Using Social Conversational Context For Detecting Users Interactions on Microblogging Sites

Rami BELKAROUI*, Rim FAIZ**, Aymen ELKHLIFI***

* LARODEC, ISG Tunis, University of Tunis, Tunisia
rami.belkaroui@gmail.com
** LARODEC, IHEC, University of Carthage, Tunisia
rim.faiz@ihec.rnu.tn
***LALIC, Paris Sorbonne University, France
aymen.elkhlifi@paris4.sorbonne.fr

Abstract. In the current era, microblogging services like Twitter, gives people the ability to communicate, interact, collaborate with each other, reply to messages from others and create conversations. These services can be seen as very large information repository containing millions of text messages usually organized into complex networks involving users interacting with each other at specific times. Several works have proposed tools for tweets search focused only to retrieve relevant tweets. Therefore, users are unable to explore the results or retrieve more relevant tweets based on the content, and may get lost or become frustrated by the information overload.

In this paper, we propose a new method to retrieve conversation on microblogging sites particularly Twitter. It's based on content analysis and content enrichment. The goal of our method is to present a more informative result compared to conventional search engine. The proposed method has been implemented and evaluated by comparing it to Google and Twitter Search engines and we obtained very promising results.

1   Introduction

Last years, People are becoming more communicative through expansion of services and multi-platform applications such as blogs, forums and social networks which establishes social and collaborative backgrounds. This behavior leads to an accumulation of an enormous amount of information. Among these platforms are so-called microblogs. Furthermore, microblogging services (Boyd et al., 2010) gives people the ability to communicate, interact, collaborate with each other, reply to messages from others and create conversations. While communicating people share different kind of information like common knowledge, opinions, emotions, information resources and their likes or dislikes. The analysis of those communications can be useful for commercial applications such as trends monitoring, reputation management and news broadcasting. In addition, one of main characteristic of microblogging services is that users are not limited to produce contents; they can get involved indirectly in conversations with other

users by liking and sharing user's posts. Several works have proposed tools for tweets search focused only to retrieve relevant tweets. Therefore, users are unable to explore the results or retrieve more relevant tweets based on the content, and may get lost or become frustrated by the information overload. In addition, finding good results concerning the given subjects needs to consider the entire context. However, context can be derived from user interactions.

This paper proposed a conversation retrieval method which can be used to extract conversation from twitter. Comparing with current methods, the new proposed not only extract directly reply tweets, but also relevant tweets which might be retweets or comments and other possible interactions. The method extract extensive posts beyond conventional conversation. The rest of the document is organized as follows: we begin by presenting related work addressing conversation retrieval on microblogging sites. In section 3, we propose our method allows extracting user's content interactions. The experimentation and evaluation results are detailed in section 4. The final section presents a summary of our work and future dicrections.

2   Related Work

Conversation retrieval is a new search paradigm for microblogging sites. It results from the intersection of Information Retrieval and Social Network Analysis (SNA). Most of microblogging services provide a way to retrieve relevant information (Jabeur et al., 2012; Cherichi and Faiz, 2013), but lack the ability to provide all tweets discussion. There have been few previous researches dealing specifically with conversation detection. In addition, existing conversation retrieval approaches for microblogging sites (Cogan et al., 2012; Magnani et al., 2010, 2011, 2012) have so far focused on the particular case of a conversation formed by directly replying tweets. Magnani et al. (2011) proposed a user-based tree model for retrieving conversations from microblogs. They considered only tweets that directly respond to other tweets by the use of @sign as a marker of addressivity. The downside is that this method does not consider tweets that do not contain the @sign. Similarly (Cogan et al., 2012) proposed a method to build conversation graphs, formed by users replying to tweets. In this case, a tweet can only directly reply to other tweet. However, users can get involved indirectly in conversations communities by commenting, liking, sharing user's posts and other possible interactions. In (Kumar et al., 2010) the authors concentrated on different microblogging conversations aspects. They proposed a simple model that produces basic conversation structures taking into account the identities of each conversation member. Other related works (Huang et al., 2010) focusing on different aspects of microblogging conversation, that deal respectively with conversation tagging and topics identification.

3   Twitter Conversation Detection Method: TCOND

We propose a method which combines a set of conversational features and the directly exchanged text messages in order to extract extensive posts beyond conventional conversation. In addition, we defined a conversation as a set of short text messages posted by a user at specific timestamp on the same topic. These messages can be directly replied to other users by using "@username" or indirectly by liking, retweeting, commenting and other possible interactions. In the next part, we will present more details about our two approach steps.

