Inferring Perceived Demographics from User Emotional Tone and User-Environment Emotional Contrast

Svitlana Volkova
Johns Hopkins University
(now at Pacific Northwest National Laboratory)
Baltimore, MD, 21218, USA
svitlana@jhu.edu

Yoram Bachrach
Microsoft Research
Cambridge, UK CB1 2FB
yobach@microsoft.com

Abstract

We examine communications in a social network to study user emotional contrast – the propensity of users to express different emotions than those expressed by their neighbors. Our analysis is based on a large Twitter dataset, consisting of the tweets of 123,513 users from the USA and Canada. Focusing on Ekman's basic emotions, we analyze differences between the emotional tone expressed by these users and their neighbors of different types, and correlate these differences with perceived user demographics. We demonstrate that many perceived demographic traits correlate with the emotional contrast between users and their neighbors. Unlike other approaches on inferring user attributes that rely solely on user communications, we explore the network structure and show that it is possible to accurately predict a range of perceived demographic traits based solely on the emotions emanating from users and their neighbors.

1 Introduction

The explosion of social media services like Twitter, Google+ and Facebook have led to a growing application potential for personalization in human computer systems such as personalized intelligent user interfaces, recommendation systems, and targeted advertising. Researchers have started mining these massive volumes of personalized and diverse data produced in public social media with the goal of learning about their demographics (Burger et al., 2011; Zamal et al., 2012; Volkova et al., 2015) and personality (Golbeck et al., 2011; Kosinski et al., 2013),[1] language variation (Eisenstein et al., 2014; Kern et al., 2014; Bamman et al., 2014),[2] likes and interests (Bachrach et al., 2012; Lewenberg et al., 2015), emotions and opinions they express (Bollen et al., 2011b; Volkova and Bachrach, 2015), their well-being (Schwartz et al., 2013) and their interactions with online environment (Bachrach, 2015; Kalaitzis et al., 2016). The recent study has shown that the environment in a social network has a huge influence on user behavior and the tone of the messages users generate (Coviello et al., 2014; Ferrara and Yang, 2015a).

People vary in the ways they respond to the emotional tone of their environment in a social network. Some people tend to send out messages with a positive emotional tone, while others tend to express more negative emotions such as sadness or fear. Some of us are likely to share peer messages that are angry, whereas others filter out such messages. In this work we focus on the problem of predicting user *perceived demographics* by examining the emotions expressed by users and their immediate neighbors. We first define the user emotional tone, the environment emotional tone, and the user-environment emotional contrast.

Definition 1 Environment emotional tone *is the proportion of tweets with a specific emotion produced by the user's neighbors. For example, if the majority of tweets sent by the user's neighbors express joy, that user has a positive environment. In contrast, a user is in a negative environment if most of his or her neighbors express anger.*

Definition 2 User emotional tone *is the proportion of tweets with a specific emotion produced by a user. If a user mostly sends sad messages, he generates a sad emotional tone, while a user who mostly sends joyful messages has a joyful tone.*

[1] https://apps.facebook.com/snpredictionapp/

[2] http://demographicvis.uncc.edu/

Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, pages 1567–1578,
Berlin, Germany, August 7-12, 2016. ©2016 Association for Computational Linguistics

Definition 3 User-environment emotional contrast *is a degree to which user emotions differ from the emotions expressed by user neighbors. We say that users express more of an emotion when they express it more frequently than their neighbors, and say they express less of an emotion when they express it less frequently than their environment.*

There are two research questions we address in this work. First, we analyze how user demographic traits are predictive of the way they respond to the emotional tone of their environment in a social network. One hypothesis stipulates that the emotional response is a *universal* human trait, regardless of the specific demographic background (Wierzbicka, 1986; Cuddy et al., 2009). For example, men and women or young and old people should not be different in the way they respond to their emotional environment. An opposite hypothesis is a *demographic dependent emotional contrast hypothesis*, stipulating that user demographic background is predictive of the emotional contrast with the environment. For example, one might expect users with lower income to express negative emotion even when their environment expresses mostly positive emotions (high degree of emotional contrast), while users with higher income are more likely to express joy even if their environment expresses negative emotions (Kahneman and Deaton, 2010).

We provide an empirical analysis based on a large dataset sampled from a Twitter network, supporting the *demographic dependent emotional contrast hypothesis*. We show that users predicted to be younger, without kids and with lower income tend to express more sadness compared to their neighbors but older users, with kids and higher income express less; users satisfied with life express less anger whereas users dissatisfied with life express more anger compared to their neighbors; optimists express more joy compared to their environment whereas pessimists express less.

Furthermore, we investigate whether user demographic traits can be predicted from user emotions and user-environment emotional contrast. Earlier work on inferring user demographics has examined methods that use lexical features in social networks to predict demographic traits of the author (Burger et al., 2011; Van Durme, 2012; Conover et al., 2011; Bergsma et al., 2013; Bamman et al., 2014; Ruths et al., 2014; Sap et al., 2014). However, these are simply features of the text a user produces, and make limited use of the social embedding of the user in the network. Only limited amount of work briefly explored the network structure for user profiling (Pennacchiotti and Popescu, 2011a; Filippova, 2012; Zamal et al., 2012; Volkova et al., 2014; Culotta et al., 2015). In contrast, we investigate the predictive value of features that are completely dependent on the network: the emotional contrast between users and their neighbors. We also combine network (context) and text (content) features to further boost the performance of our models.

Our results show that the emotional contrast of users is very informative regarding their demographic traits. Even a very small set of features consisting of the emotional contrast between users and their environment for each of Ekman's six basic emotions and three sentiment types is sufficient to obtain high quality predictions for a range of user attributes.

Carrying out such an analysis requires using a large dataset consisting of many users annotated with a variety of properties, and a large pool of their communications annotated with emotions and sentiments. Creating such a large dataset with the ground truth annotations is extremely costly; user sensitive demographics e.g., income, age is not available for the majority of social media including Twitter. Therefore, we rely our analysis on a large Twitter dataset annotated with demographics and affects using predictive models that can accurately infer user attributes, emotions and sentiments as discussed in Section 3.

