

Automatically Infer Human Traits and Behavior from Social Media Data

Shimei Pan

Department of Information Systems
University of Maryland, Baltimore County
shimei@umbc.edu

Tao Ding

Department of Information Systems
University of Maryland, Baltimore County
taoding01@umbc.edu

Abstract

Given the complexity of human minds and their behavioral flexibility, it requires sophisticated data analysis to sift through a large amount of human behavioral evidence to model human minds and to predict human behavior. People currently spend a significant amount of time on social media such as Twitter and Facebook. Thus many aspects of their lives and behaviors have been digitally captured and continuously archived on these platforms. This makes social media a great source of large, rich and diverse human behavioral evidence. In this paper, we survey the recent work on applying machine learning to infer human traits and behavior from social media data. We will also point out several future research directions.

1 Introduction

People currently spend a significant amount of time on social media to express opinions, interact with friends and families, share ideas and thoughts, provide status updates and organize/participate events and activities. According to Nielsen's 2016 Social Media report, on average, Gen X (ages 35-49) spent 6 hours and 58 minutes and Millennials (ages 18-34) spent 6 hours and 19 minutes per week on social media in the US. As a result, many aspects of their lives have been digitally captured and continuously archived on social media.

With recent advent of big data analytics and the availability of a large amount of user-generated content, data-driven human trait and behavioral analysis has increasingly been used to better understand human minds and predict human behavior. Prior research has demonstrated that by analyzing the information in a user's social media account, we can infer many latent user attributes such as political leaning [Pennacchiotti and Popescu, ; O'Banion and Birnbaum, ; Kosinski *et al.*, ; Benton *et al.*,], brand preferences [Pennacchiotti and Popescu, ; Yang *et al.*, a], emotions [Kosinski *et al.*,], mental disorders [De Choudhury *et al.*, ; Vedula and Parthasarathy,], personality [Kosinski *et al.*, ; Schwartz *et al.*, ; Liu *et al.*,], substance use [Kosinski *et al.*, ; Ding *et al.*,] and sexual orientation [Kosinski *et al.*,]. In addition, understanding individual traits and behavior has numerous real life applications including public health [Giota

and Kleftras, 2013], marketing [Yang *et al.*, a; Ding and Pan, 2016] and politics [Chirumbolo and Leone, 1].

We believe user generated data on social media is ideal for data-driven user trait and behavior analysis due to its unique characteristics: (1) large scale: it includes behaviors of a large number of social media users (e.g., millions users); (2) comprehensive: it contains a large number of behavioral markers from diverse sources (e.g., text posts, image posts, likes, and friendship); (3) longitudinal: it follows behaviors of social media users continuously over a long period of time (e.g., years); (4) objective: the analysis is based on the natural behavioral data automatically, continuously and objectively collected in an open environment.

In this survey, we summarize the recent advances in automated human traits and behavior inference from social media data. We included 24 studies published as full-length papers at one of the top AI, data mining, social media and multi-disciplinary behavior conferences and journals in the last ten years. We focus on papers that employ machine learning techniques to automatically infer or predict latent attributes or behaviors of individuals from public social media data.

2 Overview of Studies

Table 1 lists the 24 papers included in our survey. For each paper, we summarize information related to the dataset such as the "Platform" from which the social media data were collected (e.g., Twitter, Facebook, Reddit, Instagram and Quora) and "Size", which is the number of people with ground truth labels in the dataset. In addition, we also lists "Source Data Type", to indicate the different types of social media data used in the study. Here, *text* refers to user-generated text data that include user posts (e.g., tweets or status update on Facebook), profile description, and comments; *like* refers to things a social media user likes such as books, people, music, movies, TV shows, photos and products; *user profile* includes demographic information (gender, age occupation, relationship status etc.) and aggregated statistics of a user account (the number of friends, followers, followings etc.); *posting activity* which includes a set of statistics describing a user's posting behavior on social media such as the total number of tweets/retweets, the average number of posts per day, the number of likes/replies for each post; *image* which includes the profile and background photos and also the image posts shared on social media; *social network* which refers to social

