Department Of Computer Science & Information Systems

CS F441: Selected Topics from Computer Science

(Cognitive Computing)

Abnormality detection in human emotional behavior through sentiment analysis

Submitted By

Vidhi Jain

2014A7PS113P

Birla Institute of Technology & Science

Pilani

Table of Contents

Introduction

Mental disorder has been shrouded as a stigma and disregarded as a secondary issue to physical health. It has become a major contributor to morbidity, disability and at times, fatality. The information to support medical diagnosis is available. Information about patients' emotions, thoughts, behavior and physical body sensations can reveal the onset of symptoms of certain psychiatric disorders.

The data generated through our daily interaction with technology has consistent patterns to identify symptoms in prodromal phase of degrading mental health. Most of the data is unstructured and it can be analysed through natural language processing, context-based hypothesis generation and evidence based learning for computational psychiatry. Supplemented with vast amount of information in medical journals and insights from qualified experts, this information can be used to maintain fine-grained record of health data and raise 'early alarm'.

This work proposes the analysis of deep learning strategies for text based sentiment analysis. Majority of a user's data generated at interfaces is textual data. The approach has focussed upon simplistic models in terms of computation and complexity to make it feasible for deployment on mobile devices for easy user interaction. In section 2, we highlight the background and motivation for the work. We describe the relevant literature survey in section 3 and detailed mathematical basis for two architectures in section 4. We discuss about modelling of the problem along with datasets description in section 5 and the observations obtained in different experiments in section 6. In section 7, we conclude by discussing the contributions of the work and possibilities of future extensions.

Background

According to the Ministry of Health and Family Welfare in India, there are just 3500 psychiatrists for a total of 1.3 billion people. The entire workforce comprising of clinical psychiatrists, psychologists, psychiatric social workers and psychiatric nurses is nearly 7000 while the actual requirement is around 54750 [5] as shown in Fig 1.

In U.S., one in every five adults (i.e. 43.8 million or 18.5\%) experiences mental illness in a given year as per National Institute of Mental health [1].

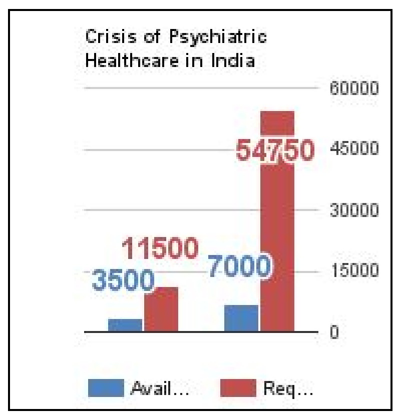


Figure 1: Crisis of Psychiatric Healthcare in India

International standard classification tools exist, namely Diagnostic and Statistical Manual of Mental Disorders (DSM-5) [6] by the American Psychiatric Association (APA) and International Statistical Classification of Diseases and Related Health Problems (ICD-10) [7] by the World Health Organization (WHO) for diagnosis of psychiatric disorders.

Literature Survey

Speech-based psychosis detection [2], emotion and disposition recognition and a variety of computational psychiatry approaches have been studied. Computational tools have been used to identify the first episode of psychotic disorder (FEP) [3].

Traditional text classification algorithms like naive Bayes, SVM (Bag of Words approach) and most TFIDF variants, required parameters to determine some closed form function. This heuristic approach to identify better parameters has often led to good results, but it is to be manually identified to optimize.

Feature extraction has been very important and challenging until deep learning methodologies came up. Deep learning allows for end-to-end training. Current techniques for text classification majorly involve Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN), particularly using Long Short Term Memory(LSTM). Though these models have been used for wide range of tasks in text understanding, the focus of this work is limited to sentiment analysis.

Web search logs are being explored to identify behavioral patterns which may indicate health conditions of the users. Cognitive healthcare systems like IBM Watson (http://www.ibm.com/watson/health/) are current revolutionary technologies which provide holistic view of a person's health along with improved patient experience.

A methodological framework for data collection and monitoring has been proposed for identifying symptoms of the individual [9] as shown in Fig 2.

