International Journal of Advanced Information Science and Technology (IJAIST) ISSN: 2319:268
Vol.2, No.6, June 2013 DOI:10.15693/ijaist/2013.v2i6.5-8

Semantic Analysis of Comparative Sentences in Text Document: *A Survey*

Saritha Khethawat
Assistant Professor
Department of Computer Sciences & Engineering
Maulana Azad national institute of technology
Bhopal, India

Trilochan Prasad Dewangan
M.Tech. Scholar,
Department of Computer Sciences & Engineering
Maulana Azad national institute of technology
Bhopal, India

***Abstract*:- This paper studies the comparative sentences in text document to identifying the product features opinion. Semantic analyses of comparative sentences contribute to describe the information about product and further use it for various purposes. Comparative sentences comparison two object bases on subjective or objective on object. An example of comparative sentence is "The Camera quality of Samsung is better than Nokia.". In this research, we basically aim to mining the text document of comparative sentences and analysis the product. On summarization task the product is analysis base on positive or negative value. This value is obtained from the product opinion on the review, blogs and comments of the product. Comparative sentences is one type of the opinion sentences but it compare thought the others product. To comparing any object with another object is the best way to analysis the features of object. In this paper, we analysis some methods to survey on comparative sentences in text document and analysis the text mining on the document.**

I. INTRODUCTION

Comparative sentences are the sentences which compare more than two objects. Comparisons are one of the best convincing ways of evaluation. Compare of any product or object is extracting the information about main feature of product or object. Extracting Features from text document is useful for many fields. For Example, in the business environment, whenever a new product comes into market, the product manufacture wants to know consumer opinion on the product in form of comparisons of other product available in market. Such information is easily available on the web in the form of product reviews, forum discussion, blogs, etc. Extracting such information can use businesses in their marketing and product efforts. In this paper, we focus the methods which extract such information on text document.

Comparative sentences are quite different from sentiments and opinion sentence, which are subjective. Comparisons on the other hand can be subjective or objective. Subjective comparative sentences may be "Car X is much better then Car Y" compare thought the particular subject. But an objective comparative sentence may be "Car X is 2 feet longer than Car Y" compare through the particular object.

We can see that in general comparative sentences use quite different language constructs from typical opinion sentences. The problem of identifying comparative sentences is challenging because although we can see that the above example sentences all contain some indicators comparative adverbs and comparative adjective, i.e. "better", "longer", many sentences that contain such words are not comparatives, e.g., "I cannot agree with you more". Similarly, many sentences that do not contain such indicators are comparative sentences, e.g., "Cell phone X has Bluetooth, but cell phone Y does not have."

In this paper, we study the some methods for mining the text document for our future work. First, Hybrid Patterns [3] are extracting the prime comparative patterns from comparison frame in CFN (Chinese FrameNet). Seconds, SRL (Semantic Role Labeling) Nitin Jindal [7] [8] work are consist in comparative sentences classification and then extract the features, Thirds, OPINE an information extraction system which mines reviews to build a product feature [2]. And CSM (Comparative Sentences Mining) is divided the work to identify the comparative sentences and extract comparative relations of the sentences [1].

II. HYBRID PATTERNS

In this method use the prime comparative patterns from the "comparison" frame in Chinese Frame Net [3]. It contain the some factors such as position of comparative entities, lexical units, part of speech and comparative result, the derivational generalized patterns are constructed. There are mainly two part Prime Pattern and Generalized pattern [3].

a. Prime pattern work on position information of entities in comparative sentences. It is very important to comparative entities recognized and labeling.

b. Generalized pattern is work on more than two entities in comparative word.

Prime pattern cannot summarize all comparative sentence types which appear in review texts. Rank information and position information of frame element is considered. Generalized pattern is based on the result of prime pattern. Prime pattern are three patterns extracted based on CFN and the position information of comparative [3].

Pattern a: {<Phen_1: CE>, <tgt: CW>, <Phen_2: CE> [,<null: CR>]};
Pattern b: {<Phen_1: CE>, <tgt: CW>, <Phen_2: CE>, <tgt: CW> [,<null: CR>]};
Pattern c: {<tgt: CW>, <Phen_1: CE>, <null: SS>, <Phen_2: CE>, <null: CR>}.