2 Data

User-Neighbor Dataset For the main analysis we collected a sample of $U = 10,741$ Twitter users and randomly sampled their neighbors $n \in N^{(u)}$ of different types including friends – u follows $n^{(u)}$, mentions – u mentions $n^{(u)}$ in his or her tweets *e.g., @modollar1*, and retweets – u retweets $n^{(u)}$ tweets *e.g., RT @GYPSY*. In total we sampled N= $141,034$ neighbors for U=$10,741$

Relation	$\subseteq U$	N_{uniq}	N_{all}	T_{total}
Retweet R	9,751	32,197	48,262	6,345,722
Mention M	9,251	37,199	41,456	7,634,961
Friend F	10,381	43,376	51,316	8,973,783
TOTAL	**10,741**	**112,772**	**141,034**	**24,919,528**

Table 1: Twitter ego-network sample stats: U=$123,513$ unique users with T=$24,919,528$ tweets, and E=$141,034$ edges that represent social relations between Twitter users.

users; on average 15 neighbors per user, 5 neighbors of each type with their 200 tweets; in total T=24,919,528 tweets as reported in Table 1. We also report the number of users with at least one neighbor of each type $\subseteq U$ and the number of unique neighbors N_{uniq}.[3]

Dataset Annotated with Demographics Unlike Facebook (Bachrach et al., 2012; Kosinski et al., 2013), Twitter profiles do not have personal information attached to the profile e.g., gender, age, education. Collecting self-reports (Burger et al., 2011; Zamal et al., 2012) brings data sampling biases which makes the models trained on self-reported data unusable for predictions of random Twitter users (Cohen and Ruths, 2013; Volkova et al., 2014). Asking social media users to fill personality questionnaires (Kosinski et al., 2013; Schwartz et al., 2013) is time consuming. An alternative way to collect attribute annotations is through crowdsourcing as has been effectively done recently (Flekova et al., 2015; Sloan et al., 2015; Preoiuc-Pietro et al., 2015).

Thus, to infer sociodemographic traits for a large set of random Twitter users in our dataset we relied on pre-trained models learned from 5,000 user profiles annotated via crowdsourcing[4] released by Volkova and Bachrach (2015). We annotated 125,513 user and neighbor profiles with eight sociodemographic traits. We only used a subset of sociodemographic traits from their original study to rely our analysis on models trained on annotations with high or moderate inter-annotator agreement. Additionally, we validated the models learned from the crowdsourced annotations on several public datasets labeled with gender as described in Section 2. Table 2 reports attribute class distributions and the number of profiles annotated.

Attribute	Class Distribution	Profiles
Age	$\leq$ 25 y.o. (65%), $>$ 25 y.o.	3,883
Children	No (84%), Yes	5,000
Education	High School (68%), Degree	4,998
Ethnicity	Caucasian (59%), Afr. Amer.	4,114
Gender	Female (58%), Male	4,998
Income	$\leq$ \$35K (66%), $>$ \$35K	4,999
Life Satisf.	Satisfied (78%), Dissatisfied	3,789
Optimism	Optimist (75%), Pessimist	3,562

Table 2: Annotation statistics of perceived user properties from Volkova and Bachrach (2015).

Validating Crowdsourced Annotations To validate the quality of perceived annotations we applied 4,998 user profiles to classify users from the existing datasets annotated with gender using approaches other than crowdsourcing. We ran experiments across three datasets (including perceived annotations): Burger et al.'s data (Burger et al., 2011) – 71,312 users, gender labels were obtained via URL following users' personal blogs; Zamal et al.'s data (Zamal et al., 2012) – 383 users, gender labels were collected via user names. Table 3 presents a cross-dataset comparison results.

We consistently used logistic regression with L2 regularization and relied on word ngram features similar to Volkova and Bachrach (2015). Accuracies on a diagonal are obtained using 10-fold cross-validation. These results show that textual classifiers trained on perceived annotations have a reasonable agreement with the alternative prediction approaches. This provides another indication that the quality of crowdsourced annotations, at least for gender, is acceptable. There are no publicly available datasets annotated with other attributes from Table 2, so we cannot provide a similar comparison for other traits.

Train\Test	Users	Burger	Zamal	Perceived
Burger	71,312	**0.71**	0.71	0.83
Zamal	383	0.47	**0.79**	0.53
Perceived	4,998	0.58	0.66	**0.84**

Table 3: Cross-dataset accuracy for gender prediction on Twitter.

Sentiment Dataset Our sentiment analysis dataset consists of seven publicly available Twitter sentiment datasets described in detail by Hassan Saif, Miriam Fernandez and Alani (2013). It includes $T_S^L = 19,555$ tweets total (35% positive, 30% negative and 35% neutral) from Stanford,[5] Sanders,[6] SemEval-2013,[7] JHU CLSP,[8] SentiStrength,[9] Obama-McCain Debate and Health Care.[10]

Emotion Dataset We collected our emotion dataset by bootstrapping noisy hashtag annota-

[3] Despite the fact that we randomly sample user neighbors, there still might be an overlap between user neighborhoods dictated by the Twitter network design. Users can be reweeted or mentioned if they are in the friend neighborhood $R \subset F, M \subset F$.

[4] Data collection and perceived attribute annotation details are discussed in (Volkova and Bachrach, 2015) and (Preoiuc-Pietro et al., 2015).

[5] http://help.sentiment140.com

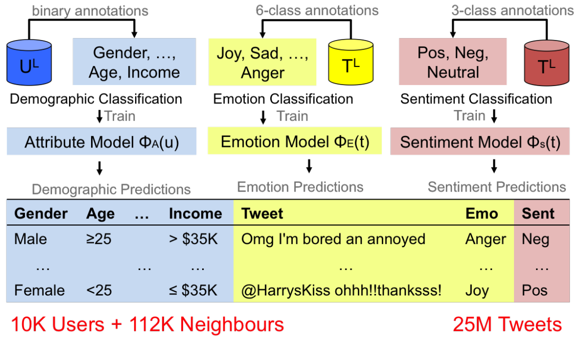
[6] http://www.sananalytics.com/lab/twitter-sentiment/

[7] http://www.cs.york.ac.uk/semeval-2013/task2/

[8] http://www.cs.jhu.edu/~svitlana/

[9] http://sentistrength.wlv.ac.uk/

[10] https://bitbucket.org/speriosu/updown/



Gender	Age	…	Income	Tweet	Emo	Sent
Male	≥25		> $35K	Omg I'm bored an annoyed	Anger	Neg
…		…	…	…	…	…
Female	<25		≤ $35K	@HarrysKiss ohhh!!thanksss!	Joy	Pos

Figure 1: Our approach for predicting user perceived sociodemographics and affects on Twitter.

tions for six basic emotions argued by Ekman[11] as have been successfully done before (De Choudhury et al., 2012; Mohammad and Kiritchenko, 2014). Despite the existing approaches do not disambiguate sarcastic hashtags e.g., *It's Monday #joy* vs. *It's Friday #joy*, they still demonstrate that a hashtag is a reasonable representation of real feelings (González-Ibáñez et al., 2011). Moreover, in this work we relied on emotion hashtag synonyms collected from WordNet-Affect (Valitutti, 2004), GoogleSyns and Roget's thesaurus to overweight the sarcasm factor. Overall, we collected $T_E^L = 52,925$ tweets annotated with anger (9.4%), joy (29.3%), fear (17.1%), sadness (7.9%), disgust (24.5%) and surprise (15.6%).

