Modeling Fine-Grained Dynamics of Mood at Scale*

Adam Sadilek[1], Christopher Homan[2], Walter S. Lasecki[3]
Vincent Silenzio[4], Henry Kautz[3]

Google[1]
sadilekadam@google.com

Rochester Institute of Technology Computer Science[2]
cmh@cs.rit.edu

University of Rochester Computer Science[3]
{wlasecki, kautz}@cs.rochester.edu

University of Rochester Medical Center[4]
vincent_silenzio@urmc.edu

ABSTRACT

Mental health affects all aspects of people's lives, yet it remains difficult to obtain accurate data about influential factors. This work investigates quantifying, infering, and predicting—via social media data—the day-to-day mental state of individuals. We develop a statistical model of the affective state (mood) of specific individuals with up to hourly temporal resolution. This model enables us to quantify, in a unified way, aggregate mood trends, as well as patterns specific to individuals and groups of friends. It finds key features of mood variation over time and allows us to decompose a person's emotional state into a weighted sum of contributing factors—shedding new light on how mood affects, and is affected by environment. We then show that individuals' mood can be accurately predicted days into the future based on online behavior.

1. INTRODUCTION

Mental health is crucial to the well-being of individuals and society as a whole, yet many emotional disorders persist at epidemic levels. For instance, the 12-month prevalence of depression is >5%, ranking third globally among all diseases [4]. Mental health disorders lead to personal health problems and broader social issues: 2009 was the first time that the number of suicides exceeded deaths by motor vehicle accident. We present an approach to monitoring, at individual-to-epidemic and hourly-to-monthly scales, the emotional state of a population using only microblog data.

Our approach provides insight beyond that of official channels, where mental health issues are often underreported [3]. The predictive aspect of our model suggests opportunities for prevention and even *proactive* intervention. Recent work has demonstrated that microblogging data can be used to predict a variety of phenomena, including movie box-office revenues [2], elections [9], and flu epidemics [7], but has mostly focused on predicting aggregate properties of the data.

Using online social data, we quantify associations between mental health and the social and physical environment. Our methods tease apart the mood dynamics among thousands of Twitter users. We contribute a scalable model, based on principal components analysis, that can predict mood 10 days into the future with more than 61% accuracy, and next-day mood with 73% accuracy. We also obtain the following descriptive results about the role social networks play in the emotional state of its members:

1. Online friends share linguistic and behavioral traits (e.g., sleeping habits), but are surprisingly uncorrelated in other ways.
2. Correlation between friends' emotional states increases with their tweeting frequency, suggesting our approach will become more powerful as social media becomes more popular.
3. Eigen- and spectral- analysis of time-series data reveals patterns in mood change related to periodic cycles, and events.
4. Mood changes in one member of a clique of friends may trigger a cascade of changes in the others (emotional coupling).

2. DATASET

Our data was collected from Twitter, a popular micro-blogging service. Users post messages of 140 characters or less, and can subscribe to receive messages others (called *following*), without being followed back. When two users mutually follow one another, we say they are *friends*. There is evidence that Twitter friendships have a substantial overlap with offline friendships [6].

We collected a sample of public tweets from the New York City (NYC) metropolitan area (100km radius) over a one month period beginning on May 18, 2010. We logged nearly 16 million tweets authored by more than $630,000$ unique users.To put these statistics in context, the NYC metropolitan area has a population of $\tilde{1}9$ million people (www.census.gov/popest/metro/). Since this work studies the effects of people's location and co-location on mood, we concentrate on the 6,237 *geo-active users* who posted more than 100 GPS-tagged tweets during the collection period.

3. THE FEATURES

From these geo-tagged tweets we derive a number of different features. First, we divide the time over which the data was collected into even intervals of a day, hour, or week. Each of the features is then derived from the set of all tweets each user makes within each time interval. We call this unit of aggregation a *user-time block*.

Our features are based on the Diagnostic and Statistical Manual IV (DSM-IV-TR)'s definition of major depressive disorder [1], commonly known as SIG E CAPS (see Table 1).