Fig1. View of prime patterns

In prime patterns "{}" is a complete comparative sentences, "CW", "CE" and "CR" are denoted by comparative word, comparative entity and comparative result word respectively. "< >" is group of binary terms, first terms (i.e. "Phen_1", "Phen_2" and "tgt")) is define the comparative pattern and second term is frame element or lexical unit of CNF [3].

III. SEMANTIC ROLE LABELING (SRL)

Semantic Role Labeling is introduces by Gildea [4] to check the problem of shallow semantic parsing by statistical machine learning methods. SRL is not described the details semantic information of a text document, it defined the relationship between predicate and its semantic roles. Example of the task of SRL in sentences: The U.S. Congress recently passed the inter-state banking law. The predicate "pass" is used with three arguments: "The U.S. Congress" as an agent; "the inter-stage banking law" as a theme; and "recently" as a temporal argument. Many interests have generated in predicate argument structure for content-related natural language processing tasks, such as Question Answering or Information Extraction.
This method two-level semantic role labeling algorithm is designed to analysis document. In this [3] first-level labeling defined the function of comparative word identifying and labeling, comparative entity identifying and labeling. Second-level identifying and label the comparative features on the basis of first-level labeling results. In semantic role labeling (SRL), first suppose that sentences is a comparative sentence then applies the word-segmented and POS-tagging. Second-level labeling classified into two types of comparative entities. In which one type search the preposition phrase in comparative features and another type search the comparative sentences those features are denoted in frame word [3].

a. First-level labeling algorithm matches the word in the sentences and checks it to comparative word set. If not match then add in the comparative word set. It used it define the precision, recall and F-value.
b. Second-level algorithm search preposition phrase in the sentences and mark in label the noun or noun phrases.

Semantic Role Labeling method is generalized the accuracy and coverage rate of the labeling results. In comparative sentences are usually two or more comparative entities and comparative features. SRL is helpful to better understand the relationship between them.

To study of mining Chining Comparative sentences by using semantic role labeling we find out the comparative opinion have six elements: Holder(which is holder of opinion), Entity1 & Entity 2 (which is entity of two object to be compared), comparative predicts, attribute(property of Entity) and sentiments(those word distinguishes the entity1 to entity2). Comparative Predicts are define to be left of the Entity1 and right of the Entity2.

IV. CSR WITH MULTIPLE MINIMUM SUPPORTS

In this paper proposed the combination of class sequential rule (CSR) mining and machine learning. Sequential patterns in the rules are used as features. Class sequential rule are basically defines form traditional sequential patterns. This method is attached the class label, which results in a rule with a sequential pattern on the left-hand-side of the rule, and a class on the right-hand-side of the rule. These methods apply the class sequential rules and mining learning algorithm to build a class based on the rules.

A Class Sequential Rule is a rule with a sequential pattern on the left and a class label on the right of the rule. This method uses a single minimum support to control the rules to be generated.

Multiple minimum supports method overcomes the single minimum supports. This algorithm needs to sort the words in each sentence in a particular order, which is not permitted in this case since that destroyed the word sequences.

In this paper, construct the dataset from the text documents than work on the sentence level. This model cannot use the raw words in each sentence because the contents of some sentences may be very different their underlying languages pattern can be the same.

a. The keyword strategy: It is easy to construct the small set of dataset of comparative sentences. We check only those sentences which contain at least one keyword, then generated the class sequential rule.
b. Sequence database: it combine the keyword and POS tag to form a single unit. Then class is attached to

each sequence according to whether the sentences are comparative or non-comparative.

V. OPINE

OPINE is the review mining system [2], OPINE include the use of relaxation labeling to find the semantic orientation of words in the context of given product features and sentences.

OPINE is the first to report its precision and recall on the task of opinion phrases polarity extraction and opinion phrases polarity determination in the context of known product features and sentences. OPINE extract the following types of product features: properties, parts, features of product parts, related concepts, parts and properties of related concepts.

