Predicting Postpartum Changes in Emotion and Behavior via Social Media

Munmun De Choudhury **Scott Counts** **Eric Horvitz**
Microsoft Research, One Microsoft Way, Redmond WA 98052
{munmund, counts, horvitz}@microsoft.com

ABSTRACT
We consider social media as a promising tool for public health, focusing on the use of Twitter posts to build predictive models about the influence of childbirth on the forthcoming behavior and mood of new mothers. Using Twitter posts, we quantify postpartum changes in 376 mothers along dimensions of social engagement, emotion, social network, and linguistic style. We then construct statistical models from a training set of observations of these measures before and after the reported childbirth, to forecast significant postpartum changes in mothers. The predictive models can classify mothers who will change significantly following childbirth with an accuracy of 71%, using observations about their prenatal behavior, and as accurately as 80-83% when additionally leveraging the initial 2-3 weeks of postnatal data. The study is motivated by the opportunity to use social media to identify mothers at risk of postpartum depression, an underreported health concern among large populations, and to inform the design of low-cost, privacy-sensitive early-warning systems and intervention programs aimed at promoting wellness postpartum.

Author Keywords
behavioral health; childbirth; depression; emotion; health; language; postpartum; PPD; social media; Twitter; wellness

ACM Classification Keywords
H.3.4; H.5.2; H.5.3

INTRODUCTION
Having a baby is a major life event that creates significant changes in the lives of new parents. Sleep and daily routines are disrupted, and adjustments must be made in personal and professional lives. First time mothers may be particularly challenged with navigating the new, complex realm of caring for their newborn. Adding to the challenges, many new mothers experience psychological changes, such as the "baby blues," a temporary condition involving mild mood instability, anxiety, and depression. Beyond such relatively mild blues, a portion of new mothers experience more extreme changes. According to the CDC[1], between 12 and 20 percent of new mothers report postpartum depression (a 13% incidence rate in a meta-analysis report [20]), a form of depression that typically begins in the first month after giving birth and is characterized by symptoms including sadness, guilt, exhaustion, and anxiety [16].

We examine social media as a tool in public health. Social media is a source of population data about behaviors, thoughts, and emotions, and can serve as record and sensor for events in peoples' lives. Whether in the form of explicit commentary, patterns of posting, or in the subtleties of language used, social media posts bear the potential to offer evidence as to how a person is affected by life events. Within this context, we investigate the feasibility of *forecasting* future behavioral changes of mothers following the important life event of childbirth. We extend our prior research that examines the value of harnessing social media signals to characterize changes in new mothers, along three dimensions: patterns of posting, linguistic style, and emotional expression [8]. These measures were used to explore the behavioral changes of a cohort of new mothers who showed large postpartum changes, including those showing increases in indicators of negative emotion and lowered posting volume.

Here, we focus on predicting significant changes in new mothers postpartum *in advance of their being exhibited*, including behavioral changes that may be associated with significant downturns in mood. We base predictions on prior behavioral patterns of new mothers as manifested on Twitter. We construct statistical models from training data to predict significant future changes in a test cohort. To construct and test the predictive models, we harness evidence from 33 different measures, spanning changes in posting behavior, ego-network, linguistic style, and emotional expression. We demonstrate that we are able to identify which new mothers will exhibit large changes several months into their postpartum phase, with a mean classification accuracy of 71% when we leverage behavioral data from only the prenatal period (i.e., *before* the birth of the baby). We obtain a considerable boost in prediction power, with accuracy in the range 80-83%, when we include in our evidential horizon data from the first few

CHI'13, April 27 – May 2, 2013, Paris, France.

[1] http://www.cdc.gov/reproductivehealth/Depression/

postnatal weeks. We find evidence that certain types of measures can be better predicted, including expression of negative affect, interpersonal pronoun use, and degree of interaction with one's network.

We see such predictions and predictive power as a step toward creating private, low-cost tools that can raise awareness among new mothers who are at the highest risk for suffering from postpartum depression (PPD), thus empowering them to seek understanding, as well as emotional support or professional assistance in a timely manner. More broadly we seek to highlight opportunities for studying the influence of major life events on people using public content shared in an online setting.

BACKGROUND

Risk and Presence of Postpartum Changes

Most attempts at understanding changes in behavior of mothers following childbirth focus on identifying the presence of PPD and on risk factors for PPD. We do not focus directly on identifying PPD, but find literature on PPD relevant, particularly from a methodological standpoint. PPD is underreported, with estimates that as many as 50% of cases of PPD go undetected [24], in part because even when mothers feel seriously depressed they do not seek help; one study reported that fewer than half of mothers suffering PPD report their depression [15]. We therefore believe that a predictive modeling effort can be especially valuable for early detection of PPD, given the major known challenges with detecting this condition [1].

Surveys such as the Postpartum Depression Predictors Inventory (PPDI) [2] reflect meta-analyses of risk factors for PPD [3], including prenatal depression, life stress, lack of social support, socioeconomic status, maternity blues, and infant temperament, among others (see also [19] for more on risks and influencers of PPD). Social support in particular has been shown to influence the attitudes, emotions, and behaviors of new mothers [11,29]. Neilson [18] identified social isolation and psychological stress, as significant predictors of PPD. Although some of these risk factors (e.g., socioeconomic status) are not easily inferred via social media posts, proxies for several factors could be monitored. For example, social support might be inferred from connectivity and the amount of social interaction a person has on social media. Infant temperament might be measured through posts the mother makes about her baby.

We do not yet understand the links between the measures we predict and PPD, nor do we propose the methods we present as a replacement for traditional PPD assessment and risk stratification. Rather, we explore the potential to use an analysis of online behavior to augment existing techniques for assessing changes in new mothers.

Social Media, Language, and Behavioral Prediction

The short history of using social media for behavioral assessment and prediction largely centers on population-scale measurement via retrospective or real-time studies. Facebook has been mined to create a happiness index that reveals the daily sentiment of people in the US [14]. Twitter posts reflect daily life patterns [12], and in the public health domain have been correlated with disease rates [22]. Sadelik et al. [25] study individual level predictions using Twitter data to estimate the likelihood that a person will become ill based on degree of co-location in the recent past, with others who became sick.

