"A pessimist sees the difficulty in every opportunity; an optimist sees the opportunity in every difficulty" – Understanding the psycho-sociological influences to it

Upendra Kumar, Vishal Singh Rana, Srinivas PYKL and Amitava Das

Indian Institute of Information Technology, Sri City, AP, India, 517588

`{upendra.k14, vishal.s14, srinivas.p amitava.das}@iiits.in`

Abstract

This paper presents an empirical study to understand *how psycho-sociological factors influence on optimism/pessimism at the individual level*. Optimists believe that future events are going to work out for the best; pessimists expect the worst. Their expectations manifest in their day-to-day behaviour and also reflects the way a person tweets or behaves in online social media. To this end, we have identified optimist/pessimist users from Twitter, analyzed their personality (psychological) and values & ethics (sociological) at the community level. Empirical analysis reveals some interesting insights and behavioral patterns related to user level optimism/pessimism in different combinations of psychological and sociological factors, are reported.

1 Introduction

Optimists are people who tend to have favourable outlook of their life whereas pessimists tend to derive negative interpretations from the events around them. Their approach towards life generally manifests in their day-to-day behaviour. For example, they have different mechanisms to cope with positive or negative events around them. The first computational model to estimate the degree of optimism/pessimism at tweet level as well as user level has been introduced by [Ruan et al.2016]. This paper could be seen as an extension to that work. Here we are trying to understand how psycho-sociological factors influence on optimism/pessimism at individual level. To understand psychological factors we have analyzed per-

sonality of each user using the *Big5 Personality Model* [Goldberg1993]. On the other hand, to understand sociological factors we have used values & ethics (*Schwartz Values Model*) [Schwartz1992] model. Therefore, both the influencing factors have been aggregated and analyzed at community level (a community in a social network is considered to be a group of nodes densely connected internally and sparsely connected externally). The presumption here was that two persons with same personality trait but under distinct sociological influences at community level may react differently and similarly people with different personalities but in different sociological circumstances may react exactly in the same way. However, [Ruan et al.2016] had also discussed that personality is correlated with optimism [Sharpe et al.2011]. Pessimism is principally associated with neuroticism and negative affect while optimism is primarily associated with extraversion and positive affect [Marshall et al.1992]. We draw motivations from all these previous works and to this end this paper reports a comprehensive empirical study to understand how user level optimism/pessimism got influenced by person level psychology i.e. personality and by societal level values & ethics at community level.

We report interesting results and correlations related to influence of psycho-sociological factors on optimists/pessimists on Twitter. A few societal factors were found to to be more influencing than other societal factors. For example, achievement and conformity classes tend to be more influencing than other classes of values in promoting optimistic views in social media. Also a few social factors (or community level factors) like achieve-

ment and stimulation were found to be more positively correlated than user level factors in optimists. On the other hand, for pessimists power-oriented and traditional settings seem to be more correlated than other factors.

The paper is structured as follows. In Section 2, we discuss the relevant research work done in the area of computational models for optimism/pessimism as well as psycho-sociological factors. The data sources used for building classifiers, fuzzy distributions of personality and values in the corpus and performance of classifiers used to infer the distribution of different psycho-sociological aspects are discussed in Section 2. Section 3 describes the methodology used for extracting the semantic representation of a community in terms of it's values. Section 4 explains the use of semantic interpretation of communities in terms of their values and psycho-sociological patterns of users for the calculation of aggregate degree of optimism/pessimism under the influence of different factors. In Section 5, important correlations related to influence of different factors on optimists/pessimists are discussed in detail.

2   Related Work

There has been a little research in the field of computational models to predict the degree of optimism. [Ruan et al.2016] developed the first computational model for measuring the optimism/pessimism of users based on their social media activity. 714 potential optimists and 614 potential pessimists were identified by searching for specific phrases. After identification of potential optimists and pessimists, they created a ground truth dataset through human annotation on a randomly selected subset of corpus. In order to rank the users according to degree of optimism and pessimism, Twitter users were sorted based on the average scores assigned to their tweets. The top 25% were labelled as optimists whereas the bottom 25% were labelled as pessimists.

Deeper understanding of human personality, beliefs, ethics, and values has been a key research agenda in Psychology and Social Science research for several decades. One of the most accepted and widely used frameworks for understanding values is Schwartz Theory of Basic Human Values [Schwartz1992]. Schwartz's 10-Values model postulated and empirically verified (on data ob-

tained from a self-assessment questionnaire) the existence of ten basic Values based on people's motivation. The Schwartz model was proved to be very successful in psychological research as well as in other fields. The ten basic Values are related to various outcomes and effects of a person's role in a society [Argandoña2003]. The Values have also proved to provide an important and powerful explanation of consumer behaviour and how they influence it [Kahle et al.1986].

