Towards Events Tweet Contextualization Using Social Influence Model and Users Conversations

Rami Belkaroui
LARODEC, ISG Tunis
University of Tunis
Bardo,Tunisia
rami.belkaroui@gmail.com

Rim Faiz
LARODEC, IHEC Carthage
University of Carthage
Carthage Presidency, Tunisia
rim.faiz@ihec.rnu.tn

ABSTRACT

Nowadays, microblogging sites have completely changed the manner in which people communicate and share information. They are among the most relevant source of knowledge where information is created, exchanged and transformed, as witnessed by the important number of their users and their activities during events or campaigns like the terror attack in Paris in 2015. On Twitter, users post messages (called tweets) in real time about events, natural disasters, news, etc. Tweets are short messages that do not exceed 140 characters. Due to this limitation, an individual tweet it's rarely self-content. However, user cannot effectively understand or consume information.

In order, to make tweet understandable to a reader, it is therefore necessary to know their context. In fact, on Twitter, context can be derived from users interactions, content streams and friendship. Given that there are rich user interactions on Twitter. In this paper, we propose an approach for tweet contextualization task which combines different types of signals from social users interactions to provide automatically information that explains the tweet. In addition, our approach aims to help users to satisfy any contextual information need. To evaluate our approach, we construct a reference summary by asking assessors to manually select the most informative tweets as a summary. Our experimental results based on this editorial data set offers interesting results and help ensure that context summaries contain adequate correlating information with the given tweet.

Categories and Subject Descriptors

H.3 [**Information Systems**]: Information Storage and Retrieval; H.3.3 [**Information Search and Retrieval**]: Retrieval models, Information filtering, Selection process

General Terms

Algorithms, Experimentation

WIMS '15, July 13 - 15, 2015, Larnaca, Cyprus

http://dx.doi.org/10.1145/2797115.2797134 ...$15.00.

Keywords

Events Tweet contextualization, Tweet understanding, Twitter Context Tree, Tweet influence, user conversations.

1. INTRODUCTION

Recent years have revealed the accession of interactive media, which gave birth to a huge volume of data produced by users called User Generated Content(UGC) in blogs and microblogs services more precisely. These Microblogging services, attract more and more users due to the ease and the speed of information sharing especially in real time.

Everyday, people are posting millions of updates under the form of short text messages on social networks, such as Twitter and Facebook. Some of these status updates describe events that people are participating in or watching through a media source such as television, including natural disasters [32], political debates [31], and sporting events [18]. The size of these statuts updates may be limited by a maximum number of characters. This limitation causes the use of a particular vocabulary that is often unusual, noisy and full of new words [7]. Indeed, the goal is to share the maximum amount of information in as few characters as possible [22]. It may thus be difficult to understand the meaning of a short text message without knowing the general context of its realization. This constraint problem is, for example, a frequent case on Twitter microblogging platform.

Twitter's data flow is examined in order to measure public sentiment [16], earthquake warning [25], reputation management, follow political activity and news [30]. According to Kwak et al., more than 85% of tweets are related to news [19]. In addition, when a news-worthy event occurs in the real world, many tweets are sent describing important information and expressing opinions about what is happening in real-time. However, an individual tweet is short and without sufficient contextual information, which makes it difficult to capture the associated information. For example, a tweet posted by Darrell (one of the most popular twitter user) just contains a single hashtag "#PrayForTunisia" during the terrorist museum attacks in Tunisia in 2015. When reading this tweet, without knowing the related news, it would be very difficult to understand this tweet topic (what is this tweet about? what happened ?). Furthermore, tweets may contain information that is not understandable to user without some context. All these obstacles impede users from effectively understanding or consuming information, which can either make users less engaged or even unfastened from using Twitter.

In this paper, we describes an approach that exploits social Twitter conversations to provide some context for a given tweet by selecting relevant tweets within a conversation of an original post, in order to help users effectively understand the tweet context. Typically, we focus, in our case on exploiting multiple different types of signals such as social metadata signals (Hashtags, URL), user-tweet influence signals, temporal signals and text based signals, which can be potentially useful to improve tweet contextualization task. The research question addressed in this paper are the following:

- What is the impact of exploring social Twitter conversations for tweet contextualization task?
- Can these signals help tweet contextualization task for guiding its users to effectively understand the contextual information?

The remainder of this paper is organized as follows: we begin by explaining the CLEF-INEX tweet contextualization task. In section 3, we describe some related works presented in INEX 2012, 2013 and 2014, and then analyze characteristics of Twitter conversations trees in Section 4. In section 5, we present our approach that exploits social Twitter conversations to provide some context for a given tweet, and introduce user-tweet influence model in section 6. In section 7, we explain different types of signals used in our work. Our experimental results are presented in Section 8. Finally, we present a summary of our work and some future dicrections.

2. TWEET CONTEXTUALIZATION TASK

In the context of Twitter microblogging service, contextualization is specifically important since 140 characters long messages are short and rarely self-content. This motivated the proposal in 2011 of a new track at Clef INEX lab of Tweet Contextualization. Furthermore, tweet contextualization task as defined by The Initiative for the Evaluation of XML Retrieval (INEX) [10] is "a readable text that provides some context (summary) about the tweet subject (query), in order to help the reader to effectively understand it, i.e., answering questions of the form What is this tweet about?" This requires combining multiple types of processing from Information Retrieval (IR) to multidocument summarization". The summary does not exceed 500 words and it is extracted from a cleaned dump of the English Wikipedia[1].