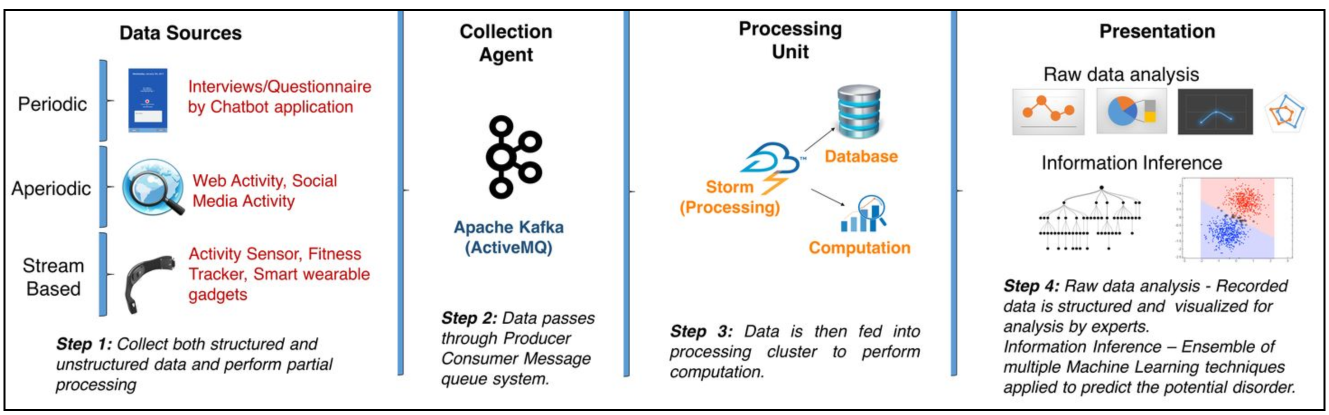


Figure 2: Methodological framework for symptom diagnosis from personal data sources

Among the three broad categories of data sources identified in Fig 2, the majority of the data is in plain textual format, especially from periodic and aperiodic sources. The objective of the current work is to explore deep learning strategies for analysis of sentiment in this textual data.

Mathematical Basis

This section discusses two mathematical models, that is, Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN), particularly using Long Short Term Memory(LSTM). Both of these deep learning strategies have been proved to be successful, with average baseline accuracy of around 80%.

Convolutional Neural Networks (CNN) 1D

The convolutional neural networks for text classification are modelled as follows. The first layers embeds words into low-dimensional vectors. The next layer performs convolutions over the embedded word vectors using multiple filter sizes. Next layers of max-pool reduce the long feature vector, Then dropout regularization (L2 norm) is added, and final layer classifies the result using a softmax layer. CNN model used as the baseline is a slight variant of that discussed in TFlearn examples, as shown in Fig 3.

