Mining User Interests to Predict Perceived Psycho-Demographic Traits on Twitter

Svitlana Volkova
Johns Hopkins University
Baltimore, MD 21218
Email: svitlana@jhu.edu

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

Benjamin Van Durme
Johns Hopkins University
Baltimore, MD 21218
Email: vandurme@cs.jhu.edu

Abstract

We analyze the relation between user interests and their perceived psycho-demographic attributes using Twitter data, training models for predicting various personal traits of users. In contrast to existing work, which bases predictions on the textual tweets produced by users, we leverage the fact that users are embedded in the Twitter social network.

We examine the accounts that our users follow, and use them to determine the high-level interests of these users, then use these areas of interest as features for predicting perceived personal traits. We cover target attributes such as gender, age, educational background, political stand and personality.

We evaluate our technique on a dataset of over 4,000 Twitter user profiles. We use crowdsourcing to annotate these user profiles with perceptions regarding their personal traits, and correlate these with user interests, as captured by the accounts they follow and their classification in the Twitter "Who To Follow" hierarchy.

We compare the accuracy of our personal trait prediction methods with the state-of-the-art approaches that solely rely on user tweets, and discuss the correlations between perceived user demographics and interests.

1. Introduction

Social media is rising in importance, taking a significant part of our everyday life, with services such as Twitter, Google+ and Facebook are used regularly by over a billion users. People using these services generate massive quantities of data, consisting of the conversations they have online, the friendship structure in the network and the pages or objects their associate themselves with.

Researchers have used these massive volumes of diverse data to investigate many issues, such as the language characterizing different groups of users [1], [2], user relationships [3],[1], how such networks affect the workplace and hiring decisions [4], [5], [6], methods for predicting user properties, such as demographic traits [7], [8], [9], [10], [11], [12], [13], [14], [15] or personality [16], [17], [18],[2], income [19] and even user mood variation [20] and well-being [21].

Information about users gained from social network analysis can be useful in recommending items for a user, improving the performance of recommender systems. While some recommendations can be made solely based on a collaborative filtering approach [22], additional information about user properties, such as personality, may improve the quality of the recommendations given [23], [24]. Previous research has already considered approaches for designing a "fingerprint" of a user in collaborative filtering systems, based on the items they had already bought. Most such approaches use forms of matrix factorizations or dimensionality reduction [25], [26], [27], [28].[3]

The majority of the existing work on predicting personal traits of users from their online social network profiles relies on natural language processing and machine learning techniques applied to the text that these users generate, such as their wall-posts and conversations on Facebook or tweets on Twitter. Such techniques typically examine the textual content that has been generated by the user, then apply relatively simple processing converting these words to features. Such features can be simple counts of the various words in the vocabulary, or various measures such as TFIDF (Term Frequency/Inverse Document Frequency). Other such features include stemmed versions of the words appearing in the posts or various stylistic or textual markers.

Although such linguistic features can be powerful in predicting user traits, this kind of an analysis does not utilize the full richness of the data in online social networks. Further, many users do not produce a lot of content, making in difficult to infer their traits based solely on such linguistic features. For instance, many Twitter users are listeners rather than talkative users, and, thus, produce only a few tweets, or tweet very sporadically. Recent work shows that only 5% of Twitter users are "loudmouths", 50% of the users produce less than a single tweet per week on average, and 20% of the users have almost empty accounts [8]. Unfortunately, the accuracy of text based attribute prediction dramatically decreases when only a limited amount of textual posts is available as input [35].

1.1. Non-Textual Features and Social Embedding

Researchers have developed prediction methods that use information other than text contained in profiles. For instance, methods that examine "Likes" in Facebook [17], [18] achieve

1. Predicting love and breakups using Facebook data: http://techcrunch.com/2014/02/14/facebook-love-data.

2. Personality prediction from Facebook likes: http://brainblogger.com/2014/06/17/facebook-likes-and-twitter-followers-predict-personality-traits-and-more.

3. Such techniques can be costly in compute time and memory, so further techniques have been designed to improve the space and time complexity of these methods, at the cost of producing less accurate recommendations [29], [30], [31], [32], [33], [34].

excellent performance in predicting personal attributes, sometimes surpassing in accuracy even labels produced by human acquaintances who are close to the target user [36].[4]

However, such signals may not be present in all social networks. For example, it is difficult to map Facebook "Likes" directly to a similar feature for Twitter. A key common denominator to all online social networking services is the *social embedding*; all these services allow users to associate themselves with others, by friending them or following them. Thus, the social network can be represented as a graph, where users are nodes in the graph, and where edges indicate an association between two nodes as formally defined for Twitter by [35].

Social embeddings has allowed researchers to leverage the network structure to improve the quality of predictions when users produce limited quantities of text [10], [38], for instance by including the text produced by user neighbors in a social graph [11], [35] or repying on their following strategies [39].

In this work we also make use of the social embedding to predict user attributes, but focus on user interest areas and their predictive power. We examine popular accounts that users may follow, and which are suggested as "recommender accounts" pertaining to specific interest areas.[5] When a user *explicitly follows* some accounts that are recommended for people interested in a specific area (such as sports, technology or art), we may infer that they are interested in that area. Thus, by examining the Twitter accounts that a user follows and their interest-area identification in the recommender system, we can find out the interests of the target user. We then correlate these explicit interests with user perceived demographic traits, and train a machine learning classifier to predict user demographics given these interests.

