

Mining Opinions in Comparative Sentences

Murthy Ganapathibhotla

Department of Computer Science
University of Illinois at Chicago
851 South Morgan Street
Chicago, IL 60607-7053
sganapat@cs.uic.edu

Bing Liu

Department of Computer Science
University of Illinois at Chicago
851 South Morgan Street
Chicago, IL 60607-7053
liub@cs.uic.edu

Abstract

This paper studies sentiment analysis from the user-generated content on the Web. In particular, it focuses on mining opinions from comparative sentences, i.e., to determine which entities in a comparison are preferred by its author. A typical comparative sentence compares two or more entities. For example, the sentence, “*the picture quality of Camera X is better than that of Camera Y*”, compares two entities “*Camera X*” and “*Camera Y*” with regard to their picture quality. Clearly, “*Camera X*” is the preferred entity. Existing research has studied the problem of extracting some key elements in a comparative sentence. However, there is still no study of mining opinions from comparative sentences, i.e., identifying preferred entities of the author. This paper studies this problem, and proposes a technique to solve the problem. Our experiments using comparative sentences from product reviews and forum posts show that the approach is effective.

1 Introduction

In the past few years, there was a growing interest in mining opinions in the user-generated content (UGC) on the Web, e.g., customer reviews, forum posts, and blogs. One major focus is sentiment classification and opinion mining (e.g., Pang et al 2002; Turney 2002; Hu and Liu 2004; Wilson et al 2004; Kim and Hovy 2004; Popescu and Etzioni 2005)

However, these studies mainly center on *direct opinions* or *sentiments* expressed on entities. Little study has been done on comparisons, which represent another type of opinion-bearing text. Comparisons are related to but are also quite different from direct opinions. For example, a typical direct opinion sentence is “*the picture quality of Camera X is great*”, while a typical comparative sentence is “*the picture quality of Camera X is better than that of Camera Y.*” We can see that comparisons use different language constructs from direct opinions. A comparison typically expresses a comparative opinion on two or more entities with regard to their shared features or attributes, e.g., “*picture quality*”. Although direct opinions are most common in UGC, comparisons are also widely used (about 10% of the sentences), especially in forum discussions where users often ask questions such as “X vs. Y” (X and Y are competing products). Discussions are then centered on comparisons.

Jindal and Liu (2006) proposed a technique to identify comparative sentences from reviews and forum posts, and to extract entities, comparative words, and entity features that are being compared. For example, in the sentence, “*Camera X has longer battery life than Camera Y*”, the technique extracts “*Camera X*” and “*Camera Y*” as entities, and “*longer*” as the comparative word and “*battery life*” as the attribute of the cameras being compared. However, the technique does not find which entity is preferred by the author. For this example, clearly “*Camera Y*” is the preferred camera with respect to the “*battery life*” of the cameras. This paper aims to solve this problem, which is useful in many applications because the preferred entity is the key piece of information in a comparative opinion. For example, a potential customer clearly wants to buy the product that is better or preferred.

In this work, we treat a sentence as the basic

information unit. Our objective is thus to identify the preferred entity in each comparative sentence. A useful observation about comparative sentences is that in each such sentence there is usually a comparative word (e.g., “better”, “worse” and *-er* word) or a superlative word (e.g., “best”, “worst” and *-est* word). The entities being compared often appear on the two sides of the comparative word. A superlative sentence may only have one entity, e.g., “Camera X is the best”. For simplicity, we use comparative words (sentences) to mean both *comparative words (sentences)* and *superlative words (sentences)*.

Clearly, the preferred entity in a comparative sentence is mainly determined by the comparative word in the sentence. Some comparative words explicitly indicate user preferences, e.g., “better”, “worse”, and “best”. We call such words *opinionated comparative words*. For example, in the sentence, “the picture quality of Camera X is better than that of Camera Y”, Camera X is preferred due to the opinionated comparative word “better”.

