are used which are follows:

i. Pattern Generation
ii. Pattern Evaluation

i. Pattern Generation

When a question is identified as a comparative question then for mining comparator pair, those comparators are replaced by $Cs symbols and/or assign a POS tag with two symbols #start and #end are attached at the beginning and ending respectively. The three types of pattern are generated which are as follows:

- Lexical Pattern: It is the pattern having only start (#start), end (#end) and words ($C) that means it contains only words and symbols. It has a restriction that a pattern should contain more than one $C.

- Generalized Pattern: It is pattern in which lexical pattern is generalize by replacing the two or more words by POS tags [3]. That is it replaces the comparators by POS tags with (#start) and (#end) at the beginning and ending respectively.

- Specialized Pattern: Sometimes, a pattern can be too general. For example, "Intel or AMD" is also a comparative question. But the pattern <$C or $C> is too general. Because there can be some non-comparative question such as "True or False" matching the pattern. So, in specialized pattern each word is replaced by POS tags and the comparators are replaced by $C [3].

Table 2:Candidate IEP example for question "Which car is faster: Audi or BMW?"

Sequential Patterns	
Lexical Pattern	<#start Which car is faster $C or $C? #end >
Generalize Pattern	<#start Which car is faster Audi/NNP or BMW/NNP? #end >
Specialize Pattern	<#start Which/WDT car/NN is/VBZ faster/JJR $C or/CC $C? #end >

ii. Pattern Evaluation

According to the assumptions that if a sequential pattern is used to mine many comparators then it is reliable IEP. The reliability score R1(pi) for the candidate pattern pi at iteration n is defined as follows:

$$R1(pi) = \frac{\sum_{\forall cpi \in CP^{n-1}} N_Q(pi \rightarrow cpi)}{N_Q(pi \rightarrow *)} \qquad (1)$$

Where, pi-can mine the known reliable comparator pair cpi.

CPn-1 –Repository of reliable comparator pair which collected the comparator pairs until n-1 iteration.

NQ (X) – means the number of questions satisfying the condition X.

(pi→cpi)– means cpi can be mine from a question by applying a pattern pi.

(pi→*) –denotes any question containing pattern pi.

Sometimes equation (1) can suffer from partial knowledge about reliable comparator pairs. It is because very few reliable pairs are normally learned in early stage of bootstrapping. So, the value of (1) is may underestimated which will affect the effectiveness of differentiating IEPs from non-reliable patterns. Therefore support S is defined for comparator pair cpj which is mined by set of candidate pattern Pn as follows:

$$S(cpj) = NQ(Pn \rightarrow cpj) \qquad (2)$$

Where, NQ (X) – means the number of questions satisfying the condition X.

(Pn → cpj) – means one of the pattern in Pn can mine cpj in certain questions.

If cpj is mined by many candidate pattern in Pn, then it is taken as reliable comparator pair in next iteration. Considering this assumption, a comparator pair is taken as a reliable one when its support S is more than the threshold value α = 3. If Support S is less than 3, then value of R2(pi) is zero. Hence by using reliable comparator pairs, another reliable score R2(pi) is defined as follows:

$$R2(pi) = \frac{\sum_{\forall cpj \in CP^{n}_{rel}} N_Q(pi \rightarrow cpi)}{N_Q(pi \rightarrow *)} \qquad (3)$$

Where, CPnrel – represents a set of reliable pairs based on the set of candidate pattern Pn

Therefore, by using equation (1) and (3) final reliability score R (pi)nfinal is defined as follows:

$$R(pi)_{nfinal} = \lambda R1(pi) + (1-\lambda) R2(pi) \qquad (4)$$

Here λ is set 0.5 for equal importance to R1(pi) and R2(pi). By using equation (4) candidate patterns are evaluated and the patterns whose reliability score more than threshold are selected as IEPs.

