Leveraging expertise in news feeds: A *Twitter* case study

Sami Belkacem*, Kamel Boukhalfa*, Omar Boussaid**

*LSI laboratory, USTHB, Algiers, Algeria

{s.belkacem,kboukhalfa}@usthb.dz

**ERIC laboratory, University of Lyon 2, Lyon, France

omar.boussaid@univ-lyon2.fr

Abstract. Due to the large amount of information posted on social media, users find themselves overwhelmed by updates displayed chronologically in their news feed. Moreover, most of them are considered irrelevant. Ranking news feeds updates in order of relevance is proposed to help beneficiary users quickly catch up with the relevant updates. Four types of features are mainly used to predict the relevance: (1) the relevance of the update's content to the beneficiary's interests; (2) the social tie strength between the beneficiary and the update's author; (3) the author's authority; and (4) the update's quality. In this work, we propose an approach that leverages another type of feature which is the author's expertise for the update's topic. Experimental results on *Twitter* highlight that judging expertise is crucial for maximizing the relevance of updates in news feeds.

1 Introduction

Social media, such as *Facebook* or *Twitter*, contribute to the concept of *Big data*. Social data are known for their volumes that can reach petabytes (10^{15} bytes), their variety (messages, articles, videos, music, images, etc.), and their velocity (arriving in real time or almost) (Xu et al., 2016a). Due to the large amount of information posted on social media (Vougioukas et al., 2017), users find themselves overwhelmed by updates displayed chronologically in their news feed (De Maio et al., 2017). Moreover, most of them are considered irrelevant (Vougioukas et al., 2017). Therefore, it becomes difficult for users to quickly catch up with relevant updates (Kuang et al., 2016). Based on the prediction of a relevance score between a beneficiary user and a new unread update in his news feed (Belkacem et al., 2016), approaches have been proposed for ranking news feeds updates in a descending relevance order (Agarwal et al., 2015). These approaches generally use 4 types of features that may influence relevance (Belkacem et al., 2016): (1) the relevance of the update's content to the beneficiary's interests; (2) the social tie strength between the beneficiary and the update's author; (3) the author's authority; and (4) the update's quality. We believe that using these features is necessary, but not sufficient. For example, updates posted by a typical user may not attract the attention such as those posted by a recognized expert in his field. Indeed, updates posted by experts are considered credible, useful and interesting (Xu et al., 2016b), and finding these experts enable users to view and interact with the relevant and trustworthy updates on a specific topic (Wei et al., 2016). Since expertise information is usually not explicitly provided by users (Liao et al., 2012), existing methods rely on expert finding, which aims at identifying users with the relevant knowledge or experiences on a given topic (Wagner et al., 2012). The main techniques used to infer a user's topical expertise leverage his behavior on social media including: the textual content he posted, his biographical information, his social relationships, his list memberships [1], etc. (Xu et al., 2017). Based on existing works, to the best of our knowledge, the author's expertise has not been used before when predicting the relevance of updates to beneficiary users. This work aims to leverage this expertise and focuses on the most popular micro-blogging site *Twitter* for the followings reasons: (1) the large flow of tweets encountered by users; (2) the irrelevance of a large part of tweets; (3) data are public by default unlike most of the other social media; and (4) the availability of API for easy crawling (Berkovsky and Freyne, 2015). However, with some adaptations, it would be possible to exploit this work in other social media.

Twitter is a microblogging social media that allows users to communicate using short messages of 280 characters called "tweets" (Shen et al., 2013). Each tweet has (see Fig. 1): (1) an author; (2) a set of beneficiaries users who can read and interact with the tweet; (3) a textual and/or multimedia content; (4) a publication date; (5) a space to perform actions on the tweet: save, report, hide, etc.; (6) hashtags which identify tweets on specific topics; (7) mentions which represent links to other users; and (8) URLs to websites or articles (Belkacem et al., 2016). Following is the only type of social relationship. When a user ***u*** follows another user ***u'***, ***u*** will receive in his news feed the tweets posted by ***u'*** and the tweets he retweeted and/or liked. Users who follow ***u*** are called *followers* of ***u*** and users that ***u*** follows are called *followings* of ***u***. In this paper, in order to study the contribution of the author's expertise to rank news feeds updates, we propose an approach that leverages, in addition to the 4 types of features used in related works, the author's topical expertise that we infer from his tweets.

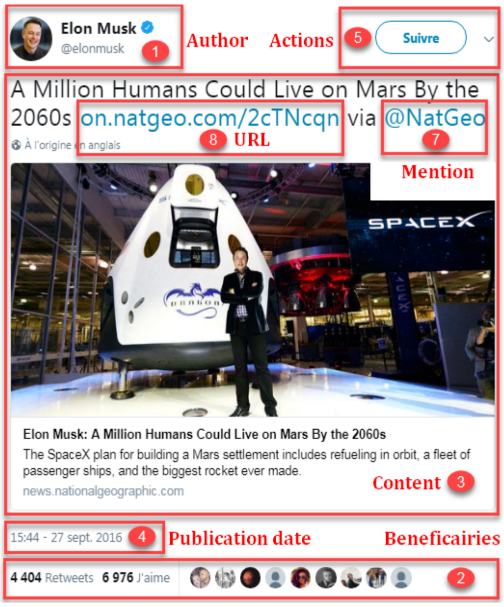
The paper is structured as follows: section 2 presents a background on ranking news feeds updates on *Twitter*, section 3 discusses related works, section 4 describes our proposed approach that leverages the author's expertise in addition to other features, section 5 presents the experiments we performed to evaluate our approach and study the contribution of the author's expertise to rank news feeds updates, and section 6 concludes and proposes future work.

2 Ranking news feeds updates on *Twitter*

A *Twitter* user's news feed, or timeline, is a list of tweets where are displayed, from the most recent to the least recent, tweets posted by his followings and tweets they retweeted and/or liked (Shen et al., 2013). When a tweet is displayed on a user's news feed, he can perform 3 actions to interact with it: (1) *Retweet*: when he finds this tweet interesting and wants to share it with his followers; (2) *Reply*: when he wants to answer or comment on this tweet; and (3) *Like*: previously named *Favorite*, when he finds this tweet interesting and wants to save it in

1. Lists that allow users organize people they follow into labeled groups.

his "Likes" section. Ranking news feeds updates on *Twitter* implies ranking and displaying, in a descending relevance order, the tweets of each user's news feed (Feng and Wang, 2013). The ranking process is done in such a way that the most relevant tweets are found at the top of the news feed and the least relevant at the bottom (Shen et al., 2013). Note that other terms can be used to refer to the ranking process, e.g.: reordering, recommendation, personalization, etc.



