Information Processing and Management xxx (2015) xxx–xxx

Contents lists available at ScienceDirect

Information Processing and Management

journal homepage: www.elsevier.com/locate/ipm

In the mood for sharing contents: Emotions, personality and interaction styles in the diffusion of news

Q1 Fabio Celli*, Arindam Ghosh, Firoj Alam, Giuseppe Riccardi

Department of Information Engineering and Computer Science, University of Trento, via Sommarive 9, 38123 Povo (TN), Italy

A R T I C L E I N F O

Article history:
Received 3 May 2014
Revised 7 July 2015
Accepted 17 August 2015
Available online xxx

Keywords:
Personality
Communication
Mood detection
Social media
Twitter
Contagion

A B S T R A C T

In this paper, we analyze the influence of Twitter users in sharing news articles that may affect the readers' mood. We collected data of more than 2000 Twitter users who shared news articles from *Corriere.it*, a daily newspaper that provides mood metadata annotated by readers on a voluntary basis. We automatically annotated personality types and communication styles of Twitter users and analyzed the correlations between personality, communication style, Twitter metadata (such as followig and folllowers) and the type of mood associated to the articles they shared. We also run a feature selection task, to find the best predictors of positive and negative mood sharing, and a classification task. We automatically predicted positive and negative mood sharers with 61.7% F1-measure.

1. Introduction and background

In online news and social media, people read and share links to news articles or other multimedia contents, that are related to their emotions, tastes and identity (Liu, 2007). The exposure to contents generated by others can give rise to different emotions like indignation, joy, anger or sadness (Cambria et al., 2012). Sometimes these contents may be shared or retweeted, indicating the users' will to participate in a diffuse conversation (Boyd et al., 2010) and share their emotions with others. Researchers (Bachrach et al., 2012; Kosinski et al., 2013) have discovered that such media consumption and sharing is affected by the personality type of the user. Different personality types are associated to different psychological dimensions (Golbeck et al., 2011b), such as linguistic functions, attentional focus, emotionality and social relationships.

In this paper, we address the question of how personality types and communication styles of Twitter users are related to the selection of contents they share in Twitter, affecting the diffusion of a positive or negative mood. We formalize this problem in 3 ways: as a correlation analysis, as a feature selection task and as a classification task. We aim at finding the relationships between personality, communicative style and mood sharing; the best predictors of mood and the performance in the classification of positive and negative mood sharers among Twitter users. We identify the data sources in Corriere[1], an Italian news platform that provides mood metadata annotated by the readers on a voluntary basis, and Twitter[2], that is widely used as an information diffusion platform. We annotate the data with personality and communication style labels, then we predict the average mood of the articles shared on Twitter by the users. The main contributions of this work to the research community are: (1) the

* Corresponding author. Tel.: +39 3381154491.
E-mail address: fabio.celli@unitn.it, fabio.celli@live.it (F. Celli).

[1] http://corriere.it
[2] http://twitter.com

http://dx.doi.org/10.1016/j.ipm.2015.08.002
0306-4573/© 2015 Published by Elsevier Ltd.

development of an aligned corpus of Tweets and news articles, automatically annotated with personality types, communication styles and gold standard mood labels; (2) the analysis of the influence of Twitter users' metadata, personality and communication style in the diffusion of mood; and (3) the prediction of mood of a news article from personal data.

The paper is structured as follows: in Section 2 we report some related works on information spread, mood, personality and emotions. Then we will describe the datasets and the annotations in Sections 3 and 4. In Section 5 we will report and discuss the results of the experiments, finally in Section 6 we will draw some conclusions.

2. Related work

It is well known that mood has an impact on social media and spreads through social networks. Bollen et al. (2011) predicted mood states (tension, depression, anger, vigor, fatigue, and confusion) from tweets and compared the results to a record of popular events gathered from media, finding a significant correlation between them. Other works focussed on information spread, virality and retweeting of messages. This kind of research reached contradictory conclusions: while some researchers concluded that the most important features to predict retweeting is the level of influence of the source of the tweet and the retweeter (Zaman et al., 2010), others discovered that message virality is connected to the content of the message being shared, rather than to the influencers who share it (Guerini et al., 2011; Suh et al., 2010).