In the recent years, there have been few initiatives to automatically identify various Big5 Personality traits of individuals from their language usage and behaviour in social media [Goldberg1993]. A milestone in this area was the 2013 Workshop and Shared Task on Computational Personality Recognition[1], repeated in 2014[2]. Further research work in the area of developing computational model for identifying Personality from language usage on Facebook and Twitter has been done in [Park et al.2015] and [Quercia et al.2011] respectively. However, no computational model for *Schwartz' Values* has been tested or examined before.

3   Obtaining   Psycho-Sociological Patterns of Users

The psycho-sociological backgrounds of an individual play a very crucial role in determining their inclination of being an optimist or pessimist. In our current work, we analyze two psycho-sociolological aspects in correlation with optimism/pessimism: personality and values. We build two classifiers (by analyzing social media content: tweets and activities) to estimate someone's inclination towards optimism/pessimism:

Personality: [Goldberg1990] to determine which personality types {*openness, conscientiousness, extroversion, agreeableness, neuroticism*} are more inclined towards optimism/pessimism.

Values & Ethics: [Schwartz1992] to estimate the nature of values possessed by optimism/pessimism, in terms of 10 values {*achievement, benevolence, conformity, hedonism, security, self-direction, stimulation, tradition, universalism*} types proposed by Schwartz.

---

[1]http://mypersonality.org/wiki/doku.php?id=wcpr13
[2]https://sites.google.com/site/wcprst/home/wcpr14

The psycho-sociological patterns thus obtained, can be used as metrics to gauge someone's inclination towards optimism/pessimism. Such finding could be helpful for various practical purposes like online recommendation system, Twitter could improve its "who to follow" suggestions, online advertisement and etc by biasing the random walks procedure used in link prediction algorithms or adding regularization terms in matrix factorization methods.

3.1 Data Sources

We used four datasets for our analysis : (1) Optimism/Pessimism dataset for replication of computational model proposed by [Ruan et al.2016], (2) Personality dataset for developing the computational model for predicting five personality traits at user-level, (3) Values & Ethics dataset for developing a model for prediction of Values & Ethics at user-level and (4) SNAP Dataset for analyzing distribution of optimists/pessimists in influence of different psycho-sociological factors.

Optimism/Pessimism: For the purpose of developing computational models for Optimism/Pessimism, we used the dataset proposed by [Ruan et al.2016]. The Twitter conversations dataset was obtained by searching for key-phrases such as "*I am optimistic*" for identifying optimistic users and keywords such as "hate", "unfair", and "disgust" for identifying potential pessimists. They identified 718 potential optimists and 640 potential pessimists with maximum of 2,000 tweets per user. A small subset of 500 potential optimists and 500 potential pessimists were selected. To create a human annotated ground truth dataset, 15 tweets for each user were randomly selected. Further, each tweet was annotated as potimist or pessimist.

Personality: The personality labeled gold corpus (10K Facebook status updates of 250 users and their Facebook network properties), released in WCPR'13[3] workshop, was used to build the personality model. From the Table 1, we can observe that the Facebook Personality corpus used in WCPR'13 is balanced corpus with almost equal distribution of users across all five different personality traits.

Values & Ethics: The Values & Ethics data for 367 users was crowd-sourced along with users'

Table 1: Flat distribution of Big5 Personality types in the Facebook Personality corpus: Openness (O), Conscientiousness (C), Extraversion (E), Agreeableness (A), Neuroticism (N), The last column gives the Average Majority Baselines.

O	C	E	A	N	Avg
70.40	52.00	38.40	53.60	39.60	**50.80**

Twitter Ids using Amazon Mechanical Turk as a service, while ensuring that the participants came from various cultures and ethnic backgrounds: the participants were equally distributed across the globe – Americans (USA, Canada, Mexico, Brazil), South Asia (India, Pakistan, Bangladesh), and a few East-Asians (Singaporeans, Malaysian, Japanese, Chinese). The selected Asians were checked to be mostly English speaking.

We obtained data from self-assessment based psychometric tests using male/female versions of PVQ, the Portrait Values Questionnaire [Schwartz et al.2001]. The PVQ asks participants to answer each question on a 1–6 Likert rating scale[4]. A rating of 1 means "*not like me at all*" and 6 means "*very much like me*". An example question is "*He likes to take risks. He is always looking for adventures.*" where the user should answer while putting himself in the shoes of "He" in the question. The standard practice is to ask a fixed number of questions per psychological dimension. Therefore, there are five questions for each of the ten Values classes, resulting in a 50 item PVQ questionnaire. Once all the questions in the PVQ have been answered, for each user and for each Values class, a score is generated by averaging all the scores (i.e., user responses) corresponding to the questions in that class, as described by [Schwartz2012]. Further, the rescaling strategy proposed by [Schwartz2012] was used to eliminate randomness from each response given by a user as follows: for each user, the mean response score was first calculated considering all the responses s/he provided, and then the mean score from each response was subtracted. See [Schwartz2012] for more details on PVQ and the score computation mechanism.