However, many comparative words are not opinionated, or their opinion orientations (i.e., positive or negative) depend on the context and/or the application domain. For instance, the word “longer” is not opinionated as it is normally used to express that the length of some feature of an entity is greater than the length of the same feature of another entity. However, in a particular context, it can express a desired (or *positive*) or undesired (or *negative*) state. For example, in the sentence, “the battery life of Camera X is longer than Camera Y”, “longer” clearly expresses a desired state for “battery life” (although this is an objective sentence with no explicit opinion). “Camera X” is thus preferred with regard to “battery life” of the cameras. The opinion in this sentence is called an *implicit opinion*. We also say that “longer” is positive in this context. We know this because of our existing domain knowledge. However, “longer” may also be used to express an undesirable state in a different context, e.g., “Program X’s execution time is longer than Program Y”. “longer” is clearly negative here. “Program Y” is thus preferred. We call comparative words such as “longer” and “smaller” *context-dependent opinion comparatives*.

Sentences with opinionated words (e.g., “better”, and “worse”) are usually easy to handle. Then the key to solve our problem is to identify the opinion orientations (positive or negative) of context-dependent comparative words. To this end, two questions need to be answered: (1) what

is a context and (2) how to use the context to help determine the opinion orientation of a comparative word?

The simple answer to question (1) is the whole sentence. However, a whole sentence as context is too complex because it may contain too much irrelevant information, which can confuse the system. Intuitively, we want to use the smallest context that can determine the orientation of the comparative word. Obviously, the comparative word itself must be involved. We thus conjecture that the context should consist of the *entity feature* being compared and the *comparative word*. Our experimental results show that this context definition works quite well.

To answer the second question, we need external information or knowledge because there is no way that a computer program can solve the problem by analyzing the sentence itself. In this paper, we propose to use the external information in customer reviews on the Web to help solve the problem. There are a large number of such reviews on almost any product or service. These reviews can be readily downloaded from many sites. In our work, we use reviews from epinions.com. Each review in epinions.com has separate Pros and Cons (which is also the case in most other review sites). Thus, positive and negative opinions are known as they are separated by reviewers. However, they cannot be used directly because Pros and Cons seldom contain comparative words. We need to deal with this problem. Essentially, the proposed method computes whether the comparative word and the feature are more associated in Pros or in Cons. If they are more associated in Pros (or Cons) than Cons (or Pros), then the comparative word is likely to be positive (or negative) for the feature. A new association measure is also proposed to suit our purpose. Our experiment results show that it can achieve high precision and recall.

2 Related Work

Sentiment analysis has been studied by many researchers recently. Two main directions are sentiment classification at the document and sentence levels, and feature-based opinion mining. Sentiment classification at the document level investigates ways to classify each evaluative document (e.g., product review) as positive or negative (Pang et al 2002; Turney 2002). Sentiment classification at the sentence-level has also been studied (e.g., Riloff and Wiebe 2003; Kim and Hovy 2004; Wilson et al 2004; Gamon et al

2005; Stoyanov and Cardie 2006). These works are different from ours as we study comparatives.

The works in (Hu and Liu 2004; Liu et al 2005; Popescu and Etzioni 2005; Mei et al 2007) perform opinion mining at the feature level. The task involves (1) extracting entity features (e.g., “*picture quality*” and “*battery life*” in a camera review) and (2) finding orientations (positive, negative or neutral) of opinions expressed on the features by reviewers. Again, our work is different because we deal with comparisons.

Discovering orientations of context dependent opinion comparative words is related to identifying domain opinion words (Hatzivassiloglou and McKeown 1997; Kanayama and Nasukawa 2006). Both works use *conjunction rules* to find such words from large domain corpora. One conjunction rule states that when two opinion words are linked by “and”, their opinions are the same. Our method is different in three aspects. First, we argue that finding domain opinion words is problematic because in the same domain the same word may indicate different opinions depending on what features it is applied to. For example, in the camera domain, “*long*” is positive in “*the battery life is very long*” but negative in “*it takes a long time to focus*”. Thus, we should consider both the feature and the opinion word rather than only the opinion word. Second, we focus on studying opinionated comparative words. Third, our technique is quite different as we utilize readily available external opinion sources.

