Modeling Temporal Progression of Emotional Status in Mental Health Forum: a Recurrent Neural Net Approach

Kishaloy Halder, Lahari Poddar, Min-Yen Kan

School of Computing
National University of Singapore
{kishaloy, lahari, kanmy}@comp.nus.edu.sg

Abstract

Patients turn to Online Health Communities not only for information on specific conditions but also for emotional support. Previous research has indicated that the progression of emotional status can be studied through the linguistic patterns of an individual's posts. We analyze a real-world dataset from the *Mental Health* section of healthboards.com. Estimated from the word usages in their posts, we find that the emotional progress across patients vary widely.

We study the problem of predicting a patient's emotional status in the future from her past posts and we propose a Recurrent Neural Network (RNN) based architecture to address it. We find that the future emotional status can be predicted with reasonable accuracy given her historical posts and participation features. Our evaluation results demonstrate the efficacy of our proposed architecture, by outperforming state-of-the-art approaches with over 0.13 reduction in Mean Absolute Error.

1 Introduction

Online mental health forums offer a medium of peer support where individuals who have endured the adversity of mental illness can share their own experiences and offer help to others facing similar conditions. While each individual goes through life, their outlook and emotional state continue to evolve over time.

Understanding the complex patterns in which an individual interacts with an online community can help us understand his or her emotional state. Our hypothesis is that individuals' online forum participation can signal that state. Previous research on social media have established the relation between an individual's psychological state and her linguistic and conversational patterns (Tamersoy et al., 2015; Paul and Dredze, 2011; De Choudhury et al., 2013a). This motivates us to study user participations in online medical communities through a linguistic lens.

We propose a framework for tracking linguistic changes of a user over time for understanding her emotional status. We use our framework to analyze user participation on a large dataset collected from the mental health forums of the website healthboards.com[1]. These forums are dedicated for users discussing mental health issues ranging from anxiety, depression, stress, to even self-injury recovery. We choose this community since it is one of the largest online mental health forums, discussing a wide range of mental health issues. Additionally it has highly active members by not only their number of posts but also by longer periods of time for which they have been participating in the forum.

Models of time-varying user preferences in the recommendation domain (Matsubara et al., 2012; Koren, 2009) generally assume that users evolve according to a 'global clock', whereas patients participating in health forums progress according to his or her own personal timeline. By observing the word usage patterns of users in the site over time, we find that there exist different classes of users. While some users go through an *improvement* over time, lessening their use of negative words in their subsequent posts, some users move on a *deteriorating* slope where increased negative emotions can be observed in their posts. Decreased social interaction and increased negativity could be early indicators of depression, which

---

[1] www.healthboards.com/boards/
mental-health-board

claims the lives of $15 - 20\%$ of its patients (Sadeque et al., 2016). Hence it will be immensely beneficial to detect such users early, to be able to prevent unfortunate life-critical situations.

We make the key observation that people who *improve* over time tend to participate more in the community for the purpose of helping others (by replying to others' posts), than seeking help for themselves (by initiating threads). This indicates a belief in social support system and is reflected through increasing positivity in their posts. On the other hand, one of the major symptoms of depression is withdrawal from social interactions. Users with decreasing levels of forum participation, indicated by the increasing gap between their consecutive posts, tend to have increased negativity in their future posts.

Building on these observations, we show that a user's patterns of participation can be predictive of her emotions in the future posts. Inspired by our empirical analysis, we design features to capture the interaction styles of a user along with the textual contents of her posts. We use these heterogeneous features in a neural architecture to build a time series predictor model.

In recent years, recurrent neural networks (RNN) have achieved remarkable success in a range of sequence modeling tasks (Lipton et al., 2015; Kuremoto et al., 2014; Qiu et al., 2014). Inspired by the success of recurrent neural networks with pre-trained word embeddings for text modeling, we use a stack of RNN layers for encoding the textual content of a post. Given the encoded textual features along with the other participation features of a series of user posts, we employ another set of RNN layers to model the temporal progression of her emotional status. We find that by using a small number of consecutive posts, we can predict the emotional status of the next post with reasonable accuracy.

The main contributions of the paper can be summarized as:

- A systematic investigation of the temporal progression of emotional status across users from a real-world large dataset crawled from an online mental health forum. We identify three different classes of users according to their emotional progress over time.
- Identification of several forum participation and textual features indicative of users' temporal progression of emotional status.

