PACMAN: Psycho and Computational Framework of an Individual (Man)

Shivani Poddar*, Sindhu Kiranmai Ernala*, Navjyoti Singh
International Institute of Information Technology - Hyderabad
Hyderabad, India - 500032
shivani.poddar@research.iiit.ac.in, sindhukiranmai.ernala@research.iiit.ac.in, navjyoti@iiit.ac.in

Abstract

Several models have tried to understand the formation of an individual's distinctive character i.e. personality from the perspectives of multiple disciplines, including cognitive science, affective neuroscience and psychology. While these models (for eg. Big Five) have so far attempted to summarize the personality of an individual as a uniform, static image, no one model comprehensively captures the mechanisms which leads to the formation and evolution of personality traits over time. This mechanism of evolving personality is what we attempt to capture by means of our framework. Through this study, we leverage the Abhidhamma tradition of Buddhism to propose a theoretical model of an individual as a stochastic finite state machine. The machine models moment to moment states of consciousness of an individual in terms of a formal ontology of mental factors that constitute any individual. To achieve an empirical evaluation of our framework, we use social media data to model a user's personality as an evolution of his/her mental states (by conducting some psycho-linguistic inferences of their Facebook (FB) statuses). We further analyze the user's personality as a composition of these recurrent mental factors over a series of subsequent moments. As the first attempt to solve the problem of evolving personality explicitly, we also present a new dataset and machine learning module for analysis of mental states of a user from his/her social media data.

Keywords: Personality Modeling, Social Media Analysis, Lexical Analysis

1. Introduction

Understanding emotion and personality profiles are a key to unlocking elusive human qualities. These qualities provide valuable insights into the interests, experiences, behaviorism and opinions of the respective individuals. Personality helps in fingerprinting an individual, which in turn is useful in decoding the human behavior, mental processes and affective reactions of people over time towards various external stimuli. Contextual systems used in a multitude of domains for instance e-commerce, advertisements, e-learning etc. could greatly benefit from such user insights (Moscoso and Salgado, 2004). While there have been a range of personality models which dominated the landscape of inferring user personality from social media platforms, Big Five model has been established as the most popular. The model was proposed by Goldberg et al (Goldberg, 1990), and studies the behavior of an individual over time to uniquely identify their Big Five Trait Dimensions: Openness, Neuroticism, Extraversion, Agreeableness and Conscientiousness. Various studies of social media have attempted to capture these traits extensively from websites such as Twitter (Golbeck et al., 2011), Facebook (Ross et al., 2009), Blog data (Poddar et al.,) etc. A recurrent underlying theme that all the research in the domain has in common is that of a constant user personality. The suggested personality of a user mined by means of most of the state of the art techniques focus on extracting the overall personality of a person. For instance, (Golbeck et al., 2011) Goldbeck et al. classify the subjects into one of the Big Five categories by means of extensive feature extraction from Facebook (Page likes, comments etc.). Although initial literature in psychology did suggest that personality remained constant after the age 30, many recent studies contradict this notion (Costa Jr and McCrae, 1980).

By means of our research we attempt to model the personality of an individual as the combination of a set of mental factors (described in Section 2.1) which have been dominant in the individual for a significant amount of time. The personality here, unlike state of the art models, is not static or of a specific type, but keeps evolving with the individual himself. For instance, if a child demonstrated acute "Selfishness" in the early years, but grew out of it eventually, their personality would manifest selfishness in the respective time span (namely childhood), and eventually evolve to get rid of deprecated traits. In essence our model attempts to capture a personality trait (or a set of mental states) from the grassroot level (i.e. the beginning moments when they start to manifest in a person) to the time when it matures and defines a person. (i.e. a series of subsequent moments when it starts recurring without fail)

Our study, thus attempts to establish coherence with the psychological theories of variability of the Big 5 across various age groups (starting from 18 towards 65). It accentuates the importance of facet-level research for understanding life span age differences in personality (Soto et al., 2011). Another study which encapsulates the importance of capturing temperamental changes in adolescence which later on can be connected to adult behavior is undertaken by McCrae (McCrae et al., 2002) , Specht (Specht et al., 2011) etc. The work undertaken to achieve this requires us to probe an individual at the atomic level of perception, awareness, cognizance and action. This also enables us to closely observe and draw relevant inferences of various other aspects which are constructive units of the personality for instance social emotions such as remorse, pride and so on. Thus, the contributions of this work and the PACMAN framework can be summarized as follows: **C1:** Formalized Ontology of mental states (adapted from Abhidhamma) and