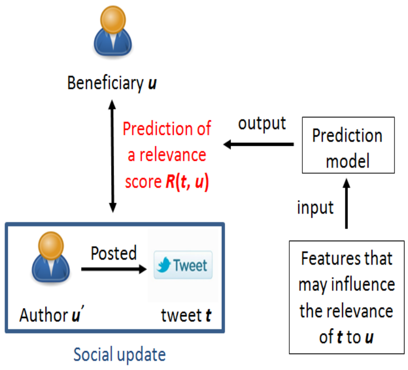


FIG. 2: Prediction of a relevance score.

FIG. 1: Tweet posted by *Elon Musk*.

Berkovsky and Freyne (2015) propose the following formalization for the problem of ranking news feeds updates: "Let $F(u)$ denotes all the tweets unread by the beneficiary user u that can potentially be included in his news feed. The ranking process implies selecting and displaying a subset $K(u) \in F(u)$, such that $|K(u)| \ll |F(u)|$, that corresponds to the most relevant tweets to u. This ranking involves 3 steps: (1) predict and assign a relevance score to each tweet $t \in F(u)$; (2) select and display, in a descending relevance order, the $|K(u)|$ tweets with the highest relevance scores in the news feed; and (3) delete the remaining $F(u) \setminus K(u)$ tweets". This work focuses on the first step, which is the most important one. The main techniques used to predict and assign a relevance score to a tweet $t \in F(u)$ are based on a prediction model, that uses as input features that may influence the relevance of t to u, to output a relevance score, denoted $R(t,u)$, that measures the relevance of t to u (Belkacem et al., 2016). Note that t is posted by an author $u' \in A(u)$, such that $A(u)$ is the set of users that u follows (see Fig. 2).

3 Related works

Ranking and predicting relevance of news feeds updates is studied in both industrial and academic community. In the industrial community, *Facebook*, *Twitter*, and *LinkedIn* make efforts to rank news feeds updates. However, their approaches are most often not disclosed due to commercial sensitivity and competition between companies (Berkovsky and Freyne, 2015). Furthermore, these companies claim that their algorithms show several limits [2] (Agarwal et al., 2015). As regards the academic community, several works have been done (Belkacem et al., 2017). Due to lack of space, we present the most representative ones.

In order to assign continuous relevance scores to news feed updates on the Chinese social media *Sina Weibo*, Kuang et al. (2016) proposed a prediction model based on weighted linear combinations with static weights. The model uses 3 types of features: social tie strength between ***u*** and ***u'***, relevance of the content of ***t*** to the interests of ***u***, and quality of ***t***. To evaluate their approach, the authors asked 1048 users to explicitly assign boolean values to updates (True for relevant and False for irrelevant). The model's *MAE* (Mean Average Precision) was 75% and was improved by 57% compared to the results of the chronological model. With the purpose of recommend relevant tweets to *Twitter* users, Shen et al. (2013) and Chen et al. (2012) proposed prediction models that use 4 types of features: social tie strength between ***u*** and ***u'***, relevance of the content of ***t*** and its mentions to the interests of ***u***, quality of ***t***, and authority of ***u'***. In (Shen et al., 2013), the authors used a supervised binary classifier model based on a *Gradient Boosted ranking* algorithm. The model's average accuracy was improved by 34.5% when comparing the results of the chronological feeds. While in (Chen et al., 2012), the authors used a probabilistic collaborative filtering model based on *latent factors* to predict binary rating scores. The model's *MAE* was 76% and the results indicated that recommended tweets attracted more attention than unrecommended ones.

With the aim of ranking tweets in order of relevance on *Twitter*, Feng and Wang (2013), De Maio et al. (2017), and Vougioukas et al. (2017) proposed prediction models that use 5 types of features: social tie strength between ***u*** and ***u'***, relevance of the content of ***t*** and its mentions to the interests of ***u***, quality of ***t***, authority of ***u'***, and activity of ***u'***. In (Feng and Wang, 2013), the model is based on matrix factorization and predict the likelihood that a beneficiary retweet a tweet from his news feed. The model's *MAE* was 76.27% and outperformed several baseline methods including the chronological model. In (De Maio et al., 2017), the authors used a supervised binary classifier model based on a *Deep Learning* method attempting to re-adapt the ranking of the tweets by preferring those that are more likely interesting to the beneficiary user. The model's *MAE* and *NDCG* (Normalized Discounted Cumulative Gain) outperformed several baseline methods including the chronological model. While in (Vougioukas et al., 2017), the authors used a supervised binary classifier model, based on *Logistic regression*, that predicts if the beneficiary will retweet the incoming tweet. In experiments with a collection of tweets received by journalists, the model's average *F1 score* was 90%.

In all previous works on *Twitter*, to obtain training and evaluation data, users' interactions with tweets, in terms of retweets and replies, were used as implicit indicators of their relevance (Shen et al., 2013; Chen et al., 2012; Feng and Wang, 2013; De Maio et al., 2017; Vougioukas et al., 2017). We note the predominance of this intuitive implicit method which assumes that user's interaction with a tweet involves its relevance. We notice also that supervised learning models, which aim to analyze users present and past behaviors to make predictive assumptions about future outcomes, have been commonly used and seem to be appropriate to rank news

2. https://longform.org/posts/who-controls-your-facebook-feed

feeds updates (Shen et al., 2013; De Maio et al., 2017; Vougioukas et al., 2017). Finally, we note that 4 types of features that may influence relevance were widely used:

— Features between ***u*** and ***t*** that measure the relevance of the content of ***t*** and its mentions to the interests of ***u*** (Kuang et al., 2016; Shen et al., 2013; Chen et al., 2012; Feng and Wang, 2013; De Maio et al., 2017; Vougioukas et al., 2017). These features are the most intuitive and may serve as direct predictors of relevance (Kuang et al., 2016).