- A proposed recurrent neural network based architecture that uses the identified features to predict the future emotional status of a user.
- A comparative study of the efficacy of our proposed architecture against state-of-the-art methods, and a complementary analysis on sensitivity of the prediction accuracy with respect to history length and variants of the architecture.

To the best of our knowledge, ours is the first work towards modeling the temporal progression of emotional status in online health forums.

2 Related Work

We start with a discussion of research efforts in understanding online textual contents related to mental health issues posted in social media as well dedicated health forums. Then we discuss works on time series forecasting which are relevant for temporal modeling of emotional status.

Detecting emotional crisis from social media outlets (e.g., Twitter) has gained significant attention in recent years (De Choudhury et al., 2013b; Coppersmith et al., 2014; De Choudhury et al., 2013a). They investigate the use of several linguistic features (choice of negative words in tweet, increased medicinal words), as well as other social features (e.g., egonetwork) to accomplish the task. However such social features are often not available in case of online health forums. In the absence of explicit signals by the users (e.g., 'mood'), the textual features can be indicative of one's emotional status.

There have been efforts from the intersection of biomedical, and NLP community to understand and analyze the textual contents users post in online health forums (Rey-Villamizar et al., 2016; Gkotsis et al., 2016; Paul and Dredze, 2011; Sadeque et al., 2016). After studying the patient community of dailystrength.org, Rey-Villamizar et al. found that on an average, the anxiety levels of patients in the community lower over time (Rey-Villamizar et al., 2016). Although they spot a global trend at the community level, there is a definite need to model the dynamics of users' emotional status over time. Sadeque et al. consider a user's linguistic and timeline features to predict whether a user will withdraw from the forum completely (Sadeque et al., 2016). In con-

trast, we are interested in modeling the temporal progression of users' emotional status.

Traditionally for time series prediction deterministic algorithms e.g., k-nearest neighbor (Wei and Keogh, 2006), ARIMA models (Hillmer and Tiao, 1982) have been used in different domains such as stock price forecasting (Pai and Lin, 2005), weather prediction (Cadenas et al., 2016) etc. Machine Learning based approaches have also been used in the literature for temporal modeling tasks in online communities (Matsubara et al., 2012; Danescu-Niculescu-Mizil et al., 2013; Cheng et al., 2015). Recently deep neural networks have shown significant progress due their capability of modeling complex sequential patterns (Ahmed et al., 2010; Lipton et al., 2015; Kuremoto et al., 2014; Qiu et al., 2014).

We propose an architecture using neural networks for modeling the temporal progression of a user using both textual and forum participation features. We believe ours is the first work to use RNNs on online health forum data and demonstrate its effectiveness over traditional machine learning models.

3 Analysis of Mental Health Forum

Online health forums provide a common platform for patients to interact with others suffering from similar diseases. Health forum websites provide a variety of functionalities. Apart from conventional discussion forum, some websites offer social media style features – e.g., "friend", "follow", virtual "hug". Although these could be indicative of a user's emotional status, in this work we focus on the most common setting: the discussion forum[2].

3.1 Dataset Description

We collected data from the *Mental Health* section of healthboards.com, a long running support group website. It comprises of individual forums for mental conditions (24 in total e.g., *Addiction & Recovery, Anger Management, Anxiety, Depression, Hypochondria, Self-injury Recovery,* and *Stress*). The website grants users three forms of participation:

• **Starting a thread:** typically contains a question about her own health.

• **Replying to own thread:** acknowledging others' advice or providing additional context to the

---

Number of posts	29,708
Number of users	1364
Average number of posts per user	21.7
Average number of words per post	140
Average life span of a user	528 days
90 percentile life span of a user	1515 days
Number of posts initiating a thread	4456
Number of posts Replying to own thread	4159
Number of posts Replying to others' thread	21,093

Table 1: Statistics of our Mental Health Discussion Forum Dataset.

original question.

• **Replying to others' thread:** providing suggestions in others' threads.

Since the objective of this work is to study the progression of emotional status over time, we have selected users who have spent at least 30 days and have posted more than 5 times in any of the above categories (statistics shown in Table 1).

