On Predicting Sociodemographic Traits and Emotions from Communications in Social Networks

Svitlana Volkova[1], **Yoram Bachrach**[2]
[1]Center for Language and Speech Processing, Johns Hopkins University, Baltimore MD 21218, USA
svitlana@jhu.edu
[2]Microsoft Research, Cambridge CB1 2FB, UK
yobach@microsoft.com

Abstract

We study[1] the relation between user demographic attributes, opinions and emotions on a large sample of Twitter users. Our methodology is based on building models for inferring coarse-grained emotions and attributes from user generated content. We examine various user attributes, including age, gender, ethnicity, income, political preference, intelligence and relationship status. We correlate these demographics with the emotional profile emanating from the users' tweets, as captured by Ekman's emotion classification.

We find that some users tend to express significantly more joy and significantly less sadness in their tweets, such as those in a relationship, with children, or with higher than average annual income or educational level. Female users tend to be more opinionated whereas male users tend to be more neutral. Finally, users who are younger, liberal and with a below average intelligence tend to project more negative opinions and emotions.

Introduction

Social media is rising in importance, taking a significant part of our everyday life, with services like Twitter and Facebook used regularly by more than 1/7[th] of the world's population. Researches have used the massive volumes of personalized and diverse data produced in such services to study various issues such as the language characterizing user attributes (Schwartz et al. 2013; Bamman, Eisenstein, and Schnoebelen 2014), romantic relationships in social networks[2] (Backstrom and Kleinberg 2014), and the relation between personality and the use of such tools (Golbeck et al. 2011; Bachrach et al. 2012; Kosinski, Stillwell, and Graepel 2013). However, less work has investigated the relation between user attributes and emotions or opinions they express in social media (Mohammad and Kiritchenko 2013; Volkova, Wilson, and Yarowsky 2013).

We propose an approach for correlating user attributes with the coarse-grained emotions and opinions they project in social media. Carrying out such an analysis requires using a large dataset consisting of many users, described by their demographic attributes, and a large pool of text generated by each such user, described by the emotion or sentiment expressed in the text. Generating such a large dataset is costly; it requires obtaining large sample of social network users, along with pieces of text they produce;[3] the users should then be tagged with their demographic attributes, which are not available or hidden due to privacy settings; finally, each piece of text should be examined to determine the emotion expressed in it.

We use Twitter as our target online social network. Our methodology is based on using crowdsourcing to get demographic labels for a medium size sample of $U^L = 5,000$ users, then training classifiers to predict these demographic traits from the textual content generated by these users. We then use the trained classifiers to get labels for a much larger sample of $U = 123,513$ users. We use a similar method for labeling the emotions expressed in user text, and train an emotion classifier on an initial sample of $T^L = 52,925$ tweets, then use the classifier to get emotion labels for a much larger sample of $T = 24,919,528$ tweets; however, rather than obtaining the tags for the initial sample through crowdsourcing, we use tweets annotated with emotional hashtags e.g., *#disgust, #anger* identifying a specific emotion. To perform a reliable analysis of the differences in the emotions expressed between users of different groups, we need our demographic and emotion predictions to be highly accurate. We show that our models outperform the existing state-of-the-art systems as described in the results section. Our high-level methodology is shown in Figure 1.

Related Work

Emotion detection in social media Emotion analysis[4] has been successfully applied to many kinds of informal and short texts including emails, blogs, chats and news headlines, but emotions in social media, including Twitter, were only recently investigated. Ekman (1992) proposed an emotion classification framework capturing 6 high-level emotions, and researchers have used supervised learning models to determine which emotions are expressed in various

[1]This is a pre-print. The full version can be found here: http://online.liebertpub.com/doi/abs/10.1089/cyber.2014.0609

[2]Predicting love and breakups with Facebook data – http://techcrunch.com/2014/02/14/facebook-love-data/

[3]Such text can be tweets on Twitter, or posts on Facebook.

[4]Emotions and sentiment are both affective states but they are not the same. Emotion is a state of consciousness in which various internal sensations are experienced. Sentiment is a thought, view, or attitude, especially one based on emotion (Desmet 2002).

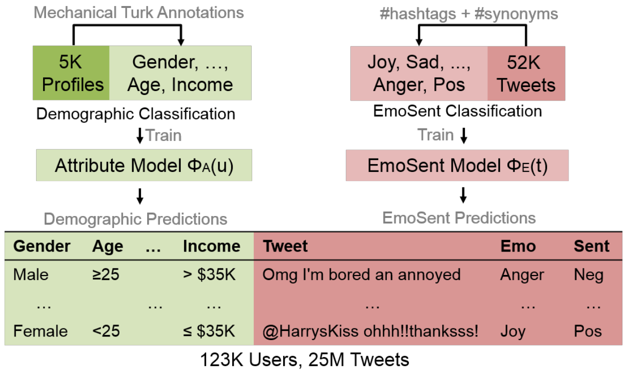


Figure 1: Our approach for predicting user demographics, emotions and opinions on Twitter.

texts (Wang et al. 2012; Roberts et al. 2012; Kim, Bak, and Oh 2012; Mohammad and Kiritchenko 2014). Due to the lack of social media data annotated with emotions and opinions, this line of work bootstraps noisy labeled data for sentiment (Pak and Paroubek 2010) and bases emotion prediction training on hashtags[5] e.g., *#happy*, *#sad* or emoticons. This bootstrapped data allows learning models using lexical or syntactic features. We use a similar idea, and build a hashtag dataset annotated with emotions, from which we build models for automatic emotion prediction, significantly outperforming previous models.