As discussed in the introduction, a closely related work to ours is (Jindal and Liu 2006). However, it does not find which entities are preferred by authors. Bos and Nissim (2006) proposes a method to extract some useful items from superlative sentences. Fiszman et al (2007) studied the problem of identifying which entity has more of certain features in comparative sentences. It does not find which entity is preferred.

3 Problem Statement

Definition (entity and feature): An *entity* is the name of a person, a product, a company, a location, etc, under comparison in a comparative sentence. A *feature* is a part or attribute of the entity that is being compared.

For example, in the sentence, “*Camera X’s battery life is longer than that of Camera Y*”, “*Camera X*” and “*Camera Y*” are entities and “*battery life*” is the camera feature.

Types of Comparatives

- 1) *Non-equal gradable*: Relations of the type *greater* or *less than* that express a total ordering of some entities with regard to their shared features. For example, the sentence, “*Camera X’s battery life is longer than that of Camera Y*”, orders “*Camera X*” and “*Camera Y*” based on their shared feature “*battery life*”.
- 2) *Equative*: Relations of the type *equal to* that state two objects as equal with respect to some features, e.g., “*Camera X and Camera Y are about the same size*”.
- 3) *Superlative*: Relations of the type *greater* or *less than all others* that rank one object over all others, “*Camera X’s battery life is the longest*”.
- 4) *Non-gradable*: Sentences which compare features of two or more entities, but do not explicitly grade them, e.g., “*Camera X and Camera Y have different features*”

The first three types are called *gradable* comparatives. This paper focuses on the first and the third types as they express ordering relationships of entities. Equative and non-gradable sentences usually do not express preferences.

Definition (comparative relation): A *comparative relation* is the following:

<ComparativeWord, Features, EntityS1, EntityS2, Type>

ComparativeWord is the keyword used to express a comparative relation in the sentence. *Features* is a set of features being compared. *EntityS1* and *EntityS2* are sets of entities being compared. Entities in *EntityS1* appear on the left of the comparative word and entities in *EntityS2* appear on the right. *Type* is *non-equal gradable*, *equative* or *superlative*. Let us see an example. For the sentence “*Camera X has longer battery life than Camera Y*,” the extracted relation is:

<longer, {battery life}, {Camera X}, {Camera Y}, non-equal gradable>.

We assume that the work in (Jindal and Liu 2006) has extracted the above relation from a comparative sentence. In this work, we aim to identify the preferred entity of the author, which is not studied in (Jindal and Liu 2006).

Our objective: Given the extracted comparative relation from a comparative sentence, we want to identify whether the entities in *EntityS1* or in *EntityS2* are preferred by the author.

4 Proposed Technique

We now present the proposed technique. As discussed above, the primary determining factors of the preferred entity in a comparative sentence are

the feature being compared and the comparative word, which we conjecture, form the context for opinions (or preferred entities). We develop our ideas from here.

4.1 Comparatives and superlatives

In English, comparatives and superlatives are special forms of adjectives and adverbs. In general, comparatives are formed by adding the suffix “-er” and superlatives are formed by adding the suffix “-est” to the *base adjectives* and *adverbs*. We call this type of comparatives and superlatives *Type 1 comparatives* and *superlatives*. For simplicity, we will use Type 1 comparatives to represent both from now on.

Adjectives and adverbs with two syllables or more and not ending in y do not form comparatives or superlatives by adding “-er” or “-est”. Instead, “more”, “most”, “less” and “least” are used before such words, e.g., “more beautiful”. We call this type of comparatives and superlatives *Type 2 comparatives* and *Type 2 superlatives*. These two types are called *regular comparatives* and *superlatives* respectively.

In English, there are also some *irregular comparatives* and *superlatives*, which do not follow the above rules, i.e., “more”, “most”, “less”, “least”, “better”, “best”, “worse”, “worst”, “further/farther” and “furthest/farthest”. They behave similarly to Type 1 comparatives and superlatives and thus are grouped under Type 1.