3.2 Capturing Emotional Status

The emotional state a user is going through is manifested by her choice of words in her posts (Park et al., 2012; De Choudhury et al., 2013b; Rey-Villamizar et al., 2016). Coppersmith et al. show that standard polarity lexicons e.g., LIWC[3] can be reliably used to identify emotional crisis in the user posts (Coppersmith et al., 2014). Inspired from their feature design, we define a metric to capture the emotional status of a user from the word usage in her posts. We note that although some websites (e.g., dailystrength.org) let users report their "mood" (e.g., *horrible, okay, good*) along with the posts which could possibly be used as an absolute metric — it is not commonly available in most of the health forum websites. Instead, we rely on a simple metric derived from the polarity word usages in the posts. We thus define the *Negative eMotion Index* (NMI) of a post as:

$$\text{NMI} = \frac{\#\text{negative words} - \#\text{positive words}}{\#\text{total words}}$$

We obtain the list of stemmed polarity words from the MPQA subjectivity lexicon[4]. Note that the NMI score of a post is in the range $\{-1, 1\}$. A high NMI score denotes more emotional crisis in a post and vice versa. Apart from the individual words, we also handle simple negation structures: we account for occurrences like "not feeling well", "not ok" by reversing the polarity of
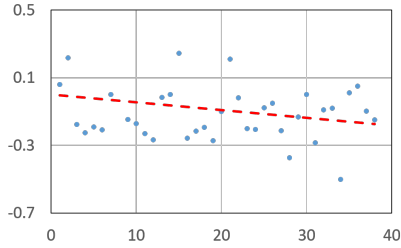

Figure 1: Temporal progression of NMI for a sample user (suffering from depression) from 38 posts made over a period of 90 days. The dashed red line denotes the trend according to linear regression model.

a *positive* word in cases where it is preceded by "not" or "no" (with distance$\leq$ 2). Since writing "n't" instead of "not" is a common practice (e.g., "haven't", "aren't"), we replace them with "not" as a part of pre-processing.

3.3 Temporal Progression of Emotional Status

The NMI progression for a sample user is shown in Figure 1. The posts (in chronological order) are along X-axis and their NMI scores are plotted along Y-axis. The trending line (based on linear regression model), is shown in red. We introduce a metric called NMI differential over time denoted by NMI':

$$\text{NMI}' = \frac{\delta \text{NMI}}{\delta \text{t}}$$

where δNMI is the difference in NMI over time period δt. Note that the slope of the trending line is same as NMI'. This admits three possible NMI' trends:

$$\text{NMI}' \begin{cases} < 0 & => \text{NMI is reducing over time} \\ > 0 & => \text{NMI is increasing over time} \\ = 0 & => \text{NMI remains constant over time} \end{cases}$$

The case NMI' < 0 points to those patients who are *improving* with time; > 0 is for those who are *deteriorating*; otherwise it denotes those patients who are *stable*. We present the CDF of NMI' across all the patients in Figure 2.

We find that the patients are Normally distributed among the three classes. Considering a soft boundary of 0.03 for NMI', we find that around 31% are in *improving* (NMI' < -0.03) class, 49% belong to the *stable* ($-0.03 < $ NMI' < 0.03) class. Interestingly, 20% of all the users fall in the *deteriorating* class.
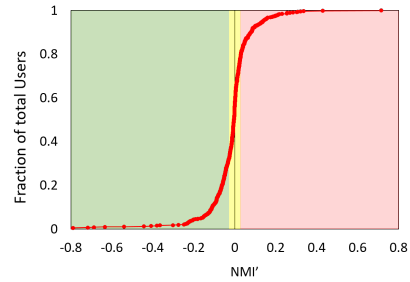

Figure 2: CDF of NMI' across all patients in Mental Health section of HealthBoards. 31% are *improving*, 20% are *deteriorating*, and 49% are *stable* with a soft threshold of NMI' $= 0.03$.

3.4 Prediction Task

The above study shows that the global trends observed on a community level do not reflect well on an individual basis. Hence we ask the following research question.

RQ: Given a user's history of forum participations, can we model the progression of her emotional status over time?