2. Comparator Extraction

When a user's question matches with IEP then it is considered as comparative question. In that comparative question, bootstrapping algorithm checks that is any singular noun or plural noun present then it is considered as comparators and these comparators are extracted. For comparator extraction three strategies are used which are as follows:

- Random Strategy: It casually selects any candidate pattern among set of patterns.
- Maximum Length Strategy: It selects the longest pattern among the set of pattern. As this pattern will be the largest one, it will have more tokens in it. So this pattern will be more suitable for user questions.
- Maximum Reliability Strategy: Here, the most reliable pattern among the set of patterns which will be used to apply to the question.

1.2 Comparator Ranking

As the question is identified as comparative question and from that question comparators are mined, it becomes very important to rank those mined comparators to be ranked for the user's entered entity. It has two methods which are as follows:

1. Comparability-Based Ranking Method

When comparators are compared with user's entity more frequently then it becomes convenient to rank them. For this, a ranking function RFreq is defined as follows:

$$RFreq(ci ; qi) = N(Qci,qi)$$

Where, ci – comparator

qi – user's entity

Qci,qi – is a set of comparative question from which comparator pairs can be mined.

This ranking function can rank the comparators by calculating that how many times comparator ci is compared with the users given entity qi.

2. Graph-Based Ranking Method:

When the user's input entity does not frequently compared with comparators stored in the reliable comparators pair's database, then the comparability-based ranking method faces problems to show result. Therefore the graph-based ranking method can be used. If any comparator is frequently compared with user's entity in the area of user's interest then that comparator can be considered as representative and it is taken as the baseline. Here PageRank is used as a graph- based method [1].

IV. EXPERIMENT

1.3 Data Source

The experiments are conducted on the datasets which are requested from Yahoo! Labs[1]. By using Yahoo! Labs, the L4, L8, L9 and L12 are requested from which up to 850 questions are accumulated as comparative questions. However to perform the existing method experiments completely, total 992 questions are needed that is 139 questions for SET A and 853 questions for SET B.

As given in the base paper that, two annotators were asked to label each questions from Yahoo! Answers 26 categories. Similarly up to 300 questions are accumulated from Yahoo! Answers as comparative questions. Thus, the base paper experiments are performed only on 992 questions as given in the base paper.

1.4 Ours Improvement

Some modification are done in weakly supervised bootstrapping method which shown in table below as "Improved Method". In modification we have set support S value at 2 by assuming it will allow the R2(pi) value to calculate early because support S was set to 3 in Existing Method experiments. This support S is used to calculate R2(pi) in equation (3). We also set parameter λ=0.7 giving R1 (pi) more importance which is used to calculate the final reliability score. The performance is also observed by setting the λ to 0.3 means giving R2 (pi) more status, but if we do so then less IEPs will be accumulated because R2(pi) is calculated only when the support S is 2. As a result, Improved Method shows slight improvements in all the experiments shown in the tables.

Also, Existing Method was only taking question which were starting by WH type questions. For example, "Which is fastest motorcycle Kawasaki Ninja H2 or Dodge Tomahawk?" But Improved Method can also identifies the questions as a comparative question which are starting by verbs and noun. For example, questions starting by verbs "Is Xylo is better than Mobilio? Which one you suggest?" and question starting by noun "Canon 60D or Sony Alpha a6000? Which is better?"

1.5 Evaluation of Data for Experiments Result

The two sets of questions are created in which SET A contains 139 questions and SET B contains 853 question which are identified as comparative question [1].

For Identifying comparative question which is given in first column of the Table 2 the total 992 questions are taken by adding the SET A and SET B questions. For Extracting comparator pairs only given in the column two, SET B is taken as the input.