Personal analytics in social media Earlier work on predicting user attributes based on Twitter data uses supervised SVM models with lexical bag-of-word features for classifying four attributes including gender (Rao et al. 2010; Burger et al. 2011; Van Durme 2012; Ciot, Sonderegger, and Ruths 2013), age (Nguyen et al. 2013), political preferences (Cohen and Ruths 2013) and ethnicity (Pennacchiotti and Popescu 2011). Other methods characterize Twitter users by applying limited amounts of network structure information (Conover et al. 2011; Zamal, Liu, and Ruths 2012; Volkova, Coppersmith, and Van Durme 2014). We focus on a variety of previously unexplored personal attributes including relationship status, having children, religious beliefs, education level, intelligence, annual income, optimism and life satisfaction. We demonstrate that these attributes can be effectively predicted using linguistic features. We investigate the relation between emotions, opinions and these different demographic attributes. Some previous work has briefly studied such correlations e.g., the relation between gender and sentiment (Mohammad and Yang 2011; Volkova, Wilson, and Yarowsky 2013), and personality and emotions (Mohammad and Kiritchenko 2013) on Twitter.

Data

We now describe the various data sources we use to generate our large scale dataset, consisting of both user demographic traits and the emotional classification of their tweets.

[5]These approaches do not take into account sarcastic hashtags e.g. *It's Monday #joy*. But as has been shown, a hashtag is a reasonable representation of real feelings. Moreover, this work relies on emotion hashtag synonyms to outweight the sarcasm factor.

Demographic Dataset

The median number of tweets produced by a random Twitter user per day is small (Volkova, Coppersmith, and Van Durme 2014). Thus, in order to exclude users who tweet very often (such as celebrities, news, bots etc.) or tweet only occasionally, we first estimate the user tweeting frequency from the 1% Twitter feed. We then randomly sample users who tweet between 4 and 10 tweets per day. Using the Twitter API, we extracted data only for users who tweet (a) exclusively in English (verified via 'language' field), (b) from the US & Canadian time zones (verified via 'time zone' field), and (c) have at least 10 friends. In total we collected 10,741 user profiles and their 200 most recent tweets.[6]

To expand our network we crawl the collection of user IDs that the target user is following, otherwise known as their "friends", using the snowball sampling strategy (Goodman 1961). We randomly sample up to 10 friends per user and extract their profiles along with their 200 most recent tweets. Similarly, for all 10,741 users we extract a collection of user names the target user has mentioned *e.g., @HarrysKiss ohhhhhh!!thanksss!* or retweeted *e.g., RT @daledaleALE: I wish the answers were simple...* in their 200 tweets.

As a result, our random sample of a Twitter social network contains $U = 123,513$ users and $T = 24,919,528$ tweets. The scale of our network sample makes this study comparable to recent noteworthy large-scale studies (Coviello et al. 2014; Kosinski, Stillwell, and Graepel 2013).

Attribute	Attribute Values
Gender	Male: 2124, Female: 2874
Age	Below 25: 2511, Above 25: 1372
Political	Conservative: 595, Liberal: 1903
Ethnicity	Afr. Amer.: 1705, Caucasian: 2409
Religion	Christian: 3388, Unaffiliated: 1610
Education	High school: 3423, College degree: 1575
Relationship	Single: 3615, In a relationship 1383
Children	Yes: 797, No: 4203
Income	Under $35K: 3324, Over $35K: 1675
Intelligence	Aver. & Below: 4087, Above Aver.: 911
Optimism	Pessimist: 907, Optimist: 2655
Life Satisfaction	Dissatisfied: 840, Satisfied: 2949

Table 1: Personal attribute annotation distributions for 5,000 random Twitter users collected using crowdsourcing.

Demographic Annotations We take a disjoint set of 5,000 random Twitter users and ask people on Amazon Mechanical Turk to look over their profiles and tweets and judge their demographics. Crowdsourcing annotations is a common practice that has been successfully used for a variety of predictive analytics (Callison-Burch 2009). In Table 1 we present 12 demographic attributes with the corresponding values and the number of annotated profiles for each attribute value used to train the models. For the purposes of this study we binarized all attributes (original annotations are more fine-grained). We take two attribute values that can be easily disambiguated e.g., optimist vs. pessimist and are representative of the US Twitter population.