As we discussed in Section 2, this question is largely unanswered by the existing literature. To this end, we formally define a prediction task. The graphical representation of the task is shown in Figure 3. Given past k post details (text, and other participation metrics), the task is to predict the next NMI score. Note that we do not observe the post text that the user would be writing next, the task focuses on estimating the next NMI for her.

All the posts written by a user within a certain time period are combined into a single *post-block*. In this work, we set this time period to be 24 hours. This is done primarily since a user's emotional status is unlikely to change within a single day. Additionally, individual posts can be short and noisy (e.g. "thank you", "take care") so combining multiple posts of the same day will be a better reflection of a user's emotional health. For a user we consider her last k post-blocks in the forum and predict the NMI score of her next post-block.

4 Method

In this section we discuss our approach towards modeling the temporal progression of a users' emotional status. Our task falls in the guise of time series forecasting. In our case, we have heterogeneous features (e.g., post types, timing of posts) generated as artifact of user participation in the on-
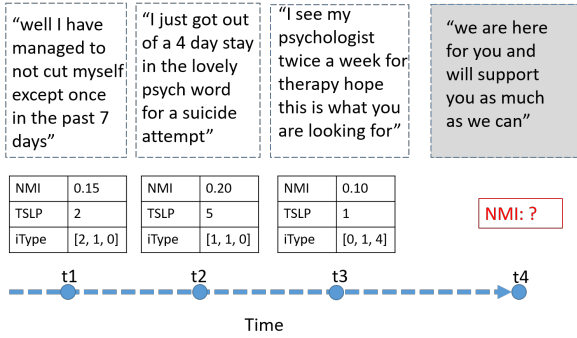
Figure 3: Graphical illustration of the prediction task. The task is to predict the next NMI score given past k posts. The shading on the text block denotes that it is not observed.
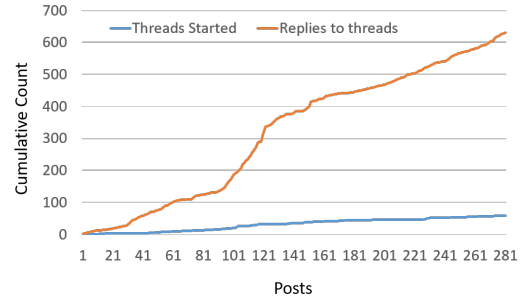

Figure 4: Temporal cumulative distribution of interaction types for a sample user in *improving* class. She keeps posting to others' threads instead of starting her own increasingly with time.

line platforms. To this end, we propose an RNN based architecture which not only takes the past NMI scores, but also incorporates other evidences seamlessly in the modeling process.

Our architecture consists of two components, namely, (1) text encoder and (2) time series encoder. The text encoder takes text of a single post-block as input and outputs a feature vector representation for it. We first encode the textual component of each post-block using the text encoder. Overall we build an ensemble style network to account for both textual and other numeric features since both these classes of features are heterogeneous in nature. One component of the network learns from the temporal sequence of feature vectors of text, while the other one from the numeric features. Both of these components consider sequence of feature vectors for the past k post-blocks in order to predict the NMI for the next to come.

In the following subsections we describe the numeric features and the two components in detail.

4.1 Numeric Features

For each post-block we consider the following numeric features.

Time Since Last Post (TSLP): The frequency with which a user engages in the forum can be indicative of her emotional health. Since people with depression often tend to withdraw from social contacts, the time gap between a user's posts can represent her diminishing social interactions (Sadeque et al., 2016). For each post-block of a user, we consider the time difference between the earliest post of the current block and the latest post of the previous block as a feature.

Interaction Type (iType): An individual user post can either be (i) initiating a thread or (ii) re-

plying to someone else's thread or (iii) replying to a self-initiated thread.

The type of interaction a user has on the forum can reflect her current role or purpose in the community. While some users seek answers to their own questions and troubles (by starting discussion threads), some users help other community members overcome theirs (by posting suggestions and advices on other's threads). The distribution of interaction type for a sample patient who has improved over time is shown in Figure 4. As we can see, with time she starts posting more on others' threads rather than starting her own. Similar trends could be observed for other patients as well whose emotional status have improved over time.

To encode this, for each post-block, we count the number of individual posts within the block that belong to the above three categories and use the counts as features.

NMI score: Apart from the participation and textual features, the past NMI scores could also be predictive of the future NMI score. Hence we use NMI scores of the post-blocks as features. Since there are multiple posts within a post-block, we take their mean NMI and consider it as the NMI score of the post-block.

