Prayas at EmoInt 2017: An Ensemble of Deep Neural Architectures for Emotion Intensity Prediction in Tweets

Pranav Goel, Devang Kulshreshta, Prayas Jain, K.K Shukla

Department of Computer Science and Engineering
Indian Institute of Technology (BHU)
Varanasi, India-221005
Email: prayas.jain.cse14@iitbhu.ac.in

Friday 8th September, 2017

Outline

Introduction - EmoInt Shared Task

- Problem Statement - Given a tweet which conveys an emotion E, what is the intensity of E?
- Intensity is a real value ranging from 0 to 1 with 1 being the strongest intensity and 0 being the weakest intensity
- Fixed Set of Emotions - Anger,Fear,Joy,Sadness
- Example - *'I hate my lawn mower. If it had a soul, I'd condemn it to the fiery pits of Hell.'* has an intensity of 0.833 for the emotion Anger
- Motivation - Complement existing sentiment analysis systems with degree of sentiment. Useful for E-Commerce Companies.

Dataset

Emotion	**Train**	**Dev**	**Test**	**All**
Anger	857	84	760	1701
Fear	1147	110	995	2252
Joy	823	74	714	1611
Sadness	786	74	673	1533
All	3613	342	3142	7097

Proposed System

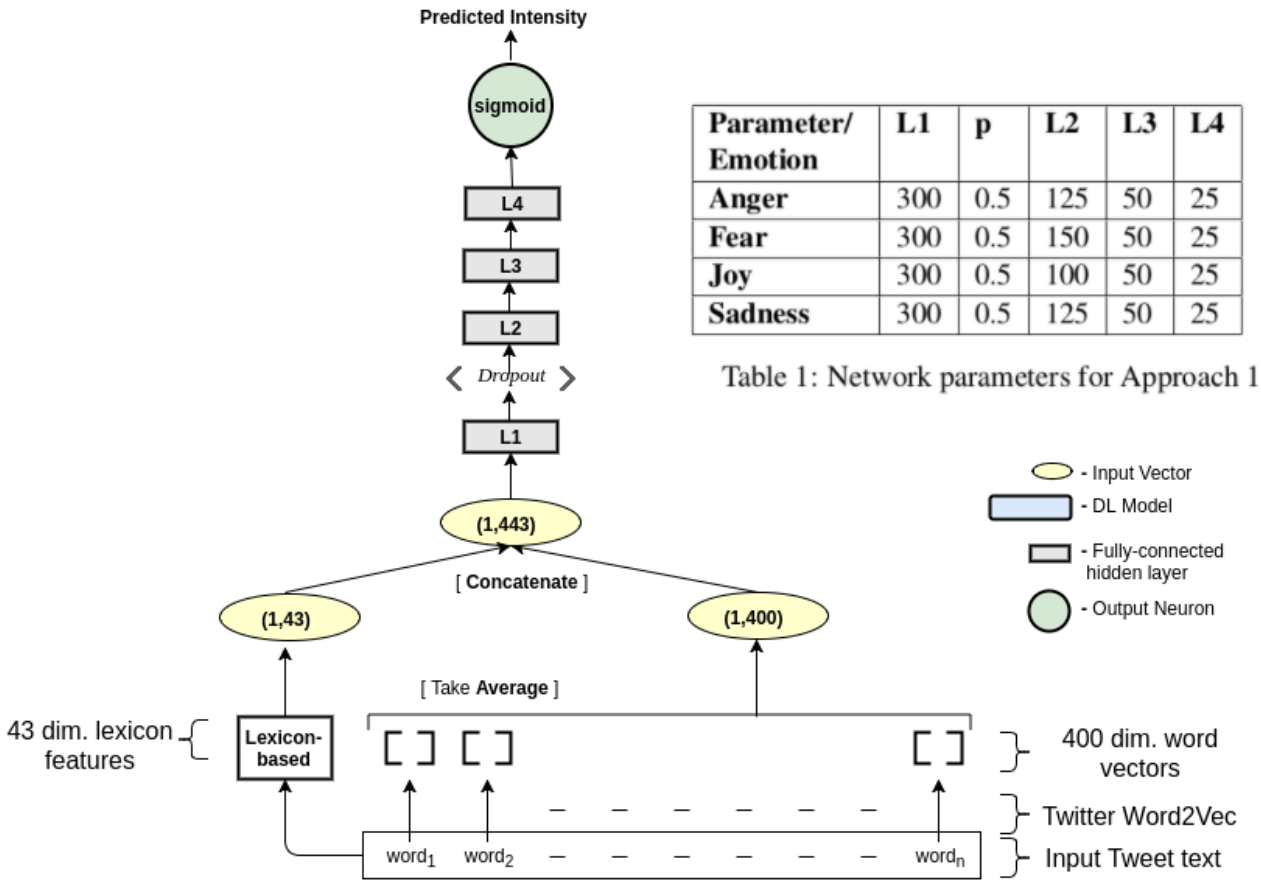
Our system is an ensemble of three sets of approaches.

