



**HAL**  
open science

# Improving Opinion Target Identification through Twitter Using Public Conversations

Rania Othman, Rami Belkaroui, Rim Faiz

► **To cite this version:**

Rania Othman, Rami Belkaroui, Rim Faiz. Improving Opinion Target Identification through Twitter Using Public Conversations. 24th Italian Symposium on Advanced Database Systems SEBD, Jun 2016, Ugento, Lecce, Italy. hal-01419210

**HAL Id: hal-01419210**

**<https://hal.science/hal-01419210v1>**

Submitted on 23 Dec 2016

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Improving Opinion Target Identification through Twitter Using Public Conversations

Rania Othman<sup>1</sup>, Rami Belkaroui<sup>1</sup> and Rim Faiz<sup>2</sup>

<sup>1</sup> Institut Supérieur de Gestion de Tunis, LARODEC, Université de Tunis, Tunisia

<sup>2</sup> IHEC Carthage, LARODEC, Université de Carthage, Tunisia

rania.othman@gmx.com, rami.belkaroui@gmail.com, rim.faiz@ihec.rnu.tn

**Abstract.** In this paper, we propose a new method for opinion target identification based on twitter conversations rather than simple individual tweets. We employ conversation interactions to effectively extract the different target features using a product review corpus involving smartphones and other electronics products. Experimental evaluations show that our proposed method is efficient and contributes to improving system performance.

**Keywords:** Twitter, Conversations, Opinion targets, Anaphora, User interactions.

## 1 Introduction

Twitter is currently one of the most popular micro-blogs which has grown at an unprecedented rate to reach over 320 million monthly active users<sup>3</sup>. Twitter is both a micro-blogging service and a conversational environment that enables people to interact, engage in daily chatter, join conversations, report news and share information. Conversations are key element in such service. Almost a quarter of Twitter users hold conversations with other users through this platform [8] and huge percentage of Twitter posts are conversational [14]. The huge volume of conversations produced everyday makes them an interesting information source and valuable tool to discover new trends and exchange opinions and feedbacks towards products, news and stories. A large number of the messages are carriers of opinions and feelings [9]. One major field that took advantage of this potential is e-reputation. It has become a common practice for business owners to allow their customers to review their products through public web sites (e.g. epinions.com, amazon.com). This enables businesses to have general overview on their consumer satisfaction and attitudes about the product or the service offered. While conversations are very rich source of information, retrieving relevant information remains challenging. The large number of messages involved in these conversations makes it hard for users to cope with rate of messages involving a lot of noise and redundancy. Thus, applying automated systems that aim to retrieve individual opinion words or phrases and what they are about, referred to as the opinion target or opinion topic, could be quite useful for business owners.

---

<sup>3</sup> <https://about.twitter.com/company>

In this paper, we propose a new method for opinion target identification based on Twitter conversations. To date, little research has been addressed to tackle conversations despite their prevalence in social streams. Most existing works on opinion analysis through Twitter, have so far focused on handling simple individual tweets rather than considering the whole conversations [12, 4, 1]. As tweets are limited to 140 characters and usually written in an informal way involving a lot of abbreviations, typos and slangs, it is frequently hard to detect the exact meaning of a tweet when taken separately. Also, a large number of features are generally referenced by anaphoric pronouns in major succeeding replies of a conversation message. Anaphora can be defined as the use of an expression the interpretation of which depends specifically upon a previous expression [15]. Consider this example sentence: “It is fantastic.”. If one wants to extract what the opinion in this sentence is about, previous replies must be analysed. Frequently, while handling conversations, a large amount of messages are left unnoticed due to their vague and unclear aspect. Therefore, we adopt a conversation-based method that employs conversation interactions, notably reply links, to effectively extract the target product features involved in the messages. To the best of our knowledge, this is the first research employing twitter conversations in the opinion target identification task.

The rest of this paper is organized as follows. Section 2 overviews the related work on opinion target identification. In section 3, we introduce our new approach for opinion target identification on Twitter conversations. Section 4 reports the experiment results. Finally, Section 5 concludes this paper with some perspectives.