For a post-block we concatenate the above mentioned numeric features to form a single numeric feature vector.

4.2 Text Encoder

For each post-block we first concatenate the raw texts of individual posts and use a text encoder to encode it into a feature vector. In the text encoder we first embed each word using an embedding layer, initialized with 50 dimensional Glove
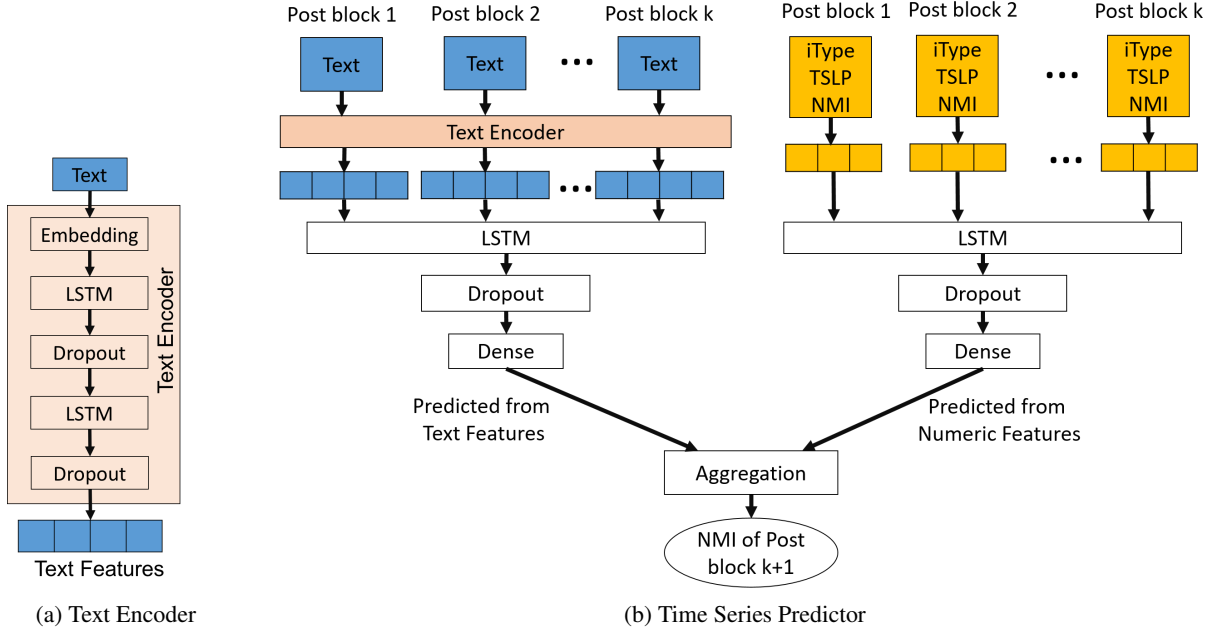
Figure 5: Illustration of model architecture. Each post-block consists of text and numeric features. The encoder for text is shown on the left side. The time series predictor, that combines both text and numeric features to predict NMI score of the next post, is shown on the right.

word embeddings [5]. The embeddings of the words are made trainable so as to reflect the domain and task dependent nature of the words. After embedding the word vectors, the sequence of words go through a stack of two LSTM layers, to encode the text into a vector. In our experiments we find that, using two stacked LSTM layers help in learning the latent representation of a text better than just a single layer. After each LSTM layer we add a Dropout layer so as to prevent overfitting.

Note that, there is only one text encoder component in the network. All the posts are encoded using the same text encoder.

4.3 Time Series Predictor

Now, given the feature vectors of the past k post-blocks we need to predict the NMI score of the next post-block. To tackle this task of time series prediction, we use a recurrent neural network architecture due to its superiority in handling short sequential data. There are two identical RNN components in our network for text, and numeric features respectively as shown in Figure 5b. The input to the RNN at each time-step i is the feature vector representation of the ith post-block – textual feature vector for one and numeric feature vector for the other. The output of the RNN at

the end of k time-steps yields the structural representation of the temporal emotional progression of the user. This is fed through a Dropout layer to prevent over-fitting. Finally a Dense layer is used to make a prediction from the output of the RNN. Given the predictions from both textual and numeric features, we aggregate (by taking mean) these two real-valued numbers to get the final NMI score of the $(k + 1)^{th}$ post-block.

