RecSys Challenge 2014: learning to rank

Daniele Loiacono
Politecnico di Milano, DEIB
p.zza Leonardo da Vinci 32
Milan, Italy
daniele.loiacono@polimi.it

Andreas Lommatzsch
TU Berlin, DAI-Labor
Ernst-Reuter-Platz 7
D-10587 Berlin, Germany
andreas@dai-lab.de

Roberto Turrin
Moviri S.p.A.
ContentWise R&D
via Schiaffino, 11 - Milan, Italy
roberto.turrin@moviri.com

ABSTRACT

The challenge focuses on the engagement generated by the tweets posted by the users of the IMDb application for smartphones. Such engagement depends on attributes concerning: the user who posts the message (e.g., his role in the social network), the tweet content (e.g., the rating), and the movie object of the tweet (e.g., the popularity of the movie).

We faced the challenge splitting the work into three sequential stages: (i) data enrichment, (ii) knowledge extraction, and (iii) engagement prediction.

Initially, we enriched the dataset with additional movie attributes extracted from IMDb and Freebase.

Successively, we analyzed the statistics of the main tweet attributes. Therefore, observing that the dataset is unbalanced towards non-engaging tweets, we re-defined the challenge ranking task as a binary classification problem; we used machine learning techniques to extract some knowledge from the data. In addition, we experimented with a linear prediction model tuned as an optimization task.

Finally, on the basis of the main outcomes of the previous analyses, we defined a predictor on attributes such as: user rating score, the presence of mentions, and whether the tweet is a retweet or it has already been retweeted. Such predictor led to a nDCG@10 equals to 0.835187.

1. INTRODUCTION

Twitter is a micro-blogging service used by millions of users. The service basically allows users to communicate via 140-character messages – referred to as tweets. Users can either post a new tweet or comment, retweet, or mark an existing tweet as favorite. Furthermore, users can follow other users (i.e., being notified of the tweets they publish), so that they eventually set up a complex social network [6].

The attitude of users to participate to the service – e.g., with tweets, retweets, and favorite bookmarking – is strictly related to the engagement level of the user and it represents a key factor for the success of the social network, granting that users' messages propagate [12] and reach interested users.

RecSysTv 2014 Foster City, Silicon Valley USA

It is worth noting that the amount of effort required for a user to reply to a tweet with a new comment is definitely higher with respect to a simple retweet or favorite mark. In fact, when commenting a tweet a person has to think about something concise to say so that it fits the 140-character length. Curiously, Zarrella's report[1] indicates a weak correlation between retweets and clicks, i.e., it is often the case that users retweet tweets they haven't even read.

In this scenario, the challenge[2] driving this work focuses on the tweets published by the users of the IMDb[3] (Internet Movie Database) application for smartphones , that have the option to share on the Twitter streamline the rating given to a certain movie. The challenge goal consists of estimating which of those tweets are the most *engaging*, defined in terms of how much engagement they will generate among users, i.e., whether and how they will be retweeted and marked as favorite.

Firstly, we enriched the dataset with further information retrieved from two external sources: IMDb and Freebase.

Therefore, we analyzed the available data, both using simple statistics and by exploiting several machine learning techniques and linear models.

Finally, we implemented a predictor based on the most relevant attributes as inferred by the previous analyses (e.g., user rating). Such predictor led to an nDCG@10 equals to 0.835187.

Outline.

The rest of the paper is organized as follows. Section 2 presents some works addressing the problem of measuring engagement. Section 3 describes the task object of the challenge, while Section 4 explains how we enriched the original data with external sources. The dataset attributes have been analyzed in Section 5 and processed with some machine learning approaches in Section 6. Section 7 presents the ultimate predictor we have implemented. Finally, Section 8 draws the conclusion.

2. BACKGROUND

The success of a social network strongly depends on the participation of users within the social community. For such a reason, any social network service implements some mechanisms to involve users [10]. The nature of contributions is

[1]http://blog.hubspot.com/blog/tabid/6307/bid/33815/New-Data-Indicates-Twitter-Users-Don-t-Always-Click-the-Links-They-Retweet-INFOGRAPHIC.aspx
[2]http://2014.recsyschallenge.com/
[3]http://www.imdb.com

usually voluntary and the choice about (i) whether to contribute or not and (ii) how to participate into the social model (e.g., which tweets retweet) is left to the users. The user participation is part of a wider topic known as engagement [9], studied by several disciplines such as play theory, information interaction, and flow theory (e.g., [3]).