In the realm of psychological health, there is evidence that social media can serve both as an aid to mental health and as a data source for studying it. For instance, Facebook use has been shown to have positive influences on psychological well-being such as helping those with lower self-esteem to attain higher social capital [28]. Brubaker et al. [5] identify language features, including both sentiment and linguistic style features, of MySpace posts that correlate with expression of grief and distress around death of loved ones.

There is a growing understanding of the relationship among linguistic analysis, human behavior, and psychology. Text analysis has been used to identify markers of emotional closeness [13], as well as depression [17], anxiety, and other psychological disorders [6,21,23,30]. Even the process of writing has been found to help people cope with difficult situations, such as the loss of a job [27].

Weaving together several threads of prior research, the connections between social media, human experience, and text analysis suggest that linguistic, activity, and network oriented analyses of social media posts can offer a novel methodology for augmenting traditional approaches to measuring the influences of the birth of a child on mothers. For example, the frequency of use of first-person pronouns in writing has been found to be a correlate of depression [7]. This finding could be leveraged to help estimate *prenatal* depression, itself a strong predictor of PPD [3]. Motivated on these lines, in our previous work, we utilized a variety of such behavioral cues from Twitter, spanning activity, emotion and language in order to understand changes underwent by new mothers. In this paper, we show the extent to which these social media-based measures can *predict* ahead of time, extreme changes in activity and correlates of mood in new mothers following childbirth.

DATA

We begin by discussing our data collection methodology for predicting new mothers' behavioral changes in postpartum. We seek female Twitter users and follow a two-stage approach to construct a high-confidence sample of new mothers. We chose Twitter because it is public and provides a longitudinal record of the events, thoughts, and emotions experienced in daily life. Other topically relevant social media sources such as support forums for mothers tend to be explicitly problem-focused and lack per-person data density and longitudinal requirements for our study.

First, we identify posts that indicate a recent birth of a child on Twitter (English language posts only). We filter the Twitter Firehose stream (made available to us via a contract with Twitter) based on phrases typically used in newspaper announcements of births. We examined announcements in four newspapers, totaling 604 birth announcement posts in 2009-2012. Note that the phrases extracted from the newspaper announcements resonate with intuitions that parents announce the birth of their children in canonical ways, often including mention of the labor experience and reporting on the physical details of their newborn child, including gender, weight, and height. These phrases also tend to be unambiguous (e.g., "Announcing the birth of a new baby boy…"). We list in Table 1 examples of these phrases that we used as search queries to find birth events on Twitter. The authors of the resulting posts (2,929 posts in all between June 2011 and April 2012) constituted an initial set of candidate new mothers.

(1) birth, weigh*, pounds/lbs, inches, length/long, baby/son/daughter/boy/girl
(2) announc*, birth/arrival of, son/daughter/brother/sister
(3) are the parents of, son/daughter/boy/girl/baby
(4) welcome* home by, brother/sister/sibling*
(5) is the proud big brother/sister
(6) after, labor, born
(7) it's a boy/girl, born

Table 1. List of queries for identifying birth events on Twitter.

In the second step, we identify a high probability set of new mothers, first by performing gender inference via a classifier of first names trained on U.S. Census data, and then obtaining confidence ratings on these inferences from crowdworkers recruited through Amazon's Mechanical Turk interface. Specifically, we showed each worker (min. 95% approval rating, English language proficient, and familiar with Twitter) a set of 10 Twitter posts from each mother in our candidate set, such that five posts were posted right before the index childbirth post, and five after the post. Additional cues from the candidate mothers' Twitter profiles were also provided to enable better judgment—e.g., their Twitter profile bio, picture, and a link to their Twitter profile. The specific question posed to the crowdworkers involved choosing from a yes/no/maybe multiple-option menu, per candidate mother, to indicate if she was truly a new mother. We collected five ratings per candidate mother from the crowdworkers, and used the majority rating as the correct label, after independent inspection from two researchers (Fleiss-Kappa was 0.71).

The final dataset consisted of 376 validated new mothers, who exhibited strong evidence (on Twitter) of having given birth to a child at a time point between June 2011 and April 2012. Finally, for each of these 376 new mothers, we queried their public Twitter timelines in the Firehose stream to collect all of their posts in two three-month periods, corresponding to prenatal and postnatal phases, around the point of childbirth. The choice of a three-month window for prenatal and postnatal data is motivated from PPD studies in the medical literature [14]. For instance, for a mother with evidence of childbirth in October 2011, the prenatal phase would correspond to data between July 2011 and September 2011, while the postnatal period would consist of data from November 2011 to January 2012. In this manner, the total timespan of our dataset is between March 2011 and July 2012, with a total of 36,948 posts from the 376 mothers during the prenatal period, and a total of 40,426 posts from the same mothers during postpartum.

MEASURES OF BEHAVIORAL CHANGE

We employ four types of measures to characterize the behavior and mood of the new mothers as below.

Engagement. A measure of overall engagement with communications in social media is *volume*, defined as the average normalized number of posts per day made by the new mothers over the prenatal and postnatal periods. We define a second engagement measure to be the mean proportion of *reply* posts (@-replies) from a mother over a day; this serves as a proxy for her level of activity in social interaction with other Twitter users. The third measure is the fraction of *retweets* from a mother per day, which indicates how the mother participates in information sharing with her followers. The proportion of *links* (urls) shared by each mother over a day comprises a fourth engagement measure. We define a fifth measure as the fraction of *question-centric* posts from a mother on a given day; this measure indicates the mother's tendency to seek information from the greater Twitter community.

Ego-network. We define two measures that characterize the nature of a mother's egocentric social network. The first measure is the number of *followers* or inlinks of a mother at a given day, while the second is the count of her *followees* or outlinks. Inlinks demonstrate her reach/popularity in the larger network (in a coarse sense), while outlinks indicate her tendency to act as an informational hub and remain connected with others.