Figure 5 shows an illustration of the architecture of our proposed of model. We also considered different variants of this architecture. The findings are discussed in Section 5.5.

5   Experiments

For our experiments, we consider a dataset from mental health forums of HealthBoards (as described in Section 3.1). In the following, we first describe how we setup the data for our prediction task. Later we describe the competitive baselines and compare our model with them in terms of the prediction accuracy. Finally we conclude with a discussion on the parameter sensitivity and other variants of our model.

5.1   Experimental Setup

Our objective is, given a history of k consecutive post-blocks of a user, predicting the NMI score of

---

her $(k+1)^{th}$ post-block. To this end, for each user we first sort her posts in chronological order. Then we combine all posts made within a 24-hour period by a user to form a single post-block. Thereafter we form tuples of length $(k + 1)$ from the sorted list of post-blocks using a sliding window method. For each such tuple of length $(k + 1)$, using the features of the first k post-blocks we predict the NMI score of the $(k + 1)^{th}$ post-block.

Consider a user with the sequence of post-blocks as shown in Table 2a. For history length $k = 3$, we reconstruct the sequence into temporal tuples as shown in Table 2b, where, given a tuple of past 3 posts ($P1, P2, P3$) we are predicting the NMI score of the next post ($P4$).

Post	NMI
P1	0.21
P2	0.24
P3	0.27
P4	0.25
P5	0.31

(a) Chronological Post-blocks of a user

Post 1	Post 2	Post 3	NMI
P1	P2	P3	0.25
P2	P3	P4	0.31

(b) Tuples of Post-blocks

Table 2: Temporal dataset construction from posts

We split our dataset in 80% tuples for training and 20% for testing and report five-fold cross validation results. We randomly selected 10% of our training data as the validation set.

To evaluate the performance of our NMI prediction task we employ the commonly used Mean Absolute Error (MAE) as our metric.

5.2 Parameter Settings

The parameters of our model include parameters for history length k, parameters for the text encoder and parameters for the time series encoder. We set the parameters using grid search on the validation set. We set the history length k to 5.

For the text encoder, the max length of a post-block text is set to 100. The embedding dimension for the words is set to 50 and is initialized with Glove embeddings. The sizes of the LSTM hidden layers are set to 64. The output of the LSTM layers go through dropout layers with 70% dropout rate to prevent over fitting.

For the time series encoder the sizes of both LSTM layers are set to 256. They are followed by dropout layers with 60% dropout rate. The predictions are made using a Dense layer with hyperbolic tangent as a non-linearity function.

Mean absolute error is used as loss function and Adam optimizer is used for optimization. Number of epochs is set to 20 but with an early stopping criteria depending on the validation accuracy. The analysis of the sensitivity of the parameters are discussed in Section 5.5.

5.3 Baselines

We compare our proposed model with traditional supervised regression models. We train the baseline models using the same history length and numeric participation features as our model and use Bag-of-Words (BOW) features to represent the textual content of a post. We consider the following models for comparison:

- **Linear Regression :** This is the basic ordinary least squares Linear Regression.

- **SVM Regression :** We experiment with support vector regression with both linear and non-linear RBF kernels.

- **Decision Tree Regression :** Learns a local linear regression approximating a sine curve. We set the max depth of the tree to be 5.

- **Random Forest Regression :** An ensemble learner that averages the predictions of a number of decision trees to improve accuracy and prevent over fitting. We use 100 decision trees to constitute the forest.

We use python's scikit-learn library[6] for the above models.

5.4 Prediction Results

We present a comparison of the results of the proposed method with the competing the state-of-the-art methods. Note that we have three sets of observed signals – text features, participation features, and NMI score. We collectively call the latter two as numeric features in this section. We perform an ablation study with numeric features, and text features across all the competing methods. The results are presented in Table 3.