[6]Van Durme (2012) and Volkova, Coppersmith, and Van Durme (2014) have demonstrated that 200 tweets per user is sufficient to predict user gender and political preferences.

Emotion Dataset

Similar to other works that bootstrapped noisy data annotated with emotions we rely on the fact that people use the hashtag *#sadness* or emoticon :-(to indicate that they are sad. For that we collect tweets from the 1% Twitter feed over the last four years with hashtags that correspond to six emotions identified by Ekman (1992): *#joy, #anger, #fear, #sadness, #disgust* and *#surprise*. In addition, we compile a synonym list using WordNet-Affect, Google Synonyms and Roget's Thesaurus. In total we expand our initial emotion hashtag set to 360 emotion hashtags and collect more tweets that contain emotion synonym hashtags.

We exclude tweets with fewer than three words, filter out non-English tweets using Roget's Thesaurus, remove retweets and tweets where hashtags are not at the end of the tweet.[7] Finally, we get about 28,656 tweets collected using the original six emotion hashtags and 24,269 tweets collected using 360 emotion synonym hashtags – $T^L = 52,925$ tweets total annotated with emotions. We present the distribution of tweets for each emotion in Table 2. Our hashtag emotion dataset is three times larger than the recently released Hashtag Emotion Corpus (Mohammad and Kiritchenko 2014) but smaller than a prior bootstrapped corpus (Wang et al. 2012). However, we show that we significantly outperform all existing emotion prediction models.

External Sentiment Datasets

Sentiment analysis in social media has been studied a lot recently and became a well established task compared to the emotion prediction task described above. Thus, it is possible to take advantage of the existing publicly available resources for sentiment classification on Twitter (Hassan Saif, Miriam Fernandez and Alani 2013). For that we rely on several publicly available sentiment analysis datasets including Stanford,[8] Sanders,[9] SemEval-2013,[10] JHU,[11] SentiStrength,[12] Obama-McCain Debate[17] and Health Care Reform.[13] In total, we aggregate $T_S^L = 19,555$ tweets labeled with *positive (35%), negative (30%)* and *neutral (35%)* sentiment over 7 datasets.[14]

Methodology

In this section we describe how to build models for user-level (attributes) and tweet-level (emotion and sentiment) predictions. We then discuss how to measure emotion and opinion distributions for users. Finally, we show how to evaluate emotion and opinion differences between users of different demographic types.

[7] Researchers show that middle-of-tweet hashtags may not be good labels (González-Ibáñez, Muresan, and Wacholder 2011).

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

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

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

[11] http://www.cs.jhu.edu/ svitlana/

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

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

[14] Twitter data sharing policy does not allow sharing the actual tweets (only tweetIDs). However, some profiles became private or deleted over time and their tweets are not available.

Tweet-Level Predictions

We assume a set of independent tweets $T = \{t_i\}$. A tweet is labeled if we know the value of the emotion function $E(t)$: $T \rightarrow$ {joy, anger, disgust, fear, surprise, sadness} and sentiment function $S(t)$: $T \rightarrow$ {positive, negative, neutral}.

We define two tweet-based models $\Phi_E(t)$ and $\Phi_S(t)$ for emotion and sentiment classification learned from an independent set of labeled tweets described in the data section. These functions map each tweet to the most likely tweet-level attribute value assignments as shown for the emotion attribute below:

$$\Phi_E(t) = \text{argmax}_e \text{P}(E(t) = e \mid t) \tag{1}$$

User-Level Predictions

We assume a set of independent users U. A user $u \in U$ is labeled if we know the value of the attribute function $A(u) : U \rightarrow \{a_0, a_1\}$, for example the ethnicity attribute defined as: $A_{ethnicity}(u) = \{\text{African American; Caucasian}\}$, children attribute $A_{children}(u) = \{\text{Yes, No}\}$ etc.

We define a set of user-based models $\Phi_A(u)$ for classifying the 12 user attributes presented in Table 1. These models are learned from user self-authored content (200 tweets per user $T^{(u)}$) and return the most likely user-level attribute value assignment as shown below:

$$\Phi_A(u) = \text{argmax}_a \text{P}(A(u) = a \mid T^{(u)}) \tag{2}$$

Measuring User Emotions

Given a set of tweets $T^{(u)}$ with predicted emotions, we estimate the proportion or normalized frequency of each emotion e per user. The example emotion distribution for a random Twitter user is shown in Figure 4a.

Finally, given the normalized distribution of emotions for each user we estimate the user's *positive emotion score* $E^+(u)$. For that we subtract four negative emotions – anger, sadness, fear and disgust from one positive emotion – joy:

$$E^+(u) = e_{joy} - e_{anger} - e_{sad} - e_{disg} - e_{fear}. \tag{3}$$

We exclude the 'surprise' emotion from the $E^+(u)$ score because it can be both positive and negative and it is hard to evaluate out of context. However, the remaining 5 emotions can be easily disambiguated.