Text contextualization [10, 28] differs from text expansion in that it aims at helping a human to understand a text rather than a system to better perform its task. For example, in the case of query expansion in IR, the idea is to add terms to the initial query that will help the system to better select the documents to be retrieved [2, 20]. On the contrary text contextualization can be viewed as a way to provide more information on the corresponding text in the aim of making it understandable and relating it to information that explains it.

3. RELATED WORK

In this section, we report related work exploiting tweet contextualization task. Moreover, there have been some studies done for this task.

In [11], the authors proposed a new method based on the local Wikipedia dump. They used TF-IDF cosine similarity measure enriched by smoothing from local context, named entity recognition and part-of-speech weighting presented at INEX 2011. They modified this method by adding bigram similarity, anaphora resolution, hashtag processing and sentence reordering at INEX 2012 [12]. Recently, authors [14] modified the method presented at INEX 2011, 2012 and 2013 [13] underlain by the product of different measures based on smoothing from local context, named entity recognition, part-of speech weighting and sentence quality analysis. Besides, they examined the influence of topic-comment relationship on contextualization. The proposed approach in [6] described a hybrid tweet contextualization system using Information Retreival (IR) and Automatic Summarization (AS). They used nutch architecture and TF-IDF based sentence ranking and sentence extracting techniques for Automatic Summarization. In the same way, Ansary et al. [1] described a pipeline system where first extracted phrases from tweets by using ArkTweet toolkit and some heuristics; then retrieved relevant documents from Wikipedia before summarizing those with MEAD toolkit.

In [34], the authors developed a statistical word stemmer which used by the CORTEX to preprocess input texts and generate readable summary. Recently, they presented three statistical summarizer systems to build tweet context applied to CLEF-INEX 2014 task [33]. The first one is Cortex summarizer based on the fusion process of several sentence selection metrics and an optimal decision module to score sentences from a document source. The second one is Artex summarizer uses a simple inner product among the topic-vector and the pseudo-word vector and the third is a performant graph-based summarizer. While, in [9], the authors used a method that allows to automatically contextualize tweets by using information coming from Wikipedia. they treat the problem of tweets contextualization as an Automatic Summarization task, where the text to resume is composed of Wikipedia articles that discuss the various pieces of information appearing in a tweet, whereas, in [24] the authors combined Information Retrieval, Automatic Summarization and Topic Modeling techniques to provide context of each tweet. They took advantage of a larger use of hashtags in topics and used them to enhance the retrieval of relevant Wikipedia articles. In [3], the authors have simply treated contextualization as a passage retrieval task. They used the textual tweet content as a query to retrieve paragraphs or sentences from the Wikipedia corpus. Another approach proposed by [21] used latent Dirichlet analysis (LDA) to obtain a tweet representation in a thematic space. This representation allows finding a set of latent topics covered by the tweet. Lately, [37] described a new method for tweet contextualization based on association rules between sets of terms. This approach allows the extension of tweets vocabulary by a set of thematically related words. In [17], the proposed approach described a sentence retrieval technique to construct answer fragments for each tweet. Three different sentence selection methodologies were used: i) language modeling score, ii) relevance modeling score and ii) topical relevance modeling score.

Traditional contextualization techniques only consider text information which is insufficient for the tweet contextualization task, since text information on twitter is very sparse. In addition, tweets are short and not always written maintaining a formal grammar and proper spelling. However, exploiting social conversations to provide some context for a

[1]https://fr.wikipedia.org/

given tweet, in order to help users effectively understand the tweet context which is the main contribution of this paper has not been researched yet.

4. TWITTER CONVERSATIONS TREE ANALYSIS

In our task, a Twitter context tree (also called Twitter conversations in [5]) is constructed by a set of tweets posted by a user at specific timestamp on the same topic. These tweets can be directly replied to other users by using "@username" or indirectly by liking (favorite), retweeting, commenting and other possible interactions. In this section, we conduct further analysis on the Twitter conversations trees with respect to temporal growth and depth distributions.

4.1 Temporal Growth Analysis

We first analyze the temporal growth of the Twitter conversation tree in Figure 1, where y-axis is the number of tweets and x-axis is the relative temporal distance from the original tweets, measured by hours.

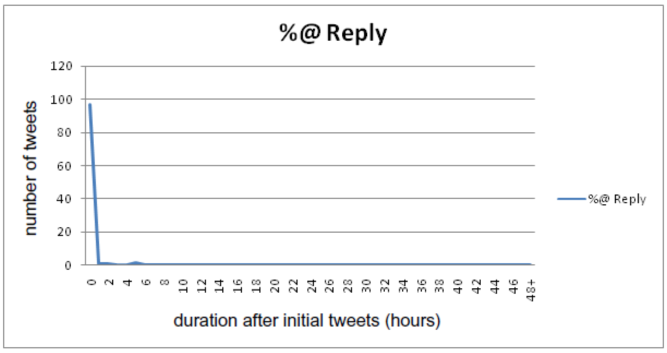

Figure 1: Number of Tweets over Time

Given that Twitter is a real-time service, overall, about 97.87% of replies are generated within the first hour, while an additional 0.98% of replies happen in the second hour, which shows that Twitter can propagate information quite fast and a meaningful context tree can be formed very quickly. Consequently, the temporal growth of the context tree prove the importance of exploiting twitter conversations in our approach for tweet contextualization task.