Several works have addressed the problem of measuring the engagement from the user perspective, i.e., how much the user influences the community (e.g., [2]). A similar problem has been faced by Lee et al. [7], which try to model user's willingness and readiness to retweet information. They explore dimensions such as social network properties (i.e., user social interaction), the content of tweets, and a wait-time model that estimates the time when the user is expected to retweet a tweet. In [8], Macskassy and Michelson look at what drives information diffusion processes in social media. Their experiments on a set of Twitter users revealed that the majority of retweet behavior was explained by a content-based propagation model (i.e., based on the topic of the tweet).

Note that, differently from existing approaches, our task does not look at the problem from the point of view of the user (i.e., how much a user will be engaged), but the interest is on the tweet (i.e., how much a tweet will engage users).

3. TASK DEFINITION

The task of the challenge focuses on forecasting the engagement of users' tweets. More precisely, given a set of tweets of several users, participants are required to rank the tweets of each user according to the engagement that each tweet will receive. The engagement is defined as the sum between the number of users that retweet it and the number of users that mark it as favorite. The capability of sorting the tweets of a user is measured in terms of the quality metric nDCG@10 (normalized Discounted Cumulative Gain[11], computed on the top 10 elements), which measures the ranking quality. Finally, the overall quality is obtained by averaging the nDCG@10 values of each single user. In the following we will refer to the aggregate metric simply as nDCG. We can observe that:

- each user has the same impact on the overall performance, disregarding of how many tweets he/she posted
- the influence of the single user within the social network does not affect the nDCG@10 for that user. In fact, while the absolute engagement of each tweet (e.g., how many retweets it receives) is likely to depend on the role of the user in the social community [4] (e.g., an influencer user), ranking his/her tweets does only consider the relative engagement, not the absolute value.

For each tweet, the dataset reports the user identifier, the tweet identifier, the timestamp of the message, the rating shared by the user (in a 1-to-10 rating scale), and a data structure containing, among the others, information such as: whether the tweet has mentions (i.e., it cites some users), if it is a retweet, the IMDb url of the movie, the tweet language, some properties of the user (e.g., the number of followers and the number of user friends) etc.

The **training set** consists of 170,285 tweets posted by 22,079 users about 13,618 movies, while the **test set** made available to practitioners to experiment with their predictor consists of 21,285 tweets posted by 5,717 about 4,226 movies.

4. DATA ENRICHMENT

The dataset provided by the organizer of the challenge has been *enriched* with additional information either inferred from the original data – as a transformation of one or multiple attributes – or retrieved from external sources.

For instance, starting from the tweet timestamp, we computed further temporal attributes, such as: day of week, month, and time of day of the tweet, number of days between the original tweet and a retweet.

Retrieving knowledge from external data sources requires entities to be reconciled, i.e., that the same entity must be recognized as the same object in both sources. In our scenario, each tweet contains the IMDb url of the rated movie, and we used this attribute to reconcile the movie two external sources: IMDb and Freebase (FB)[4]. In particular, we parsed the related IMDb Web page to get further data about the movie and we used the Freebase APIs both to reconcile the movie and to extract additional information from such knowledge base, retrieving most movies (94%). Among the others, the most promising properties we have added to the dataset are:

- average IMDb rating and number of IMDb raters (IMDb)
- type, category, and genre, runtime and censure rating of the movie (FB)
- movie release date (FB) and the related number of days between movie release and tweet
- main movie language and main country (FB)
- number of won awards and estimated budget (FB)
- if the movie was adapted from a book (FB)
- number of festivals the movies attended (FB)
- if the movie is part of a series or a prequel/sequel (FB)

The enriched dataset is the basis for our statistical dataset analysis.

5. DATASET ANALYSIS

In this section we analyze the available data, focusing on the properties of the training set (enriched with the additional information presented in Section 4).

We can observe from Table 1 – that shows the number of tweets per month – that only 4.8% of tweets (8,178 out of 170,285) have a positive engagement, i.e., the dataset is unbalanced towards non-engaging tweets.

