2018 IEEE International Conference on Big Data (Big Data)

Situation-Based Interpretable Learning for Personality Prediction in Social Media

Lei Zhang
Department of Computer Science
Virginia Tech
zhanglei@vt.edu

Liang Zhao
Department of Information Science and Technology
George Mason University
lzhao9@gmu.edu

Xuchao Zhang
Department of Computer Science
Virginia Tech
xuczhang@vt.edu

Wenmo Kong
Department of Psychology
George Mason University
wkong2@gmu.edu

Zitong Sheng
Department of Management and Entrepreneurship
Virginia Commonwealth University
shengz@vcu.edu

Chang-Tien Lu
Department of Computer Science
Virginia Tech
ctlu@vt.edu

***Abstract*—Predicting individuals personality traits with their social media profile has proved to be feasible, but researchers recently have run into bottlenecks on further improving the prediction accuracy. One major limitation is that existing studies failed to consider context information in predicting social media users' behaviors. In this paper, we adopted the DIAMONDS situation theory in psychology to capture the context information in Facebook posts. To solve this issue, we proposed a novel situation-based feature interaction learning model. In this study, we extracted situation features according to the DIAMONDS lexicon and computed the interaction values between these situation features and the commonly used n-gram features at the post level. Features at the post level were aggregated up to the user level using the averaging strategy. A group lasso penalty was employed to enforce strong heredity in the model, which addressed the overfitting challenge introduced by the interaction features. Empirical tests on a large-scale data set have demonstrated the effectiveness of the proposed method.**

***Index Terms*—personality, situation, psychology, social media, interaction learning**

I. Introduction

Research on online behaviors and online selves has largely benefited from the rapid development of social media and AI techniques. Social media extended the boundary of human behaviors. The time people spend on social media is continuously increasing these years. According to a report on www.socialmediatoday.com, teens now spend up to nine hours a day on social platforms. As a result, an inseparable part of people's daily life now is on the cyberspace. The digital traces in cyberspace, compared to offline behaviors, are easy to record and analyze, which is extremely valuable for research purposes. Previous studies have proved that online behaviors reflect actual personality instead of self-idealization [1]. Other studies have already started developing personality prediction models with social media data, albeit with limitations. Exploring the limitation of such personality prediction models is not only helpful for preventing privacy leaks but also of great value for psychological research.

Social media texts are widely used as indicators to predict users' traits or behaviors [2]–[9]. Existing studies, however, typically use simple bag-of-words models for feature extraction, thereby artificially limiting performance. One such limitation is that, feature extraction methods based on the bag-of-words model cannot capture the context from sentences. Without considering context information, features extracted from the text can be misleading. For example, introverted people generally tend to use more negative emotion words than extroverted people. In a situation that elicits unpleasant feelings, however, a person can express a lot of negative emotions just because he/she is in that situation, regardless of his/her personality. It would be problematic to predict ones introversion/extraversion level using their emotion words without considering the context.

In this paper, we propose a hierarchical interaction model incorporating domain knowledge of 'situations' from psychology. 'Situation' is another factor that can affect people's behaviors besides personality. The notion of 'situation' in psychology is naturally similar to the notion of 'context' in text mining. Figure 1 shows the intuition of our method. Personality is the variable to be predicted. The online digital traces can be seen as behaviors of social media users. All existing studies in this area were based on the intuition that personality is related to behaviors. However, most of them did not take situation into account, which is also related to personality.

Incorporating the situation factor in psychology, we propose a situation-based interaction regression model. This model utilizes the interactions between basic features (n-grams) and situation features to predict personality traits of social media users. The personality traits are defined based on the Five Factor Model (FFM), including openness to experience, conscientiousness, extraversion, agreeableness, and neuroticism. Situations are defined based on the Situational Big Eight DIAMONDS research. Interaction features will capture the co-occurrence information between basic behaviors and situations in social media users' posts. The use of interaction features can increase the prediction power of models while interpretability remains.

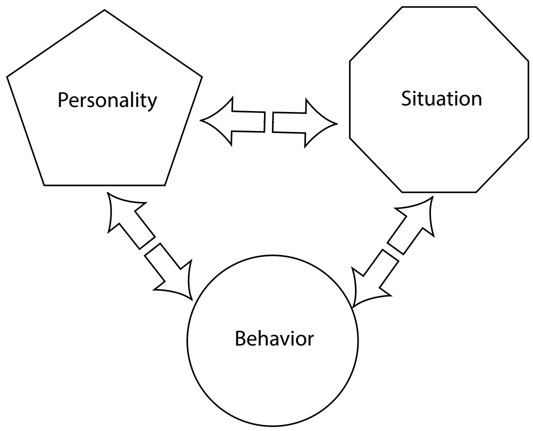

Fig. 1. Personality, situation, and behaviors are three strongly correlated factors in psychology [10]

The proposed situation-based approach will conjunctly address two critical challenges brought by the interaction features–sparsity of features and interpretability of model.

Sparsity of features: The number of text features can be tremendous. The features related to personality are only a small subset of all the features. The model should perform both variable selection and regularization to enhance the prediction accuracy and interpretability of the statistical model it produces.

Interpretability of model: Interpretability of the text regression model is vital for providing insights on psychology research (e.g., verifying lexical hypothesis). However, more complex models usually yield poor interpretation.

The main contributions of our study are summarized as follows:

- We are the first to apply interaction feature learning on social media users' personality prediction tasks.
- We explore the influence of adding DIAMONDS situations, a psychology situation theory framework, on prediction performance.
- We proposed a hierarchical interaction approach to model the co-occurrence of keywords and situations.
- Our method outperforms existing methods while it keeps the interpretability.

The rest of this paper is organized as follows. Section 2 introduces the psychology background of two domain-specific concepts used in this paper. Section 3 reviews background and related works. Section 4 presents our situation-based interaction learning model. Experiments on a real Facebook dataset are presented in Section 5, and the paper concludes with a summary of our research in Section 6.