Emotion. We consider four measures of the emotional state of mothers: *positive affect* (PA), *negative affect* (NA), *activation*, and *dominance*. PA and NA are computed using the psycholinguistic lexicon LIWC (http://www.liwc.net/). LIWC's emotion categories have been scientifically validated to perform well for determining affect with Internet language [23], as well as from short text data, e.g., Twitter [8,12]. For PA computation, we focus on words in the positive emotion category of LIWC [12]. For NA, we consider the negative affect categories: *negative emotion, anger, anxiety, sadness.*

Like in [8], we use the ANEW lexicon [4] for computing activation and dominance. This resource provides a set of normative emotional ratings (pleasure, arousal, and dominance) for a large number of words (~2000) in the English language. Activation measures the intensity of an emotion, an important dimension beyond the valence captured by PA and NA. As an example, while *frustrated* and *infuriated* are both negative emotions, *infuriated* is higher in activation. Dominance represents the controlling and dominant nature of an emotion. For instance while both *fear* and *anger* are negative emotions, *anger* is a dominant emotion, while *fear* is a submissive emotion.

Linguistic Style. We also use measures to characterize change based on the use of linguistic styles in posts from new mothers during the prenatal and the postnatal periods [8]. Linguistic styles capture how language is used by individuals and provide information about their behavioral characteristics subject to their social environment [7,23]. We again referred to LIWC for determining 22 specific linguistic styles: *articles, auxiliary verbs, conjunctions, adverbs, impersonal pronouns, personal pronouns, prepositions, functional words, fillers, assent, negation, certainty* and *quantifiers.*

OBSERVATIONS OF POSTPARTUM CHANGES

We share several empirical observations prior to focusing on the task of constructing classifiers to predict postpartum changes in new mothers. The observations illustrate the manner in which mothers change in their behavior as manifested through our different measures. Figure 1 shows heat maps of individual-level changes for five measures: volume, replies, negative affect, activation and first-person pronoun use. For brevity, we focus on these measures as illustrative examples of change, though we note that most measures showed similar patterns. Greater details of changes in different measures across the cohorts of mothers can be referred to in [8].

The heat maps in Figure 1 display a variety of shifts in patterns, including how some new mothers may show noticeable changes in their behavior on Twitter. Changes are observed in the decreasing and increasing directions. Some new mothers exhibit only small changes while others show more extreme shifts in one or more measures.

For instance, the heat maps show a considerable decrease in volume, replies, and activation following childbirth for a subset of the mothers. Such decreases indicate that these women are posting less, suggesting a possible loss of social connectedness following childbirth. We also observe a noticeable increase in NA for a portion of the mothers. This finding may be attributable to these mothers' physical, mental, and emotional exhaustion [19], as well as the sleep deprivation typical of parenting a newborn. Similarly, the drastic reduction in activation during the postnatal phase for some mothers may indicate emotions of low intensity, perhaps based in fatigue from handling daily tasks around care of the newborn. Finally, we find that the use of the first-person pronoun increases considerably for some mothers, possibly reflecting increases in attention to self and emotional distancing from others after childbirth [7,30].

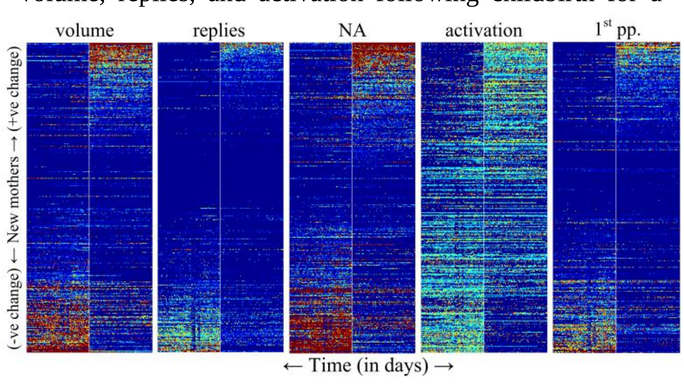

Figure 1. Heat maps (scaled on RGB spectrum with red=high, blue=low) showing changes in five measures behavior in new mothers during postpartum, compared to prenatal period. Center white line in each figure represents estimated time of childbirth.

The cohort of mothers showing extreme changes (in either direction, depending on the measure, e.g., decreasing for volume, increasing for NA) is of particular interest to us, as these significant changes could indicate difficulty adjusting to new motherhood, including emotional changes seen in maternity blues. In fact, prior literature establishes that the above-observed signs of considerable decrease in social interaction (e.g., lowered volume), generally unhappy postings (e.g., lowered PA, high NA), and psychological distancing (e.g., high 1st person pronoun use) may point to emotional instability, depression vulnerability, or existing depression [7]. In this light, classifying and predicting the behavior of mothers who will later show extreme negative postpartum changes *ahead of time*, may be useful in flagging risk of forthcoming behavioral health problems.

PREDICTION FRAMEWORK

We now pursue the use of supervised learning to construct classifiers trained to predict postpartum behavioral and emotional changes of new mothers. To this end, we use observations of our measures during the prenatal phase: engagement, emotion, ego-network, and linguistic style.

Classification Setup

Given observed data during the prenatal period, we frame prediction as a binary classification problem per measure, where we discriminate the following two classes:

- *Extreme-changing mothers*: the first group (*class C1*) comprises mothers whose mean value of a measure in postpartum *after* childbirth is considerably less than (or greater than, depending on the measure of interest) that *before* childbirth, with respect to a suitably chosen empirical threshold τ that is discussed below;
- *Standard-changing mothers*: the second group (*class C0*) comprises those mothers not in the extreme change class.

To construct these two classes, we need to estimate the expected directionality of change[2] of a particular measure indicating the manner in which the mothers deviate in their

[2]Normalized change per mother is defined as the difference between mean value of a measure postpartum and the mean value of the same measure in the prenatal period, divided by the latter.

behavior postpartum. To this end, we leverage insights from prior literature that examine association between the linguistic expression of individuals and their responses to traumatic context and crises, including depression vulnerability [7,30]. For instance, increases in NA, and decreases in activation are known to be indicative of emotional instability [6,8]. Thus, in our class definition for NA, we consider increasing directional changes for demarcating C1, while decreasing directional changes corresponding to activation. The directionality of change for other measures is discussed in greater detail in Table 2.