Interestingly, sometimes after news events such as earthquakes or other natural disasters, Wikipedia informations are not immediately available. Even, after a few hours, the available content is often inaccurate or highly redundant. For example: the attack against french newspaper Charlie Hebdo, there were no articles on Wikipedia describing the topic #jesuischarlie. Indeed, the first article that explains this event was available 7 hours after the terrorist attack. While, the twitter context tree was launched at 11:52h (less than an hour after the attack). At the same time, these events demonstrates a scenario where users urgently need information, especially if they are directly affected by the event. So, unexpected news events such as earthquakes represent information access problem where the approaches using Wikipedia to contextualize a tweet fail.

4.2 Depth Distributions Analysis

The next question is whether the tree structure can help contextualization task. Figure 2 shows the cumulative distribution of the number of tweets over depth in context trees, assuming the root of each tree has the depth of 0. Surprisingly, the structures of these Twitter context trees are highly skewed, and more than 80% of tweets are at depth 1. In addition, 10.7% received two levels depth (a reply to the original conversation reply). Only 1.53% of Twitter conversations are three levels depth after the original tweet (there is a reply, reply to the reply, and reply to the reply of reply).

This means that most tweets reply to their corresponding root tweets directly, and the conversation root can have several reply lead to an informative content.

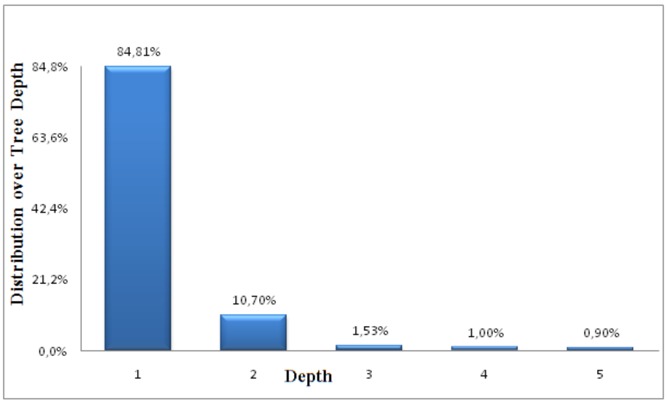

Figure 2: Distributions over Tree Depth

5. ETCA: EVENTS TWEET CONTEXTUALIZATION APPROACH

On Twitter, many users post freely tweets in order to express what they are thinking about any event or topic, followed by some comments, retweet or favorite. These posts can be written with a particular vocabulary that is often unusual, noise, which makes it difficult to detect the meaning of these messages. For example, posts such as "#end7 Check this out.reut.rs/XZv3L1" are even for humans difficult to understand without knowing the context. However, users' interactions essentially reflect the importance of different tweets and can be used to improve the quality of tweet contextualization.

The key idea of our approach, called ETCA, as depicted in Figure 3, is to extract the Twitter messages context by selecting relevant tweets within a conversation of an original post can be applied to clarify the meaning of tweets. Indeed, we defined a social tweet context as follow:

Social tweet context

Given a tweet t its social context C_t is defined as $<I_t, U_t>$ where I_t is a set of interactions (comment, retweet,...) on t written by users U_t in a social network.

6. SOCIAL INFLUENCE GENERATION BASED USER-TWEET INTERACTION MODEL

On Twitter, there are several interactional relationships between users and tweets, such as post, reply, follow, fa-

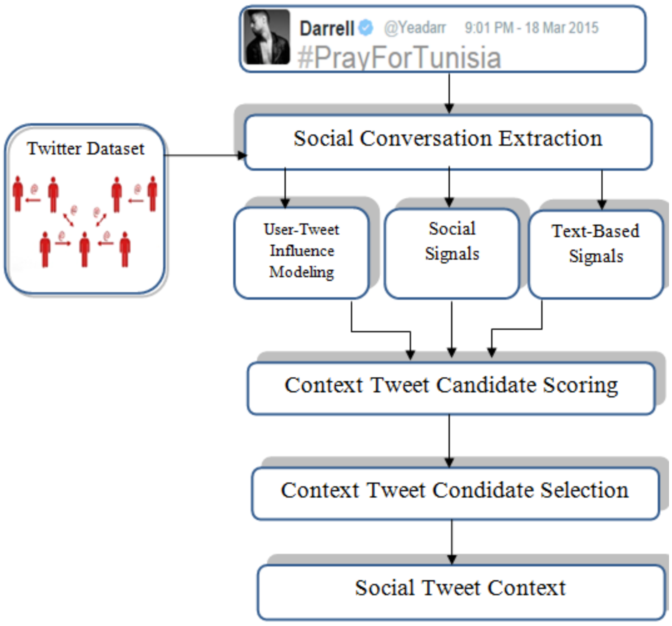

Figure 3: Overview of Our Proposed Approach

vorite, mention and retweet. We take these relationships into account for measuring the tweet influence score in order to select context candidate tweets. The motivations using the tweet influence are:

- If we know that the user A has a strong influence on a user B within the same conversation, in this case, when A publish a tweet (conversation root) and causes a big twitter conversation, those tweets in the conversation published by B are more likely to be a context candidate tweet.
- If a tweet published in the same conversation has been replied, favorite and retweeted by many users, a natural assumption is that this tweet is most likely to influence all those other tweets and be a context candidate tweet.

