User-Tweet Interaction Model and Social Users Interactions for Tweet Contextualization

Rami Belkaroui, Rim Faiz, Pascale Kuntz

User-Tweet Interaction Model and Social Users Interactions For Tweet Contextualization

Rami Belkaroui[1], Rim Faiz[2], and Pascale Kuntz[3]

[1] LARODEC, ISG Tunis, University of Tunis
Bardo,Tunisia
rami.belkaroui@gmail.com
[2] LARODEC, IHEC Carthage, University of Carthage
Carthage Presidency, Tunisia
rim.faiz@ihec.rnu.tn
[3] LINA, la Chantrerie-BP 50609,
44360 Nantes cedex France
pascale.kuntz@univ-nantes.fr

Abstract. In the current era, microblogging sites have completely changed the manner in which people communicate and share information. They give users the ability to communicate, interact, create conversations with each other and share information in real time about events, natural disasters, news, etc. On Twitter, users post messages called tweets. Tweets are short messages that do not exceed 140 characters. Due to this limitation, an individual tweet it's rarely self-content. However, users cannot effectively understand or consume information.
In order, to make tweets 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. 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 ensure that context summaries contain adequate correlating information with the given tweet.

Keywords: Tweet contextualization, Tweet influence, Conversation aspects, user interactions.

1 Introduction

Recent years have revealed the exponential growth of microblogging platforms, offering to users an easy way to share different kinds of information like common knowledge, opinions, emotions under the form of short text messages. Twitter, the microblogging service addressed in our work, is a communication mean and a collaboration system that allows users to share short text messages (tweets), which doesn't exceed 140 characters with a defined group of users called followers. Twitter's data flow is examined in order to measure public sentiment, trends monitoring, reputation management, follow political activity and news. However, tweets in its raw form can be less informative, but also overwhelming. For both end-users and data analysts, it is a nightmare to plow through millions of tweets which contain a lot of noise and redundancy. Furthermore, an individual tweet is short and without sufficient contextual information, it is often hard 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 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.

Traditional contextualization techniques only consider text information which is insufficient for 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. Given that there are rich user interactions on Twitter called social conversations [3], this paper describes an approach that exploits social conversations to provide some context for a given tweet, 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 signals, user-tweet influence signals, temporal signals and text based signals, which can be potentially useful to improve tweet contextualization task.

The remainder of this paper is organized as follows: we begin by describing some related works presented in INEX 2012, 2013 and 2014. In section 3, we present our approach that exploits social conversations to provide some context for a given tweet, and introduce tweet-user influence model in section 4. In section 5, we explain different types of signals used in our work. Our experimental results are presented in Section 6. Finally, we conclude and present some future works.

2 Related Work

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

In [9], 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. Recently, [10] modified the method presented at INEX 2011, 2012 and 2013 [12,13] by adding the influence of topic-comment relationship on contextualization. The approach proposed 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. While, in the same way [1] described a pipeline system where first extracted phrases from tweets by using ArkTweet toolkit and some heuristics; then retrieved relevant documents for these phrases from Wikipedia before summarizing those with MEAD toolkit.

In [22], the authors developed and tested 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 the tweet context applied to CLEF-INEX 2014 task [21]. 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 [8], the authors used a method that allows to automatically contextualize tweets by using information coming from Wikipedia. They treat the problem of tweet 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 [17] the authors combined Information Retrieval, Automatic Summarization and Topic Modeling techniques to provide the context of each tweet. They took advantage of a larger use of hashtags in the topics and used them to enhance the retrieval of relevant Wikipedia articles. In [2], 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 Wikipedia corpus. Another approach proposed by [14] 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, [25] described a new method for tweet contextualization based on association rules between sets of terms. This approach allows the extension of tweet's vocabulary by a set of thematically related words.

These works presented in INEX Tweet Contextualization track are based on the assumption that it is possible to overcome the tweets' lack of knowledge by providing a bunch of sentences that give some context or additional information extracted from Wikipedia about a given tweet. However, sometimes after news events, the Wikipedia information is not immediately available. For example: After, the terrorist attack against newspaper Charlie Hebdo, there were no articles on Wikipedia describing the topic #jesuischarlie. Indeed, the first article that explains this topic was available 7 hours after this attack. While, the first tweet was launched at 11:52h (less than an hour after the attack). At the same time, this event demonstrates a scenario where users urgently need

information, especially if they are directly affected by the event. Unexpected news events represent an information access problem where the approaches using Wikipedia to contextualize a tweet fail.

3 Our Proposed Approach

On Twitter, users are posting millions of tweets in order to express what they are thinking about natural disasters, political debates and sporting events followed by some comments, retweet or favorite. The users' interactions essentially reflect the importance of different tweets and can be used to improve the tweet contextualization quality.