## 2 RELATED WORK

Opinion target extraction is crucial for opinion mining (OM) and summarization especially given that this task provides the foundation for opinion summarization [5]. Opinion target can be defined as the entity ( i.e., person, object, feature, event or topic) about which the user expresses his opinion. Extensive approaches and techniques have been addressed to mine opinion components or targets from unstructured reviews. These works can be very broadly divided into two main categories supervised and unsupervised. Other works have also employed the semi-supervised approach. In the supervised learning approaches, a machine-learning model is trained on manually labeled data to extract and classify the feature set in the reviews. Although these techniques provide good results for opinion target extraction, they require extensive manual work for the training set preparation, they are also time consuming, and sometimes domain dependent. The most common techniques employed in supervised approaches are decision tree, support vector machine (SVM), K-nearest neighbor (KNN), naive Bayesian classifier and neural network [7, 11, 17]. On the other hand, unsupervised approaches automatically extract product features using syntactic and contextual patterns without the need of labeled data [6, 9]. A challenge which is frequently encountered in the opinion target extraction task is, that entities can be sometimes implicit and therefore hard to find. In the case of explicit target identification, we generally employ noun phrases with syntactic rules [6, 13]. While for implicit targets, context dependency or distribution similarity are

employed. To the best of our knowledge, there is currently only two systems that integrate coreference information and apply anaphora resolution (AR) in OM. Stoyanov and Cardie [16] develop an algorithm that identifies coreferring targets in newspaper articles. They rely on manually annotated targets thus, a candidate selection phase for the opinion targets is not required. The authors focus only on the coreference resolution but they do not resolve pronominal anaphora in order to achieve this purpose. For their part, [7] adapt the rule based AR algorithm CogNIAC to extract opinion targets on a movie review corpus. They have shown that extending an OM algorithm with AR for opinion target extraction can achieve significant improvements. In this paper, we apply AR to extract opinion targets ( i.e., product features) from Twitter conversations. To the best of our knowledge, this is the first research employing twitter conversations and AR for the feature extraction task.

### 3 PROPOSED APPROACH

In our work, we aim to extract product features reviewed by users in our tweet collection based on Twitter conversations. Our approach is based on three main modules namely, pre-processing, conversation retrieval and opinion target identification. Figure 1 presents an overview of the proposed approach. Given a tweet

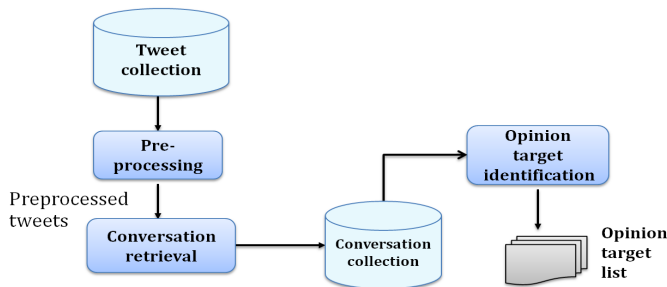


Fig. 1. Overview of our proposed approach

corpus on electronic products, we clean our collection via the pre-processing phase, than we proceed to retrieve conversations from the collection of separate tweets. The final step consists on extracting opinion targets commented by users through the conversation collection. The three modules are illustrated in details in the following subsections.

#### 3.1 Pre-processing

The collection of tweets is preprocessed before the conversation retrieval module. First of all, the URL in each message is analyzed. While different shortened URLs might redirect to the same end URL, it is necessary to replace them with the real URLs they redirect. Thereafter, we employ an API service <sup>4</sup> for HTML text extraction which removes comments, links, ads, and other unrelated parts of a web page and returns key contents in plain text. Then, we proceed to the second step of the processing involving the text analysis. In this step,

<sup>4</sup> <http://www.alchemyapi.com/>

both the text content of the tweet and the text retrieved from the URLs are taken into account. We cleaned the text by removing ASCII characters, numbers, punctuation, and stop words. In the end, we convert the text to lower case and we tokenize it. The remaining tweet features such as the ID of the author and other social information are also extracted from the API and stored. We indexed the collection of tweets with Apache Lucene<sup>5</sup> which is full-featured text search engine library written entirely in Java.

### 3.2 Conversation Retrieval Module

In this section, we describe the process of constructing conversations from a collection of tweets. We consider the conversation definition presented by [3]. They defined a conversation as a set of short text messages posted by a set of users at specific timestamps on the same topic. This messages can be directly replied to other users by using “@username” or indirectly by liking, retweeting, commenting and other possible interactions. We apply the same method developed by [3] who proposed a user-based tree model for retrieving conversations from microblogs. They do not only retrieve direct messages based on reply links but also consider indirect messages that can be related to the conversation via other links such as retweet and mention interactions. To avoid bias related to features existing with a very small number, we cleaned our collection of conversations by filtering out the conversations that involve less than 3 participants and containing less than 7 tweets.