1.2. Our Contribution

We propose an approach for predicting perceived user demographic traits given their interest areas. We train a classifier on a dataset consisting of Twitter accounts annotated with a variety of perceived demographics mapped to their interest areas.[6]

The demographic trait annotations for these profiles were obtained via crowdsourcing by asking workers on Mechanical Turk to examine these accounts and provide their opinion regarding the demographic traits of the account owners. The interest area tags for the accounts were obtained by examining highly popular accounts that these users follow and their classification in the "Who to Follow?" Twitter recommender system.

Our dataset contains 4,129 Twitter accounts annotated with 10 psycho-demographic attributes and mapped to the relative degree of interests these users exhibit in 26 Twitter interest categories shown in Figure 1.

We qualitatively and quantitatively evaluate the power of user interests in predicting various user properties, and compare our results with the state-of-the-art models that rely solely on language. Further, we investigate the interest areas that users of different demographic groups tend to have. For example, we examine gender and age differences in interest areas (as reflected by accounts followed in Twitter).

Our research has several important applications, including online advertising, personalized marketing, personalized recommendation systems and online search. Furthermore, knowing the demographics of the speaker (extra-linguistic or socio-linguistic features) can help a variety of downstream natural language processing tasks such as language generation [42], sentiment analysis and topic classification [15].

2. Perceived Psycho-Demographic Traits and Interests Dataset

We now describe how we created the dataset for correlating user interests and perceived psycho-demographic traits.

2.1. Collecting Perceived Demographics

We sampled 5,000 Twitter accounts of users residing in North America from the 1% Twitter firehouse along with their latest 200 tweets.

We requested Mechanical Turk workers to examine several of these Twitter accounts. Each worker had to report their opinion regarding properties of the owner of the Twitter account: gender, age, political stance, religion, educational background, relationship status, whether they have children, their income, as well as their degree of life satisfaction and whether they appear to be narcissistic. As these properties were not the ground-truth or self-reports of the profile owners, but rather the opinions or impressions regarding them formed by others, we refer to them as *perceived* demographic traits.

We also sourced redundant labels for some of the accounts, and estimated the inter-rater agreement on these properties, measured by Cohen's Kappa. [7]

2.1.1. Traits and Inter-Annotator Agreement. Table 1 shows the demographic class distributions for the attributes

4. Facebook "Likes" are not unique their ability to predict demographic traits. For example, recent work shows it is possible to predict a user's demographic traits from the video clips they watch on Youtube [37].

5. For a detailed explanation on recommended accounts to follow see http://www.twelveskip.com/tutorials/twitter/597/discover-people-to-follow-by-categories.

6. Our goal is not to build a recommender system for interest areas [40], [41]. Rather, we propose a way to infer user demographic traits given the areas they are interested in, and to find out what different demographic groups tend to be interested in.

7. We have only obtained redundant labels for few accounts to measure inter-rater agreement. While we could source redundant labels for all Twitter accounts, this would have very high costs. Further, once multiple labels are sourced for the same account, one must decide on how to aggregate these into a single label when the multiple annotators do not agree. A simple approach is using majority vote (which has been shown to significantly improve quality in some domains [43], [44]), but there are alternative more sophisticated methods [45], [46], [47], [48]. Rather than obtaining multiple annotators per account, we chose to spend our budget on annotating more Twitter accounts for our analysis.

in our dataset. The agreement level varied from perfect (gender Cohen's $\kappa = 0.81$), through fair (age, having children, education and income) to slight (the remaining attributes) as reported in the third column in Table 1.[8]

Attribute	Attribute Values	κ
Gender	Male: 1,797, Female: 2,329	0.81
Age	Below 25: 2,017, Above 25: 1,178	0.37
Education	School: 2,723, Degree: 1,403	0.26
Children	Yes: 657, No: 3,471	0.21
Income	$\leq$ \$35K: 1,480, $\geq$ \$35K: 2,648	0.21
Relationship	Single: 2,979, Relations: 1,147	0.13
Narcissist	Yes: 1,505, No: 1,773	0.11
Life Satisf	Dissatisfied: 648, Satisfied: 2,484	0.09
Religion	Christian: 2,822, Unaffiliated: 1,305	0.08
Political	Conservative: 514, Liberal: 1,577	0.06

TABLE 1: Attribute class distributions for 4,129 Twitter users for whom we were able to determine interests.

As described later in this section, our methodology only allows us to extract interest area labels for users who follow enough popular accounts (which are classified in the "Who to Follow" recommender). We thus had to eliminate some of the 5,000 users for which we could not determine interests from our dataset, leaving us with only 4,129 users.

2.2. Validating Crowdsourced Annotations

Many online services keep track of user demographic properties such as gender, age or eduction as fields in a structured profile information. For these services, one can simply extract these demographic features from the profile information, as done in work on Facebook [17], [18], Youtube [37], Flickr [50] or Google+ [51]. In contrast, Twitter does not keep such personal information as an integral part of the profile, so these traits of a user must be inferred through other means.

Various techniques, such as collecting self-reports [52], [53] or labeling users based on political candidates they follow [11], [54] brings data sampling biases, making the trained models achieve reduced performance when applied to users different from the population they were trained on [35], [54]. Other methods, such as asking users to fill questionnaires [1], [18] can be time consuming and costly. Our analysis is based on perceived traits: we ask annotators to provide their opinions regarding other users' demographic traits.