* have contributed equally to the paper

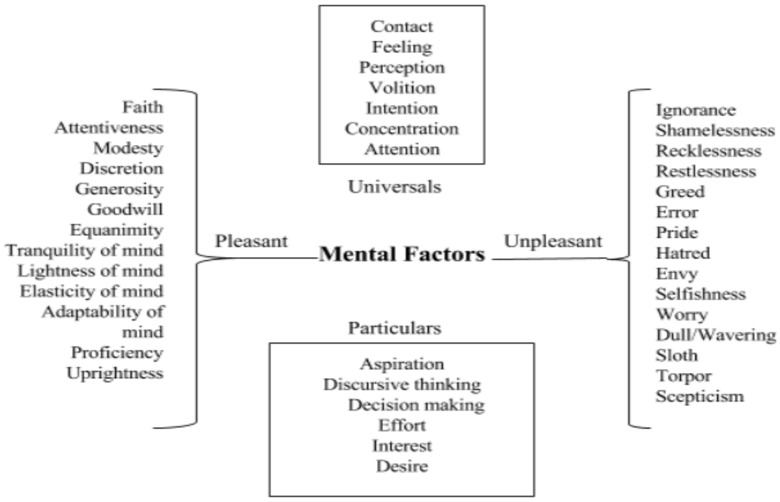

Figure 1: Represents a conceptualization of the mental states that populate the model of an individual.

a Stochastic Model of an individual to capture the evolving user personality from these moment-by-moment mental states. **C2:** A Multi-Label machine learning module to enable training of individual user social-media statuses with the respective mental factors. **C3:** Labeled dataset of 4,179 Facebook statuses (of the mypersonality dataset(Celli et al., 2013)) with the respective annotated mental factors. **C4:** An in-depth analysis of the mental factors and evolving personalities of 50 users from the above mentioned dataset (Celli et al., 2013) and the comparisons of these with the Big Five traits of the same users.

Organization of the paper: *Section 2* discusses the theoretical constructs used to formulate the evolving model of the personality of an individual. Here, Section 2.1 discusses the Ontology and description of the mental states and the moments in an individual's life where they are embedded (as inspired from the Abhidhamma discourse of Buddhism). Section 2.2 discusses the Stochastic Finite State Machine which helps us to populate these mental states at the given moments of time. *Section 3* is descriptive of the dataset used and the methodology (i.e pre-processing techniques, features, and machine learning algorithm) used to train the PACMAN Computational Learning Model to so as to predict the respective mental states from a user status. *Section 4* briefly states the results achieved by means of this model and *Section 5* Discusses these results and their implications. Finally by means of *Section 6* we present our conclusions and future.

2. PACMAN: Formal Model of an Individual

This section briefly discusses the inspiration of our adapted ontology of the mental states of an individual. It also presents a brief outline of our stochastic finite state machine representation of an individual.

2.1. Ontology of Cognitive Procedures

Abhidhamma scholarship in Buddhism (Mon, 1995) has long deliberated on the mechanisms of reality. In the Abhidhamma both mind and matter, which constitute the complex machinery of man, are microscopically analysed. The analysis provides descriptions of sentient experience as a succession of physical and mental processes that arise and cease subject to various causes and conditions. These sequential processes (mental and physical) formulated as discrete, momentary events are referred to as tropes (defined as dhammas in the original text) (Lancaster, 1997). Tropes are thus seen as psycho-physical events that provide mental cognitive awareness. The doctrine also presents the concept of a moment (khana) which is a kind of synchronic duration of each conscious event. In this sense, Abhidhamma visualizes the time scale of these mental/physical processes so they can be seen as operating from moment to moment. The Abhidhamma thus attempts to provide an exhaustive account of every possible type of experience, every type of occurrence that may possibly present itself in one's consciousness in terms of its constituent tropes.(Cox, 2004)

Further, the doctrine provides a taxonomy of tropes and their relational schema whereby each acknowledged experience, phenomenon, or occurrence can be determined and identified by particular definition and function. There are two kinds of tropes that constitute reality according to this doctrine - ultimate tropes (paramattha dhamma) and conventional tropes (samutti dhamma). Conventional tropes are complexes constituted by ultimate tropes and include social and psychological reality. Ultimate tropes are organized into a fourfold categorization. The first three categories include 1) the bare phenomenon of consciousness (citta) that encompasses a single trope type and of which the essential characteristic is the cognizing of an object; 2) associated mental factors (cetasika) that encompasses fifty-two trope types; and 3) materiality or physical phenomena (rupa) that include twenty-eight trope types that make up all physical occurrences . The fourth category that neither arises nor ceases through causal interaction is nibbana.

For our conception of modeling an individual based on Abhidhamma, we build a discrete line of moments, wherein each moment stands for a consciousness trope or citta. An individual is then conceived as a formal arrangement of these conscious tropes on a discrete line. This line of moments compulsively passes to the next moment as a result of previous cognition and action. Each moment has 2 categories of tropes embedded in it. 1) mental factors related to the cognition and 2) material cognition and actions. This in a nutshell is a basic mechanism of individual for which in the next section we write a stochastic finite state machine (LaViers and Egerstedt, 2011) (Nomura, 1996) which takes the line from one moment state to the next moment state.