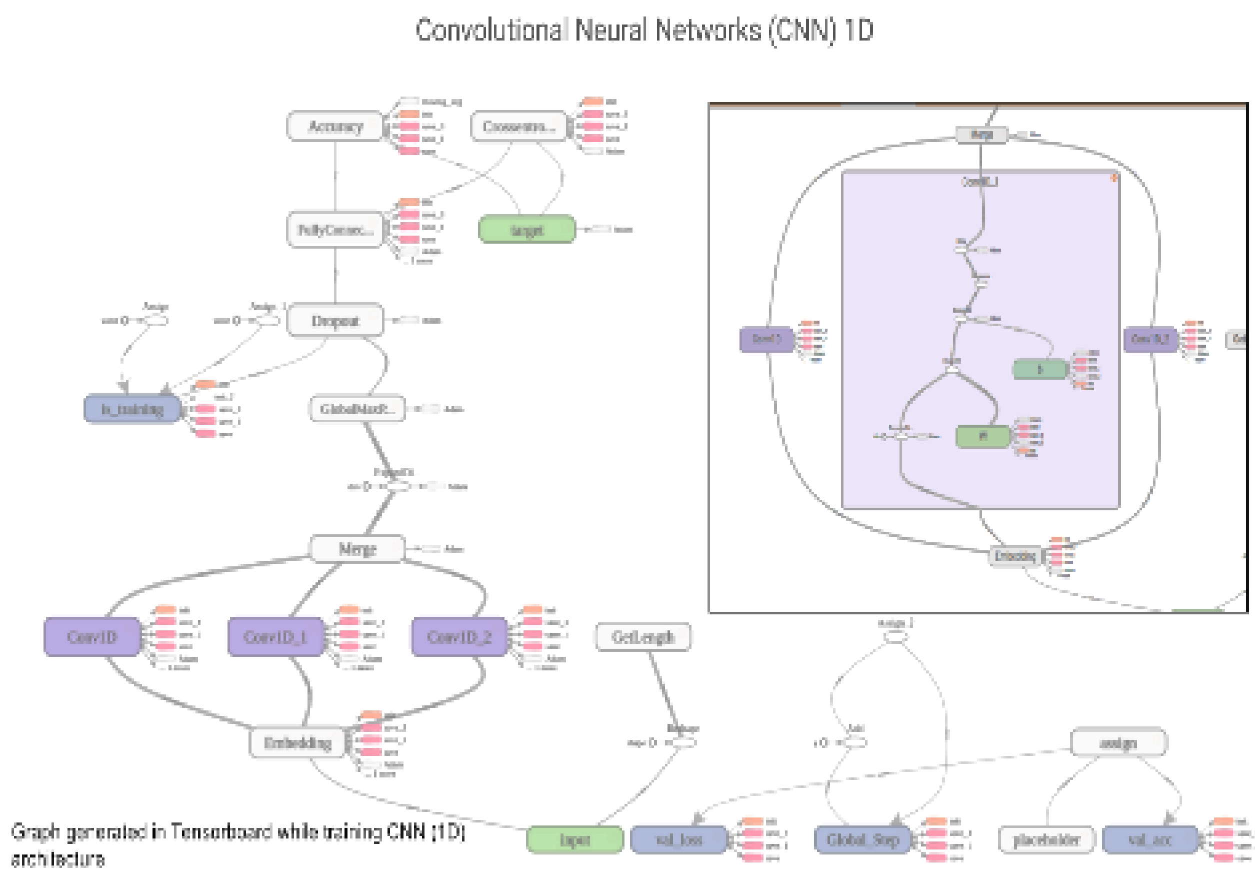


Figure 3: Graph of model architecture for baseline convolutional neural networks for sentiment classification

Representation of Word Vector

Let xi ∈ R k be the k-dimensional word vector corresponding to the i-th word in the sentence. The sentences can be made of length n by padding. This length n vector can be represented as

$$\mathbf{x}_{1:n} = \mathbf{x}_1 \oplus \mathbf{x}_2 \oplus \ldots \oplus \mathbf{x}_n, \qquad (1)$$

Here, xi:i+j refer to the concatenation of words. Later, we will also focus on character level modelling.

Operation

Convolutional operation involves a filter $\mathbf{w} \in \mathbb{R}^{hk}$, which is applied to window of h words. The filter yields us the features. Say, feature ci is computed by

$$c_i = f(\mathbf{w} \cdot \mathbf{x}_{i:i+h-1} + b).$$

where b is the bias term and f is an activation function, which is ReLU for this work.

Feature Map: Features computed for each window of word $\mathbf{c} = [c_1, c_2, \ldots, c_{n-h+1}]$,

Max Pooling: Selecting the maximum value feature from each feature map.

Softmax Layer: Classifier layer taking input as features (obtained from multiple filters) and giving output as probability distribution over labels.

Recurrent Neural Networks (RNN): Long Short Term Memory (LSTM)

A recurrent neural network can be thought of as multiple copies of the same network, each passing a message to a successor.

Long Short Term Memory networks are capable of learning long-term dependencies [11]. The working of an LSTM cell is illustrated in Fig 4.