Apart from these comparatives and superlatives, there are *non-standard words* that express gradable comparisons, e.g., “prefer”, and “superior”. For example, the sentence, “in term of battery life, Camera X is superior to Camera Y”, says that “Camera X” is preferred. We obtained a list of 27 such words from (Jindal and Liu 2006) (which used more words, but most of them are not used to express gradable comparisons). Since these words behave similarly to Type 1 comparatives, they are thus grouped under Type 1.

Further analysis also shows that we can group comparatives into two categories according to whether they express increased or decreased values:

Increasing comparatives: Such a comparative expresses an increased value of a quantity, e.g., “more”, and “longer”.

Decreasing comparatives: Such a comparative expresses a decreased value of a quantity, e.g., “less”, and “fewer”.

As we will see later, this categorization is very useful in identifying the preferred entity.

Since comparatives originate from adjectives and adverbs, they may carry positive or negative sentiments/opinions. Along this dimension, we can divide them into two categories.

1. *Opinionated comparatives*: For Type 1 comparatives, this category contains words such as “better”, “worse”, etc, which has explicit opinions. In sentences involving such words, it is normally easy to determine which entity is the preferred one of the sentence author.

In the case of Type 2 comparatives, formed by adding “more”, “less”, “most”, and “least” before adjectives or adverbs, the opinion (or preferred entity) is determined by both words. The following rules apply:

“increasing comparative” Negative → Negative Opinion
“increasing comparative” Positive → Positive Opinion
“decreasing comparative” Negative → Positive Opinion
“decreasing comparative” Positive → Negative Opinion

The first rule says that the combination of an increasing comparative word (e.g., “more”) and a negative opinion adjective/adverb (e.g., “awful”) implies a negative Type 2 comparative. The other rules are similar. These rules are intuitive and will not be discussed further.

2. *Comparatives with context-dependent opinions*: These comparatives are used to compare gradable quantities of entities. In the case of Type 1 comparatives, such words include “higher”, “lower”, etc. Although they do not explicitly describe the opinion of the author, they often carry *implicit sentiments* or *preferences* based on contexts. For example, in “Car X has higher mileage per gallon than Car Y”, it is hard to know whether “higher” is positive or negative without domain knowledge. It is only when the two words, “higher” and “mileage”, are combined we know that “higher” is desirable for “mileage” from our domain knowledge.

In the case of Type 2 comparatives, the situation is similar. However, the comparative word (“more”, “most”, “less” or “least”), the adjective/adverb and the feature are all important in determining the opinion or the preference. If we know whether the comparative word is increasing or decreasing (which is easy since there are only four such words), then the opinion can be determined by applying the four rules above in (1).

For this work, we used the *opinion word* list from (Hu and Liu 2004), which was compiled using a bootstrapping approach based on WordNet. For opinionated comparatives, due to the observation below we simply convert the opinion

adjectives/adverbs to their comparative forms, which is done automatically based on grammar (comparative formation) rules described above and WordNet.

Observation: If a word is positive (or negative), then its comparative or superlative form is also positive (or negative), e.g., “good”, “better” and “best”.

After the conversion, these words are manually categorized into increasing and decreasing comparatives. Although this consumes some time, it is only a one-time effort.

4.2 Contexts

To deal with comparatives with context dependent opinions, we need contexts. It is conjectured that the comparative and the feature in the sentence form the context. This works very well. For a Type 2 comparative, we only need the feature and the adjective/adverb to form a context. For example, in the sentence, “*Program X runs more quickly than Program Y*”, the context is the pair, (“run”, “quickly”), where “run” is a verb feature. If we find out that (“run”, “quickly”) is positive based on some external information, we can conclude that “*Program X*” is preferred using one of the four rules above since “more” is an increasing comparative.

We will use such contexts to find opinion orientations of comparatives with regard to some features from the external information, i.e., Pros and Cons in online reviews.

4.3 Pros and Cons in Reviews

Figure 1 shows a popular review format. The reviewer first describes Pros and Cons briefly, and then writes a full review.