Linguistic. We we measure the percent of words within each user-time block that appears in a Linguistic Inquiry and Word Count (LIWC) lexicon. LIWC is a collection of behavior-related lexicons developed by behavioral scientists. We focus on eight lexicons related emotional state: positive emotional (posemo), negative emotional (negemo), social, swear, anxiety, anger, sad, and death.

Geo-temporal. Described in Table 2.

*This research was funded by ARO grant W911NF-08-1-0242, ONR grant N00014-11-10417, OSD grant W81XWH-08-C0740, NSF grant 1111016, and Microsoft Research Ph.D. Fellowship.

WSDM 2014 New York, NY USA

	Criterion	Features
S	Sleep disturbance	*after_midnight, max_delta, max_2nd_delta, health*
I	Interest	*social, health*
G	Guilt feelings	*anxiety, sadness*
E	Energy level	*gyration, mileage, health*
C	Concentration	*anxiety, health*
A	Appetite / weight	*health*
P	Psychomotor activity	*gyration, mileage*
S	Psychomotor activity	*sadness, anxiety,, death*

Table 1: SIG E CAPs criteria (from DSM-IV-TR) mapped to our geo-tagged tweet features. (Table 2 explains the features.)

Name	Definition
num_tweets	number of tweets
after_midnight	number of tweets after midnight
max_delta	longest time interval between tweets
2nd_max_delta	second longest time interval
gyration	$\sqrt{(\sum_i (x_i - x_c)^2 + (y_i - y_c)^2)/n}$
mileage	$\sum_i \sqrt{(x_i - x_{i-1})^2 + (y_i - y_{i-1})^2}$

Table 2: The geo-temporal features used in our study. Here x_c and y_c are the mean position in space of an individual's tweets (taken as an estimate of the individual's home), x_i and y_i represent the location of a particular tweet, and n the number of tweets.

The features below are derived from those in the above categories. The next two categories describe the emotional environments of a user's social and physical spaces, respectively.

Friends. For each user-time block and linguistic or geo-temporal feature described, we take the associated "friends" feature to be the mean of the feature over all of a user's Twitter friends in the set.

Aggregate LIWC. A final feature, which is central to our predictive model, reduces all of the LIWC features into three discrete states σ: positive, neutral, and negative given by

$$\sigma = \text{sign}\left(\text{norm}\left(\frac{\text{posemo + social}}{\text{negemo + swear + anxiety + anger + sad + death}}\right)\right) \quad (1)$$

Equation 1 captures the ratio between positive and negative features, and is normalized per-user by subtracting out the mean and dividing by standard deviation. Equation 1 reduces the LIWC feature space to a single finite state variable, dramatically increasing the scalability of our predictive model in a way that attenuates the emotional aspects of the data, and reduces variation in the LIWC values of individual users. Note that anger, anxiety, and sad are subsets of the negemo lexicon. Including them separately allows each (including the part of negemo not in these subsets) to independently influence σ. By normalizing per user, our model learns dynamic patterns in mood that are invariant despite user biases.

4. METHODS

In the remainder of this paper, we address such questions as, "Given that 20% of your online friends are sad today, and that you have recently met eight strangers, who were in a bad mood as well, can we accurately quantify your increased likelihood of sadness in the near future?" We observe **correlations** between various features, using standard correlation coefficients. Next, we study how these correlations scale as Twitter use increases; for $w \in \{0, 1, \ldots, 50\}$ we remove those user-days with fewer than w tweets from the dataset.

We then use **Fourier analysis** to extract in a principled way significant time periods in emotional signals. We map a function f from the time to the frequency domain via discrete the Fourier transform (DFT), given by

$$Z_k = \mathscr{F}_t\left[\{x_t\}_{t=0}^{T-1}\right](k) = \sum_{t=0}^{T-1} x_t e^{\left(-2\pi k \frac{t}{T}\right)i} \quad (2)$$

where $x_0, \ldots, x_{T-1}$ is a series denoting mood over T time chunks.