— Features between ***u*** and ***u'*** that measure social tie strength between them (Kuang et al., 2016; Shen et al., 2013; Chen et al., 2012; Feng and Wang, 2013; De Maio et al., 2017; Vougioukas et al., 2017). The assumption is that ***t*** could be relevant to ***u*** if he has a strong social relationship with ***u'*** (De Maio et al., 2017). Certainly, people who have a strong relationship tend to have common interests (Vougioukas et al., 2017).

— Features of ***u'*** that measure his authority (Shen et al., 2013; Chen et al., 2012; Feng and Wang, 2013; De Maio et al., 2017; Vougioukas et al., 2017). The assumption is that ***t*** could be relevant to ***u*** if ***u'*** has authority. Indeed, Nagmoti et al. (2010) states that if a user is important, i.e. has authority, then his tweets are also important.

— Features of ***t*** that measure his quality: formal, informative, popular, etc. (Kuang et al., 2016; Shen et al., 2013; Chen et al., 2012; Feng and Wang, 2013; De Maio et al., 2017; Vougioukas et al., 2017). The assumption is that ***t*** could be relevant to ***u*** if it is of good quality, independently of his interests (Chen et al., 2012).

Nonetheless, based on existing works, to the best of our knowledge, the features that measure the expertise of ***u'*** for the topics of ***t*** (features between ***u'*** and ***t***) have not been considered in related works. The assumption is that ***t*** could be relevant to ***u*** if ***u'*** is an expert in the topics of ***t***. Certainly, tweets posted by experts, i.e. users with the relevant knowledge or experiences on a given topic (Wagner et al., 2012), are often considered credible, valuable and interesting (Xu et al., 2017), and leverage this expertise might allow beneficiary users to catch up with the relevant and trustworthy updates. For example, *Elon Musk*, one of the most popular heroes in the tech culture, is known for his warning about the risks of Artificial Intelligence, and his tweets about that often attract users' attention [3]. In the next Section, we present our approach that leverages the author's expertise in addition to other features considered in related works.

4 Proposed approach

The proposed approach takes as an input a set of tweets ***F(u)*** unread by a beneficiary user ***u***, and outputs a relevance score to each tweet ***t*** $\in$ ***F(u)***. This approach uses *Decision Trees* models and leverages, in addition to other features considered in related works, the author's expertise that we infer from tweets he posted. Let denote by ***S*** the subset of beneficiary users for whom we apply the proposed approach, and ***D(u)*** a subset of tweets previously read by a user ***u***. With a view to use supervised prediction models based on *Decision Trees*, we first create a training database for each user ***u*** $\in$ ***S***. We recall that supervised learning has been commonly used in related works and seem to be appropriate to rank news feeds updates (Shen et al., 2013; De Maio et al., 2017; Vougioukas et al., 2017). The training database is a set of input-output pairs, such that an input represents the features that may influence the relevance of a tweet ***t*** $\in$ ***D(u)*** to ***u***, and the output represents the implicit relevance score, denoted ***R(t,u)***,

3. www.wired.co.uk/article/elon-musk-artificial-intelligence-world-war-3

that measures the relevance of $\boldsymbol{t}$ to $\boldsymbol{u}$. The proposed approach involves 3 steps: (1) assigning relevance scores to tweets; (2) defining features that may influence relevance; and (3) training of the prediction model. In this section, we describe each step of our approach.

4.1 Relevance scores

For each user $\boldsymbol{u} \in \boldsymbol{S}$, and each previously read tweet $\boldsymbol{t} \in \boldsymbol{D(u)}$, to assign an implicit relevance score $\boldsymbol{R(t,u)}$ that measures the relevance of $\boldsymbol{t}$ to $\boldsymbol{u}$, we assume that $\boldsymbol{t}$ is relevant to $\boldsymbol{u}$ if $\boldsymbol{u}$ interacted with $\boldsymbol{t}$. As shown by equation 1, predicting relevance scores becomes a binary classification problem with $\boldsymbol{R(t,u)} \in \{0,1\}$. We believe that *likes*, as well as retweets and replies which have been used in related work, are also implicit indicators of relevance.

$$\boldsymbol{R(t,u)} = \begin{cases} 1 & \text{if } \boldsymbol{u} \text{ interacted with } \boldsymbol{t} \text{ (retweet or reply or like)} \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

We use the implicit method, which has been used in several works (Shen et al., 2013; Chen et al., 2012; Feng and Wang, 2013; De Maio et al., 2017; Vougioukas et al., 2017), because the explicit method used by Kuang et al. (2016) brings several limits. It is not included on social media on one hand (the authors asked users to assign relevance scores to updates), and on the other, it is binding since it asks users to assign relevance scores to a large amount of updates (Belkacem et al., 2016). Moreover, we split relevance scores into 2 bins, relevant and not relevant, because train a finer-grained classifier (e.g. $\boldsymbol{t}$ is very relevant to the beneficiary user $\boldsymbol{u}$ if he retweeted, liked, and replied to it) would have been difficult since users' multiple interactions with the same tweet are not common (Vougioukas et al., 2017).

4.2 Features that may influence relevance

We define 9 features that may influence the relevance score $\boldsymbol{R(t,u)}$ that measures the relevance of a tweet $\boldsymbol{t}$, posted by an author $\boldsymbol{u'}$, to the beneficiary $\boldsymbol{u}$. These features are summarized in Table 1 and divided into 5 categories: (1) features that measure the relevance of the content of $\boldsymbol{t}$ and its mentions to the interests of $\boldsymbol{u}$; (2) features that measure social tie strength between $\boldsymbol{u}$ and $\boldsymbol{u'}$; (3) features that measure the expertise of $\boldsymbol{u'}$ for the topics of $\boldsymbol{t}$; (4) features that measure the authority of $\boldsymbol{u'}$; and (5) features that measure the quality of $\boldsymbol{t}$. For each category, we consider the most relevant features according to related work. Note that the features 1,3,4 and 6 are gradually updated as tweets are injected into the news feed of $\boldsymbol{u}$, from the least recent to the most recent. We thus simulate an evolution of the social media over time. In the rest of this Section, we provide a detailed description of each feature.