Recent works that put together emotions and information spread, found that emotionally charged tweets tend to be retweeted more often and more quickly compared to neutral ones (Stieglitz & Dang-Xuan, 2013). Viral messages containing the six primary emotions (surprize, joy, sadness, anger, fear, and disgust) are very effective on recipients' emotional responses to viral marketing campaigns. However, emotional content can evoke different reactions based also on the gender of the audience. Dobele et al. (2007) discovered that male recipients were more likely to forward disgust-based and fear-based campaigns that their female counterparts. The effectiveness of mood as a feature has been proven for tasks like author profiling (Argamon et al., 2009) and cyberpedophilia (Bogdanova et al., 2014). Hill et al. provided formal evidence that positive and negative emotional states behave like infectious diseases spreading across social networks over long periods of time (Hill et al., 2010). As for the relationship between sentiment and personality, previous literature (Celli & Zaga, 2013) reports a little improvement in the classification of sentiment exploiting personality types.

Unlike previous works, this one does not make use of resources for sentiment analysis (Cambria et al., 2012), mood annotation (Staiano & Guerini, 2014), or mood assessment (Shahid et al., 2012). We exploit mood metadata annotated directly by news readers in *Corriere.it* on a voluntary basis, to analyze the role of the users in spreading moods in a social network like Twitter. In corriere there are 5 context-independent mood states: *amused, satisfied, disappointed, worried and indignated*. Each one of them can have a strength value between 0 and 100. To define personality types, we adopt the most popular personality model in psychology: the Big Five (Costa & McCrae, 2008), that defines 5 bipolar traits: *extroversion* (sociable *vs* shy); *emotional stability/neuroticism* (secure *vs* neurotic); *agreeableness* (friendly *vs* ugly); *conscientiousness* (organized *vs* careless) and *openness to experience* (insightful *vs* unimaginative). To define communication styles we adopt the classes provided by Analyzewords, a tool for tweet analysis based on Linguistic Inquiry and Word Count (LIWC) (Tausczik & Pennebaker, 2010). Analyzewords defines 11 communicative dimensions, namely: *upbeat* (positive words and large use of "we"), *worried* (use of anxious language and short questions), *angry* (large use of captions and hostile words), *depressed* (use of self-reference and negative words), *plugged-in* (use energy words and include many mentions in tweets), *personable* (use positive words and often refers to others), *distant* (use action words and do not refer to self much), *spacy* (use excited words and a lot of exclamation marks), *analytic* (use long words and complex conjunctions) *sensory* (use many feeling words and reference to self), *in the moment* (use mainly verbs at present and hashtags). In the next section we describe the collection and annotation of the dataset, in Section 4 we will evaluate the automatic annotation of personality.

3. Data collection and annotation

Twitter is a very popular micro-blogging web service that allows users to post short text messages, called "tweets", up to 140 characters. Common practices in Twitter are the "mentions", to converse with other users, "retweets" - to share information (Boyd et al., 2010), and "hashtags" - to aggregate messages by topic. In recent years a lot of works have focussed on data mining from Twitter. For example, for sentiment analysis from emoticons (Pak & Paroubek, 2010), irony detection (Reyes et al., 2013), ranking algorithm for extracting topic keyphrases from tweets (Zhao et al., 2011) and of course personality recognition (Celli & Rossi, 2012; Quercia et al., 2011; Golbeck et al., 2011a). Corriere is one of the most popular Italian daily newspapers, and the online platform is structured as a social network, according to the definition in Boyd and Ellison (2007). In particular, the website of corriere provides (1) a semi-public profile for each registered user, (2) articulates a list of users connected by a relationship of interest and (3) allows to view their list of connections to other registered users.