II. Psychology Background

What makes people different? This is a question that psychologists have been trying to answer since half a century ago. Two camps of research, the personality trait psychologists and situationists, have different opinions on this question.

From the interactionist perspective, people's behaviors are highly related to both personality and situation. As our target is to predict personality with people's online behaviors, it is necessary to measure and quantify both personality and situation in a systematic way.

A. Personality

Personality is a set of stable inner features of people. There are various personality theories, but the Big Five model has emerged as one of the most well researched and widely accepted measures of personality structure. It has been proven that Big Five personality is one of the most useful factors for explaining and predicting human behaviors [11].

There are five dimensions of traits as defined by the Big Five model. Each dimension can be represented as a continuous value with high and low bounds (e.g., 0.0-5.0). Polar traits on one personality dimension represent opposite personalities. The Big Five model defines the five most fundamental dimensions of human traits, including:

- Openness: curious, intelligent, imaginative. High scorers tend to be artistic and sophisticated in taste and appreciate diverse views, ideas, and experiences.
- Conscientiousness: responsible, organized, persevering. Conscientious individuals are extremely reliable and tend to be high achievers, hard workers, and planners.
- Extraversion: outgoing, amicable, assertive. Friendly and energetic, extroverts draw inspiration from social situations.
- Agreeableness: cooperative, helpful, nurturing. People who score high in agreeableness are peace-keepers who are generally optimistic and trusting of others.
- Neuroticism: anxious, insecure, sensitive. Neurotics are moody, tense, and easily tipped into experiencing negative emotions.

B. Situation

Situation represents the outer factor that is not under the direct control of people. Situationists declared that people behave differently depending on the situation. The DIAMONDS model defined eight robust dimensions of situations characteristics identified by situationists [12]. The eight dimensions are

- Duty (*Does something need to be done?*)
- Intellect (*Is deep thinking required or desired?*)
- Adversity (*Are there external threats?*)
- Mating (*Is the situation sexually and/or romantically charged?*)
- Positivity (*Is the situation enjoyable?*)
- Negativity (*Does the situation elicit unpleasant feelings?*)
- Deception (*Is someone being untruthful or dishonest?*)
- Sociality (*Are social interaction and relationship formation possible, required, or desired?*)

The most recent studies on the personality-situation debate mainly have two directions: the interactionism and the syntheticism. Both of them suggest that people's behavior is highly correlated with both people's personality and situation.

The personality-situation debate and the new directions all inspired us to take the situation into account on the personality prediction tasks.

III. Related Work

Three research areas are mostly related to our work. First, as our task is to predict personality, we briefly introduce various personality measurement works, including traditional methods and the prediction models. Second, our study is closely related to the social media users' trait prediction/forecasting/reference studies. Last but not least, the technical part of our approach is related to the hierarchical interaction learning methods.

A. Personality Measurements

The traditional way to measure personality is asking people to fill out self-reported questionnaires/inventories such as the Revised NEO Personality Inventory (NEO PI-I) [13] or BFI [14]. Most of these inventories comprise either items that are self-descriptive sentences or, in the case of lexical measures, items that are single adjectives.

Languages, as the most common types of data, have been used in multiple papers as indicators for predicting personality. A typical method for linking language to psychological variables involves counting the number of words that belong to pre-defined lexicons such as LIWC [15]. Golbeck et al. conducted the first study that used LIWC features to predict Facebook users' personality [16]. The impressiveness of their results was limited by the small sample size. A similar dataset was used in a competition on Kaggle for personality prediction over Twitter messages, in which participants were provided language cues based on LIWC [17]. Many researchers tried linear and non-linear algorithms in this competition, yet results were limited by the LIWC features.

Instead of using domain knowledge in a pre-defined lexicon, Andrew et al. designed Open Vocabulary, a data-driven framework for predicting social media users' traits [18]. They defined the old lexicon-based feature engineering methods as 'closed vocabulary' and the data-driven n-grams/LDA topics as 'open vocabulary'. As the number of features is larger than the number of users, they used PCA for dimension reduction. The prediction power of open vocabulary outperforms the old ones. They improved their method in another paper [19] and published the toolkit based on open vocabulary [20].

There are also some papers that used website specific features to predict personality [21]–[23]. David and Michal created regression models to predict Big Five personality dimensions only using Facebook users' Likes data. On the 1-of-N encoding Likes data sorted by each user, they applied singular value decomposition on the user-Likes matrix as a dimension reduction method. Youyou et al. further compared the prediction power of Facebook Likes with people's judgment. Their conclusion was that computer-based personality judgments were more accurate than those made by humans [22]. Liu et al. tried to use profile pictures as the prediction input for both Facebook and Twitter users [23]. Similarly, Plank et al. built an MBTI personality prediction model using Twitter tweets [24].

B. Social Media User Properties Mining

Besides personality, social media can also reveal other characteristics of users [25]. The feasibility of predicting demographic information such as age [2], [3] and gender [4]–[9] has been proved in early years and improved in recent years. Another widely researched yet more difficult task is to predict mental health issues such as depression [26]–[28], post-traumatic stress disorder (PTSD) [29], [30], and suicide ideation [31], [32]. Some of these studies are contributive in helping people with mental health problems. It has been proved that severe diseases such as schizophrenia can also be quantified [33]. Other attributes that can be predicted include religion, race [21], occupation [34], political preferences [35], and so on. LIWC features and n-grams are also widely-used language features in these studies. LIWC can hardly capture context information as it uses the simple bag-of-words model. N-Gram features can cover a context window with n words. Extracting n-gram features with large n is extremely time consuming and not practical.

C. Interaction Features

High order feature interactions can significantly increase the complexity of models and thus improve the performance, yet there is also the challenge of feature selection to avoid overfitting problems [36]–[39]. Feature selection by considering feature interactions has long been attracting research interests [40], [41]. For example, to overcome the dimensionality issues introduced by interaction effects, two types of heredity constraints have been studied [42]: strong heredity in which an interaction effect can be selected into the model only if both of its corresponding linear effects have been selected; and weak heredity, in which an interaction effect can be selected if at least one of its corresponding linear effects has been selected [43]. Similarly, Haris et al. [44] explored different types of norms of the constraints. Lin et al. proposed a multiple-task interaction learning framework to aggregate more data to avoid overfitting [45].