6.1 User-Tweet Interaction Model

To construct a user-tweet graph, we define a user-tweet schema graph, as illustrated in Figure 2 similar to the graph in [36]. A user-tweet schema graph is a directed graph G = (V, E). V is a set of nodes which are of two kinds. Let $V_t = \{t_1, t_2, \ldots, t_m\}$ be the set of tweet nodes representing tweets and $V_u = \{u_1, u_2, ..., u_n\}$ be the set of user nodes representing users. $V = V_t \cup V_u$ nodes. E is the edge set consisting of post, reply, follow, retweet, mention and favorite edges.

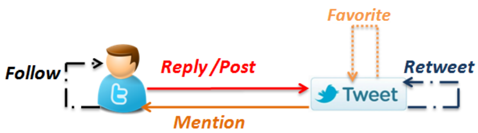

Figure 4: User-Tweet Schema Graph

A reply edge is from a user u to a tweet t posted by u. A follow edge is from a user u to another user who follows u. A retweet edge is from a tweet t to another tweet which retweets t. A mention edge is from a tweet t to a user u who comments t. A favorite edge is from a tweet t to another tweet which favorite's t. A user-tweet interaction model is shown in Figure 5. In the case of user-tweet interaction model, it consists of user nodes, tweets nodes and six kinds of edges.

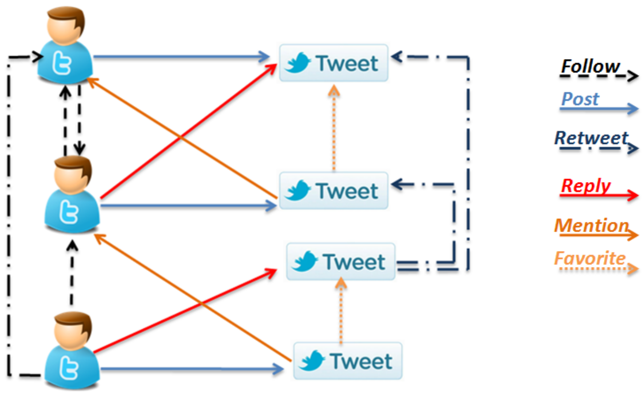

Figure 5: User-Tweet Interaction Model

6.2 Social Influence Measuring

In Twitter microblog, the tweet of a user who has more followers always draw more attention, so they are evidently exists a correlation between tweet characteristics influence and tweet's author influence. We exploit two types of score for social influence measuring:

- Measuring tweet influence score refers to those features which represent the particular characteristics of tweet such as reply influence.
- Measuring tweet's author influence score refers to those features which represent the influence of tweet's author such as follow influence.

6.2.1 Tweet Influence Measuring

The tweet influence is calculated from reply influence, retweet influence and favorite influence.

- **Reply Influence**

  When a user replied to tweet, it means she/he has taken time to react to the posted content. She/he is reacting to what this user tweeted and is most likely sharing her personal opinion in published content.

  Reply influence score(t): The action here is replying. The more replies a tweet receives, the more influential it is. This influence can be quantified by the number of replies the tweet receives. The reply influence is defined as follow:

$$\mathrm{ReplyInfluence(t)} = \alpha \times \mathrm{NumberReply(t)}. \tag{1}$$

$\alpha \in (0, 1]$. It is adjustable and indicates the weight of reply edge.

.

- **Retweet Influence**

  Generally, a user retweets a tweet if it appears to contain useful information, because he/she wants to share it with his/her followers.

  Retweet influence score(t): The action here is retweeting. The more frequently user's messages are retweeted by others, the more influential it is. This can also be quantified by the number of retweets. It is defined as follow:

$$\mathrm{RetweetInfluence(t)} = \beta \times \mathrm{NumberRetweet(t)}. \tag{2}$$

  $\beta \in (0,\ 1]$. It is adjustable and indicates the weight of retweet edge.

  .

- **Favorite Influence**

  Favorites are described as indicators that a tweet is well-liked or popular among online users. A tweet can be identified as a favorite by the small star icon seen beside the post. When a user mark tweets as favorites, she/he can easily find useful and relevant information. In addition, she/he can also spark the interest of other online users to start a conversation or comment on the tweet.

  Favorite influence score(t): The action here is favoriting. The more favorites a tweet receives, the more influential it is. This influence can be quantified by the number of favorite the tweet receives. It is defined as follows:

$$\mathrm{FavoriteInfluence(t)} = \gamma \times \mathrm{NumberFavorite(t)}. \tag{3}$$

  $\gamma \in (0,\ 1]$. It is adjustable indicates the weight of favorite edge.

  .

According to the experience, α is bigger than β and γ, it means that the users who reply on tweet t are more interested in it than others who only retweet or favorite it.