For the purposes of classification, we represent each mother as a vector of features, where the features consist of daily values of all of the measures we track during the prenatal period (33 measures in all). The high dimensionality of the feature space can lead to overfitting to the training data. To avoid overfitting and to eliminate feature redundancy and interaction, we employ principal component analysis (PCA) and regularized random forest procedures [10]. We compare several different parametric and non-parametric classifiers to empirically determine the best suitable classification technique, including linear, quadratic, discriminant classifiers, naïve Bayes, *k*-nearest neighbor, decision trees, and Support Vector Machines with a radial-basis function (RBF) kernel [10]. The best performing classifier was found to be the SVM across all measures, which outperformed the other methods with prediction accuracy improvements in the range of 10-35%. We use SVMs with an RBF kernel as the classification method for the rest of this paper. For all of our analyses, we train and test one classifier for each measure. We use five-fold cross validation on the set of 376 mothers, over 100 randomized experimental runs.

Estimating Optimal Threshold for Class Definition

We now discuss our method of identifying a threshold τ per each behavioral measure for defining the two classes of mothers as defined above. The ideal threshold is a conceptual boundary that would distinctly separate the extreme-changing mothers, from the mothers who show smaller changes. We define the threshold τ as the minimum normalized change[2] in the value of the measure *after* childbirth, compared to that *before* birth.

We follow an empirical strategy involving the optimization of the threshold τ for class definition. Specifically, given a measure that we intend to predict, we step through a range of threshold values (τ=0, .01, .05, .1, .15, …, .3, etc.) to define the classes. For each such case, we train an SVM classifier and attempt to maximize the log likelihood of the learned model using expectation-maximization (EM) [10].

In conducting this series of model fittings in pursuit of the optimal τ for each measure, the best model fit for τ is observed in the range .05 to .2 across all of our measures. Corresponding to these values of τ, we report, in Table 2, the median increase/decrease in changes per measure exhibited by the two classes of mothers. Qualitatively, the changes for the extreme-changing mothers (C1) stand out against the background behaviors seen in other mothers. As an example, for the volume measure, mothers in the extreme change class (C1), exhibit median change of -0.88 postpartum, indicating an 88% drop in posts per day, while the standard changing mothers actually increase in posting volume by 84%. While both groups change in volume, the extreme changing mothers are distinguished by a relative change of 2.05 times that of standard changers and in a decreasing direction.

Measures	C1	C0	Measures	C1	C0
volume	-0.878	0.838	verbs	-0.911	0.951
Replies	-0.983	1.417	aux-verbs	-0.889	0.914
RT	-0.847	2.188	adverbs	3.025	-0.678
Links	-0.791	2.347	preposition	-0.899	1.020
Questions	2.110	-0.749	conjunction	-0.890	1.042
PA	-0.798	0.428	negate	2.867	-0.690
NA	1.726	-0.593	quantifier	-0.896	0.950
activation	-0.743	0.525	swear	3.258	-0.751
dominance	-0.717	0.543	tentative	3.710	-0.700
Followers	-0.941	1.451	certain	-0.997	1.386
Followees	-0.928	1.013	inhibition	3.425	-0.722
functional words	4.574	-0.865	inclusive	-0.914	1.048
1st pp.	1.924	-0.426	exclusive	-0.822	2.304
2nd pp.	-0.901	0.872	assent	-0.904	1.073
3rd pp.	-0.894	0.990	non-fluency	3.028	-0.724
indefinite pronoun	-0.810	2.082	filler	2.929	-0.786
article	3.940	-0.701			

Table 2. Median changes in measures for two classes of new mothers with C1 corresponding to extreme-changing mothers and C0 corresponding to mothers showing smaller changes. Directionality of changes reflects for each measure whether extreme-changing mothers are considered to be changing in increasing or decreasing directions. *Volume* decreases postpartum for C1, but *negative affect* (NA) increases [6,8].

This and similar observations from Table 2 indicate that the optimal τ's for different measures obtained through the learning technique separate the two classes well and help us construct reliable ground truth labels for the mothers. We note that, over the range of optimal τ, the sizes of the two classes of behavior across all of the measures fall in the range 8.8-28.2% for class C1 (median: 15.7%; standard deviation 4.3%), and the rest for class C0. In the next section, we use these optimal threshold τ (referred to as τ^*) for the class definitions corresponding to each of our behavioral measures.

RESULTS ON PREDICTION

Prediction with Prenatal Period Data Alone

We begin by first examining the performance of the classifiers in identifying the two classes of mothers. Here, we train an SVM using only the mothers' behavioral measures during the prenatal period. The goal of this particular model is to predict postpartum changes before the birth of a child. We use six different performance metrics:

accuracy, precision and recall, F1, specificity and area under curve (AUC). We present the results of this prediction model in Table 3. We train one classifier to predict each of the behavioral measures (e.g. volume, replies, PA, NA etc.). For brevity, we report the mean performance per measure category—*engagement*, *emotion*, *ego-network* and *linguistic style*.

Measures	% Acc.	Prec.	Recall	F1	Spec.	AUC
engagement	71.70%	0.757	0.708	0.727	0.742	0.734
Mean accuracy of extreme-changing mothers (eng)						74.21%
emotion	71.21%	0.696	0.740	0.714	0.688	0.689
Mean accuracy of extreme-changing mothers (emo)						70.94%
ego-net.	68.63%	0.725	0.797	0.757	0.510	0.633
Mean accuracy of extreme-changing mothers (ego-net)						69.25%
style	72.90%	0.738	0.713	0.722	0.743	0.712
Mean accuracy of extreme-changing mothers (style)						74.30%
Mean	71.11%	0.729	0.740	0.730	0.671	0.692
Std. dev.	0.023	0.032	0.020	0.075	0.075	0.032
Mean accuracy for predicting extreme changers						71.46%

Table 3. Mean performance metrics (accuracy, precision, recall, F1, specificity, and AUC) of measure types, in predicting behavioral change classes of new mothers. Mean accuracy in predicting the class of extreme-changing mothers is also shown separately for each category. We only use data from the prenatal period.