IV. Situation-based Interaction Learning

A. Multiple-Instance Learning for Facebook Posts

The training set $D = (X, Y)$ consist of m bags $X =< X_1, ..., X_m >$ and their corresponding real-valued labels $Y =< y_1, ..., y_m >$. Each bag X_i has n_i instances $x_{i1}, ..., x_{in_i}$ and each instance x_{ij} is described by p_1 features. The goal is to determine a function f over the bag space $\mathbb{N}^{\mathbb{X}}$ which can make predictions

$$\hat{y_i} = f(X_i)$$

of label y_i of new bags X_i as accurately as possible.

There are various ways to define the search space of the function f mapping from the bag space to the output space, including linear models and non-linear models. The non-linear models always perform better because they have larger search spaces. However, non-linear models always have difficulties in interpreting the relationship between raw features and the dependent variables. The advantage of linear models is that they are simple and have good interpretability. The interpretability is valuable for psychological research and applications. The classic linear regression model can be written as

$$\hat{y_i} = x_i^T \beta + \epsilon_i, i = 1, ..., m$$

Where β is the weight vector for features and ϵ_i is the disturbance term or error variable. However, the label for each Facebook user corresponds to the information in multiple posts, which makes the problem a Multi-instance learning (MIL) problem.

In the multiple instance learning (MIL) paradigm, we are given labels for sets of instances. These sets are also known as bags or groups. The bag-level labels are assumed to be an association function (e.g., OR, average) of the unknown instance level labels. Treating our personality prediction problem as a MIL problem, one Facebook post is an instance, all the instances from a bag, either one dimension of personality score is the label for the bag. By using the average MIL assumption, the model can be written as:

$$\hat{y_i} = \frac{1}{n_i} \sum_{j=1}^{n_i} x_{ij}^T \beta + \epsilon_i$$

B. Fitting Personality Prediction Model with Interactions

To introduce the situation-based interaction into our model, three types of features will be used, including basic features, situation features, and interaction features. The basic features are the features used by most other works which represent the behavior of a Facebook user. The dimension of basic features is p_1. The situation features are new features introduced in this paper. The dimension of situation features is defined as p_2. The interaction of basic features and situation features will form the third type of feature with a resulting dimension of $p_1 * p_2$. The final model will be a linear mapping from the three types of features to the dependent variable.

$$\hat{y_i} = \frac{1}{n_i} \sum_{j=1}^{n_i} (x_{ij}^T \beta_1 + z_{ij}^T \beta_2 + < x_{ij} z_{ij}^T, \beta_3 >) + \epsilon_i \quad (1)$$

$x_{ij} z_{ij}^T$ are interaction terms between all the features in x and z. β_3 is a matrix of coefficients for interaction features. $< x_{ij} z_{ij}^T, \beta_3 >$ is the Frobenius inner product of $x_{ij} z_{ij}^T$ and β_3

Given a labeled training dataset, the loss function is

$$min_{\beta_1, \beta_2, \beta_3} \frac{1}{m} \sum_{i=1}^{m} (y_i - \hat{y}_i)^2$$

In equation (1), the features are extracted on the instance level. As the aggregation method for the multi-instance learning assumption is averaging, however, we can pre-compute

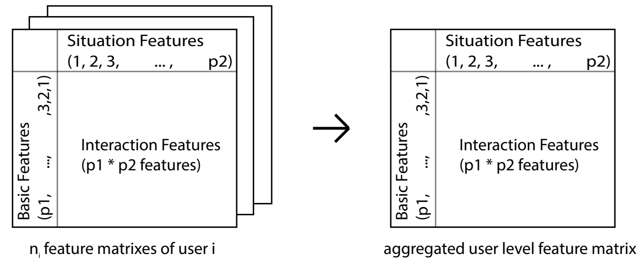

Fig. 2. Features extracted on instance level can be aggregated into user/bag level

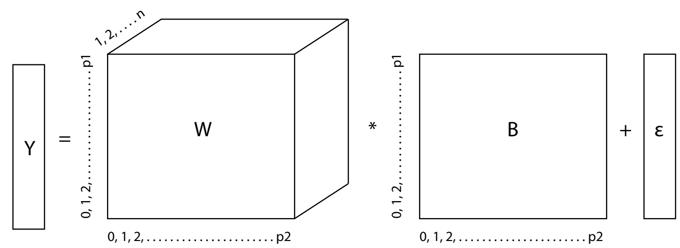

Fig. 3. Array notation of interaction modeling.

the features on bag level before feeding to the optimization algorithm. In other words, the model can be reduced and simplified as

$$\hat{y_i} = \frac{1}{n_i} x_i^T \beta_1 + z_i^T \beta_2 + < x_i z_i^T, \beta_3 > + \epsilon_i \quad (2)$$

Figure 2 shows the feature matrix aggregation process for user i. The raw data extracted for user i were n_i feature matrices with dimension of $(p_1 + 1) \times (p_2 + 1)$. The aggregated data for user i is a single $(p_1 + 1) \times (p_2 + 1)$ matrix.

For brevity, we write the model using array notation. We construct the $n \times (p_1 + 1) \times (p_2 + 1)$ array W as follows: for $i \in \{1, ..., m\}$, $j \in \{0, ..., p_1\}$, $k \in \{0, ..., p_2\}$,

$$W_{i,j,k} = \begin{cases} X_{i,j} Z_{i,k}, & \text{for } j \neq 0 \text{ and } k \neq 0 \\ X_{i,j}, & \text{for } k = 0 \text{ and } j \neq 0 \\ Z_{i,k}, & \text{for } j = 0 \text{ and } k \neq 0 \\ 1, & \text{for } j = k = 0 \end{cases}$$

Then (2) is equivalent to the model

$$y = W * B + \varepsilon$$

where B is the matrix of coefficients as in (2), and $W * B$ denotes the m-vector whose i^{th} element takes the form
$(W * B)i \equiv \sum_{j=0}^{p_1} \sum_{k=0}^{p2} W_{i,j,k} B_{j,k}$.