To validate the quality of perceived annotations obtained through crowdsourcing, one can use a dataset of user profiles tagged with perceived traits to train a machine learning classifier to predict personal traits from the language contained in the accounts, and then apply this classifier to a set of accounts for which demographic tags were obtained through a different methodology.

A disagreement between the classifiers does not necessarily mean that the perceived trait dataset has a low quality. It could be the case that the alternative method misclassified users, or that the machine learning classifier does not have a high enough accuracy (i.e. even if most labels in the perceived trait dataset are indeed correct, the classifier for mapping the text users generate to their traits may still have a significant error on another user population).[9]

To validate the quality of our perceived annotations we apply our data to classify users from the existing datasets annotated with gender and political preferences using approaches other than crowdsourcing. For gender, we run experiments across three datasets (including our data): Burger's data (71,312 users, gender labels were obtained via URL following) [8]; Zamal's data (383 users, gender labels were collected via user names) [11]. For political preference we used several alternative approaches: Volkova's Geo data (270 users, labels were obtained using self-reports) [35]; Zamal's data (371 users, political labels were collected by following political accounts) [11].

Tables 2 and 3 present a cross-dataset comparison results for gender and political preferences. We consistently used logistic regression models learned using binary word unigram features. Accuracies on a diagonal are obtained using 10-fold cross-validation.

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

TABLE 2: Cross-dataset classification accuracy for gender.

Train\Test	Users	Geo	Active	Our data
Geo	270	**0.66**	0.53	0.31
Active	371	0.60	**0.87**	0.76
Our data	2,498	0.56	0.61	**0.76**

TABLE 3: Cross-dataset classification accuracy for political preference attribute.

The results in Tables 2 and 3 show that textual classifiers trained on our data have a reasonable agreement with alternative prediction approaches for gender and political preferences. This provides another indication that the quality of our crowdsourced annotations, at least for these two traits, is acceptable. Unfortunately, no public datasets annotated with other attributes from Table 1 are available, so we cannot provide a similar comparison for other traits.[10]

2.3. Extracting Interest Areas

To determine user interests, we first examined each of 5,000 annotated Twitter profiles, and extracted a list of Twitter accounts that each of them follows. The "Who to Follow"

8. A commonly used scale for Cohen's Kappa values [49] characterized values < 0 as indicating no agreement and 0 to 0.20 as slight, 0.21 to 0.40 as fair, 0.41 to 0.60 as moderate, 0.61 to 0.80 as substantial, and 0.81 to 1 as almost perfect agreement.

9. Further, if *ground truth* labels were collected while constructing a dataset for an alternative attribute classification method, we can test our approach on this alternative dataset.

10. There are datasets with age annotations, but they use very different age bins (e.g. above or below 25 years old vs. 18-23, 23-25 y.o. etc.), so we could not perform the analysis for age.

recommender categorizes highly popular accounts into the 26 interest categories shown in Figure 1. For example, there are 48 Twitter accounts in the News category including the accounts CNN, MSNBC, FoxNews etc.[11]

We found that 871 of the 5,000 tagged Twitter accounts did not follow any of the popular accounts classified under the 26 Twitter interest categories. Thus, we excluded these users from our dataset.

We are interested in quantifying a user's degree of interest in some area, as reflected in the accounts that they follow. As we do not have interest labels for all followed accounts, but only for the popular ones, we restrict our attention to the interest-classified accounts that a user follows (i.e. the accounts that are part of the "Who to Follow" list).

Denote a user u's list of interest-tagged followed accounts as F_u, and denote by F_u^i the accounts that u follows and who are tagged with the interest i (i.e. the accounts in F_u that are classified in the interest area i in the "Who to Follow" list).

We quantify a user's degree of interest in an area as the proportion of followed accounts that deal with the interest i, given by:

$$\frac{|F_u^i|}{|F_u|}$$

Since $\sum_i \frac{|F_u^i|}{|F_u|} = 1$, we can talk of the proportional interest of a user in an interest area.

3. Methodology

We have a set of independent users $U = \{u\}$ for whom we want to predict a variety of perceived properties. We denote the function mapping a user to an attribute as $A(u)$, where the range of the function depends on the attribute (male/female for gender, yes/no for having children etc.)

3.1. Predicting User Properties

To predict user properties we train attribute classifiers $\Phi(u)$ using two feature types $f^{(u)}$: I. Context-based features learned from user interests $f_i^{(u)}$, II. Content-based features learned from user tweets $f_t^{(u)}$.

We define $\Phi(u)$ as a function mapping a user to the most likely attribute value assignment:

$$\Phi(u) = \text{argmax}_a \text{P}(\text{A}(u) = a | f^{(u)}). \quad (1)$$

To obtain interest features $f_i^{(u)}$ we get the distribution over the 26 interest categories for each user. We average these distributions in a specific user demographic sub-population (for example, the mean distribution for males, for those with perceived salary above the median etc.)

The example mean distributions over interests for male versus female users are shown in Figure 2. We then use statistical hypothesis testing to determine whether the differences in these distributions are statistically significant. For example, we can check whether male users are more interested in sports than female users. For such an analysis, we have used a non-parametric Mann Whitney U test to check for differences in interests between groups of users with contrasting attributes.