3 Methodology

Annotating User-Neighbor Data with Sociodemographics and Affects As shown in Figure 1, to perform our analysis we developed three machine learning components. The first component is a user-level demographic classifier $\Phi_A(u)$, which can examine a set of tweets produced by any Twitter user and output a set of predicted demographic traits for that user, including age, education etc. Each demographic classifier relies on features extracted from user content. The second and third components are tweet-level emotion and sentiment classifiers $\Phi_E(t)$ and $\Phi_S(t)$, which can examine any tweet to predict the emotion and sentiment expressed in the tweet.

For inferring user demographics, emotions and sentiments we trained log-linear models with L2 regularization using scikit-learn.[12] Our models rely on word ngram features extracted from user or neighbor tweets and affect-specific features described below.

Perceived Attribute Classification Quality In Section 2 we compared attribute prediction models trained on crowdsourced data vs. other datasets. We showed that models learned from perceived annotations yield higher or comparable performance using the same features and learning algorithms. Given Twitter data sharing restriction,[13] we could only make an indirect comparison with other existing approaches. We found that our models report higher accuracy compared to the existing approaches for gender: +0.12 (Rao et al., 2010), +0.04 (Zamal et al., 2012); and ethnicity: +0.08 (Bergsma et al., 2013), +0.15 (Pennacchiotti and Popescu, 2011b).[14] For previously unexplored attributes we present the ROC AUC numbers obtained using our log-linear models trained on lexical features estimated using 10-fold c.v. in Table 6.

Affect Classification Quality For emotion and opinion classification we trained tweet-level classifiers using lexical features extracted from tweets annotated with sentiments and six basic emotions. In addition to lexical features we extracted a set of *stylistic features including emoticons, elongated words, capitalization, repeated punctuation, number of hashtags and took into account the clause-level negation* (Pang et al., 2002). Unlike other approaches (Wang and Manning, 2012), we observed that adding other linguistic features e.g., higher order ngrams, part-of-speech tags or lexicons did not improve classification performance. We demonstrate our emotion model prediction quality using 10-fold c.v. on our hashtag emotion dataset and compare it to other existing datasets in Table 4. Our results significantly outperform the existing approaches and are comparable with the state-of-the-art system for Twitter sentiment classification (Mohammad et al., 2013; Zhu et al., 2014) (evaluated on the official SemEval-2013 test set our system yields F1 as high as 0.66).

Correlating User-Environment Emotional Contract and Demographics We performed

[11] We prefer Ekman's emotion classification over others e.g., Plutchik's because we would like to compare the performance of our predictive models to other systems.

[12] Scikit-learn toolkit: http://scikit-learn.org/stable/ Email svitlana.volkova@pnnl.gov to get access to pre-trained scikit-learn models and the data.

[13] Twitter policy restricts to sharing only tweet IDs or user IDs rather than complete tweets or user profiles. Thus, some profiles may become private or get deleted over time.

[14] Other existing work on inferring user attributes rely on classification with different categories or use regression e.g., age (Nguyen et al., 2011), income (Preoiuc-Pietro et al., 2015), and education (Li et al., 2014).

#Emotion	Wang (2012)		Roberts (2012)		Qadir (2013)		Mohammad (2014)		This work	
#anger	457,972	0.72	583	0.64	400	0.44	1,555	0.28	4,963	0.80
#disgust	–	–	922	0.67	–	–	761	0.19	12,948	0.92
#fear	11,156	0.44	222	0.74	592	0.54	2,816	0.51	9,097	0.77
#joy	567,487	0.72	716	0.68	1,005	0.59	8,240	0.62	15,559	0.79
#sadness	489,831	0.65	493	0.69	560	0.46	3,830	0.39	4,232	0.62
#surprise	1,991	0.14	324	0.61	–	–	3849	0.45	8,244	0.64
ALL:	**1,991,184**	–	**3,777**	**0.67**	**4,500**	**0.53**	**21,051**	**0.49**	**52,925**	**0.78**

Table 4: Emotion classification results (one vs. all for each emotion and 6 way for ALL) using our models compared to others.

our *user-environment emotional contrast analysis* on a set of users U and neighbors N, where $N^{(u)}$ are the neighbors of u. For each user we defined a set of incoming T^{in} and outgoing T^{out} tweets. We then classified T^{in} and T^{out} tweets containing a sentiment $s \in S$ or emotion $e \in E$, e.g. T_e^{in}, T_e^{out} and T_s^{in}, T_s^{out} where $E \rightarrow$\{*anger, joy, fear, surprise, disgust, sad*\} and $S \rightarrow$ \{*positive, negative, neutral*\}.

We measured the proportion of user's incoming and outgoing tweets containing a certain emotion or sentiment e.g., $p_{sad}^{in} = |T_{sad}^{in}|/|T^{in}|$. Then, for every user we estimated *user-environment emotional contrast* using the normalized difference between the incoming p_e^{in} and outgoing p_e^{out} emotion and sentiment proportions:

$$\Delta e = \frac{p_e^{out} - p_e^{in}}{p_e^{out} + p_e^{in}}, \forall e \in E. \quad (1)$$

We estimated *user environment emotional tone* and *user emotional tone* from the distributions over the incoming and outgoing affects e.g., $D_s^{in} = \{p_{pos}^{in}, \ldots, p_{neut}^{in}\}$ and $D_e^{in} = \{p_{joy}^{in}, \ldots, p_{fear}^{in}\}$. We evaluated *user environment emotional tone* – proportions of incoming emotions D_e^{in} and sentiments D_s^{in} on a combined set of friend, mentioned and retweeted users; and *user emotional tone* – proportions of outgoing emotions D_e^{out} and sentiment, D_s^{out} from user tweets. We measure similarity between *user emotional tone* and *environment emotional tone* via Jensen Shannon Divergence (JSD). It is a symmetric and finite KL divergence that measures the difference between two probability distributions.

$$\text{JSD}(D^{in}||D^{out}) = \frac{1}{2}I(D^{in}||D) + \frac{1}{2}I(D^{out}||D), \quad (2)$$

where $D = \frac{1}{2}I(D^{in}||D^{out})$, $I = \sum_e D^{in}\ln\frac{D^{in}}{D^{out}}$.

