Resolving Object and Attribute Coreference in Opinion Mining

Xiaowen Ding
Department of Computer Science
University of Illinois at Chicago
dingxwsimon@gmail.com

Bing Liu
Department of Computer Science
University of Illinois at Chicago
liub@cs.uic.edu

Abstract

Coreference resolution is a classic NLP problem and has been studied extensively by many researchers. Most existing studies, however, are generic in the sense that they are not focused on any specific text. In the past few years, opinion mining became a popular topic of research because of a wide range of applications. However, limited work has been done on coreference resolution in opinionated text. In this paper, we deal with object and attribute coreference resolution. Such coreference resolutions are important because without solving it a great deal of opinion information will be lost, and opinions may be assigned to wrong entities. We show that some important features related to opinions can be exploited to perform the task more accurately. Experimental results using blog posts demonstrate the effectiveness of the technique.

1 Introduction

Opinion mining has been actively researched in recent years. Researchers have studied the problem at the document level (e.g., Pang et al., 2002; Tuney, 2002; Gamon et al., 2005; sentence and clause level (Wilson et al., 2004; Kim and Hovy, 2004), word level (e.g., Hatzivassiloglou and McKeown, 1997; Esuli and Sebastiani, 2006; Kanayama and Nasukawa, 2006; Qiu et al., 2009), and attribute level (Hu and Liu 2004; Popescu and Etzioni, 2005; Ku et al., 2006; Mei et al., 2007; Titov and McDonald 2008). Here attributes mean different aspects of an object that has been commented on. Let us use the following example blog to illustrate the problem: "*I bought a Canon S500 camera yesterday. It looked beautiful. I took a few photos last night. They were amazing*". "*It*" in the second sentence refers to "*Canon S500 camera*", which is called an *object*. "They" in the fourth sentence refers to "photos", which is called an *attribute* of the object "*Canon S500 camera*". The usefulness of coreference resolution in this case is clear. Without resolving them, we lose opinions. That is, although we know that the second and fourth sentences express opinions, we do not know on what. Without knowing the opinion target, the opinion is of limited use. In (Nicolov et al., 2008), it was shown based on manually annotated data that opinion mining results can be improved by 10% if coreference resolution is used (the paper did not provide an algorithm).

In this paper, we propose the problem of *object and attribute coreference resolution* – the task of determining which mentions of objects and attributes refer to the same entities. Note that here *entities* refer to both objects and attributes, not the traditional named entities. To our knowledge, limited work has been done on this problem in the opinion mining context apart from a prior study on resolving opinion sources (or holders) (Stoyanov and Cardie 2006). Opinion sources or holders are the persons or organizations that hold some opinions on objects and attributes. In this paper, we do not deal with source resolution as we are mainly interested in opinion texts on the web, e.g., reviews, discussions and blogs. In such environments opinion sources are usually the authors of the posts, which are displayed in Web pages.

This work follows the *attribute-based opinion mining model* in (Hu and Liu 2004; Popescu and Etzioni, 2005). In their work, attributes are called *features*. We do not use the term "feature" in this paper to avoid confusion with the term "feature" used in machine learning.

Our primary interests in this paper are opi-

nions expressed on products and services, which are called *objects*. Each object is described by its parts/components and attributes, which are all called *attributes* for simplicity.

This paper takes the supervised learning approach to solving the problem. The key contribution of this paper is the design and testing of two novel opinion related features for learning. The first feature is based on sentiment analysis of normal sentences (non-comparative sentences), comparative sentences, and the idea of sentiment consistency. For example, we have the sentences, "*The Sony camera is better than the Canon camera. It is cheap too.*" It is clear that "It" means "Sony" because in the first sentence, the opinion on "Sony" is positive (comparative positive), but negative (comparative negative) on "Canon", and the second sentence is positive. Thus, we can conclude that "It" refers to "Sony" because people usually express sentiments in a consistent way. It is unlikely that "It" refers to "Canon". This is the idea of *sentiment consistency*. As we can see, this feature requires the system to have the ability to determine positive and negative opinions expressed in normal and comparative sentences.

