# Generating the Structured Answer for Feature-based Sentiment Questions from Customer Reviews

Nu Nu War
*University of Computer Studies, Mandalay*
*nuwar99@gmail.com*

## Abstract

*With the rapid growth the rich opinionated information on various features of product, people are more interested in opinion questions that can immediately reflect others' opinions. Opinion question answering is an important research area in opinion mining and Natural Language Processing. Question analysis and answer generating are also the important keys for this system. A new approach, pattern knowledge in linguistic with the POS tagging are developed for feature based sentiment questions analysis on the products. This paper also introduces a Wordnet based methodology to retrieve the review sentences from the customer reviews. Finally, the system aggregates the quality review sentences by using polarity grouping and a simple ranking approach to get the complete set of answer for sentiment questions.*

## 1. Introduction

Due to rapidly increasing the scale of user generated content (e.g. customer reviews, forum posts, and blogs) on the web, numerous consumer reviews are now available online, and these reviews contain rich opinionated information on various aspects of products. They are naturally valuable resources for answering opinion questions about products, such as "How do people think about the battery of Nokia 6600?"[6].

Question Answering (sentiment-QA) on products seeks to uncover consumers' thinking and feeling about the products or aspects of products. It is different from traditional factual QA, where the questions ask for the fact. Earlier research in this field mainly focused on factual question answering which is insufficient for most real-life applications.

One way in which opinion questions differ from many types of fact-based questions is that, rather than having a single best answer, opinion questions often have many relevant answers, which may reflect a variety of different viewpoints. Opinion questions have very different characteristics when compared with fact-based questions: opinion questions are often much longer, more likely to represent partial answers rather than complete answers and vary much more widely.

Most of research in QA systems [5], [2] typically invoke an information retrieval (IR) subsystem to retrieve and rank document fragments with respect to the question. However, the objective of QA systems is to find answers to factual questions, such as "What is the longest river in the United States?" and "Who is Andrew Carnegie?" Such factual question answering approaches are not effective enough to retrieve answers for opinion questions.

For a product opinionated question, the answer should not be just a best answer. The answer should reflect the opinions of various segments of users, and incorporate both positive and negative viewpoints. Hence the answer should be a summarization of public opinions and comments on the product or specific aspect asked in the question [7].

In addition, it should also include public opinions and comments on the product or feature of product. For example, the question "What do people think the camera of Nokia 6610?" asks for public positive and negative opinions on the feature "camera" of product "Nokia 6610."

In fact, rather than factual information, people would also like to know about others' opinions, toward some specific objects. Factual question answering thus cannot deal with yet comparative/superlative questions where the asker wants to compare the quality of products such as "Which is better Canon G3 and Nikon Coolpix 4300?" and "What is the best digital camera?" unless a reviewer explicitly compared these items in the review like that " Based on my experiences, Canon G3 works better than Nikon Coolpix 4300." for the comparative question and "Canon Powershoot SD500 is the best digital camera." for superlative question. It is not enough to answer such questions in factual question answering because the answer is inaccurate and only based on one person's idea. These are challenges in this paper.

Motivated by the above observations, we propose a framework for sentiment question answering. The system first adopts an opinion mining technique in the preprocessing phase to extract features and estimate their quality. Target features are attributes or components of the target product (entity) that have been commented on in the review, e.g. 'zoom' and 'battery life' for a digital camera that is target entity on the product sold online. Our proposed system including question analysis that identifies aspects and opinions asked in the questions, review sentence retrieval, and answer generation which aggregates the opinions of retrieved sentences.

In this paper, we investigate a linguistic approach for sentiment question analysis to support answering of opinion-based questions. The purpose is to get the accurate result of question analysis with the use of pattern based approach.

Then, in paper, we study how to analysis sentiment question to get good results and how to retrieve opinionated review sentences correspondence to the asked target items (entity and feature) for sentiment questions with the Wordnet based similarity approach.

Finally the system groups the opinions with high quality of the retrieved relevant review sentences with a ranking method instead of filtering method. It is helpful to enhance the quality of a summary to produce the optimal result.

Experimental evaluation conducts using web evaluative texts; including consumer reviews dataset. There are three contributions.

(1) Firstly, Analyzing the sentiment questions for produce the accurate answering review sentences.

(2) Retrieving the opinionated review sentences relevant to the asked target entity and features.

(3) Grouping these retrieved sentences according to sentiment polarity to form an appropriate answer to question.

This paper is organized as follows. Section 2 contains the related works. Section 3 gives the background theory about sentiment question answering. Section 4 describes the sentiment question answering framework. Section 5 presents discussion and results. Conclusion remarks are given in section 6.