Next, we compared emotion and sentiment differences for the groups of users with different demographics $A = \{a_0; a_1\}$ e.g., a_0 = Male and a_1 = Female using a non-parametric Mann-Whitney U test. For example, we measured the means $\mu_{\Delta e=joy}^{Male}$ and $\mu_{\Delta e=joy}^{Female}$ within the group of users predicted to be Males or Females, and estimated whether these means are statistically significantly different. Finally, we used logistic regression to infer a variety of attributes for $U = 10,741$ users using different features below:

- outgoing emotional tone p_e^{out}, p_s^{out} – the overall emotional profile of a user (regardless the emotions projected in his environment);
- user-environment emotional contrast $\Delta e, \Delta s$ – show whether a certain emotion Δe or sentiment Δs is being expressed more or less by the user given the emotions he has been exposed to within his social environment;
- lexical features extracted from user content – represent the distribution of word unigrams over the vocabulary.

4 Experimental Results

For sake of brevity we will refer to a user *predicted* to be male as a male, and a tweet predicted to contain surprise as a simply containing surprise. Despite this needed shorthand it is important to recall that a major contribution of this work is that these results are based on *automatically predicted* properties, as compared to ground truth. We argue here that while such automatically predicted annotations may be less than perfect at the individual user or tweet level, they provide for meaningful analysis when done on the aggregate.

4.1 Similarity between User and Environment Emotional Tones

We report similarities between *user emotional tone* and *environment emotional tone* for different groups of Twitter users using Jensen Shannon Divergence defined in the Eq. 2. We present the mean JSD values estimated over users with two contrasting attributes e.g., predicted to be a_0=Male vs. a_1=Female in Table 5.

	Sentiment Similarities						Emotion Similarities					
Attribute $[a_0, a_1]$	**Retweet**		**Friend**		**All**		**Retweet**		**Friend**		**All**	
Income [$\geq$ \$35K, $<$ \$35K]	22.1	19.4	23.7	21.1	18.6	15.1	18.7	17.8	33.6	33.3	20.0	17.6
Age [$<$ 25 y.o, $\geq$ 25 y.o.]	19.0	22.7	20.2	25.3	14.3	19.7	17.2	19.9	32.8	34.7	17.0	21.1
Education [School, Degree]	19.4	22.1	21.1	23.8	15.2	18.5	18.0	18.1	33.9	32.1	18.1	18.9
Children [Yes, No]	24.2	19.9	28.4	21.4	23.2	15.6	20.9	17.8	35.6	33.2	22.6	18.0
Gender [Male, Female]	19.7	20.5	22.0	21.9	16.5	15.9	18.3	17.9	31.6	34.6	18.2	18.5
Ethnicity [Caucas., Afr. American]	20.5	19.4	21.7	22.5	15.8	16.9	17.2	19.8	32.5	35.2	17.5	20.1
Optimism [Pessimist, Optimist]	19.9	20.3	23.1	21.7	16.8	16.0	18.9	17.9	33.6	33.3	18.6	18.3
Life Satisfaction [Dissatis., Satisfied]	19.4	20.3	21.6	22.0	15.3	16.3	18.6	18.0	33.1	33.4	18.5	16.5

Table 5: Mean Jensen Shannon Divergences (displayed as percentages) between the incoming D^{in} and outgoing D^{out} affects for contrastive attribute values a_0 and a_1. MannWhitney test results for differences between a_0 and a_1 JSD values are shown in blue (p-value $\leq$ 0.01), green (p-value $\leq$ 0.05), and gray (p-value $\leq$ 0.1).

In Table 5 *user environment emotional tones* are estimated over different user-neighbor environments e.g., retweet, friend, and all neighborhoods including user mentions. We found that if *user environment emotional tones* are estimated from mentioned or retweeted neighbors the JSD values are lower compared to the friend neighbors. It means that users are more emotionally similar to the users they mention or retweet than to their friends (users they follow).

We show that *user incoming and outgoing sentiment tones* D_s^{in} and D_s^{out} estimated over all neighbors are significantly different for the majority of attributes except ethnicity. The divergences are consistently pronounced across all neighborhoods for income, age, education, optimism and children attributes (p-value $\leq$ 0.01). When the *incoming and outgoing emotional tones* D_e^{in} and D_e^{out} are estimated over all neighbors, they are significantly different for all attributes except education and life satisfaction.

4.2 User-Environment Affect Contrast

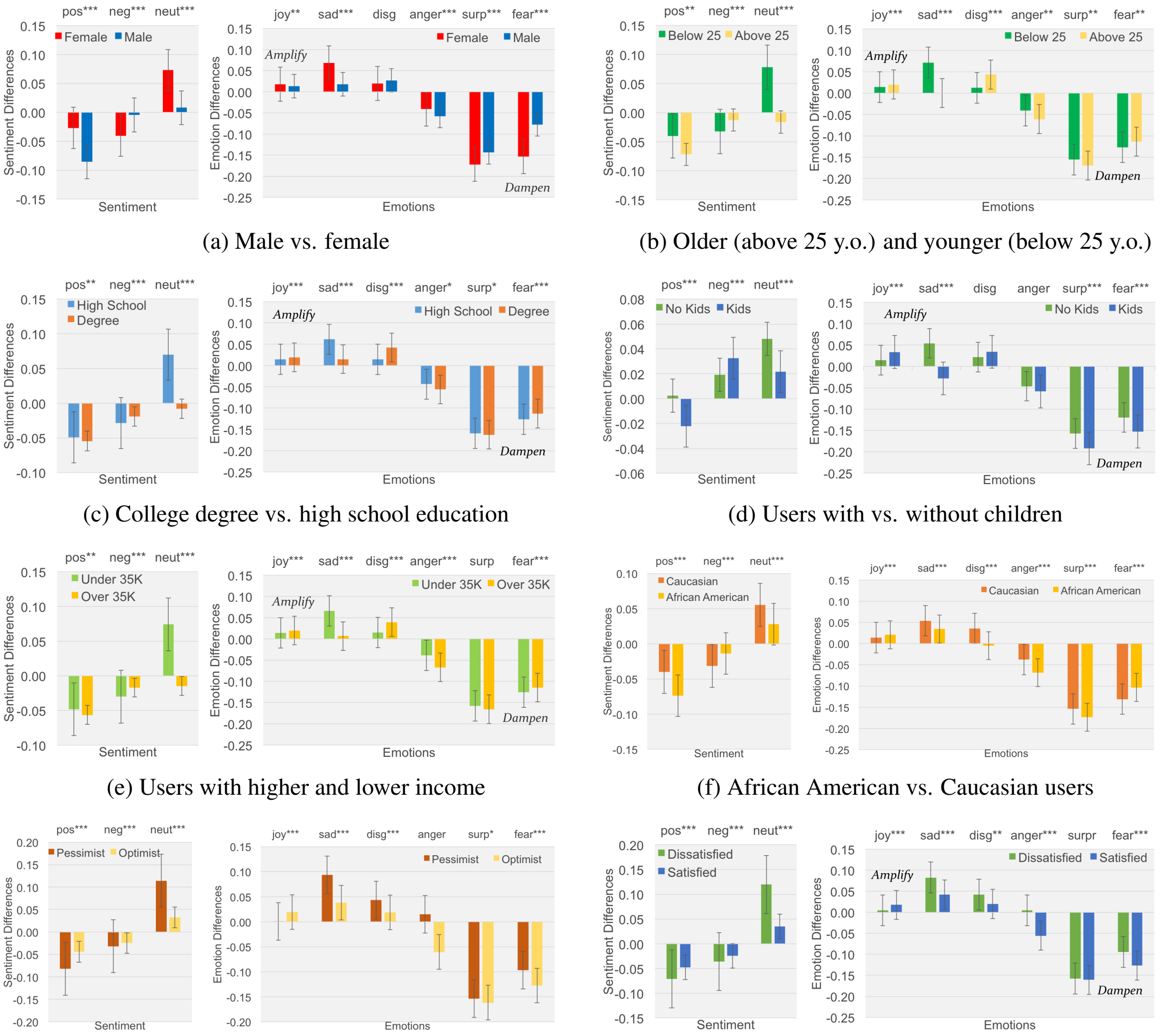
Our key findings discussed below confirm the *demographic dependent emotional contrast hypothesis*. We found that regardless demographics Twitter users tend to express more $(U > N)$ sadness↑, disgust↑, joy↑ and neutral↑ opinions and express less $(U < N)$ surprise↓, fear↓, anger↓, positive↓ and negative↓ opinions compared to their neighbors except some exclusions below.