Table 2 provides some statistics about the distribution of the number of tweets posted by users. About 43% of users (9,477 out of 22,079) posted a single message, while only 20% of users (4,577 out of 22,079) posted some engaging tweets.

In addition, Table 3 reports the distribution of the number of distinct values of engagement per user. For example, if we consider only engaging tweets, 4174 users have posted tweets with the same value of engagement, while there exist 320 users whose posted tweets have two distinct values of engagement. We can note that most users have posted tweets with a single level of engagement.

We approached the problem studying what characterizes engaging tweets with respect to non-engaging tweets. Let us firstly observe that engaging tweets tend to be related

[4]http://www.freebase.com/

month	overall #tweets	proportion of tweets non-engaging	engaging
Feb 2013	220	96.82%	3.18%
Mar 2013	15826	96.40%	3.60%
Apr 2013	14232	96.15%	3.85%
May 2013	14757	95.60%	4.40%
Jun 2013	15199	95.49%	4.51%
Jul 2013	15610	94.44%	5.56%
Sep 2013	21888	95.98%	4.02%
Aug 2013	18029	95.45%	4.55%
Oct 2013	16105	94.59%	5.41%
Nov 2013	15532	94.17%	5.83%
Dec 2013	16749	94.14%	5.86%
Jan 2014	6138	93.60%	6.40%
	170285	95.20%	4.80%

Table 1: Number of tweets per month distinguishing between: non-engaging, engaging, and overall.

#tweets	#Users all	engaging
1	9477	3394
2-5	6406	1020
6-10	2319	106
11-20	1837	35
21-50	1482	16
51+	562	52
	22079	4577

Table 2: Number of tweets posted by users, distinguishing between all tweets and only tweets that generate engagement (engaging). Note that 17502 users out of 22079 (79%) have no engaging tweets.

#Engagement scores	#Users all	engaging
1	19175	4174
2	2551	320
3	278	57
4	50	14
5+	25	12
	22079	4577

Table 3: Number of engaging levels per user (i.e., number of distinct values of engagement).

to higher user ratings, as shown in Table 4. We proceeded studying the correlation between the main tweet attributes and the fact that the tweet generates some engagement. The main results are reported in the Tables 5.

rating value	overall #ratings	proportion non-engaging	engaging
1	1988	91.10%	8.90%
2	1885	96.23%	3.77%
3	3134	96.90%	3.10%
4	5798	97.55%	2.45%
5	11853	97.49%	2.51%
6	22488	98.08%	1.92%
7	38122	96.83%	3.17%
8	40773	95.01%	4.99%
9	23695	92.24%	7.76%
10	20476	90.89%	9.11%
	170212	95.20%	4.80%

Table 4: Distribution of ratings. Non-valid ratings have been excluded.

Table 5(a) reports the correlation between the main nominal attributes and the engagement: language of the tweet, language of the movie (curiously, Arabic movies tend to have a high engagement), month of movie release, and weekday of movie release (Saturday's tweets result above average engagement).

In Table 5(b) we studied the impact of some numeric and boolean attributes on the engagement of tweets. Separately between engaging and non-engaging tweets, we reported either the average value of tweets (numeric attributes) or the percentage of tweets having such property (boolean attributes). We also computed the information gain of each attribute. The attributes the most relevant to differentiate the two classes of tweets are: the presence of mentions in the tweet and whether the tweet is a retweet itself or it has already been retweeted. Indeed, if a tweet has mentions or it is a retweet, it is very likely to generate engagement. Other significant attributes are: the user rating, the IMDb average rating and popularity, the difference between the user rating and the IMDb average rating, the number of won awards, and the number of attended festival. Furthermore, more recent tweets tend to have a higher engagement with respect to old tweets.

6. DATA KNOWLEDGE EXTRACTION

In this section we experiment with several machine learning techniques to extract some further knowledge from the existing data.

6.1 Ranking as a binary prediction problem

The analysis performed in Section 5 pointed out that 95% of tweets generate no engagement. Thus, we opted to redefine the original problem consisting of ranking the tweets of a user as a *binary* problem. Practically, we approached the problem by building classifier able to distinguish whether a tweet is engaging (i.e., with an engagement greater than 0) or not (i.e., with an engagement equals to 0).