The second feature considers what objects and attributes are modified by what opinion words. Opinion words are words that are commonly used to express positive or negative opinions, e.g., good, best, bad, and poor. Consider the sentences, "*The picture quality of the Canon camera is very good. It is not expensive either.*" The question is what "It" refers to, "Canon camera" or "picture quality". Clearly, we know that "It" refers to "Canon camera" because "picture quality" cannot be expensive. To make this feature work, we need to identify what opinion words are usually associated with what objects or attributes, which means that the system needs to discover such relationships from the corpus.

These two features give significant boost to the coreference resolution accuracy. Experimental results based on three corpora demonstrate the effectiveness of the proposed features.

2 Related Work

Coreference resolution is an extensively studied NLP problem (e.g., Morton, 2000; Ng and Cardie, 2002; Gasperin and Briscoe, 2008). Early knowledge-based approaches were domain and linguistic dependent (Carbonell and Brown 1988), where researchers focused on diverse lexical and grammatical properties of referring expressions (Soon et al., 2001; Ng and Cardie, 2002; Zhou et al., 2004). Recent research relied more on exploiting semantic information. For example, Yang et al. (2005) used the semantic compatibility information, and Yang and Su (2007) used automatically discovered patterns integrated with semantic relatedness information, while Ng (2007) employed semantic class knowledge acquired from the Penn Treebank. Versley et al. (2008) used several kernel functions in learning.

Perhaps, the most popular approach is based on supervised learning. In this approach, the system learns a pairwise function to predict whether a pair of noun phrases is coreferent. Subsequently, when making coreference resolution decisions on unseen documents, the learnt pairwise noun phrase coreference classifier is run, followed by a clustering step to produce the final clusters (coreference chains) of coreferent noun phrases. For both training and testing, coreference resolution algorithms rely on feature vectors for pairs of noun phrases that encode lexical, grammatical, and semantic information about the noun phrases and their local context.

Soon et al. (2001), for example, built a noun phrase coreference system based on decision trees and it was tested on two standard coreference resolution data sets (MUC-6, 1995; MUC-7, 1998), achieving performance comparable to the best-performing knowledge based coreference engines at that time. The learning algorithm used 12 surface-level features. Our proposed method builds on this system with additional sentiment related features. The features inherit from this paper includes:

Distance Feature: Its possible values are 0, 1, 2, 3 and so on which captures the sentence distance between two entities.

Antecedent-pronoun feature, anaphor-pronoun feature: If the candidate antecedent or anaphor is a pronoun, it is true; false otherwise.

Definite noun phrase feature: The value is true if the noun phrase starts with "the"; false otherwise.

Demonstrative noun phrase feature: The value is true if the noun phrase starts with the word "this", "that", "these", or "those"; false otherwise.

Number agreement feature: If the candidate antecedent and anaphor are both singular or both plural, the value is true; otherwise false.

Both-proper-name feature: If both the candidates are proper nouns, which are determined by capitalization, return true; otherwise false.

Alias feature: It is true if one candidate is an alias of the other or vice versa; false otherwise.

Ng and Cardie (2002) expanded the feature set of Soon et al. (2001) from 12 to 53 features. The system was further improved by Stoyanov and Cardie (2006) who gave a partially supervised clustering algorithm and tackled the problem of opinion source coreference resolution.

Centering theory is a linguistic approach tried to model the variation or shift of the main subject of the discourse in focus. In (Grosz et al., 1995; Tetreault, 2001), centering theory was applied to sort the antecedent candidates based on the ranking of the forward-looking centers, which consist of those discourse entities that can be interpreted by linguistic expressions in the sentences. Fang et al. (2009) employed the centering theory to replace the grammatical role features with semantic role information and showed superior accuracy performances.

Ding et al. (2009) studied the *entity assignment* problem. They tried to discover the product names discussed in forum posts and assign the product entities to each sentence. The work did not deal with product attributes.

Unsupervised approaches were also applied due to the cost of annotating large corpora. Ng (2008) used an Expectation-Maximization (EM) algorithm, and Poon and Domingos (2008) applied Markov Logic Network (MLN).

Another related work is the indirect anaphora, known as bridging reference. It arises when an entity is part of an earlier mention. Resolving indirect anaphora requires background knowledge (e.g. Fan et al., 2005), and it is thus not in the scope of this paper.