- Feed Forward Neural Networks
- MultiTask Deep Learning
- Sequence Modelling using CNN and LSTM

Feed Forward Neural Network

- Convert each tweet to a 443 dimensional vector by using *Word2Vec* (400 dimensional vector) and *TweetToLexiconFeatureVector* (43 dimensional vector) and concatening the results
- Architecture - A four hidden layer neural network with dropout after the first layer. *ReLU* activation function is used.
- Training - Minimize negative pearson correlation using Adam algorithm with mini-batches of size 8 and training epochs 30.

Feed Forward Neural Network



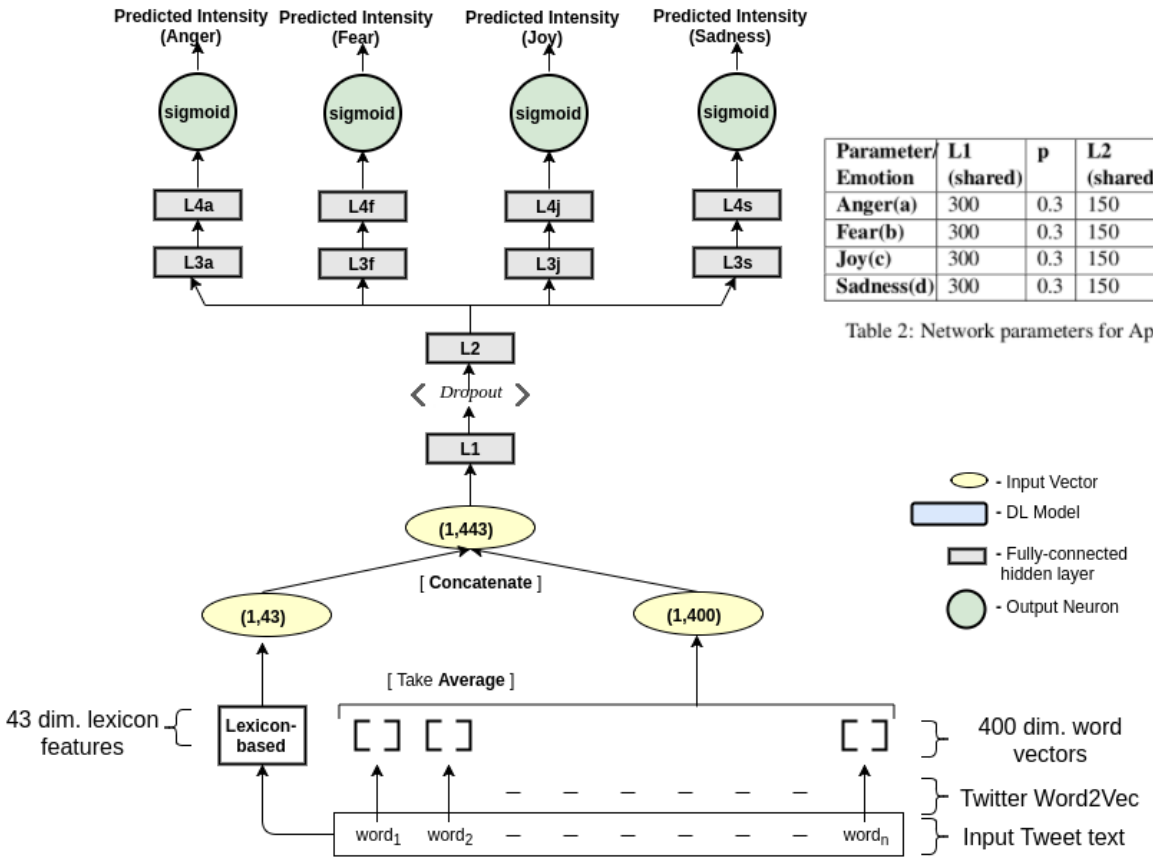
Parameter/ Emotion	L1	p	L2	L3	L4
Anger	300	0.5	125	50	25
Fear	300	0.5	150	50	25
Joy	300	0.5	100	50	25
Sadness	300	0.5	125	50	25