LSTM has three gates whose corresponding equations are:

input_gate = tanh(dot(input_vector, W_input) + dot(prev_hidden, U_input) + b_input)

forget_gate = tanh(dot(input_vector, W_forget) + dot(prev_hidden, U_forget) + b_forget)

output_gate = tanh(dot(input_vector, W_output) + dot(prev_hidden, U_output) + b_output)

Here, W_xxx and b_xxx refer to the weight and bias of the gate xxx. The LSTM cell also has a candidate state corresponding to each cell which is used to calculate the new hidden state for the next adjacent cell.

candidate_state = tanh(dot(x, W_hidden) + dot(prev_hidden, U_hidden) + b_hidden)

memory_unit = prev_candidate_state * forget_gate + candidate_state * input_gate

new_hidden_state = tanh(memory_unit) * output_gate

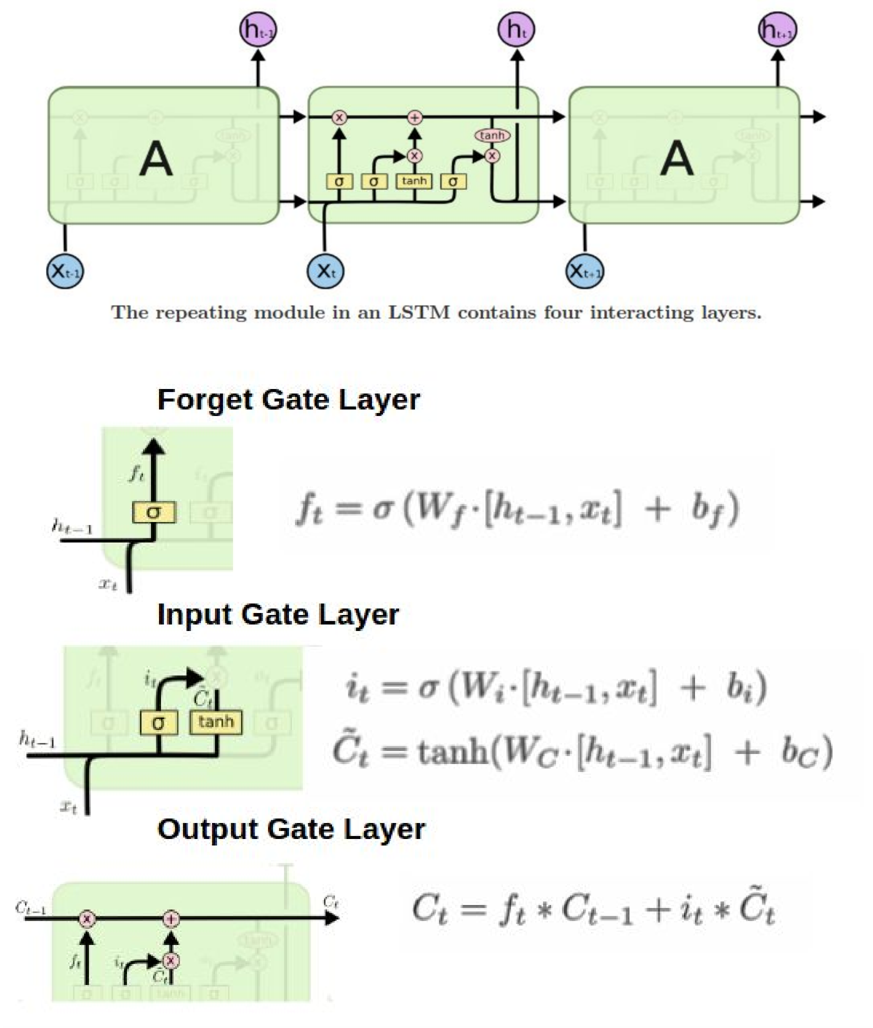


Figure 4: Structure of an LSTM and mathematical equations [11][12]

LSTM have been useful in capturing the dependency of word in a phrase and neglect the non-dependent connections [14]. Thus, LSTM are suitable capturing sentiment in text.