Compared to traditional contexts that are defined based on textual information, we modeled social tweet context using various dynamic social relationships such as following relationships between users, retweeting and replying relationships between tweets. In this paper, our proposed approach (Figure 1) consists of extracting social tweet context by means of social Twitter conversations. In addition, we defined a social tweet context and social Twitter conversations as follows:

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.

Social Twitter conversations[5]

There is 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 (favorite).

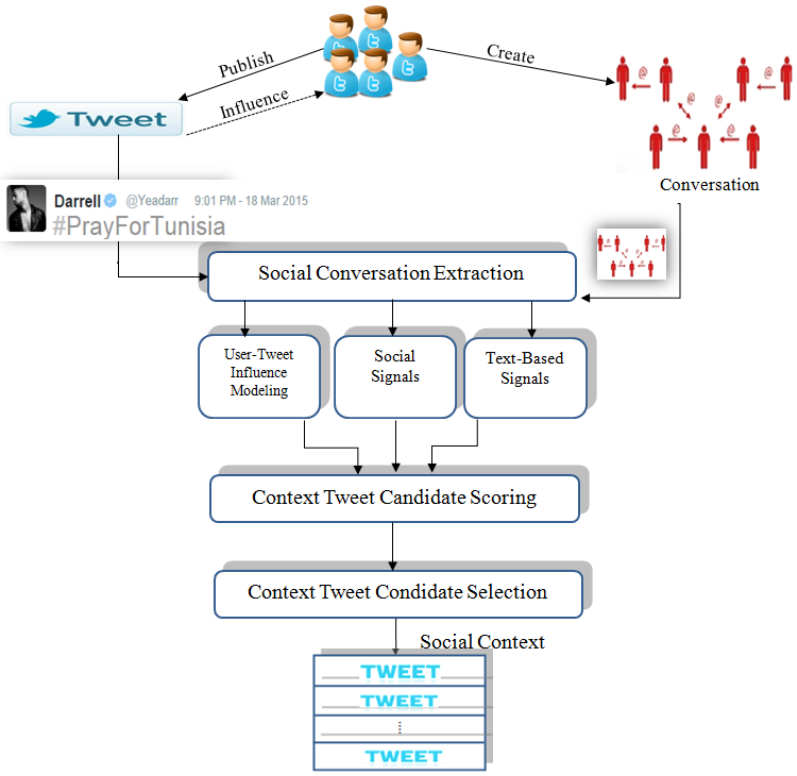

Fig. 1. Overview of our proposed approach

4 Social Influence Generation

On Twitter, there are several interactional relationships between users and tweets such as post, reply, follow, favorite, mention and retweet. We take these relationships into account for measuring

tweet influence score in order to select context candidate tweets. The motivations of using tweet influence are:

- If we know that the user A has a strong influence on a user B within in 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 context candidate tweet.

4.1 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 [24]. 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.

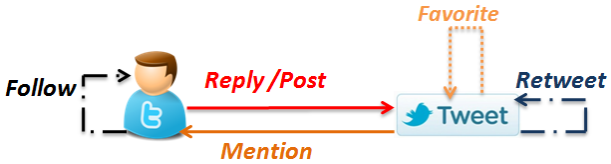

Fig. 2. user-tweet schema graph

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 3. In the case of user-tweet interaction model, it consists of user nodes, tweets nodes and six kinds of edges.

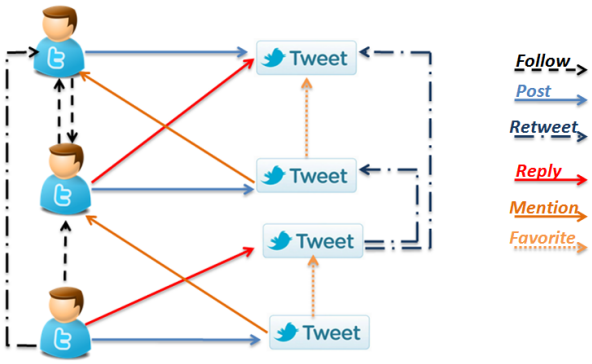

Fig. 3. user-tweet interaction model

4.2 Influence Measuring Based on User-Tweet Interaction Model

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 tweet 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.