Our work differs from these existing studies as we work in the context of opinion mining, which gives us extra features to enable us to perform the task more effectively.

3 Problem of Object and Attribute Co-reference Resolution

In general, opinions can be expressed on anything, e.g., a product, an individual, an organization, an event, a topic, etc. Following (Liu, 2006), we also use the term *object* to denote an named entity that has been commented on. The object has a set of *components* (or parts) and also a set of *attributes*. For simplicity, *attribute* is used to denote both component and attribute in this paper. Thus, we have the two concepts, *object* and *attribute*.

3.1 Objective

Task objective: To carry out coreference resolution on objects and attributes in opinion text.

As we discussed in the introduction section, coreference resolution on objects and attributes is important because they are the core entities on which people express opinions. Due to our objective, we do not evaluate other types of co-references. We assume that objects and entities have been discovered by an existing system (e.g., Hu and Liu 2004, Popescu and Etzioni 2005). Recall that a coreference relation holds between two noun phrases if they refer to the same entity. For example, we have the following three consecutive sentences:

s_1: I love *the nokia n95* but not sure how good *the flash* would be?

s_2: and also *it* is quite expensive so anyone got any ideas?

s_3: I will be going on contract so as long as i can get a good deal of *it*.

"it" in s_2 refers to the entity "the nokia n95" in s_1. In this case, we call "the nokia n95" the *antecedent* and pronoun "it" in s_2 the *anaphor*. The referent of "it" in s_3 is also "the nokia n95", so the "it" in s_3 is coreferent with the "it" in s_2.

Our task is thus to decide which mentions of objects and attributes refer to the same entities.

3.2 Overview of Our Approach

Like traditional conference resolution, we employ the supervised learning approach by including additional new features. The main steps of our approach are as follows:

Preprocessing: We first preprocess the corpus by running a POS tagger [1], and a Noun Phrase finder [2]. We then produce the set O-NP which includes both possible objects, attributes and other noun phrases. The noun phrases are

---

[1] http://nlp.stanford.edu/software/tagger.shtml

[2] http://crfchunker.sourceforge.net/

found using the Noun Phrase finder and the object names are consecutive NNPs. O-NP thus contains everything that needs to be resolved.

Feature vector construction: To perform machine learning, we need a set of features. Similar to previous supervised learning approaches (Soon et al., 2001), a feature vector is formed for every pair of phrases in O-NP extracted in the preprocessing step. We use some of the features introduced by Soon et al. (2001) together with some novel new features that we propose in this work. Since our focus is on products and attributes in opinionated documents, we do not use personal pronouns, the gender agreement feature, and the appositive feature, as they are not essential in blogs and forum posts discussing products.

Classifier construction: Using the feature vectors obtained from the previous step, we construct the training data, which includes all pairs of manually tagged phrases that are either object names or attributes. More precisely, each pair contains at least one object or one attribute. Using the training data, a decision tree is constructed using WEKA[3].

Testing: The testing phase employs the same preprocessing and feature vector construction steps as described above, followed by the application of the learnt classifier on all candidate coreference pairs (which are represented as feature vectors). Since we are only interested in coreference information for objects and attribute noun phrases, we discard non-object and non-attribute noun phrases.

4 The Proposed New Features

On surface, object and attribute coreference resolution seems to be the same as the traditional noun phrase coreference resolution. We can apply an existing coreference resolution technique. However, as we mentioned earlier, in the opinion mining context, we can have a better solution by integrating opinion information into the traditional lexical and grammatical features. Below are several novel features that we have proposed. We use α_i to denote an antecedent candidate and α_j an anaphor candidate. Note that we will not repeat the features used in previous systems, but only focus on the new features.