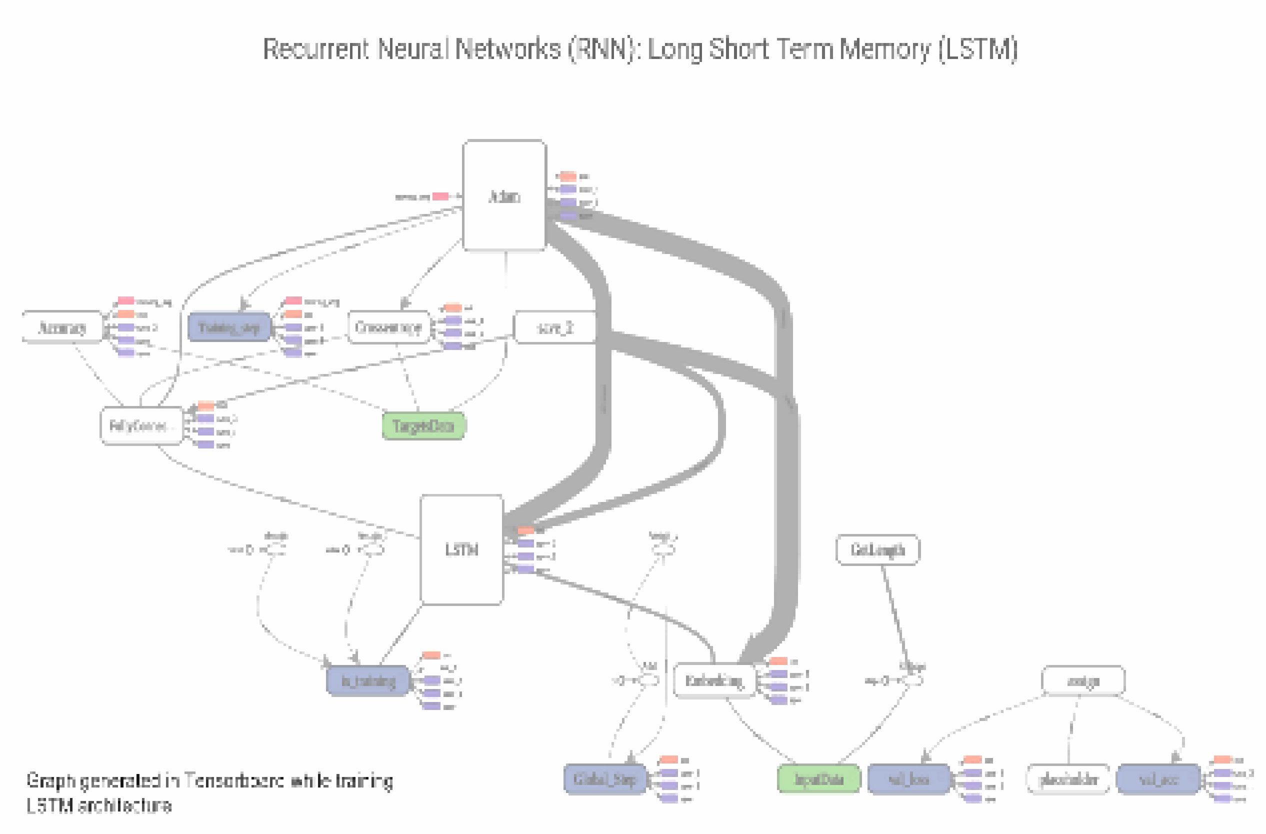


Figure 5: Graph showing the model architecture for LSTM used for sentiment classification

Modelling of the Problem

The patterns of behavior and emotion can be tracked by our mobile phones. Several variety of sources to track the emotional, behavioral and physical factors of a person can be integrated to this application to aggregate all kind of data about the user. The primary motive of the application is to collect data regularly from the user about his/her experiences of the day(s). The user shall describe about anything ranging from his daily routine to anything particular incident and how he/she have been feeling. The data shall also be collected as per individual's personalisation, to include other sources like social media, Internet activity, fitness tracking gadgets. These sources can be added and permitted for analysis as per user's discretion.

The objective of the current work is to identify the deep learning strategies for different datasets and evaluate their performance. The polar sentiment analysis recorded over time may yield patterns that indicate whether an individual has been exhibiting a prolonged negative sentiment. Fine grain sentiment analysis requires more categories for classification task, which significantly affects the accuracy. To improve the accuracy, we require sufficiently large dataset which is reliably labelled.

Dataset

The current work has utilized three major datasets:

IMDB Movie Reviews (MR)

2 labels, Credits:http://www.iro.umontreal.ca/~lisa/

Twitter Sentiment Dataset (Crowdflower)

11 labels, Credits: Sentiment Analysis: Emotion in Text (July 15, 2016 by CrowdFlower | Data Rows: 40000)

ISEAR

6 labels, Credits: Swiss Center for Affective Sciences http://www.affective-sciences.org/en/home/research/materials-and-online-research/research-material/

Observations

The following observations are recorded by us for the baseline architectures performance without fine tuning of hyperparameters. Detailed experiments have been conducted by Ma et al [13], and Yoon Kim [10].