## 2. Related Works

Most of the previous studies have mainly addressed the problem of opinion summarization and opinion retrieval, question answering.

Most of the research on QA systems has been developed for factual questions. Research focused on building factoid QA systems has a long tradition; however, it is only recently that researchers have started to focus on the development of Opinion Question Answering systems. In [9], author summarizes the opinion to support a Multi- Perspective QA system, aiming at identifying the opinion-oriented answers for a question.

Question analysis often has to distinguish the opinion question from the factual one, and find the key points asked in the questions, such as the product aspect and product name. In [2], the author proposes a method for separating facts form opinions and identifying polarity of opinion sentences. They first present a Bayesian classifier for discriminating between subjective (opinion-based) documents and objective (fact-based) documents and then apply three unsupervised techniques for detecting opinion sentences in documents: similarity approach, na¨ıve Bayes classifier, and multiple na¨ıve Bayes classifier. In the next step they determine the polarity of each sentence by computing the occurrence of its words with words from a known positive and negative seed set. For evaluation, they compute the precision and recall of classifiers. Gold standards for evaluating sentences and polarity are made manually.

In [3], opinion holders are identified, which are a key component in retrieving the correct answers to opinion questions. Similarly, in [8], the author utilized a SVM classifier for recognizing opinion and fact-based sentences. Their method first retrieves answer sentences based on the keyword matching and then the answers are re-ranked based on question type and answer sentences. They also discuss that traditional frameworks for evaluating QA systems fail to measure the effectiveness of opinion QA systems. They evaluate their method by computing the average precision and argue that this metric is a better evaluation framework for opinion QA methods. In [1], author proposed a two-layered classifier for question analysis, and retrieved the answer-fragments by keyword matching. In particular, they first identified the opinion questions, and classified them into six predefined question types, including holder, target, attitude, reason, majority, and yes/no. These question types and corresponding polarity on the questions were used to filter non-relevant sentences in the answer fragments. F1-measure was employed as the evaluation metric.

The authors propose a slightly different method for opinion QA in [5]. They consider each sentence as a node and used the similarity of each two sentences as weight of the corresponding edge between them. They first built a graph on the retrieved sentences, with each sentence as the node, and the similarity (i.e. Cosine similarity) between each sentences pair as the weight of the corresponding edge. After building this graph, given a question, its similarity to each sentence in the graph was computed. Such similarity was viewed as the relevant score to the corresponding sentence. They compute the sentence score by combining similarity values. They evaluate their method by computing the F-measure.

## 3. Natural Language Processing

Natural Language Processing (NLP), also known as computational linguistics, is a field of computer science that studies interactions of human languages with computers. The main goal of NLP is to enable effective human-machine communication, which could be either as spoken or written form. Here, only the written form will be addressed. For many applications, is desirable to automatically process texts written in natural language.

Computers can parse and automatically generate natural language texts, extract semantics from them and identify real world objects. Some examples include search engines understanding natural language text queries and data information extraction applications which could interpret a large amount of text and store just the significant parts in a database. Part-of-Speech tagging is an important application of a NLP technique used in text mining.

### 3.1. Part-of-Speech Tagging

One special application of natural language processing is determining the part of speech of each word in a sentence, known as part-of-speech (POS) tagging. The part-of-speech is a category used in linguistics that is defined by a syntactic or morphological behavior of a word. Due to the availability of large corpora which have been manually annotated with POS information, many taggers use annotated text to learn the rules and use them to automatically assign POS tags to unseen text. Part-of-speech tags are assigned to a single word according to its role in the sentence. Traditional grammar classifies words based on eight parts of speech: the verb (VB), the noun (NN), the pronoun (PR+DT), the adjective (JJ), the adverb (RB), the preposition (IN), the conjunction (CC), and the interjection (UH).

The reason why POS tagging is so important to information extraction is the fact that each category plays a specific role within a sentence. Nouns give names to objects, beings or entities from the world. An adjective qualifies or describes nouns. Also some adverbs can play pretty much the same role as an adjective, however under different circumstances, as they rarely modify a noun.

### 3.2. Opinion Mining

Opinion mining is a type of natural language processing for tracking the mood of the public about a particular product. Opinion mining, which is also called sentiment analysis, involves building a system to collect and examine opinions about the product made in blog posts, comments, reviews or tweets. Automated opinion mining often uses machine learning, a component of artificial intelligence (AI).