Firstly, we measured the upper bound performance of the binary approach by creating an *ideal* binary classifier, i.e., a sort of omniscient classifier that classifies a tweet as engaging if its engagement is greater than one. The nDCG of such classifier applied to the test set results equal to 0.987657; worth noting that the small ratio of error (i.e., the nDCG is lower than 1) depends on the fact that such binary classifier can distinguish engaging and non-engaging tweets, but it is not capable to correctly sort multiple engaging tweets of the same user.

Therefore, starting from the original training set, we created a binary training dataset where tweets have been labeled either as *engaging* (if the engagement value is positive) or *non-engaging* (if there is no engagement). Such dataset has been balanced, randomly sub-sampling the non-engaging tweets, i.e., we randomly picked up (without replacements) a subset of about 8000 (i.e., about 5%) non-engaging tweets, so that the resulting training dataset was composed by an even number of positive (engaging) and negative (non-engaging) samples.

Finally, we experimented with several supervised learning techniques (using the toolkit Weka [5]), the most encouraging being: Naïve Bayes (NB), Bayes Networks (BN), and J48

(a) Nominal attributes

attribute	source	value	overall #tweets	proportion of tweets non-engaging	proportion of tweets engaging	average engagement	information gain
Tweet lang	tweet	**Arabic**	1125	64.62%	35.38%	1.01	0.0268
		English	153342	95.53%	4.47%	0.22	
		Vietnamese	663	94.87%	5.13%	0.19	
		French	511	83.37%	16.63%	0.19	
		Spanish	2379	93.74%	6.26%	0.07	
		Slovak	4650	96.15%	3.85%	0.05	
		Danish	804	96.27%	3.73%	0.04	
		Italian	1033	96.22%	3.78%	0.04	
		Slovene	717	97.49%	2.51%	0.03	
Month	inferred	**Jan**	6138	93.60%	6.40%	0.09	0.0020
		Feb	220	96.82%	3.18%	0.03	
		Mar	15826	96.40%	3.60%	0.05	
		Apr	14232	96.15%	3.85%	0.05	
		May	14757	95.60%	4.40%	0.07	
		Jun	15199	95.49%	4.51%	0.06	
		Jul	15610	94.44%	5.56%	1.60	
		Aug	18029	95.45%	4.55%	0.08	
		Sep	21888	95.98%	4.02%	0.08	
		Oct	16105	94.59%	5.41%	0.08	
		Nov	15532	94.17%	5.83%	0.08	
		Dec	16749	94.14%	5.86%	0.09	
Weekday	derived	Sun	35151	95.46%	4.54%	0.08	0.0004
		Mon	23405	95.26%	4.74%	0.09	
		Tue	18938	95.45%	4.55%	0.15	
		Wed	19245	95.23%	4.77%	0.08	
		Thu	19870	94.89%	5.11%	0.09	
		Fri	22948	94.88%	5.12%	0.09	
		Sat	30728	95.12%	4.88%	0.74	
Movie lang	Freebase	Arabic	506	90.91%	9.09%	0.29	0.0268
		English	148056	95.46%	4.54%	0.23	
		French	2423	94.35%	5.65%	0.07	
		German	1608	89.05%	10.95%	0.16	
		Hindi	1127	94.50%	5.50%	0.07	
		Italian	705	95.46%	4.54%	0.06	
		Japanese	1521	95.46%	4.54%	0.07	
		Korean	601	94.68%	5.32%	0.07	
		Latin	1339	94.32%	5.68%	0.07	
		Russian	651	95.39%	4.61%	0.05	
		Spanish	4098	92.90%	7.10%	0.10	