Users predicted to be older and having kids express less sadness whereas younger users and user without kids express more. It is also known as the *aging positivity effect* recently picked up in social media (Kern et al., 2014). It states that older people are happier than younger people (Carstensen and Mikels, 2005). Users predicted to be pessimists express less joy compared to their neighbors whereas optimists express more. Users predicted to be dissatisfied with life express more anger compared to their environment whereas users predicted to be satisfied with life produce less. Users predicted to be older, with a degree and higher income express neutral opinions compared to their environment whereas users predicted to be younger, with lower income and high school education express more neutral opinions. Users predicted to be male and having kids express more positive opinions compared to their neighbors whereas female users and users without kids express less. We present more detailed analysis on user-environment emotional contrast for different attribute-affect combinations in Figure 2.

Gender Female users have a stronger tendency to express more surprise and fear compared to their environment. They express less sadness compared to male users, supporting the claim that female users are more emotionally driven than male users in social media (Volkova et al., 2013). Male users have a stronger tendency to express more anger compared to female users. Female users tend to express less negative opinions compared to their environment.

Age Younger users express more sadness but older users express similar level of sadness compared to their environment. It is also known as the *aging positivity effect* recently picked up in social media (Kern et al., 2014). It states that older people are happier than younger people (Carstensen and Mikels, 2005). They have a stronger tendency to express less anger but more disgust compared to younger users. Younger users have a stronger tendency to express less fear and negative sentiment compared to older users.

Education Users with a college degree have a weaker tendency to express less sadness but stronger tendency to express more disgust from



(a) Male vs. female

(b) Older (above 25 y.o.) and younger (below 25 y.o.)

(c) College degree vs. high school education

(d) Users with vs. without children

(e) Users with higher and lower income

(f) African American vs. Caucasian users

(g) Optimists vs. pessimists

(h) Satisfied vs. dissatisfied with life

Figure 2: Mean differences in affect proportions between users with contrasting demographics. Error bars show standard deviation for every e and s; p-values are shown as $\leq 0.01^{***}$, $\leq 0.05^{**}$ and $\leq 0.1^{*}$.

their environment compared to users with high school education. They have a stronger tendency to express less anger but weaker tendency to express less fear. Users with high school education are likely to express more neutral opinions whereas users with a college degree express less.

Children Users with children have a stronger tendency to express more joy, less surprise and fear from their environment compared to users without children. Users with children express less sadness and less positive opinions whereas users without children express more.

Income Users with higher annual income have a weaker tendency to express more sadness and have a stronger tendency to express more disgust, less anger and fear from their environment. They tend to express less neutral opinions whereas users with lower income express more.

Ethnicity Caucasian users have a stronger tendency to express more sadness and disgust from their environment whereas African American users have a stronger tendency to express more joy and less disgust. African American users have a stronger tendency to express less anger and surprise, but a weaker tendency to express less fear.

Optimism Optimists express more joy from their environment whereas pessimists do not. Instead, pessimists have a stronger tendency to express more sadness and disgust compared to optimists. Optimists tend to express less fear. Pessimists tend to express less positive but more neutral opinions.

Life Satisfaction User-environment emotional contrast for the life satisfaction attribute highly correlates with the optimism attribute. Users dissatisfied with life have a weaker tendency to express more joy but a stronger tendency to express more sadness and disgust. They express more anger whereas users satisfied with life express less anger. Users satisfied with life have a stronger tendency to express less fear but weaker tendency to express less positive and negative opinions.

In addition to our analysis on user-environment emotional contrast and demographics, we discovered which users are more "opinionated" relative to their environment on Twitter. In other words, users in which demographic group amplify less neutral but more subjective tweets e.g., positive, negative. As shown in Figure 2 male users are significantly more opinionated $\gg$ than female users, users with kids $>$ users without kids, users with a college degree $\gg$ users with high school education, older users $\gg$ younger users, users with higher income $\gg$ users with lower income, optimists $\gg$ pessimists, satisfied $\gg$ dissatisfied with life, and African American $>$ Caucasian users.

4.3 Inferring User Demographics From User-Environment Emotional Contrast

Our findings in previous sections indicate that predicted demographics correlate with the emotional contrast between users and their environment in social media. We now show that by using user emotional tone and user-environment emotional contrast we can quite accurately predict many demographic properties of the user.

Table 6 presents the quality of demographic predictions in terms of the area and the ROC curve based on different feature sets. These results indicate that most user traits can be quite accurately predicted using solely the emotional tone and emotional contrast features of the users. That is, given the emotions expressed by a user, and contrasting these with the emotions expressed by user environment, one can accurately infer many interesting properties of the user without using any additional information. We note that the emotional features have a strong influence on the prediction quality, resulting in significant absolute ROC AUC improvements over the lexical only feature set.

Furthermore, we analyze correlations between users' emotional-contrast features and their demographic traits. We found that differences between users and their environment in sadness, joy, anger and disgust could be used for predicting whether these users have children or not. Similarly, negative and neutral opinions, as opposed to joy, fear and surprise emotions can be predictive of users with higher education.