Sentiment analysis, also called opinion mining, is the field of study that analyzes people's opinions, sentiments, evaluations, appraisals, attitudes, and emotions. An opinion mining is capable of extracting knowledge from examples in a database and incorporating new data to improve performance over time. The process can be as simple as learning a list of positive and negative words, or as complicated as conducting deep parsing of the data in order to understand the grammar and sentence structure used.

## 4. Sentiment Question Answering

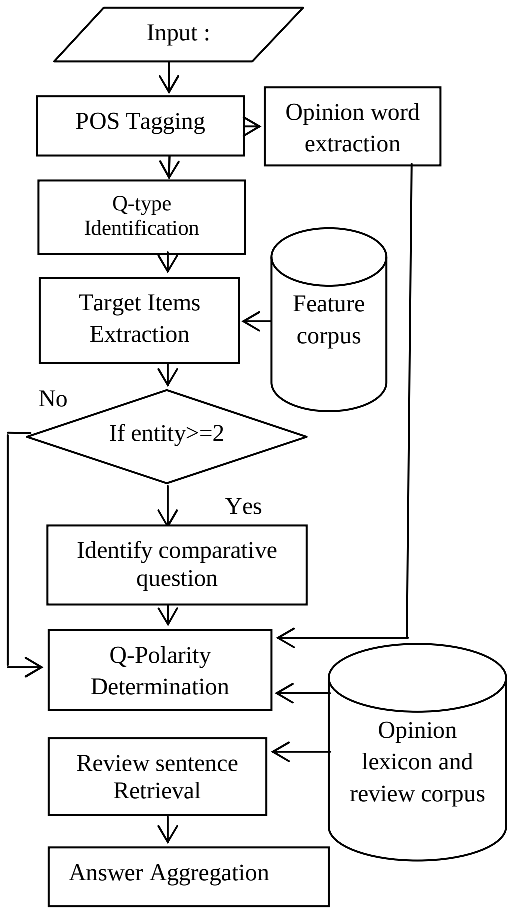


**Figure 1. A framework for sentiment question answering system**

Question Answering aims to provide answers to human-generated questions automatically. Sentiment question answering system is similar to opinion retrieval task only that instead of returning a set of opinions, answers have to be a summary of those opinions. A sentiment question answering framework can be seen as shown in Figure 1.

Sentiment question answering tasks are:

(1) Question Analysis with the POS tagging,

(2) Review Answer retrieval from the customer reviews and,

(3) Answer generating to get high quality retrieved sentences.

### 4.1. Question Analysis

This is the first step in Sentiment QA system. By analyzing a question, given a collection of question set $=\{q_1,q_2,q_3,\ldots q_n\}$ about products, opinions about product entities wanted by the user can be known to extract. It is different from the tradition textual question analysis in the area of web search. Various techniques have been developed for Sentiment QA system already.

In order to be able to extract the correct answer to opinion questions, different elements on the questions must be considered. There are four elements in question analysis as follows: (i) Question types classification (ii) Target items extraction (iii) Question forms classification (iv) Question polarity determination.

**Question type classification:** The system first uses "POS tagging" to identify part of speech of each word. For example, "what_WP is_ VBZ the_ DT best_ JJS MP3_NNP Player _NNP ? ." where 'WP', 'JJS' and 'NNP' indicates wh-pronoun, adjective and nouns, respectively. Since different types of questions have different POS patterns and question words (e.g. why, what, who, when etc), in system, opinionated four types of questions (target, attitude, reason, yes/no) are determined based on tagged patterns and question words.

The question word patterns with POS description are MD:Modal, VB:Verb, base form, VBD:verb, past tense, VBG: Verb, gerund or present participle, VBN: Verb, past participle, VBP: Verb, non-3rd person singular present, VBZ: Verb, 3rd person singular present, WDT: Wh-determiner, WP: Wh-pronoun, WRB: Wh-adverb.

**Target items extraction:** Target entities can be a product (e.g. Canon, Nikon) or product category (Digital camera). It has a set of attributes, e.g.; picture quality, size, weight and a set of parts, e.g.; lens, viewfinders, battery. It is also called features or aspects.

From tagged the input opinionated question, the system extracts nouns, verb, and adjective, etc.; as predefined patterns for target entities and feature extraction and identified it with annotated training corpus for domain product. Based on the predefined pattern, system extracts the target entities and feature for comparative and single questions. Example predefined patterns of target items (product names and their features) are shown in Table 1 and Table 2.