3.1. Dataset for the experiments

We sampled about 2500 users from Twitter who shared at least two articles from *corriere.it*. We limited the number of tweets sampled from the APIs to 3000 per user. We computed the ratio between the number of articles shared and the number of tweets posted, cutting the tail in the fourth quartile (tweet-shared articles ratio above 0.32), in order to remove the accounts of Corriere.it, journalists of Corriere and bots that retweet corriere articles. To compute average mood class, first we subtracted the

Please cite this article as: F. Celli et al., In the mood for sharing contents: Emotions, personality and interaction styles in the diffusion of news, Information Processing and Management (2015), http://dx.doi.org/10.1016/j.ipm.2015.08.002

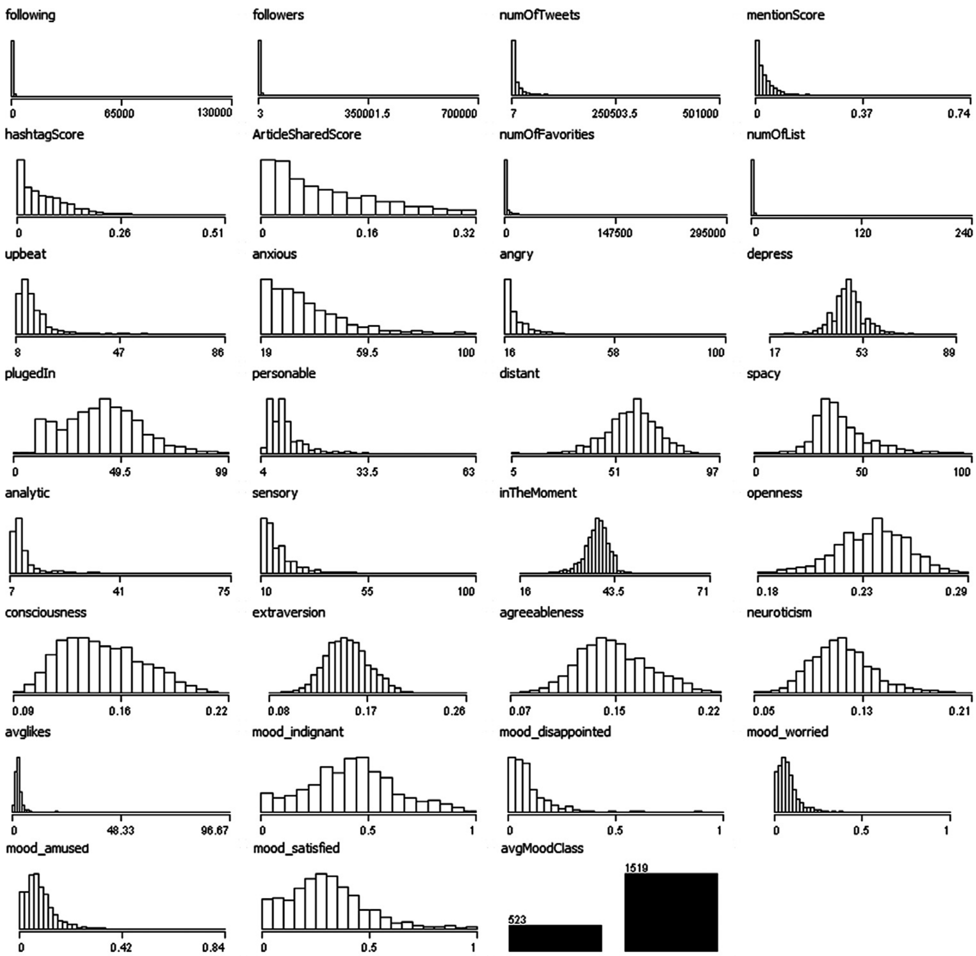

Fig. 1. Distribution of features in the dataset for experiments.

sum of "disappointed", "worried" and "indignated" scores from the sum of "amused" and "satisfied", obtaining a unique polarity score. Then we turned this polarity score into two classes: above and below zero, remiving 21 instances with score equal to 0. After this process we have 2042 unique users. A summary of the distribution of all features is reported in Fig. 1. Hashtag score, mention score and articles shared score are computed as the ratios of hashtags ($\frac{hashtags}{tweets}$), mentions ($\frac{all@-self@}{tweets}$) and Corriere articles ($\frac{articles}{tweets}$) over the number of Tweets sampled.