We observe that our method outperforms other models comfortably. It achieves the best accuracy when it considers both set of features. Interestingly we find that the numeric feature set alone is quite predictive about the future, whereas if we only use the text features – the accuracy degrades. The traditional ML based baseline models yield

---

[6] http://scikit-learn.org/stable/index.html

Model	MAE		
	Numeric Features	Text Features	Numeric + Text Features
Linear Regression	0.2034	8.3553	3.4914
SVM (linear kernel)	0.2022	3.1513	0.2125
SVM (RBF kernel)	0.2724	0.2072	0.2071
Decision Tree	0.2106	0.2078	0.2106
Random Forest	0.2046	0.2032	0.2031
Our Model	0.0788	0.0802	**0.0781**

Table 3: Prediction results of different models

far less accurate results. Specifically we find that both linear regressor model and the SVM regressor with linear kernel model are unable to use the BOW features for the prediction task. Overall we can conclude that our architecture leveraging RNNs, is able to capture the temporal progression of emotional status with reasonable accuracy.

5.5 Parameter Sensitivity Analysis

We now study the sensitivity of our model by varying the history length from 1 to 5. Table 4 presents the accuracy scores obtained by our model with varying history lengths across different feature combinations.

Generally the performance improves with increasing history length, which is intuitive. We also observe that the numeric feature consistently appear to be more predictive compared to text feature alone. However we achieve best score with a combination of both while considering a history length of 5.

History Length	MAE		
	Numeric Features	Text Features	Numeric + Text Features
1	0.0813	0.0824	0.0810
2	0.0813	0.0818	0.0808
3	0.0807	0.0814	0.0803
4	0.0797	0.0806	0.0798
5	0.0788	0.0802	**0.0781**

Table 4: Effect of history length and features on the performance of our model.

Discussion on Model Architecture Variants: Apart from the architecture presented in Section 4, we experimented with a few other variants as mentioned below.

- For the RNN we experimented with both LSTM (Hochreiter and Schmidhuber, 1997) and GRU (Cho et al., 2014) and got similar results. Furthermore, we did not observe any significant improvement by replacing the RNN with a Bidirectional RNN (Schuster and Paliwal, 1997).

- We tried with larger embedding dimensions for words and larger neuron counts in the RNN layers but that led to over-fitting, possibly due to the dataset size.

- Instead of using a simple mean as the aggregation function, we experimented with using another Dense layer for predicting the final score. The Dense layer takes as input the concatenation of the outputs of the previous two Dense layers (from textual and numeric features) and outputs the final NMI score. This increased the number of parameters in the model but did not improve performance.

- Instead of using the textual and numeric features separately in the time series predictor, we also experimented with concatenating all the features into a single post feature vector. Thereafter the sequence of post feature vectors were fed into an RNN followed by a Dense layer to make the prediction. The performance of this model was slightly worse with MAE 0.0787.

6 Conclusion

In this paper we have presented a framework towards understanding temporal progression of users' emotional status in online mental health forums. We identify several forum participation features that are indicative of a user's temporal emotional progression. Our proposed neural network architecture uses textual content as well as participation features from a user's past posts to predict her future emotional status. Empirical evaluations on a large real world dataset of online mental health forum demonstrate the superiority of recurrent neural network for temporal modeling, as our model outperforms state-of-the-art approaches significantly.

In future, we would like to explore how our model can be extended to capture progression of other physical illnesses especially long term ones e.g., ALS, Multiple Sclerosis. Incorporating social features into the model could be another interesting direction. Social media and other online platforms will play an important role in providing healthcare in the 21st century (Dredze, 2012). With the constant influx of users seeking help from online health outlets, we believe our generic framework would be applicable to a wide spectrum of online mental health forums.

References

Nesreen K. Ahmed, Amir F. Atiya, Neamat El Gayar, and Hisham El-Shishiny. 2010. An empirical comparison of machine learning models for time series forecasting. *Econometric Reviews* 29(5-6):594–621.

Erasmo Cadenas, Wilfrido Rivera, Rafael Campos-Amezcua, and Christopher Heard. 2016. Wind speed prediction using a univariate arima model and a multivariate narx model. *Energies* 9(2):109.

Justin Cheng, Cristian Danescu-Niculescu-Mizil, and Jure Leskovec. 2015. Antisocial behavior in online discussion communities. In *ICWSM*.

Kyunghyun Cho, Bart Van Merriënboer, Dzmitry Bahdanau, and Yoshua Bengio. 2014. On the properties of neural machine translation: Encoder-decoder approaches. *arXiv preprint arXiv:1409.1259* .