To get lexical features $f_t^{(u)}$, we aggregate 200 tweets per user and treat them as a single document (similarly to many earlier works on social network analytics). We remove hyperlinks (URLs), punctuation, user mentions and tokens that appear fewer than 5 times in the texts.

For each user we build a feature vector using binary word unigram features. We have also experimented with other features including POS tags, punctuation, LIWC lexicon, communication behavior, user metadata etc. However, we did not find any statistically significant improvements over the simple bag-of-words features.

We implement log-linear models $\Phi(u)$ using scikit-learn toolkit [57]. We preferred logistic regression over reasonable alternatives e.g., SVM or perceptron, following results in previous work on predictive analytics and text classification in social media [18], [37], [53], [58], [59], which indicate they have a good accuracy in such prediction tasks.

4. Results

We now present the results obtained using our prediction models when applied on our dataset.

4.1. Differences in Interests Across Demographic Trait Groups

We first measure the differences in interests between groups of users with contrastive demographics. Our results are given in Table 4, showing mean values for every interest category. As discussed earlier we determine whether two interest level means for users with contrastive attributes a_0 and a_1 (e.g. males versus females) are statistically significantly different using a Mann-Whitney U test. Our analysis shows that all differences reported in Table 4 are statistically significant with p-value $\leq$ 0.001 (or p-value $\leq$ 0.05 for some highlighted interest-attribute combinations).

The results in Table 4 show that most demographic sub-populations differ in their degree of interest in many areas. This indicates that it should be possible to make predictions regarding demographic traits given information about the areas a user is interested in.

4.2. User Attribute Prediction Model Performance

We now discuss the attribute classification task. Table 5 shows the performance of our classifiers for each attribute, measured in in terms of the area under the ROC curve (AuC). We use AuC rather than prediction accuracy as the accuracy may be artificially inflated by the fact that the dataset is

11. Our approach is based on the "Who to Follow" structure. An alternative method for determining user interests is through crowdsourced annotation [55], which is more costly.

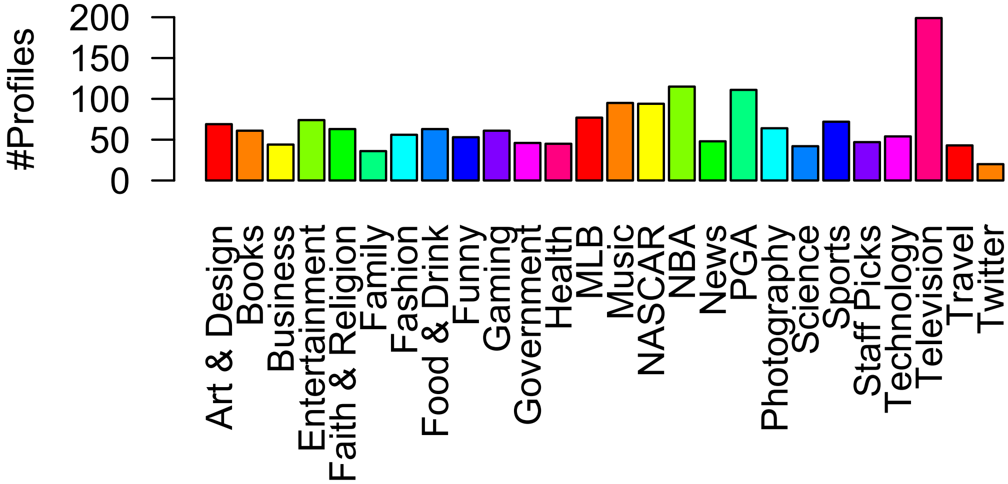


Fig. 1: "Who to follow" profile distribution over 26 Twitter interest categories (https://twitter.com/who_to_follow/) as of September 2014 [56].

Attribute	Funny	Entertainment	Sports	Government	News	NBA	Television	Music
Male	7	6	**12**	**3**	**5**	**15**	8	23
Female	**9**	**10**	5	2	3	6	**11**	**32**
$>$ 25y.o.	6	8	-	**4**	**6**	-	-	20
$\leq$ 25y.o.	**9**	**9**	-	2	3	-	-	**32**
Liberal	-	-	7	3	4	**10**	**10**	**28**
Conserv	-	-	**10**	3	6	8	9	18
Degree	-	**9**	-	**4**	**6**	8	9	19
School	-	8	-	2	3	**11**	**10**	**33**
Relation	-	**9**	-	**3**	**4**	8	-	27
Single	-	8	-	2	3	**10**	-	**29**
Children	-	-	-	**3**	**5**	**10**	-	25
No	-	-	-	2	4	9	-	**29**
$\geq \$35K$	6	8	7	**4**	**6**	-	-	21
$< \$35K$	**9**	**9**	**8**	2	3	-	-	**32**
Satisfied	-	-	**8**	**3**	4	-	**10**	26
Dissatisf	-	-	6	2	4	-	9	**33**
Narcissist	-	-	-	2	3	**11**	9	**32**
Not	-	-	-	**3**	**4**	9	**10**	24