Measuring User Sentiment

We estimate the proportion or normalized frequency of sentiment expressed by each user. The example sentiment distribution for a random Twitter user is shown in Figure 4b. Given the proportion of sentiment per user, we estimate the user's *positive sentiment score* $S^+(u)$. For that we subtract the proportion of negative opinions from positive opinions as shown below:

$$S^+(u) = s_{pos} - s_{neg}. \tag{4}$$

We exclude neutral sentiment from $S^+(u)$ because we are only interested in measuring the difference between the most extreme sentiment polarities.

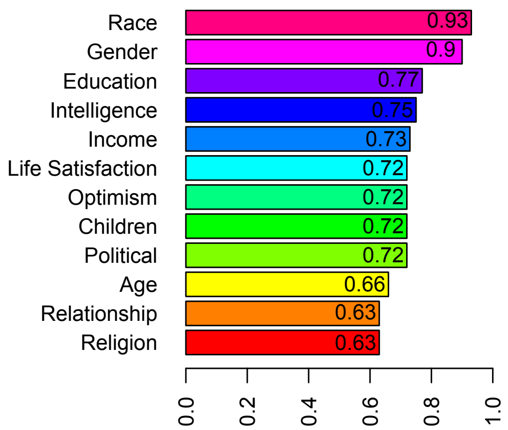
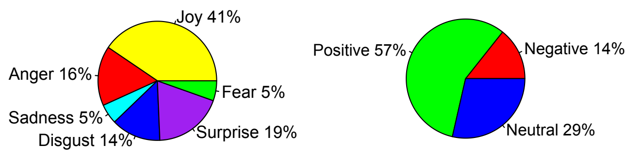


Figure 2: User demographic attribute classification performance (ROC AUC).

(a) Single (b) Above 35K

(c) In a relationship (d) Below 35K

Figure 3: Predictive lexical markers for income and relationship attributes.



(a) Emotions (b) Sentiment

Figure 4: Emotion and sentiment distributions for a user.

Estimating Emotion and Opinion Differences

We first evaluate whether Twitter users with different attribute values e.g., a_0 = Male vs. a_1 = Female express emotions and sentiment on the same or different levels. For that we group emotion $e \in E$ and opinion $s \in S$ distributions for a_0 and a_1 users as well as E^+ and sentiment S^+ positive scores calculated using the Eq. 3 and 4. We apply a *non-parametric Mann-Whitney U test*. Our null hypothesis H_0 is that a_0 emotion and sentiment proportions tend to be distributed similarly to a_1 values; H_a is that they tend to be distributed differently.

We then quantitatively measure emotion and opinion differences between a_0 and a_1 users. For that we estimate the averaged a_0 and a_1 distributions over emotions and opinions as $\mu_k^{(a_0)} = \frac{\sum_{a_0} |e_k|}{U}$ and $\mu_k^{(a_1)} = \frac{\sum_{a_1} |e_k|}{U}, e_k \in E$ (similarly $\mu_l^{(a_0)}$ and $\mu_l^{(a_1)}, s_l \in S$). We then take the difference between emotion means $\Delta\mu_k = \mu_k^{(a_0)} - \mu_k^{(a_1)}$ and opinion means $\Delta\mu_l = \mu_l^{(a_0)} - \mu_l^{(a_1)}$.

Experimental Setup and Results

Our experiment consists of three stages: (1) building emotion, sentiment and personal attribute classifiers to predict twelve attributes for 123K users and label their 25M tweets with emotions e.g., joy, sadness, anger, disgust, surprise or fear and sentiment e.g., positive, negative or neutral; (2) studying diversities between users of different demographic types and the emotions/sentiment they express; (3) measuring the strength of the relationships between demographics and emotions by fitting a regression model to infer demographic attributes from emotion and opinions.

We learn our classifiers using log-linear models $\Phi_A(u)$ defined in the Eq. 2 using the *scikit-learn toolkit*.[15] We prefer logistic regression over reasonable alternatives e.g., SVM or perceptron following previous work on predictive analytics and text classification in social media (Smith 2004).

