Is Readability a Valuable Signal for Hashtag Recommendations?

Ion Madrazo Azpiazu
Computer Science Dept.
Boise State University
Boise, Idaho, USA
ionmadrazo@boisestate.edu

Maria Soledad Pera
Computer Science Dept.
Boise State University
Boise, Idaho, USA
solepera@boisestate.edu

ABSTRACT

We present an initial study examining the benefits of incorporating readability indicators in social network-related tasks. In order to do so, we introduce TweetRead, a readability assessment tool specifically designed for Twitter and use it to inform the hashtag prediction process, highlighting the importance of a readability signal in recommendation tasks.

CCS Concepts

•**Human-centered computing** → **Social recommendation; Social networks;**

Keywords

Hashtag Recommendation; Readability

1. INTRODUCTION

Readability is a measure of the ease with which a text can be read. Usually represented by a number, it is an indicator used by teachers to classify and find appropriate resources for students. Several studies have demonstrated the benefits of using readability indicators in educational-related applications, such as book recommendation, text simplification, or automatic translation. However, applying readability indicators outside this environment remains relatively unexplored. Social networks could benefit from readability assessment. Twitter is a social network where users and texts are the main focus. For this reason, it is natural to think that for Twitter the ease with which a tweet can be understood by a user may affect his interest in it, and therefore influence actions taken, such as re-tweeting, giving a like or replying to the tweet.

The authors of [6] examined the degree to which the age of a user, a feature strongly correlated with readability, influences who people follow on Twitter, and demonstrated that Twitter users have a higher chance to follow people of similar age. Using standard readability measures in text from Twitter, which constrains tweets to be of at most 140 characters in length, is not a trivial task. The lack of structure and shortness of those texts make standard natural language analysis techniques inefficient. With that in mind, we developed TweetRead, a novel readability assessment tool specifically designed for tweets. TweetRead takes advantage of social information, such as hashtags or mentions, for predicting the text complexity levels of tweets. Furthermore, in order to highlight the usefulness of such a tool in social networking environments, we developed a simple, yet effective, hashtag recommendation strategy that takes advantage of TweetRead-generated complexity levels of tweets to inform the hashtag recommendation process.

WOODSTOCK '97 El Paso, Texas USA

DOI: 10.475/123_4

2. TWEETREAD

TweetRead's goal is to estimate readability of any given tweet T. However, traditional Natural language processing techniques are known not to work properly on short and unstructured text such us the one contained in tweets. Therefore, TweetRead avoids using traditional NLP strategies and relies on simpler models based on content and tweet-specific information. TweetRead is based on a logistic regression technique[1] that fuses simple indicators describing T from different perspectives and determines its text complexity. The indicators considered by TweetRead are described below:

- **Extended Flesch.** Flesch reading ease formula) see Figure 3) is widely used by teachers for estimating readability of texts for their students. However, this formula requires the input text to be sufficiently long to give accurate predictions. Given that tweets are only 140 character long, the accuracy of Flesch is low for predicting the readability of tweets, as shown in our assessment in Figure 2. To address the issue of the textual content, we explore various strategies that consider tweets that may have similar readability levels. As the number of tweets considered increases, so does the amount of text we have, increasing with it the expected precision of Flesch. We considered 3 tweet groups that may serve as indicators of readability:

[1]We empirically verified that among numerous supervised techniques, logistic regression was the most promising one.

- **Other tweets by the user.** A user is expected to be consistent in his writing in terms of readability. Therefore, to determine the readability of a tweet, it may be useful to take advantage of other tweets written by the same user. For this, the average Flesch of all tweets of the user is used as readability predictor.
- **Tweets by users mentioned** Homophily stands for the nature of users to relate to similar users. It is a principle that widely manifests in social networks, in terms of aspects such as age, hobbies or profession of users. Following this principle, our hypothesis for this feature is that users with same readability also tend to relate to each other more frequently that random users. Therefore, we consider flesch readability of tweets written by users mentioned in the tweet as indicator of the readability of the tweet.
- **Tweets that contain same hash-tags** Following the same principle homophily, users we hypothesize that users with similar readability also share the hashtags they usually user. Therefore, we consider the Flesch readability of all tweets containing the same hash-tags that the tweet contains.