Relevance of the keywords of $\boldsymbol{t}$ to $\boldsymbol{u}$ This feature measures the relevance of the textual content of $\boldsymbol{t}$ to the interests of $\boldsymbol{u}$. According to Shen et al. (2013), keywords of tweets previously posted by a user and/or with which he has previously interacted reflect implicitly his topics of interests, and may serve as direct predictors of whether $\boldsymbol{t}$ is relevant to $\boldsymbol{u}$. First, after removing HTML characters and URLs in the textual content of $\boldsymbol{t}$, we represent its topics with keywords defined using *DBpedia Spotlight* annotation service [4] which is based on *DBpedia* [5]. When as-

4. http://demo.dbpedia-spotlight.org/
5. Project aiming to extract structured content from *Wikipedia*.

Features that may influence relevance		N°
Relevance of the content of ***t*** and its mentions to ***u***	Relevance of the keywords of ***t*** to ***u***	f_1
	Relevance of the mentions of ***t*** to ***u***	f_2
Social tie strength between ***u*** and ***u'***	Interaction rate of ***u*** with ***u'***	f_3
Expertise of ***u'*** for ***t***	Publishing rate of ***u'*** for keywords of ***t***	f_4
Authority of ***u'***	Followers count / Followings count	f_5
	Seniority	f_6
Quality of ***t***	Presence of hashtags	f_7
	Presence of a URL	f_8
	Presence of an image or a video	f_9

TAB. 1: Features that may influence relevance.

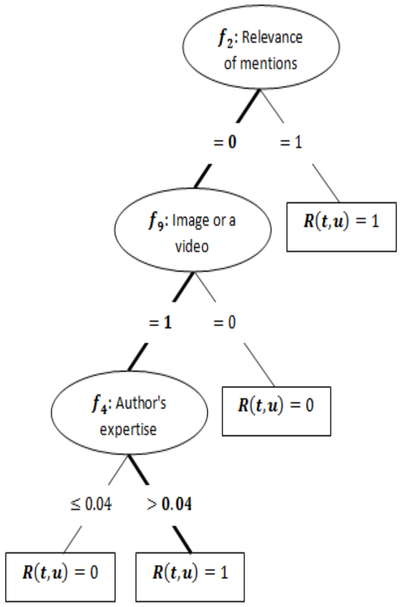


FIG. 3: *Decision Tree* of *NASAKepler*.

signing numerical values to the variables *support* and *confidence*, it allows us to represent each keyword by the URI[6] (Uniform Resource Identifier) corresponding to the annotated resource. This method, unlike methods used in related works which are mainly based on TF-IDF (Feng and Wang, 2013; Vougioukas et al., 2017) or a topic model (Chen et al., 2012; Shen et al., 2013), is not compromised by the short and informal nature of tweets. For example, applied to the tweet in Fig. 1 with a *support* = 20 and a *confidence* = 0.32, the annotation service returns these keywords: *2060s, National_Geographic_Channel, Human, Mars*. After that, we propose to calculate this feature with the following equation:

$$f_1(\boldsymbol{u},\boldsymbol{t}) = \sum_{i=1}^{nbk(\boldsymbol{t})} P(\boldsymbol{u}, k_i(\boldsymbol{t})) \tag{2}$$

Such that:

— $k_i(\boldsymbol{t})$ is the i^{th} keyword of ***t***

— $nbk(\boldsymbol{t})$ is the number of keywords of ***t***

— $P(\boldsymbol{u}, k_i(\boldsymbol{t}))$ is the number of times ***u*** has previously posted and/or interacted with $k_i(\boldsymbol{t})$

E.g., if ***t*** has 2 keywords k_1 and k_2, and ***u*** has previously posted and/or interacted 10 times with k_1 and 5 times with k_2, the sum of the two values, i.e. 15, will be assigned to this feature.

Relevance of the mentions of *t* to *u* A tweet that mentions ***u*** is likely to match his interests. Feng and Wang (2013) state that a mention serves to draw a user's attention to the tweet in which he was mentioned. The authors propose to calculate this feature with a boolean variable.

6. String of characters used to identify a resource.

Interaction rate of *u* with *u'* As reported by Vougioukas et al. (2017), if ***u*** interacted frequently with tweets posted by ***u'*** in the past, i.e. he found his tweets relevant, he tend to have a strong social relationship with ***u'*** and keep interacting with his tweets in the future. We propose to calculate this feature with the following equation:

$$f_3(\boldsymbol{u},\boldsymbol{u'}) = \frac{|\text{Tweets posted by } \boldsymbol{u'} \text{ with which } \boldsymbol{u} \text{ interacted}|}{|\text{Tweets posted by } \boldsymbol{u'} \text{ that } \boldsymbol{u} \text{ previously read}|} \quad (3)$$

Publishing rate of *u'* for keywords of *t* Tweets posted by a user have been used in several works to infer his expertise (Xu et al., 2016b, 2017). According to Xu et al. (2016b), if a user can express frequently his opinion on a topic, he is likely to have a strong knowledge of that topic. Since we represent each tweet's topics by keywords defined using *DBpedia Spotlight*, we propose to calculate this feature with equation 4. We recall that our aim is not to propose a method that infers users' expertise, but to study its contribution to rank news feeds updates.

$$f_4(\boldsymbol{u'},\boldsymbol{t}) = \frac{\sum_{i=1}^{nbk(\boldsymbol{t})} Post(\boldsymbol{u'}, k_i(\boldsymbol{t}))}{nbk(\boldsymbol{t}) \times nbp(\boldsymbol{u'})} \quad (4)$$

Such that:
— $k_i(\boldsymbol{t})$ is the i^{th} keyword of ***t***
— $nbk(\boldsymbol{t})$ is the number of keywords of ***t***
— $nbp(\boldsymbol{u'})$ is the number of tweets previously posted by ***u'***
— $Post(\boldsymbol{u'}, k_i(\boldsymbol{t}))$ is the number of times ***u'*** has previously posted $k_i(\boldsymbol{t})$

E.g., if ***t*** has 3 keywords k_1, k_2, k_4 and ***u'*** has previously posted 2 tweets t_1 and t_2 which have respectively the keywords k_1, k_2 and k_1, k_3, this value will be assigned to this feature:

$$f_4(\boldsymbol{u'},\boldsymbol{t}) = \frac{2+1+0}{3 \times 2} = 0.5 \quad (5)$$

Followers count / Followings count According to Pan et al. (2013), users who have authority on social media tend to have more followers than followings. The authors propose to calculate this feature by dividing the followers count of ***u'*** by his followings count.

