ATTENTION-BASED LSTM FOR PSYCHOLOGICAL STRESS DETECTION FROM SPOKEN LANGUAGE USING DISTANT SUPERVISION

Genta Indra Winata, Onno Pepijn Kampman, Pascale Fung

Human Language Technology Center, Department of Electronic and Computer Engineering

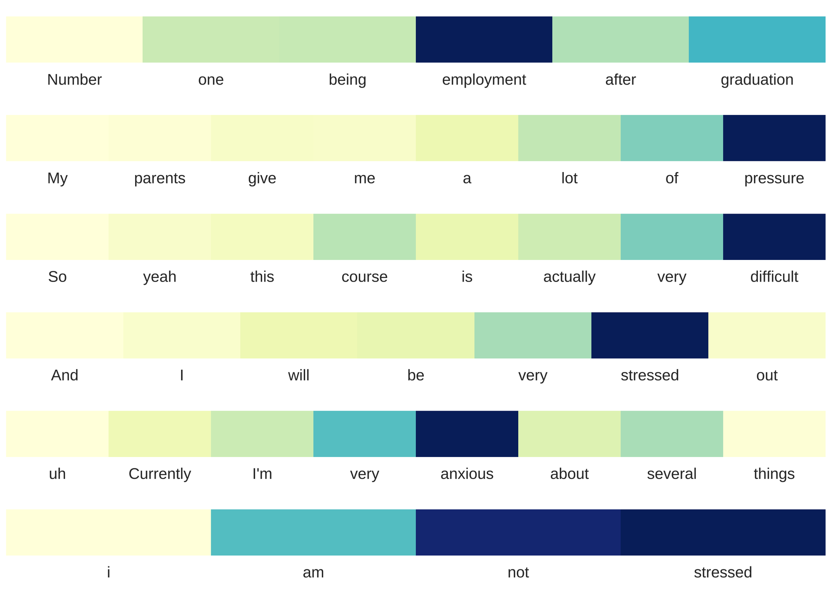
香港科技大學 THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

Introduction

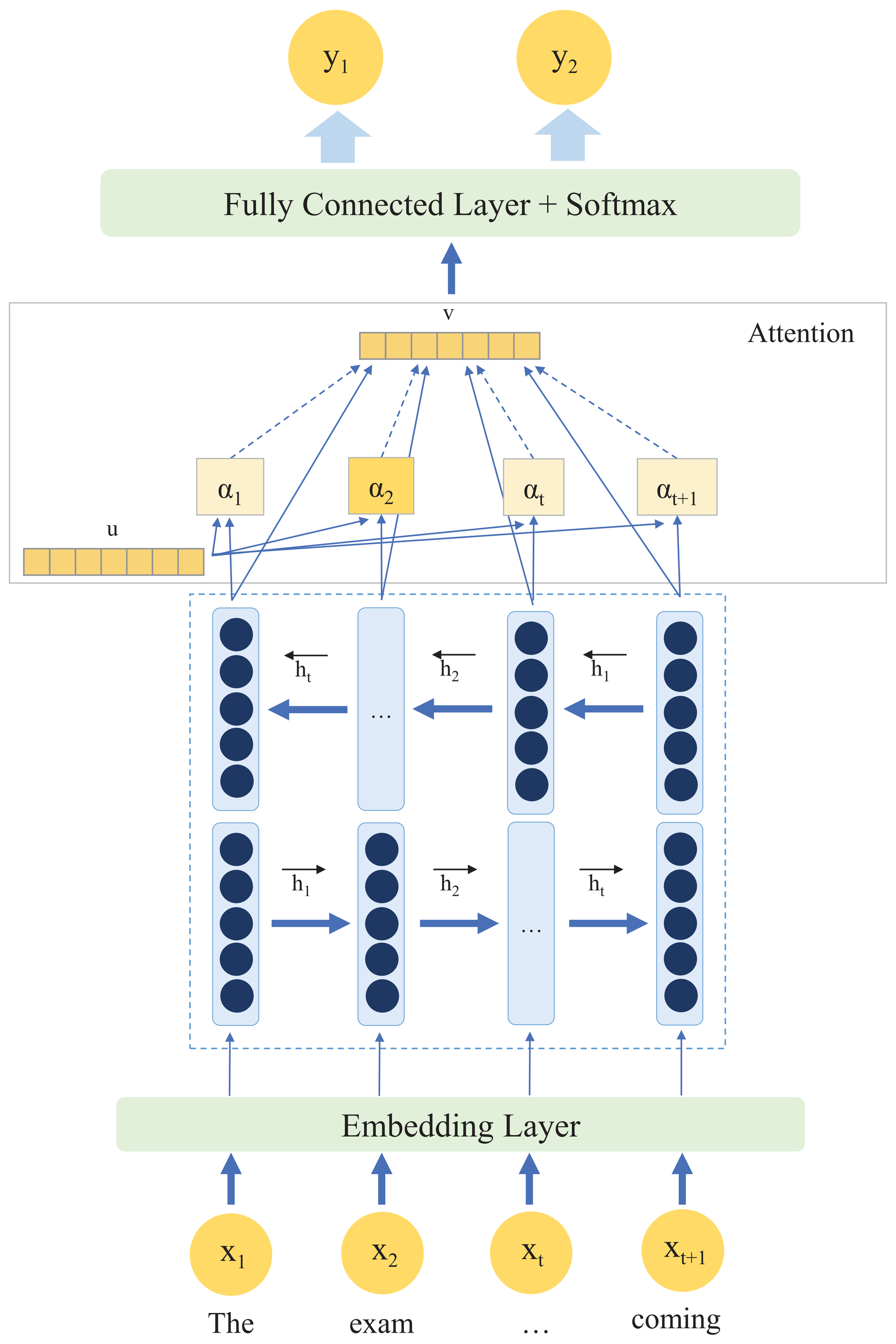
- Linguistic studies have shown that language choice contains pointers to levels of stress and mental health. Research on sentence-level stress detection has been mostly focused on written text collected from social media and utilized stress-related words from dictionary [1, 2].
- Recently, attention-based LSTM is a model that learns long dependency across words in an utterance and weighs the importance of every word in the memory [3].
- Distant supervision is known to be useful in utilizing noisy labels in tweets.