	Movie Reviews (IMDB)	22500/2500	**ISEAR**	6900/766	**Twitter (Crowd-flower)**	36000/4000
	Accuracy	Loss	Accuracy	Loss	Accuracy	Loss
CNN	0.7972	0.53883	0.5561	1.3832	0.33775	1.99468
LSTM	0.8072	0.76406	-	-	-	-
Bi-LSTM	0.7947	0.95425	-	-	-	-
	2 labels		6 labels		11 labels	

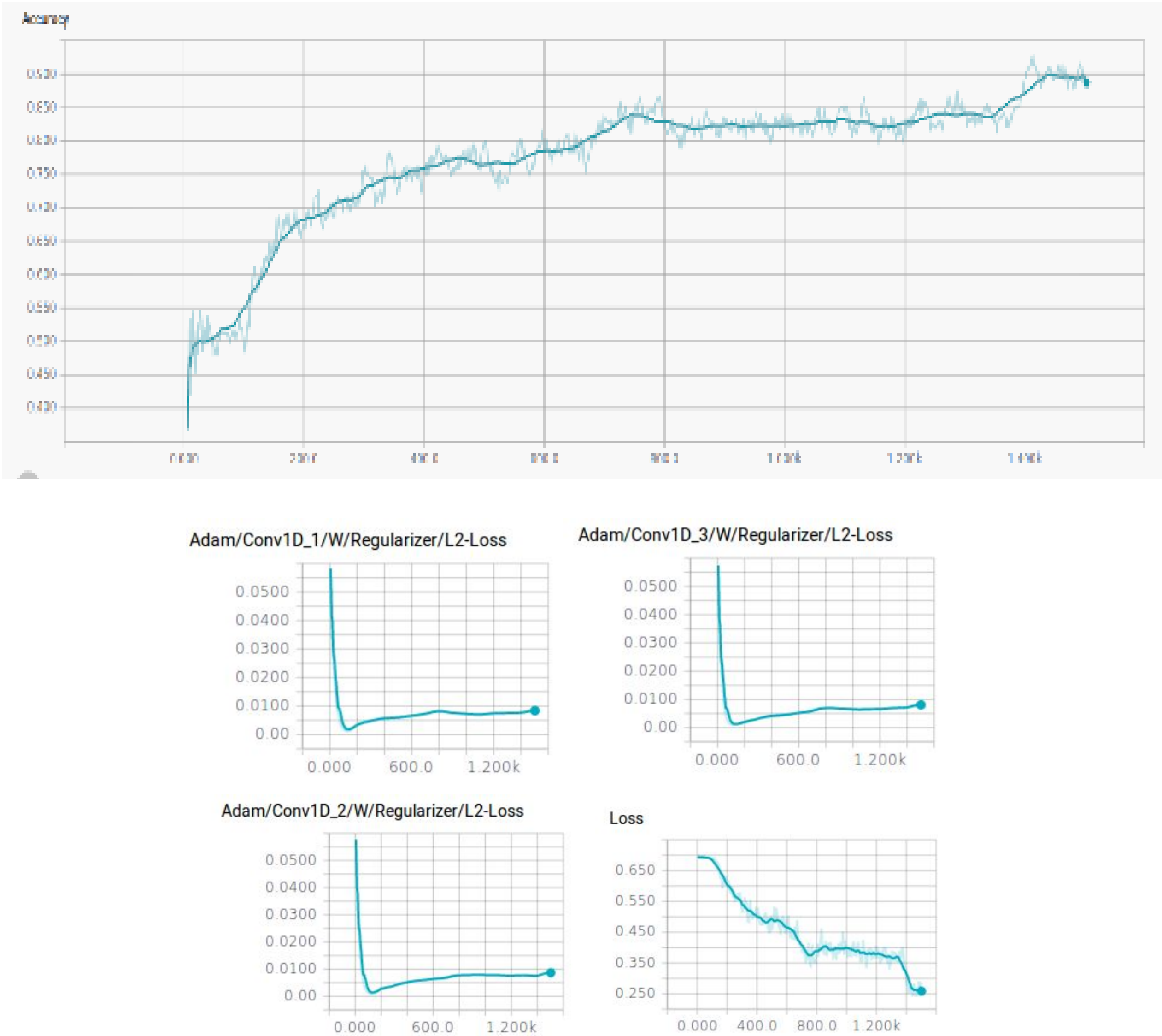


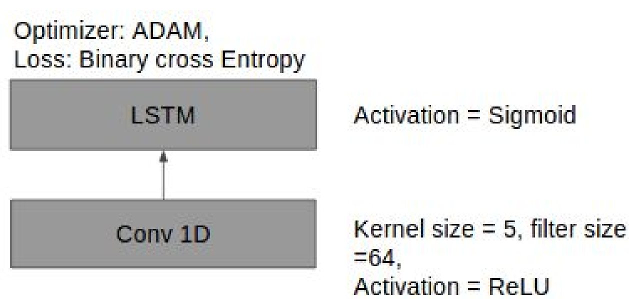
Figure 6: Graph showing accuracy, L-2 loss for each convolutional layer and overall loss function

The performance of CNN and LSTM are comparable in terms of accuracy. Both are different techniques but address the same purpose. The proposed model is a combination of CNN 1D layers with LSTM layer on top.

The proposed model obtained the following metrics on imdb dataset.

Validation Loss : 0.3761, Validation Accuracy : 0.8537

Figure 7: Basic hybrid model with CNN and RNN (LSTM)



Contributions