- **N-gram models.** Studies [6] demonstrate that users of same age, tend to use similar terminology when writing tweets. Considering that age is a very correlated metric to readability, we take advantage of these writing trends for readability prediction. For doing so, we create one feature $fgram_r$ for each existing readability group r. Each of feature $fgram_r$ is intended to measure the similarity of term distribution between the collection d_r of all the tweets of readability r and the given tweet. For doing so, we take advantage of the well known tfIdf formula, considering each d_r a document containing all the tweets of readability r and D the collection of all tweets.

$$tfidf(t,d,D) = tf(t,d) \times idf(t,D) \tag{1}$$

$$tf(t,d) = f(t,d)| \tag{2}$$

$$idf(t,D) = \log \frac{|D|}{|\{d \in D : t \in d\}|} \tag{3}$$

  The $fgram_r$ similarity of a tweet to the group r is computed as the sum of all tfIdf values of all the terms contained on it.

- **Metatags** Tweets contain more information than just raw text, they also contain Twitter specific information such as hash-tags or mentions and social information such as emoticons. In order to take advantage of this information we also consider frequencies of this tags as predictors for readability.

Unlike traditional readability formulas that tend to map readability levels with school grades, to tailor TweetRead to the Twittersphere, we consider six levels of text complexity following Levinston's [3] adult development stages.

$$Flesch = 206.835 - 1.015(\frac{totalwords}{totalsentences}) - 84.6(\frac{totalsyllables}{totalwords})$$

Figure 1: Flesch reading ease

3. HASHTAG RECOMMENDATION

Hashtags are character strings used to represent concepts on Twitter, starting with a # symbol. They are a core Twitter feature and serve classification and search purposes. Their unrestricted nature, however, creates difficulties, including the fact that the same concept can be represented by different hashtags, hindering the search process of a concept [5]. For example, tweets related to the Monaco Formula 1 Grand Prix can be searched using #monacoGP, #monacoF1GP or #monacoF1 retrieving different results. Hashtag recommendation aims at identifying suitable hashtags a user can include in his tweet to reduce the space of tags generated [5] and facilitate the ease with which he and other users can locate the corresponding tweet.

Given that (i) the scope of this paper is to validate the importance of considering a text complexity signal to enhance a recommendation task and (ii) multiple and increasingly complex systems have been developed for hashtag recommendation [2], we base our study on an existing framework for hashtag recommendation presented in [5]. Given a tweet T, the proposed framework identifies existing hashtags to recommend by following two major steps: (1) generate candidate hashtags by recommending hashtags present in similar tweets, using tf-idf a based similarity and (2) rank hashtags from retrieved candidate tweets using different strategies. The strategies presented in [5] include:

- **Similarity**. Prioritizes hashtags included on tweets that have the closes similarity to T, as estimated using the well-known tf-idf similarity measure.
- **Global popularity**. Prioritizes hashtags based on their respective frequency of occurrence on Twitter.
- **Local popularity**. Prioritizes hashtags based on their frequencies of occurrence among the tweets retrieved in response to T.

We enhance the proposed strategies by taking advantage of TweetRead, as follows:

- **TweetRead**. Prioritizes candidate hashtags that have the same or similar text complexity (estimated using TweetRead) with respect to T.
- **PopularityTweetRead**. Prioritizes hashtags based on their frequencies of occurrence among tweets whose readability level is estimated to match T's.
- **SimilarityTweetRead**. Prioritizes candidate hashtags based on their respective ranking scores computed using Similarity only on tweets whose readability level is estimated to match T's .

	Similarity	GlobalPopularity	LocalPopularity	TweetRead	SimilarityTweetRead	PopularityTweetRead
Mean Reciprocal Rank	0.47	0.19	0.40	0.23	0.52	0.50
First Relevant doc. on avg.	2.14	5.14	2.51	4.39	1.93	2.02

Table 1: Comparison of hash-tag recommendation strategies

Flesch	Spache	TweetRead
27%	31%	81%

Table 2: Performance evaluation of TweetRead vs. baselines.

4. INITIAL ASSESSMENT

In this section, we discuss an initial evaluation on TweetRead, as well as its applicability for suggesting hashtags.