It is obvious that microblogging users mainly focus on the current tweets. However, temporal aspect can also provide valuable information for tweet contextualization due to the real-time characteristics of Twitter. Therefore, the tweet timestamp plays an important role on the tweet influence .i.e. a recent tweet has larger chance to have bigger influences compared to old published tweet. So to cope with, we use Gaussian Kernel [23] to estimate a distanceΔt between tweet conversation root time d and other tweet time d' within the same conversation, i.e., $\Delta \mathrm{t} = \left|\mathrm{d}' - \mathrm{d}\right|$. It is defined as follow:

$$\Gamma(\Delta \mathrm{t}) = \exp\left\lfloor \frac{-\Delta \mathrm{t}^2}{2\sigma^2} \right\rfloor \mathrm{with}\sigma \in \mathbb{R}+. \tag{4}$$

Finally, the tweet influence score is defined as follows:

$$\begin{aligned}\mathrm{TweetInfluence(t)} &= \Gamma(\Delta \mathrm{t}) \times \mathrm{ReplyInfluence(t)} \\ &+ \mathrm{RetweetInfluence(t)} + \mathrm{FavoriteInfluence(t)}.\end{aligned} \tag{5}$$

6.2.2 *Tweet's Author Influence Measuring*

People may be particularly interested in tweets from celebrities or opinion leaders. Indeed, if there are two tweets stating the same information, we assume that the tweet published by an influential user is more important than the other tweet, on account of two interesting observations about Twitter users. First, users with a high influence have a larger audience. In addition, their tweets are apt to be read by more users than those of non-influential users. Second, encouraged by the interactions with their followers, influential users are more likely to publish informative tweets of better readability, less error, and preferable completeness than common users. In this part, the tweet's author influence consists of follow influence and mention influence.

- **Mention Influence** which we measure through the number of mentions containing one's name, indicates the ability of that user to engage others in a conversation. The mention influence score is defined as follows:

$$\mathrm{MentionInfluence(u)} = \delta \times \mathrm{NumberMention(u)}. \tag{6}$$

  $\delta \in (0,\ 1]$. It indicates the weight of mention edge.

.

- **Follow Influence**

A user followed by many users is likely to be an authoritative user and their post is also likely to be useful. In addition, the followers number of a user directly indicates the audience size for that user. The follow influence score is defined as follows:

$$\mathrm{FollowInfluence(u)} = \omega \times \mathrm{NumberFollow(u)}. \tag{7}$$

$\omega \in (0,\ 1]$. It indicates the weight of follow edge.

.

Finally, the tweet's author influence score is defined as follows:

$$\begin{aligned}\mathrm{TweetAuthorInfluence(u)} &= \mathrm{MentionInfluence(u)} \\ &+ \mathrm{FollowInfluence(u)}\end{aligned} \tag{8}$$

7. CONTEXT CANDIDATES TWEETS EXTRACTION

Besides user-tweet interaction model we also included text based signals and social signals.

7.1 Text-based signals

In this section, we assign score to a candidate tweet based on the similarity between different tweets in the whole conversation. Therefore, From each tweet t in a conversation C, we derive a vector $\vec{\mathrm{V}}$ using the vector space model [26]. Thus, the set of conversation is viewed as a set of vector.

- **Similarity to tweet root**

We used cosine similarity to measure the similarity between the tweet root vector $\vec{V_{t_{root}}}$ and other tweets vector $\vec{V_t}$ within the same conversation. In addition, we aim to measure how much a tweet would be related to tweet root's content.

$$\text{cosine}(\vec{V_t}, \vec{V_{t_{root}}}) = \frac{\vec{V_t}.\vec{V_{t_{root}}}}{||\vec{V_t}||.||\vec{V_{t_{root}}}||} \tag{9}$$

- **Similarity of content [8]**

In our case, it measures how many tweet of the whole conversation C are similar in content with current tweet $\vec{t_{current}}$. We calculate cosine similarity score for every pair of tweets. The similarity is calculated using Lucene similarity function [2]. We denote current tweet modeled as a vector:

$$\text{cosine}(\vec{t_{current}}, C) = \frac{\sum_{\vec{t_{current}} \neq \vec{t'}} \text{sim}(\vec{t_{current}}, \vec{t'})}{|C| - 1} \tag{10}$$

7.2 Social metadata Signals

- **Context Candidate Tweets Regarding the URLs**

  By sharing an URL, an author would enrich the information published in his tweet. When a URL is present in the tweet root, we download the page and extract its title as well as the body content. For each candidate tweet t we computed:

  - The word overlap between candidate tweet t and the title of the web page, and between t and the body content of the web page.
  - The cosine similarity between t and the title of the web page, and between t and the body content of the web page.

- **Context Candidate Tweets Regarding the Hashtags**

  The # symbol, called hashtags in tweets is very important pieces of information, since they are tags that were generated by the user. Hashtags is used to mark a topic in a tweet or to follow conversation. In addition, publishers can use hashtags to provide implicit context of the tweet. We used this feature to collect candidate tweets that share the same root tweet hashtags.

$$\text{F1}(t, t_{root}) = \begin{cases} 1 \text{ if t contains the same hashtag.} \\ 0 \text{ otherwise.} \end{cases} \tag{11}$$

7.3 Supervised Learning Framework

Given the above signals, we could convert them as features, then cast the Twitter context summarization task into a supervised learning problem. After training a model, we could predict a few tweets as its summary for all tweets in a new context tree. In this paper, we choose Gradient Boosted Decision Tree (GBDT) algorithm [15] to learn a non-linear model. GBDT is an additive regression algorithm consisting of an ensemble of trees, fitted to current residuals, gradients of the loss function, in a forward step-wise manner.

[2] http://lucene.apache.org/core/3_6_1/scoring.html

8. EXPERIMENTS AND RESULTS

In order to answer the research questions introduced in Section 1, we have designed experiments. In this section we detail our experimental setup, including how we sample tweets, the Twitter conversations dataset, reference summary, and how we evaluate relating individual tweets to their contexts.