The mental factors embedded in the subsequent moments of an individual have a defined ontology as suggested by the Buddhist literature on personality. They are primarily divided into 3 main classes: Pleasant, Unpleasant and Neutral (Universals and Particulars) as illustrated in Figure 1. There are various other models of psychology also which leverage from these traditional theories of Buddhism. For instance Buddhist Personality Model (BPM) (Grabovac et al., 2011)

2.2. Stochastic Finite State Automaton for an Individual

In this section we formally define a stochastic automaton of an individual based on the conception of a formal model of individual as described in Section 2.1. A central concept to this doctrine is that, there is a total ordered temporal sequence of moments that captures the consciousness of an individual. We model this sequence of moments as states of a finite state automaton. Each state is a temporal moment defined in terms of the mental factors and actions embedded in it. This embedding of a particular set of mental factors and actions in each moment is defined through transition functions of the automata. Upon this basic architecture, to populate each moment as a bag of word representation from individual's web data, we write stochastic processes to help in modeling, predicting and refining rules governing the persona of the individual.

Formally speaking we define our automaton as a finite state machine. Let Q = $\{Q_1, Q_2, Q_3 \dots\}$ be a set of symbols that represent moment states, A = $\{A_1, A_2, A_3 \dots\}$ be a set of symbols that represent actions and material cognition, and T = $\{T_1, T_2, T_3 \dots\}$ be a set of symbols that represent the mental concomitants of an individual. We define our stochastic automaton whose internal state space is Q and whose input and output spaces as a Cartesian product AxT.

$$I(r,f) = \{Q, A, T, r, f, \pi(f,r,.), M(f,.), AT(r,..), E\}$$

$$r \in [0,1])^D, f \in [0,1]$$

$$AT : [0,1]^D \times Q \times A \times T \to [0,1]$$

$$AT(r_i, Q_i, A_j \times T_j) :$$

Probability that the output is $A_j \times T_j$ *when the internal state is* Q_i.

It is important to note here that which Q (a moment state) is an embedding of A (action and material cognition) and T (mental concomitants of the social machine), it's structure varies by means of it's temporality and the personality/persona (f, r) of an individual.

$$M : [0,1] \times (A \times T) \times (A \times T) \times Q \to [0,1]$$

$$M(f, A_j \times T_j, A_k \times T_k, Q_l) :$$

Probability that the next moment state is Q_l *when the input is* $A_j \times T_j$*and the output is*$A_k \times T_k$

$E(\in Q)$: *Halting state*

*i.e. when the moment state moves on to *empty state**

$$\pi(f, r, Q_i) :$$

*Probability that the initial state (after *empty state*) is* Q_i

$$\sum_{j=1}^{n} AT(r, Q_i, A_j \times T_j) = 1$$

$$\sum_{l=1}^{m} M(f, A_j \times T_j, A_k \times T_k, Q_l) = 1$$

$$\sum_{l=1}^{m} \pi(f, r, Q_l) = 1$$

Here, f represents the personality parameter and r represents the attitude of the given individual towards an object for output.

Let m(t) ∈ Q be a moment state at any discrete time 't', at_out(t) be any output set of A & T at time 't' and at_in(t) be any input set of A & T at 't'. Then the relation m(t), at_out(t) and at_in(t) share is as follows:

$$Prob(em(0) = Q_i) = \pi(f, r, Q_l) \quad (1)$$
$$Prob(em(t+1) = Q_i) = M(f, at\_in(t), at\_out(t), Qi)$$
$$Prob(ac\_out(t) = A_j \times T_j) = AT(r, em(t), A_j \times T_j)$$

Let $TRM_k(f,r) \in Mat_m(R)$ be the state transition probability matrix in the case the input is $A_k \times T_k$. From (1), we can get $TRM_k(f,r)$ as follows:

$$TRM_k(f,r) = (trm_k(f,r)_{ij}) \in Mat_m(\mathbb{R})$$

$$trm_k(f,r)_{ij} = Prob(E_i \to E_j | input = A_k \times T_k) \quad (2)$$
$$= \sum_{l=1}^{m} AT(r, f, Q_i, A_j \times T_j).$$
$$(f, A_k \times T_k, A_j \times T_j, Q_l)$$

3. PACMAN: Computational Learning Model

By means of this section we aim to present the machine learning module by means of which we will be able to transcend the above defined theoretical constructs (of an evolving personality) into a usable model for personality observation (and eventually, prediction). We empirically verify our model on the dataset described in section 3.1. The following section 3.2 illustrates the methodology used to train a multi-label classifier to predict the 40 mental states (Figure 1) populating moment-by-moment data of an individual (here, consecutive user statuses on social media). Our aim is to use these mental states as descriptors of the change in user personality over time.

