

Inferring Latent User Properties from Texts Published in Social Media

Svitlana Volkova¹, Yoram Bachrach², Michael Armstrong² and Vijay Sharma²

¹Center for Language and Speech Processing, Johns Hopkins University, Baltimore MD 21218, USA
svitlana@jhu.edu

²Microsoft Research, Cambridge CB1 2FB, UK
{yobach, a-mica, a-vishar}@microsoft.com

Abstract

We demonstrate an approach to predict latent personal attributes including user demographics, online personality, emotions and sentiments from texts published on Twitter. We rely on machine learning and natural language processing techniques to learn models from user communications. We first examine individual tweets to detect emotions and opinions emanating from them, and then analyze all the tweets published by a user to infer latent traits of that individual. We consider various user properties including age, gender, income, education, relationship status, optimism and life satisfaction. We focus on Ekman's six emotions: anger, joy, surprise, fear, disgust and sadness. Our work can help social network users to understand how others may perceive them based on how they communicate in social media, in addition to its evident applications in online sales and marketing, targeted advertising, large scale polling and healthcare analytics.

Introduction

Online social networks like Twitter, Google+ and Facebook contain much information that can potentially reveal many traits, preferences and opinions of the profile owner. This resulted in research on personal analytics – automatically inferring such latent author attributes in social media. Recent work show that many traits of individuals can be accurately predicted from language on Twitter, including gender (Van Durme 2012), age (Zamal, Liu, and Raths 2012), and political views (Volkova, Coppersmith, and Van Durme 2014). Similar analysis has been applied to other information sources such as browsing history (Kosinski et al. 2012), likes or posts on Facebook¹ (Bachrach et al. 2012; Kosinski, Stillwell, and Graepel 2013). Moreover, previous work on social dynamics and big data analytics demonstrated the role of social media in understanding emotional and mood changes over time in large populations such as user emotional reactions to sad or happy events, user well-being² (Schwartz et al. 2013) and even depression and stress level (De Choudhury et al. 2013).

Copyright © 2015, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

¹Demographics and personality prediction from Facebook likes – <https://apps.facebook.com/snpredictionapp/>

²World Well Being Project – <http://wwbp.org/>

Our approach examines information on Twitter, determining the emotions expressed in tweets based on the words used in these tweets. It also infers various properties for a given user based on a set of aggregated tweets for that user. Predictions from our system have several applications. In *personalized online advertising*, the advertiser can personalize content based on user features, so as to match the emotional tone that the user expects. In *online marketing*, one could detect opinions and emotions users express in social media about products or services within targeted populations. In *large scale passive polling* or *real-time polling* one could predict political opinions and voting intentions for the users with certain demographics. Finally, this work could be used in *healthcare analytics*, for example in identifying depression or mental illness.

Models

Inferring Demographics and Online Personality Existing approaches for personal analytics typically rely on supervised models trained on a set of user profiles annotated with latent properties. For example, some demographic traits like gender, age and relationship status may be public in Facebook profiles or self-reported preferences may be expressed in the Twitter biography field or tweets e.g., *I am a Republican* or *It is my 25th birthday*. Such models can then be used to make new predictions for unseen users given features extracted from their profiles. However, these “explicit/self-reported” publicly available annotations are very sparse, so the existing datasets are very small including only few hundred users. Creating a larger dataset using surveys can be costly. Our methodology uses crowdsourcing to annotate user profiles.³ We asked workers on Amazon Mechanical Turk to glance through 5,000 Twitter profiles, all available metadata and tweets and make subjective judgments about a variety of their latent properties.

We trained *log-linear models* using *lexical features* extracted from 200 tweets per user profile annotated as de-

³We note that crowdsourcing subjective judgments of latent user properties is not a trivial task; some attributes are more difficult to recognize than others. But such crowdsourced annotations have been proven to be a reasonable approximation of the ground truth as shown for tasks at least as challenging, such as gender (Ciot, Sondererger, and Raths 2013) and depression prediction in social media (De Choudhury et al. 2013).

scribed above. Our models outperform the state-of-the-art approaches for inferring various traits, mostly due to better feature engineering and the larger training data size for commonly explored attributes.⁴

Our models output the probabilities (classifier confidence) for every user attribute being one or another class as shown in Figure 1 e.g., Male: 0.8 (Female: 0.2), Caucasian: 0.95 (African American: 0.05) etc. In total we classify 10 demographic attributes, 5 personality traits and 3 types of controlled impression behavior.