### 3.3 Opinion Target Identification Module

In this stage, we aim to extract the opinion targets or features customers expressed their opinions on. For example, if we want to generate an OM about iPod, some of the common features are “battery life”, “sound quality” and “ease of use”. Given a conversation collection, our system splits the reviews into sentences, then, it converts them to lower case and remove the non literal characters at the beginning and the end of each word (e.g. “#IPhone#” becomes “iphone”). Steinberger et al.[15] reveal that noun and noun phrases in the sentence are likely to be the features customers expressed their opinions on. We therefore perform the Part-of-Speech (POS) tagging of the whole document to identify the grammatical class of each word using TreeTagger<sup>6</sup>. We extract nouns from the reviews and we move on to the feature decider step. We construct our stop word list and we filter out the extracted nouns existing in the stop word list. Then, we construct noun phrases which are composed of two successive nouns (e.g. Click wheel, Battery Life). We extract all noun phrases but we only keep those appearing together at least 3 times in the reviews. We remove sentence redundancy, if a noun appears more than one time in the same sentence, we consider as if it appears one time. We then compute frequency of occurrences in the reviews for the whole set of extracted nouns and we only keep those whose frequency is greater than 0.02.

To decide if the noun phrases we collected are meaningful, we apply a similarity measure called point-wise mutual information (PMI-IR) [18] that uses

<sup>5</sup> <https://lucene.apache.org/>

<sup>6</sup> <http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/>

page counts returned by a web search engine to recognize synonyms. Like [9], we detect the compactness of a noun phrase using the number of tweets concerning a given product instead of the search engine page counts. We prune noun phrases having  $\text{PMI} < 0$ .

Having our feature list, we proceed to the next step which aims to affect each review (or post/message) in our corpus to the feature the author comment on in his post. To do so, we apply our proposed Target Identification Algorithm based on conversation interactions notably the reply-links. Indeed, a large number of features appearing as noun phrases and even simple nouns in reviews are generally referenced by anaphoric pronouns in major succeeding sentences of a review document [10, 7]. The anaphoric term is called an anaphor. For example, in the sentence: “I’m so glad with my new iPhone 5, it’s just amazing”, the pronoun “it” is an anaphor, referring back to the antecedent “my new iPhone 5”. Actually, the overwhelming majority of the opinion targets ( i.e., product features) are pronouns in the datasets [10]. Thus, AR is crucial for binding feature-review pairs, otherwise a very big number of opinion reviews will be left unnoticed due to their ambiguous aspect. In order to identify the associations of such reviews with correct features, a conversation-interactions based algorithm has been developed. To the best of our knowledge, this is the first such algorithm that employs conversation interactions notably reply-links for effective binding of feature-review pairs. Our algorithm proceeds as follows, for a tweet  $t_{i,l}$ , that does not contain any feature words existing in the feature list but involves opinionated words ( i.e., adjectives and adverbs) along with some pronouns, all anaphora pronouns present in this sentence that require mapping are extracted, and a set of anaphora  $P = \{p_1, p_2, p_3, \dots, p_n\}$  is compiled for proper context determination. We employ a backtracking mechanism in which review documents are accessed in reverse order based on reply-links to extract precedent sentences  $t_{i-1,j}$  that  $t_{i,l}$  replies on. For each anaphora pronoun  $p_i \in P$ , proper context is determined to compile a set  $A = \{a_{k,1}, a_{k,2}, a_{k,3}, \dots, a_{q,m}\}$  consisting of candidate antecedents. The best antecedent  $a_{k,t} \in A$  is selected for binding with  $a_o$  using CogNIAC algorithm [2], a publically available algorithm for AR that employs a rule based approach for antecedent identification. This approach can be an adequate strategy for our OM task, since in our corpus, a small percentage of the total number of pronouns are actual product features (only 6%). We denote the  $l^{th}$  tweet at iteration  $i$   $t_{i,l}$  while the  $j^{th}$  tweet at iteration  $i$  is denoted  $t_{i,j}$ . As each anaphora pronoun is replaced by the selected antecedent, backtracking process terminates with the iteration  $i-1$  for each sentence. At the end of this phase, we obtain a set of opinionated sentences and each sentence is associated to the corresponding target feature that the user commented on.