Table 1: Summary of Studies

Paper	Platform	Source Data Type	Size	Predicted Target
[Pennacchiotti and Popescu,]	Twitter	tweet, user profile	10,338 6,000 10,000	political leaning ethnicity user preference
[Yang <i>et al.</i> , b]	Twitter	tweet	1,029 8,029 15,038	hashtag adoption (Politic) (Movie) (Random)
[De Choudhury <i>et al.</i> ,]	Twitter	post, social network, user profile	476	depression
[Zhang and Pennacchiotti,]	Facebook	like	13,619	purchase behavior
[O’Banion and Birnbaum,]	Twitter	tweet, social network	4,200 1,643	political leaning (Election day) (Tweetcast)
[Kosinski <i>et al.</i> ,]	Facebook	likes	58,000 46,027 1,200 18,833 9,752	age, gender, personality relationship substance use religion political leaning
[Schwartz <i>et al.</i> ,]	Facebook	post	75,000	age, gender, personality
[Chen <i>et al.</i> ,]	Reddit	post	799	human value
[Gao <i>et al.</i> , 2014]	Twitter	tweet, retweet	5,387 2,593	attitude (Topic:Fracking) (Topic:Vaccination)
[Wu <i>et al.</i> ,]	Facebook	like	70,520	personality
[Yang <i>et al.</i> , a]	Twitter	tweet	608	brand preference
[Preoȃiuc-Pietro <i>et al.</i> , a]	Twitter	tweet, user profile	5,191	occupation
[Lee <i>et al.</i> ,]	Twitter	user profile, tweet, social network	1,902 1,859	retweet action (Public safety) (Bird flu)
[Shen <i>et al.</i> ,]	Facebook	like, tweet	1,327	personality
[Bhargava <i>et al.</i> , 2015]	Facebook	post	488	interest
[Song <i>et al.</i> , b]	Twitter, Facebook, Quora	post, bio description	1,607	interest
[Hu <i>et al.</i> ,]	Twitter	tweet	9,800	occupation
[Song <i>et al.</i> , a]	Twitter, Facebook, LinkedIn	tweet, social network	5,436	volunteerism
[Liu <i>et al.</i> ,]	Twitter	profile picture	66,000	personality
[Benton <i>et al.</i> ,]	Twitter	tweet, social network	102,328 500 383 396	topic engagement friend recommendation age, gender political leaning
[Ding <i>et al.</i> ,]	Facebook	status, like	1,200	substance use
[Vedula and Parthasarathy,]	Twitter	tweet, social network	3215	depression
[Preoȃiuc-Pietro <i>et al.</i> , b]	Twitter	tweet	3,938	political leaning
[Singh <i>et al.</i> ,]	Instagram	caption, comments, image	699	cyberbullying

connections between different user accounts such as friendship networks on Facebook and follower/following relations on Twitter.

Although social media provide us an opportunity to easily track a large number of heterogeneous user behavioral data, the characteristics of social media data also bring significant challenges to data analysis. For example, the text and images are unstructured data. Making sense out of unstructured data is always a big challenge. User likes is very sparse and high dimensional. For example, the likes data in [Ding *et al.*,] includes 10 millions unique like dimensions. It is also not easy to search and analyze a large social network graph efficiently.

Finally, “Predicted Target” in Table 1 describes the predicted individual characteristics. We categorize them into

either explicit user characteristics or latent user characteristics. *Explicit User Characteristics* refer to observable user attributes that can be easily collected from a social media platform. The most common explicit user characteristics include demographics (e.g., age, gender, race etc.), and user preferences (e.g., likes). *Latent User Characteristics* refer to user attributes that cannot be observed easily, and thus need to be assessed by more sophisticated psychometric evaluations. Typical latent user characteristics include personality, human values, depression and addiction. Since typical psychometric tests include dozens or even hundreds of questions (e.g. to assess detailed personality, the IPIP-300 test has 300 questions), it is often difficult to collect the ground truth to assess latent user characteristics at a large scale.

In summary, as shown in table 1, Twitter is the most common social media platform used in these studies. This is mainly due to the relative ease of accessing user data on Twitter using its APIs. In terms of the type of user data involved in these studies, text (e.g., tweets or status updates on Facebook) is most common, followed by likes and social networks. Since different types of user data require different analysis techniques (e.g., using graph analytics for social networks and natural language processing for text data), user data type will have significant impact on the machine learning algorithms employed. The size of the datasets varies significantly from as many as 100K people to as few as 383. In general, user attributes declared directly on social media (e.g., topic engagement on Twitter) can be obtained relatively easily at a large scale. In contrast, for user traits/behavior that require sophisticated psychometric evaluation (e.g., depression), their ground truth datasets tend to be smaller. Finally, the predicted user characteristics are quite diverse, ranging from demographics (e.g., age, gender, race, income and occupation), latent user traits (e.g., personality and values), mental disorders (e.g., addiction and depression) and online or real world behaviors (e.g., purchase behavior and cyberbullying).