Attribute	Lexical	EmoSent	All	Δ
Age	0.63	0.74 (+0.11)	0.83	+0.20
Children	0.72	0.67 (–0.05)	0.80	+0.08
Education	0.77	0.78 (+0.01)	0.88	+0.11
Ethnicity	0.93	0.75 (–0.18)	0.97	+0.04
Gender	0.90	0.77 (–0.13)	0.95	+0.05
Income	0.73	0.77 (+0.04)	0.85	+0.12
Life Satisf.	0.72	0.77 (+0.05)	0.84	+0.12
Optimism	0.72	0.77 (+0.05)	0.83	+0.11

Table 6: Sociodemographic attribute prediction results in ROC AUC using Lexical, EmoSent (user emotional tone + user-environment emotional contrast), and All (EmoSent + Lexical) features extracted from user content.

5 Discussion

We examined the expression of emotions in social media, an issue that has also been the focus of recent work which analyzed emotion contagion using a controlled experiment on Facebook (Coviello et al., 2014). That study had important ethical implications, as it involved manipulating the emotional messages users viewed in a controlled way. It is not feasible for an arbitrary researcher to reproduce that experiment, as it was carried on the proprietary Facebook network. Further, the significant criticism of the ethical implications of the experimental design of that study (McNeal, 2014) indicates how problematic it is to carry out research on emotions in social networks using a controlled/interventional technique.

Our methodology for studying emotions in social media thus uses an *observational* method, focusing on Twitter. We collected subjective judgments on a range of previously unexplored user properties, and trained machine learning models to predict those properties for a large sample of Twitter users. We proposed a concrete quantitative definition of the emotional contrast between users and their network environment, based on the emotions emanating from the users versus their neighbors.

We showed that various demographic traits correlate with the emotional contrast between users and their environment, supporting the *demographic-dependent emotional contrast hypothesis*. We also demonstrated that it is possible to accurately predict many perceived demographic traits of Twitter users based solely on the emotional contrast between them and their neighbors. This suggests that the way in which the emotions we radiate differ from those expressed in our environment reveals a lot about our identity.

We note that our analysis and methodology have several limitations. First, we only study *correlations* between emotional contrast and demographics. As such we do not make any *causal* inference regarding these parameters. Second, our labels regarding demographic traits of Twitter users were the result of subjective reports obtained using human annotations – subjective impressions (Flekova et al., 2016) of people rather than the true traits. Finally, we crawled both user and neighbor tweets within a short time frame (less than a week) and made sure that user and neighbor tweets were produced at the same time. Despite these limitations, our results do indicate higher performance compared to earlier work. Due to the large size of our dataset, we believe our findings are correct.

6 Related Work

Personal Analytics in Social Media Earlier work on predicting latent user attributes based on Twitter data uses supervised models with lexical features for classifying four main attributes including gender (Rao et al., 2010; Burger et al., 2011; Zamal et al., 2012), age (Zamal et al., 2012; Kosinski et al., 2013; Nguyen et al., 2013), political preferences (Volkova and Van Durme, 2015) and ethnicity (Rao et al., 2010; Bergsma et al., 2013).

Similar work characterizes Twitter users by using network structure information (Conover et al., 2011; Zamal et al., 2012; Volkova et al., 2014; Li et al., 2015), user interests and likes (Kosinski et al., 2013; Volkova et al., 2016), profile pictures (Bachrach et al., 2012; Leqi et al., 2016).

Unlike the existing work, we not only focus on previously unexplored attributes e.g., having children, optimism and life satisfaction but also demonstrate that user attributes can be effectively predicted using emotion and sentiment features in addition to commonly used text features.

Emotion and Opinion Mining in Microblogs Emotion analysis[15] has been successfully applied to many kinds of informal and short texts including emails, blogs (Kosinski et al., 2013), and news headlines (Strapparava and Mihalcea, 2007), but *emotions* in social media, including Twitter and Facebook, have only been investigated recently. Researchers have used supervised learning models trained on lexical word ngram features, synsets, emoticons, topics, and lexicon frameworks to determine which emotions are expressed on Twitter (Wang et al., 2012; Roberts et al., 2012; Qadir and Riloff, 2013; Mohammad and Kiritchenko, 2014). In contrast, *sentiment* classification in social media has been extensively studied (Pang et al., 2002; Pang and Lee, 2008; Pak and Paroubek, 2010; Hassan Saif, Miriam Fernandez and Alani, 2013; Nakov et al., 2013; Zhu et al., 2014).

Emotion Contagion in Social Networks Emotional contagion theory states that emotions and sentiments of two messages posted by friends are more likely to be similar than those of two randomly selected messages (Hatfield and Cacioppo, 1994). There have been recent studies about emotion contagion in massively large social networks (Fan et al., 2013; Ferrara and Yang, 2015b; Bollen et al., 2011a; Ferrara and Yang, 2015a).

Unlike these papers, we do not aim to model the spread of emotions or opinions in a social network. Instead, given both homophilic and assortative properties of a Twitter social network, we study how emotions expressed by user neighbors correlate with user emotions, and whether these correlations depend on user demographic traits.

7 Summary

We examined a large-scale Twitter dataset to analyze the relation between perceived user demographics and the emotional contrast between users and their neighbors. Our results indicated that many sociodemographic traits correlate with user-environment emotional contrast. Further, we showed that one can accurately predict a wide range of perceived demographics of a user based solely on the emotions expressed by that user and user's social environment.

Our findings may advance the current understanding of social media population, their online behavior and well-being (Nguyen et al., 2015). Our observations can effectively improve personalized intelligent user interfaces in a way that reflects and adapts to user-specific characteristics and emotions. Moreover, our models for predicting user demographics can be effectively used for a variety of downstream NLP tasks e.g., text classification (Hovy, 2015), sentiment analysis (Volkova et al., 2013), paraphrasing (Preotiuc-Pietro et al., 2016), part-of-speech tagging (Hovy and Søgaard, 2015; Johannsen et al., 2015) and visual analytics (Dou et al., 2015).

[15] EmoTag: http://nil.fdi.ucm.es/index.php?q=node/186

References

Yoram Bachrach, Michal Kosinski, Thore Graepel, Pushmeet Kohli, and David Stillwell. 2012. Personality and patterns of Facebook usage. In *Proceedings of ACM WebSci*, pages 24–32.

Yoram Bachrach. 2015. Human judgments in hiring decisions based on online social network profiles. In *Data Science and Advanced Analytics (DSAA), 2015. 36678 2015. IEEE International Conference on*, pages 1–10. IEEE.

David Bamman, Jacob Eisenstein, and Tyler Schnoebelen. 2014. Gender identity and lexical variation in social media. *Journal of Sociolinguistics*, 18(2):135–160.

Shane Bergsma, Mark Dredze, Benjamin Van Durme, Theresa Wilson, and David Yarowsky. 2013. Broadly improving user classification via communication-based name and location clustering on Twitter. In *Proceedings of NAACL-HLT*, pages 1010–1019.