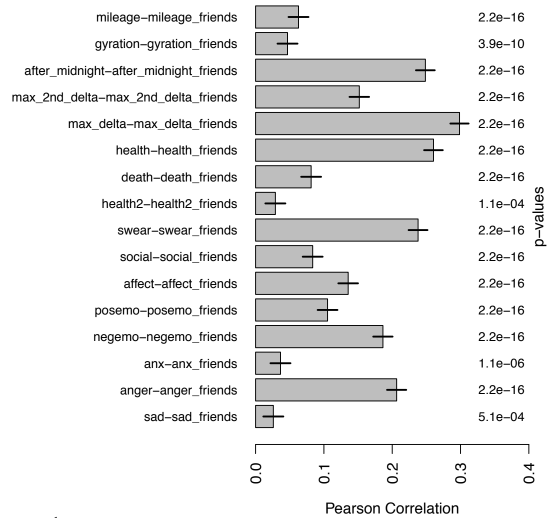

Figure 1: Pearson correlations between users and the mean of their friends. Error bars are 95% confidence intervals and p-values are shown on the right.

Finally, we use **Principal component analysis (PCA)** to perform descriptive time-series and well as predictive analysis. PCA is a dimensionality reduction technique that transforms the original data into a new basis, where the basis vectors are, in turn, aligned with the directions of the highest remaining variance of the data. PCA can be performed by eigendecomposition (also called eigenanalysis) of the data covariance matrix, or by applying singular value decomposition (SVD) directly on the data matrix. Our implementation uses the latter approach, as it is more numerically stable. We leverage PCA's probabilistic interpretation as a latent variable model, which endows our model with all the practical advantages stemming from this relationship, including efficient learning and dealing with missing data [8].

5. DESCRIPTIVE RESULTS

Our data shows correlations between an individual's affective state and those of their friends. Figure 1 summarizes our results. In terms of the SIG E CAPS criteria, it shows that, while major depression may be transmitted individual-to-individual, its underlying components may do so independently.

5.1 More Data Yields Stronger Correlations

To explore the effect of data volume on our model, we measured how correlated and uncorrelated features grew with increasing amounts of data. Using a threshold of between 1 and 50 daily tweets, and dividing features into this with above 0.1 correlation (correlated) and those below 0.1 (uncorrelated), we found the average Spearman correlation was 0.205 (Pearson correlation of 0.129) when all workers were considered, but grew to 0.416 (Pearson: 0.416) for users that had at least 50 tweets. The trend was also strongly logarithmic ($R^2 = 0.996$ for Spearman and $R^2 = 0.973$ for Pearson), suggesting that increases in per-user tweet density can be modest over time and still benefit our approach.

5.2 Time-series Analysis

Fig. 2 shows the aggregate daily affect over all users in our dataset. Significant dates are highlighted. Intuitively, we see that Monday is consistently the most negative day in a week, except for Memorial Day Monday, which is a U.S. national holiday.

Similarly, Fig. 3 shows hourly fluctuations in positive affect, where Memorial Day again exhibits a considerable increase in happiness. In agreement with prior work [5], we see that mornings are

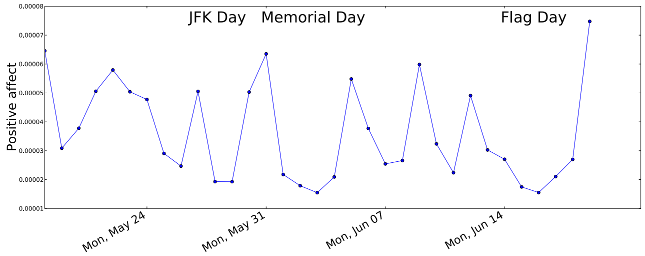

Figure 2: Daily trends in positive affect aggregated over all users. Monday is consistently the most "depressing" day in a week, expect for Memorial Day Monday, which is the second most positive day in the dataset.

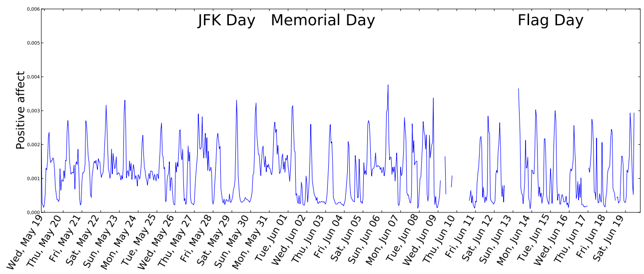

Figure 3: Hourly trends in positive affect aggregated over all users. Note the strong effect of Memorial Day weekend.

the happiest time of day. We also see that most days have an affective "low" mid-day with a slight improvement in the evening.