---

[3] http://www.cs.waikato.ac.nz/ml/weka/

4.1 Sentiment Consistency

Intuitively, in a post, if the author starts expressing opinions on an object, he/she will continue to have the same opinion on that object or its attributes unless there are contrary words such as "but" and "however". For example, we have the following blog (an id is added before each sentence to facilitate later discussion):

"(1) I bought Camera-A yesterday. (2) I took a few pictures in the evening in my living room. (3) The images were very clear. (4) They were definitely better than those from my old Camera-B. (5a) It is cheap too. (5b) The pictures of that camera were blurring for night shots, but for day shots it was ok"

The comparative sentence (4) says that Camera-A is superior to Camera-B. If the next sentence is (5a) ((5a) and (5b) are alternative sentences), "it" should refer to the superior product/object (Camera-A) because sentence (5a) expresses a positive opinion. Similarly, if the next sentence is sentence (5b) which expresses a negative opinion in its first clause, "that camera" should refer to the inferior product (Camera-B). We call this phenomenon *sentiment consistency* (*SC*), which says that consecutive sentiment expressions should be consistent with each other unless there are contrary words such as "but" and "however". It would be ambiguous if such consistency is not observed.

Following the above observation, we further observe that if the author wants to introduce a new object o, he/she has to state the name of the object explicitly in a sentence s_{i-1}. The question is what happens to the next sentence s_i if we need to resolve the pronouns in s_i.

We consider several cases:

1. s_{i-1} is a normal sentence (not a comparative sentence). If s_i expresses a consistent sentiment with s_{i-1}, it should refer to the same object as s_{i-1}. For example, we have

   s_{i-1}: *The N73* is my favorite.

   s_i: *It* can produce great pictures.

   Here "It" in s_i clearly refers to "The N73" in the first sentence s_{i-1}.

2. s_{i-1} is a normal sentence and s_i does not express a consistent sentiment, then α_i and α_j introduced in these two sentences may not be coreferenced. For example, we have

   s_{i-1}: *The K800* is awesome.

   s_i: *That phone* has short battery life.

Here "The K800" and "That phone" may not be a coreference pair according to sentiment consistency. "That phone" should refer to an object appeared in an earlier sentence.

3. s_{i-1} is a comparative sentence. If s_i expresses a positive (respectively negative) sentiment, the pronoun in s_i should refer to the superior (or inferior) entity in s_{i-1} to satisfy sentiment consistency. This situation is depicted in the earlier example blog. For completeness, we give another example.

> s_{i-1}: *The XBR4* is brighter than *the 5080*.
> s_i: Overall, *it* is a great choice.

Here "it" in s_i should refer to "The XBR4" in s_{i-1} since they both have positive sentiments expressed on them.

Opinion Mining of Comparative Sentences: To deal with case (3), we need to identify superior entities from comparative sentences. In fact, we first need to find such comparative sentences. There is a prior work on identifying comparative sentences (Jindal and Liu. 2006). Since our focus is not to identify such sentences, we used several heuristic rules based on some comparative keywords, e.g. *than*, *win*, *superior*, etc. They achieve the F-score of 0.9. We then followed the opinion mining method introduced in (Ding et al. 2009) to find superior entities. Since a comparative sentence typically has entities on the two sides of a comparative keyword, i.e., "*Camera-X is better than Camera-Y*", based on opinion mining, if the sentence is positive, then the entities before the comparative keyword is superior and otherwise they are inferior (with the negation considered).

SC Feature: The possible value for this feature is 0, 1, or 2. If α_i and α_j have the same opinion, return 1; different opinions, return 0; and if the opinions cannot be identified for one or both of them, return 2. Here is an example explaining how the feature is used in our system:

> "*My wife has currently got a Nokia 7390, which is terrible. My 6233 would always get great reception, hers would get no signal*."

Using our algorithm for opinion mining, "hers" gets a negative opinion in the second sentence. So the value for this feature for the pair, "hers" and "a Nokia 7390", is 1. The feature value for the pair "hers" and "My 6233" is 0. The idea is that because the first sentence expresses a negative sentiment on "a Nokia 7390", and there is no discourse connective (such as "but" and "however") between these two sentences. "Hers" should be talking about "a Nokia 7390" so as to satisfy sentiment consistency.

4.2 Entity and Opinion Word Association

One of the most important factors determining the orientation of opinions is the opinion words that opinion holders use to express their opinions. Different entities may be modified by different opinion words. We can use their association information with entities (both objects and attributes) to identify their coreferences.

Opinion Words: In most cases, opinions in sentences are expressed using *opinion words*. For example, the sentence, "*The picture quality is amazing*", expresses a positive opinion on the "picture quality" attribute because of the positive opinion word "amazing".