4.1 TweetRead

Given that readability of social content is an unexplored area, benchmark datasets that can be used for evaluation purposes are unavailable. For this reason, we built our own dataset. We initially gathered 172M tweets over an 8-month period using Twitter streaming API. For the purpose of this experiment we assume that the age of people exactly corresponds to their readability level, and that each tweet written by a user will have the same readability level as its author. With that in mind, we followed the framework presented in [6], which examines patterns such as "happy xth birthday", for determining the age of Twitter users. In doing so, we eliminated from our dataset, users (and their corresponding tweets) from whom age could not be determined. Thereafter, we grouped labeled tweets into 6 age groups, which translates into a uniformly distributed dataset of 22k tweets with their corresponding readability levels. We followed a 10-cross-fold validation strategy and measured the accuracy of the predicted readability levels with respect to the ground truth. As shown in Table 2, TweetRead significantly outperforms the baselines considered for this assessment: Flesch [1] and Spache [4], which are two well-known, traditional readability measures. The reported results demonstrate the need for readability strategies that examine information beyond standard text analysis, if they are meant to be successfully used in the social networking context.

4.2 Hashtag recommendation

For evaluating the strategies for hashtag recommendation presented in Section 3, we used the aforementioned dataset. We treated the hashtag of each corresponding tweet as the ground truth. In other words, for each tweet T, we generated the corresponding top-N hashtag recommendations and considered relevant the ones matching the hashtags in T. As in [5], we used the recall measure to evaluate performance and determine to which extend the correct hashtags were recommended within the top N generated suggestions. As shown in Figure 2, even if readability on its own is not a sufficient factor to suggest hashtags, when combined in-tandem with other content-based and/or popularity strategies, it leads to the improvement of the overall hashtag recommendation process.

To further highlight the improvement achieved by the use of readability, we computed the Mean Reciprocal Rank (MRR) metric for each of the ranking strategy considered. This metric represents in which rank is the first relevant document found in average. We consider hash-tags as documents and the only relevant hash-tag is the one that appears in the ground-truth for the input tweet.

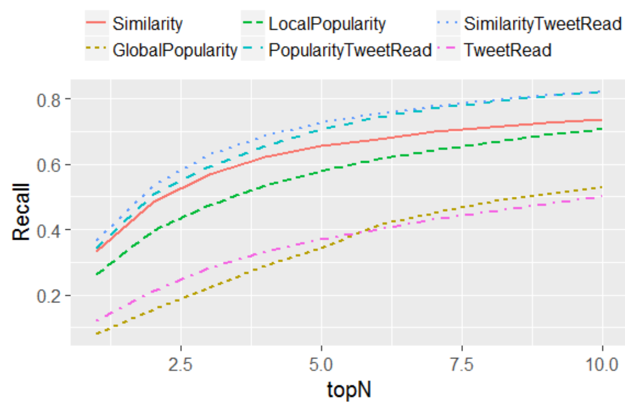

Figure 2: Hashtag recommendation assessment.

$$MRR = \frac{1}{|Q|}\sum_{i=1}^{|Q|}\frac{1}{rank_i}$$

Figure 3: Mean Reciprocal Rank (MRR), where Q represents the tested tweets and $rank_i$ the position of the first relevant hash-tag

The results, show that on average the first relevant hashtag is found at position 1.93 on average for the strategy that combines readability and similarity, being the one that best results achieved in terms of MRR. The best non-readability based strategy is the one that relies on similarity which on average retrieves documents on 2.14 position.

5. CONCLUSION AND FUTURE WORK

In this paper, we presented TweetRead, a novel readability assessment tool specifically designed to predict the readability of tweets. We also discussed the initial study conducted to demonstrate the benefit of using a readability signal in the hashtag recommendation task, which yielded promising results. In the future, we plan to explore other applications of readability in social networks, such as user recommendation, advertisement targeting or re-tweet prediction. We will also explore techniques to further enhance TweetRead and adapt it to other social networks beyond Twitter.

6. REFERENCES

[1] R. Flesch. A new readability yardstick. *Journal of Applied Psychology*, 32(3):221, 1948.

[2] F. Godin, V. Slavkovikj, W. De Neve, B. Schrauwen, and R. Van de Walle. Using topic models for twitter hashtag recommendation. In *WWW*, pages 593–596. ACM, 2013.

[3] D. J. Levinson. A conception of adult development. *American psychologist*, 41(1):3, 1986.

[4] G. Spache. A new readability formula for primary-grade reading materials. *The Elementary School Journal*, 53(7):410–413, 1953.

[5] E. Zangerle, W. Gassler, and G. Specht. Recommending #-tags in twitter. In *SASWeb 2011*, volume 730, pages 67–78, 2011.

[6] J. Zhang, X. Hu, Y. Zhang, and H. Liu. Your age is no secret: Inferring microbloggers' ages via content and interaction analysis. In *AAAI ICWSM*, 2016.