3.1. Dataset Used

myPersonality (Celli et al., 2013) is a sample of personality scores and Facebook profile data that has been used in recent years for several different researches (Bachrach et al., 2012). It has been collected by David Stillwell and Michal Kosinski by means of a Facebook application that implements the Big5 test (Costa Jr and McCrae, 1995), among other psychological tests. The application obtained the consent from its users to record their data and use it for the research purposes. We randomly picked a set of 50 users from this dataset (who had more than 20 status updates) to analyse and validate PACMAN.

As the first attempt to solve the problem of evolving personality explicitly, we also contribute a labeled data-set named PACMAN dataset, [1] which can be used for further exploration in the field of evolving user personality. So as to achieve an extensive and unbiased set of annotations, we had a set of 3 independent annotators to tag each of the FB statuses of a random user in our dataset with a set of relevant mental factors (out of the 40 factors suggested in Figure 1). We then computed the MASI (Measuring Agreement in Set-Valued Items) to evaluate the disagreement amongst these annotations. Given two sets, A and B, the formula for MASI is:

$$1 - J_{A,B} \times M_{A,B}$$

where J is the Jaccard metric (Blackburn, 1980) for comparing two sets: a ratio of the cardinality of the intersection of two sets to their union. M (for monotonicity) is a four-point scale that takes on the value 1 when two sets are identical, 2/3 when one is a subset of the other, 1/3 when the intersection and both set differences are non-null, and 0 when the sets are disjoint. MASI ranges from zero to one. It approaches 0 as two sets have more members in common and are more nearly equal in size. An average value of 0.376 as suggested in Table 1 is, thus reflects that the sets of the labels under consideration are a close intersection of one-another.

G $\leftrightarrow$ A	**G $\leftrightarrow$ B**	**A $\leftrightarrow$ B**	**Avg. MASI**
0.306	0.386	0.435	0.376

Table 1: Inter-annotator values. G is the labeled gold data by annotator 1, A is the labeled data set from annotator 2 and B is the labeled data set from annotator 3.

3.2. Methodology

We used the following methodology to first pre-process the given data so as to filter out any noise. We then extracted the relevant features for our model and finally trained the multi-label classifier with the help of these features.

3.2.1. Pre-Processing

To preprocess the data available to us from the myPersonality dataset (Celli et al., 2013), we extract each individual based on the unique authentication ID provided in the dataset. This data is inclusive the statuses posted by the user, the dates of these posts and his/her Big Five traits. For our analysis we extract these FB statuses and the corresponding dates from original dataset and chronologically sort them. By means of language filtering, we then process this dataset to retain only those statuses that are using English language. So as to feed the statuses into the LIWC API, we were then required to also (for an improved analysis) determine the gender of the given user. We extracted Pronouns (such as “herself”, “himself”, “hers” etc.) from the Stanford POS tagger and mapped these pronouns to their respective gender usage as defined in English Language. This helped us to heuristically determine the grammatical gender of each user effectively. We mapped the gender of users with no gender specific pronoun usage to be 0 in the LIWC API.

3.2.2. Features Used

Feature extraction from short texts such as FB statuses, requires extensive linguistic analysis. So as to achieve an effective feature generation, we leverage the psycholinguistic tool, Linguistic Inquiry and Word Count (LIWC) (Pennebaker et al., 2001). It is adept in reflecting the various features relevant to the linguistic and psychological processes of a user in the context of social media (meaning shorter texts and more noise pertaining to faulty usage of English). The LIWC API includes a text analysis module along with a group of built-in dictionaries. The dictionaries are used to identify which words are associated with which psychologically-relevant categories. These categories include psychological features such as Analytical thinking, Emotional Tone, Social words and Informal speech as well as linguistic features such as Functional Words, Personal Pronouns and Punctuation. We use this API to extract the respective psycho-linguistic features for each FB status of a given user. To enhance the predicted 180 LIWC features by means of this API, we also specified (as additional parameters) the content-type as “Social Media” and the user-gender obtained via pre-processing.

3.2.3. Multi-Label Classification - Binary Relevance Method

Determining 40 mental factors from a linguistic unit, such as an FB status (here), can be cast as a multi-label classification problem. We propose using the Problem Transform approach to train our multi-label classifier. For training purposes, we transformed the extracted LIWC features for each status as a $M_{i,j}$ matrix, where i $\in$ (0,length of LIWC feature vector f) and j $\in$ (0,No of FB statuses of each user). We then appended this matrix with a set of 40 columns each that represented each of the mental factors mf by a value of 0 (for marking absence of mf) and 1 (for the presence of the mf).