3.2. *Dataset for the evaluation of personality*

In order to evaluate the annotation of personality types, we recruited 210 Twitter users with an advertising campaign targeted at the followers of Corriere in Twitter, we assessed their personality types by means of the short BFI-10 personality test (Rammstedt & John, 2007) online[3]. In this way we obtained gold standard personality labels for training and evaluation. We used the short test (it takes less than 5 min to be completed) and we recruited only volunteers in order to have the full attention of the users (Buchanan et al., 2005). In the sample we have 118 males and 92 females aged between 14 and 65 years. A summary of the distribution of gold standard personality types is reported in Table 1. In the next section we will describe how we automatically annotated personality types and communication styles for the experiments and evaluated the automatic annotation of personality in the dataset for the evaluation.

4. Tools and evaluation

In order to perform the automatic annotation of personality types, we trained a supervized model on the gold standard labelled dataset we collected from Twitter. We split the data into training (180 Twitter users) and test set (30 users) using bag

[3] http://personality.altervista.org/personalitwit.php

Please cite this article as: F. Celli et al., In the mood for sharing contents: Emotions, personality and interaction styles in the diffusion of news, Information Processing and Management (2015), http://dx.doi.org/10.1016/j.ipm.2015.08.002

Table 1
Summary of gold standard personality types distribution.

Trait	Min	Mean	Max
Open.	−0.2	0.21	0.5
Cons.	−0.2	0.18	0.5
Extr.	−0.3	0.18	0.5
Agre.	−0.3	0.14	0.5
Neuro.	−0.3	0.12	0.5

Table 2
Results of personality score evaluation.

Class	RMSE	Baseline
Open.	0.18	0.19
Cons.	0.15	0.16
Extr.	0.17	0.22
Agre.	0.17	0.17
Neuro.	0.24	0.24
Avg.	**0.18**	**0.19**

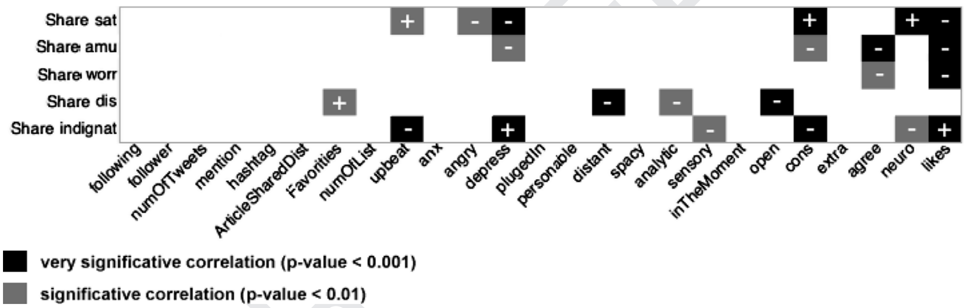

Fig. 2. Heatmap of the correlations between all the dimensions we retrieved (Twitter metadata, corriere metadata) and generated (personality types, communication styles).

of n-grams as features and Random Forest as learning algorithm. We obtained an average Root mean Squared Error of 0.18, as reported in detail in Table 2. This result is state-of-the-art, comparable to Golbeck et al. (2011a), who obtained an average Mean Absolute Error of 0.15.

We also labeled the dataset with communication styles, defined in Section 2, exploiting another tool freely available online[4] Analyzewords. This tool provides a representation of Tweets based on the psycholinguistic dimensions in LIWC (see Section 2), it does not require evaluation, as it is designed based on expert knowledge. In the next section we will report and discuss the results of the experiments.