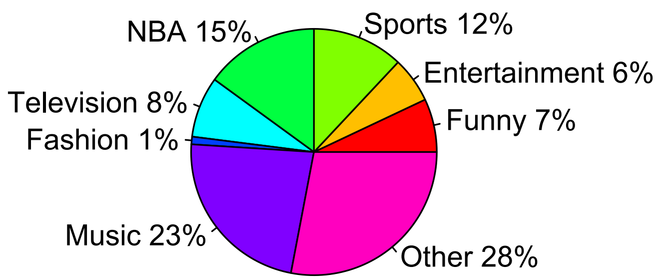
TABLE 4: Mean interest values in percentages % for 8 most followed interest categories (followed by more than 1/4 of users in the dataset) averaged within the groups of users with contrastive demographics e.g., Male vs. Female ("-" represents no statistical significance; highlighted cells stand for p-value ≤ 0.05; p-value ≤ 0.001 otherwise).

not balanced across properties. For instance, we had about three times more people perceived to be liberals than people perceived to be conservative in our dataset. Thus, the accuracy of a classifier that predicts "liberal" rather than conservative without even looking at the data would be far higher than 50%. The AuC metric takes inbalanced data into account, and is equivalent to the probability of correctly classifying two randomly selected users one from each class (e.g. one liberal and one conservative user).

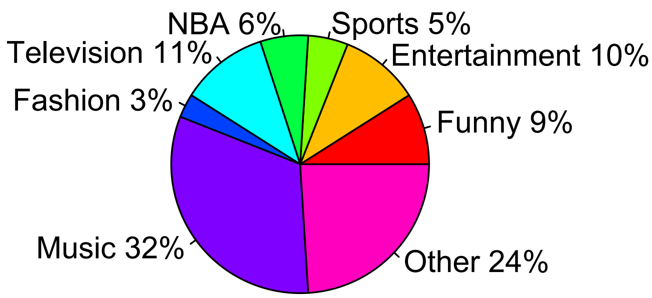
We present the results obtained using **Context** (interest) vs. **Content** (tweet) features. We further compare our results with the state-of-the-art performance achieved using **Content**† features alone (binary word unigrams extracted from user tweets) reported by [60].

Our results show that even when we do not have access to the text that users produce, and only know which popular accounts they follow, we are still able to accurately predict many of their personal traits, in quality comparable with state-of-the-art approaches that access textual information. Accuracy comparable to textual methods is achieved for many traits (highlighted in bold in Table 5), including age, education, relationship status and income level. However, for the highly verbose attribute of gender content features still yield better results compared to context features.

There has been lots of work on inferring user attributes for texts published in social media. However, the above results indicate even if a user is completely passive in online social network, and only "listens" to content generated by others,



(a) Male



(b) Female

Fig. 2: The distribution of the mean interests for male vs. female users.

we may still accurately predict user demographics by simply examining the kind of content they consume.

4.3. Feature Impact Analysis

Figure 3 visualizes the predictive power of various user interests, showing the regression coefficients learned by our models for predicting psycho-demographic attributes. Features that have higher absolute coefficients have a stronger impact on the prediction.

The figure shows that some interests are predictive of one attribute value (red), and some of an opposite value (blue). It also shows a dendrogram for attributes (rows) and interests (columns), and groups data based on row and column similarities using a hierarchical clustering algorithm. We observe that the most similar interests are gaming and sports, humor and music, PGA and business, health and books, Twitter and travel. We find that the most similar attributes (in terms of interest area coefficients) are income, age, education and life satisfaction.

The coefficient analysis indicates that:

- Humor, gaming and sport interests are predictive of users perceived to be young; Family, technology and business interests are predictive of users perceived to be older;
- Business and PGA interests are predictive of users perceived to be conservative; Government, music and gaming interests are predictive of users perceived to be liberal;
- Unsurprisingly, faith and religious interests are predictive of users perceived to be religious; Gaming and health are predictive of users perceived not to be religious;
- Travel interests and MLB and NASCAR are predictive of users perceived to be satisfied with life, while humor, news and music interests are predictive of users perceived to be dissatisfied with life;
- Technology (as well as "staff picks") correlate with users perceived to have higher education, while television and music interests are correlated with users perceived to be less educated (high-school level of education or less);
- Books, travel, health and technology interests correlate with higher income perceptions; Music, humor and entertainment correlate with lower income perceptions.

The annotators did not have any access to user interests during their annotation. Our findings that perceived user properties can be accurately predicted using user interests alone highlight the importance and strength of correlations between demographics and user interests.

5. Conclusions

We studied the relations between user interests, as captured by Twitter accounts they follow, and their perceived psycho-demographics traits. Our results indicate that such interests, which relate to the social embedding of the user in the network, can indeed provide important insights regarding the properties of the user.

Our technique is based on the fact that there are significant differences between what users of different demographic groups tend to be interested in. We have used these differences to build a system for predicting psycho-demographic traits.

Our results show that in fact such interest areas features are as predictive of some demographic traits as the texts generated by users. Some users who are privacy-aware may limit access to the text they produce, for example by restricting access to their posts or messages to other users. Similarly, other users may express themselves in a very careful manner, in an attempt to make a certain desired impression on others, or in an attempt to limit the information about them available to strangers. despite this, our results indicate that it is sometimes sufficient to observe what a user follows and what content they consume, to be able to accurately infer many traits of a user. Such content may appear to users are less "intrusive", and they may not be as cautious with them as they are with the textual content they produce.

Attribute	Context	Content	Content†
Gender	0.76	0.90	0.90
Age	**0.61**	0.61	0.66
Political	**0.67**	0.68	0.72
Religion	**0.57**	0.57	0.63
Education	**0.70**	0.71	0.77
Relationship	**0.57**	0.57	0.63
Children	0.58	0.66	0.72
Income	**0.67**	0.68	0.73
Life Satisf	0.60	0.67	0.72
Narcissist	**0.60**	0.61	0.68

TABLE 5: Attribute classification results reported as AuC (area under the ROC curve) using context (interest) vs. content (tweet) features.