8.1 Tweets

The tweets dataset has been obtained by monitoring Twitter microblogging system over the period of January-March 2015. In particular, we used a sample of about 2000 posts using Twitter's streaming API. So, a significant portion of all tweets are non informative and only a small fraction contains topics of general interest.

8.2 Twitter Conversations Dataset

As celebrities are highly influential in Twitter [35], celebrities initiated tweets would lead to large context trees. We extract 50 Twitter context trees from January 7th to March 22th, 2015, using our conversations trees detection system [4] to construct a data set in our work. These 50 context trees are initiated by many celbrities as like as Lady Gaga, who is the most popular elite user on Twitter, Manuel Valls, Olivia Wilde, J. k. Rowling, Norman Thavaud, Francois Hollande, etc. Furthermore, 26 out of 50 context trees are about the terrorist attack on charlie Hebdo, another 16 context trees are related to the Tunisian museum terror attack, while the remaining 8 are about different topics.

8.3 Reference Summary

To the best of our knowledge, there is no data set available to evaluate social tweet contextualization. Thus, we conduct a pilot study to construct such an editorial data set in our work. The goal of this study is to construct a reference summary generating by humans which can be useful to evaluate our results. Thus, we only focus on 15 context trees about three different topics, but ask 10 assessors for judgments for every context tree. In addition, we ask each assessor to first read the root tweet and open any URL inside to have a sense of what the root tweet is about. Then, the assessor reads through all contexts candidate tweets to get a sense of the overall set of data. Thus, for each context tree, we will have 10 independent judgments. Finally, the assessor selects 5 to 10 tweets ordered sequentially as the summary, which respond or extend the original tweets by providing extra information about it.

8.4 Evaluation Metrics

Tweet contextualization is evaluated on both informativeness and readability [27]. Informativeness aims at measuring how well the summary explains the tweet or how well the summary helps a user to understand the tweet content. On the other hand, readability aims at measuring how clear and easy is to undersatnad the summary.

- **Informativeness**: The objective of this metric is to evaluate the selection of relevant tweets. Informativeness aims at measuring how well the summary helps a user to understand the tweet content. Therefore, for each tweet root, each candidate tweet will be evaluated independently from the others, even in the same summary.

	Unigrams	Bigrams	Skipgrams
Topic1			
Human Summary	0.7263	0.8534	0.9213
Our Proposed Summary	**0.7909**	**0.8865**	**0.9355**
Topic2			
Human Summary	0.7932	0.9137	0.9361
Our Proposed Summary	**0.8105**	**0.9408**	**0.9592**
Topic3			
Human Summary	0.7786	0.9172	0.9426
Our Proposed Summary	**0.8272**	**0.9438**	**0.9617**

Table 1: Table of Informativeness Results

	Relevance	Non Redundancy	AVG
Topic1			
Human Summary	88.65%	66.33%	77.49%
Our Proposed Summary	**89.72%**	**69.78%**	**79.75%**
Topic2			
Human Summary	90.72%	65.82%	78.27%
Our Proposed Summary	**91.03%**	**67.49%**	**79.26%**
Topic3			
Human Summary	90.23%	69.06%	79.64%
Our Proposed Summary	**92.24%**	**69.72%**	**80.98%**

Table 2: Table of Readability Results

The 10 best tweets summary for each tweet root are selected for evaluation. This choice is made based on the score assigned by the automatic system tweets contextualization (high scores).

The dissimilarity between a human selected summary (constructed using a pilot study) and the proposed summary (using our approach) is given by:

$$\mathrm{Dis}(\mathrm{T},\mathrm{S}) = \sum_{t\in \mathrm{T}} (\mathrm{P}-1) \times \left(1 - \frac{\min(\log(\mathrm{P}),\log(\mathrm{Q}))}{\max(\log(\mathrm{P}),\log(\mathrm{Q}))}\right) \tag{12}$$

where $\mathrm{P} = \frac{f_T(t)}{f_T} + 1$ and $\mathrm{Q} = \frac{f_S(t)}{f_S} + 1$.

T is the set of terms presented in reference summary. For each term $t \in T$, $f_T(t)$ represents the frequency of occurrence of t in reference summary and $f_S(t)$ its frequency of occurrence in the proposed summary. More Dis (T,S) is low, more the proposed summarry is similar to the reference. T may take three distinct forms:

– Unigrams made of single lemmas.

– Bigrams made of pairs of consecutive lemmas (in the same sentence).

– Bigrams with 2-gaps as well as the bigram, but can be separated by two lemmas.

Our results in the informativeness evaluation presented in Table 1 .

- **Readability**: readability aims at measuring how clear and easy it is to understand the summary. By contrast, readability is evaluated manually (cf.Table 2). Each summary has been evaluated by considering the following two parameters [29]:

  – Relevance: judge if the tweets make sense in their context (i.e. after reading the other tweets in the same context). Each assessor had to evaluate relevance with three levels, namely highly relevant (value equal to 2), relevant (value equal to 1) or irrelevant (value equal to 0).

  – Non-Redundancy: evaluates the ability of the context does not contain too much redundant information, i.e. information that has already been given in a previous tweets. Each assessor had to evaluate redundancy with three levels, namely not redundancy (value equal to 2), redundancy (value equal to 1) or highly redundancy (value equal to 0).

8.5 Experimental Outcomes and Interpretation Results

A good summary should have good quality but with less redundancy. The obtained informativeness (cf.Table 1) evaluation results shed light that our proposed approach offers interesting results and help that context summaries contain adequate correlating information with the root tweet. In addition, based on the editorial data set, our experimental results show that user influence information is very helpful to generate a high quality summary for each Twitter context tree.