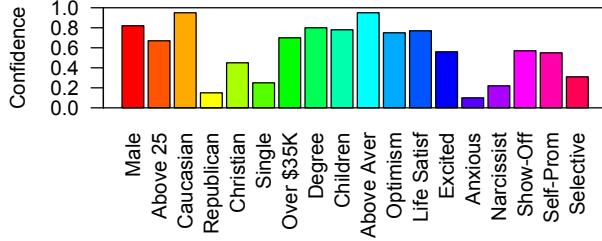


Figure 1: System output for automatically inferred latent demographics, personality and controlled impression behavior for a random user.

Predicting Emotions and Sentiments Training a tweet-level machine learning classifier for emotion detection required an initial tagged dataset. Obtaining a large pool of user tweets annotated with emotions and sentiments via crowdsourcing is extremely costly. Instead, we use distant supervision and bootstrap noisy hashtag annotations for the six basic emotions argued by Ekman.⁵ To get a training dataset, we find tweets which include an emotion hashtag (e.g. #joy to express happiness or #surprise to express surprise). We then treat the hashtag as the correct classification for the tweet, and these hashtagged tweets as an initial dataset to train a machine learning model for classifying emotions based on the words included in a text. For emotion and sentiment classification we combine various features including *lexical*, *stylistic* and take into account the *clause-level negation*. We report aggregated statistics about user emotions and sentiments as shown in Figure 2 along with their predicted demographics and personality aspects as shown in Figure 1.

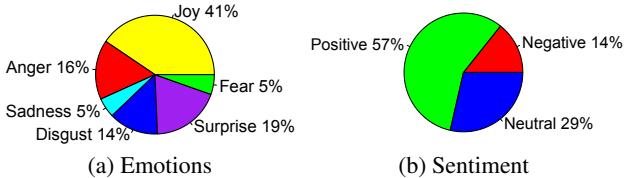


Figure 2: Emotion and sentiment distributions for a user.

⁴Gender: (Zamal, Liu, and Ruths 2012) +0.14, (Bergsma et al. 2013) +0.04; ethnicity: (Bergsma et al. 2013) +0.08, (Pennacchiotti and Popescu 2011) +0.15; political: (Volkova, Coppersmith, and Van Durme 2014) +0.05, (Cohen and Ruths 2013) +0.04.

⁵Similar methodology has been recently applied for affect analysis in Twitter (Mohammad and Kiritchenko 2014).

Conclusion

In this work we demonstrate how social media communications can be effectively used to analyze users' emotions and sentiments, and to infer latent online identities such as demographic attributes or personality traits. Our approach can be effectively used as a building block for studying emotion propagation in social networks, and to determine how user demographic traits affect their behavior in this media. We hope that this work would allow users to see how they may be perceived by their peers in social networks, so they can better understand what drives the image they project online.

References

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

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.

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*.

De Choudhury, M.; Gamon, M.; Counts, S.; and Horvitz, E. 2013. Predicting depression via social media. In *Proceedings of ICWSM*.

Kosinski, M.; Stillwell, D.; Kohli, P.; Bachrach, Y.; and Graepel, T. 2012. Personality and website choice. In *Proceedings of ACM Web Sciences*.

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

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

Schwartz, H. A.; Eichstaedt, J. C.; Kern, M. L.; Dziurzynski, L.; Lucas, R. E.; Agrawal, M.; Park, G. J.; Lakshminarayanan, S. K.; Jha, S.; Seligman, M. E.; et al. 2013. Characterizing geographic variation in well-being using tweets. In *Proceedings of ICWSM*.

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.

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.