The ranges of scores obtained from the previous rescaling method may vary across different Values classes. For instance, the ranges of the rescaled scores for the Twitter Values corpus are as

---

[3] http://mypersonality.org/wiki/doku.php?id=wcpr13

[4] http://www.simplypsychology.org/likert-scale.html

Table 2: Flat distribution of Schwartz' value types in the corpora: Achievement (AC), Benevolence (BE), Conformity (CO), Hedonism (H), Power (PO), Security (SE), Self-Direction (SD), Stimulation (ST), Tradition (TR), Universalism (UN). The last column gives the Average Majority Baselines.

AC	BE	CO	HE	PO	SE	SD	ST	TR	UN	Avg
81.00	78.70	73.30	77.10	50.10	76.30	83.40	73.60	52.60	82.00	**72.80**

follows: Achievement $[-4.12, 3.36]$, Benevolence $[-1.56, 3.39]$, Conformity $[-3.35, 3.01]$, Hedonism $[-5.18, 4.35]$, Power $[-6.0, 2.27]$, Security $[-2.60, 2.40]$, Self-Direction $[-1.61, 3.40]$, Stimulation $[-5.0, 2.63]$, Tradition $[-4.49, 3.35]$, and Universalism $[-3.33, 3.30]$.[5] Hence the following normalisation formula was applied to move the ranges of the different Values classes to the $[-1, 1]$ interval:

$$x_{scaled} = \frac{2 * (x - x_{min})}{x_{max} - x_{min}} - 1$$

Further, for obtaining text data to train computational model we collected tweets from users' Twitter accounts. However, several challenges have to be addressed when working with Twitter. For example, several users had protected Twitter accounts, so that their tweets were not accessible when using the Twitter API. In addition, many users had to be discarded since they had tweeted less than 100 tweets, making them uninteresting for statistical analysis. In addition, some extreme cases when users mentioned someone else's (some celebrity's) Twitter account, had to be discarded. Finally after filtering the data, we obtained a text dataset of 367 users consisting of at least 100 tweets and maximum of 3200 tweets. Categorical flat distributions of Values types are reported in the Table 2.

SNAP Dataset: In order to analyze the behaviour of optimists/ pessimists at societal level, the egocentric twitter network released by SNAP is used. For the purpose of investigating the sociological factors operating at societal level, 1,562 ground-truth communities spanning over 81306 nodes and $1, 768, 149$ edges were considered for further analysis (other communities having size less than 5 and with number of tweets less than 100 were discarded).

3.2 Corpus Statistics

Categorical flat distributions of Personality and Values types are reported in Table 1 and Table 2, respectively. It is noteworthy that the Facebook Personality corpus used in the Workshop on Computational Personality Recognition shared task [Celli et al.2013] is quite balanced because it was judiciously chosen from a larger 10K user data corpus collected in the myPersonality project.[6] The shared task organisers chose the right (desired) distribution for the released corpus, but in a real-life setting it is almost impossible to get a balanced data from any user population or social media platform. On the other hand, the Twitter corpus, collected by us is skewed.

Moreover, both the Personality model and Schwartz' Values model support fuzzy membership, which means that anyone having Open personality can have Agreeable nature as well, and similarly that someone with Power orientation also can have Achievement orientation. To understand this notion, the fuzzy membership statistics of Facebook Personality corpus and Twitter Values corpus are reported in Figures 1 and 2 respectively.

On a careful analysis of Figure 1 one can clearly observe how each Personality trait is overlapped with other Personality traits. For example, for Openness (O), we can clearly see that almost equal amount of people are positively oriented towards all the remaining four Personality dimensions. Similarly, Agreeable people are evenly distributed among three traits: Openness, Extroversion and Conscientiousness, but very few of them have in neurotic nature. On the other hand, the class distribution of Neuroticism (N) is highly imbalanced, as most of the people who are neurotic are always eager to experience new things always to satisfy their ever-changing mood. Further, it can be inferred from the visualisation that very few extroverts have neurotic nature, but many of them are positively oriented towards Conscientiousness (C) trait. It indeed makes sense as extrovert people are outgoing and would like to mingle in different circles of the society, and thus they are accommodative and less sentimental, i.e., less neurotic but they are rather methodical i.e., conscientious.