Figure 3 shows the formulation of the underlying interaction learning model.

The loss function is:

$$min_{B \in \mathbb{R}(p_1+1)(p_2+1)} \frac{1}{2n} ||y - W * B||_2^2$$

To overcome the dimensionality issues introduced by interaction effects, a simple way is to apply lasso penalty on all main effects and all interactions. This method is also called all pair lasso:

$$min_{B\in\mathbb{R}(p_1+1)(p_2+1)}\frac{1}{2n}||y-W*B||_2^2+\lambda||B||_1$$

All pair lasso can apply feature selection based on training data. However, it ignores the differences between the main effects and interactions. We adopt group Lasso penalties to induce strong heredity.

$$\begin{aligned}&min_{B\in\mathbb{R}(p_1+1)(p_2+1)}\{\frac{1}{2n}||y-W*B||_2^2+\\&+\lambda_1\sum_{j=1}^{p_1}||B_{j,.}||_2^2\\&+\lambda_2\sum_{k=1}^{p_2}||B_{.,k}||_2^2\\&+\lambda_3||B_{-0,-0}||_1\}\end{aligned}\tag{3}$$

The group lasso penalties can yield an estimator that obeys strong heredity. Different from Haris's FAMILY framework [44], the interaction features in the feature matrix are not the products of the first row and first column because we extracted interaction features on the instance (single Facebook post) level and aggregated them to the user level.

C. Parameter Optimization

The objective function in Equation (3) is convex because the loss function, regularization terms, and constraints are all convex. To solve the convex optimization problem with constraints, the alternating method of multipliers(ADMM) is a good option. ADMM can break the original large problem into smaller sub-problems that can be solved easily and fast. In ADMM form, our problem can be written as:

$$\begin{aligned}&min_{B\in\mathbb{R}(p_1+1)(p_2+1)}\{\frac{1}{2n}||y-W*B||_2^2+\\&+\lambda_1\sum_{j=1}^{p_1}||D_{j,0}||_2^2\\&+\lambda_2\sum_{k=1}^{p_2}||E_{0,k}||_2^2\\&+\lambda_3||F_{-0,-0}||_1\}\end{aligned}$$

subject to

$$\begin{aligned}&B-D=0\\&B-E=0\\&B-F=0\end{aligned}$$

The augmented Lagrangian can be rewritten as

$$\begin{aligned}&L_\rho(B,D,E,F,\Gamma_1,\Gamma_2,\Gamma_3)=\frac{1}{2n}||y-W*B||_2^2\\&+\lambda_1\sum_{j=1}^{p_1}||D_{j,0}||_2^2\\&+\lambda_2\sum_{k=1}^{p_2}||E_{0,k}||_2^2\\&+\lambda_3||F_{-0,-0}||_1\\&+<\Gamma_1,B-D>+<\Gamma_2,B-E>+<\Gamma_3,B-F>\\&+\rho/2||B-D||_F^2+\rho/2||B-E||_F^2+\rho/2||B-F||_F^2\end{aligned}\tag{4}$$

where $\Gamma_1,\Gamma_2,\Gamma_3$ are $(p_1+1)\times(p_2+1)$ dimensional dual variables of ADMM and ρ is the step size of the dual step.

By following the updating strategy proposed by Boyd et al. [46], the optimization problem can be divided into the following sub-tasks.

$$\begin{aligned}B^{k+1}&=argmin_B L_\rho(B,D^k,E^k,F^k,\Gamma_1^k,\Gamma_2^k,\Gamma_3^k)\\D^{k+1}&=argmin_D L_\rho(B^k,D,E^k,F^k,\Gamma_1^k,\Gamma_2^k,\Gamma_3^k)\\E^{k+1}&=argmin_E L_\rho(B^k,D^k,E,F^k,\Gamma_1^k,\Gamma_2^k,\Gamma_3^k)\\F^{k+1}&=argmin_F L_\rho(B^k,D^k,E^k,F,\Gamma_1^k,\Gamma_2^k,\Gamma_3^k)\\\Gamma_1^{k+1}&=argmin_{\Gamma_1} L_\rho(B^k,D^k,E^k,F^k,\Gamma_1,\Gamma_2^k,\Gamma_3^k)\\\Gamma_2^{k+1}&=argmin_{\Gamma_2} L_\rho(B^k,D^k,E^k,F^k,\Gamma_1^k,\Gamma_2,\Gamma_3^k)\\\Gamma_3^{k+1}&=argmin_{\Gamma_3} L_\rho(B^k,D^k,E^k,F^k,\Gamma_1^k,\Gamma_2^k,\Gamma_3)\end{aligned}$$

The parameters B,D,E,F,Γ are alternately solved by the proposed algorithm. It alternately optimizes each of the parameters in until an acceptable residual is achieved. The detailed optimization steps are described in more detail below.

1) Update B

$$\begin{aligned}B^{i+1}&=argmin_B\frac{1}{2n}||y-W*B||_2^2\\&+<\Gamma_1,B-D>+<\Gamma_2,B-E>\\&+\rho/2||B-D||_F^2+\rho/2||B-E||_F^2+\rho/2||B-F||_F^2\\&=argmin_B\frac{1}{2n}||y-W*B||_2^2\\&+\frac{2\rho^i}{2}||\frac{1}{3\rho^i}[\rho^i(D^i+E^i+F^i)\\&-(\Gamma_1+\Gamma_2+\Gamma_3)]-B||_F^2\end{aligned}\tag{5}$$

The sub-task of optimizing for B is similar to ridge regression problem except the coefficients are in a matrix, not a vector. The optimal solution can be efficiently computed as a closed-form solution.