[1] https://researchweb.iiit.ac.in/ shivani.poddar/PACMAN_Dataset

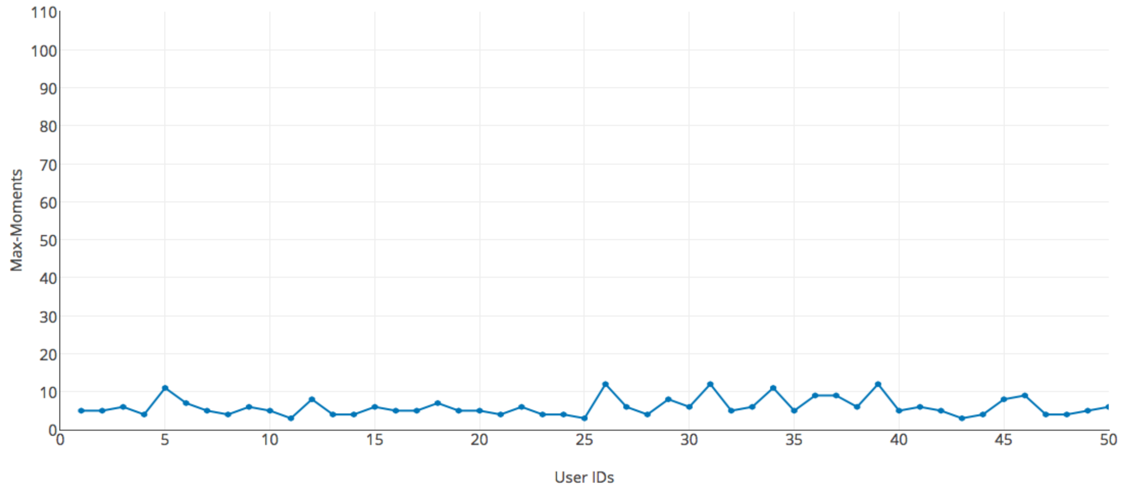

Figure 2: Maximum Moments reflect the number of statuses which have a consecutive set of similar mental factors.

f_1	f_2	..	f_{180}	mf_1	mf_2	..	mf_{40}
25.7	0.2	..	7	1	0	..	1

Table 2: Matrix $M_{i,j}$ for training and testing the Multi-Label Classifier Model.

Adapting to the Problem Transformation Method (Tsoumakas and Katakis, 2006), this problem is then approached as a joint set of binary classification tasks. These are expressed with label binary indicator array: each sample is one row of a 2d array of shape (nsamples, nclasses) with binary values: the one, i.e. the non zero elements, corresponding to the subset of labels. Using the binary relevance approach, we then use the One-vs-All SVM classifier to discriminate the data points of one class versus the others. Since our labels are not exclusive this works well for us since each classifier essential would answer the question : "Does it contain mental state x?" and so on for all all x belongs to (total 40 mental states). A brief representation of the $M_{i,j}$ matrix is as illustrated in Table 2.

4. Results

We analyse each individual and assert that any mental states which have a sustained cognition for more than a threshold x of the states is contributing to the personality of an individual. So as to arrive at the threshold x, we analyse the temporal mental states of n individuals and work out the intersection of states which imply sustained mental states in a given time span. This time span would be contributory to the defining personality of an individual. Threshold is the average of all the maximum moments of the sustained mental states (of the listed users). Here, empirically our threshold came out to be approximately 6.02 moments (elaborated in Section 5).

Since each instance in the multilabel data is not a single label but a vector of different label, established evaluation metrics such as accuracy, precision-recall, f-measure etc cannot be used directly (Gao and Zhou, 2013). Based on the learning problem we are addressing, Hamming loss: the fraction of the wrong labels to the total number of labels, i.e.

$$HammingLoss(x_i, y_i) = \frac{1}{|D|}\sum_{l=1}^{|D|}\frac{xor(x_i, y_i)}{|L|}$$

where $|D|$ is the number of samples, $|L|$ is the number of labels, y_i is the ground truth and x_i is the prediction. The average value of Hamming Loss is 10.455, which means that approximately every 10 out of 100 labels are predicted wrongly.

5. Discussions

Our results show that we can analyse and predict the evolving mental states contributing to the composition/change in the personality of an individual. We can also predict mental states to within just over 10%, a resolution that is likely fine-grained enough enough for many applications. A loss of 0.1 labels in a dataset which is being analysed moment by moment will not have many implications in various practical applications of our framework.