The results in Table 3 indicate that, in our test set, the predictive models yield an average accuracy of more than 71% corresponding to the class of mothers showing extreme changes (C1). Good performance of this classifier is also evident from the receiver-operator characteristic (ROC) curves in Figure 2. These curves depict relationship of the true positive (extreme-changing mothers) and false positive (mothers with less extreme changes predicted as extreme-changing mothers) for different thresholds of the inferred probability of extreme-changing mothers required to admit these mothers into the extreme-changing class.

We find that the measures of linguistic styles, engagement, and emotion are more accurately predicted than the ego-network measures. Mean accuracies of classifying extreme-changing mothers per measure category are displayed in Table 3. We explore the differences in discriminatory power of measures in greater detail in the next subsection.

Prediction with an Optimal Training Window

Estimating Optimal Training Window k^*

So far, we have investigated predicting new mothers' future behavioral changes using data from the prenatal period alone. However, clinical literature on PPD marks the typical onset of PPD at about *one month following childbirth* [16]; hence a few days or weeks of training data in the early postnatal phase may contain valuable clues about future changes in behavior that can be additionally leveraged to boost prediction performance.

We now consider harnessing evidence for additional periods of up to three weeks (21 days) following childbirth for each mother. Together with the three month prenatal phase (91/92 days prenatal plus up to 21 days postnatal), we refer to this period as the *training window*. We note here that, due to the intrinsic differences in the characteristics of the different measures (e.g., see Figure 1 for illustrative examples), it is likely there will be an optimal training window for each measure, perhaps shorter than the 21 day period, which can make the best predictions of change in mothers. To obtain such an optimal training window k for a certain measure (referred to as k^*), we follow a similar optimization strategy using the training data and EM, as we used for inferring τ^* in the previous section. We observe a relatively narrow range of optimal k, corresponding to the best model fit for all measures. This narrow range is between 12-19 days following estimated day of childbirth.

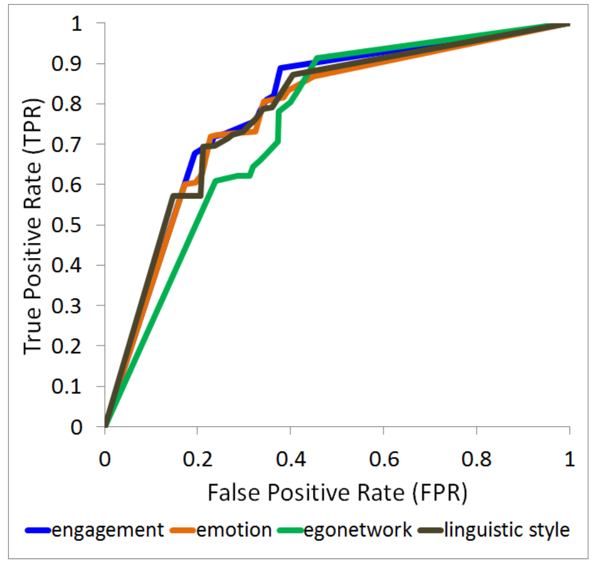

Figure 2. ROC curves for prediction using prenatal data. Aggregated trends over measure categories shown.

Overall Classification Performance

We now present classification performance over the two classes of mothers for the optimal training window size k^* for each of the measures. Table 4 reports the performance of the classifier along the different metrics. Overall, we observe good performance from our classifiers, with an average accuracy of more than 80% and F1 measure of 0.82 (standard deviations over all the performance metrics are shown at the bottom of Table 4). Note that this constitutes a considerable improvement over our previous prediction that used only the data over the prenatal period for training (more than 9% improvement in accuracy and F1 scores).

Specifically, we observe an accuracy of 81.62% in classifying the relatively small fraction of extreme-changers in the entire population of our dataset (mean size of class extreme-changing mothers is 16%). As before, we find that classifiers for changes in emotion, engagement, and linguistic style show the best performance. In essence, we conclude that two to three weeks into postpartum, the mothers show enough evidence in their behavioral and emotional changes, that it can be used along with the data from the prenatal period, to make more accurate predictions for remainder of the postpartum phase studied.

Measures	Acc.	Prec.	Recall	F1	Spec.	AUC
volume	*82.72%*	*0.819*	*0.893*	*0.853*	*0.745*	*0.835*
replies	*82.64%*	*0.796*	*0.884*	*0.836*	*0.735*	*0.854*
retweets	81.44%	0.807	0.823	0.812	0.828	0.838
links	77.46%	0.718	0.774	0.743	0.717	0.769
questions	80.06%	0.743	0.808	0.770	0.800	0.825
mean	80.86%	0.777	0.837	0.803	0.765	0.824
PA	80.99%	0.785	0.906	0.840	0.616	0.833
NA	*81.05%*	*0.824*	*0.789*	*0.805*	*0.813*	*0.820*
activation	*82.33%*	*0.784*	*0.908*	*0.840*	*0.666*	*0.842*
dominance	80.34%	0.799	0.891	0.842	0.677	0.820
mean	81.18%	0.798	0.874	0.832	0.693	0.829
followers	78.45%	0.747	0.918	0.822	0.629	0.829
followees	*80.32%*	*0.821*	*0.877*	*0.848*	*0.675*	*0.837*
mean	79.39%	0.784	0.898	0.835	0.652	0.833
Functional words	81.38%	0.795	0.880	0.835	0.731	0.822
1st pp.	*82.44%*	*0.811*	*0.918*	*0.860*	*0.686*	*0.819*
2nd pp.	*83.26%*	*0.869*	*0.887*	*0.876*	*0.733*	*0.832*
3rd pp.	*83.39%*	*0.819*	*0.879*	*0.845*	*0.720*	*0.835*
indefinite pronoun	78.46%	0.758	0.784	0.766	0.787	0.763
article	*82.20%*	*0.802*	*0.894*	*0.845*	*0.738*	*0.801*
verbs	80.53%	0.789	0.861	0.820	0.722	0.779
aux-verbs	80.59%	0.832	0.842	0.836	0.745	0.793
adverbs	81.38%	0.843	0.857	0.849	0.741	0.843
prepos.	80.88%	0.753	0.863	0.801	0.654	0.778
conjunct.	81.38%	0.789	0.900	0.837	0.724	0.809
negate	80.31%	0.786	0.886	0.831	0.705	0.792
quantifier	80.85%	0.844	0.850	0.845	0.740	0.822
swear	78.99%	0.767	0.883	0.819	0.682	0.835
tentative	80.83%	0.775	0.896	0.829	0.715	0.847
certain	79.40%	0.817	0.796	0.804	0.737	0.818
inhibition	*81.39%*	*0.788*	*0.891*	*0.835*	*0.729*	*0.806*
inclusive	80.06%	0.791	0.877	0.828	0.710	0.780
exclusive	78.90%	0.738	0.744	0.740	0.787	0.732
assent	79.80%	0.793	0.865	0.827	0.704	0.828
non-fluency	80.87%	0.777	0.886	0.825	0.728	0.813
filler	78.93%	0.816	0.714	0.761	0.842	0.843
mean	80.74%	0.798	0.857	0.823	0.730	0.809
overall mean	**80.54%**	**0.789**	**0.866**	**0.823**	**0.726**	**0.824**
Std. dev.	0.007	0.009	0.022	0.013	0.042	0.009
Accuracy of class of extreme-changing mothers					81.62%	