Pros and Cons are used in our work for two main reasons. First, the brief information in Pros and Cons contains the essential information related to opinions. Each phrase or sentence segment usually contains an entity feature and an opinion word. Second, depending on whether it is in Pros or in Cons, the user opinion on the product feature is clear.

To use the Pros and Cons phrases, we separate them use punctuations and words, i.e., ‘,’; ‘,’; ‘and’, and ‘but’. Pros in Figure 1 can be separated into 5 phrases or segments,

great photos	<photo>
easy to use	<use>
good manual	<manual>
many options	<option>
takes videos	<video>

We can see that each segment describes an entity

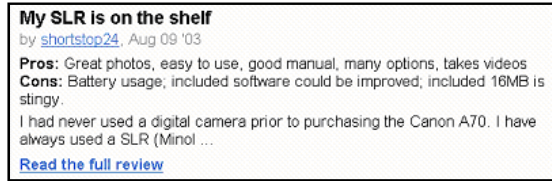


Figure 1: An example review

feature on which the reviewer has expressed an opinion. The entity feature for each segment is listed within $\langle \rangle$.

4.4 Identifying Preferred Entities: The Algorithm

Since we use Pros and Cons as the external information source to help determine whether the combination of a comparative and an entity feature is positive or negative, we need to find comparative and entity features words in Pros and Cons. However, in Pros and Cons, comparatives are seldom used (entity features are always there). Thus we need to first convert comparatives to their base forms. This can be done automatically using WordNet and grammar rules described in Section 4.1. We will not discuss the process here as it is fairly straightforward.

We now put everything together to identify the preferred entity in a comparative sentence. For easy reference, we denote the comparative word as C and the feature being compared as F . After obtaining the base forms of C , we work on two main cases for the two types of comparatives:

Case 1. Type 1 Comparative or Superlative:

There are four sub-cases.

1.A. C is opinionated: If the comparative or superlative C has a positive orientation (e.g., “better”), EntityS1 (which appears before C in the sentence) is temporarily assigned as the preferred entity. Otherwise, EntityS2 is assigned as the preferred entity. The reason for the temporary assignment is that the sentence may contain negations, e.g., “not”, which is discussed below.

1.B. C is not opinionated but F is opinionated: An example is, “*Car X generates more noise than Car Y*”, which has the feature F “noise”, a negative noun. If the orientation of F is positive and C is an increasing comparative word, we assign EntityS1 as the preferred entity. Otherwise, we assign EntityS2 as the preferred entity. The possibilities are listed as four rules below, which are derived from the 4 rules earlier:

- “increasing C ” + Positive \rightarrow EntityS1 preferred
- “decreasing C ” + Positive \rightarrow EntityS2 preferred

“increasing C ” + Negative \rightarrow EntityS2 preferred
“decreasing C ” + Negative \rightarrow EntityS1 preferred

“Positive” and “Negative” stand for the orientation of feature F being positive and negative respectively.

1.C. Both C and F are not opinionated: In this case, we need external information to identify the preferred entity. We use phrases in Pros and Cons from reviews.

In this case, we look for the feature F and comparative word C , (i.e., the context) in the list of phrases in Pros and Cons. In order to find whether the combination of C and F indicates a positive or negative opinion, we compute their associations in Pros and in Cons. If they are more associated in Pros than in Cons, we conclude that the combination indicates a positive sentiment, and otherwise a negative sentiment. The result decides the preferred entity. Point-wise mutual information (PMI) is commonly used for computing the association of two terms (e.g., Turney 2002), which is defined as:

$$PMI(F, C) = \log \frac{Pr(F, C)}{Pr(F)Pr(C)}$$

However, we argue that PMI is not a suitable measure for our purpose. The reason is that PMI is symmetric in the sense that $PMI(F, C)$ is the same as $PMI(C, F)$. However, in our case, the feature F and comparative word C association is not symmetric because although a feature is usually modified by a particular adjective word, the adjective word can modify many other features. For example, “long” can be used in “long lag”, but it can also be used in “long battery life”, “long execution time” and many others. Thus, this association is asymmetric. We are more interested in the conditional probability of C (including its synonyms) given F , which is essentially the confidence measure in traditional data mining. However, confidence does not handle well the situation where C occurs frequently but F appears rarely. In such cases a high conditional probability $Pr(C|F)$ may just represent some pure chance, and consequently the resulting association may be spurious. We propose the following measure, which we call *one-side association (OSA)*, and it works quite well:

$$OSA(F, C) = \log \frac{Pr(F, C)Pr(C|F)}{Pr(F)Pr(C)}$$

The difference between OSA and PMI is the conditional probability $Pr(C|F)$ used in OSA,

which biases the mutual association of F and C to one side.

Given the comparative word C and the feature F , we first compute an OSA value for positive, denoted by $OSA_P(F, C)$, and then compute an OSA value for negative, denoted by $OSA_N(F, C)$. The decision rule is simply the following:

If $OSA_P(F, C) - OSA_N(F, C) \geq 0$ then
EntityS1 is preferred
Otherwise, EntityS2 is preferred

Computing $OSA_P(F, C)$: We need to compute $Pr_P(F, C)$, for which we need to count the number of times that comparative word C and the feature F co-occur. Instead of using C alone, we also use its base forms and synonyms and antonyms. Similarly, for F , we also use its synonyms. If C (or a synonym of C) and F (or a synonym) co-occur in a Pros phrase, we count 1. If an antonym of C and F (or a synonym) co-occur in a Cons phrase, we also count 1. Thus, although we only evaluate for positive, we actually use both Pros and Cons. This is important because it allows us to find more occurrences to produce more reliable results. Synonyms and antonyms are obtained from WordNet. Currently, synonyms and antonyms are only found for single word features.

We then count the number of occurrences of the comparative word C and the feature F separately in both Pros and Cons to compute $Pr_P(F)$ and $Pr_P(C)$. In counting the number of occurrences of C , we consider both its synonyms in Pros and antonyms in Cons. In counting the number of occurrences of F , we consider its synonyms in both Pros and Cons.

Computing $OSA_N(F, C)$: To compute $Pr_N(F, C)$, we use a similar strategy as for computing $Pr_P(F, C)$. In this case, we start with Cons.

1.D. C is a feature indicator: An example sentence is “Camera X is smaller than Camera Y”, where “smaller” is the *feature indicator* for feature “size”. In this case, we simply count the number of times (denoted by n_+) that C appears in Pros and the number of times (denoted by n_-) that C appears in Cons. If $n_+ \geq n_-$, we temporarily assign EntityS1 as the preferred entity. Otherwise, we assign EntityS2 as the preferred entity. Note that in some sentences, the entity features do not appear explicitly in the sentences but are implied. The words that imply the features are called *feature indicators*.

Case 2: Type 2 Comparative or Superlative:

There are two sub-cases:

2.A. *Adjective/adverb in the comparison is opinionated*: In this case, the feature F is not important. An example sentence is:

“*Car X has more beautiful interior than Car Y*”, “*more*” is an increasing comparative, and “*beautiful*” is the adjective with a positive orientation (the feature F is “*interior*”). “*Car X*” is clearly preferred in this case.

Another example is: “*Car X is more beautiful than Car Y*”. In this case, “*beautiful*” is a feature indicator for the feature “*appearance*”. Again, “*Car X*” is preferred. This sub-case can be handled similarly as case 1.B.

2.B. *adjective/adverb in the comparison is not opinionated*: If the adjective/adverb in comparison is a feature indicator, we can use the method in 1.D. Otherwise, we form a context using the feature and adjective/adverb, and apply the method in 1.C. We then combine the result with the comparative word before the adjective/adverb to decide based on the rules in 1.B.

Negations: The steps above temporarily determine which entity is the preferred entity. However, a comparative sentence may contain a negation word or phrase (we have compiled 26 of them), e.g., “*Camera X’s battery life is not longer than that of Camera Y.*” Without considering “*not*”, “*Camera X*” is preferred. After considering “*not*”, we assign the preferred entity to “*Camera Y*”. This decision may be problematic because “*not longer*” does not mean “*shorter*” (thus it can also be seen to have no preference).