---

[5]The distribution of a particular value type over a corpus was analysed using the Bienaymé-Chebyshev Inequality [Bienaymé1853, Tchébichef1867], showing that, for example, most of the Achievement instances (89%) were in the range $[-2.96, 2.84]$.

258

[6]http://mypersonality.org

Personality Traits	O	C	E	A	N
Openness(O)		52.27	40.90	60.22	38.63
Conscientiousness(C)	70.76		46.92	57.69	28.46
Extraversion(E)	75.00	63.54		59.37	22.91
Agreeableness(A)	79.10	55.97	42.53		25.37
Neuroticism(N)	68.68	37.38	22.25	34.56	

Figure 1: Fuzzy distributions of Big5 Personality traits in the myPersonality corpus. Similar to Schwartz's fuzzy distribution this table attempts to present the interconnection between different Personality traits.

Schwartz Values	AC	BE	CO	HE	PO	SE	SD	ST	TR	UN
Achievement(AC)		28.31	19.49	29.41	41.54	15.81	11.77	19.85	41.91	17.28
Benevolence(BE)	24.12		19.84	31.12	52.92	18.68	10.51	22.18	42.8	7
Conformity(CO)	18.59	23.42		35.32	47.58	12.64	17.47	24.91	35.32	15.99
Hedonism(HE)	17.63	24.04	25.32		43.35	21.03	9.01	14.6	45.92	12.88
Power(PO)	12.64	33.52	22.53	27.47		17.58	13.74	17.03	41.21	20.33
Security(SE)	17.63	24.82	15.47	33.81	46.04		13.31	21.94	38.49	14.39
Self-Direction(SD)	21.05	24.34	26.97	30.26	48.35	20.72		20.72	47.04	12.5
Stimulation(ST)	18.66	25.37	24.63	25.75	43.66	19.03	10.08		42.91	16.04
Tradition(TR)	18.13	23.83	9.84	34.72	44.56	11.4	16.58	20.73		17.1
Universalism(UN)	24.75	20.07	24.41	32.11	51.51	20.4	11.04	24.75	46.49	

Figure 2: Fuzzy distributions of Schwartz' values in the Twitter corpus. Schwartz' theory explains how the values are interconnected and influence each other, since the pursuit of any of the values results in either an accordance with one another (e.g., Conformity and Security) or a conflict with at least one other value (e.g., Benevolence and Power).

It is also clear that the fuzziness is much higher among the Values classes than among the Personality traits. One possible reason is that Personality has fewer number of classes than Schwartz Values. Such overlapping nature of psychological classes makes the computational classification problem much more challenging than the classical sentiment analysis problem.

3.3 Psycholinguistic and Network Features

We explored exhaustive set of features including – (**f1**) Word N-grams; (**f2**) Lingustic Features (LIWC[7]; Harvard General Inquirer, MRC psycholinguistic feature; Sensicon[8]); (**f3**) Speech-Act classes; (**f4**) Sentiment Amplifiers (Exclamation Marks, Quotes, Ellipses, Interjections, Emoticons, Word/Sentence Length); (**f5**) Sentiment/ Emotion lexica (NRC emotion Lexicon [Mohammad et al.2013], Sentiwordnet [Esuli and Sebastiani2007]); (**f6**) Topics words obtained from topic model. A brief overview about personality, values,

Table 3: Features used in psycho-sociological models

Models	f1	f2	f3	f4	f5	f6	F-Score
Personality	+	+	+	-	-	-	79.35%
Values	-	+	+	-	-	+	80.10%
Optimism/Pessimism	+	+	-	-	+	-	77.89%

and optimism/pessimism models and features used are illustrated in 3.

3.4 Building Classifiers

We collected data from several sources to build three classification models. Here for each model we report the best classifier. The feature engineering required for improving performance of Personality and Values models is discussed in detail in [Tushar Maheshwari2016]. All the results reported in Table 3 are based on 10-fold cross validation on respective datasets.

Personality: The personality labeled gold corpus (10K Facebook status updates of 250 users and their Facebook network properties), released in WCPR'13[9] workshop [Celli et al.2013], is used

---

[7]http://www.liwc.net/
[8]https://hlt-nlp.fbk.eu/technologies/sensicon

259

[9]http://mypersonality.org/wiki/doku.php?id=wcpr13

to build the personality model. Our SVM-based model outperforms the state-of-the-art [Verhoeven et al.2013] by 10%, achieving average F-Score of 79.35%. Features used in this model are reported in Table 3.

Values & Ethics: For the values model we crowd-sourced a Twitter corpus using the Amazon Mechanical Turk[10]. Self-assessments were obtained using the Portrait Values Questionnaire (PVQ) [Schwartz et al.2001]. At the end of the data collection process, data from 367 unique users had been gathered, having 1,608 average tweets per user (see supp. for details). The SVM-based values classifiers achieves an average F-Score of 80% using features reported in Table 3.