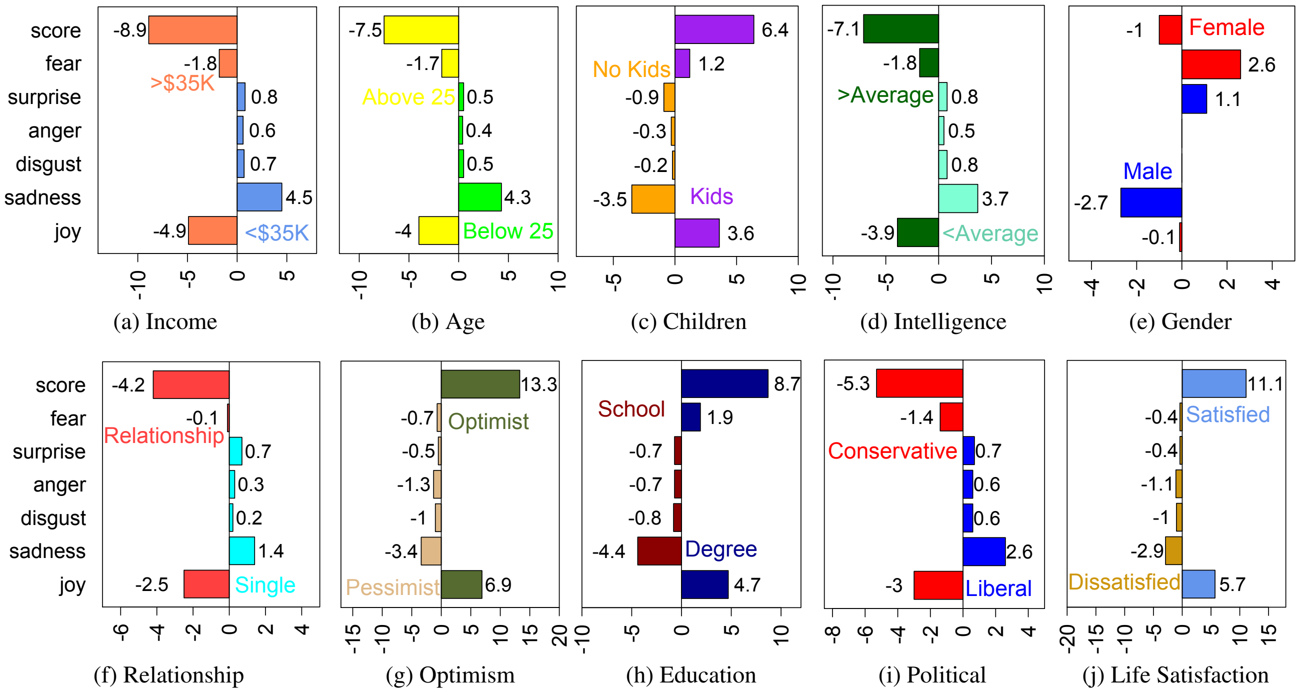
Predicting User Emotions, Sentiment and Demographics from Tweets

For opinion and emotion classification we train $\Phi_E(t)$ and $\Phi_S(t)$ models from the Eq. 1 using combined features as:

- LEXICAL: binary unigram bag-of-word features (using higher order ngrams or normalized frequency count-based features does not improve classification results).
- STYLISTIC: elongated words *e.g., Yaaay, noooo*; capitalization *e.g., COOL, MAD*; positive and negative emoticons;[16] punctuation *e.g., !!!!, ????*; number of hashtags.
- NEGATION: append a _NEG suffix to every word appearing between a negation and a clause-level punctuation mark (Pang, Lee, and Vaithyanathan 2002).[17]
- LEXICON: presence and score for unigram features from the Hashtag Emotion Lexicon (Mohammad and Kiritchenko 2014).
- POSTAGS: part-of-speech tags obtained using Twitter POS tagger.[18]

In Table 2 we present emotion classification results obtained using lexical, stylistic and negation features (using lexicon and part-of-speech tag features does not improve performance as also reported by Wang et al.). We compare our models to the results reported in (Mohammad and Kiritchenko 2014) and (Wang et al. 2012). Compared to Mohammad and Kiritchenko, our emotion predictors yield significantly higher results for all emotions when trained on

[15]http://scikit-learn.org/stable/
[16]http://sentiment.christopherpotts.net/tokenizing.html
[17]http://sentiment.christopherpotts.net/lingstruc.html
[18]http://www.ark.cs.cmu.edu/TweetNLP/



(a) Income (b) Age (c) Children (d) Intelligence (e) Gender

(f) Relationship (g) Optimism (h) Education (i) Political (j) Life Satisfaction

Figure 5: Emotion differences $\Delta\mu_k$ for Twitter users of different demographic types.

our and their combined data using 10-fold cross-validation ΔF_{M+T}: +0.23 F1 improvement in 6 way classification. Moreover, getting more data and adding stylistic features dramatically improves classification for individual emotions e.g., disgust +0.60 F1 and anger +0.43 F1. Furthermore, even though our models are learned from 40 times less data than (Wang et al. 2012) we significantly outperform their models – from +0.08 to +0.26 F1 gain.

We learn our sentiment classifier using the same features we used for our emotion classifier. We train our model on 19,555 tweets and test on 3,223 tweets from the official SemEval-2013 test set. Our sentiment classifier yields F1=0.66 (3 classes) which is comparable the top ranked system in the SemEval-2013 official ranking F1=0.69 (Nakov et al. 2013) and the current state-of-the-art with more advanced features F1=0.70 (Mohammad, Kiritchenko, and Zhu 2013).

For personal attribute prediction we learn models from user content e.g., tweets. For that we rely on *binary word unigram features*. We found that adding higher order ngrams, count-based features or part-of-speech tags does not improve performance. In Figure 2 we present classification results in terms of area under the ROC curve for the twelve attributes outlined in Table 1. Our results for gender and race (ethnicity) prediction demonstrate significantly higher performance compared to previous work (Rao et al. 2010; Burger et al. 2011; Pennacchiotti and Popescu 2011; Bergsma et al. 2013). For predicting political beliefs of 'average' Twitter users our results outperform (Cohen and Ruths 2013) models.