Measuring of Tweet Influence 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:

$$Reply\_influence(t) = \alpha \times number\_reply(t). \tag{1}$$

$\alpha \in (0, 1]$. 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:

$$Retweet\_influence(t) = \beta \times number\_retweet(t). \tag{2}$$

$\beta \in (0, 1]$. 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:

$$Favorite\_influence(t) = \gamma \times number\_favorite(t). \tag{3}$$

$\gamma \in (0, 1]$. 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 [16] to estimate a distanceΔt between tweet conversation root time d and other tweet time d' within the same conversation, i.e., $\Delta t = |d' - d|$. It is defined as follow:

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

Finally, the tweet influence score is defined as follows:

$$Tweet\_influence(t) = \Gamma(\Delta t) \times Reply\_influence(t) + Retweet\_influence(t) + Favorite\_influence(t). \tag{5}$$

Measuring of Tweet's Author Influence In Twitter, many celebrities post relevant messages but got many followers simply because of their influence in real life. Thus, considering only the number of followers can not show real influence in microblogs. It has been proved that attributes expressing engaging audience links such as mention relationship are better to represent user influence. Furthermore, in our work we consider both the follow relationship and the mention relationship.

– **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:

$$Mention\_influence(u) = \delta \times number\_Mention(u). \tag{6}$$

$\delta \in (0, 1]$. 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:

$$Follow\_influence(u) = \omega \times number\_follow(u). \tag{7}$$

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

.

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

$$Tweetauthor\_influence(u) = Mention\_influence(u) + Follow\_influence(u) \tag{8}$$

5 Context Candidates Tweets Extraction

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

5.1 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 $\boldsymbol{V}$ using the vector space model [18]. 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 tweet root vector $\boldsymbol{V_{t_{root}}}$ and other tweets vector $\boldsymbol{V_t}$ within the same conversation. In addition, we aim to measure how much a tweet would be related to tweet root's content.

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

– Similarity of content [7]

In our case, it measures how many tweet of the whole conversation C are similar in content with current tweet $\boldsymbol{t_{current}}$. We calculate cosine similarity score for every pair of tweets. The similarity is calculated using Lucene similarity function [4]. We denote current tweet modeled as a vector:

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

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

5.2 Social Signals

- **Context Candidates Tweets Regarding the URLs**
  By sharing an URL, an author would enrichment 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 a candidate tweet t and the web page title, and between t and the body content of the web page.
  - The cosine similarity between t and the web page title, and between t and the body content of the web page.

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

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

5.3 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 [11] 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.

6 Experiments and Results

6.1 Twitter Conversations Data Set

As celebrities are highly influential in Twitter [23], 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, François Hollande. From another perspective, 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.

6.2 Reference Summary

To the best of our knowledge, there is no data set available to evaluate social Twitter contextualization. Thus, we conduct a pilot study to construct an editorial data set in our work. The pilot study goal 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.The assessors selected among students and colleagues of the authors (with backgrounds in computing and social sciences). In addition, we ask each assessor to first read the tweet root and open any URL inside to have a sense of what the tweet root is about. Then, the assessor reads through all contexts candidates 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.

6.3 Evaluation Metrics

Tweet contextualization is evaluated on both informativeness and readability [19]. Informativeness aims at measuring how well the summary explains tweet or how well the summary helps user to understand 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 relevant tweets selection. Informativeness aims at measuring how well summary helps user to understand tweet content. Therefore, for each tweet root, each candidate tweet will be evaluated independently from the others, even in the same summary. 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:

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

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

	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

- **readability**: readability aims at measuring how clear and easy it is to understand summary. By contrast, readability is evaluated manually (cf.Table 2). Each summary has been evaluated by considering the following two parameters [20]:
  - Relevance: judge if the tweet 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 context does not contain too much redundant information, i.e. information that has already been given in a previous tweet. 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).

6.4 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 ensure that context summaries contain adequate correlating information with the tweet root. In addition, based on 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 tweet contextualization approach based on social Twitter conversations leads to the context informativeness improvement ; we note also that the tweets selection impacts the context quality. The contexts are less readable; it may be that they contain some noises which need to be cleaner.

	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

7 Conclusion

we explored in this paper the tweet contextualization problem. We proposed an approach 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 influence signals 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 or comments on Facebook pages.