In the following, we summarize the typical data analysis methods that are used to infer individual characteristics from social media data.

3 Inference Methods

Figure 1 shows the typical architecture of such a system. One or more types of user-generated data are extracted from a social media account. For each type of user data such as text or image, a set of features are extracted (we call this “single-view feature extraction/learning”). The features from each view are then combined together to form a single user representation (we call this “multiview feature learning”). Finally, the combined user features are used to predict various human traits and behaviors using typical machine learning methods such as SVM classification or linear/logistic regression.

There are three main challenges when inferring user traits and behavior from social media data: (1) small labeled training datasets since the ground truth human trait and behavior data are hard to collect at a large scale. (2) unstructured and high dimensional user data. For example, there could be millions of unigram and bigram features to represent the texts from a user. (3) heterogeneous user data. Frequently we need to combine user data from different modality (e.g., text and images) and types (e.g., structured and unstructured) to paint a complete picture of a user.

Based on how these challenges are addressed, we categorize the studies in our survey along three dimensions. As shown in Table 2, first, we categorize them based on whether they employ a separate unsupervised feature learning stage to take advantage of a large amount of unsupervised data (2-stage) or they simply use labeled data directly in supervised feature extraction and prediction (1-stage). In general, when the labeled ground truth dataset is small, unsupervised feature learning can boost a system’s performance significantly. To address the second challenge and avoid the “curse of dimensionality” problem, we categorize these sys-

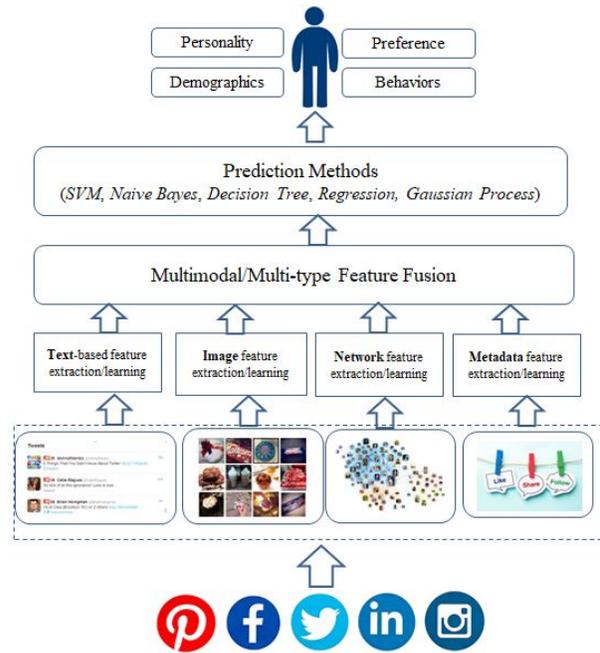


Figure 1: A Typical System Architecture

tems based on how they extract, select and learn a small number of features from unstructured data (e.g., text and image). The features can be (a) “human engineered” where existing domain/linguistic knowledge is used to construct and select a small number of features or (b) constructed via “supervised selection”, which refers to systems that select relevant features based on their correlations with the ground truth; or (c) constructed via “unsupervised feature learning”, which refers to systems that automatically learn a small number latent features based on unlabeled social media data. In terms of the third challenge, we categorize these studies based on how they combine features from different views together. If they simply concatenate features from different views together, we label them as “concatenate”. If they employ machine learning to fuse information from different views together, we label them as “fusion”. We also list different machine learning algorithms employed for user traits and behavior prediction.

Since feature extraction, unsupervised feature learning and multiview fusion play important roles in developing such a system, in the following , we describe each topic in detail.

3.1 Basic Feature Extraction

Basic Feature extraction is often the first step that maps the information in one’s social media account into a meaningful and easy to manipulate feature representation. Here, we focus on extracting features from text and images since they are unstructured information and thus more difficult to represent. We will also briefly discuss how to extract features from social networks.