Table 1: Network parameters for Approach 1

MultiTask Deep Learning

- The tweet is converted to 443 dimensional input vector like in Feed Forward NN (Approach 1)
- We consider predicting the intensity of the different emotions as different sub-tasks.
- Architecture - We again use a four hidden layer NN like Approach 1. The network's initial layers (layer 1 and 2) are shared among various sub-tasks, whereas layer 3 and layer 4 are learnt independently.
- Layer 1 & 2 learn task general features, whereas Layer 3 & 4 learn sub-task specific features.
- Training - We use same cost function, optimization algorithm, number of epochs etc as that in Approach 1. We iterate on the emotions in a cyclic order.

MultiTask Deep Learning



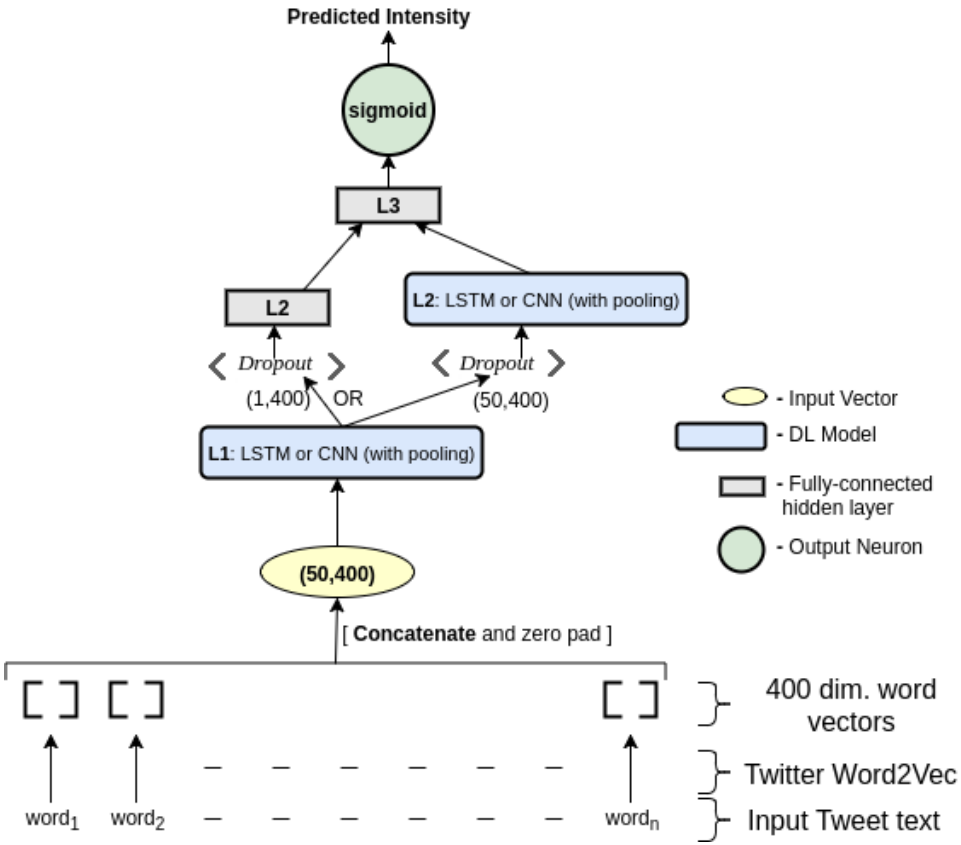
Parameter/ Emotion	L1 (shared)	p	L2 (shared)	L3	L4
Anger(a)	300	0.3	150	50	20
Fear(b)	300	0.3	150	75	25
Joy(c)	300	0.3	150	50	15
Sadness(d)	300	0.3	150	50	20