Since this research relied heavily on studying the mental states w.r.t Buddhist tradition of Abhidhamma, we define our heuristics for these analysis inspired by the same doctrine. Tapping into the dynamic nature of user persona would require us to study the persistent mental states which dominate any timespan in a user's timeline, changes in these mental states, and finally new emerging mental states. Drawing on these research ideas, the work also chalks potential in the field of studying external situational conditions which affect the presence and frequency of certain mental states affecting user personality. By means of this section we attempt to present a two-fold analysis. Firstly, elaborate on the in-depth insights per user for a small subset of users (4 users) that we analysed as a part

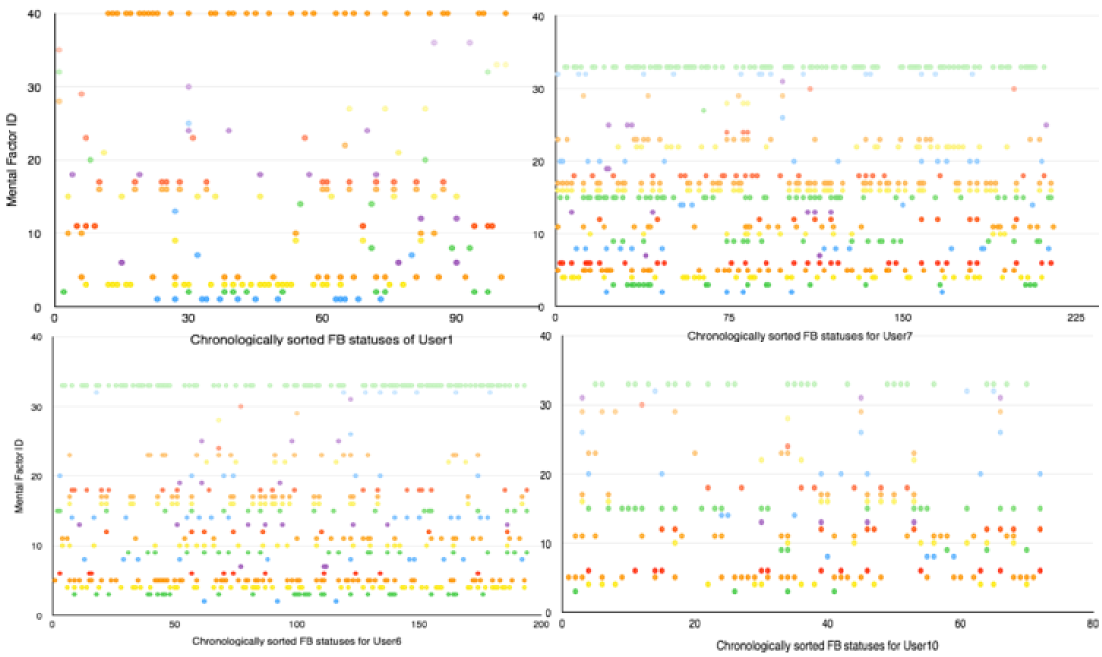

Figure 3: The moment by moment representation of 40 mental states for the given users from their FB statuses. The mental states unlike the Big 5 traits of the same users evolve over time. While few can be mapped to appear occasionally, some mental states also appear inherently in the users. The states are represented by the following values:
0: Aspiration, 1: Discursive Thinking, 2: Effort/Energy, 3: Desire, 4: Decision Making, 5: Greed, 6: Hate, 7: Dullness/Wavering, 8: Error, 9: Selfishness, 10: Worry, 11: Conceit/Pride, 12: Envy, 13: Shamelessness, 14: Recklessness, 15: Restlessness, 16: Sloth, 17: Torpor, 18: Skepticism/Doubt/Perplexity, 19: Generosity, 20: Faith/Confidence, 21: Discretion, 22: Equanimity, 23: Tranquility, 24: Lightness, 25: Adaptability, 26: Elasticity, 27: Proficiency, 28: Right Speech, 29: Right Action, 30: Right Livelihood, 31: Wisdom, 32: Goodwill, 33: Insight, 34: Sympathetic Joy, 35: Compassion, 36: Ignorance, 37: Attentiveness, 38: Modesty, 39: Uprightness, 40: Interest

of this study. Second, discuss some inferences which we found salient for the users we analysed by means of a bigger subset (50 users).

As illustrated in Figure 4 we present the maximum number of occurrences of particular mental factors in consecutive moments. This provides a statistical estimate of the number of times a mental factor has to occur to become part of a personality trait. Over a sample of 50 users and their FB data of an year, we find this estimate at an average of 6.2 moments (say μ is approximately 6 moments). We use this in identifying two important properties of the occurrence of mental factors leading to changes in personality. First are the ones which occur consecutively upto μ are identified as *Inherent mental states* of an individual. These states represented as a bag of words form a static image of our individual's personality traits. Secondly, we further identify *dynamic mental states* which occur in bursts and are contributory to the evolving personality of an individual. From the sample data, we find that the distance between the previous occurrence of a mental state and its current occurrence is at $1/5^{th}$ *of the total number of statuses/data points*. With this value, for each of the four users we identify their dynamic occurring mental states as described in Figure 3 and Table 4. For example, For User 6 we found the mental states: Decision Making (4, Yellow), Greed (5, Orange), Sloth (16, Orange), Torpor (17, Yellow), to be persistent and thus contributing to their personality. Whereas states such as Worry (10, Yellow), Confidence (20, Blue), Equanimity (22, Yellow), Tranquility (23, Orange) occurring in bursts causing the dynamically changing attributes of his/her personality to vary.