5 Evaluation

A system, called PCS (*Preferred entities in Comparative Sentences*), has been implemented based the proposed method. Since there is no existing system that can perform the task, we could not compare with an existing approach. Below, we first describe the evaluation datasets and then present the results.

5.1 Evaluation Datasets

Our comparative sentence dataset consists of two subsets. The first subset is from (Jindal and Liu 2006), which are product review and forum discussion sentences on *digital cameras*, *DVD players*, *MP3 players*, *Intel vs AMD*, *Coke vs Pepsi*, and *Microsoft vs Google*. The original dataset used in (Jindal and Liu 2006) also contains many non-gradable comparative sentences, which are

not used here as most such sentences do not express any preferences.

To make the data more diverse, we collected more forum discussion data about mobile phones from <http://www.howardforums.com/>, and reviews from amazon.com and cnet.com on products such as *laptops*, *cameras* and *mobile phones*. Table 1 gives the number of sentences from these two sources. Although we only have 837 comparative sentences, they were collected from thousands of sentences in reviews and forums. About 10% of the sentences from them are comparative sentences.

Skewed Distribution: An interesting observation about comparative sentences is that a large proportion (based on our data) of them (84%) has EntityS1 as the preferred entity. This means that when people make comparisons, they tend to put the preferred entities first.

Pros and Cons corpus: The Pros and Cons corpus was crawled from reviews of epinions.com. It has 15162 Pros and 15162 Cons extracted from 15162 reviews of three types of products, i.e., digital cameras (8479), and printers (5778), and Strollers (905).

Table 1. Sentences from different sources

Data Sources	No. of Comparative Sentences
(Jindal and Liu 2006)	418
Reviews and forum posts	419
Total	837

5.2 Results

The results on the whole dataset are given in Table 2. Note that 84% of the sentences have EntityS1 as the preferred entity. If a system does nothing but simply announces that EntityS1 is preferred, we will have the accuracy of 84%. However, PCS using the OSA measure achieves the accuracy of 94.4%, which is much better than the baseline of taking the majority. Since in skewed datasets accuracy does not reflect the prediction well, we will mainly use precision (Prec.), recall (Rec.) and F-score (F) in evaluation. For the case that EntityS1 is preferred, the algorithm does extremely well. For the case that EntityS2 is preferred, the algorithm also does well although not as well as for the EntityS1 case. Based on our observation, we found that in such cases, the sentences are usually more complex.

Next, we compare with the case that the system does not use Pros and Cons (then OSA or PMI is not needed) (row 2). When a sentence requires context dependency handling, the system simply takes the majority as the default, i.e.,

assigning EntityS1 as the preferred entity. From the results in Table 2, we can see that F-scores are all worse. In the case that EntityS1 is the preferred entity, taking defaults is not so bad, which is not surprising because of the skewed data distribution. Even in this case, the precision improvement of PCS(OSA) is statistically significant at the 95% confidence level. The recall is slight less but their difference is not statistically significant. When EntityS2 is the preferred entity, its F-score (row 2) is much worse, which shows that our technique is effective. The recall improvement of PCS (OSA) is dramatic (statistically significant at the 95% confidence level). The two precisions are not statistically different. For OSA vs. PMI, see below.

Table 2: Preferred entity identification: whole data

	EntityS1 Preferred			EntityS2 Preferred		
	Prec.	Rec.	F	Prec.	Rec.	F
PCS (OSA)	0.967	0.966	0.966	0.822	0.828	0.825
PCS: No Pros & Cons	0.925	0.980	0.952	0.848	0.582	0.690
PCS (PMI)	0.967	0.961	0.964	0.804	0.828	0.816

Now let us look at only the 187 sentences that need context dependency handling. The data is still skewed. 72.2% of the sentences have EntityS1 as the preferred entities. Table 3 shows the results of PCS with and without using Pros and Cons. The results of PCS without Pros and Cons (OSA or PMI is not needed) are based on assigning EntityS1 as preferred for every sentence (taking the majority). Again, we can see that using external Pros and Cons (PCS(OSA)) helps dramatically. Not surprisingly, the improvements are statistically significant except the recall when EntityS1 is preferred.