Table 2: Network parameters for Approach 2

Sequence Labelling using CNN & LSTM

- The tweet is converted to a (50,400) dimensional input vector using Word2Vec. We concatenate the vectors of each words, and zero pad them to get a constant size 0f (50,400).
- Architecture - The input is first fed to a CNN or LSTM. These models give a sequential output. We can either feed this sequential output to another CNN/LSTM or pool the vectors to form a single 400 dimensional vector which can be fed to fully connected layers. The number of fully connected layers to use or whether to stack LSTMs/CNNs or not is decided based on cross validation results. Finally, a neuron with sigmoid activation gives the intensity.
- Training - We optimize the mean absolute error using Adam algorithm with batch size 8 and 15 training epochs.

Sequence Labelling using CNN & LSTM -2



Sequence Labelling using CNN & LSTM -3

- The variations we tried - Layers, max pooling versus average pooling for CNNs, LSTM vs Bi-LSTM vs GRU vs RNN, number of neurons, and dropout value. The best three architectures based on cross validation results are reported.

Parameter/ Emotion	L1	p	L2	L3
Anger (1)	CNN (250,Max)	0	125	50
Anger (2)	CNN (256,Avg)	0	100	-
Anger (3)	LSTM (300)	0	CNN (200,Avg)	100
Fear (1)	LSTM (256)	.2	CNN (150,Avg)	100
Fear (2)	CNN (250,Max)	0	125	50
Fear (3)	LSTM (250)	.2	CNN (120,Avg)	50
Joy (1)	CNN (256,Max)	0	100	-
Joy (2)	LSTM (300)	0	CNN (200,Avg)	100
Joy (3)	LSTM (300)	.2	CNN (200,Avg)	100
Sadness(1)	CNN (250,Max)	0	125	50
Sadness(2)	CNN (250,Max)	.2	125	50
Sadness(3)	CNN (256,Max)	0	100	-

Table: Network Parameters for Approach 3

Ensembling the Models

- We combine 5 models (1 each from approach 1 & 2 and 3 from Approach 3) for each emotion.
- The final prediction is a weighted average of all the 5 models
- The weights are 1 for Approach 1, 3 for Approach 2, 3 each for top 2 systems from Approach 3, and 2 for the last system from Approach 3.
- Ensembling improved the performance by 2% over individual models.

Results

Our submitted ensemble model achieves an average (or overall) score of 75.26% and 74.70%, which beats the baseline model by about 14% and 10% on cross validation and test sets respectively.

Approach	Average		Anger		Fear		Joy		Sadness	
	CV	Test	CV	Test	CV	Test	CV	Test	CV	Test
Feed Forward NN	69.75	69.58	66.22	67.88	72.71	72.42	72.08	68.26	67.99	69.77
Multitask DL	66.30	66.20	63.73	64.49	68.07	67.74	66.80	65.37	66.65	67.22
CNN+LSTM Seq. Modeling	70.70	71.79	69.22	70.15	72.08	72.95	73.22	69.14	68.29	74.93
CNN+LSTM Seq. Modeling	70.25	72.15	69.08	69.86	70.95	73.27	72.93	69.86	68.04	75.6
CNN+LSTM Seq. Modeling	70.03	71.81	68.90	69.71	70.67	72.92	72.81	69.57	67.74	75.06
Ensemble Model	**75.26**	**74.70**	**72.94**	**73.2**	**76.78**	**76.20**	**74.42**	**73.20**	**76.90**	**76.50**
Baseline	61.10	64.8	60.50	63.9	57.40	65.2	70.30	65.4	56.20	64.8

Table: Results showing Pearson Correlation

Results

The ensemble model achieves about 3-5% improvement over the average scores.