## 4 EXPERIMENTS AND RESULTS

### 4.1 Dataset Description

Due to the lack of test collections for Twitter conversations, we have created our own collection. We crawled 221 663 English tweets using Twitter Application Programmable Interface (API) <sup>7</sup>. The Twitter API allows developers around the

<sup>7</sup> <https://dev.twitter.com/overview/api>

world to have free and open access to Twitter’s database. The tweet collection was crawled over a period of 4 months from April 25th 2015 to July 25th 2015.

We only search popular tweets talking about a given product involving character description, promotion information and comments about new products. After removing the repeated ones, 211 350 tweets remained. From our collection of tweets, we have constructed 8 720 conversations involving 64 370 tweets and 13 827 bloggers. We employ statuses/lookup.json files accessed through Twitter API, which contain all information related to the tweets. Table 1 outlines some statistics on our Conversation collection. As shown, there exist over 120K pronouns and roughly 11.13 % of the opinion targets are referred to by pronouns.

**Table 1.** Corpus size statistics

|                            |           |
|----------------------------|-----------|
| Tweets                     | 64 370    |
| Tokens                     | 2 568 160 |
| Target + Opinion Pairs     | 7 960     |
| Targets which are Pronouns | 886       |
| Pronouns                   | > 120 350 |

## 4.2 Evaluation Results

We implement our approach using java language and we conducted our experiments using the Twitter conversations of five electronics products: 2 digital cameras and 3 smartphones. As our work evaluation requires substantial human effort to identify product features and subjective reviews, we reduced the set of 8K conversations to just 4K. For each product, we extract the first 800 conversations. After preprocessing, our system is applied to perform opinion summarization. For evaluation, we manually read all the extracted conversations. For each tweet, if it involves user opinions, all the features on which the author has expressed his opinion are tagged. For each product, we manually produced a feature list. Column “Nbr of features” in Table 2 shows the number of manual features for each product. The features generated by our system are compared with the manually tagged results. In this subtask, we used precision and recall which are among the main evaluation measures employed in the feature selection task. In our case, TP is the number of relevant features identified, TP+TN represent the number of relevant features and TP+FP gives the number of features identified. Table 2 shows the evaluation results for the 5 products obtained in the feature identification phase. The highest values reached by our system are 77.81% precision and 82.62% recall. Our system presents high precision and recall scores. The recall value is higher than precision indicating that the majority of correct features were correctly recognized by the system. This can show the efficacy of the use of conversation interactions in extracting product features that have been commented by users. The precision value is lower than recall indicating that some identified features are not correct. This can be justified since most of the reviewers do not follow grammatical rules strictly while writing tweets due to which the parser fails to assign correct POS tag and thereby correct dependency relations between words. On analysis, we observed that, when a pronoun to be

**Table 2.** Evaluation results for the feature detection phase

| Product          | Nbr of features | Feature identification on individual tweets |              | Feature identification on conversations (this research) |              |
|------------------|-----------------|---|--------------|---|--------------|
|                  |                 | Precision(%)                                | Recall(%)    | Precision(%)  | Recall(%)    |
| Digital camera 1 | 27              | 68.33                                       | 62.13        | 75.26   | 79.61        |
| Digital camera 2 | 31              | 63.52                                       | <b>64.40</b> | 76.78   | 77.98        |
| Smartphone 1     | 52              | 55.97                                       | 51.83        | 75.22   | 78.23        |
| Smartphone 2     | 25              | <b>71.54</b>                                | 57.74        | 69.46   | <b>82.62</b> |
| Smartphone 3     | 43              | 57.88                                       | 53.73        | <b>77.81</b>  | 81.73        |
| <b>Average</b>   |                 | 63.44                                       | 57.96        | 74.90   | 80.03        |

resolved has more than two or three candidate antecedents, the occurrence of noisy anaphora-antecedent pairs increases. This leaves scope for enhancing our method for detecting features to reach better precision level. To further illustrate the effectiveness of our feature identification phase, we compared the features generated using our method with features found by the same method applied to the same set of tweets taken separately rather than extracting conversations and without the use of AR between messages. This method is closely similar to the feature detection process employed by [9] which extract product features from customer reviews based on product reviews collected via electronic commerce websites and Twitter. The average recall of opinion sentence extraction is nearly 63% while the average precision is around 57%. We observe that both the average recall and precision of the second method are significantly lower than those of our method. On analysis, we observed that, in testing dataset, a total of 13 824 anaphoric pronouns are present, in which 886 pronouns correctly refer to product features. By applying the second method, almost 60% of pronouns are left unnoticed or erroneously extracted. Comparing the results in Table 2, we can clearly see that the proposed method is much more effective for our task.