(b) Numeric and boolean attributes

attribute	source	Engaging tweet non-engaging	Engaging tweet engaging	overall	information gain
User rating	tweet	7.27	7.99	7.31	0.0220
Has mentions	tweet	0%	22.11%	2.03%	0.2017
Is a retweet	tweet	0.89%	24.72%	1.06%	0.2390
Has been retweeted	tweet	0%	17.36%	0.83%	0.2168
IMDb avg rating	IMDb	7.00	7.24	7.01%	0.0117
delta between user rating and IMDb avg rating	derived	0.27	0.74	0.29	0.0128
IMDb popularity (thousands of ratings)	IMDb	161.60	184.03	162.68	0.0121
Movie runtime (minutes)	Freebase	114.92	116.89	115.02	0.0067
Number of won awards	Freebase	2.28	2.74	2.30	0.0025
Estimated budged (millions of $)	Freebase	62.17	61.54	62.14	0.0019
Number of attended festivals	Freebase	1.54	2.02	1.57	0.0052
Number of days the movie is released	Freebase	2093.36	2129.34	2095.08	0.0095
Number of days the tweet is posted	Freebase	173.53	159.93	172.87	0.0056
Is adapted from book	Freebase	30.91%	29.21%	30.83%	0.0000
Is a series	Freebase	21.48%	21.23%	21.74%	0.0000
Is a sequel	Freebase	16.83%	14.19%	16.70%	0.0000
Is a prequel	Freebase	16.72%	16.11%	16.69%	0.0000
Is a standalone movie	Freebase	71.44%	72.61%	71.50%	0.0001

Table 5: Analysis of main nominal (a) and numeric/boolean (b) tweet properties with respect to the engagement of tweets. Last column details the information gain of attributes. Numeric attributes are reported as average value, boolean attributes as percentage of present data. The data source indicates whether the property has been obtained either from the original tweet, IMDb, Freebase, or derived from existing fields. Singular data and most promising attributes to identify engaging tweets are highlighted in bold. The listed values of tweet languages and of movie languages are restricted to languages having more than 500 tweets.

Algo	Accuracy		Confusion matrix			
	correct	incorrect	TP	TN	FP	FN
NB	72.28%	27.72%	4639	6438	1206	3042
BN	73.36%	26.64%	4798	6444	1200	2883
DT	70.28%	29.72%	3195	7575	69	4486

Table 6: 10-fold cross-validation performance of the binary predictors - Naïve Bayes (NB), Bayes Networks (BN), and Decision Trees (DT) - in terms of percentage of classification accuracy and confusion matrix, i.e., true positive (TP), true negative (TN), false positive (FP), and false negative (FN).

Decision Trees (DT). Classification performance computed using 10-fold cross-validation is reported in Table 6.

Among the others, it is worthy reporting an example of the decision tree learned by J48, which confirms some outcomes of the analysis performed in Section 5: a tweet is classified as engaging in one of the following cases: (i) if the tweet is a retweet, (ii) if the tweet has already been retweeted, or (iii) if the tweet has mentions. In all other cases the tweet is classified as non-engaging.

6.2 Pair learning

We experimented also with a *pair learning* approach [1], as described in the following. Again, the original ranking problem has been converted into a binary classification task. The ground idea consists of creating a new training dataset where each instance is composed by merging the attributes of a couple of tweets t_1 and t_2. From pairs of numeric attributes, we enrich the instance model by deriving additional attributes as the difference between the original numerical value (e.g., the difference between the user rating of t_1 and the user rating of t_2). Therefore, the instance has been classified as positive if the engagement of t_1 is greater than the engagement of t_2, otherwise it is classified as negative. Once we built the dataset with all possible pairs of tweets posted by users, we experimented with canonical binary classification techniques to learn to rank pairs of tweets, in particular we focused again on Naïve Bayes, Bayesian Networks, and Decision Trees.

Regardless the approach seemed promising in the theory, it did not significantly improve the performance of the previous one (see Section 6.1), with the disadvantage of a higher complexity.

6.3 Linear model

In previous sections we explored the corresponding binary problem studying the classification accuracy in terms of the confusion matrix. In this section we move forward and, on the basis of previous analyses, we experiment with linear models, evaluated using the nDCG metric defined by the challenge directly on the test set.

Let us assume that the engagement of a tweet can be directly computed based on the attributes of a tweet (e.g., rating value, rating count, etc.). Formally, let t be a tweet characterized by the attributes $a_0, a_1, \ldots, a_n$. We estimate the engagement for the tweet t by computing the weighted sum of the attributes a_i:

$$\text{eng}(t) = \sum_{i=1}^{n} w_i a_i \tag{1}$$

The objective consists of computing the weights w_i that maximize the nDCG score for the list consisting of the tweets ordered by the predicted engagement scores.