5. Experiments and discussion

5.1. Correlation analysis

First of all we computed correlations between all the dimensions we retrieved, and we report the heatmap in Fig. 2. Many interesting relationships emerge from this experiment: first of all, the correlations between Twitter metadata and the action of sharing a specific mood are very few and weak. The only significative correlation is between the number of favorite Tweets and the tendency to share articles that arouse disappointment. An explanation of this may be that these users tend to read and collect news and tweets that attract their attention arousing disappointment.

Among communication styles, it is very interesting to note that the upbeat style is in a strong negative correlation to sharing articles that arouse indignation, and in a positive correlation with the action of sharing satisfaction. On the contrary, a depressed communicative style is strongly correlated to sharing indignation and negatively correlated to sharing satisfaction. Surprisingly, a distant communicative style is negatively correlated to sharing disappointing articles. We find the same negative correlation, although weaker, also for the users with an analytic communication style. Moreover, an angry communicative style is not correlated to sharing indignation, but it is just negatively correlated to sharing satisfaction.

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

Table 3
Results of feature selection.

InfoGain	Feature
0.0886	Avglikes
0.0706	NumOfTweets
0.0689	ArticleSharedScore
0.0681	Depress
0.0653	Consciousness
0.0604	Angry
0.0544	PlugedIn
0.0528	Upbeat
0.0499	NumOfFavorities
0.0477	HashtagScore

Table 4
Results of classification of positive and negative mood sharers in Twitter.

Class	P	R	F1
Baseline	0.5	0.5	0.5
Positive	0.608	0.663	0.634
Negative	0.629	0.572	0.599
Avg.	0.618	0.617	0.617

Among personality types, openness to experience is negatively correlated to sharing disappointment, just like the distant communicative style. An explanation for this, is that open-minded users like to understand things and do not like to share articles arousing disappointment. Conscientiousness is positively correlated to sharing satisfaction and negatively correlated to sharing indignation, and also negatively correlated to sharing amusement, although with less strength. A surprize is that also agreeableness is negatively correlated to sharing articles arousing amusement, but it is also negatively correlated to sharing articles that arouse worry or concern. Unsurprisingly, emotional stability/neuroticism is strongly correlated to sharing satisfaction and negatively correlated to sharing indignation. Surprisingly, extraversion is not correlated to any mood sharing action, although strongly correlated to an upbeat communication style.

Crucially, the number of likes on the articles is strongly correlated to articles that arouse indignation, while is negatively correlated to articles arousing worry, amusement and satisfaction. It is not easy to explain why the “like” action is strongly associated to a negative emotion. We suggest this may be connected to the fact that indignation is a social emotion (Miller, 2000) triggered by people’s tendency to view others’ behavior in relation to self-behavior. Under this perspective, the “like” action is an expression of support (Gerlitz & Helmond, 2011) to indignated people.

5.2. Predictors of positive and negative mood sharing

In the feature selection experiment we want to find the best predictors of the average mood shared on Twitter. We ran feature selection with information gain ranking as algorithm and 10-fold cross validation as the evaluation method. This algorithm evaluates the worth of the features by measuring the information gain of each attribute with respect to the class:

$$InfoGain(Class, Attribute) = H(Class) - H(Class|Attribute)$$

where H is the entropy. The results, reported in Table 3, show that the best features are the average article like score, which is not really surprising, because it depends directly from the article content. Crucially, the best communication style predictor is depression and the best personality predictor is conscientiousness, in line with the findings in previous work (Celli & Zaga, 2013).

5.3. Classification of positive and negative mood sharers

We performed a classification task to predict the average mood class and recognize automatically the positive and negative mood sharers on Twitter. As classification algorithm we used a Logistic Regression, with 66% training and 33% test split. We balanced the two classes with a weighting scheme, in order to preserve the number of instances, and used all the features. The results, reported in Table 4, show that it is possible to predict correctly about 60% of positive and negative mood sharers in Twitter using personality types and communication styles. In particular, positive mood sharers can be detected with more recall and negative mood sharers with more precision.

6. Conclusions and future work

In this paper we analyzed the role of personality and communication styles in the diffusion of news articles that may affect the readers’ mood. We explored the correlations between personality, communication style and Twtter metadata and we successfully