The most commonly used text features are unigrams [O’Banion and Birnbaum, ; PreoŃiu-Pietro *et al.*, b] and the LIWC (Linguistic Inquiry and Word Count) features. A unigram is the term frequency computed for each vocab-

Table 2: Summary of Analysis Methods

Paper	Stage	Dimension Reduction	Fusion	Prediction Method
[Pennacchiotti and Popescu,]	2-stage	human engineered unsupervised learning	concatenation	Decision Trees
[Yang <i>et al.</i> , b]	1-stage	human engineered	NA	SVM
[De Choudhury <i>et al.</i> ,]	2-stage	human engineered unsupervised learning	concatenation	SVM
[Zhang and Pennacchiotti,]	2-stage	human engineered unsupervised learning	concatenation	SVM, Naive Bayes Logistic regression
[O’Banion and Birnbaum,]	1-stage	human engineered	concatenation	SVM
[Kosinski <i>et al.</i> ,]	2-stage	unsupervised learning	NA	Logistical/Linear regression
[Schwartz <i>et al.</i> ,]	2-stage	human engineered unsupervised learning	concatenation	SVM, Linear regression
[Chen <i>et al.</i> ,]	1-stage	human engineered	NA	Logistic regression Naive Bayes
[Gao <i>et al.</i> , 2014]	1-stage	supervised selection	NA	Lasso regression Least squares regression
[Wu <i>et al.</i> ,]	1-stage	human engineered supervised selection	NA	Lasso regression
[Yang <i>et al.</i> , a]	1-stage	human engineered	NA	SVM
[PreoŃiuç-Pietro <i>et al.</i> , a]	2-stage	human engineered unsupervised learning	concatenation	Gaussian Processes
[Lee <i>et al.</i> ,]		human engineered	concatenation	SVM, Random Forest Naive Bayes, Adaboost Logistic regression
[Shen <i>et al.</i> ,]	1-stage	human engineered	concatenation	Linear Regression
[Bhargava <i>et al.</i> , 2015]	1-stage	human engineered	NA	Hierarchical clustering
[Song <i>et al.</i> , b]	2-stage	unsupervised learning	fusion	SVM, Least squares regression Multi-view multi-task learning
[Hu <i>et al.</i> ,]	2-stage	human engineered unsupervised learning	NA	SVM
[Song <i>et al.</i> , a]	2-stage	unsupervised learning	fusion	SVM, Decision Trees Random Forest Graph-based Learning
[Liu <i>et al.</i> ,]	1-stage	human engineered	NA	Linear Regression
[Benton <i>et al.</i> ,]	2-stage	human engineered unsupervised learning	fusion	SVM, Multi-task learning
[Ding <i>et al.</i> ,]	2-stage	unsupervised learning	fusion	SVM
[Vedula and Parthasarathy,]	1-stage	human engineered	concatenation	Decision Trees
[PreoŃiuç-Pietro <i>et al.</i> , b]	2-stage	human engineered unsupervised learning	concatenation	Logistic regression
[Singh <i>et al.</i> ,]	1-stage	human engineered supervised selection	concatenation	Bagging classifier

ulary word in a text corpus (e.g., a corpus of all the Facebook posts). Sometimes, unigrams can be weighted based on their informativeness (e.g., based on Inverse Document Frequency or IDF [Benton *et al.*,]). Since robust inference often requires repeated word occurrence, low-frequency

words are frequently filtered out. In addition to individual words, meaningful phrases can be extracted by keeping only ngrams (e.g., bigrams) with high point-wise mutual information (PMI) [Schwartz *et al.*, ; Hu *et al.*,].

LIWC features are human engineered features that are

constructed based on the psycholinguistic dictionary LIWC. It groups words into psychologically meaningful categories. Empirical results have confirmed that LIWC features are capable of detecting meaning and providing a broader range of social and psychological insights such as feelings, personality, values and motivations. LIWC includes 81 features in five categories such as psychological Processes (e.g., emotional, cognitive, sensory, and social processes), Relativity (e.g., words about time, the past, the future), Personal Concerns (e.g., occupation, financial issues, health), and other dimensions (e.g., punctuation and swear words). LIWC also includes writing style features such as word complexity (e.g., words with more than 6 characters). Many of the systems in our survey used LIWC features [Pennacchiotti and Popescu, ; De Choudhury *et al.*, ; Schwartz *et al.*, ; Chen *et al.*, ; Preoțiu-Pietro *et al.*, b; Vedula and Parthasarathy, ; Singh *et al.*,].

Sometimes, customized vocabulary is used to extract informative words related to a prediction task. For example, [Preoțiu-Pietro *et al.*, b] defined 12,000 political terms in order to select informative unigrams pertaining to politics.