Furthermore, our tweets contextualization approach based on social Twitter conversations leads to the improvement of the context informativeness ; we note also that the selection of tweets impacts the quality of the contexts. The contexts are less readable; it may be that they contain some noises which need to be cleaner.

9. CONCLUSION

we explored in this paper the problem of the tweet contextualization. We proposed an approach for twitter contextualization task that combined different types of signals from social user interactions and exploited a set of conversational features, which help users to get more context information when using Twitter.

Traditional contextualization methods only consider text information and we focused on exploiting multiple types of signals such as social signals, user-tweet interaction model, social influence and text based signals. All signals are converted into features, and we throw tweet contextualization into a supervised learning problem. Our approach was evaluated by using an editorial data set in which 10 assessors are employed to generate a reference summary for each context tree.

Future work will further research the conversational aspects by including human communication aspects, like the degree of interest in the conversation by gathering data from multiple sources such as comments on news articles, comments on Facebook pages and web pages. In addition, we will also exploit other social metadata such as expertise of conversation participants (authors).

10. REFERENCES

[1] K. H. Ansary, A. T. Tran, and N. K. Tran. A pipeline tweet contextualization system at INEX 2013. In *Working Notes for CLEF 2013 Conference , Valencia, Spain, September 23-26*, 2013.

[2] A. Bandyopadhyay, M. Mitra, and P. Majumder. Query expansion for microblog retrieval. In *Proceedings of The Twentieth Text REtrieval Conference, TREC 2011, Gaithersburg, Maryland, November 15-18*, 2011.

[3] A. Bandyopadhyay, S. Pal, M. Mitra, P. Majumder, and K. Ghosh. Passage retrieval for tweet contextualization at INEX 2012. In *CLEF 2012 Evaluation Labs and Workshop, Online Working Notes, Rome, Italy, September 17-20*, 2012.

[4] R. Belkaroui, R. Faiz, and A. Elkhlifi. Social users interactions detection based on conversational aspects. In *New Trends in Intelligent Information and Database Systems*, volume 598 of *Studies in Computational Intelligence*, pages 161–170. Springer International Publishing, 2015.

[5] R. Belkaroui, R. Faiz, and A. Elkhlifi. Using social conversational context for detecting users interactions on microblogging sites. *Revue des Nouvelles Technologies de l'Information*, Extraction et Gestion des Connaissances, RNTI-E-28:389–394, 2015.

[6] P. Bhaskar, S. Banerjee, and S. Bandyopadhyay. A hybrid tweet contextualization system using IR and summarization. In *CLEF 2012 Evaluation Labs and Workshop, Online Working Notes, Rome, Italy, September 17-20*, 2012.

[7] M. Choudhury, R. Saraf, V. Jain, A. Mukherjee, S. Sarkar, and A. Basu. Investigation and modeling of the structure of texting language. *International Journal of Document Analysis and Recognition (IJDAR)*, 10(3-4):157–174, 2007.

[8] F. Damak, K. Pinel-Sauvagnat, M. Boughanem, and G. Cabanac. Effectiveness of state-of-the-art features for microblog search. In *Proceedings of the 28th Annual ACM Symposium on Applied Computing*, SAC '13, pages 914–919, New York, NY, USA, 2013. ACM.

[9] R. Deveaud and F. Boudin. Contextualisation automatique de tweets à partir de wikipédia. In *CORIA 2013 - Conférence en Recherche d'Infomations et Applications - 10th French Information Retrieval Conference, Neuchâtel, Suisse, April 3-5*, pages 125–140, 2013.

[10] S. Eric, V. Moriceau, X. Tannier, P. Bellot, and J. Mothe. Overview of the INEX 2011 question answering trackQA@INEX. In *Focused Retrieval of Content and Structure, 10th International Workshop of the Initiative for the Evaluation of XML Retrieval, INEX 2011, Saarbrücken, Germany, December 12-14*, pages 188–206, 2011.

[11] L. Ermakova and J. Mothe. IRIT at INEX: question answering task. In *Focused Retrieval of Content and Structure, 10th International Workshop of the Initiative for the Evaluation of XML Retrieval, INEX 2011, Saarbrücken, Germany, December 12-14, 2011,*, pages 219–226, 2011.

[12] L. Ermakova and J. Mothe. IRIT at INEX2012: Tweet contextualization. In *Conference on Multilingual and Multimodal Information Access Evaluation (CLEF),Rome, Italy*, 2012.

[13] L. Ermakova and J. Mothe. IRIT at INEX 2013: Tweet contextualization track. In *Conference on Multilingual and Multimodal Information Access Evaluation (CLEF), Valencia, Spain, September 23-26*, 2013.

[14] L. Ermakova and J. Mothe. IRIT at INEX 2014: Tweet contextualization track. In *Conference on Multilingual and Multimodal Information Access Evaluation (CLEF), Sheffield, UK, September 15-18*, pages 557–564, 2014.

[15] J. H. Friedman. Greedy function approximation: A gradient boosting machine. In *The Annals of Statistics*, volume 29, pages 1189–1232. The Institute of Mathematical Statistics, 2000.

[16] D. Gaffney. #iranelection: quantifying online activism. In *WebSci10: Extending the Frontiers of Society On-Line*, 2010.