Seniority Shen et al. (2013) state that senior users (whose accounts were created early) tend to have authority. We propose to calculate this feature with the following equation:

$$f_6(\boldsymbol{u'},\boldsymbol{t}) = \text{Year in which } \boldsymbol{t} \text{ was created} - \text{Year in which the account of } \boldsymbol{u'} \text{ was created} \quad (6)$$

Presence of hashtags A tweet with hashtags can provide more information and be of better quality. Indeed, Chen et al. (2012) state that the author spends time on tagging the tweet because he thinks it may be useful. We propose to calculate this feature with a boolean variable.

Presence of a URL As reported by De Maio et al. (2017), since a tweet is limited to 280 characters, it is quite common to see users include a URL to a website containing more details. According to the authors, a tweet with a URL can provide more information and be of better quality. They propose to calculate this feature with a boolean variable.

Presence of an image or a video Feng and Wang (2013) state that a tweet with an image or a video, in addition to the textual content, can provide more information and be of better quality. The authors propose to calculate this feature with a boolean variable.

4.3 Prediction model

First, considering each previously read tweet $t \in D(u)$ from the least recent to the most recent, we create in the form of input-output pairs, training database instances of each user $u \in S$, such that an input represents the features that may influence the relevance of t to u, and the output represents the relevance score $R(t,u)$ that measures the relevance of t to u. Then, we divide the training database of each user $u \in S$ into 2 sets: a training set of his prediction model for 70% of the first instances (the least recent ones); and a test set for 30% of the remaining instances (the most recent ones). The latter will be used to evaluate the prediction model.

In order to create a prediction model for each user $u \in S$, we train a *Decision Tree* classifier on his training set using the CART (Classification and Regression Trees) algorithm (Breiman et al., 1984) and entropy criterion for information gain. The purpose is to predict relevance scores for tweets unread by u using relevance classification rules learned from tweets previously read of the training set. The CART algorithm constructs binary trees using the feature and threshold that yield the largest information gain at each node (Loh, 2014). More details are provided in (Breiman et al., 1984). We choose *Decision Trees* as prediction models because they: (1) tend to produce powerful prediction models for binary classification problems; (2) don't require data preprocessing; (3) are fast to train and to predict outcomes; (4) require little data for training; (5) implicitly perform feature selection; and (6) allow us to compute importance scores of features in judging the relevance of tweets by beneficiary users, and therefore, study the importance of the author's expertise (computed using equation 4) compared to the other features considered in related works (Loh, 2014). Note that other supervised learning algorithms are also applicable and that it is out of the scope of this paper to compare them.

In our case, a *Decision Tree* of a user $u \in S$ is a flowchart-like structure in which each node represents a test on a feature that may influence relevance, each branch represents the outcome of the test, and each leaf node represents the class label (relevance score) of a tweet t. The paths from the root the to leaf represent relevance classification rules. An example of a classification rule from the *Decision Tree* of the *Twitter* user *NASAKepler* is presented in bold in Fig. 3: if the mentions of t are not relevant to u, and t has an image or a video, and the author's expertise score for the topics of t is greater than 0.04, then t is relevant to u.

5 Experimentation and results

In order to evaluate our approach and study the contribution of the author's expertise to rank news feeds updates, we describe in this section: (1) the dataset used in the experiments we performed; (2) the measures used to evaluate the performances; and (3) the obtained results.

5.1 Dataset

First, we randomly selected a subset ***S*** of 20 beneficiary users for whom we apply the proposed approach. Each user $\boldsymbol{u} \in \boldsymbol{S}$ has the following criteria: (1) interacts (retweet, reply, like) frequently with tweets from his news feed (interaction rate greater than or equal to 10%). This criterion is motivated by the use of the implicit training and evaluation method which assumes that user's interaction with a tweet implies its relevance; (2) English-speaking with a view to use the English version of *DBpedia Spotlight* which is the most complete. Then, using *Twitter Rest API*[7], we collected over a period of 10 months all data needed for the proposed approach. We point out that tweets' keywords were defined using *DBpedia Spotlight* with a *support* = 20 and a *confidence* = 0.32 (values determined after experiments).

On *Twitter*, it is impossible to retrieve users' news feeds directly and say, in case of non-interaction, if a given tweet was read by ***u*** (Chen et al., 2012). Therefore, in order to constitute the news feed of each user $\boldsymbol{u} \in \boldsymbol{S}$, we used a variant of the principle proposed by Feng and Wang (2013) to select, ***D(u)***, the subset of tweets posted by his followings that he may have read. We consider these tweets irrelevant. The variant is as follows: "Firstly, sort all the tweets posted by the followings of ***u*** in chronological order, from the least recent to the most recent. After that, for each tweet ***t*** with which ***u*** interacted (retweet, reply, like), keep the chronological session defined by: the tweet ***t***, the tweet before ***t*** and the tweet after ***t***. Finally, after deleting duplicates, sort the selected tweets again in chronological order, from the least recent to the most recent". Note that the application of our variant resulted in an interaction rate of approximately 35% for each beneficiary user and an average number of instances of 639 instances in the training database of each beneficiary user.