References

1. Ansary, K.H., Tran, A.T., Tran, N.K.: A pipeline tweet contextualization system at INEX 2013. In: Working Notes for CLEF 2013 Conference , Valencia, Spain, September 23-26, 2013. (2013)
2. Bandyopadhyay, A., Pal, S., Mitra, M., Majumder, P., Ghosh, K.: Passage retrieval for tweet contextualization at INEX 2012. In: CLEF 2012 Evaluation Labs and Workshop, Online Working Notes, Rome, Italy, September 17-20, 2012 (2012)
3. Belkaroui, R., Faiz, R., Elkhlifi, A.: Conversation analysis on social networking sites. In: Tenth International Conference on Signal-Image Technology and Internet-Based Systems, SITIS 2014, Marrakech, Morocco, November 23-27, 2014. pp. 172–178 (2014)
4. Belkaroui, R., Faiz, R., Elkhlifi, A.: Social users interactions detection based on conversational aspects. In: Barbucha, D., Nguyen, N.T., Batubara, J. (eds.) New Trends in Intelligent Information and Database Systems, Studies in Computational Intelligence, vol. 598, pp. 161–170. Springer International Publishing (2015)
5. Belkaroui, R., Faiz, R., Elkhlifi, A.: 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. Bhaskar, P., Banerjee, S., Bandyopadhyay, S.: 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 (2012)
7. Damak, F., Pinel-Sauvagnat, K., Boughanem, M., Cabanac, G.: Effectiveness of state-of-the-art features for microblog search. In: Proceedings of the 28th Annual ACM Symposium on Applied Computing. pp. 914–919. SAC '13, ACM, New York, NY, USA (2013)
8. Deveaud, R., Boudin, F.: 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, 2013. pp. 125–140 (2013)
9. Ermakova, L., Mothe, J.: 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, Revised Selected Papers. pp. 219–226 (2011)
10. Ermakova, L., Mothe, J.: IRIT at INEX 2014: Tweet contextualization track. In: Conference on Multilingual and Multimodal Information Access Evaluation (CLEF), Sheffield, UK, September 15-18, 2014. pp. 557–564 (2014)
11. Friedman, J.H.: Greedy function approximation: A gradient boosting machine. In: The Annals of Statistics. vol. 29, pp. 1189–1232. The Institute of Mathematical Statistics (2000)
12. Liana, E., Josiane, M.: IRIT at INEX2012: Tweet contextualization. In: Conference on Multilingual and Multimodal Information Access Evaluation (CLEF),Rome, Italy, 17/09/2012-20/09/2012 (2012)
13. Liana, E., Josiane, M.: IRIT at INEX 2013: Tweet contextualization track. In: Conference on Multilingual and Multimodal Information Access Evaluation (CLEF), Valencia, Spain, September 23-26, 2013. (2013)
14. Morchid, M., Linarès, G.: 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 (2012)
15. Omatu, S., Malluhi, Q.M., Rodríguez-González, S., Bocewicz, G., Bucciarelli, E., Giulioni, G., Iqba, F. (eds.): Distributed Computing and Artificial Intelligence, 12th International Conference, DCAI 2015, Salamanca, Spain, June 3-5, 2015, Advances in Intelligent Systems and Computing, vol. 373. Springer (2015), http://dx.doi.org/10.1007/978-3-319-19638-1
16. Phillips, J.M., Venkatasubramanian, S.: A gentle introduction to the kernel distance. Computing Research Repository abs/1103.1625 (2011)
17. Romain, D., Florian, B.: Effective tweet contextualization with hashtags performance prediction and multi-document summarization. In: Working Notes for CLEF 2013 Conference , Valencia, Spain, September 23-26, 2013. (2013)
18. Salton, G., Wong, A., Yang, C.S.: A vector space model for automatic indexing. In: Communications of the ACM. pp. 613–620. No. 11, ACM, New York, NY, USA (Nov 1975)
19. SanJuan, E., Bellot, P., Moriceau, V., Tannier, X.: Overview of the inex 2010 question answering track (qa@inex). In: Geva, S., Kamps, J., Schenkel, R., Trotman, A. (eds.) Comparative Evaluation of Focused Retrieval, Lecture Notes in Computer Science, vol. 6932, pp. 269–281. Springer Berlin Heidelberg (2011)
20. SanJuan, E., Moriceau, V., Tannier, X., Bellot, P., Mothe, J.: Overview of the inex 2012 tweet contextualization track. In: Forner, P., Karlgren, J., Womser-Hacker, C. (eds.) CLEF (Online Working Notes/Labs/Workshop) (2012)
21. Torres-Moreno, J.: Three statistical summarizers at CLEF-INEX 2013 tweet contextualization track. In: Working Notes for CLEF 2014 Conference, Sheffield, UK, September 15-18, 2014. pp. 565–573 (2014)
22. Torres-Moreno, J., Velázquez-Morales, P.: 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 (2012)
23. Wu, S., Hofman, J.M., Mason, W.A., Watts, D.J.: Who says what to whom on twitter. In: Proceedings of the 20th International Conference on World Wide Web. pp. 705–714. WWW '11, ACM, New York, NY, USA (2011)
24. Yamaguchi, Y., Takahashi, T., Amagasa, T., Kitagawa, H.: Turank: Twitter user ranking based on user-tweet graph analysis. In: Proceedings of the 11th International Conference on Web Information Systems Engineering. pp. 240–253. WISE'10, Springer-Verlag, Berlin, Heidelberg (2010)
25. Zingla, M.A., Ettaleb, M., Latiri, C.C., Slimani, Y.: INEX2014: tweet contextualization using association rules between terms. In: Working Notes for CLEF 2014 Conference, Sheffield, UK, September 15-18, 2014. pp. 574–584 (2014)