Table 3: Performance Comparison between Existing Method and Improved method

	Identification Only (SET A + SET B)		Extraction Only (SET B)		All (SET B)	
	Existing Method	Improved Method	Existing Method	Improved Method	Existing Method	Improved Method
Recall	0.817	0.825	0.760	0.767	0.760	0.770
Precision	0.833	0.839	0.916	0.919	0.776	0.785
F-Score	0.825	0.831	0.833	0.836	0.768	0.777

Table 4: Effect of Different Pattern between Existing Method and Improved method

Patterns	Recall		Precision		F-Score	
	Existing Method	Improved Method	Existing Method	Improved Method	Existing Method	Improved Method
Original	0.689	0.692	0.449	0.450	0.544	0.545
Specialize Pattern	0.731	0.742	0.602	0.619	0.665	0.674
Generalize Pattern	0.760	0.761	0.776	0.777	0.768	0.769

Also for "All" which is shown in the column three of Table 3. SET B is used as input. Here "All" means overall performance of bootstrapping method that is identifying comparative question, generation of pattern, evaluation of patterns and mining comparators.

The effect of specialization and generalization is also analysed in Table 4. In which improved method shows slight improvements in performance. This result shows flexibility of the learning pattern to identify and accumulate various expressions of comparative questions.

Table 5: Performance Comparison of Different Extraction strategies between Existing Method and Improved method

	Recall		Precision		F-Score	
Extraction Strategies	Existing Method	Improved Method	Existing Method	Improved Method	Existing Method	Improved Method
Random	0.744	0.750	0.891	0.895	0.810	0.816
Max Length	0.760	0.767	0.916	0.919	0.833	0.836
Max Reliable	0.747	0.751	0.891	0.895	0.813	0.819

The Table 5. Shows the efficiency of different strategies of comparator extraction. From the table below, it can be said that Max Length Strategy works better where Random Strategy and Max Reliability Strategy shows similar performance. This is because most of applicable IEPs have the same reliability score.

V. CONCLUSION

When people uses search engines to know the better one in two or more options he/she will get many links as search engine shows multiple results. But Weakly Supervised Bootstrapping Method focuses only on comparative question and comparable entities which means it directly gives which option is better as well as it also gives better alternative options. Consequently, this Weakly Supervised Bootstrapping Method is very useful for everyone who are comparing two or more products, services, companies, colleges, places like wise. The Weakly Supervised Bootstrapping Method effectively mine the comparable items from a comparative questions and give best result for the online posted entity instead of using separate Sequential Rules. It significantly improves recall in both tasks that is identifying comparative question and extracting comparators while maintaining high precision. This method can effectively be used for commerce search or product recommendation system. Also, this Approach can provide useful information to companies which want to identify their competitors.

VI. ACKNOWLEDGMENTS

I express my sincere thanks to Yahoo! Labs for permitting me datasets for my experiments and my guide Prof. Ujwala M. Patil. The faith & confidence shown by her in me, boosted me and motivated me to perform better.

REFERENCES

[1] S. Li, C.-Y. Lin. Y-I. Song, and Z. Li., "Comparable Entity Mining from Comparative Question's", IEEE Transactions on Knowledge and Data Engineering, vol 25, no. 7, July 2013.

[2] Z. Kozareva, E. Rilof, 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.

[3] N. Jindal and B. Liu. "Mining Comparative Sentences and Relations", Proc. 21st Nat'l Conf. Artificial Intelligence (AAAI '06), 2006.

[4] 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.

[5] Kennedy C, "Comparatives semantics of. In Enclycopedia of Language and Linguistics", Second Edition, Elsevier, 2005.

[6] 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.

[7] Hu, M., and Liu, B., "Mining and summarizing customer reviews",
KDD'04, 2004.

[8] 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.

[9] Turney, P., "Thumbs Up or Thumbs Down? Semantic Orientation applied to Unsupervised Classification of Reviews", *ACL'02*.

[10] E. Rilof and R. Jones, "Learning Dictionaries for Information Extraction by Multi-Level Bootstrapping", Proc. 16th Natl Conf. Artificial Intelligence and the 11th Innovative Applications of Artificial Intelligence Conf. (AAAI '99/IAAI '99). pp. 474-479, 1999.

[11] E. Riloff, "Automatically Generating Extraction Patterns from Untagged Text", Proc. 13th Nat'l Conf. Artificial Intelligence, pp. 1044-1049, 1996.