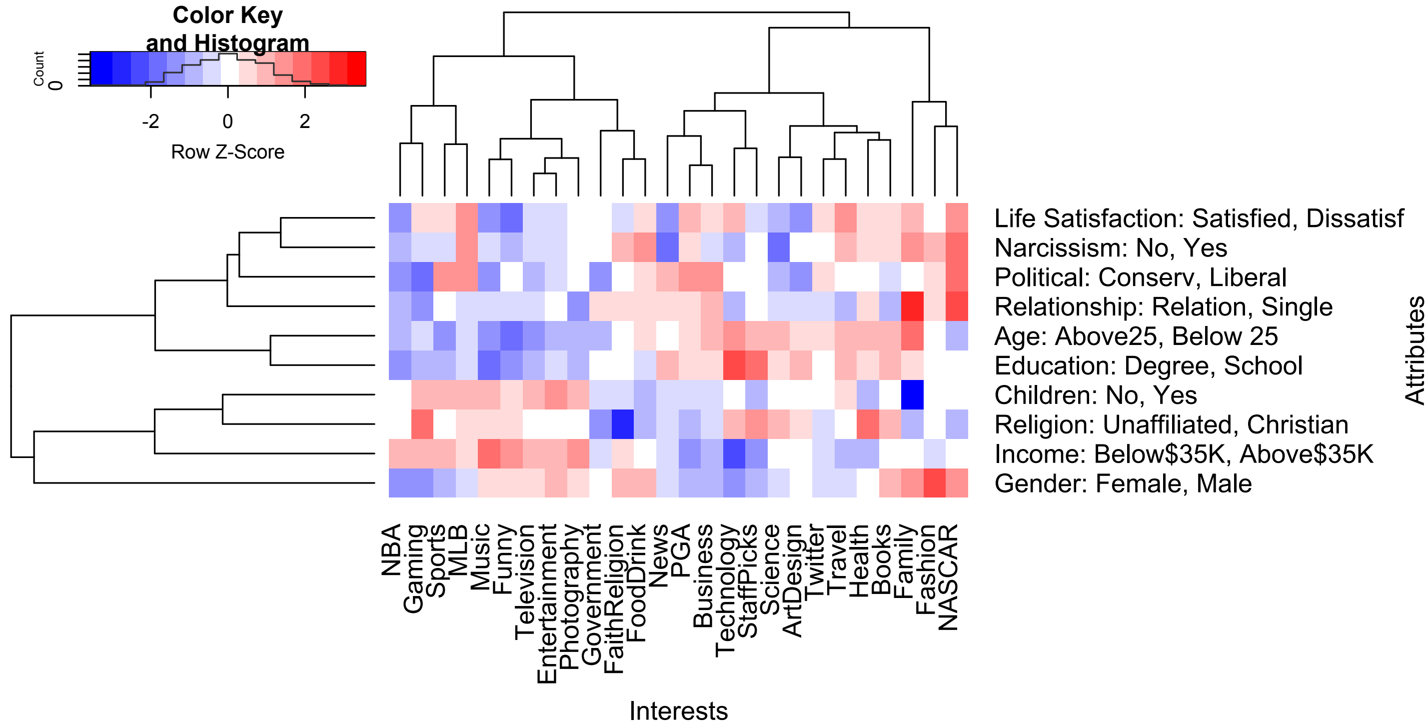


Fig. 3: The relation between user interests and perceived psycho-demographic attributes on Twitter. Every attribute (row) has two values ordered by color red (left) and blue (right) e.g., Satisfied (red), Dissatisfied (blue). The dendrograms shown on the top and on the left cluster attributes (rows) and interests (columns) by similarity using a hierarchical clustering algorithm.

Further, there are many settings where we have access to user interests but not texts produced by the user. For example, a user's browsing history is an indication of things they are interested in, but contains no text produced by the user. Our results show that in such settings it may still be possible to predict a user's psycho-demographic profile from these interests. Our models, relying on interest features, are therefore, an excellent alternative or complementary approach to the existing methods that rely on user communications.

Many questions are left open for future research. First, can a more fine-grained hierarchy of user interest improve the accuracy of our methods? For instance, it might be possible to cluster followed accounts based on the information contained in them, and use this more detailed feature as a predictor.

Second, could we use transfer learning methods to take interests learned in one domain (such as interest areas inferred based on browsing history or purchased products), map them to the interest areas examined in this work, and use this mapping to predict psycho-demographic traits?

Finally, could one uncover additional relations between user interest areas and other behaviors in online social networks, such as the sentiments or emotions displayed?

References

[1] H. A. Schwartz, J. C. Eichstaedt, M. L. Kern, L. Dziurzynski, R. E. Lucas, M. Agrawal, G. J. Park, S. K. Lakshmikanth, S. Jha, M. E. Seligman *et al.*, "Characterizing geographic variation in well-being using tweets," in *ICWSM*, 2013.

[2] D. Bamman, J. Eisenstein, and T. Schnoebelen, "Gender identity and lexical variation in social media," *Journal of Sociolinguistics*, vol. 18, no. 2, pp. 135–160, 2014.