Table 4. Performance metrics (accuracy, precision, recall, F1, specificity and AUC) for predicting behavioral change classes of mothers, using optimal training window k^*.

Best Predicted Measures

We now discuss the performance of individual measures in more detail. In the *engagement* category, we find that volume and replies are best predicted (shown in italics), with higher accuracy, precision, recall, and F1 by 2-11%. One-way ANOVA reveals statistically significant differences in performance among the different engagement measures ($F(4,300)=2.54$; $p<.01$), while pairwise comparisons (Tukey range test) confirm that volume and replies show significant improvements in prediction compared to most other engagement measures. We note here that the volume of postings would signal how involved a mother is with her contacts, while replies would indicate the degree of social support that maybe available to a mother via her one-to-one interactions with other users. Given that both of these measures go down in the case of extreme-changing mothers (see Table 2), it appears that their trends during the training window are powerful predictors of future drops in these two key measures of Twitter-based social engagement.

Among the *emotion* measures, the best prediction is observed in the cases of NA and activation (margin of 2-20% improvement across different metrics; one-way ANOVA and corresponding pairwise Tukey range tests yields significant improvements over other emotion measures: $F(3,225)=3.25$; $p<.01$). We speculate that lower arousal may be based in exhaustion, anxiousness, and the general overwhelming routine of new motherhood. In order to obtain some qualitative observations on this conjecture, we examined several randomly selected posts shared by the mothers who we classified as showing extreme changes during postpartum. Some of the excerpts indicate notable negative emotional expression and concur with our above conjecture: "*Anxiety/panic attacks need to eff off!!!!!!!!!!!!!! I'm trying to lead a somewhat normal life with my baby!!!!*" (high NA); "*My first time being alone with my baby and I cant stop crying. What is wrong with me?*" (low activation).

For the *ego-network* measures, #followees (outlinks) is better predicted than #followers (inlinks). Outlinks could indicate how a mother is attempting to socialize or her tendency to consume external information and remain connected with others. In the light of the social isolation that some of the extreme-changing mothers appear to undergo, it appears that the shrinkage of ego-networks provides valuable cues in predicting their postpartum behavioral changes along this measure.

For the *linguistic style* measures, we observe high prediction performance for the three pronoun uses (first-, second- and third-person usages) and the use of articles, by a margin of 2-17% over other linguistic style metrics (statistically significant differences with the most number of other measures, based on one-way ANOVA: $F(21,1575)=2.97$; $p<.01$, followed by Tukey range test). Results in prior literature suggest that use of these styles provide information about how individuals respond to psychological triggers. For instance, we observe one of the extreme changing mothers in this category posting the following during postpartum: "*No lie I fuckin miss all socializing..... my daughter keeps me occupied and exhausted*." This excerpt involving high first-person pronoun usage shows high self-focus on the part of the

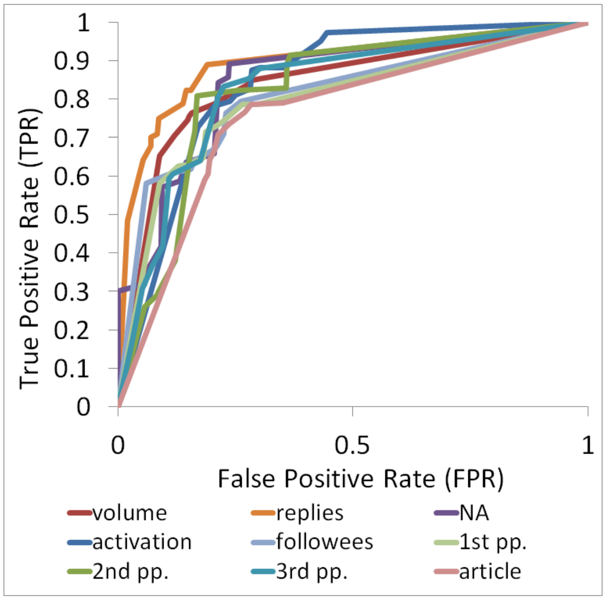

Figure 3. ROC curves for top predictable measures using composite features based classifier. Mean AUC values corresponding to these curves are shown in Table 4.

mother [7,30]. We conjecture that the general social and psychological distancing characterizing the circumstances of new motherhood is linked to such high attentional focus on oneself, and turns out to be a strong predictor of postpartum change for the extreme-changing mothers for these measures of linguistic style.

As a summary graphical statistic, Figure 3 presents the ROC curves corresponding to the best predicted measures discussed above. The curves demonstrate improvements in Figure 2 (upward trends towards left with growing area under the curve), with increases in the true positive rate (i.e., the extreme changers) versus false positives.

Broadly, the observation, that certain measure categories (emotion, NA, pronoun, article use) can be better predicted when demarcating mothers with extreme change, reveals interesting artifacts about postpartum behavioral signatures of these mothers. For instance, the behavioral trends of these particular measures bear resemblance to heightened levels of emotional pain or distress, up to even depression as indicated in prior literature [6,23]. In the following section, we discuss possibilities of leveraging these postpartum behavioral signatures to aid new mothers as well as the broader implications of our prediction methodology and outcomes in public health.