We apply Fourier analysis to find the periodicities present in both daily and hourly trends (see Fig. 4). We see strong periodicities at 12 and 24 hours, and 3, 5 and 7 days. Hourly and daily resolutions produce consistent patterns (*e.g.*, we detect a strong 5-day period and a corresponding 120-hour period). Interestingly, a significant fraction of power is concentrated in the hourly fluctuation of mood.

These findings suggest that our measure of affect is consistent with prior work and captures affective signals.

5.3 Descriptive Eigenanalysis

Eigenanalysis enables us to decompose the sentiment signal into factors and quantify the contributions of each factor in an unsupervised fashion. We apply probabilistic PCA on our data matrix D (rows are user-weeks, columns are days of the week). Each element σ in D is given by Equation 1. We call the eigenvectors of the covariance matrix of D *eigenweeks*. Fig. 5 shows eigenweeks learned from our dataset, ranked by the percentage of variance explained.

Eigenweeks show long-range patterns, but we are also interested in the broader interactions between days within a month. We transform the data matrix into D' so that each row represents a user-month (columns are days). We call the eigenvectors of the covariance matrix of D' *eigenmonths*. Fig. 6 shows learned eigenmonths ranked by the amount of variance explained. One can think of the eigenmonths as a "library" of prototypical months from which any actual user-months can be reconstructed. Since the behavior of users can be accurately described as a weighted sum of only a few

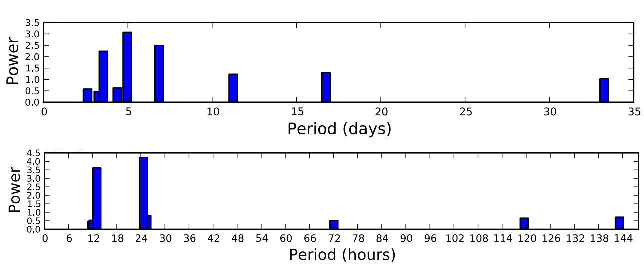

Figure 4: Power spectrum of daily (top) and hourly (bottom) positive affect aggregated over all users. We see strong periodicities at 12 and 24 hours, and 3, 5 and 7 days (but much of power lies in the hourly fluctuations).

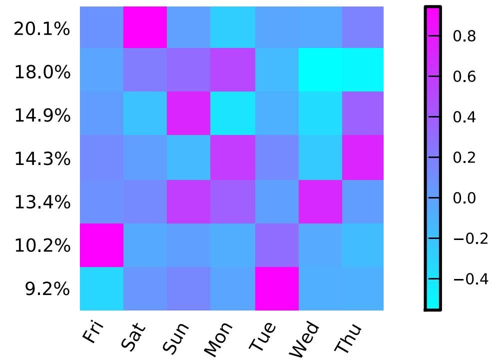

Figure 5: Each row is an eigenweek – an eigenvector of the covariance matrix of σ values for each user-day by day of the week. The first two eigenweeks model "happy" (pink) Saturday and Sunday followed by a "sad" (blue) Monday and alone explain nearly 38.1% of the variance in *all* weeks.

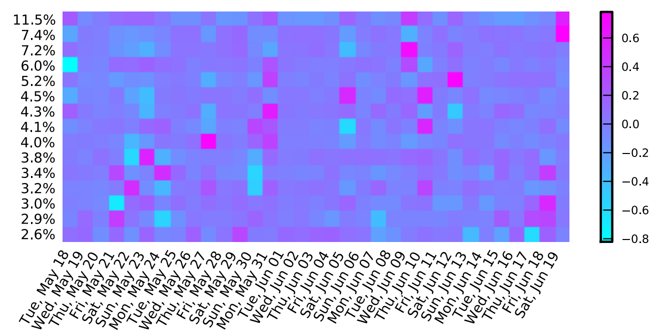

Figure 6: Top-15 most significant eigenmonths jointly explain over 73% of the variance. An eigenmonth is an eigenvector of the covariance matrix of σ values for each user-day by day of the month. The first eigenmonth captures the essence of a typical month for most users: happy weekends and less happy weekdays. But different clusters of users are happier at different times. The second eigenmonth models individuals who may have to work on weekends and have time off mid-week. Eigenmonths #4-9 capture enjoyment of Memorial Day Weekend, including Monday, May 31.