## 5 Conclusion

In this paper, we proposed a new method for opinion target identification that handles twitter conversations rather than single tweets. We employ conversation interactions, notably reply links, to effectively extract the target product features from customer reviews. Our experimental results indicate that the proposed method is very promising in performing its task. In particular, we have proved that incorporating conversation structure in the opinion target identification task contributes to improving system performance.

In our future work, we plan to further experiment it with other entities not only products. We will also look into employing our approach on other customer opinion resources rather than Twitter. We also intend to enhance the AR process.

## References

1. Bahrainian, S.A., Dengel, A.: Sentiment analysis and summarization of twitter data. In: Computational Science and Engineering (CSE), 2013 IEEE 16th Interna-



- tional Conference on. pp. 227–234. IEEE (2013)
2. Baldwin, B.: Cogniac: high precision coreference with limited knowledge and linguistic resources. In: Proceedings of a Workshop on Operational Factors in Practical, Robust Anaphora Resolution for Unrestricted Texts. pp. 38–45. Association for Computational Linguistics (1997)
  3. Belkaroui, R., Faiz, R., Elkhelifi, A.: Conversation analysis on social networking sites. In: Signal-Image Technology and Internet-Based Systems (SITIS), 2014 Tenth International Conference on. pp. 172–178. IEEE (2014)
  4. Bora, N.N.: Summarizing public opinions in tweets. *International Journal of Computational Linguistics and Applications* 3(1), 41–55 (2012)
  5. Feldman, R., Fresko, M., Goldenberg, J., Netzer, O., Ungar, L.: Extracting product comparisons from discussion boards. In: Data Mining, 2007. ICDM 2007. Seventh IEEE International Conference on. pp. 469–474. IEEE (2007)
  6. Hu, M., Liu, B.: Mining and summarizing customer reviews. In: Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining. pp. 168–177. ACM (2004)
  7. Jakob, N., Gurevych, I.: Using anaphora resolution to improve opinion target identification in movie reviews. In: Proceedings of the ACL 2010 Conference Short Papers. pp. 263–268. Association for Computational Linguistics (2010)
  8. Java, A., Song, X., Finin, T., Tseng, B.: Why we twitter: understanding microblogging usage and communities. In: Proceedings of the 9th WebKDD and 1st SNA-KDD. pp. 56–65. ACM (2007)
  9. Jmal, J., Faiz, R.: Customer review summarization approach using twitter and sentiwordnet. In: Proceedings of the 3rd International Conference on Web Intelligence, Mining and Semantics. p. 33. ACM (2013)
  10. Kamal, A., Abulaish, M.: Statistical features identification for sentiment analysis using machine learning techniques. In: Computational and Business Intelligence (ISCBI), 2013 International Symposium on. pp. 178–181. IEEE (2013)
  11. Kessler, J.S., Nicolov, N.: Targeting sentiment expressions through supervised ranking of linguistic configurations. In: ICWSM. AAAI Press (2009)
  12. Meng, X., Wei, F., Liu, X., Zhou, M., Li, S., Wang, H.: Entity-centric topic-oriented opinion summarization in twitter. In: Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining. pp. 379–387. ACM (2012)
  13. Popescu, A.M., Etzioni, O.: Extracting product features and opinions from reviews. In: Natural language processing and text mining, pp. 9–28. Springer (2007)
  14. Ritter, A., Cherry, C., Dolan, B.: Unsupervised modeling of twitter conversations. In: HLT-NAACL. pp. 172–180. The Association for Computational Linguistics (2010)
  15. Steinberger, J., Poesio, M., Kabadjov, M.A., Ježek, K.: Two uses of anaphora resolution in summarization. *Information Processing & Management* 43(6), 1663–1680 (2007)
  16. Stoyanov, V., Cardie, C.: Topic identification for fine-grained opinion analysis. In: Proceedings of the 22nd International Conference on Computational Linguistics-Volume 1. pp. 817–824. Association for Computational Linguistics (2008)
  17. Toh, Z., Su, J.: Nlangp: Supervised machine learning system for aspect category classification and opinion target extraction. In: Proceedings of the 9th International Workshop on SemEval. pp. 719–724 (2015)
  18. Turney, P.: Mining the web for synonyms: PMI-IR versus LSA on TOEFL. In: Machine Learning: ECML. pp. 491–502 (2001)