DISCUSSION

Theoretical Implications

Through our experimental findings on prediction, we have demonstrated how sets of behavioral markers manifested in social media during the prenatal period and in early postpartum periods, can be harnessed to predict future changes. We are able to identify with good accuracy (71%), a cohort of mothers who later show extreme changes (the median class size of extreme-changing mothers was 16% across all measures), by leveraging their behavioral and emotional signals exhibited in social media before the birth of the child. The accuracy of the predictive models rises (~80%) when the evidential horizon is extended to the early postnatal period.

The ability to predict significant changes in behavior and mood postpartum has broad implications. The postpartum behavioral markers exhibited by the subset of mothers who we identified as showing extreme changes, resonate with the feelings of hopelessness, dejection, and depressive tendencies seen in postpartum depression [16]. Further, the 16% of extreme-changing mothers aligns with reported rates of postpartum depression in the United States [19,20]. We conjecture that the predictive models may be able to reveal valuable clues about forthcoming shifts in mood ahead of time or soon after childbirth; thereby providing sufficient time for mothers and caregivers to take appropriate and valuable action.

In follow-on research, we seek to collect ground truth on PPD in new mothers, wherein we can empirically validate the relationship between the measures we predict and blues and deeper depression experienced by some mothers. Along these lines, we believe that our methodology has implications for public health. The findings can potentially assist agencies, support groups, or the larger medical community with the measurement of PPD in large populations, providing a new lens on a traditionally underreported illness.

We note that the clinical literature [2] indicates that the risk factors of postpartum depression include history of depression or psychological disturbance during pregnancy, experience of stressful events during the past year, problems in the woman's relationship with the partner, and weakness in the available support system(s). Such observations may be identifiable via longitudinal records of social media postings made by mothers during the prenatal period or earlier and could be used as higher-level evidence in the training classifiers that predict postpartum behavioral changes in mothers.

Our result that including two to three weeks of data following childbirth improves our prediction significantly may be explained by findings in the literature. Studies have shown that PPD typically arises after the fading away of the less severe baby blues, which usually appears in the week or two following childbirth [11]. Other known PPD triggers, such as difficult infant temperament and childcare stress [3] only manifest after the birth of the baby. Thus, mothers who consistently show extreme changes following childbirth and who suffer these early postnatal stressors may be at highest risk of PPD. However, as mentioned above, concrete validation of such potential risks will require ground truth data on PPD from a sufficiently large population of new mothers.

Directions in Design and Deployment

We have demonstrated the feasibility of predicting the future appearance of extreme changes in the behavior of new mothers, up to three months into the future with relatively high accuracy, using only prenatal behavioral data. The predicted behavioral changes in the measures we study can enable adjuvant diagnosis of postnatal disorders, complementary to survey based approaches (e.g., Edinburgh Postnatal Depression Scale [19]), helping promote wellness of women postpartum. In another direction, we believe that it is possible to develop software applications and services that serve as early warning systems, one day providing pregnant women and new mothers with personalized information on their risk of encountering significant changes in activity and mood. Such systems could be designed as privacy preserving applications that are deployed by and for individuals.

Predictive models running in a personalized service or within a smartphone application may provide value over traditional methods used for keeping track of new mothers' health and vitality. Traditionally, new mothers are asked to fill out surveys or are interviewed about PPD-like symptoms that they may be experiencing following childbirth, as well as the duration and severity of the symptoms. A tool motivated by this research could automate this process via monitoring behavioral trends of new mothers, and could be part of broader diary-centric systems that capture a self-narrative about postpartum life. Such an application is loosely akin to systems for Web-based depression treatment, which have been shown to be engaging and effective [9]. Per our study and results, applications could work privately in making *predictions* about potential risks of future changes as well, considering such measures in social media activity, as volume, replies, NA, activation, etc. The application might, for example, assign a "PPD risk score" to mothers based on predictions made about forthcoming extreme changes in their behavior and mood. In operation, if inferred likelihoods of forthcoming extreme changes surpass a threshold, mothers could be warned or engaged, and information might be provided about professional assistance and/or such guidance to mothers as noting the value of social and emotional support from friends and family.

Ethics and Privacy

Analysis of publicly shared data that could reveal existing or future mental health challenges raises concerns regarding the preservation of personal data privacy, as well as more general ethical considerations with pursuing research in this realm. On the research itself, we note that our studies have leveraged publicly available data (public expression by mothers on Twitter) with no personally identifiable information used in the analysis. As we discussed earlier, the privacy of mothers can be honored with applications that restrict the sharing of such information to the user herself and optionally to a trained medical practitioner or support group.

Nevertheless, we note that this type of research, and also results on the kinds of inferences that can be made from publicly available data pose interesting questions for individuals and for society more broadly. We have demonstrated that it is possible to make inferences from publicly available feeds about future psychological states that people may not wish to share with others. The predictions we make are similar to predictions about people made by systems in common use, including recommender systems that make inferences, e.g., about the titles of books that online users may wish to purchase given their history of purchases, search engines which guess the intentions of people performing online searches, given past behavior and terms input in online searches, and predictions made by online services about the likelihood that a user will click on a particular advertisement. Although the methods may be based in a similar mathematics of statistics and large-scale data analysis about people and their online activities, predictions about future changes in psychological wellbeing may be viewed by many as qualitatively different. People may be uncomfortable with the possibility that third parties might have the ability to predict future psychological states, especially when relatively accurate predictions can be made about future illness and disability. We believe it is important to bring the possibilities to the fore, so as to leverage the benefits of these methods and ideas to enhance the quality of life for people, as well as to stimulate discussion and awareness of potential concerns that need to be addressed at the individual and societal levels.