3.1 Step 1: Constructing Direct Conversation

In this step, we aim to collect all tweets in reply directly to other tweets. Obviously, a reply to a user will always begin with "@username". Our goal in this step is to create reply tree. The reply tree construction process consists of two algorithms run in parallel recursive root finder algorithm and iterative search algorithm.

Algorithm 1 Recursive Root Finder (A: twitter)

Let T be a tweet collected from Twitter (ID tweet)
while (T_i !=root) **do**
    Extract T_i- 1 by matching field "in reply to status id"
end while
A : twitter = A : twitter – 1

Let T0 is the root (first tweet published) of the conversation C and T is a single tweet of the conversation retrieved. Let consider T_i the type of tweet T. A tweet can have three types: root, reply or retweet. The goal of the Recursive Root Finder Algorithm is to identify the conversation root T0 given T. Note that when the algorithm starts, $|T|$ is not known. Once, the conversation root T0 has been established, the Iterative Search Algorithm is used to seek the remainder of conversation C by searching all tweets addressed to T_i using matching field "in reply to status id". It is run repeatedly until some conditions, indicating that the conversation has ended, are met.

3.2 Step 2: Relevant Indirect Tweets

We define new features that may help to detect tweets related indirectly to a same conversation. The goal is to extract tweets that may be relevant to the conversation without the use of the @symbol. We use the following notations in the sequel:

- t_i is a set of tweets present in direct conversation (tweets in reply to other tweets directly).

- t_j is a tweet that can be linked indirectly to conversation.

The features we used are:

- ***Using the same URL***:

By sharing an URL, an author would enrichment the information published in his tweet. This feature is applied to collect tweets that share the same URL.

$$P1(t_i, t_j) = \begin{cases} 1 \text{ if t contains the same URL.} \\ 0 \text{ otherwise.} \end{cases} \tag{1}$$

- ***Hashtags Similarity***:

The # symbol, called hashtag, is used to mark a topic in a tweet or to follow conversation. Any user can categorize or follow topics with hashtags. We used this feature to collect tweets that share the same hashtags.

$$P2(t_i, t_j) = \begin{cases} 1 \text{ if t contains the same hashtag.} \\ 0 \text{ otherwise.} \end{cases} \quad (2)$$

- ***Tweets Time Difference***:

The time difference is highly important feature for detecting tweets linked indirectly to conversation. We use the time attribute to efficiently remove tweets having a large distance in terms of time compared to conversation root.

- ***Tweets Publication dates***:

Date attribute are highly important for detecting conversations. Users tend to post tweets about conversational topic within a short time period. The Euclidean distance has been used to calculate how similar two posts publication dates are.

- ***Content***:

We compute the textual similarity between each element in t_j, t_i taking the maximum value as the similarity measure between two messages. The similarity between two elements is calculated using the well-known tf-idf cosine similarity, $sim(t_i, t_j)$.

- ***Similarity Function***:

Finally, the similarity between tweets indirectly linked to conversation and tweets which are present in the reply tree is calculated by a linear combination between their attributes.

4   Experiments and Results

The following experiment has been designed to gather some knowledge on the impact of our results on end-users. For this experiment we have selected three events and queried our dataset using Google [1], Twitter search engine [2] and our method (TCOND). Then, we have asked a set of 100 assessors to rate the top-10 results of every search task with three relevance levels, namely highly relevant (value equal to 2), relevant (value equal to 1) or irrelevant (value equal to 0). In order to measure the results quality, we use the Normalized Discounted Cumulative Gain (NDCG) at 10 for all the judged events. In addition, we used a second metric which is the Precision at top 10.

The dataset has been obtained by monitoring microblogging system Twitter posts over the period of July-August 2013. In particular, we used a sample of about 113 000 posts containing trending topic keywords using Twitter's streaming API. Trending topics have been determined directly by Twitter, and we have selected the most frequent ones during the monitoring period.

---

1. www.google.com
2. Search.twitter.com.

4.1 Experimental Outcomes and Interpretation Results

We compare our conversation retrieval method with the results returned by Google and by Twitter search engine using two metrics namely the P@10 and the NDCG@10. From this comparison, we obtained the values summarized in Table 1 where we notice that our method overcomes the results given by both of Google and Twitter. The reason of these promising values is the fact that we combine a set of conversational features and direct replies method to retrieve conversation may have a significant impact on the users' evaluation.