2) Update D

$$D^{k+1} = argmin_D \lambda_1 \sum_{j=1}^{p_1} ||D_{j,.}||_2^2$$
$$+ < \Gamma_1, B - D^k > +\rho/2||B - D^k||_F^2$$
$$= argmin_D \frac{\rho}{2}||D - (B + \frac{\Gamma_1}{\rho})||_F^2 + \lambda_1 \sum_{j=1}^{p_1} ||D_{j,.}||_2^2 \tag{6}$$

We use the proximal algorithm to solve the optimization problem for D.

3) Update E

$$E^{k+1} = argmin_D \lambda_1 \sum_{k=1}^{p_2} ||E_{.,k}||_2^2$$
$$+ < \Gamma_1, B - E > +\rho/2||B - E||_F^2$$
$$= argmin_E \frac{\rho}{2}||E - (B + \frac{\Gamma_2}{\rho})||_F^2 + \lambda_2 \sum_{k=1}^{p_2} ||E_{.,k}||_2^2 \tag{7}$$

Minimizing fuction (4) with respect to E follows the same method as optimizing for D.

4) Update F

$$F^{k+1} = argmin_F \lambda_3 ||F_{-0,-0}|| + < \Gamma_3, B - F > +\rho/2||B - F||_F^2 \tag{8}$$

$$\begin{aligned} F_{0,.} &= B_{0,.} + \frac{\Gamma_{3_{0,.}}}{\rho} \\ F_{.,0} &= B_{.,0} + \frac{\Gamma_{3_{.,0}}}{\rho} \\ F_{j,k} &= sign\Big(B_{j,k} + \frac{\Gamma_{3_{j,k}}}{\rho}\Big)\Big(|B_{j,k} + \frac{\Gamma_{3_{j,k}}}{\rho}| - \frac{\lambda_3}{\rho}\Big)_+ \\ &\text{for } j \neq 0, k \neq 0. \end{aligned} \tag{9}$$

Optimizing for the main effects (the first row and first column of the coefficient matrix) have closed-form solutions.

Optimizing for the interaction variables ($W_{-0,-0}$) is a simple soft-thresholding problem.

5) Update Γ

The updating of the dual variables $\Gamma_1, \Gamma_2, \Gamma_3$ are as follows:

$$\begin{aligned} \Gamma_1^{k+1} &= \Gamma_1 + \rho(B - D) \\ \Gamma_2^{k+1} &= \Gamma_2 + \rho(B - E) \\ \Gamma_3^{k+1} &= \Gamma_3 + \rho(B - F) \end{aligned} \tag{10}$$

Algorithm 1 shows the basic steps of the optimization using ADMM.

V. Experiments

In this section, we evaluate the proposed situation-based interaction learning model on the myPersonality dataset. After the data set and experimental setup have been introduced, the effectiveness of the methods is evaluated against several existing methods.

ALGORITHM 1: Parameters optimization based on ADMM

Input: Data tensor W, one arbitrary dimension personality scores y
Output: Solution B
Initialize $\rho = 1$, $B, D, E, F, \Gamma_1, \Gamma_2, \Gamma_3$,;
Choose $\varepsilon^{pri} > 0$, $\varepsilon^{dual} > 0$;
repeat
  Update B by equation (5);
  Update D and E by equation (6) (7);
  Update F by equation (9);
  Update $\Gamma_1, \Gamma_2, \Gamma_3$ by equation (10);
  if $r > 10s$ **then**
    $\rho \leftarrow 2\rho$;
  else if $10r < s$ **then**
    $\rho \leftarrow \rho/2$;
  else
    $\rho \leftarrow \rho$;
  end
until $r < \varepsilon^{pri}, s < \varepsilon^{dual}$;

A. Dataset

We did all the experiments on the myPersonality dataset. myPersonality was a popular Facebook application that allowed users to take real psychometric tests and allowed researchers to record their psychological and Facebook profiles with explicit opt-in consent for reuse for research purposes. Currently, the database contains more than 6,000,000 test results, together with more than 4,000,000 individual Facebook profiles. The respondents came from various age groups, backgrounds, and cultures. They were highly motivated to answer honestly and carefully, as the only gratification that they receive for their participation was feedback on their results. The personality score we used in this study were measured with the International Personality Item Pool proxy for the NEO Personality Inventory Revised (NEO-PI-R). Participants were free to choose measures of different lengths, ranging from 20 to 100 items.

For our specific task of personality predicting, data were filtered by following the same rules described in the state-of-the-art paper [18]. Each instance/user must have both demographic information and big-5 personality labels. They must have posted more than 1000 words. Their age must be smaller than 65. In the dataset, 3,137,694 users have big5 labels. After applied all the above rules, we ended up getting 55,835 users in total as the data set for the experiments.

Same with the preprocessing method in Schwartz's paper, Happier Fun Tokenizer was applied on users' Facebook post texts. Happier Fun Tokenizer is an improved version of Happy Fun Tokenizer with optimization for Facebook emotions [1].

B. Experimental Setup

We mainly use two types of features in this paper. The basic features are n-gram features which represent the behaviors of Facebook users. The additional features are situation features which describe the situation of the Facebook users when they submit the posts.

[1] https://github.com/dlatk/happierfuntokenizing

For n-gram features, we first scan all the Facebook posts and keep only the n-grams used by more than 5% users in our dataset. Next, all the remaining n-grams will be filtered again according to a pointwise mutual information (PMI) threshold.

$$pmi(\text{n-gram}) = log\frac{p(\text{n-gram})}{\prod_{w\in\text{n-gram}} p(w)} \tag{11}$$

As suggested by the open-vocabulary approach [18], we kept n-grams with PMI values higher than $2 * length$, where $length$ is the number of words contained in the n-gram. For example, we keep all the bi-grams whose $pmi > 4.0$.