#Emotion	Mohammad		This work		ΔF_M	$\Delta F'_M$
#anger	1,555	0.28	4,963	0.80	+0.52	+0.43
#disgust	761	0.19	12,948	0.92	+0.73	+0.60
#fear	2,816	0.51	9,097	0.77	+0.26	+0.21
#joy	8,240	0.62	15,559	0.79	+0.17	+0.13
#sadness	3,830	0.39	4,232	0.62	+0.23	+0.10
#surprise	3849	0.45	8,244	0.64	+0.19	+0.15
ALL:	**21,051**	**0.49**	**52,925**	**0.78**	**+0.29**	**+0.23**

Table 2: Emotion detection results using our emotion classifier compared to ΔF_M (Mohammad and Kiritchenko 2014). $\Delta F'_M$ are the absolute improvements over their results when we combine both datasets to train a joint model.

Analyzing Emotion and Opinion Differences for Contrastive Demographics

Here we analyze emotional differences between users with contrasting a_0 and a_1 attributes and present the results in Figure 5. We present opinion differences in Table 3. Our key findings are discussed in detail below.

- **Gender** Female users generate more happy as well as sad tweets, while male users produce more surprise and fear tweets. Female users have higher positive emotion scores. In line with other work, our results further confirm that female users are more emotional online (Mohammad and Yang 2011; Volkova, Wilson, and Yarowsky 2013).
- **Age** Older (above 25 y.o.) users are 7.5% more positive, generate 4% more joy and 4% fewer sad tweets. Younger users produce more disgust, anger and surprise tweets. Our results are in line with the recently explored "aging positivity effect" in social media that states that older people are happier than younger people (Kern et al. 2014).
- **Relationship** Users in a relationship produce 4% more positive emotions, generate 2.5% more joy and 1.4% fewer sad tweets compared to single users. Single users produce more surprise, anger and disgust tweets.

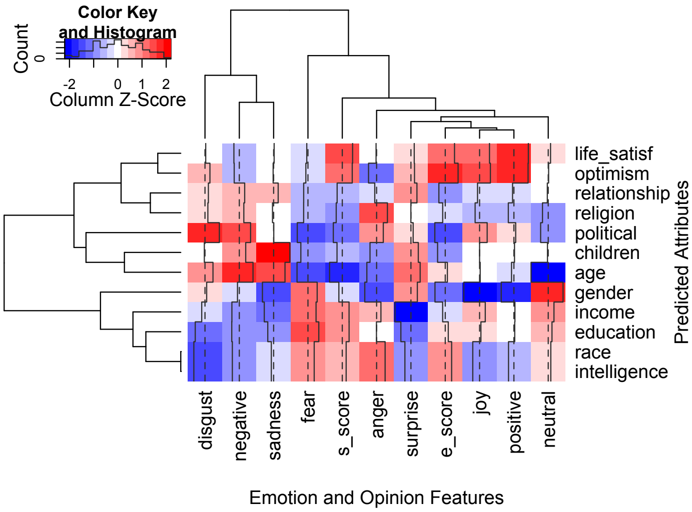


Figure 6: Predicting hidden attributes from user emotions and opinions. Colors represent regression coefficients for each feature e.g., red stands for male, satisfied, optimist, single, non religious, liberal, no kids, below 25 y.o., over \$35K, Caucasian, college degree, above average intelligence, and blue – opposite attributes values.

- **Children** Users without children produce 3.5% more sad and 3.6% fewer joy tweets. Users with children have 6.4% higher positive emotion scores. They produce fewer disgust, anger and surprise tweets but more fear tweets.
- **Education** Users with a college degree produce 4.7% more joy tweets and have 8.7% higher positive emotion score. In contrast, users with only high school education generate 4.4% more sad, disgust and anger tweets.
- **Political** Conservative users produce 3% more joy and 1.8% more fear tweets. Liberal users generate 2.6% more sad tweets and have a 7% lower positive emotion score.
- **Income** Users with higher income (above \$35K) tweet 4.9% more joy tweets and have an 8.9% higher positivity score. Users with lower income tweet 4.5% more sad tweets, almost 1% more anger and disgust tweets.
- **Intelligence** Users with above average intelligence create 3.9% more joy tweets and 3.7% fewer sad tweets. In contrast, users with average and below average intelligence have a 7.1% lower positive emotion score.
- **Ethnicity** Caucasian users produce 4.2% more joy tweets and have 7.5% higher positive emotion score. African American users generate 2% more sad, 1% more disgust and surprise and about 1% more fear and anger tweets.
- **Religion** Christian users produce more joy tweets and have higher positive emotion score compared to unaffiliated users, who produce more disgust and anger tweets.
- **Optimism** Optimists produce 7% more joy tweets and have 13% higher E^+ whereas pessimists generate 3.4% more sad, about 1% more anger and disgust tweets.
- **Life Satisfaction** Users satisfied with life produce 6% more joy and 3% fewer sad tweets, and 11% higher E^+.