Optimism/Pessimism: A Multinomial Naive Bayes classifier is trained on the dataset introduced by [Ruan et al.2016], the state-of-the-art for this work. We obtain an F-Score of 77.89% using features reported in Table 3), which is comparable to the state-of-the-art i.e. 81% [Ruan et al.2016].

4 Semantic Interpretation of Communities

In order to analyze the behaviour of optimists/pessimists at societal level, the egocentric twitter network released by SNAP is used. The Twitter network, released by SNAP [Leskovec and Krevl2015] (nodes: 81,306, edges: 1,768,149) has been used to study community structure. We considered 1,562 ground-truth communities (after discarding communities having size less than 5 and with tweets less than 100).

In order to analyse whether people within the same community tend to be homogeneous with respect to their background values/ethics, we measure Shannon's Entropy (measure of the uncertainty) [Lin1991] for each dimension separately.

Higher entropy scores suggest lower similarity. To calculate the entropy score vector $X_{(i)}$ for a community $C_{(i)}$ consisting of n users as $u_{(1)}, u_{(2)}, u_{(3)}...u_{(n)}$, a matrix $A_{(i)}$ is created where $A_{(i,j)}$ row vector represents the estimated scores of each of the ten values for a user $u_{(j)}$ and $A_{(i,:,k)}$ column vector represents the estimated scores of kth class for all n users. The $A_{(i,:,k)}$ column vector was transformed to a probability distribution vector $N_{(i,:,k)}$ using softmax

normalization:

$$N_{(i,j,k)} = \frac{exp(A_{(i,j,k)})}{||exp(A_{(i,:,k)})||_1}$$

The entropy score $X_{(i,k)}$ for $N_{(i,:,k)}$ can be calculated using the following formulation:

$$X_{(i,k)} = -\sum_{j=1}^{n} N_{(i,j,k)} * \log N_{(i,j,k)}$$

Table 4: Illustrates entropy calculation for values model. Here $T_{(i)}$ represents the binary estimate of fuzzy distribution of values and $S_{(i)}$ represents the zero-mean unit-variance scaled values of $X_{(i)}$ for a community $C_{(i)}$. Similarly, binary estimates for five personality traits $P_{(i)}$ of user $u_{(i)}$ are calculated.

	AC	BE	CO	HE	PO	SE	SD	ST	TR	UN
$u_{(1)}$	0.91	0.47	0.02	0.07	0.32	0.24	0.65	0.78	0.94	0.10
$u_{(2)}$	0.97	0.40	0.49	0.50	0.56	0.83	0.62	0.73	0.04	0.08
$u_{(3)}$	0.99	0.75	0.50	0.72	0.38	0.60	0.75	0.02	0.57	0.62
$u_{(4)}$	0.77	0.44	0.40	0.16	0.19	0.55	0.73	0.08	0.53	0.25
$u_{(5)}$	0.29	0.02	0.26	0.56	0.41	0.23	0.95	0.02	0.79	0.86
$X_{(i)}$	1.54	1.40	1.40	1.39	1.55	1.50	1.59	0.99	1.42	1.28
$S_{(i)}$	0.87	-0.12	-0.12	-0.19	0.95	0.57	1.26	-2.35	0.00	-0.87
$T_{(i)}$	1.0	0.0	0.0	0.0	1.0	1.0	1.0	0.0	0.0	0.0