Approach	Average		Anger		Fear		Joy		Sadness	
	CV	Test	CV	Test	CV	Test	CV	Test	CV	Test
Feed Forward NN	69.75	69.58	66.22	67.88	72.71	72.42	72.08	68.26	67.99	69.77
Multitask DL	66.30	66.20	63.73	64.49	68.07	67.74	66.80	65.37	66.65	67.22
CNN+LSTM Seq. Modeling	70.70	71.79	69.22	70.15	72.08	72.95	73.22	69.14	68.29	74.93
CNN+LSTM Seq. Modeling	70.25	72.15	69.08	69.86	70.95	73.27	72.93	69.86	68.04	75.6
CNN+LSTM Seq. Modeling	70.03	71.81	68.90	69.71	70.67	72.92	72.81	69.57	67.74	75.06
Ensemble Model	**75.26**	**74.70**	**72.94**	**73.2**	**76.78**	**76.20**	**74.42**	**73.20**	**76.90**	**76.50**
Baseline	61.10	64.8	60.50	63.9	57.40	65.2	70.30	65.4	56.20	64.8

Table: Results showing Pearson Correlation

Results

Approach 2 (Multitask DL) achieves the lowest scores among the three sets of approaches. Among Approach 1 (Feed Forward NN) and Approach 3 (CNN+LSTM Seq. Modeling), approach 3 has a best test score of 72.15 compared to approach 1's 69.58.

Approach	Average		Anger		Fear		Joy		Sadness	
	CV	Test	CV	Test	CV	Test	CV	Test	CV	Test
Feed Forward NN	69.75	69.58	66.22	67.88	72.71	72.42	72.08	68.26	67.99	69.77
Multitask DL	66.30	66.20	63.73	64.49	68.07	67.74	66.80	65.37	66.65	67.22
CNN+LSTM Seq. Modeling	70.70	71.79	69.22	70.15	72.08	72.95	73.22	69.14	68.29	74.93
CNN+LSTM Seq. Modeling	70.25	72.15	69.08	69.86	70.95	73.27	72.93	69.86	68.04	75.6
CNN+LSTM Seq. Modeling	70.03	71.81	68.90	69.71	70.67	72.92	72.81	69.57	67.74	75.06
Ensemble Model	**75.26**	**74.70**	**72.94**	**73.2**	**76.78**	**76.20**	**74.42**	**73.20**	**76.90**	**76.50**
Baseline	61.10	64.8	60.50	63.9	57.40	65.2	70.30	65.4	56.20	64.8

Table: Results showing Pearson Correlation

Results

Among the individual emotions, our ensemble model gives the best performance for 'Sadness', followed very closely by 'Fear', then 'Joy' and finally 'Anger'.

Approach	Average		Anger		Fear		Joy		Sadness	
	CV	Test	CV	Test	CV	Test	CV	Test	CV	Test
Feed Forward NN	69.75	69.58	66.22	67.88	72.71	72.42	72.08	68.26	67.99	69.77
Multitask DL	66.30	66.20	63.73	64.49	68.07	67.74	66.80	65.37	66.65	67.22
CNN+LSTM Seq. Modeling	70.70	71.79	69.22	70.15	72.08	72.95	73.22	69.14	68.29	74.93
CNN+LSTM Seq. Modeling	70.25	72.15	69.08	69.86	70.95	73.27	72.93	69.86	68.04	75.6
CNN+LSTM Seq. Modeling	70.03	71.81	68.90	69.71	70.67	72.92	72.81	69.57	67.74	75.06
Ensemble Model	**75.26**	**74.70**	**72.94**	**73.2**	**76.78**	**76.20**	**74.42**	**73.20**	**76.90**	**76.50**
Baseline	61.10	64.8	60.50	63.9	57.40	65.2	70.30	65.4	56.20	64.8

Table: Results showing Pearson Correlation

Future Work

- User hand-crafted features as additional input to the model
- Use other filters provided in the AffectiveTweets Package, provided with the shared task. Example - TweetToSentiStrengthFeatureVector, TweetNLPTokenizer
- Try Ensembling the models in a different fashion

Thank You

⋮

Questions?

For any questions or queries, email us

pranav.goel.cse14@iitbhu.ac.in
prayas.jain.cse14@iitbhu.ac.in
devang.kulshreshta.cse14@iitbhu.ac.in