5.2 Measures

For each user $\boldsymbol{u} \in \boldsymbol{S}$, we first train a *Decision Tree* classifier on his training set (70% of the least recent instances in his training database, see Section 4.3). Then, in order to evaluate the performances of the proposed approach with his test set (30% of the most recent instances), and study the contribution of the author's topical expertise to rank news feeds updates, we define the following concepts (Sammut and Webb, 2011):

— TP (True Positive): number of relevant tweets correctly predicted relevant to ***u***
— TN (True Negative): number of irrelevant tweets correctly predicted irrelevant to ***u***
— FP (False Positive): number of irrelevant tweets incorrectly predicted relevant to ***u***
— FN (False Negative): number of relevant tweets incorrectly predicted irrelevant to ***u***

After that, we use the weighted *F1 score* measure, denoted *F*, and described in equation 7 (Sammut and Webb, 2011). It calculates *F1 score* for each class and find their average weighted by support (the number of true instances for each class label). This measure is appropriate to evaluate the classification performances since classes are slightly unbalanced (interaction rate of approximately 35% for each beneficiary user) and we are interested to measure the performances of predicting both of them: relevant tweets class and irrelevant tweets class.

$$F = \frac{(F_r \times (TP + FN)) + (F_i \times (TN + FP))}{TP + TN + FP + FN} \tag{7}$$

7. https://dev.twitter.com/rest/public

Such that:
— F_r is the standard *F1 score* for relevant tweets class
— F_i is the standard *F1 score* for irrelevant tweets class
Finally, for each user $\boldsymbol{u} \in \boldsymbol{S}$, we perform experiments with his test set using the *F score* and compare two approaches: our approach that leverages the author's topical expertise and a classical approach that is the same as ours, except that it does not leverage it. Moreover, to study the importance of the author's expertise compared to the other features considered in related works, we compute importance scores of features for $\boldsymbol{u}$. As described by Breiman and Cutler (2007), a feature's importance is computed as the normalized total reduction of the criterion brought by that feature and is also known as the *Gini importance*. Note that the higher value, the more important the feature is in judging the relevance of tweets by $\boldsymbol{u}$. More details about the importance of features are provided in (Breiman and Cutler, 2007).

5.3 Results

Firstly, experimental results of the comparison between our proposed approach and the classical one are presented in Table 2. The results show that our approach often succeeds in predicting relevance scores of tweets with an average *F score* of 71.48%. Furthermore, we point out that the latter applied to several beneficiary users gives excellent results with an *F score* of more than 80%: 90% for *SfNtweets*, 83.97% for *JHUBME*, 83.58% for *elonmusk*, and 80.62% for *demishassabis*. Indeed, the features and *Decision Trees* models we used are fully suitable to predict the relevance of tweets to these users. The results indicate also that the gain brought by our approach has reached up to +13.86%. This highlights that judging the author's expertise is crucial for maximizing the relevance of updates in news feeds. Moreover, we note that this feature makes a very significant contribution to several users, e.g.: +13.86% for *NASAKepler*, +9.35% for *microphilosophy*, +5.48% for *bamwxcom* and *sxbegle*, etc. This confirms that infer expertise enables beneficiary users, especially who attach importance to it, to catch up with the valuables and trustworthy tweets on a specific topic.

Secondly, we computed the average feature importance scores for all beneficiary users, which are presented in Fig. 4. The results show that all features we used are important in judging the relevance of tweets. The top feature is the feature f_3 (0.29) which measures interaction rate of $\boldsymbol{u}$ with ***u'***. Certainly, as stated by Vougioukas et al. (2017), if $\boldsymbol{u}$ found tweets posted by ***u'*** relevant in the past, he tend to have a strong social relationship and common interests with ***u'***, and may find his tweets relevant in the future. The results indicate also that the features f_5 and f_6 which measure the author's authority are very important (0.18 and 0.12 respectively). Indeed, these features are consistent with the observations in (Nagmoti et al., 2010) that state that if a user is important, i.e. has authority, then his tweets are also important. The third most important feature is the feature f_4 (0.15) which we used to infer the author's topical expertise from tweets he posted. This confirms two assumptions: (1) if an author can express frequently his opinion on a topic, he is likely to have a strong knowledge of that topic (Xu et al., 2016b); and (2) the author's expertise, which has not been considered in the academic community, neither in the industrial community, is very important in judging the relevance of tweets by beneficiary users. Moreover, we notice that the features f_1 and f_2, which measure respectively the relevance of the textual content of $\boldsymbol{t}$ and its mentions to $\boldsymbol{u}$, are surprisingly not the most important features (0.12 and 0.06 respectively). This proves that predicting relevance scores

User's screen name	Number of instances	F with expertise	F without expertise	C
Astro_Pam	837	67.9	66.84	**1.06**
bamwxcom	123	71.18	65.7	**5.48**
demishassabis	319	80.62	81.59	**-0.97**
elonmusk	303	83.58	83.69	**-0.11**
gwern	771	56.3	61.64	**-5.34**
homebrew	1115	69.11	67.66	**1.45**
jadelgador	1401	72	72.54	**-0.54**
JHUBME	516	83.97	83.97	**0**
microphilosophy	112	58.82	49.47	**9.35**
NASAKepler	86	79.63	65.77	**13.86**

User's screen name	Number of instances	F with expertise	F without expertise	C
PattiPiatt	1173	74.59	76.47	**-1.88**
rafat	782	76.67	76.77	**-0.1**
realDonaldTrump	140	66.67	68.8	**-2.13**
Red_or_MC1R	306	56.65	56.65	**0**
scimichael	2940	71.13	70.76	**0.37**
SfNtweets	198	90	85.4	**4.6**
SLSingh	560	63.31	59.26	**4.05**
sxbegle	382	72.34	66.86	**5.48**
TheRickDore	202	71.88	73.3	**-1.42**
USDISA	522	63.29	67.46	**-4.17**

C = *F* with expertise – *F* without expertise

TAB. 2: Experimental results.

is a difficult task because the most important features are not necessarily the most intuitive ones. We note lastly that the features f_7, f_8, and f_9 which measure tweet's quality are not very important (0.03, 0.03, and 0.02 respectively) since they are non-personalized features which does not take into consideration the preferences of each user (Feng and Wang, 2013).

Finally, we computed the features importance scores for each beneficiary user, which are listed in Table 3. We first note that features importance differs according to users, i.e. features that are important for one user are not necessarily important for another, e.g. social tie strength is very important for *elonmusk* when judging the relevance of tweets (0.54) but not for *NASAKepler* (0.03). Certainly, users' preferences are different and this highlights the importance of a personalized prediction model for each beneficiary user, like *Decision Trees*, which take into consideration individual preferences (Belkacem et al., 2016). We report also that the author's expertise is very important for several users, compared to the other features, in judging the relevance of tweets, e.g. 0.25 for *bamwxcom* and 0.19 for *microphilosophy*. This proves once again that information about the authors' expertise is very important for informing credibility judgments of tweets, especially for the users who attach importance to it.