For situation features, we compute the per category usage percentage based on a predefined DIAMONDS[2] lexicon S8-LIWC [47]. The S8-LIWC contains 433 words chosen by domain experts to capture the eight situational dimensions. The percentage is calculated as:

$$\begin{aligned} &p(category|subject) \\ =&\frac{\sum\limits_{word\in category} freq(word, subject)}{\sum\limits_{word\in vocab(subject)} freq(word, subject)} \end{aligned} \tag{12}$$

After feature extraction, each Facebook user has 3459 uni-gram features, 1579 bi-gram features, 161 tri-gram features, and 8 situation features. The total of 5199 n-gram features and 8 situation features will generate 41,592 interaction features. As the purpose of the interaction features is to capture the context information of each Facebook post, the per-user interaction features were extracted at post level and averaged at the user level. The 41,592 features will be the input of our proposed method.

The methods used in the experiments are listed as follows:

- **LR_ngrams**: Least square linear regression without any penalties is the most basic baseline method. The feature set is n-gram. This method is in the range of open-vocabulary approach but doesn't use the context information. It should be considered as the most basic baseline.
- **Lassongrams**: Different Lasso models are built for corresponding dimensions of personality. The parameter of Lasso will be selected by grid search. The feature set is n-gram. This method is an improved version of LR_ngrams by adding the feature selection function. It is the major competing method.
- **AllPairLasso_interaction**: All pair Lasso is an interaction learning regression model with L1 norm on all the features. Basic features, situation features, and the interaction features are all used in this setting. The weight of the penalty term is selected by grid search. This method is a simplified version of the proposed method.
- **Hierarchical_interaction**: Our proposed method. This method considers all the three types of features. Group lasso penalties are applied to the rows and columns of feature matrix to achieve strong heredity. The weight of the penalty term is selected by grid search.

[2]https://www.bigeightdiamonds.com/

C. Results and Analysis

We train all the models on the same training set and evaluate them on the same test set. The training set was created by randomly selecting 75% of the whole dataset (41,876 participants). The remaining 25% data (13,959 participants) were used as the test set for evaluation. The main metric for comparison is Pearson correlation coefficients (or Pearson's r) between predicted values and the ground truth values of personalities. The larger the Pearson correlation coefficient is, the better the model performs.

Table I shows the comparison of the proposed model and baseline methods. All the Pearson correlation coefficients are significant ($p < 0.01$). Mean Absolute Error (MAE) is also reported in the table. The smaller the MAE is, the better the model is. The first two rows show the performance of n-gram based methods. The last two rows show the results after situation is considered. By comparing the results of n-gram based methods (in 1st and 2nd rows in table I) and the interaction learning methods (in 3rd and 4th rows in table I), it is clear that interaction features do have additional predictive power on personality dimensions. The improvements from Lasso_ngrams to LR_ngrams indicate that sparsity is still important even the data are extensive. It is reasonable because there are 41,876 instances and 41,592 features while training. The improvement from AllPairLasso_interaction to Hierarchical_interaction illustrate that the heredity/hierarchical formulation contributes to better performance as well. All the results follow the similar patterns as previous studies, which verified some conclusions drawn by these studies. Openness and Extraversion dimensions are relatively easy to predict. Agreeableness and Neuroticism are very difficult to predict using Facebook data.

For a better understanding, the results our model achieved can be compared with some existing standards. The Pearson correlation coefficient between predicted Openness and self-reported Openness is almost 0.4. By contrast, the participants' Facebook friends can judge more accurate (r=0.49) [22], but they may consider other information when judging (e.g., images, Facebook likes, friends). On the other hand, the reported test-retest correlations of Big Five self-report questionnaires typically range from 0.65 to 0.85. The test-retest correlations defined how inaccurate the labels are when used for training and evaluation. By using self-reported questionnaire results as the golden standard, it is certain that 0.65 defined the upper bound of any models can perform no matter what features are used. We admit that 0.65 may be reached in future model, at the same time we expect such model to also use tons of features that are not limited to texts. The closer the Pearson's r approaches 0.65, the more difficult it is to build the model. In this paper, we successfully updated the benchmark towards the upper bound by combining the psychology domain knowledge with machine learning techniques. Compared to Lasso_ngrams (the primary competing method), the proposed method Hierarchical_interaction improved the Pearson correlation coefficients by 0.023-0.043. These improvements are

TABLE I
COMPARISON OF PROPOSED MODEL AND BASELINE METHODS

Methods	Openness		Conscientiousness		Extraversion		Agreeableness		Neuroticism	
	r	MAE	r	MAE	r	MAE	r	MAE	r	MAE
LR_ngrams	0.344	0.504	0.289	0.575	0.316	0.627	0.228	0.566	0.278	0.647
Lasso_ngrams	0.372	0.490	0.316	0.561	0.341	0.613	0.249	0.553	0.298	0.634
AllPairLasso_interaction	0.374	0.487	0.326	0.559	0.335	0.610	0.262	0.544	0.298	0.631
Hierarchical_interaction	**0.399**	**0.478**	**0.348**	**0.550**	**0.364**	**0.599**	**0.292**	**0.533**	**0.321**	**0.619**

TABLE II
EXAMPLE OF INTERACTION FEATURES

High Openness	Low Openness
Situation + n-gram	Situation + n-gram
Mating + upon	Adversity +coming
Sociality + hold	Negativity +outta
Positivity +faster	Sociality + annoying
Mating + missed	Deception + often
Deception + book	Duty + lol

substantial when compared with the upper bound.

Another major advantage of our proposed method is the interpretability of our model. By using the interaction feature between n-grams and situations, the complexity of the model is increased while interpretability remains. The final model can be interpreted through the weights of the features. Table II showed example interactions between situation features and n-gram features for people who are high versus low in openness. As an illustration, a combination of the sociality 'situation' feature and the 'hold' n-gram feature identifies people who are high in openness, while a combination of the 'sociality' situation feature and the 'annoying' n-gram feature identifies people who are low in openness. This indicates that people who are high in openness may be more likely to 'hold' social events, and individuals who are low in openness may find social events 'annoying'. With the situation-based interaction learning model, we are able to go into details about how people express themselves differently under different situations, therefore increasing the accuracy of prediction.