Along with these individual analysis, an extensive exploration of the dataset of another 50 randomly selected users helped us encounter some interesting findings which also validate the claims made by the Abhidhamma tradition. For instance, the doctrine suggests that the unpleasant and the pleasant mental factors occur exclusive of one another. The mental factors predicted by means of PACMAN adhered to this theory. For example, in one of the users (from the PACMAN dataset of predicted user states), while we did see an overall fluctuation in factors such as "faith" (belonging to pleasant mental factors) and "skepticism" (belonging to unpleasant mental factors), they never occurred at same instance (moment/status). Another interesting observation that can be made on the basis of the inherent and sporadic mental factors of all the

USER ID	MAXIMUM MOMENTS/TOTAL MOMENTS	INHERENT STATES	DYNAMIC STATES	BIG FIVE (enaco)
User 1	5/101	Desire(3, Yellow), Interest(40, Orange)	Aspiration (1, Blue), Decision Making (4, Orange)	nynny
User 6	7/194	Decision Making (4, Yellow), Greed (5, Orange), Sloth (16, Orange), Torpor(17, Yellow),	Worry (10, Yellow), Confidence (20, Blue), Equanimity (22, Yellow), Tranquility (23, Orange)	nnnny
User 7	5/215	Hate (6, Red), Decision Making (4, Yellow), Sloth (16, Orange), Torpor(17, Yellow), Restlessness (15, Green)	Desire (3, Green), Shamelessness (13, Violet)	nynny
User 10	5/72	Greed (5, Orange), Restlessness (15, Green)	Sloth (16, Orange), Torpor(17, Yellow), Envy (12, Red), Skepticism (18, Red), Faith (20, Blue)	nynny

Table 3: Analysis of Mental factors, (enaco) tuple represents the Big Five traits in the order of Extraversion, Neuroticism, Agreeableness, Conscientiousness and Openness, here **y** means that the trait is present and **n** means it is absent.

users (like those covered Table 4) are the co-occurrence of certain mental factors with one another. For instance, “sloth” always accompanies “torpor”, selfishness is usually present with an inherent state of greed, sharing informative resources (for instance news articles) helped in suggesting a basic level of the mental factor “insight” amongst users, and so on. Figure 3. is an illustration of the analysis of these basic phenomenon shown for 4 out of the 50 users we analysed.

In comparison to the state of art, we observe that while the Big Five characteristics of the user remain constant over the course of this year, PACMAN helps in mining certain dynamic mental states for the user’s persona. For instance, for User 1, while Interest (40, Orange) might be an inherent mental factor for the user, we do encounter a sudden change in the presence of other mental factors such as Aspiration (1, Blue), Decision Making (4, Orange). These are factors directly contributory (by definition) to one of the Big Five traits such as Agreeable defined to be absent in User 1 (this absence is perceived to be constant for the personality of the user). Various modern literature suggests that personality is a construct of various external stimuli and a different adaptive process for each one of us. Since, by most means the experiences we have are starkly different from one another, our personalities are also varied. In keeping with this theoretical foundation, we observe that while the Big Five personalities for 3 of the 4 randomly chosen users shown in Table 3 are the same, the PACMAN model accommodates different inherent and dynamic states for each one of them. We witnessed such changes in all the 50 users we analysed by means of our experimentation. These time-spans (for 50 users) which record the continuous presence of these dynamic mental factors (thus transcending them into inherent factors) have also been illustrated by means of Figure 2.

6. Conclusion & Future Work

The results of our initial investigation in dynamic personality analysis from social media provide encouraging evidence which backs the theoretical foothold of evolving user persona in psychology. Extracting and modeling mental states from lexical resources is just the beginning of our exploration into the plausible dynamics of personality change over time. By means of this study we attempt to propose an initial stochastic model of an individual, a theoretical foundation inspired from the Abhidhamma meditations of Buddhism to ascertain the transitional heuristics of the model (transition matrix and so on), and a machine learning framework to populate and analyse the dynamic personality model of a social media user. We envision extending this work by understanding the transition from one mental state to another by means of learning algorithms trained over a large influx of data. This will also help us to predict the various futuristic mental states given a substantial amount of (past and present) data for a user. We believe that our model will eventually accommodate not only applications focused on observing dynamic user persona, but also those which want to leverage from predicting user behavior, mentality, actions, and thoughts. The dataset we contribute by means of this work, a first annotated dataset for user mental states based on the Buddhist Model of Personality, would also be a useful resource helping researchers to conduct explorations in the domain. As a part of our future efforts we want to incorporate a predictive edge to our baseline model. We also hope to tap into the various psychological and social phenomenon that one can look into based on the trends observed in our mapping of an individual. We believe that this model can potentially be extended to address and recognise various clinical, social, psychological issues at an early stage by effectively learning the respective personality trends of people.