The image data such as profile pictures and photo posts, may contain rich information about individual characteristics. Each image is often represented as a vector of pixels, each pixel is represented by a number (in black and white photos) or three numbers using the RGB color scheme. In addition to the raw image features, meaningful high-level features such as color, facial expressions and postures can be extracted from images [Liu *et al.*,], which are then correlated with a user's traits and behavior. Besides general image features, [Singh *et al.*,] extracts demographics and a taxonomy-based object categories (presence of tattoos, graffiti, drug) from Instagram images to facilitate the detection of cyberbullying.

Egocentric social network features are frequently used to characterize the social relations of a user [De Choudhury *et al.*, ; Hong *et al.*, ; Benton *et al.*, ; Vedula and Parthasarathy,]. An egocentric network is defined as a network containing a single actor (ego), all the actors that an ego is connected to (alters), and all the links between the alters. For each ego, a set of network features are frequently extracted such as network size, betweenness centrality, normalized ego betweenness, cluster coefficient and normalized brokerage. We may also compute measures to assess network homophily such as average age difference between the ego and the alters.

3.2 Unsupervised Single View Feature Learning

Although the ground truth user traits and behavior data are costly to collect at a large scale, it is relatively easy to obtain a large amount of unlabeled user data from social media. The unsupervised feature learning algorithms are used to discover latent features from unlabeled social media data. Here we review typical unsupervised feature learning algorithms.

Singular Value Decomposition (SVD) is a mathematical technique that is frequently used for dimension reduction [De Lathauwer *et al.*,]. Given any $m * n$ matrix A , the algorithm will find matrices U , V and W such that $A = UWW^T$. Here U is an orthonormal $m * n$ matrix, W is a diagonal $n * n$ metric and V is an orthonormal $n * n$ matrix. Dimensionality reduction is done by computing $R = U * W_r$

where W_r neglects all but the r largest singular values in the diagonal matrix W .

Principle Component Analysis (PCA) is a popular dimensionality reduction mechanism used to eliminate highly correlated variables. PCA can be implemented using SVD. SVD and PCA have been used to learn a low-dimension representation from a bag-of-words representation of social media posts [Benton *et al.*, ; Ding *et al.*,], likes [Kosinski *et al.*, ; Ding *et al.*,], and social networks [Benton *et al.*,].

Latent Dirichlet Allocation (LDA) is a generative graphical model that allows sets of observations to be explained by unobserved latent groups [Blei *et al.*,]. In natural language processing, LDA is frequently used to learn a set of topics from a large number of documents. The topics are distributions of words that are frequently interpretable. In [Schwartz *et al.*, ; Ding *et al.*,], LDA is employed to learn topics from a user's social media posts.

GloVe is an unsupervised learning algorithm originally designed to learn vector representations of words based on aggregated global word-word co-occurrence statistics from a text corpus [Pennington *et al.*,]. GloVe employs a global log bilinear regression model that combines the advantages of global matrix factorization with that of local context window-based methods. GloVe has been applied to Facebook status updates [Preoțiu-Pietro *et al.*, 2015; Ding *et al.*,] and likes [Ding *et al.*,] to learn a dense feature vector for each word/like. To summarize all the words or likes from a user, we can use a vector aggregation functions such as *average*.

Recently, there is a new generation of neural network-based feature learning methods that employs self-taught learning to automatically derive a feature representation from examples automatically constructed from a large amount of unlabeled social media data.

Autoencoder (AE) is a neural network-based feature learning method [Hinton and Salakhutdinov,]. It learns an identity function so that the output is as close to the input as possible. Although an identity function seems a trivial function to learn, by placing additional constraints (e.g., to make the number of neurons in the hidden layer much smaller than that of the input), we can still force the system to uncover latent structures in the data.

Word Embedding with Word2Vec is a neural network-based method originally designed to learn dense vector representations for words [Mikolov *et al.*,]. The intuition behind the model is the Distributional Hypothesis, which states words that appear in the same context have similar meanings. There are two models for training a representation of word: continuous bag of word (CBOW) and skipgram (SG) model. CBOW predicts target word from one or more context words, while SG predicts one or more context words from target word. The models are frequently trained using either a hierarchical softmax function (HS) or negative sampling (NS) [Mikolov *et al.*,]. To process social media posts, the word2vec model is applied to learn a vector representation of each word. Then a simple average of all the word vectors by the same user is used to represent all the posts of a user [Benton *et al.*, ; Ding *et al.*,]. In addition, word embeddings can be used to produce word clusters [Preoțiu-Pietro *et al.*, a; Preoțiu-Pietro *et al.*, b] and domain lexicons (e.g., depres-

sion lexicon) [Vedula and Parthasarathy,].