VI. CONCLUSION

This paper presents a novel situation-based interaction learning model to further improve personality prediction with social media data. Existing methods either were not interpretable or failed to take context into consideration. Our work considers both basic features and situation features by using the interactions between them in prediction. Hierarchical constraints have been applied to the interaction learning problem to achieve sparsity and to avoid overfitting. We implement an efficient algorithm based on ADMM to get the optimal solution in a time-efficient manner. Overall, our model has successfully advanced the accuracy of existing personality prediction methods by incorporating situation information into personality prediction.

AI-based personality prediction techniques can serve as an important complement to traditional survey methods because of their advantages of rapidity and inexpensiveness, but these methods can only be functional when they are accurate enough. Following the great success of the Open-Vocabulary approach announced in 2013 [18], our work is the first attempt to update the benchmark in this area. Although skepticism still exists regarding the utility and appropriateness of using social media data in personality prediction, our study demonstrated some benefits of such methods and enriched this research field.

In the future, we plan to extend our framework by exploring more advanced feature extraction methods and situation measurement approaches, to further improve the accuracy of personality prediction with social media data.

ACKNOWLEDGEMENT

This work was supported by the National Science Foundation grant: #1755850.

REFERENCES

[1] M. D. Back, J. M. Stopfer, S. Vazire, S. Gaddis, S. C. Schmukle, B. Egloff, and S. D. Gosling, "Facebook profiles reflect actual personality, not self-idealization," *Psychological science*, vol. 21, no. 3, pp. 372–374, 2010.

[2] S. Goswami, S. Sarkar, and M. Rustagi, "Stylometric analysis of bloggers age and gender," in *Third International AAAI Conference on Weblogs and Social Media*, 2009.

[3] S. Rosenthal and K. McKeown, "Age prediction in blogs: A study of style, content, and online behavior in pre-and post-social media generations," in *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1*. Association for Computational Linguistics, 2011, pp. 763–772.

[4] R. Sarawgi, K. Gajulapalli, and Y. Choi, "Gender attribution: tracing stylometric evidence beyond topic and genre," in *Proceedings of the Fifteenth Conference on Computational Natural Language Learning*. Association for Computational Linguistics, 2011, pp. 78–86.

[5] M. Sap, G. Park, J. Eichstaedt, M. Kern, D. Stillwell, M. Kosinski, L. Ungar, and H. A. Schwartz, "Developing age and gender predictive lexica over social media," in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2014, pp. 1146–1151.

[6] M. Ciot, M. Sonderegger, and D. Ruths, "Gender inference of twitter users in non-english contexts," in *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, 2013, pp. 1136–1145.

[7] W. Liu and D. Ruths, "What's in a name? using first names as features for gender inference in twitter." in *AAAI spring symposium: Analyzing microtext*, vol. 13, no. 1, 2013, pp. 10–16.

[8] J. S. Alowibdi, U. A. Buy, and P. Yu, "Empirical evaluation of profile characteristics for gender classification on twitter," in *Machine Learning and Applications (ICMLA), 2013 12th International Conference on*, vol. 1. IEEE, 2013, pp. 365–369.

[9] D. Hovy, "Demographic factors improve classification performance," in *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, vol. 1, 2015, pp. 752–762.

[10] R. Wood and A. Bandura, "Social cognitive theory of organizational management," *Academy of management Review*, vol. 14, no. 3, pp. 361–384, 1989.

[11] R. R. McCrae and O. P. John, "An introduction to the five-factor model and its applications," *Journal of personality*, vol. 60, no. 2, pp. 175–215, 1992.

[12] J. F. Rauthmann, D. Gallardo-Pujol, E. M. Guillaume, E. Todd, C. S. Nave, R. A. Sherman, M. Ziegler, A. B. Jones, and D. C. Funder, "The situational eight diamonds: A taxonomy of major dimensions of situation characteristics." *Journal of Personality and Social Psychology*, vol. 107, no. 4, p. 677, 2014.

[13] P. T. Costa and R. R. MacCrae, *Revised NEO personality inventory (NEO PI-R) and NEO five-factor inventory (NEO-FFI): Professional manual*. Psychological Assessment Resources, Incorporated, 1992.

[14] O. P. John, L. P. Naumann, and C. J. Soto, "Paradigm shift to the integrative big five trait taxonomy," *Handbook of personality: Theory and research*, vol. 3, no. 2, pp. 114–158, 2008.

[15] J. W. Pennebaker, M. E. Francis, and R. J. Booth, "Linguistic inquiry and word count: Liwc 2001," *Mahway: Lawrence Erlbaum Associates*, vol. 71, no. 2001, p. 2001, 2001.

[16] J. Golbeck, C. Robles, M. Edmondson, and K. Turner, "Predicting personality from twitter," in *Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third Inernational Conference on Social Computing (SocialCom), 2011 IEEE Third International Conference on*. IEEE, 2011, pp. 149–156.

[17] C. Sumner, A. Byers, R. Boochever, and G. J. Park, "Predicting dark triad personality traits from twitter usage and a linguistic analysis of tweets," in *Machine learning and applications (icmla), 2012 11th international conference on*, vol. 2. IEEE, 2012, pp. 386–393.

[18] H. A. Schwartz, J. C. Eichstaedt, M. L. Kern, L. Dziurzynski, S. M. Ramones, M. Agrawal, A. Shah, M. Kosinski, D. Stillwell, M. E. Seligman *et al.*, "Personality, gender, and age in the language of social media: The open-vocabulary approach," *PloS one*, vol. 8, no. 9, p. e73791, 2013.

[19] G. Park, H. A. Schwartz, J. C. Eichstaedt, M. L. Kern, D. J. Stillwell, M. Kosinski, L. H. Ungar, and M. E. Seligman, "Automatic personality assessment through social media language," *Journal of Personality and Social Psychology*, vol. 108, pp. 934–952, 2014.

[20] H. A. Schwartz, S. Giorgi, M. Sap, P. Crutchley, L. Ungar, and J. Eichstaedt, "Dlatk: Differential language analysis toolkit," in *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, 2017, pp. 55–60.