[17] D. Ganguly, J. Leveling, and G. J. F. Jones. Dcu@inex-2012: Exploring sentence retrieval for tweet contextualization. In *CLEF 2012 Evaluation Labs and Workshop, Online Working Notes, Rome, Italy, September 17-20*, 2012.

[18] J. Hannon, K. McCarthy, J. Lynch, and B. Smyth. Personalized and automatic social summarization of events in video. In *Proceedings of the 16th International Conference on Intelligent User Interfaces*, IUI '11, pages 335–338, New York, NY, USA, 2011. ACM.

[19] H. Kwak, C. Lee, H. Park, and S. Moon. What is twitter, a social network or a news media? In *Proceedings of the 19th International Conference on World Wide Web*, WWW '10, pages 591–600, New York, NY, USA, 2010. ACM.

[20] C. H. Lau, Y. Li, and D. Tjondronegoro. Microblog retrieval using topical features and query expansion. In *Proceedings of The Twentieth Text REtrieval Conference, TREC 2011, Gaithersburg, Maryland, November 15-18*, 2011.

[21] M. Morchid and G. Linarès. INEX 2012 benchmark a

semantic space for tweets contextualization. In *CLEF 2012 Evaluation Labs and Workshop, Online Working Notes, Rome, Italy, September 17-20*, 2012.

[22] M. Morchid and G. LinarÃÍs. Inex 2012 benchmark a semantic space for tweets contextualization. In P. Forner, J. Karlgren, and C. Womser-Hacker, editors, *CLEF (Online Working Notes/Labs/Workshop)*, 2012.

[23] J. M. Phillips and S. Venkatasubramanian. A gentle introduction to the kernel distance. *Computing Research Repository*, abs/1103.1625, 2011.

[24] D. Romain and B. Florian. Effective tweet contextualization with hashtags performance prediction and multi-document summarization. In *Working Notes for CLEF 2013 Conference , Valencia, Spain, September, 23-26*, 2013.

[25] T. Sakaki, M. Okazaki, and Y. Matsuo. Earthquake shakes twitter users: Real-time event detection by social sensors. In *Proceedings of the 19th International Conference on World Wide Web*, WWW '10, pages 851–860, New York, NY, USA, 2010. ACM.

[26] G. Salton, A. Wong, and C. S. Yang. A vector space model for automatic indexing. In *Communications of the ACM*, volume 18, pages 613–620, New York, NY, USA, Nov. 1975. ACM.

[27] E. SanJuan, P. Bellot, V. Moriceau, and X. Tannier. Overview of the inex 2010 question answering track (qa@inex). In S. Geva, J. Kamps, R. Schenkel, and A. Trotman, editors, *Comparative Evaluation of Focused Retrieval*, volume 6932 of *Lecture Notes in Computer Science*, pages 269–281. Springer Berlin Heidelberg, 2011.

[28] E. SanJuan, V. Moriceau, X. Tannier, P. Bellot, and J. Mothe. Overview of the INEX 2012 tweet contextualization track. In *CLEF 2012 Evaluation Labs and Workshop, Online Working Notes, Rome, Italy, September 17-20*, 2012.

[29] E. SanJuan, V. Moriceau, X. Tannier, P. Bellot, and J. Mothe. Overview of the inex 2012 tweet contextualization track. In P. Forner, J. Karlgren, and C. Womser-Hacker, editors, *CLEF (Online Working Notes/Labs/Workshop)*, 2012.

[30] N. Savage. Twitter as medium and message. volume 54, pages 18–20, New York, NY, USA, Mar. 2011. ACM.

[31] D. A. Shamma, L. Kennedy, and E. F. Churchill. Tweet the debates: Understanding community annotation of uncollected sources. In *Proceedings of the First SIGMM Workshop on Social Media*, WSM '09, pages 3–10, New York, NY, USA, 2009. ACM.

[32] K. Starbird, L. Palen, A. L. Hughes, and S. Vieweg. Chatter on the red: What hazards threat reveals about the social life of microblogged information. In *Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work*, CSCW '10, pages 241–250, New York, NY, USA, 2010. ACM.

[33] J. Torres-Moreno. Three statistical summarizers at CLEF-INEX 2013 tweet contextualization track. In *Working Notes for CLEF 2014 Conference, Sheffield, UK, September 15-18*, pages 565–573, 2014.

[34] J. Torres-Moreno and P. Velázquez-Morales. Two statistical summarizers at INEX 2012 tweet contextualization track. In *CLEF 2012 Evaluation Labs and Workshop, Online Working Notes, Rome, Italy, September 17-20*, 2012.

[35] S. Wu, J. M. Hofman, W. A. Mason, and D. J. Watts. Who says what to whom on twitter. In *Proceedings of the 20th International Conference on World Wide Web*, WWW '11, pages 705–714, New York, NY, USA, 2011. ACM.

[36] Y. Yamaguchi, T. Takahashi, T. Amagasa, and H. Kitagawa. Turank: Twitter user ranking based on user-tweet graph analysis. In *Proceedings of the 11th International Conference on Web Information Systems Engineering*, WISE'10, pages 240–253, Berlin, Heidelberg, 2010. Springer-Verlag.

[37] M. A. Zingla, M. Ettaleb, C. C. Latiri, and Y. Slimani. INEX2014: tweet contextualization using association rules between terms. In *Working Notes for CLEF 2014 Conference, Sheffield, UK, September 15-18, 2014.*, pages 574–584, 2014.