	P@10 (Average%)	NDCG (Average%)
Task1		
Google	59.62	56.86
Twitter	65.73	59.71
TCOND	**73.28**	**64.52**
Task2		
Google	57.31	56.02
Twitter	62.78	58.45
TCOND	**67.27**	**62.73**
Task3		
Google	63.21	66.52
Twitter	65.88	68.46
TCOND	**77.27**	**69.33**

TAB. 1 – *Table of Values for Computing our Worked Example*

Focusing on the three messages selections, we observe that all conversations obtained with our method receive higher scores with compared to Google and Twitter's selection. According to the free comments of some users and following the qualitative analysis of the posts in the three selections we can see that Google and twitter received lower scores not because they contained posts judged as less interesting, but because some posts were considered not relevant with regard to the searched topic.

Concentration on the three messages selections we observe that all conversations selections obtained with twitter search has higher scores with respect to Google's selection. These results lead us toward a more general interpretation of the collected data. It appears that the social metrics usage have a significant impact on the users' degree interest in the retrieved posts. In addition, the retrieving conversations process from Social Network differs from traditional Web information retrieval; it involves human communication aspects, like the degree interest in the conversation explicitly or implicitly expressed by the interacting people.

5 Conclusion

This work explored a new method for detecting conversation on microblogging sites: an information retrieval activity exploiting a set of conversational features in addition to the directly exchanged text messages to retrieve conversation. Our experimental results have highlighted many interesting points. First, including social features and the concept of direct conversation

in the search function improves the relevance of tweets informativeness and also provides results that are considered more satisfaction with respect to a traditional tweet search task. Future work will further research the conversational aspects by including human communication aspects, like the degree of interest in the conversation and their influence/popularity by gathering data from multiple sources from Social Networks in real time.

References

Boyd, D., S. Golder, and G. Lotan (2010). Tweet, tweet, retweet: Conversational aspects of retweeting on twitter. In *Proceedings of the 2010 43rd Hawaii International Conference on System Sciences*, HICSS '10, Washington, DC, USA, pp. 1–10. IEEE Computer Society.

Cherichi, S. and R. Faiz (2013). New metric measure for the improvement of search results in microblogs. In *Proceedings of the 3rd International Conference on Web Intelligence, Mining and Semantics*, WIMS '13, New York, NY, USA, pp. 24:1–24:7. ACM.

Cogan, P., M. Andrews, M. Bradonjic, W. S. Kennedy, A. Sala, and G. Tucci (2012). Reconstruction and analysis of twitter conversation graphs. In *Proceedings of the First ACM International Workshop on Hot Topics on Interdisciplinary Social Networks Research*, HotSocial '12, New York, NY, USA, pp. 25–31. ACM.

Huang, J., K. M. Thornton, and E. N. Efthimiadis (2010). Conversational tagging in twitter. In *Proceedings of the 21st ACM conference on Hypertext and hypermedia*, HT '10, New York, NY, USA, pp. 173–178. ACM.

Jabeur, L. B., L. Tamine, and M. Boughanem (2012). Uprising microblogs: A bayesian network retrieval model for tweet search. In *Proceedings of the 27th Annual ACM Symposium on Applied Computing*, New York, NY, USA, pp. 943–948. ACM.

Kumar, R., M. Mahdian, and M. McGlohon (2010). Dynamics of conversations. In *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD '10, New York, NY, USA, pp. 553–562. ACM.

Magnani, M., D. Montesi, G. Nunziante, and L. Rossi (2011). Conversation retrieval from twitter. In *Proceedings of the 33rd European conference on Advances in information retrieval*, ECIR'11, Berlin, Heidelberg, pp. 780–783. Springer-Verlag.

Magnani, M., D. Montesi, and L. Rossi (2010). Information propagation analysis in a social network site. In N. Memon and R. Alhajj (Eds.), *ASONAM*, pp. 296–300. IEEE C.S.

Magnani, M., D. Montesi, and L. Rossi (2012). Conversation retrieval for microblogging sites. In *Information.Retrieval Journal*, Volume 15, pp. 354–372. Springer Netherlands.

Résumé

Dans ce travail, nous proposons une nouvelle méthode de détection des conversations sur les sites des réseaux sociaux. Cette méthode est basée sur l'analyse et l'enrichissement de contenu dans le but de présenter un résultat informatif basé sur les interactions des utilisateurs. Nous avons évalué notre méthode sur corpus recueillis de réseau social lié à des sujets spécifiques, et nous avons obtenu des bons résultats.