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

[4] C. Brandenburg, "Newest way to screen job applicants: A social networker's nightmare, the," *Fed. Comm. LJ*, vol. 60, p. 597, 2007.

[5] A. Acquisti and C. M. Fong, "An experiment in hiring discrimination via online social networks," *Available at SSRN 2031979*, 2013.

[6] Y. Bachrach, "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*. IEEE, 2015, pp. 1–10.

[7] D. Rao, D. Yarowsky, A. Shreevats, and M. Gupta, "Classifying latent user attributes in Twitter," in *Proceedings of SMUC*, 2010, pp. 37–44.

[8] J. D. Burger, J. Henderson, G. Kim, and G. Zarrella, "Discriminating gender on Twitter," in *Proceedings of EMNLP*, 2011, pp. 1301–1309.

[9] D. Rao, M. Paul, C. Fink, D. Yarowsky, T. Oates, and G. Coppersmith, "Hierarchical Bayesian models for latent attribute detection in social media," in *Proceedings of ICWSM*, 2011.

[10] M. Pennacchiotti and A. M. Popescu, "A machine learning approach to Twitter user classification," in *Proceedings of ICWSM*, 2011, pp. 281–288.

[11] F. A. Zamal, W. Liu, and D. Ruths, "Homophily and latent attribute inference: Inferring latent attributes of Twitter users from neighbors," in *Proceedings of ICWSM*, 2012.

[12] Y. Bachrach, T. Graepel, P. Kohli, M. Kosinski, and D. Stillwell, "Your digital image: factors behind demographic and psychometric predictions from social network profiles," in *AAMAS*, 2014, pp. 1649–1650.

[13] S. Volkova, Y. Bachrach, M. Armstrong, and V. Sharma, "Inferring latent user properties from texts published in social media (demo)," in *Proceedings of AAAI*, 2015.

[14] S. Volkova and Y. Bachrach, "On predicting sociodemographic traits and emotions from communications in social networks and their implications to online self-disclosure," *Cyberpsychology, Behavior, and Social Networking*, vol. 18, no. 12, pp. 726–736, 2015.

[15] D. Hovy, "Demographic factors improve classification performance," *Proceedings of ACL*, 2015.

[16] J. Golbeck, C. Robles, M. Edmondson, and K. Turner, "Predicting personality from Twitter," in *Proceedings of SocialCom/PASSAT*, 2011.

[17] Y. Bachrach, M. Kosinski, T. Graepel, P. Kohli, and D. Stillwell, "Personality and patterns of Facebook usage," in *Proceedings of ACM WebSci*, 2012, pp. 24–32.

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

[19] D. Preoţiuc-Pietro, S. Volkova, V. Lampos, Y. Bachrach, and N. Aletras, "Studying user income through language, behaviour and affect in social media," *PloS one*, vol. 10, no. 9, p. e0138717, 2015.

[20] J. Bollen, H. Mao, and X. Zeng, "Twitter mood predicts the stock market," *Journal of Computational Science*, vol. 2, no. 1, pp. 1–8, 2011.

[21] D. Preotiuc-Pietro, J. Eichstaedt, G. Park, M. Sap, L. Smith, V. Tobolsky, H. A. Schwartz, and L. Ungar, "The role of personality, age and gender in tweeting about mental illnesses," in *Proceedings of the NAACL Workshop*, 2015.

[22] P. Resnick and H. R. Varian, "Recommender systems," *Communications of the ACM*, vol. 40, no. 3, pp. 56–58, 1997.

[23] D. M. Pennock, E. Horvitz, S. Lawrence, and C. L. Giles, "Collaborative filtering by personality diagnosis: A hybrid memory-and model-based approach," in *Proceedings of the Sixteenth conference on Uncertainty in artificial intelligence*. Morgan Kaufmann Publishers Inc., 2000, pp. 473–480.

[24] Y. Bachrach, S. Ceppi, I. A. Kash, P. Key, F. Radlinski, E. Porat, M. Armstrong, and V. Sharma, "Building a personalized tourist attraction recommender system using crowdsourcing," in *Proceedings of the 2014 international conference on Autonomous agents and multi-agent systems*. International Foundation for Autonomous Agents and Multiagent Systems, 2014, pp. 1631–1632.

[25] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl, "Application of dimensionality reduction in recommender system-a case study," DTIC Document, Tech. Rep., 2000.

[26] Y. Koren, R. Bell, and C. Volinsky, "Matrix factorization techniques for recommender systems," *Computer*, no. 8, pp. 30–37, 2009.

[27] Y. Bachrach, R. Herbrich, and E. Porat, "Sketching algorithms for approximating rank correlations in collaborative filtering systems," in *SPIRE*, 2009.

[28] Y. Bachrach, E. Porat, and J. S. Rosenschein, "Sketching techniques for collaborative filtering," in *IJCAI*, Pasadena, California, July 2009.

[29] Y. Bachrach and R. Herbrich, "Fingerprinting ratings for collaborative filteringtheoretical and empirical analysis," in *SPIRE*. Springer, 2010, pp. 25–36.

[30] P. Li and C. König, "b-bit minwise hashing," in *Proceedings of the 19th international conference on World wide web*. ACM, 2010, pp. 671–680.

[31] A. Dasgupta, R. Kumar, and T. Sarlós, "Fast locality-sensitive hashing," in *KDD*. ACM, 2011, pp. 1073–1081.