Glen Coppersmith, Mark Dredze, and Craig Harman. 2014. Quantifying mental health signals in twitter. *ACL 2014* 51.

Cristian Danescu-Niculescu-Mizil, Robert West, Dan Jurafsky, Jure Leskovec, and Christopher Potts. 2013. No country for old members: User lifecycle and linguistic change in online communities. In *Proceedings of the 22nd international conference on World Wide Web*. ACM, pages 307–318.

Munmun De Choudhury, Scott Counts, and Eric Horvitz. 2013a. Social media as a measurement tool of depression in populations. In *Proceedings of the 5th Annual ACM Web Science Conference*. WebSci '13, pages 47–56.

Munmun De Choudhury, Michael Gamon, Scott Counts, and Eric Horvitz. 2013b. Predicting depression via social media. AAAI.

Mark Dredze. 2012. How social media will change public health. *IEEE Intelligent Systems* 27(4):81–84.

George Gkotsis, Anika Oellrich, Tim JP Hubbard, Richard JB Dobson, Maria Liakata, Sumithra Velupillai, and Rina Dutta. 2016. The language of mental health problems in social media. In *Third Computational Linguistics and Clinical Psychology Workshop (NAACL)*. pages 63–73.

Steven C Hillmer and George C Tiao. 1982. An arima-model-based approach to seasonal adjustment. *Journal of the American Statistical Association* 77(377):63–70.

Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural computation* 9(8):1735–1780.

Yehuda Koren. 2009. Collaborative filtering with temporal dynamics. In *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. KDD '09, pages 447–456.

Takashi Kuremoto, Shinsuke Kimura, Kunikazu Kobayashi, and Masanao Obayashi. 2014. Time series forecasting using a deep belief network with restricted boltzmann machines. *Neurocomputing* 137:47–56.

Zachary C Lipton, David C Kale, Charles Elkan, and Randall Wetzell. 2015. Learning to diagnose with lstm recurrent neural networks. *arXiv preprint arXiv:1511.03677* .

Yasuko Matsubara, Yasushi Sakurai, Christos Faloutsos, Tomoharu Iwata, and Masatoshi Yoshikawa. 2012. Fast mining and forecasting of complex time-stamped events. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, pages 271–279.

Ping-Feng Pai and Chih-Sheng Lin. 2005. A hybrid arima and support vector machines model in stock price forecasting. *Omega* 33(6):497–505.

Minsu Park, Chiyoung Cha, and Meeyoung Cha. 2012. Depressive moods of users portrayed in twitter. In *Proceedings of the ACM SIGKDD Workshop on healthcare informatics (HI-KDD)*. pages 1–8.

Michael J Paul and Mark Dredze. 2011. You are what you tweet: Analyzing twitter for public health. *Icwsm* 20:265–272.

X. Qiu, L. Zhang, Y. Ren, P. N. Suganthan, and G. Amaratunga. 2014. Ensemble deep learning for regression and time series forecasting. In *2014 IEEE Symposium on Computational Intelligence in Ensemble Learning (CIEL)*. pages 1–6.

Nicolas Rey-Villamizar, Prasha Shrestha, Farig Sadeque, Steven Bethard, Ted Pedersen, Arjun Mukherjee, and Thamar Solorio. 2016. Analysis of anxious word usage on online health forums. *EMNLP 2016* page 37.

Farig Sadeque, Ted Pedersen, Thamar Solorio, Prasha Shrestha, Nicolas Rey-Villamizar, and Steven Bethard. 2016. Why do they leave: Modeling participation in online depression forums. In *Proceedings of the 4th Workshop on Natural Language Processing and Social Media*. pages 14–19.

Mike Schuster and Kuldip K Paliwal. 1997. Bidirectional recurrent neural networks. *IEEE Transactions on Signal Processing* 45(11):2673–2681.

Acar Tamersoy, Munmun De Choudhury, and Duen Horng Chau. 2015. Characterizing smoking and drinking abstinence from social media. In *Proceedings of the 26th ACM Conference on Hypertext & Social Media*. ACM, pages 139–148.

Li Wei and Eamonn Keogh. 2006. Semi-supervised time series classification. In *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, pages 748–753.