Researchers have compiled sets of such words for adjectives, adverbs, verbs, and nouns respectively. Such lists are collectively called the *opinion lexicon*. We obtained an opinion lexicon from the authors of (Ding et al. 2009).

It is useful to note that opinion words used to express opinions on different entities are usually different apart from some general opinion words such as good, great, bad, etc, which can express opinions on almost anything. For example, we have the following passage:

> "*i love the nokia n95 but not sure how strong the flash would be? And also it is quite expensive, so anyone got any ideas*?"

Here "strong" is an opinion word that expresses a positive opinion on "the flash", but is seldom used to describe "the nokia n95". "expensive", on the other hand, should not be associated with "the flash", but is an opinion word that indicates a negative opinion on "the nokia n95". So "the nokia n95" is more likely to be the antecedent of "it" in the second sentence.

The question is how to find such associations of entities and opinion words. We use their co-occurrence information to measure, i.e., the *pointwise mutual information* of the two terms. First, we estimate the probability of $P(NP)$, $P(OW)$ and $P(NP\&OW)$. Here NP means a noun phrase, e.g., an object (attribute) after removing determiners, and OW means an opinion word. To compute the probability, we first count the occurrences of the words. Then the probability is computed as follow:

$$P(NP, OW) = \frac{NumofS(NP \& OW)}{TotalNumofSentence}$$

where *NumofS* is a function that gives the number of sentences that contain the particular word string. *P(NP, OW)* is computed in the same way. Let us use the previous example again. We compute *P*("nokia n95","expensive") as the number of sentences containing both "nokia n95" and "expensive" divided by the total number of sentences in the whole corpus.

Then we use the pointwise mutual information between a noun phrase and an opinion word to measure the association.

$$PMI(NP, OW) = \log\frac{P(NP, OW)}{P(NP)P(OW)}$$

However, this PMI value cannot be encoded directly as a feature as it only captures the local information between antecedent candidates and opinion words. That is, it cannot be used as a global feature in the classifier. We thus rank all possible antecedents of anaphor α_j based on their *PMI* values and use the ranking as the feature value. The highest ranked antecedent α_i has value 1; the second one has value 2 and so on. The candidates ranked below the fourth place all have the value 5. In the example above, if *PMI*("nokia n95", "expensive") is greater than *PMI*("flash", "expensive"), the feature for "nokia n95" and "it" pair will have a smaller value than the feature for the "flash" and "it" pair.

One may ask if we can use all adjectives and adverbs to associate with objects and attributes rather than just opinion words since most opinion words are adjectives and adverbs. We tested that, but the results were poor. We believe the reason is that there are many adjectives and adverbs which are used for all kinds of purposes and may not be meaningful for our task.

4.3 String Similarity Feature

Soon et al. (2001) has a string match feature (SOON STR), which tests whether the two noun phrases are the same string after removing determiners from each. Ng and Cardie (2002) split this feature into several primitive features, depending on the type of noun phrases. They replace the SOON STR feature with three features — PRO STR, PN STR, and WORDS STR — which restrict the application of string matching to pronouns, proper names, and non-pronominal

noun phrases, respectively.

In the user generated opinion data, these may not be sufficient. For a certain product, people can have a large number of ways to express it. For example, we have

"Panasonic TH50PZ700U VS TH50PZ77U, Which Plasma tv should I go for. The TH77U is about $500.00 more than the 700U."

Here "TH77U" is the same entity as "Panasonic TH50PZ77U", and "TH50PZ700U" is the same as "700U". But they cannot be easily identified by "same string" features mentioned above. Although "700U" can be solved using substring features, "TH77U" is difficult to deal with.

We employ a modified edit distance to computing a similarity score between different mentions and use that as a feature in our system. When one candidate is a substring of another, return 1; otherwise, 1 plus the edit distance.

4.4 Other Useful Features

In the machine learning approach introduced by Soon et al. (2001), they had several general features that can deal with various kinds of entities, e.g., semantic class agreement features dealing with different semantic classes like date, location, etc., and the gender agreement feature related to personal entities. However, these features are not so useful for our task because the semantic class of a product in one domain is usually consistent, and dates and locations are unlikely to be of any products that people will express their opinions. Moreover, we do not study opinion holders (as they are known in the Web environment), so personal entities are not the aspect that we concentrate on. Thus we did not use the following features: semantic class agreement features, the gender agreement feature, and appositive feature.