Sentiment divergence results in Table 3 (similar to emotion differences) demonstrate that users who are female, above 25 years old, in a relationship, with kids, have a degree, get over \$35K a year, are conservative, with above average intelligence, Caucasian and Christian produce significantly more positive opinions.

Attribute	s_{pos}	s_{neut}	s_{neg}	S^+
Male+, Female–	–3.7	**+7.2**	–3.5	–0.3
Below 25+, Above 25–	–1.6	**+5.3**	**+6.9**	**–8.4**
Single+, Relationship–	+1.1	+0.7	+1.8	–3.0
Kids+, No kids–	+0.8	+3.3	–4.1	**+4.9**
Degree+, School–	+3.3	+4.1	**–7.5**	**+10.8**
$\leq$\$35K+, >\$35K–	–3.0	–4.5	**+7.5**	**–10.6**
Liberal+, Conservative–	–1.1	–3.4	+4.5	**–5.6**
$\leq$Average+, >Average–	–1.9	–4.5	**+6.4**	**–8.2**
Afr. Amer.+, Caucasian–	**–5.3**	+1.2	+4.1	**–9.4**
Christian+, Unaffiliated–	+1.2	–0.0	–1.2	+2.3
Optimist+, Pessimist–	**+5.3**	+2.3	**–7.6**	**+12.8**
Satisfied+, Dissatisfied–	**+5.2**	+2.6	**–7.8**	**+13.0**

Table 3: Sentiment differences $\Delta\mu_l$ (the most pronounced differences are highlighted in bold).

Inferring User Demographics from Emotions and Opinions

In Figure 6 we present demographic classification results using emotion and opinion distributions as features. We show that some emotions and opinions are predictive of one attribute value (red), some of an opposite value (blue). For instance, negative sentiment and sadness are predictive for users with no children and users below 25 years old; anger for non Christian users; surprise for single users; neutral sentiment vs. joy, positive sentiment and sadness for gender etc. In addition, we show a dendrogram for attributes (rows) and emotions (columns). It groups data based on row and column similarities using a hierarchical clustering algorithm. We observe that children and age attributes, life satisfaction and optimism are the most similar; negative sentiment and sadness, positive sentiment and joy are quite similar too.

Finally, we compare prediction performance using emotions and opinions vs. lexical features only. We observe that some attributes are more linguistically expressed, and therefore, are better predicted using lexical features e.g., gender (–0.16), race (–0.19), relationship (–0.05), children (–0.04), political beliefs (–0.05), religion (–0.05). However, some attributes are better predicted from emotions and opinions extracted from tweets e.g., age (+0.9), income (+0.05), education (+0.02), optimism (+0.06), and life satisfaction (+0.06).

Conclusions

We presented an approach to automatically infer demographic attributes for Twitter users and detect their opinions and emotions on a large scale. We demonstrated that users of different demographic types project different emotions. We found that rich, conservative, Caucasian, religious, educated, older users, with higher intelligence, kids and in a relationship are significantly more positive and produce more joyful tweets vs. users with opposite demographics, who are significantly more negative and generate more sad tweets.

We hope that this work will increase awareness of how public information on social networks can be used to learn many personal things about users. In addition, there are evident applications in online sales and targeted advertising.

References

Bachrach, Y.; Kosinski, M.; Graepel, T.; Kohli, P.; and Stillwell, D. 2012. Personality and patterns of Facebook usage. In *Proceedings of WebSci*, 24–32.

Backstrom, L., and Kleinberg, J. 2014. Romantic partnerships and the dispersion of social ties: A network analysis of relationship status on Facebook. In *Proceedings of CSCW*.

Bamman, D.; Eisenstein, J.; and Schnoebelen, T. 2014. Gender identity and lexical variation in social media. *Journal of Sociolinguistics* 18(2):135–160.

Bergsma, S.; Dredze, M.; Van Durme, B.; Wilson, T.; and Yarowsky, D. 2013. Broadly improving user classification via communication-based name and location clustering on Twitter. In *Proceedings of NAACL-HLT*, 1010–1019.

Burger, J. D.; Henderson, J.; Kim, G.; and Zarrella, G. 2011. Discriminating gender on Twitter. In *Proceedings of EMNLP*, 1301–1309.

Callison-Burch, C. 2009. Fast, cheap, and creative: Evaluating translation quality using Amazon's Mechanical Turk. In *Proceedings of EMNLP*, 286–295.

Ciot, M.; Sonderegger, M.; and Ruths, D. 2013. Gender inference of Twitter users in non-english contexts. In *Proceedings of EMNLP*, 1136–1145.

Cohen, R., and Ruths, D. 2013. Classifying Political Orientation on Twitter: It's Not Easy! In *Proceedings of ICWSM*.