Despite the improvements we presented, we notice that our approach shows limits for some users. Firstly, from Table 2, experimental results of the comparison, between our approach and the classical one show that our approach applied to a few users gives modest results with an *F score* of less than 60%: 58.82% for *microphilosophy*, 56.65% for *Red_or_MC1R*, and 56.3% for *gwern*. Since the number of instances in the training database seems to have no effect on results obtained (higher number of instances does not necessarily involves better results and

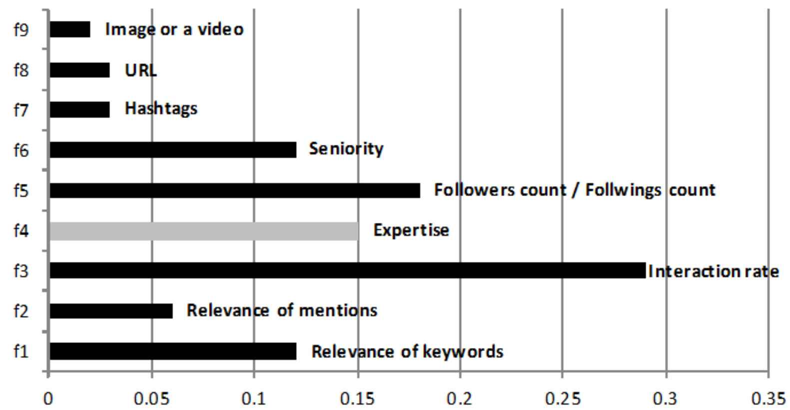


FIG. 4: Average importance of features for users.

vice-versa), e.g. results for *gwern* and *rafat*, we believe that these modest results are due to: (1) other features that we did not use but may influence relevance to these users e.g.: tweet's popularity (the number of interactions it received), tweet's length, and the number of lists to which the author has been added (Shen et al., 2013); (2) *Decision Trees* models which tend to overfit sometimes by creating over-complex trees which do not generalize the data well (Loh, 2014); and (3) the implicit training and evaluation method, and more especially the FP measure (irrelevant tweets incorrectly predicted relevant), may wrongly penalize our approach. Indeed, non-interaction is not always synonym of irrelevance. A user can, for example, find a tweet relevant and deliberately choose not to interact with it (Belkacem et al., 2016).

Secondly, the results of Table 2 show that the author's expertise decreases results to some users, e.g.: -5.34% for *gwern* and -4.17% for *USDISA*. Furthermore, from features importance scores for each beneficiary user, which are listed in Table 3, we note that the author's expertise is not very important for a few users, compared to other features, in judging the relevance of tweets, e.g. 0.12 for *elonmusk* and 0.13 for *jadelgador*. Since the number of instances seems to have no effect on the contribution of the author's topical expertise or its importance (higher number of instances does not necessarily involves a better contribution or a greater importance and vice-versa), e.g. contributions of the author's expertise for *homebrew* and *PattiPiatt* and its importance for *demishassabis* and *elonmusk*, we believe that these results are due to: (1) users information needs which are different on social media (Liao et al., 2012). Certainly, a minority may not give importance to the author's expertise when judging the relevance of tweets; and (2) other features that we did not use to infer expertise but may be important for these users when judging the relevance of tweets posted by an author e.g.: his biographical information, his social relationships, his list memberships, etc. (Xu et al., 2017).

6 Conclusion and future work

In this work, we proposed using a prediction model based on *Decision Trees* an approach that predicts the relevance of news feeds updates and leverages, in addition to the four types

User's screen name	Number of instances	*f1*	*f2*	*f3*	*f4*	*f5*	*f6*	*f7*	*f8*	*f9*
Astro_Pam	837	0.13	0.02	0.27	**0.13**	0.23	0.15	0.03	0.02	0.02
bamwxcom	123	0.05	0.08	0.07	**0.25**	0.2	0.18	0.04	0.04	0.08
demishassabis	319	0.02	0.08	0.37	**0.21**	0.14	0.16	0.01	0.01	0.0
elonmusk	303	0.06	0.0	0.54	**0.12**	0.13	0.09	0.01	0.02	0.04
gwern	771	0.12	0.0	0.33	**0.18**	0.18	0.1	0.02	0.05	0.01
homebrew	1115	0.08	0.01	0.41	**0.15**	0.13	0.11	0.06	0.02	0.02
jadelgador	1401	0.16	0.0	0.37	**0.13**	0.18	0.08	0.04	0.01	0.02
JHUBME	516	0.14	0.1	0.34	**0.12**	0.13	0.11	0.02	0.03	0.0
microphilosophy	112	0.11	0.0	0.25	**0.19**	0.2	0.17	0.07	0.01	0.0
NASAKepler	86	0.14	0.43	0.03	**0.1**	0.11	0.18	0.0	0.0	0.0
PattiPiatt	1173	0.06	0.0	0.53	**0.13**	0.09	0.12	0.02	0.02	0.04
rafat	782	0.08	0.03	0.35	**0.13**	0.15	0.12	0.03	0.07	0.03
realDonaldTrump	140	0.1	0.0	0.1	**0.17**	0.46	0.08	0.0	0.09	0.0
Red_or_MC1R	306	0.05	0.0	0.17	**0.16**	0.31	0.21	0.04	0.04	0.02
scimichael	2940	0.21	0.0	0.41	**0.14**	0.12	0.07	0.03	0.01	0.01
SfNtweets	198	0.24	0.1	0.24	**0.11**	0.1	0.15	0.05	0.01	0.01
SLSingh	560	0.14	0.01	0.45	**0.12**	0.18	0.05	0.02	0.03	0.01
sxbegle	382	0.19	0.1	0.24	**0.17**	0.15	0.08	0.03	0.03	0.01
TheRickDore	202	0.03	0.25	0.13	**0.19**	0.22	0.1	0.05	0.02	0.01
USDISA	522	0.1	0.09	0.17	**0.17**	0.27	0.1	0.03	0.04	0.02
USDISA	522	0.1	0.09	0.17	**0.17**	0.27	0.1	0.03	0.04	0.02

TAB. 3: Importance of features for users

of features used in related works: (1) the relevance of the update's content to the beneficiary's interests; (2) the social tie strength between the beneficiary and the update's author; (3) the author's authority; and (4) the update's quality, the author's expertise that we inferred from tweets he posted. Experimental results on *Twitter* highlight that judging expertise is crucial for maximizing the relevance of updates in news feeds. However, given the limits we identified for some users, efforts must still be made to improve the proposed approach.