However, we added some specific features, which are based on two extracted entities, α_i and α_j, where α_i is the potential antecedent and α_j is the potential anaphor:

Is-between feature: Its possible values are true and false. If the words between α_i and α_j have an is-like verb (i.e., is, are, was, were, and be) between them and there is no comparative indicators, this feature has the value of true, e.g., **"***The nokia e65* is *a good handset***."**

In sentences similar to this example, the entities before and after "is" usually refer to the same object or attribute by a definition relation.

And the value of this feature will be true.

If "is" appears together with a comparative word, it is probably an indication that the two entities are different, and the value for this feature will be false, e.g., **"*Overall the K800 is far superior to the W810.*"**

Has-between feature: Its possible values are also true and false. If the words between α_i and α_j have a has-like verb (i.e., has, have, and had), the value is true, and otherwise false, e.g., **"*The k800* has *a 3.2 megapixel camera.*"**

This feature usually indicates a "part-of" relation if "has" appears between two entities. They do not refer to the same entity. Table 1 gives a summary of all the features used in our system.

5 Experiments and Discussions

5.1 Datasets

For evaluation, we used forum discussions from three domains, mobile phones, plasma and LCD TVs, and cars. Table 2 shows the characteristics of the three data sets. Altogether, we downloaded 64 discussion threads, which contain 453 individual posts with a total of 3939 sentences. All the sentences and product names were annotated strictly following the MUC-7 coreference task annotation standard[4]. Here is an example:

"*Phil had <COREF ID = "6" TYPE = "OBJ">a z610</COREF> which has <COREF ID = "7" TYPE = "ATTR">a 2MP cemara</COREF>, and he never had a problem with <COREF ID = "8" TYPE = "OBJ" REF = "6">it</COREF>.*"

ID and REF features are used to indicate that there is a coreference link between two strings. ID is arbitrary but uniquely assigned to each noun phrase. REF uses the ID to indicate a coreference link. "TYPE" can be "OBJ" (an object or a product), or "ATTR" (an attribute of an object). The annotation was done by the first author and another student before the algorithm construction, and the annotated data sets will be made public for other researchers to use.

For our experiments, we used the J48-decision tree builder in WEKA, a popular of machine learning suite developed at the University of Waikato. We conducted 10-fold cross validation on each dataset.

The performances are measured using the standard evaluation measures of precision (p), recall (r) and F-score (F), $F = 2pr/(p+r)$. As we stated in Section 3, we are only interested in object and attributes noun phrases. So in the testing phrases, we only compute the precision and recall based on those pairs of candidates that contain at least one object or attribute noun phrase in each pair. If both of the candidates are not an object or an attribute, we ignore them.

5.2 Baseline

As the baseline systems, we duplicated two representative systems. Baseline1 is the decision tree system in Soon et al. (2001). We do not use the semantic class agreement feature, gender agreement feature and appositive feature in the original 12 features for the reason discussed in Section 4.4. Thus, the total number of features in Baseline1 is 9. The second baseline (baseline2) is based on the centering theory from the semantic perspective introduced by Fang et al. (2009). Centering theory is a theory about the local discourse structure that models the interaction of referential continuity and the salience of discourse entities in the internal organization of a text. Fang et al. (2009) extended the centering theory from the grammar level to the semantic level in tracking the local discourse focus.

5.3 Results Analysis

Table 3 gives the experimental results of the two baseline systems and our system with different features included. From Table 3, we can make several observations.

(1) Comparing the results of Baseline1 and our system with all features (Our System (All)), the new features introduced in this paper improves Baseline1 on average by more than 9% in F-score.

(2) Comparing the results of Baseline2 and our system with all features (Our System (All)), our system performs better than Baseline2 by about 3 - 5%. We also observe that centering theory (Baseline2) is indeed better than the traditional decision tree.

(3) Our system with sentiment consistency (SC) makes a major difference. It improves Baseline1 (our method is based on Baseline1) by 5-6% in F-score.