Johan Bollen, Bruno Gonçalves, Guangchen Ruan, and Huina Mao. 2011a. Happiness is assortative in online social networks. *Artificial life*, 17(3):237–251.

Johan Bollen, Huina Mao, and Xiaojun Zeng. 2011b. Twitter mood predicts the stock market. *Journal of Computational Science*, 2(1):1–8.

John D. Burger, John Henderson, George Kim, and Guido Zarrella. 2011. Discriminating gender on Twitter. In *Proceedings of EMNLP*, pages 1301–1309.

Laura L Carstensen and Joseph A Mikels. 2005. At the intersection of emotion and cognition aging and the positivity effect. *Current Directions in Psychological Science*, 14(3):117–121.

Raviv Cohen and Derek Ruths. 2013. Classifying political orientation on Twitter: It's not easy! In *Proceedings of ICWSM*.

Michael D. Conover, Bruno Gonçalves, Jacob Ratkiewicz, Alessandro Flammini, and Filippo Menczer. 2011. Predicting the political alignment of Twitter users. In *Proceedings of Social Computing*.

Lorenzo Coviello, Yunkyu Sohn, Adam DI Kramer, Cameron Marlow, Massimo Franceschetti, Nicholas A Christakis, and James H Fowler. 2014. Detecting emotional contagion in massive social networks. *PloS one*, 9(3):e90315.

Amy JC Cuddy, Susan T Fiske, Virginia SY Kwan, Peter Glick, Stéphanie Demoulin, Jacques-Philippe Leyens, Michael Harris Bond, Jean-Claude Croizet, Naomi Ellemers, Ed Sleebos, et al. 2009. Stereotype content model across cultures: Towards universal similarities and some differences. *British Journal of Social Psychology*, 48(1):1–33.

Aron Culotta, Nirmal Kumar Ravi, and Jennifer Cutler. 2015. Predicting the demographics of Twitter users from website traffic data. In *Proceedings of AAAI*.

Munmun De Choudhury, Michael Gamon, and Scott Counts. 2012. Happy, nervous or surprised? Classification of human affective states in social media. In *Proceedings of ICWSM*.

Wenwen Dou, Isaac Cho, Omar ElTayeby, Jaegul Choo, Xiaoyu Wang, and William Ribarsky. 2015. Demographicvis: Analyzing demographic information based on user generated content. In *Visual Analytics Science and Technology (VAST), 2015 IEEE Conference on*, pages 57–64. IEEE.

Jacob Eisenstein, Brendan O'Connor, Noah A Smith, and Eric P Xing. 2014. Diffusion of lexical change in social media. *PloS one*, 9(11):e113114.

Rui Fan, Jichang Zhao, Yan Chen, and Ke Xu. 2013. Anger is more influential than joy: sentiment correlation in Weibo. *arXiv preprint arXiv:1309.2402*.

Emilio Ferrara and Zeyao Yang. 2015a. Measuring emotional contagion in social media. *PloS one*, 10(11):e0142390.

Emilio Ferrara and Zeyao Yang. 2015b. Quantifying the effect of sentiment on information diffusion in social media. *PeerJ Computer Science*, 1:e26.

Katja Filippova. 2012. User demographics and language in an implicit social network. In *Proceedings of EMNLP-CoNLL*.

Lucie Flekova, Salvatore Giorgi, Jordan Carpenter, Lyle Ungar, and Daniel Preotiuc-Pietro. 2015. Analyzing crowdsourced assessment of user traits through Twitter posts. *Proceedings of the Third AAAI Conference on Human Computation and Crowdsourcing*.

Lucie Flekova, Jordan Carpenter, Salvatore Giorgi, Lyle Ungar, and Daniel Preotiuc-Pietro. 2016. Analyzing biases in human perception of user age and gender from text. In *Proceedings of the Association for Computational Linguistics*.

Jennifer Golbeck, Cristina Robles, Michon Edmondson, and Karen Turner. 2011. Predicting personality from Twitter. In *Proceedings of SocialCom/PASSAT*.

Roberto González-Ibáñez, Smaranda Muresan, and Nina Wacholder. 2011. Identifying sarcasm in Twitter: A closer look. In *Proceedings of ACL*, pages 581–586.

Yulan He Hassan Saif, Miriam Fernandez and Harith Alani. 2013. Evaluation datasets for Twitter sentiment analysis: A survey and a new dataset, the stsgold. *First ESSEM workshop*.

Elaine Hatfield and John T Cacioppo. 1994. *Emotional contagion*. Cambridge university press.

Dirk Hovy and Anders Søgaard. 2015. Tagging performance correlates with author age. In *Proceedings of the Association for Computational Linguistics (ACL)*, pages 483–488.

Dirk Hovy. 2015. Demographic factors improve classification performance. *Proceedings of ACL*.

Anders Johannsen, Dirk Hovy, and Anders Søgaard. 2015. Cross-lingual syntactic variation over age and gender. In *Proceedings of CoNLL*.

Daniel Kahneman and Angus Deaton. 2010. High income improves evaluation of life but not emotional well-being. *Proceedings of the National Academy of Sciences*, 107(38):16489–16493.

Alfredo Kalaitzis, Maria Ivanova Gorinova, Yoad Lewenberg, Yoram Bachrach, Michael Fagan, Dean Carignan, and Nitin Gautam. 2016. Predicting gaming related properties from twitter profiles. In *2016 IEEE Second International Conference on Big Data Computing Service and Applications (BigDataService)*, pages 28–35. IEEE.

Margaret L Kern, Johannes C Eichstaedt, H Andrew Schwartz, Gregory Park, Lyle H Ungar, David J Stillwell, Michal Kosinski, Lukasz Dziurzynski, and Martin EP Seligman. 2014. From sooo excited!!! to so proud: Using language to study development. *Developmental psychology*, 50(1):178.

Michal Kosinski, David Stillwell, and Thore Graepel. 2013. Private traits and attributes are predictable from digital records of human behavior. *National Academy of Sciences*.

Liu Leqi, Daniel Preoţiuc-Pietro, Zahra Riahi, Mohsen E. Moghaddam, and Lyle Ungar. 2016. Analyzing personality through social media profile picture choice. *ICWSM*.

Yoad Lewenberg, Yoram Bachrach, and Svitlana Volkova. 2015. Using emotions to predict user interest areas in online social networks. In *Data Science and Advanced Analytics (DSAA), 2015. 36678 2015. IEEE International Conference on*, pages 1–10. IEEE.

Jiwei Li, Alan Ritter, and Eduard Hovy. 2014. Weakly supervised user profile extraction from Twitter. *Proceedings of ACL*.