**Table 1. Predefined patterns of target entities**

| Product Names | Patterns |
|---|---|
| Apex AD2600 | NNP CD |
| Nokia 6610 | |
| Nokia 6600 | |
| Canon S100 | |
| iPhone 3GS | NNP NNP |
| Nokia N95 | |
| Canon G3 | |
| Canon EOS 450D | NNP NNP CD |
| BlackBerry Bold 9700 | |
| Nikon coolpix 4300 | NNP NN CD |
| Canon PowerShot SD500 | NNP NNP NNP CD |
| Creative Labs noMad 40GB | NNP NNP NNP NNP CD |

**Table 2. Predefined feature patterns**

| Patterns | Features |
|---|---|
| NN | battery, Memory, screen, size, zoom, ringtone, eight, lcd, vibration, earpiece, keypad, headphone, etc. |
| VB | use, focus, look, made, learn, read, delete, rewind, recognize, etc. |
| VB RP | Set up, look up |
| DT NN | no disc |
| NN NN | music quality, jail breaking, lens adapter, battery life, movie mode, sunset feature, memory card, mp3 player, volume range, etc. |
| JJ NN | audio device, optical mode, indoor image, macro mode, rechargeable battery, mobile service, etc. |
| JJ VB NN | progressive scan player |
| NN VB NN | autofocus assist light |
| JJ NN NN | continuous shot mode, dual layer dvd, etc |

**Question form identification:** If a question has two entities or comparative adjectives or adverbs (i.e. better, cheaper, etc.) or, it is a comparative. Otherwise it is a single question.

The POS tagger tags comparative adjectives and adverbs with tag "JJR" and tag "RBR", for example,"Why_WRB Canon_NNP G3_NNP is_VBZ better_JJR than_IN Nikon_NNP coolpix_NN 4300_CD ?_." Therefore, if the system extracts more than one entity or comparative adjectives or adverbs (i.e. better, cheaper, etc.) from tagged question, it is a comparative. Otherwise it is a single question**.**

**Question polarity determination:** Question polarity indicates the direction of the question (positive, negative, or both) that the asker wants to find out. For example, when a user asks "Why do people recommend NikonX?" some positive features of that target item is produced. On the other hand, questions like "What is wrong with NikonX?" are asked to find out negative of the target item. However, sometimes users just want to get general information (both positive and negative aspects) about an item, e.g. "How do people think about NikonX?.

VB: verb, RB: Adverb, RBR: Adverb, comparative, RBS: Adverb superlative, JJ: Adjective or numeral, ordinal, JJR: Adjective, comparative, JJS: Adjective, superlative are the POS tags for opinion words. For example, a tagged question is "Is_ VBZ the_ DT battery_ NN of_ IN Apex_ NNP AD2600_ CD great_ JJ ?_."

Two seed sets of positive and negative words (opinion lexicon) are used to find out the polarity of the question. In this case, the word great_ JJ is extracted as possible opinion and match it with opinion lexicon for positive or negative. For the above example, the opinion word "great" is bearing positive opinion. Thus, the question can be determined as positive.

## 4.2. Answer Generating

Answer generation aims to generate an appropriate answer for a given opinion question based on the retrieved review sentences. An answer is essentially a sequence of sentences.

The system retrieves all of the reviews for a given question about the target items (entities, features) from the customer review dataset. Feature and sentiment scores with strongly or weakly opinion strength in positive/negative answer from the labeled dataset are extracted in preprocessing step. The feature is designed according to their polarity in strong/weak positive/negative opinionated answers, for instance, zoom [+3] and picture quality [-1].

For each discovered feature, related opinion sentences are put into positive and negative categories according to the opinion sentences' orientations. In case, some feature of products may be same meaning but different spelling, for instance, "photo". "Photo" may be feature of a product. Another words of "photo" are "picture" and" image". Moreover singular and plural nouns are also considered with Wordnet, for instance, "picture" and "pictures". To overcome this, Wordnet based similarity is approached by using Wordnet lexicon.

In answer aggregation, a count is computed to show how many review sentences give the positive/negative opinions to the feature. Then, all features in particular product (entity) with the retrieved review sentences are ranked according to the opinion strength of the feature, (in that case, [+3] is positively strong and [-1] is negatively strong, etc.;) and the frequency of their appearances in the reviews sentences.

In fact, Different products for comparative/superlative questions such as "Which is better Canon G3 and Nikon Coolpix 4300?" and "What is the best digital camera?" can be compared and stated by comparing their common features with the high quality rating (in case, opinion strength). Finally, the system uses question polarity for grouping answer sentences to provide a complete set of answers for the user.