Document Embedding with Doc2Vec is an extension of Word2Vec, which produces a dense low dimensional feature vector for a sentence or a paragraph. There are two Doc2Vec models: Distributed Memory (DM) and Distributed Bag-of-Words (DBOW). Given a sequence of tokens, DM can simultaneously learn a vector representation for each individual word token and a vector for the entire sequence. In DM, each sequence of words (e.g. a paragraph) is mapped to a sequence vector (e.g., paragraph vector) and each word is mapped to a unique word vector. The paragraph vector and one or more word vectors are aggregated to predict a target word in the context. DBOW learns a global sequence vector to predict tokens randomly sampled from a sequence. Unlike DM, DBOW only learns a vector for the entire sequence. It does not learn vectors for individual tokens (e.g., words). Neither does it use a local context window since the words for prediction are randomly sampled from the entire sequence. To characterize an individual's social media posts, a document for each user is created by aggregating every post of the same user.

3.3 Multi-view Feature Fusion

To obtain a single, comprehensive and coherent user representation based on all the social media data available, we need to combine user features from different views together. In addition to simply concatenating features extracted from different views, we can also apply machine learning algorithms to systematically fuse them together. We categorize these fusion methods into two types: (a) unsupervised feature learning that doesn't require any supervised data and (b) supervised multi-task learning.

There are two types unsupervised learning algorithms for multiview feature fusion : (1) Canonical Correlation Analysis and (b) Deep Canonical Correlation Analysis.

Canonical Correlation Analysis (CCA) CCA is a statistical method that explores the relationships between two multivariate sets of variables (vectors) [Hardoon *et al.*,]. Given two feature vectors, CCA tries to find a linear transformation of each feature vector so that they are maximally correlated. CCA has been used in [Sargin *et al.*, ; Chaudhuri *et al.*, ; Kumar and Daumé, ; Sharma *et al.*, ; Ding *et al.*,].

Deep Canonical Correlation Analysis (DCCA) DCCA aims to learn highly correlated deep architectures, which can be a non-linear extension of CCA [Andrew *et al.*,]. The intuition is to find a maximally correlated representation of two feature vectors by passing them through multiple stacked layers of nonlinear transformation [Andrew *et al.*,]. Typically, there are three steps to train DCCA: (1) using a denoising autoencoder to pretrain each single view. (2) computing the gradient of the correlation of top-level representation. (3) tuning parameters using back propagation to optimize the total correlation.

The features learned from multiple views are often more informative than those from a single view. Comparing with single-view user feature extraction and learning, multi-view learning achieved significantly better performance in predicting demographics [Benton *et al.*,], politic leaning [Benton

et al.,] and substance use [Ding *et al.*,].

Multi-task learning (MTL) is a supervised learning method to combine multi-view user data together. It tries to jointly train multiple prediction tasks at the same time to exploit the commonalities and differences across tasks. In [Song *et al.*, a], the authors collected multi-view user data from different platforms (e.g., Twitter, Facebook, LinkedIn accounts of the same user) and predict volunteerism from two tasks: user-centric analysis and network-centric analysis. Finally, they linearly fused these two components to enhance the final prediction.

4 Discussion and Future Directions

Large-scale social media-based user trait and behavior analysis is an emerging multidisciplinary field with the potential to transform human trait and behavior analysis from controlled small scale experiments to large scale studies of natural human behavior in an open environment. This provides us a new opportunity to explore the interactions of a large number of individual, social and environmental factors simultaneously, which would not be possible with traditional methods that use small samples. The insight gained from these studies could be valuable to help better understand the human minds and decision making process. It will also provide empirical evidence to help public health providers and police makers to improve mental health care and to combat public health threats (e.g., addiction and obese). Due to the privacy concerns on accessing user data on social media and the sensitive nature of the inferred user characteristics, if not careful, there could be significant privacy consequences and ethical implications. So far, most of the studies in our survey focused primarily on technical contributions. There should be more discussions on ethical considerations when conducting research in this field.