After normalization, $N_{(i,:,k)}$ vector represents the probability distribution of kth Value class across n users where entropy score $X_{(i,k)}$ represents the randomness in community along k^{th} Value class. The lower the randomness, higher the kth class is dominant in the $C_{(i)}$ community. Now, in order to obtain binary estimates $T_{(i)}$ for each of the ten values and classes in $C_{(i)}$ community, the entropy score vector $X_{(i)}$ is scaled using zero-mean unit-variance method and for numerical values greater than 0, 1 was assigned and for numerical values less than 0, 0 was assigned as class label for $C_{(i)}$ community. Instead of labelling a community $C_{(i)}$ with a class having minimum entropy, the scaling approach is used for the purpose of preserving the fuzzy distribution of values at community level. The obtained $T_{(i)}$ vector represents the fuzzy distribution of values and is thus a representation to capture the semantic information about the community.
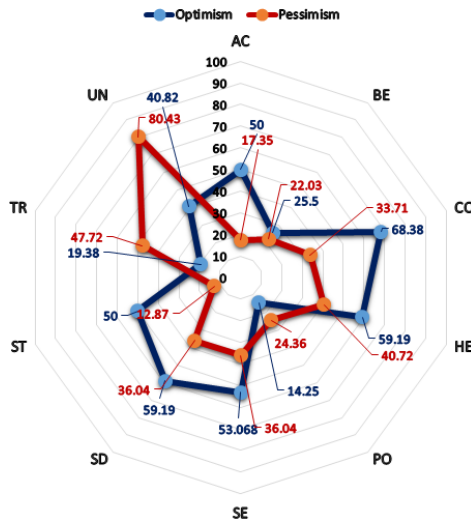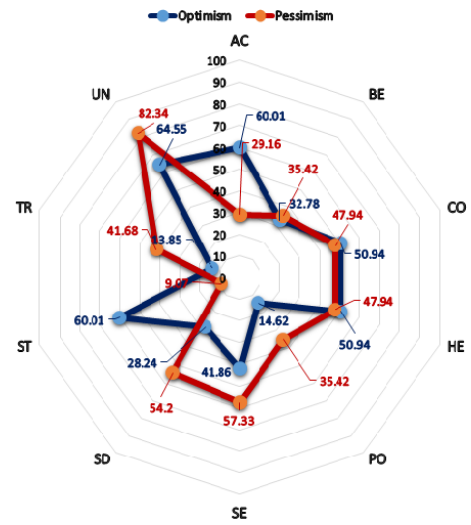
Figure 3: Openness
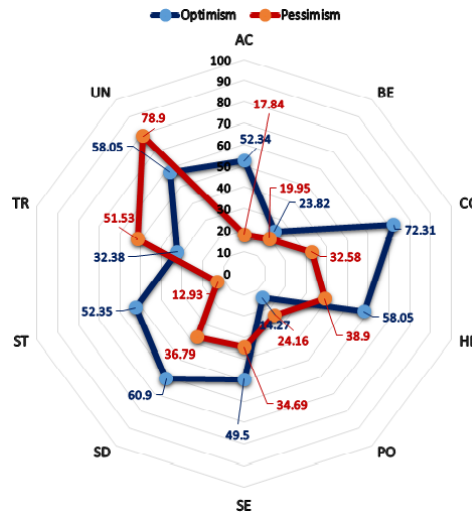

Figure 4: Conscientiousness
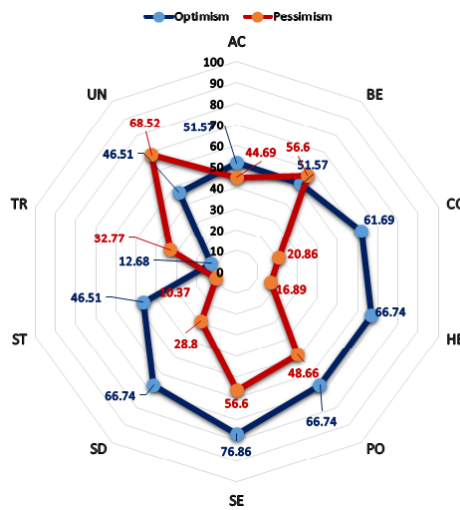

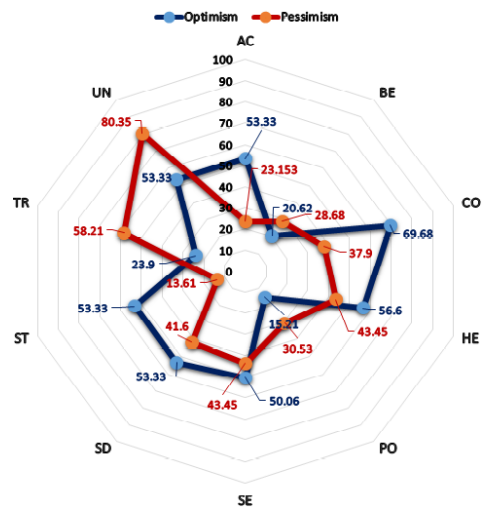Figure 5: Extraversion

Figure 6: Agreeableness

Figure 7: Neuroticism

Figure 8: Psycho-Sociological Influences on Optimism/Pessimism over Twitter. For i^{th} personality trait, the radar chart visualizes $I_{(i,:,0)}$ and $I_{(i,:,1)}$ represents degree of optimism and pessimism respectively over 10 classes of Values & Ethics.

5 Understanding Psycho-Sociological Influences of Optimism/Pessimism

The analysis for user-level estimation starts with each user $u_{(i)}$ and it's the binary estimates of all five traits of the Personality ($P_{(i)}$), probability estimates of ten classes of Values ($V_{(i)}$) and probability estimates of degree of optimism/pessimism ($O_{(i)}$), obtained using the pre-trained Big5 Personality model, Schwartz Values model and optimism/pessimism models respectively. The user-level estimation for the degree of optimism/pessimism was done by averaging the optimism/pessimism scores for each user. Further the estimated scores for five personality traits, ten values traits and degree of optimism/pessimism was scaled using zero-mean unit-variance method for each of the 16 classes independently.