In this case, for the question "Why do people recommend CanonX?" the ranked list of answers shows positive opinionated sentences describing the quality of the features of that product entity first, and then produce the sentences with a negative point at the end . That's why the user can see and decide about the quality of the target items for the important decision making process.

## 5. Discussion and Results

We intend to present not only effective question analysis on the sentiment question but also how to retrieve and aggregate the review sentences to get high quality of them with a ranking approach by using polarity measuring.

Experiment are conducted using the customer reviews of six products such as three digital camera (Canon G3, Nikon coolpix 4300, Canon Powershoot SD500), Two cellular phone(Nokia 6600, Nokia 6610), and a DVD player. The reviews public is available from www.amazon.com. 220 questions are created for these products by referring to real questions in Yahoo! Answer service. Table 3 shows the evaluation results of answer generation for the hot features of each product asked in the system. Statistics of positive and negative of the retrieved opinionated answers sentences are also illustrated as shown in table 4.

Q type is used to define the template for the answers. For reason and attitude questions, answers are generated by summarizing corresponding answer fragments. For the target questions, related opinionated sentences that compare these products are output based on the majority voting of the opinions in the retrieved answer. (Most of target questions are comparative questions). Finally, answers for (yes, no) questions are produced based on the asked opinions and the major opinions in the answer, and then summarize them.

**Table 3. Evaluation results for answer generation**

| Product name | Precision | Recall | F-Measure |
|---|---|---|---|
| Nokia 6610 | 0.501 | 0.972 | 0.661 |
| Nokia 6600 | 0.760 | 0.942 | 0.841 |
| Canon G3 | 0.776 | 0.755 | 0.765 |
| Nikon coolpix 4300 | 0.714 | 0.853 | 0.777 |
| Canon PowerShot SD500 | 0.806 | 0.956 | 0.875 |
| Apex AD2600 | 0.788 | 0.954 | 0.863 |
| Average | 0.724 | 0.905 | 0.797 |

**Table 4. Statistic of the positive and negative sentences for each product**

| Product name | Opinionated | | Non-opinionated |
|---|---|---|---|
| | POS | Neg | |
| Nokia 6610 | 46% | 15% | 39% |
| Nokia 6600 | 55% | 26% | 19% |
| Canon G3 | 37% | 10% | 53% |
| Nikon coolpix 4300 | 49% | 10% | 41% |
| Canon PowerShot SD500 | 51% | 12% | 37% |
| Apex AD2600 | 27% | 32% | 41% |

## 6. Conclusion

In this paper, a new framework develops for generating the structured answer for the sentiment questions. In addition, a new approach, linguistic pattern with POS tagging presents to give accurate results for the question analysis. An idea the wordnet based similarity is also applied for generating the answers and then ranks the retrieved sentences correspondence to the target entities and features of asked questions. Finally, Answers aggregate with high quality sentences to solve current question answering's challenges. The experimental results demonstrate more effective evaluation methods to get the optimal results.

## References


[1] L. wei Ku, Y. ting Liang, and H. hsi Chen. Question analysis and answer passage retrieval for opinion question answering systems. *In Proceedings of 19th Conference on CLSP*, 2007.

[2] H. Yu and V. Hatzivassiloglou. "Towards answering opinion questions: separating facts from opinions and identifying the polarity of opinion sentences". *In Proceeding of the EMNLP '03.*

[3] S.-M. Kim and E. Hovy. Identifying opinion holders for question answering in opinion texts. *In Proceedings of the AAAI Workshop on Question Answering in Restricted Domains*, 2005.

[4] Kim, S. and Hovy, E. Determining the Sentiment of Opinions. *Proceedings of the 20th International Conference on (COLING '04)*, 2004.

[5] F. Li, Y. Tang, M. Huang, and X. Zhu. Answering opinion questions with random walks on graphs. *Proceeding of the 47th Annual Meeting of ACL* , 2009.

[6] Jianxing Yu, Zheng-Jun Zha, Tat-Seng Chua, Answering Opinion Questions on Products by Exploiting Hierarchical Organization of Consumer Reviews, *In Proceeding of EMNLP*, 2012.

[7] P.Jiang,H. Fu, C. Zhang, and Z. Niu. A Framework for Opinion Question Answering.*IMS*, 2010.

[8] Somasundaran, T. Wilson, J. Wiebe and V. Stoyanov. QA with Attitude: Exploiting Opinion Type Analysis for Improving Question Answering in Online Discussions and the News. *ICWSM*, 2007.

[9] Cardie, J. Wiebe, T. Wilson, and D. Litman. Combining Low-level and Summary Representations of Opinions for Multi-Perspective Question Answering. *AAAI*, 2003.