Visualization

To interpret the trained model, we extract the attention weights from the best model and evaluate several stressed and unstressed utterances. Darker colors represent stronger word contributions to the classification task. Interestingly, it captures key terms related to stress.



Model Architecture



References

[1] Huijie Lin, Jia Jia, Quan Guo, Yuanyuan Xue, Jie Huang, Lianhong Cai, and Ling Feng. Psychological stress detection from cross-media microblog data using deep sparse neural network. In *Multimedia and Expo (ICME), 2014 IEEE International Conference on*, pages 1–6. IEEE, 2014.

[2] Huijie Lin, Jia Jia, Liqiang Nie, Guangyao Shen, and Tat-Seng Chua. What does social media say about your stress?. In *IJCAI*, pages 3775–3781, 2016.

[3] Zichao Yang, Diyi Yang, Chris Dyer, Xiaodong He, Alexander J Smola, and Eduard H Hovy. Hierarchical attention networks for document classification. In *HLT-NAACL*, pages 1480–1489, 2016.

Methodology

We build a bidirectional LSTM (BLSTM) taking word embedding as input. We denote V as the number of unique words in our corpus and k as the dimension of the word embedding vectors. Each word is a one-hot vector $\boldsymbol{x} \in \mathbb{R}^{|V|}$ and performs a multiplication with the embedding layer $A \in \mathbb{R}^{|V| \cdot k}$, where $k = 100$. The resulting vector is $\boldsymbol{b} \in \mathbb{R}^k$. The LSTM consists of one recurrent layer that propagates the embedding vector b_t for the word at time t where $t \in [1, T]$.

$$\boldsymbol{b} = A^T \boldsymbol{x}$$
$$\overrightarrow{h_t} = LSTM(b_t)$$
$$\overleftarrow{h_t} = LSTM(b_{T-t})$$
$$h = [\overrightarrow{h_t} : \overleftarrow{h_{T-t}}]$$

All hidden states from both directions are concatenated and fed into a subsequent attention layer [3]. The word importance vector u_t is calculated. The normalized word weight α_t is obtained through a softmax. The aggregate of all the information in the sentence **v** is the weighted sum of each h_t with α_t as corresponding weights.

$$u_t = tanh(W h_t + b)$$
$$\alpha_t = \frac{exp(u_t^T u)}{\Sigma_t exp(u_t^T u)}$$
$$\boldsymbol{v} = \sum_t \alpha_t h_t$$

This vector **v** is then fed to a fully connected layer with softmax activation to perform the final classification. The prediction is a vector $\boldsymbol{y} \in \mathbb{R}^2$ with the probabilities of being unstressed and stressed. We choose the highest probability by using $argmax$ as the model's prediction.

Distant Supervision

A method to add unlabeled Twitter tweets to our training set. This technique refers to extracting noisy signals from text as label. We manually pick hashtags that indicate either a stressed or unstressed state of mind of the author, and use them to scrape stressed and unstressed tweets. They are included because our interview corpus is relatively small and covers a limited number of topics, mostly related to academia.

Model performance

method	accu.	prec.	recall	f-score
SVM	68.7	72.0	61.2	66.2
LSTM	70.0	70.3	68.1	69.2
LSTM w/ attention	**73.8**	**74.7**	**71.9**	**73.2**
BLSTM	72.2	74.5	67.5	70.8
BLSTM w/ attention	72.5	73.1	71.2	72.2

After applying Distant Supervision

method	accu.	prec.	recall	f-score
LSTM	73.4	73.6	73.1	73.4
LSTM w/ attention	73.8	74.4	72.5	73.4
BLSTM	73.8	**74.7**	71.9	73.2
BLSTM w/ attention	**74.1**	73.6	**75.0**	**74.3**

Conclusion

- The best performance was found for our bidirectional LSTM model, which outperformed the other models in terms of accuracy, recall, and f-score.
- The two-phase training method with the out-of-domain stress tweets dataset improves the learning performance and robustness.