(4) With the additional feature of entity and opinion association (EOA), the results are

Feature category	Feature	Remark
Opinion mining based features	Opinion consistency	1, if the opinion orientation of α_i is the same as α_j, 0 if the opinions are different, else 2
	Entity and opinion words association	1, 2, 3, 4, 5 which indicate the rank positive based on the *PMI* value introduced in Section 4.2
grammatical	i-Pronoun feature	1, if α_i is a pronoun, else 0
	j-Pronoun feature	1, if α_j is a pronoun, else 0
	Number agreement feature	1, if both of the noun phrases agree in numbers, else 0
	Definite feature	1, if α_j starts with the word "the", else 0
	Demonstrative feature	1, if α_j starts with the word "this", "that", "those", or "these", else 0
	Both proper-name feature	1, if α_i and α_j are both proper names, else 0
lexical	String similarity	The string similarity score between α_i and α_j
	Alias feature	1, If α_i is an alias of α_j or vice versa, else 0
Others	Distance feature	The sentence distance between the pair of noun phrases, 0 if they are in the same sentence
	Keywords between features	1, if some keywords exist between α_i and α_j, else 0. Details are discussed in Section 4.5

Table 1: Feature list: α_i denotes the antecedent candidate and α_j the anaphor candidate

	Posts	Sentences
Phone	168	1498
TVs	173	1376
Cars	112	1065
Total	453	3939

Table 2: Characteristics of the datasets

		Cellphone			TVs			Cars		
		p	r	**F**	p	r	**F**	p	r	**F**
1	Baseline1	0.66	0.57	**0.61**	0.67	0.61	**0.64**	0.70	0.63	**0.66**
2	Baseline2	0.70	0.64	**0.67**	0.72	0.65	**0.68**	0.76	0.70	**0.73**
3	Our System (SC)	0.71	0.64	**0.67**	0.73	0.66	**0.69**	0.74	0.69	**0.72**
4	Our System (SC+EOA)	0.74	0.68	**0.71**	0.74	0.68	**0.71**	0.77	0.71	**0.74**
5	Our System (All)	0.75	0.70	**0.72**	0.76	0.70	**0.73**	0.78	0.73	**0.75**

Table 3: Results of object and attribute coreference resolution

improved further by another 2-4%.

(5) Our system with all features (row 5) performs the best.

Paired *t*-tests were performed on the three systems, i.e., baseline1, baseline2, and our system (row 5). The tests show that the improvements of our method over both Baseline1 and Baseline2 are significant at the confidence level of 95% for the first two datasets. For the third dataset, the improvement over Baseline1 is also significant at the confidence level of 95%, while the improvement over Baseline2 is significant at the confidence level of 90%.

In summary, we can conclude that the new technique is effective and is markedly better than the existing methods. It is clear that the new features made a major difference.

6 Conclusion

This paper investigated the coreference resolution problem in the opinion mining context. In particular, it studied object and attribute resolutions which are crucial for improving opinion mining results. Although we still took the supervised learning approach, we proposed several novel features in the opinion mining context, e.g., sentiment consistency, and object/attribute and opinion word associations. Experimental results using forum posts demonstrated the effectiveness of the proposed technique. In our future work, we plan to further improve the method and discover some other opinion related features that can be exploited to produce more accurate results.

References

A. Andreevskaia and S. Bergler. 2006. *Mining WordNet for Fuzzy Sentiment: Sentiment Tag Extraction from WordNet Glosses*. EACL'06.

J. Carbonell and R. Brown. 1988. *Anaphora resolution: a multi-strategy approach. COLING'1988*.

G. Carenini, R. Ng, and A. Pauls. 2006. *Interactive multimedia summaries of evaluative text*. IUI'06.

X. Ding, B. Liu and L. Zhang. 2009. *Entity Discovery and Assignment for Opinion Mining Application*. KDD'09.

A. Esuli and F. Sebastiani. 2006. *Determining Term Subjectivity and Term Orientation for Opinion Mining*, EACL'06.

J. Fan, K. Barker and B. Porter. 2005. *Indirect anaphora resolution as semantic path search*. K-CAP'05.

C. Gasperin and T. Briscoe. 2008. *Statistical anaphora resolution in biomedical texts*. COLING'08

B. J. Grosz, A. K. Joshi and S. Weinstein. 1995. *Centering: a framework for modeling the local coherence of discourse*. Computational Linguistics, 21(2).