7. Acknowledgements

We thank Tejaswini Yeleswarapu, Sneha Nanawati for assistance with annotating datasets, and Professor Vasudev Varma for comments that greatly improved the manuscript.

8. Bibliographical References

Bachrach, Y., Kosinski, M., Graepel, T., Kohli, P., and Stillwell, D. (2012). Personality and patterns of facebook usage. In *Proceedings of the 4th Annual ACM Web Science Conference*, pages 24–32. ACM.

Blackburn, D. T. (1980). A generalized distance metric for the analysis of variable taxa. *Botanical Gazette*, pages 325–335.

Celli, F., Pianesi, F., Stillwell, D., and Kosinski, M. (2013). Workshop on computational personality recognition (shared task). In *Proceedings of the Workshop on Computational Personality Recognition*.

Costa Jr, P. T. and McCrae, R. R. (1980). Still stable after all these years: Personality as a key to some issues in adulthood and old age. *Life-span development and behavior*.

Costa Jr, P. T. and McCrae, R. R. (1995). Domains and facets: Hierarchical personality assessment using the revised neo personality inventory. *Journal of personality assessment*, 64(1):21–50.

Cox, C. (2004). From category to ontology: The changing role of dharma in sarvāstivāda abhidharma. *Journal of Indian philosophy*, 32(5):543–597.

Gao, W. and Zhou, Z.-H. (2013). On the consistency of multi-label learning. *Artificial Intelligence*, 199:22–44.

Golbeck, J., Robles, C., Edmondson, M., and Turner, K. (2011). 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*, pages 149–156. IEEE.

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

Grabovac, A. D., Lau, M. A., and Willett, B. R. (2011). Mechanisms of mindfulness: A buddhist psychological model. *Mindfulness*, 2(3):154–166.

Lancaster, B. L. (1997). On the stages of perception: Towards a synthesis of cognitive neuroscience and the buddhist abhidhamma tradition. *Journal of Consciousness Studies*, 4(2):122–122.

LaViers, A. and Egerstedt, M. (2011). The ballet automaton: A formal model for human motion. In *American Control Conference (ACC), 2011*, pages 3837–3842. IEEE.

McCrae, R. R., Costa Jr, P. T., Terracciano, A., Parker, W. D., Mills, C. J., De Fruyt, F., and Mervielde, I. (2002). Personality trait development from age 12 to age 18: Longitudinal, cross-sectional and cross-cultural analyses. *Journal of personality and social psychology*, 83(6):1456.

Mon, M. T. (1995). Buddha abhidhamma: Ultimate science.

Moscoso, S. and Salgado, J. F. (2004). 'dark side'personality styles as predictors of task, contextual, and job performance. *International Journal of Selection and Assessment*, 12(4):356–362.

Nomura, T. (1996). Generation of relations between individuals based on a stochastic automaton and an analogy from social psychology. *Proceedings of the ALIFE V Poster Presentations*, pages 125–132.

Pennebaker, J. W., Francis, M. E., and Booth, R. J. (2001). Linguistic inquiry and word count: Liwc 2001. *Mahway: Lawrence Erlbaum Associates*, 71:2001.

Poddar, S., Kattagoni, V., and Singh, N.). Personality mining from biographical data with the "adjectival marker".

Ross, C., Orr, E. S., Sisic, M., Arseneault, J. M., Simmering, M. G., and Orr, R. R. (2009). Personality and motivations associated with facebook use. *Computers in human behavior*, 25(2):578–586.

Soto, C. J., John, O. P., Gosling, S. D., and Potter, J. (2011). Age differences in personality traits from 10 to 65: Big five domains and facets in a large cross-sectional sample. *Journal of personality and social psychology*, 100(2):330.

Specht, J., Egloff, B., and Schmukle, S. C. (2011). Stability and change of personality across the life course: the impact of age and major life events on mean-level and rank-order stability of the big five. *Journal of personality and social psychology*, 101(4):862.

Tsoumakas, G. and Katakis, I. (2006). Multi-label classification: An overview. *Dept. of Informatics, Aristotle University of Thessaloniki, Greece*.