To this extent, we have obtained $P_{(j)}$, $V_{(j)}$ & $O_{(j)}$ representing the personality traits, values and degree of optimism respectively for each user $u_{(j)}$ and binary estimates $T_{(i)}$ representing the fuzzy distribution of values for each community $C_{(i)}$.

Aggregate Analysis over Personality Traits: For the purpose of understanding the distribution of $O_{(j)}$ under the influence of $P_{(j)}$ and $T_{(i)}$ for a user $u_{(j)}$ in community $C_{(i)}$, we divide our analysis in five parts according to five different traits of personality. The division of analysis according to five personality traits will help us better understand the influence of each of these personality traits on optimists/pessimists. Therefore, for all users $u_{(j)}$ having a particular personality trait, we go into the communities the user $u_{(j)}$ belongs to and lookup the values of that community using $T_{(i)}$ as well as obtain the degree of optimism/pessimism for that user from $O_{(j)}$. In this way, we *aggregate the degree of optimism for each pair of personality trait and values class* and obtain the distribution of degree of optimism/pessimism under the influence of socio-psychological factors. The aggregation of degree of optimism/pessimism can be formulated as follows :

$$I_{(i,j,0)} = \sum_{u_{(k)}|(P_{(k,i)}=1)\cap(T_{(k,j)}=1)} O_{(k,j,0)}$$

where $I_{(i,j,0)}$ represents the aggregate degree of optimism for i^{th} personality trait and j^{th} value

class. Similarly, $I_{(i,j,1)}$ representing the degree of pessimism is calculated by aggregating over $O_{(k,j,1)}$.

6 Obtained psycho-sociological patterns

On careful investigation of five radar plots (Figure 8) for each of the personality traits we discover interesting patterns on how combination of personality trait and values influence the behaviour of optimist/pessimist users on Twitter.

Influence of Psycho-Sociological Factors on Optimists: In Figure 3, we can infer that people with open personality are generally optimistic irrespective of the dominating values of their community as individuals with high openness tend to seek euphoric experiences resulting in positive expectations. From Figure 4, it is evident that individuals high in conscientiousness are positively correlated with very optimistic people in both achievement as well as stimulation oriented communities. Since, people high in conscientiousness tend to have obsession, they are expected to be optimistic when they are surrounded by people aspiring to achieve and face challenges. From Figure 5, it can be observed that extroversion has positive correlation with very optimistic people in communities highly oriented towards achievement, conformity, hedonism, security, self-direction and stimulation owing to their high energy and positive emotions. Further from Figure 6, we can observe that agreeable people have very distinct radar profile portraying them as very optimistic people irrespective of their social settings because of their compassionate and cooperative nature.

Overall, from the radar charts, we propose the order of dominance of societal factors (values & ethics) over individual factors (personality) by observing that in certain social factors dominate over all user-level factors : AC > CO > ST. For example, achievement and stimulation oriented settings encourage more optimistic views on social media irrespective of different personality traits.

Influence of Psycho-Sociological Factors on Pessimists: Open people in traditional or power oriented settings seem to be more pessimistic owing to many restrictions posed in a traditional community and quest for prestige in a power oriented community. Similarly, for all other personality traits, the traditional settings influence more

than an individual's personality traits in promoting pessimistic views. Except for agreeable people, we can observe that power-oriented community also increases the sense of pessimism among people. In addition, for conscientious users belonging to security or self-direction oriented communities are very pessimistic. The pessimistic nature of the conscientious users can be explained by the fact that their sense of responsibility coupled with pressure for maintaining stability, security and making independent choices in life may lead to lowering their positive expectations.

On the other hand, a few outliers in the analysis can be seen due to accumulation of inaccuracies in the trained models of Values & Ethics as a result of biased training data which is clear from flat distribution of values in Table 2. For example, on an average, optimism in users belonging to communities high in universalism as well as pessimism in users belonging to communities high in conformity is relatively high irrespective of their personality traits.

7 Conclusion and future direction

This paper presents an empirical study to understand *how psycho-sociological factors influence on optimism/pessimism at individual level.*

However, we have only analyzed intra-community psycho-sociological patterns in this study, but we strongly believe that neighbouring communities also have influential roles to play on person level optimism/pessimism. In addition, we need to study the influence of communities which are not power-oriented or not achievement oriented along with different personality traits. We are working on analyzing and inferring other influencing factors using computational models.

References

[Argandoña2003] Antonio Argandoña. 2003. Fostering values in organizations. *Journal of Business Ethics*, 45(1–2):15–28.