Comparison of existing state-of-the-art deep learning architectures for sentiment analysis on three independent datasets.

Evaluation of each model based on various metrics to achieve the objective of the problem

Analysis of combined baseline architectures of CNN and LSTMs to obtain higher accuracy for the same dataset.

Designing of prototype of mobile application as shown in Fig 7 for sentiment analysis of textual input from user and integration with social media, and browser history.

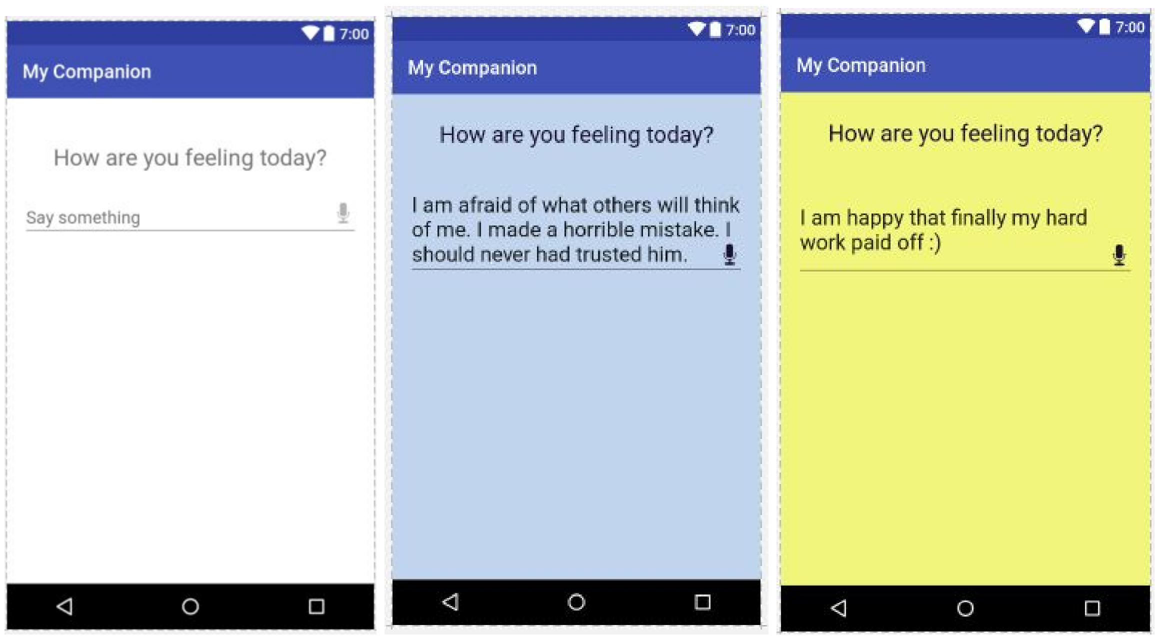


Figure 7: Prototype of mobile application for tracking a person's sentiment. As soon as the user expresses a strong emotional sentiment, the color of the screen is changed : Blue (Cool colours) to soothe and relax the person if he/she is in negative sentiment and Yellow (Warm colours) to stimulate the positivity expressed even further. [15]