[21] M. Kosinski, D. Stillwell, and T. Graepel, "Private traits and attributes are predictable from digital records of human behavior," *Proceedings of the National Academy of Sciences*, vol. 110, no. 15, pp. 5802–5805, 2013.

[22] W. Youyou, M. Kosinski, and D. Stillwell, "Computer-based personality judgments are more accurate than those made by humans," *Proceedings of the National Academy of Sciences*, vol. 112, no. 4, pp. 1036–1040, 2015.

[23] L. Liu, D. Preotiuc-Pietro, Z. R. Samani, M. E. Moghaddam, and L. H. Ungar, "Analyzing personality through social media profile picture choice." in *ICWSM*, 2016, pp. 211–220.

[24] B. Plank and D. Hovy, "Personality traits on twitterorhow to get 1,500 personality tests in a week," in *Proceedings of the 6th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, 2015, pp. 92–98.

[25] X. Zhang, Z. Chen, W. Zhong, A. P. Boedihardjo, and C.-T. Lu, "Storytelling in heterogeneous twitter entity network based on hierarchical cluster routing," in *Big Data (Big Data), 2016 IEEE International Conference on*. IEEE, 2016, pp. 1522–1531.

[26] M. De Choudhury, M. Gamon, S. Counts, and E. Horvitz, "Predicting depression via social media." *ICWSM*, vol. 13, pp. 1–10, 2013.

[27] G. Coppersmith, M. Dredze, C. Harman, K. Hollingshead, and M. Mitchell, "Clpsych 2015 shared task: Depression and ptsd on twitter," in *Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*, 2015, pp. 31–39.

[28] H. A. Schwartz, J. Eichstaedt, M. L. Kern, G. Park, M. Sap, D. Stillwell, M. Kosinski, and L. Ungar, "Towards assessing changes in degree of depression through facebook," in *Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*, 2014, pp. 118–125.

[29] G. Harman and M. H. Dredze, "Measuring post traumatic stress disorder in twitter," *In ICWSM*, 2014.

[30] T. Pedersen, "Screening twitter users for depression and ptsd with lexical decision lists," in *Proceedings of the 2nd workshop on computational linguistics and clinical psychology: from linguistic signal to clinical reality*, 2015, pp. 46–53.

[31] M. De Choudhury, E. Kiciman, M. Dredze, G. Coppersmith, and M. Kumar, "Discovering shifts to suicidal ideation from mental health content in social media," in *Proceedings of the 2016 CHI conference on human factors in computing systems*. ACM, 2016, pp. 2098–2110.

[32] G. Coppersmith, R. Leary, E. Whyne, and T. Wood, "Quantifying suicidal ideation via language usage on social media," in *Joint Statistics Meetings Proceedings, Statistical Computing Section, JSM*, 2015.

[33] M. Mitchell, K. Hollingshead, and G. Coppersmith, "Quantifying the language of schizophrenia in social media," in *Proceedings of the 2nd workshop on Computational linguistics and clinical psychology: From linguistic signal to clinical reality*, 2015, pp. 11–20.

[34] D. Preoţiuc-Pietro, V. Lampos, and N. Aletras, "An analysis of the user occupational class through twitter content," in *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, vol. 1, 2015, pp. 1754–1764.

[35] S. Volkova, G. Coppersmith, and B. Van Durme, "Inferring user political preferences from streaming communications," in *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, vol. 1, 2014, pp. 186–196.

[36] X. Wu, B. Shi, Y. Dong, C. Huang, L. Faust, and N. V. Chawla, "Restful: Resolution-aware forecasting of behavioral time series data," in *International Conference on Information and Knowledge Management (CIKM)*. ACM, 2018, pp. 1073–1082.

[37] X. Zhang, L. Zhao, A. P. Boedihardjo, C.-T. Lu, and N. Ramakrishnan, "Spatiotemporal event forecasting from incomplete hyper-local price data," in *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*. ACM, 2017, pp. 507–516.

[38] B. Wang, X. Zhang, C.-T. Lu, and F. Chen, "Water disaggregation via shape features based bayesian discriminative sparse coding," *arXiv preprint arXiv:1808.08951*, 2018.

[39] L. Zhao, J. Ye, F. Chen, C.-T. Lu, and N. Ramakrishnan, "Hierarchical incomplete multi-source feature learning for spatiotemporal event forecasting," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 2016, pp. 2085–2094.

[40] C. Huang, X. Wu, and D. Wang, "Crowdsourcing-based urban anomaly prediction system for smart cities," in *International Conference on Information and Knowledge Management (CIKM)*. ACM, 2016, pp. 1969–1972.

[41] X. Wu, Y. Dong, C. Huang, J. Xu, D. Wang, and N. V. Chawla, "Uapd: Predicting urban anomalies from spatial-temporal data," in *Joint European Conference on Machine Learning and Knowledge Discovery in Databases (ECML/PKDD)*. Springer, 2017, pp. 622–638.

[42] J. Bien, J. Taylor, and R. Tibshirani, "A lasso for hierarchical interactions," *Annals of statistics*, vol. 41, no. 3, p. 1111, 2013.

[43] X. Wu, B. Shi, Y. Dong, C. Huang, and N. Chawla, "Neural tensor factorization," *arXiv preprint arXiv:1802.04416*, 2018.

[44] A. Haris, D. Witten, and N. Simon, "Convex modeling of interactions with strong heredity," *Journal of Computational and Graphical Statistics*, vol. 25, no. 4, pp. 981–1004, 2016.

[45] K. Lin, J. Xu, I. M. Baytas, S. Ji, and J. Zhou, "Multi-task feature interaction learning," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 2016, pp. 1735–1744.

[46] S. Boyd, N. Parikh, E. Chu, B. Peleato, J. Eckstein *et al.*, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends® in Machine learning*, vol. 3, no. 1, pp. 1–122, 2011.

[47] D. G. Serfass and R. A. Sherman, "Situations in 140 characters: assessing real-world situations on twitter," *PloS one*, vol. 10, no. 11, p. e0143051, 2015.