eigenmonths, there is a large amount of *coupling* across individuals. Even though our dataset contains 6,237 user-months, just 15 eigenmonths are enough to explain more than 73% of the variance.

5.4 Cascading Emotions in Cliques

We use eigendecomposition to tease apart the *global* signal that shapes the mood of virtually everybody in a given locale (*e.g.*, those driven by weather, elections, etc.) and the *local* fluctuations that affect only subsets of the population, namely cliques of friends. To quantify patterns internal only to a clique c, we first learn dominant eigenmonths E_m from all user-days, excluding users in c. Optimal projection of c onto E_m enables us to "subtract out" the global signal and highlight the dynamics within the clique.

Fig. 7 shows the residuals (*i.e.*, the projection error) for one clique. Turquoise (pink) color shows user-days that are unexpectedly negative (positive), respectively. We observe an interesting phenomenon, where a change in a person's mood triggers a cascade of similar changes in other users in the clique. This effect has not been, to our knowledge, measured before. It is evident that nonparametric statistical analysis reveals important patterns that would be otherwise very difficult to see.

Using eigenanalysis, we can now tease apart the patterns of mood specific to a target person from the background fluctuations of sentiment observed in a broader population. This is crucial for suicide prevention and other mental health applications where anomalous behavior needs to be recognized quickly and effectively.

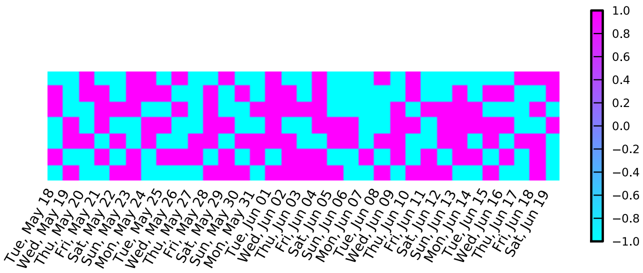

Figure 7: Clique residuals showing the dynamics of affective state within a clique of seven friends. Each row is user-month. Note the "diagonal" patterns that indicate a change in mood of a person is immediately followed by a cascade of corresponding changes in other members of the clique.

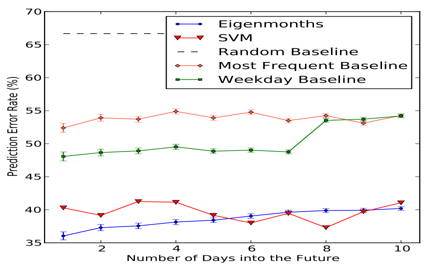

Figure 8: Error rates of predicting emotional state ten days into the future. As another model for comparison, we trained a support vector machine (SVM) with a linear kernel on the LIWC features of each person for the last 11 days. For longer-term predictions, SVM mostly dominates. Eigenanalysis has the ability to capture long-range statistical patterns, whereas SVMs learn more robust margins.

6. PREDICTIVE RESULTS

We now turn to the evaluation of the predictive power of our eigendecomposition model. The target to predict is which future user-days will be positive, neutral, or negative.

To predict future mood using eigendecomposition (PCA), we first learn eigen-3-weeks using the first 21 days of all user data. The affective state of each person p over the last (withheld) 10 days in the dataset is then predicted by the following process. We project the last 11 days of p's observed data onto the space induced by the first 11 days of the eigen-3-weeks. This results in a set of weights—one weight for each eigen-3-week. The prediction is the weighted sum of the last 10 days of the eigen-3-weeks. The length of the eigenvectors is limited only by the size of the dataset (*e.g.*, we used eigen-3-weeks and withheld the last 10 days of data for validation). In a real-time application, the patterns can be of arbitrary length.