Due to the fact that the linear model requires numeric attribute values, we preprocessed the nominal attributes (e.g., `tweet language`). We split nominal attributes having m distinct values in m numeric attributes of those exactly one is set to 1, whereas all other attribute values are set to 0.

Due to the big number of attributes, we use a heuristic for computing the attribute weights w_i. Consequently, we faced the problem as an optimization task – i.e., we explored the space of weights w_i – starting from simple models composed by single attributes and incrementally adding more tweet properties to refine the predictor.

Firstly, we set all attributes weights to zero, with exception of one that is set to 1. The most interesting results show that a model uniquely based on the user rating value leads to a nDCG equals to 0.8131. All the others attributes – used singularly – leed to a lower accuracy (e.g., if we consider only the number of followers of a user we reach a nDCG performance equals to 0.7521 – only slightly, better than random).

We incrementally extend the linear model; at each step, we add to the previous model a new attribute and we optimize its weight with respect to the nDCG score. We start from a model M_1 composed only by the user rating value, and, for the sake of convenience, we fixed its weight w_1 to 1000. Then, we take into consideration, one by one, one among the remaining attributes, in order to construct a model M_2 composed by the rating count and such attribute. For each tested attribute, we tried several possible weights, spanning the space with a grid search (e.g., testing all weights w_2 from -100 to 100 with step 1). We discovered that the optimal attribute to add to M_1 was the number of user followers, that led to a nDCG equals to 0.8146. We proceed with this algorithm with the other attributes in order to incrementally improve the complexity of the model and its related prediction power. Table 7 reports the performance of the incremental models, with the attribute added with respect to previous model and its contribute. The best model reaches a nDCG equals to 0.8212.

#	Added attrib.	w_i	nDCG	Δ
1	User rating	1000	0.8131	0.8131
2	# User followers	10	0.8146	0.0150
3	# User favorites	1	0.8168	0.0022
4	# User friends	-3	0.8200	0.0032
5	Tweet language	n.a.	0.8212	0.0012

Table 7: Incremental linear model and related nDCG scores. The last column shows the increment in nDCG of a model with respect to the previous one. As for the tweet language attribute, some languages (e.g., Arabic) positively contribute to the score and the weight is not reported since it depends on the language.

7. PREDICTING THE ENGAGEMENT

In this section we present the final predictor we have used to face with the challenge task. In order to have a reference performance score, we first implemented a trivial classifier that estimates an engagement equals to zero (i.e., the value of engagement with the highest frequency) for any tweet.

The nDCG of such classifier results 0.750932. We expect any advanced predictor to score higher than this lower bound.

The predictor we have implemented leverages the multiple factors discovered during the analyses performed in Sections 5 and 6, mainly: the user rating value, the fact the the tweet content mentions other users, the fact the tweet is a retweet or whether it has been already retweeted.

It is worth noting that we do not need to estimate the exact engagement, but we can lean on an arbitrary **ranking score** to correctly sort tweets on the bases of their engagement. Among the several approaches we have experimented with, let us present three simple predictors - (i) the rating-based, (ii) the mention-based, and (iii) retweet-based predictor - followed by the final predictor that linearly merges the single models.

The rating-based predictor estimates the ranking score as the user rating score, reaching a nDCG equals to 0.817433.

The mention-based predictor estimates a binary ranking score, assigning 1 whether the tweet mentions other users, 0 otherwise. Its performance is equals to 0.75835, only slightly higher than the baseline predictor.

The retweet-based model takes into consideration whether a tweet is a retweet or it has already been retweeted. We assign a ranking score equals to 2 if the tweet is both a retweet and it has already been retweeted, a score equals to 0 if the tweet is neither a retweet nor it has already been retweet, 0 elsewhere, i.e.,

$$\mathrm{eng}(t) = \mathrm{has\_rts}(t) + \mathrm{is\_rt}(t) \tag{2}$$

where $\mathrm{has\_rts}(t)$ and $\mathrm{is\_rt}(t)$ are two boolean properties - equals to 1 when true, to 0 when false - representing whether the tweet has retweets and whether the tweet is a retweet, respectively. The model reaches a nDCG equals to 0.8009801. Worth noting that using only the property $\mathrm{has\_retweets}(t)$ would lead to a nDCG equals to 0.754922, while using only the property $\mathrm{is\_retweet}(t)$ would reach a performance, in terms of nDCG, equals to 0.795936.