Conover, M. D.; Ratkiewicz, J.; Francisco, M.; Gonc, B.; Flammini, A.; and Menczer, F. 2011. Political polarization on Twitter. In *Proceedings of ICWSM*, 89–96.

Coviello, L.; Sohn, Y.; Kramer, A. D.; Marlow, C.; Franceschetti, M.; Christakis, N. A.; and Fowler, J. H. 2014. Detecting emotional contagion in massive social networks. *PloS one* 9(3):e90315.

Desmet, P. M. 2002. Designing emotion.

Ekman, P. 1992. An argument for basic emotions. *Cognition & Emotion* 6(3-4):169–200.

Golbeck, J.; Robles, C.; Edmondson, M.; and Turner, K. 2011. Predicting personality from Twitter. In *Proceedings of SocialCom/PASSAT*.

González-Ibáñez, R.; Muresan, S.; and Wacholder, N. 2011. Identifying sarcasm in Twitter: A closer look. In *Proceedings of ACL HLT*, 581–586.

Goodman, L. A. 1961. Snowball sampling. *The annals of mathematical statistics* 148–170.

Hassan Saif, Miriam Fernandez, Y. H., and Alani, H. 2013. Evaluation datasets for Twitter sentiment analysis: A survey and a new dataset, the STS-Gold. *First ESSEM Workshop*.

Kern, M. L.; Eichstaedt, J. C.; Schwartz, H. A.; Park, G.; Ungar, L. H.; Stillwell, D. J.; Kosinski, M.; Dziurzynski, L.; and Seligman, M. E. 2014. From sooo excited!!! to so proud: Using language to study development. *Developmental psychology* 50(1):178.

Kim, S.; Bak, J.; and Oh, A. H. 2012. Do you feel what I feel? Social aspects of emotions in Twitter conversations. In *Proceedings of ICWSM*.

Kosinski, M.; Stillwell, D.; and Graepel, T. 2013. Private traits and attributes are predictable from digital records of human behavior. *National Academy of Sciences*.

Mohammad, S., and Kiritchenko, S. 2013. Using nuances of emotion to identify personality. In *Proceedings of ICWSM*.

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

Mohammad, S., and Yang, T. 2011. Tracking sentiment in mail: How genders differ on emotional axes. In *Proceedings of WASSA*, 70–79.

Mohammad, S. M.; Kiritchenko, S.; and Zhu, X. 2013. NRC-canada: Building the state-of-the-art in sentiment analysis of tweets. In *Proceedings of SemEval*.

Nakov, P.; Rosenthal, S.; Kozareva, Z.; Stoyanov, V.; Ritter, A.; and Wilson, T. 2013. Semeval-2013 task 2: Sentiment analysis in twitter. In *Proceedings of SemEval*, 312–320.

Nguyen, D.; Gravel, R.; Trieschnigg, D.; and Meder, T. 2013. "How old do you think I am?" A study of language and age in Twitter. In *Proceedings of ICWSM*, 439–448.

Pak, A., and Paroubek, P. 2010. Twitter as a corpus for sentiment analysis and opinion mining. In *LREC*.

Pang, B.; Lee, L.; and Vaithyanathan, S. 2002. Thumbs up?: sentiment classification using machine learning techniques. In *Proceedings of EMNLP*, 79–86.

Pennacchiotti, M., and Popescu, A. M. 2011. A machine learning approach to Twitter user classification. In *Proceedings of ICWSM*, 281–288.

Rao, D.; Yarowsky, D.; Shreevats, A.; and Gupta, M. 2010. Classifying latent user attributes in Twitter. In *Proceedings of SMUC*, 37–44.

Roberts, K.; Roach, M. A.; Johnson, J.; Guthrie, J.; and Harabagiu, S. M. 2012. Empatweet: Annotating and detecting emotions on Twitter. In *Proceedings of LREC*.

Schwartz, H. A.; Eichstaedt, J. C.; Kern, M. L.; Dziurzynski, L.; Ramones, S. M.; Agrawal, M.; Shah, A.; Kosinski, M.; Stillwell, D.; Seligman, M. E.; et al. 2013. Personality, gender, and age in the language of social media: The open-vocabulary approach. *PloS one* 8(9):e73791.

Smith, N. A. 2004. Log-linear models.

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

Volkova, S.; Coppersmith, G.; and Van Durme, B. 2014. Inferring user political preferences from streaming communications. In *Proceedings of ACL*, 186–196.

Volkova, S.; Wilson, T.; and Yarowsky, D. 2013. Exploring demographic language variations to improve sentiment analysis in social media. In *Proceedings of EMNLP*.

Wang, W.; Chen, L.; Thirunarayan, K.; and Sheth, A. P. 2012. Harnessing Twitter" big data" for automatic emotion identification. In *Proceedings of SocialCom*, 587–592.

Zamal, F. A.; Liu, W.; and Ruths, D. 2012. Homophily and latent attribute inference: Inferring latent attributes of Twitter users from neighbors. In *Proceedings of ICWSM*, 387–390.