As cross-validation, we repeat the evaluation process with all users and report the mean error rate in Fig. 8. Three baselines are considered: random guess, most frequent baseline (predict the mode of a user's past states), and weekday (predict the mode for each day of week). Eigendecomposition outperforms all these methods for predictions within 5 days into the future. The experiments show that PCA takes advantage of the complex global correlations in mood and is successful at near-term predictions. As the noise begins to dominate for temporally distant query times, the SVM achieves slightly higher accuracy.

7. DISCUSSION

Although the final predictive model does not include any friend or colocation feature, we did try adding them to the models, but discovered no significant difference in performance. This may be due to limitations in our methods. Given how little Twitter tells us about a user's social or spatial context, it is somewhat surprising that as many correlations in Table 1 were as strong as they were. One interpretation is that these unnormalized features are relatively static, even as the normalized data used in the predictive model is more volatile. This could be seen as evidence of the social support model, in which social structure creates an environment for managing emotional problems, as opposed to the viral model, where one can be "infected" with the emotional state of others.

7.1 Limitations

The problem of assessing the mental state of an individual is hard, even for an expert observer with frequent firsthand interactions with them. Direct, medically valid diagnosis of mental state is fraught with ethical considerations. Instead, we estimate the mood of an individual solely based on secondary information from that individual's microblogging activity. We base our feature set on the DSM-IV-TR's definition of major depressive disorder [1], which provides an imperfect but principled approach. We are less concerned here with grounding our models in a particular medical framework than we are in capturing the dynamic relationships that emerge between affect and social and environmental factors, using social media as a proxy.

Noisy and incomplete data pose a major challenge to all research in this area. Users participate in online social networks in varying degrees of intensity. The distribution of the number of messages posted by the users follows a power law. Moreover, usage statistics exhibit chaotic fluctuations [10]. Even an otherwise typical user may go for a week without posting anything. We explore model quality as a function of the level of activity of the users considered—an important problem that is only beginning to emerge in social network research.

8. CONCLUSION

This paper explores prediction of individuals' affective state on the basis of fine-grained social network data—an important instance of the general problem of modeling emergent properties of large real-world dynamical systems. We use geo-tagged Twitter posts as a noisy proxy of users' emotion, mobility, and overall context. We quantify the impact of a number of otherwise elusive factors on people's mental health, and show that we can accurately predict one's future mood from past interactions with others.

9. REFERENCES

[1] A. P. Association. *DSM-IV-TR.* 2000.
[2] S. Asur and B. Huberman. Predicting the future with social media. *WI-IAT 2010*, volume 1, 492–499. 2010.
[3] J. M. Bertolote and A. Fleischmann. Suicide and psychiatric diagnosis: a worldwide perspective. *World Psychiatry*, 1:181, 2002.
[4] P. Cuijpers, A. T. Beekman, and C. F. Reynolds III. Preventing depression a global priority. *The Journal of the American Medical Association*, 307(10):1033–1034, 2012.
[5] S. Golder and M. Macy. Diurnal and seasonal mood vary with work, sleep, and daylength across diverse cultures. *Science*, 333(6051):1878–1881, 2011.
[6] A. Gruzd, B. Wellman, and Y. Takhteyev. Imagining Twitter as an imagined community. *American Behavioral Scientist, Special issue on Imagined Communities*, 2011.
[7] A. Sadilek, H. Kautz, and V. Silenzio. Predicting disease transmission from geo-tagged micro-blog data. *AAAI 2012*.
[8] M. Tipping and C. Bishop. Probabilistic principal component analysis. *Journal of the Royal Statistical Society. Series B, Statistical Methodology*, p611–622, 1999.
[9] A. Tumasjan, T. Sprenger, P. Sandner, and I. Welpe. Predicting elections with Twitter: What 140 characters reveal about political sentiment. *ICWSM 2010*, 178–185, 2010.
[10] S. Ye and S. Wu. Measuring message propagation and social influence on twitter. *Social Informatics*, 216–231, 2010.