There are also several promising directions for future research. Since individual traits and behavior are highly correlated, building a prediction model that simultaneous infer multiple correlated traits and behavior should yield better performance than predicting each trait/behavior separately. Most existing studies only predict one user attribute/behavior at a time. More research should be conducted to jointly train and predict multiple user attributes together for better performance.

It is also common for a user to have multiple accounts on different social media platforms. Recently, new technologies have been developed to link different social media accounts of the same user together [Abel *et al.*,]. With this linked data, it is possible to perform novel cross-platform user trait and behavior analysis such as (1) domain bias analysis that focuses on studying the impact of domain or social media platform on user trait and behavior prediction, (2) domain adaptation that addresses how to adjust prediction models trained on one platform (e.g., Twitter) to predict the traits and behavior on another platform (e.g., Facebook). So far, there is some initial work on domain bias analysis and correction in personality prediction [Kiliç and Pan,]. More research is needed in order to develop more robust tools for human trait and behavior analysis.

References

- [Abel *et al.*,] F. Abel, E. Herder, G. Houben, N. Henze, and D. Krause. Cross-system user modeling and personalization on the social web. *UMUAI*, 23(2-3).
- [Andrew *et al.*,] G. Andrew, R. Arora, J. Bilmes, and K. Livescu. Deep canonical correlation analysis. In *ICML*.
- [Benton *et al.*,] A. Benton, R. Arora, and M. Dredze. Learning multiview embeddings of twitter users. In *ACL*, volume 2.
- [Bhargava *et al.*, 2015] P. Bhargava, O. Brdiczka, and M. Roberts. Unsupervised modeling of users' interests from their facebook profiles and activities. In *IUI*, pages 191–201. ACM, 2015.
- [Blei *et al.*,] D. Blei, A. Ng, and M. Jordan. Latent dirichlet allocation. *JMLR*, 3(Jan).
- [Chaudhuri *et al.*,] K. Chaudhuri, S. Kakade, K. Livescu, and K. Sridharan. Multi-view clustering via canonical correlation analysis. In *ICML*.
- [Chen *et al.*,] J. Chen, G. Hsieh, J. Mahmud, and J. Nichols. Understanding individuals' personal values from social media word use. In *CSCW*.
- [Chirumbolo and Leone,] A. Chirumbolo and L. Leone. Personality and politics: The role of the hexaco model of personality in predicting ideology and voting. *Personality and Individual Differences*, 49(1).
- [De Choudhury *et al.*,] M. De Choudhury, M. Gamon, S. Counts, and E. Horvitz. Predicting depression via social media. *ICWSM*, 13.
- [De Lathauwer *et al.*,] L. De Lathauwer, B. De Moor, and J. Vandewalle. A multilinear singular value decomposition. *SIMAX*, 21(4).
- [Ding and Pan, 2016] T. Ding and S. Pan. Personalized emphasis framing for persuasive message generation. *EMNLP*, 2016.
- [Ding *et al.*,] T. Ding, W. Bickel, and S. Pan. Multi-view unsupervised user feature embedding for social media-based substance use prediction. In *EMNLP*.
- [Gao *et al.*, 2014] H. Gao, J. Mahmud, J. Chen, J. Nichols, and Michelle X. Zhou. Modeling user attitude toward controversial topics in online social media. In *ICWSM*, 2014.
- [Giota and Kleftharas, 2013] K. Giota and G. Kleftharas. The role of personality and depression in problematic use of social networking sites in greece. *Cyberpsychology: Journal of Psychosocial Research on Cyberspace*, 7(3), 2013.
- [Hardoon *et al.*,] D. Hardoon, S. Szedmak, and J. Shawe-Taylor. Canonical correlation analysis: An overview with application to learning methods. *Neural computation*, 16(12).
- [Hinton and Salakhutdinov,] G. Hinton and R. Salakhutdinov. Reducing the dimensionality of data with neural networks. *science*, 313(5786).
- [Hong *et al.*,] L. Hong, A. Doumith, and B. Davison. Co-factorization machines: modeling user interests and predicting individual decisions in twitter. In *WSDM*.