Jiwei Li, Alan Ritter, and Dan Jurafsky. 2015. Learning multi-faceted representations of individuals from heterogeneous evidence using neural networks. *arXiv preprint arXiv:1510.05198*.

Gregory McNeal. 2014. Facebook manipulated user news feeds to create emotional responses. *Forbes*.

Saif M. Mohammad and Svetlana Kiritchenko. 2014. Using hashtags to capture fine emotion categories from tweets. *Computational Intelligence*.

Saif M. Mohammad, Svetlana Kiritchenko, and Xiaodan Zhu. 2013. NRC-Canada: Building the state-of-the-art in sentiment analysis of tweets. In *Proceedings of SemEval*, June.

Preslav Nakov, Sara Rosenthal, Zornitsa Kozareva, Veselin Stoyanov, Alan Ritter, and Theresa Wilson. 2013. Semeval-2013 task 2: Sentiment analysis in Twitter. In *Proceedings of SemEval*, pages 312–320.

Dong Nguyen, Noah A. Smith, and Carolyn P. Rosé. 2011. Author age prediction from text using linear regression. In *Proceedings of LaTeCH*, pages 115–123.

Dong Nguyen, Rilana Gravel, Dolf Trieschnigg, and Theo Meder. 2013. ”How old do you think I am?” A study of language and age in Twitter. In *Proceedings of ICWSM*, pages 439–448.

Dong Nguyen, A Seza Doğruöz, Carolyn P Rosé, and Franciska de Jong. 2015. Computational sociolinguistics: A survey. *arXiv preprint arXiv:1508.07544*.

Alexander Pak and Patrick Paroubek. 2010. Twitter as a corpus for sentiment analysis and opinion mining. In *LREC*.

Bo Pang and Lillian Lee. 2008. Opinion mining and sentiment analysis. *Foundations of Trends in IR*, 2(1-2):1–135.

Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan. 2002. Thumbs up?: sentiment classification using machine learning techniques. In *Proceedings of EMNLP*, pages 79–86.

Marco Pennacchiotti and Ana-Maria Popescu. 2011a. Democrats, republicans and starbucks afficionados: user classification in Twitter. In *Proceedings of KDD*, pages 430–438.

Marco Pennacchiotti and Ana Maria Popescu. 2011b. A machine learning approach to Twitter user classification. In *Proceedings of ICWSM*, pages 281–288.

Daniel Preoiuc-Pietro, Svitlana Volkova, Vasileios Lampos, Yoram Bachrach, and Nikolaos Aletras. 2015. Studying user income through language, behaviour and affect in social media. *PLoS ONE*, 10(9):e0138717, 09.

Daniel Preotiuc-Pietro, Wei Xu, and Lyle Ungar. 2016. Discovering user attribute stylistic differences via paraphrasing.

Ashequl Qadir and Ellen Riloff. 2013. Bootstrapped learning of emotion hashtags #hashtags4you. *WASSA 2013*.

Delip Rao, David Yarowsky, Abhishek Shreevats, and Manaswi Gupta. 2010. Classifying latent user attributes in Twitter. In *Proceedings of SMUC*, pages 37–44.

Kirk Roberts, Michael A Roach, Joseph Johnson, Josh Guthrie, and Sanda M Harabagiu. 2012. Empatweet: Annotating and detecting emotions on Twitter. In *Proceedings of LREC*.

Derek Ruths, Jürgen Pfeffer, et al. 2014. Social media for large studies of behavior. *Science*, 346(6213):1063–1064.

Maarten Sap, Gregory Park, Johannes Eichstaedt, Margaret Kern, David Stillwell, Michal Kosinski, Lyle Ungar, and Hansen Andrew Schwartz. 2014. Developing age and gender predictive lexica over social media. In *Proceedings of EMNLP*.

Hansen Andrew Schwartz, Johannes C Eichstaedt, Margaret L Kern, Lukasz Dziurzynski, Richard E Lucas, Megha Agrawal, Gregory J Park, Shrinidhi K Lakshmikanth, Sneha Jha, Martin EP Seligman, et al. 2013. Characterizing geographic variation in well-being using tweets. In *ICWSM*.

Luke Sloan, Jeffrey Morgan, Pete Burnap, and Matthew Williams. 2015. Who tweets? deriving the demographic characteristics of age, occupation and social class from twitter user meta-data. *PloS one*, 10(3):e0115545.

Carlo Strapparava and Rada Mihalcea. 2007. Semeval-2007 task 14: Affective text. In *Proceedings of SemEval*, pages 70–74.

Ro Valitutti. 2004. Wordnet-affect: an affective extension of wordnet. In *Proceedings of LREC*, pages 1083–1086.

Benjamin Van Durme. 2012. Streaming analysis of discourse participants. In *Proceedings of EMNLP*, pages 48–58.

Svitlana Volkova and Yoram Bachrach. 2015. On predicting sociodemographic traits and emotions from communications in social networks and their implications to online self-disclosure. *Cyberpsychology, Behavior, and Social Networking*, 18(12):726–736.

Svitlana Volkova and Benjamin Van Durme. 2015. Online bayesian models for personal analytics in social media. In *Proceedings of AAAI*.

Svitlana Volkova, Theresa Wilson, and David Yarowsky. 2013. Exploring demographic language variations to improve multilingual sentiment analysis in social media. In *Proceedings of EMNLP*.

Svitlana Volkova, Glen Coppersmith, and Benjamin Van Durme. 2014. Inferring user political preferences from streaming communications. In *Proceedings of ACL*, pages 186–196.

Svitlana Volkova, Yoram Bachrach, Michael Armstrong, and Vijay Sharma. 2015. Inferring latent user properties from texts published in social media (demo). In *Proceedings of AAAI*.

Svitlana Volkova, Yoram Bachrach, and Benjamin Van Durme. 2016. Mining user interests to predict perceived psycho-demographic traits on Twitter.

Sida Wang and Christopher D Manning. 2012. Baselines and bigrams: Simple, good sentiment and topic classification. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Short Papers-Volume 2*, pages 90–94.

Wenbo Wang, Lu Chen, Krishnaprasad Thirunarayan, and Amit P Sheth. 2012. Harnessing Twitter "big data" for automatic emotion identification. In *Proceedings of SocialCom*, pages 587–592.

Anna Wierzbicka. 1986. Human emotions: universal or culture-specific? *American Anthropologist*, 88(3):584–594.

Faiyaz Al Zamal, Wendy Liu, and Derek Ruths. 2012. Homophily and latent attribute inference: Inferring latent attributes of Twitter users from neighbors. In *Proceedings of ICWSM*.

Xiaodan Zhu, Svetlana Kiritchenko, and Saif M Mohammad. 2014. NRC-Canada-2014: Recent improvements in the sentiment analysis of tweets. *SemEval*.