For further works, we first intend to make a deeper analysis with a greater number of users and compare the proposed approach with other ones from related works. Moreover, to improve the results, we project to extend it by using other features that may infer expertise. Finally, we plan to use other prediction models that address the limits of *Decision Trees*.

References

Agarwal, D., B.-C. Chen, Q. He, Z. Hua, G. Lebanon, Y. Ma, P. Shivaswamy, H.-P. Tseng, J. Yang, and L. Zhang (2015). Personalizing linkedin feed. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1651–1660. ACM.

Belkacem, S., K. Boukhalfa, and O. Boussaid (2016). News feeds triage on social networks: A survey. In *Proceedings of The 2nd International Conference on Computing Systems and Applications (CSA)*, pp. 34–43.

Belkacem, S., K. Boukhalfa, and O. Boussaid (2017). Tri des actualités sociales: Etat de l'art et Pistes de recherche. In *Journées francophones sur les Entrepôts de Données et l'Analyse en ligne (EDA)*, Volume 13, pp. 85–100. Revue des Nouvelles Technologies de l'Information.

Berkovsky, S. and J. Freyne (2015). Personalised Network Activity Feeds: Finding Needles in the Haystacks. In *Mining, Modeling, and Recommending'Things' in Social Media*, pp. 21–34. Springer.

Breiman, L. and A. Cutler (2007). Random forests-classification description: Random forests, Available at http://www.stat.berkeley.edu/~breiman/RandomForests/cc_home.htm [Accessed September 20, 2017].

Breiman, L., J. Friedman, C. J. Stone, and R. A. Olshen (1984). *Classification and regression trees*. CRC press.

Chen, K., T. Chen, G. Zheng, O. Jin, E. Yao, and Y. Yu (2012). Collaborative personalized tweet recommendation. In *Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval*, pp. 661–670. ACM.

De Maio, C., G. Fenza, M. Gallo, V. Loia, and M. Parente (2017). Time-aware adaptive tweets ranking through deep learning. *Future Generation Computer Systems*.

Feng, W. and J. Wang (2013). Retweet or not?: personalized tweet re-ranking. In *Proceedings of the sixth ACM international conference on Web search and data mining*, pp. 577–586. ACM.

Kuang, L., X. Tang, M. Yu, Y. Huang, and K. Guo (2016). A comprehensive ranking model for tweets big data in online social network. *EURASIP Journal on Wireless Communications and Networking 2016*(1), 1.

Liao, Q. V., C. Wagner, P. Pirolli, and W.-T. Fu (2012). Understanding experts' and novices' expertise judgment of twitter users. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 2461–2464. ACM.

Loh, W.-Y. (2014). Fifty years of classification and regression trees. *International Statistical Review 82*(3), 329–348.

Nagmoti, R., A. Teredesai, and M. De Cock (2010). Ranking approaches for microblog search. In *Web Intelligence and Intelligent Agent Technology (WI-IAT), 2010 IEEE/WIC/ACM In-*

ternational Conference on, Volume 1, pp. 153–157. IEEE.

Pan, Y., F. Cong, K. Chen, and Y. Yu (2013). Diffusion-aware personalized social update recommendation. In *Proceedings of the 7th ACM conference on Recommender systems*, pp. 69–76. ACM.

Sammut, C. and G. I. Webb (2011). *Encyclopedia of machine learning*. Springer Science & Business Media.

Shen, K., J. Wu, Y. Zhang, Y. Han, X. Yang, L. Song, and X. Gu (2013). Reorder user's tweets. *ACM Transactions on Intelligent Systems and Technology (TIST) 4*(1), 6.

Vougioukas, M., I. Androutsopoulos, and G. Paliouras (2017). Identifying Retweetable Tweets with a Personalized Global Classifier. *CoRR abs/1709.06518*.

Wagner, C., V. Liao, P. Pirolli, L. Nelson, and M. Strohmaier (2012). It's not in their tweets: Modeling topical expertise of twitter users. In *Privacy, Security, Risk and Trust (PASSAT), 2012 International Conference on and 2012 International Confernece on Social Computing (SocialCom)*, pp. 91–100. IEEE.

Wei, W., G. Cong, C. Miao, F. Zhu, and G. Li (2016). Learning to find topic experts in Twitter via different relations. *IEEE Transactions on Knowledge and Data Engineering 28*(7), 1764–1778.

Xu, Y., D. Zhou, and S. Lawless (2016b). Inferring Your Expertise from Twitter: Integrating Sentiment and Topic Relatedness. In *Web Intelligence (WI), 2016 IEEE/WIC/ACM International Conference on*, pp. 121–128. IEEE.

Xu, Y., D. Zhou, and S. Lawless (2017). Inferring your expertise from Twitter: combining multiple types of user activity. In *Proceedings of the International Conference on Web Intelligence*, pp. 589–598. ACM.

Xu, Z., Y. Liu, N. Yen, L. Mei, X. Luo, X. Wei, and C. Hu (2016a). Crowdsourcing based description of urban emergency events using social media big data. *IEEE Transactions on Cloud Computing*.

Résumé

En raison de la grande quantité d'informations produites sur les réseaux sociaux, les utilisateurs se trouvent submergés par les actualités affichées chronologiquement dans leur fil d'actualité. De plus, la plupart d'entre elles sont considérés non pertinentes. Le tri des fils d'actualité par ordre de pertinence est proposé pour aider les utilisateurs bénéficiaires à rattraper rapidement les actualités pertinentes. Quatre types de caractéristiques sont principalement utilisées pour prédire la pertinence: (1) la pertinence du contenu de l'actualité pour les intérêts du bénéficiaire; (2) la force de la relation sociale entre le bénéficiaire et l'auteur de l'actualité; (3) l'autorité de l'auteur; et (4) la qualité de l'actualité. Dans ce travail, nous proposons une approche qui exploite un autre type de caractéristique qui est l'expertise de l'auteur pour la thématique de l'actualité. Les résultats expérimentaux sur *Twitter* soulignent que le jugement de l'expertise est crucial pour maximiser la pertinence dans les fils d'actualité.