[Bienaymé1853] Irénée-Jules Bienaymé. 1853. *Considérations à l'appui de la découverte de Laplace sur la loi de probabilité dans la méthode des moindres carrés.* Imprimerie de Mallet-Bachelier.

[Celli et al.2013] Fabio Celli, Fabio Pianesi, David Stillwell, and Michal Kosinski. 2013. The workshop on computational personality recognition 2013. In *Proceedings of the AAAI*, pages 2–5. AAAI.

[Esuli and Sebastiani2007] Andrea Esuli and Fabrizio Sebastiani. 2007. Sentiwordnet: A high-coverage lexical resource for opinion mining. *Evaluation*, pages 1–26.

[Goldberg1990] Lewis R Goldberg. 1990. An alternative" description of personality": the big-five factor structure. *Journal of personality and social psychology*, 59(6):1216.

[Goldberg1993] Lewis R Goldberg. 1993. The structure of phenotypic personality traits. *American psychologist*, 48(1):26.

[Kahle et al.1986] Lynn R Kahle, Sharon E Beatty, and Pamela Homer. 1986. Alternative measurement approaches to consumer values: The list of values (lov) and values and life style (vals). *Journal of consumer research*, pages 405–409.

[Leskovec and Krevl2015] Jure Leskovec and Andrej Krevl. 2015. {SNAP Datasets}:{Stanford} large network dataset collection.

[Lin1991] Jianhua Lin. 1991. Divergence measures based on the shannon entropy. *IEEE Transactions on Information theory*, 37(1):145–151.

[Marshall et al.1992] Grant N Marshall, Camille B Wortman, Jeffrey W Kusulas, Linda K Hervig, and Ross R Vickers Jr. 1992. Distinguishing optimism from pessimism: Relations to fundamental dimensions of mood and personality. *Journal of personality and social psychology*, 62(6):1067.

[Mohammad et al.2013] Saif M Mohammad, Svetlana Kiritchenko, and Xiaodan Zhu. 2013. Nrc-canada: Building the state-of-the-art in sentiment analysis of tweets. *arXiv preprint arXiv:1308.6242.*

[Park et al.2015] Gregory Park, H Andrew Schwartz, Johannes C Eichstaedt, Margaret L Kern, Michal Kosinski, David J Stillwell, Lyle H Ungar, and Martin EP Seligman. 2015.

Automatic personality assessment through social media language. *Journal of personality and social psychology*, 108(6):934.

[Quercia et al.2011] Daniele Quercia, Michal Kosinski, David Stillwell, and Jon Crowcroft. 2011. Our twitter profiles, our selves: Predicting personality with twitter. In *Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third Inernational Conference on Social Computing (SocialCom), 2011 IEEE Third International Conference on*, pages 180–185. IEEE.

[Ruan et al.2016] Xianzhi Ruan, Steven R Wilson, and Rada Mihalcea. 2016. Finding optimists and pessimists on twitter. In *The 54th Annual Meeting of the Association for Computational Linguistics*, page 320.

[Schwartz et al.2001] Shalom H Schwartz, Gila Melech, Arielle Lehmann, Steven Burgess, Mari Harris, and Vicki Owens. 2001. Extending the cross-cultural validity of the theory of basic human values with a different method of measurement. *Journal of cross-cultural psychology*, 32(5):519–542.

[Schwartz1992] Shalom H. Schwartz. 1992. Universals in the content and structure of values: theoretical advances and empirical tests in 20 countries. In *Advances in Experimental Social Psychology*, pages 1–65. In M. Zanna (Ed.), Advances in experimental social psychology (Vol. 25), New York: Academic Press.

[Schwartz2012] Shalom H Schwartz. 2012. An overview of the schwartz theory of basic values. *Online Readings in Psychology and Culture*, 2(1):11.

[Sharpe et al.2011] J Patrick Sharpe, Nicholas R Martin, and Kelly A Roth. 2011. Optimism and the big five factors of personality: Beyond neuroticism and extraversion. *Personality and Individual Differences*, 51(8):946–951.

[Tchébichef1867] Pafnuty Lvovich Tchébichef. 1867. Des valeurs moyennes (translated into French by N.V. Khanykov). *Journal de Mathématiques Pures et Appliquées*, 12(2):177–184.

[Tushar Maheshwari2016] Upendra Kumar Amitava Das Tushar Maheshwari, Aishwarya Reganti. 2016. Semantic interpretation of community in social networks. *AAAI*.

[Verhoeven et al.2013] Ben Verhoeven, Walter Daelemans, and Tom De Smedt. 2013. Ensemble methods for personality recognition. *Proceedings of WCPR13, in conjunction with ICWSM-13*.