V. Hatzivassiloglou and K. McKeown. 1997. *Predicting the Semantic Orientation of Adjectives*. ACL-EACL'97.

M. Hu and B. Liu. 2004. *Mining and summarizing customer reviews*. KDD'04.

N. Jindal, and B. Liu. 2006. *Mining Comparative Sentences and Relations*. AAAI'06.

H. Kanayama and T. Nasukawa. 2006. *Fully Automatic Lexicon Expansion for Domain-Oriented Sentiment Analysis*. EMNLP'06.

S. Kim and E. Hovy. 2004. *Determining the Sentiment of Opinions*. COLING'04.

N. Kobayashi, K. Inui and Y. Matsumoto. 2007. *Extracting Aspect-Evaluation and Aspect-of Relations in Opinion Mining*. EMNLP-CoNLL'07.

F. Kong, G. Zhou, Q. Zhu and P. Qian. 2009. *Employing the Centering Theory in Pronoun Resolution from the Semantic Perspective*. EMNLP'09.

L.-W. Ku, Y.-T. Liang and H.-H. Chen. 2006. *Opinion Extraction, Summarization and Tracking in News and Blog Corpora*. CAAW'06.

B. Liu. 2006. *Web Data Mining*, Springer.

R. McDonald, K. Hannan, T Neylon, M. Wells, and J.Reynar. 2007. *Structured Models for Fine-to-Coarse Sentiment Analysis*. ACL-07.

Q. Mei, X. Ling, M. Wondra, H. Su, and C. Zhai. 2007. *Topic Sentiment Mixture: Modeling Facets and Opinions in Weblogs*. WWW'07.

T. S. Morton. 2000. *Coreference for NLP applications*. ACL'00.

V. Ng and C. Cardie. 2002. *Improving machine learning approaches to coreference resolution*. ACL'02.

V. Ng. 2007. *Semantic Class Induction and Coreference Resolution*. ACL'07.

V. Ng. 2008. *Unsupervised Models for Coreference Resolution*. EMNLP'08.

N. Nicolov, F. Salvetti and S. *Ivanova, Sentiment analysis: Does coreference matter?* AISB'2008.

B. Pang, L. Lee, and S. Vaithyanathan. 2002. *Thumbs up? Sentiment Classification Using Machine Learning Techniques*. EMNLP'02.

H. Poon and P. Domingos. 2008. *Joint Unsupervised Coreference Resolution with Markov Logic*. EMNLP'08, 650–659.

A-M. Popescu and O. Etzioni. 2005. *Extracting product features and opinions from reviews*. EMNLP'05.

Qiu, Guang, B. Liu, J. Bu and C. Chen. 2009 *Expanding Domain Sentiment Lexicon through Double Propagation*. IJCAI 2009.

W. M. Soon, H. T. Ng and D. Lim. 2001. *A machine learning approach to coreference resolution of noun phrase*. Computational Linguistics, 27(4).

V. Stoyanov, C. Cardie. 2006. *Partially supervised coreference resolution for opinion summarization through structured rule learning.* EMNLP'06.

I. Titov and R. McDonald. 2008, *A joint model of text and aspect ratings for sentiment summarization*, ACL'08.

J. Tetreault. 2001. *A corpus-based evaluation of centering and pronoun resolution*. Computational Linguistics. 27(4):507-520.

P. Turney. 2002. *Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews.* ACL'02.

Y. Versley, A. Moschitti, M. Poesio and X. Yang. 2008. *Coreference systems based on kernels methods.* COLING'08.

T. Wilson, J. Wiebe, and R. Hwa. 2004. *Just how mad are you? Finding strong and weak opinion clauses.* AAAI'04.

X. F. Yang, J. Su and C. L. Tan. 2005. *Improving Pronoun Resolution Using Statistics - Based Semantic Compatibility Information*. ACL'05.

X. F. Yang and J. Su. 2007. *Coreference Resolution Using Semantic Relatedness Information from Automatically Discovered Patterns.* ACL'07.

G. D. Zhou and J. Su. 2004. *A high-performance coreference resolution system using a multi-agent strategy*. COLING'04.