OPINE also extract opinion phrases, which is adjective, noun, verb or adverb phrases representing customer opinions. Opinion can be positive or negative and very in strength. OPINE decompose the problem of review mining into subtasks. These tasks identify the product features then identify opinion regarding product features and determine the polarity of opinions. OPINE is divide the problem in subtask and analysis the product features, strength of features.[2]

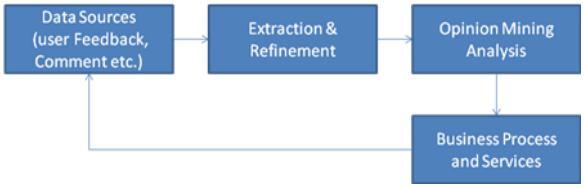

Fig2. OPINE overview

OPINE main goal to find the feature and opinion of product. OPINE extract the explicit features for the given product class. First check the part and properties of the product class. Then, system finds the related concept of the part and properties to be collect in set of class. Result will be contributes the set of final explicit features.

VI. COMPARATIVE SENTENCES MINING

The objective of this method is to extract and to analyze comparative sentences in evaluative texts on the web. It work in two steps [7], first identify the comparative sentences from the texts and classifying the identified comparative sentences into different classes. Second, extract relation from the identified sentences.

This task in solved in class sequential rule and label sequential rules. Label sequential is used for extraction. Comparative sentences classified into two types gradable and non-gradable. Gradable is based on such relationships as greater or less than, equal to, and greater or less than all others. Non-Gradable comparatives express implicit comparisons [7].

In this paper, first directly apply the class sequential rule (which are define above) and used the machine learning algorithm to build a class base on the rule. In this paper, CSR used with the multiple minimum supports, which support the frequent word to generate the various number of CSR rules. It constructs the dataset for mining the text document. There are some methods to constructs the dataset.

The Keyword Strategy: To designing the strategy easily find out small set of keywords to learning stage. We only consider those sentences which contain at least one keyword and generate some class sequential rule s to filter the sentences into two categories comparative and non-comparative. It find some keyword manually and collect in subset of comparative sentences then use the WordNet [5] to find their synonyms and some phrases to be add in dataset of comparative sentence.

Building the Sequence Database: to follow some steps to generating the database.

1. Each sentences containing some at least one keyword or key phrases, we used that word to add in our database.
2. We combine the word with their POS tag to form a single word. Because some keyword have multiple POS tags depending in their use.
3. A class is set according to sequence it used in comparative or non-comparative sentences.

ACKNOWLEDGMENTS

In this paper we try to involve all the possible methods of comparative sentences to extract the information on text document. Special thanks to my guide Saritha Khethawat for many useful discussions, annotation of data and helping me in this literature survey.

REFERENCES

[1] Feng hou, guo-hui li, Mining Chinese Comparative sentences by Semantic role labeling, Proceedings of the Seventh International Conference on Machine Learning and Cybernetics, Kunming, 12-15 July 2008.

[2] Ana-Maria Popescu and Oren Etzioni. Extracting Product Features and Opinions from Reviews. Proceedings of the con ference on Human Language Technology and Empirical Methods in Natural Language Processing. October 2005.

[3] Suge Wang, Hongxia Li, Xiaolei Song. Automatic Semantic Role Labeling for Chinese Comparative Sentences Based on Hybrid Patterns. 2010 International Conference on Artificial Intelligence and Computational Intelligence. Oct. 2010.

[4] Gildea D, Jurafsky D. Automatic labeling of semantic roles. Computational Linguist, 2002, 28(3):245-288.

[5] Fellbaum, C. WordNet: an electronic lexical database, MIT Press, 1998.

[6] Michelle Annett and Grzegorz Kondrak, A Comparison of Sentiment Analysis Techniques: Polarizing Movie Blogs. Proceedings of the Canadian Society for computational studies of intelligence, 21st conference on Advances in artificial intelligence. May 2008.

[7] Nitin Jindal and Bing Liu. Identifying Comparative Sentences in Text Documents. In Proceedings of the 29th Annual International ACM SIGIR Conference on Research & Development on Information Retrieval (SIGIR-06), Seattle 2006.

[8] Nitin Jindal and Bing Liu. Mining Comparative Sentences and Relations. In Proceedings of 21st National Conference on Artificial Intelligence (AAAI-2006), July16.20, 2006,Boston, Massachusetts, USA.