[32] Y. Bachrach and E. Porat, "Sketching for big data recommender systems using fast pseudo-random fingerprints," in *ICALP*, 2013, pp. 459–471.

[33] Y. Bachrach, Y. Finkelstein, R. Gilad-Bachrach, L. Katzir, N. Koenigstein, N. Nice, and U. Paquet, "Speeding up the xbox recommender system using a euclidean transformation for inner-product spaces," in *RecSys*, 2014.

[34] Y. Bachrach and E. Porat, "Fingerprints for highly similar streams," *Information and Computation*, 2015.

[35] S. Volkova, G. Coppersmith, and B. Van Durme, "Inferring user political preferences from streaming communications," in *Proceedings of ACL*, 2014, pp. 186–196.

[36] W. Youyou, M. Kosinski, and D. Stillwell, "Computer-based personality judgments are more accurate than those made by humans," *PNAS*, 2015. [Online]. Available: http://www.pnas.org/content/early/2015/01/07/1418680112.abstract

[37] K. Filippova, "User demographics and language in an implicit social network," in *Proceedings of EMNLP-CoNLL*, 2012.

[38] M. D. Conover, B. Gonçalves, J. Ratkiewicz, A. Flammini, and F. Menczer, "Predicting the political alignment of Twitter users," in *Proceedings of Social Computing*, 2011.

[39] A. Culotta, N. K. Ravi, and J. Cutler, "Predicting the demographics of twitter users from website traffic data," in *Proceedings of AAAI*, 2015.

[40] P. Bhattacharya, M. B. Zafar, N. Ganguly, S. Ghosh, and K. P. Gummadi, "Inferring user interests in the twitter social network," in *Proceedings of the 8th ACM Conference on Recommender systems*. ACM, 2014, pp. 357–360.

[41] P. Kapanipathi, P. Jain, C. Venkataramani, and A. Sheth, "User interests identification on twitter using a hierarchical knowledge base," in *The Semantic Web: Trends and Challenges*. Springer, 2014, pp. 99–113.

[42] J. Eisenstein, B. O'Connor, N. A. Smith, and E. P. Xing, "Diffusion of lexical change in social media," *PloS one*, vol. 9, no. 11, p. e113114, 2014.

[43] Y. Bachrach, T. Graepel, G. Kasneci, M. Kosinski, and J. Van Gael, "Crowd iq: aggregating opinions to boost performance," in *AAMAS*, 2012, pp. 535–542.

[44] M. Kosinski, Y. Bachrach, G. Kasneci, J. Van-Gael, and T. Graepel, "Crowd iq: Measuring the intelligence of crowdsourcing platforms," in *WebSci*. ACM, 2012, pp. 151–160.

[45] P. Welinder, S. Branson, P. Perona, and S. J. Belongie, "The multidimensional wisdom of crowds," in *NIPS*, 2010, pp. 2424–2432.

[46] Y. Bachrach, T. Graepel, T. Minka, and J. Guiver, "How to grade a test without knowing the answers—a bayesian graphical model for adaptive crowdsourcing and aptitude testing," *ICML*, 2012.

[47] M. Salek, Y. Bachrach, and P. Key, "Hotspotting-a probabilistic graphical model for image object localization through crowdsourcing." in *AAAI*, 2013.

[48] B. Shalem, Y. Bachrach, J. Guiver, and C. M. Bishop, "Students, teachers, exams and moocs: Predicting and optimizing attainment in web-based education using a probabilistic graphical model," in *ECML/PKDD*. Springer, 2014, pp. 82–97.

[49] J. R. Landis and G. G. Koch, "The measurement of observer agreement for categorical data," *biometrics*, pp. 159–174, 1977.

[50] M. Eltaher and J. Lee, "User profiling of flickr: Integrating multiple types of features for gender classification," *Journal of Advances in Information Technology Vol*, vol. 6, no. 2, 2015.

[51] Q. Fang, J. Sang, C. Xu, and M. Hossain, "Exploiting interaction with multimedia information for relational user attribute inference," 2015.

[52] C. Beller, R. Knowles, C. Harman, S. Bergsma, M. Mitchell, and B. Van Durme, "I'm a belieber: Social roles via self-identification and conceptual attributes," in *Proceedings of ACL*, 2014.

[53] G. Coppersmith, M. Dredze, and C. Harman, "Quantifying mental health signals in twitter," in *ACL Workshop on Computational Linguistics and Clinical Psychology*, 2014.

[54] R. Cohen and D. Ruths, "Classifying political orientation on Twitter: It's not easy!" in *Proceedings of ICWSM*, 2013.

[55] Y. Lewenberg, Y. Bachrach, and S. Volkova, "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*. IEEE, 2015, pp. 1–10.

[56] P. Gupta, A. Goel, J. Lin, A. Sharma, D. Wang, and R. Zadeh, "WTF: The who to follow service at Twitter," in *Proceedings of WWW*, 2013, pp. 505–514.

[57] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and d. Duchesnay, "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011. [Online]. Available: http://scikit-learn.org/

[58] N. A. Smith, "Log-linear models," 2004.

[59] D. Bamman, J. Eisenstein, and T. Schnoebelen, "Gender in Twitter: Styles, stances ands social networks," Tech. Rep., 2012.

[60] S. Volkova, B. Van Durme, D. Yarowsky, and Y. Bacharach, "Tutorial on social media predictive analytics," in *Proceedings of NAACL*, June 2015.