CONCLUSION

We have found that online activities, including the nature and changes of peoples' social networks, statistics of engagement, and the expression of thoughts and emotions in social media can be harnessed to make predictions about future behaviors and moods. We introduced a methodology for constructing and evaluating predictive models that forecast forthcoming changes in the behavior and mood of new mothers postpartum. We explored attributes of new mothers as manifested in their interactions on Twitter before and shortly after childbirth, including those of engagement, emotion, ego-network, and linguistic styles. We found that we could identify a group of new mothers who are likely to exhibit extreme changes in their behavior during the three-month period following childbirth. We constructed predictive models with the ability to predict with 71% accuracy which mothers will show extreme changes postpartum, using only prenatal observations. With adding consideration of additional short periods of postnatal data, the predictive accuracy rises to 80%.

We believe that such predictive models could be valuable tools in public health and could be used one day in the design and deployment of new kinds of early warning systems that could bring people timely information and assistance. At the same time, we foresee that our results will stimulate useful discussions about privacy and ethics with regard to the feasibility of forecasting forthcoming mental or physical illness from information that people share publicly. We can predict future behavior and mood with well-characterized confidence. However, people may be uncomfortable with others performing and sharing these predictions, even if the inferences are based solely on data that they have shared openly with the public.

REFERENCES

1. Bagehahl-Strindlund, M. & Borjesson, K.M. (1998). Postnatal depression: A hidden illness. *Acta Psychiatrica Scandinavica, 98*, 272-275.
2. Beck, C.T. (1998). A checklist to identify women at risk for developing postpartum depression. *Journal of Obstetric, Gynecologic, & Neonatal Nursing, 27*, 39-46.
3. Beck, C.T. (2001). Predictors of Postpartum Depression: An Update. *Nursing Research*, 50 (5), 275-285.
4. Bradley, M.M., & Lang, P.J. (1999). Affective norms for English words (ANEW). Gainesville, FL. *The NIMH Center for the Study of Emotion and Attention.*
5. Brubaker, J. R., Kivran-Swaine, F., Taber, L., and Hayes, G. R. (2012). Grief-Stricken in a Crowd: The language of bereavement and distress in social media. In *Proc. ICWSM 2012*.
6. Bucci W, Freedman N. (1981). The language of depression. *Bull. Menninger Clin.* 45:334–58.
7. Chung, C.K., & Pennebaker, J.W. (2007). The psychological functions of function words. In K. Fielder (Ed.), *Social communication* (pp. 343-359).
8. De Choudhury, M., Counts, S., and Horvitz, E. (2013). Major Life Changes and Behavioral Markers in Social Media: Case of Childbirth. In *Proc. CSCW 2013*.
9. Doherty, G., Coyle, D., & Sharry, J. (2012). Engagement with Online Mental Health Interventions: An Exploratory Clinical Study of a Treatment for Depression. In *Proc. CHI'12*.
10. Duda, Richard O., Hart, Peter E., & Stork, David G. (2000). *Pattern Classification*. 2nd Edition, Wiley.
11. Fleming, A. S., Klein, E. and Corter, C. (1992). The Effects of a Social Support Group on Depression, Maternal Attitudes and Behavior in New Mothers. *Journal of Child Psych. & Psychiatry*, 33: 685–698.
12. Golder, S. A., & Macy, M. W. (2011). Diurnal and Seasonal Mood Vary with Work, Sleep and Daylength Across Diverse Cultures. *Science*. 30 Sep 2011.
13. Kapoor, A., Horvitz, E. & Basu, S. (2007). Selective Supervision: Guiding Supervised Learning with Decision-Theoretic Active Learning. In *Proc. IJCAI*.
14. Kramer, A. (2010). An Unobtrusive Behavioral Model of "Gross National Happiness". In *Proc. CHI 2010*.
15. MacLennan, A., Wilson, D., & Taylor, A. (1996). The self-reported prevalence of postnatal depression in Australia and New Zealand. *Journal of Obstetrics and Gynecology, 36*, 313.
16. Miller, Laura J. (2002). Postpartum Depression. *J. of American Med. Association* (JAMA) 287 (6): 762–765.
17. Neuman, Y., Cohen, Y., Assaf, D., & Kedma, G. (2012). Proactive screening for depression through metaphorical and automatic text analysis. (2012). *Artificial Intelligence in Medicine*, 56 (1), 19-25.
18. Nielson, Forman D., Videbech, P., Hedegaard, M., Dalby Slavig, J. & Secher, N.J. (2000). Postnatal depression: identification of women at risk. *British Journal of Obstetrics and Gynaecology* (BJOG) 107 (10): 1210–1217.
19. O'Hara, M.W. (1995). Postpartum Depression: Causes and Consequences. *New York: Springer-Verlag*.
20. O'Hara, M.W., & Swain, A.M. (1996). Rates and risks of postpartum depression – a meta-analysis. *Intl. Review of Psychiatry, 8 (1)*, 37-54.
21. Oxman T.E., Rosenberg S.D., & Tucker G.J. (1982). The language of paranoia. *American J. Psychiatry* 139:275–82.
22. Paul, M., J., & Dredze, M. (2011). You are What You Tweet: Analyzing Twitter for Public Health. In *Proc. ICWSM '11*.
23. Pennebaker, J.W., Mehl, M.R., and Niederhoffer, K.G. (2002). Pyschological aspects of natural language use: Our words, ourselves. *Annual Review of Psychology* 54: 547-477.
24. Ramsay, R. (1993). Postnatal depression. *Lancet, 341, 1358*.
25. Sadilek, A., Kautz, H., & Silenzio, V. (2012). Modeling Spread of Disease from Social Interactions. In *Proc. ICSWM '11*.
26. Scott KD, Klaus PH, Klaus MH.(1999). The obstetrical and postpartum benefits of continuous support during childbirth. *J Womens Health Gend Based Med*. 8(10):1257-64.
27. Spera SP, Buhrfeind ED, Pennebaker JW. (1994). Expressive writing and coping with job loss. *Acad. Manag.* J. 37:722–33.
28. Steinfeld, C., Ellison, N., Lampe, C. (2008). Social capital, self-esteem, and use of online social network sites: A longitudinal study. *J. of Applied Developmental Psychology, 29*, 434-445.
29. Tarkka, M.-T. & Paunonen, M. (1996). Social support and its impact on mothers'experiences of childbirth. *Journal of Advanced Nursing*, 23: 70–75.
30. Weintraub W. (1981). Verbal Behavior: Adaptation and Psychopathology. New York: Springer.