Table 3: Preferred entity identification with 187 sentences that need context dependency handling

	EntityS1 Preferred			EntityS2 Preferred		
	Prec.	Rec.	F	Prec.	Rec.	F
PCS (OSA)	0.896	0.877	0.886	0.696	0.736	0.716
PCS: No Pros & Cons	0.722	1.000	0.839	0.000	0.000	0.000
PCS (PMI)	0.894	0.855	0.874	0.661	0.736	0.696

OSA vs. PMI: Comparing PCS(OSA) with PCS (PMI) (Table 3), OSA is better in F-score when EntityS1 is preferred by 1.2%, and better in F-score when EntityS2 is preferred by 2%. Although OSA's improvements over PMI are not large, we believe that in principle OSA is a more suitable measure. Comparing with PMI when the whole dataset is used (Table 2), OSA's gains are less because the number of sentences requiring context dependency handling is small (22%).

6 Conclusions

This paper studied sentiments expressed in comparative sentences. To our knowledge, no work has been reported on this topic. This paper proposed an effective method to solve the problem, which also deals with context based sentiments by exploiting external information available on the Web. To use the external information, we needed a measure of association of the comparative word and the entity feature. A new measure, called *one-side association* (OSA), was then proposed. Experimental results show that the technique produces accurate results.

Acknowledgement: The project is partially funded by Microsoft Corporation.

References

- Bos, J. and Nissim, M. An Empirical Approach to the Interpretation of Superlatives. *EMNLP-06*, 2006.
- Esuli, A and Sebastiani, F. Determining Term Subjectivity and Term Orientation for Opinion Mining. *EACL'06*, 2006.
- Fizman, M., Demner-Fushman, D., Lang, F., Goetz, P., and Rindfleisch, T. Interpreting Comparative Constructions in Biomedical Text. *BioNLP*, 2007.
- Gamon, M., Aue, A., Corston-Oliver, S. and Ringger, E.K. Pulse: Mining customer opinions from free text. *IDA'2005*.
- Hatzivassiloglou, V. and McKeown, K. Predicting the Semantic Orientation of Adjectives. *ACL-EACL'97*.
- Hu, M and Liu, B. Mining and summarizing customer reviews. *KDD'04*, 2004.
- Jindal, N. and Liu, B. Mining Comparative Sentences and Relations. *AAAI'06*, 2006.
- Kanayama, H and Nasukawa, T. Fully automatic lexicon expansion for domain-oriented sentiment analysis. *EMNLP'06*.
- Kim, S. and Hovy, E. Determining the Sentiment of Opinions. *COLING'04*, 2004.
- Liu, B, Hu, M. Cheng, J. Opinion Observer: Analyzing and Comparing Opinions on the Web. *WWW'05*.
- Mei, Q., Ling, X., Wondra, W., Su, H. and Zhai, C. Topic Sentiment Mixture: Modeling Facets and Opinions in Weblogs. *WWW'07*, 2007.
- Pang, B., Lee, L. and Vaithyanathan, S. Thumbs up? Sentiment Classification Using Machine Learning Techniques. *EMNLP'02*, 2002.
- Popescu, A.-M. and Etzioni, O. Extracting Product Features and Opinions from Reviews. *EMNLP'05*.
- Riloff, E & Wiebe, J. Learning extraction patterns for subjective expressions. *EMNLP'03*, 2003.
- Stoyanov, V. and Cardie, C. Toward opinion summarization: Linking the sources. In *Proc. of the Workshop on Sentiment and Subjectivity in Text*, 2006.
- Turney, P. Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews. *ACL-2002*.
- Wiebe, J. and Mihalcea, R. Word Sense and Subjectivity. *ACL'06*, 2006.
- Wilson, T., Wiebe, J. and Hwa, R. Just how mad are you? Finding strong and weak opinion clauses. *AAAI'04*, 2004.