Finally, we combined together the three approaches in a single linear model, where the ranking score is defined as:

$$\mathrm{eng}(t) = \mathrm{rating}(t) + \mathrm{mentioned}(t) + \mathrm{has\_rts}(t) + \mathrm{is\_rt}(t) \tag{3}$$

where $\mathrm{rating}(t)$ is the user rating in a 1-to-10 rating scale and $\mathrm{mentioned}(t)$ is a boolean property equals to 1 if the tweet has mentions, 0 otherwise. This model led to a nDCG equals to 0.835187.

8. CONCLUSION

The several experiments ran so far to define a good engagement predictor pointed out that the most relevant attributes for the challenge task are those related to the content of the tweet – i.e., rating posted by the user, presence of some user mentions - and to the retweet status – i.e., whether it is a retweet or it has already been retweeted. Although promising from the preliminary analyses, all additional attributes related to the user (e.g., his/her role/influence in the social network) and to the item (e.g., the movie genre) led only to minor (or no) improvements.

Acknowledgments

This research is supported by funding from the European Commission's 7th Framework Program under grant agreements no. 610594 (CrowdRec).

9. REFERENCES

[1] Z. Cao, T. Qin, T.-Y. Liu, M.-F. Tsai, and H. Li. Learning to rank: from pairwise approach to listwise approach. In *Proc. of the 24th Intl. conf. on Machine learning*, pages 129–136. ACM, 2007.

[2] M. Cha, H. Haddadi, F. Benevenuto, and K. P. Gummadi. Measuring user influence in twitter: The million follower fallacy. In *in ICWSM '10: Proc. of intl. AAAI Conf. on Weblogs and Social*, 2010.

[3] C. Finneran and P. Zhang. A person-artifact-task (pat) model of flow antecedents in computer-mediated environments. *International Journal of Human-Computer Studies*, 59:475–496, 2003.

[4] M. Forestier, A. Stavrianou, J. Velcin, and D. A. Zighed. Roles in social networks: Methodologies and research issues. *Web Intelligence and Agent Systems*, 10(1):117–133, 2012.

[5] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten. The weka data mining software: An update. *SIGKDD Explor. Newsl.*, 11(1):10–18, Nov. 2009.

[6] R. Kumar, J. Novak, and A. Tomkins. Structure and evolution of online social networks. In P. S. Yu, J. Han, and C. Faloutsos, editors, *Link Mining: Models, Algorithms, and Applications*, pages 337–357. Springer New York, 2010.

[7] K. Lee, J. Mahmud, J. Chen, M. Zhou, and J. Nichols. Who will retweet this?: Automatically identifying and engaging strangers on twitter to spread information. In *Proc. of the 19th Intl. Conf. on Intelligent User Interfaces*, IUI '14, pages 247–256, New York, NY, USA, 2014. ACM.

[8] S. A. Macskassy and M. Michelson. Why do people retweet? anti-homophily wins the day! In L. A. Adamic, R. A. Baeza-Yates, and S. Counts, editors, *ICWSM*. The AAAI Press, 2011.

[9] H. L. O'Brien and E. G. Toms. What is user engagement? a conceptual framework for defining user engagement with technology. *Journal of the American Society for Information Science and Technology*, 59(6):938–955, 2008.

[10] V. Singh, R. Jain, and M. Kankanhalli. Mechanism design for incentivizing social media contributions. In S. C. Hoi, J. Luo, S. Boll, D. Xu, R. Jin, and I. King, editors, *Social Media Modeling and Computing*, pages 121–143. Springer London, 2011.

[11] H. Valizadegan, R. Jin, R. Zhang, and J. Mao. Learning to rank by optimizing ndcg measure. In *Advances in neural information processing systems*, pages 1883–1891, 2009.

[12] S. Ye and S. Wu. Measuring message propagation and social influence on twitter.com. In L. Bolc, M. Makowski, and A. Wierzbicki, editors, *Social Informatics*, volume 6430 of *Lecture Notes in Computer Science*, pages 216–231. Springer Berlin Heidelberg, 2010.