- [Hu *et al.*,] T. Hu, H. Xiao, J. Luo, and T. Nguyen. What the language you tweet says about your occupation. In *ICWSM*.
- [Kiliç and Pan,] D. Kiliç and S. Pan. Analyzing and preventing bias in text-based personal trait prediction algorithms. In *ICTAI*.
- [Kosinski *et al.*,] M. Kosinski, D. Stillwell, and T. Graepel. Private traits and attributes are predictable from digital records of human behavior. *PNAS*, 110(15).
- [Kumar and Daumé,] A. Kumar and H. Daumé. A co-training approach for multi-view spectral clustering. In *ICML*.
- [Lee *et al.*,] K. Lee, J. Mahmud, J. Chen, M. Zhou, and J. Nichols. Who will retweet this? detecting strangers from twitter to retweet information. *TIST*, 6(3).
- [Liu *et al.*,] L. Liu, D. Preotiuc-Pietro, Z. Samani, M. Moghaddam, and L. Ungar. Analyzing personality through social media profile picture choice. In *ICWSM*.
- [Mikolov *et al.*,] T. Mikolov, I. Sutskever, K. Chen, G. Corrado, and J. Dean. Distributed representations of words and phrases and their compositionality. In *NIPS*.
- [O'Banion and Birnbaum,] S. O'Banion and L. Birnbaum. Using explicit linguistic expressions of preference in social media to predict voting behavior. In *ASONAM*.
- [Pennacchiotti and Popescu,] M. Pennacchiotti and A. Popescu. A machine learning approach to twitter user classification. *ICWSM*, 11(1).
- [Pennington *et al.*,] J. Pennington, R. Socher, and C. Manning. Glove: Global vectors for word representation. In *EMNLP*, volume 14.
- [Preotiuc-Pietro *et al.*, a] D. Preotiuc-Pietro, V. Lampos, and N. Aletras. An analysis of the user occupational class through twitter content. In *ACL*, volume 1.
- [Preotiuc-Pietro *et al.*, b] D. Preotiuc-Pietro, Y. Liu, D. Hopkins, and L. Ungar. Beyond binary labels: political ideology prediction of twitter users. In *ACL*, volume 1.
- [Preotiuc-Pietro *et al.*, 2015] D. Preotiuc-Pietro, S. Volkova, V. Lampos, Y. Bachrach, and N. Aletras. Studying user income through language, behaviour and affect in social media. *PLoS one*, 10(9):e0138717, 2015.
- [Sargin *et al.*,] M. Sargin, E. Erzin, Y. Yemez, and A. Tekalp. Multimodal speaker identification using canonical correlation analysis. In *ICASSP*, volume 1.
- [Schwartz *et al.*,] A. Schwartz, J. Eichstaedt, M. Kern, L. Dzurzynski, S. Ramones, M. Agrawal, A. Shah, M. Kosinski, D. Stillwell, M. Seligman, et al. Personality, gender, and age in the language of social media: The open-vocabulary approach. *PLoS one*, 8(9).
- [Sharma *et al.*,] A. Sharma, A. Kumar, H. Daume, and D. Jacobs. Generalized multiview analysis: A discriminative latent space. In *CVPR*.
- [Shen *et al.*,] J. Shen, O. Brdiczka, and J. Liu. A study of facebook behavior: What does it tell about your neuroticism and extraversion? *Computers in Human Behavior*, 45.
- [Singh *et al.*,] V. Singh, S. Ghosh, and C. Jose. Toward multimodal cyberbullying detection. In *CHI*.
- [Song *et al.*, a] X. Song, Z. Ming, L. Nie, Y. Zhao, and T. Chua. Volunteerism tendency prediction via harvesting multiple social networks. *TOIS*, 34(2).
- [Song *et al.*, b] X. Song, L. Nie, L. Zhang, M. Liu, and T. Chua. Interest inference via structure-constrained multi-source multi-task learning. In *IJCAI*.
- [Vedula and Parthasarathy,] N. Vedula and S. Parthasarathy. Emotional and linguistic cues of depression from social media. In *Proceedings of the 2017 International Conference on Digital Health*.
- [Wu *et al.*,] Y. Wu, M. Kosinski, and D. Stillwell. Computer-based personality judgments are more accurate than those made by humans. *PNAS*, 112(4).
- [Yang *et al.*, a] C. Yang, S. Pan, J. Mahmud, H. Yang, and P. Srinivasan. Using personal traits for brand preference prediction. In *EMNLP*.
- [Yang *et al.*, b] L. Yang, T. Sun, M. Zhang, and Q. Mei. We know what@ you# tag: does the dual role affect hashtag adoption? In *WWW*.
- [Zhang and Pennacchiotti,] Y. Zhang and M. Pennacchiotti. Predicting purchase behaviors from social media. In *WWW*.