Future Scope and Challenges

The following are a few areas for improved text sentiment classification and understanding :

- Character level encoding for text classification in CNN
- Dynamic memory networks with Attention mechanisms
- Use of GRUs (Gated Recurrent Units) in RNNs

Collection of context-specific (psychiatric/counsellor) dataset in a challenge. In order to apply advanced deep learning algorithms, the dataset must be large (about in ten thousands or million data points) and must be accurately labelled or verified. We need even larger as the categories of sentiment/emotion to be considered increases.

References

1. Center for Behavioral Health Statistics and Quality. 2015. Behavioral health trends in the United States: Results from the 2014 National Survey on Drug Use and Health Results from the 2014 National Survey on Drug Use and Health. (2015).

2. Gillinder Bedi, Facundo Carrillo, Guillermo Cecchi, Diego Fernández Slezak, Mariano Sigman, Natália B Mota, Sidarta Ribeiro, Daniel C Javitt, Mauro Copelli, and Cheryl M Corcoran. 2015. Automated analysis of free speech predicts psychosis onset in high-risk youths. Schizophrenia 1, May (2015), 1–7.

3. Gorrell et al. 2016. Identifying First Episodes of Psychosis in Psychiatric Patient Records using Machine Learning. Proceedings of the 15th Workshop on Biomedical Natural Language Processing (2016), 196–205

4. Rajamannar Ramasubbu, Matthew R G Brown, Filmeno Cortese, Ismael Gaxiola, Bradley Goodyear, Andrew J. Greenshaw, Serdar M. Dursun, and Russell Greiner. 2016. Accuracy of automated classification of major depressive disorder as a function of symptom severity. NeuroImage: Clinical 12 (2016), 320–331.

5. J. Sampathi. 2016. Shortage of psychiatrists hits treatment - Times of India. (2016), 1–2.

6. Diagnostic and Statistical Manual of Mental Disorders: DSM-5. Washington: American Psychiatric, 2014. Print.

7. International Statistical Classification of Diseases and Related Health Problems: ICD-10. Geneva: World Health Organization, 2009. Print.

8. John Paparrizos, Ryen W White, and Eric Horvitz. 2016. Screening for Pancreatic Adenocarcinoma Using Signals From Web Search Logs: Feasibility Study and Results. Journal of oncology practice / American Society of Clinical Oncology 12,8 (2016), 737–44.

9. Vidhi Jain and Prakhar Agarwal. 2017. Symptomatic Diagnosis and Prognosis of Psychiatric Disorders through Personal Gadgets. In Proceedings of CHI'17 Extended Abstracts, Denver, CO, USA, May 06-11, 2017, 6 pages. DOI: http://dx.doi.org/10.1145/3027063.3048417

10. Yoon Kim. 2014. Convolutional Neural Networks for Sentence Classification. Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP) (2014).

11. Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long Short-Term Memory. Neural Computation 9, 8 (1997), 1735–1780. DOI:http://dx.doi.org/10.1162/neco.1997.9.8.1735

12. Christopher Olah's blog on Understanding LSTMs, Available at : http://colah.github.io/posts/2015-08-Understanding-LSTMs/

13. Ma et al, Dependency-based Convolutional Neural Networks for Sentence Embedding http://www.aclweb.org/anthology/P15-2029

14. Cheng, Jianpeng, Li Dong, and Mirella Lapata. "Long Short-Term Memory-Networks for Machine Reading." Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, 2016. doi:10.18653/v1/d16-1053.

15. Psychological properties of colors, http://www.colour-affects.co.uk/psychological-properties-of-colours

All the codes used for analysis are slight variants of open-source examples of TFlearn and Keras. A repository containing the collection of source codes used along with the datasets is : https://github.com/vidhiJain/Deep-learning-for-sentiment-analysis