MATH 6380q Project 1: Whether and why are people feeling happy? Mining Affective Events Based on Text-based Information

Ziming WU[1] , Feng HAN[1] , Song LIU[2]

[1]Department of Computer Science and Engineering, The Hong Kong University of Science and Technology,

[2]Department of Chemistry, The Hong Kong University of Science and Technology,

Abstract

Happiness Science is a field of positive psychology that focuses on understanding what behaviours make people happy in a sustainable way. Statistical and machine learning methods have opened a new door for understand how people express their happy moments. In this project, we use the power of Deep Learning method: Cruz-Affect in terms of their effectiveness of extracting features for the datamining of HappyDB dataset.

Introduction

HappyDB is a corpus with 100,000 crowdsourcing happy moments. The goal of the corpus is to advance the artistic state of understanding the causes of happiness that can be gathered from the text.

Figure 1. Demonstration of HappyDB dataset

HappyDB is a collection of sentences in which crowdworkers answered the question: what made you happy in the past 24 hours (or alternatively, the past 3 months). Naturally, the descriptions of happy moments exhibit a high degree of linguistic variation.

Some examples from HappyDB dataset:

1. Unexpected presents or gifts.
2. Driving in my car, with my girl by my side, my hand in hers, and singing to the radio.
3. I managed to get to work early and bring in donuts for all the coworkers.
4. I bought fresh fruits and vegetables with low price.
5. Waking up to things being clean and tidy

More details are available on this site:
https://rit-public.github.io/HappyDB/

Methodologies

Cruz-Affect is designed for effective classification tasks, and it contains several types of robust and efficient models. In this work, a deep learning Convolutional Neural Networks (CNN) classifier is used for the classification of HappyDB dataset. We utilize several sentiment lexicons and explore the emotional features, to investigate essential indicators of social involvement and control that a subject might exercise in their happy moments, which was described in textual snippets from the HappyDB database. The data contains a labeled set (10K), and an unlabeled set (70K). Firstly, we use supervised methods on the 10K dataset, and then we use a bootstrapped semi-supervised algorithm on the 70K dataset. We explore these models for binary classification of agency and social labels, as well as a multi-class prediction for concepts labels. We obtained promising results on the held-out data, suggesting that the proposed feature sets effectively represent the data for effective classification jobs. We also build concepts models to discover general themes recurring in happy moments. Our results indicate that general characteristic is shared between the classes of the agency, concepts and social, suggesting it should be possible to build general models for effective classification tasks.

For the CNN model with word embedding, we explore its performance with different parameter settings. The best hyperparameters of the CNN model include filter size 3, multiple region size (2, 3, 4) and max pooling size 1, or filter size 4, multiple region size (2, 3, 4, 5) and max pooling size 1. The region size implies a windows size for N-grams. After getting the best hyperparameters, we train the model with word embeddings, and word embeddings concatenated with syntactic and emotional features, to test whether syntactic and emotional features improve performance. Figure 2 illustrates the CNN model with region size (2, 3, 4).

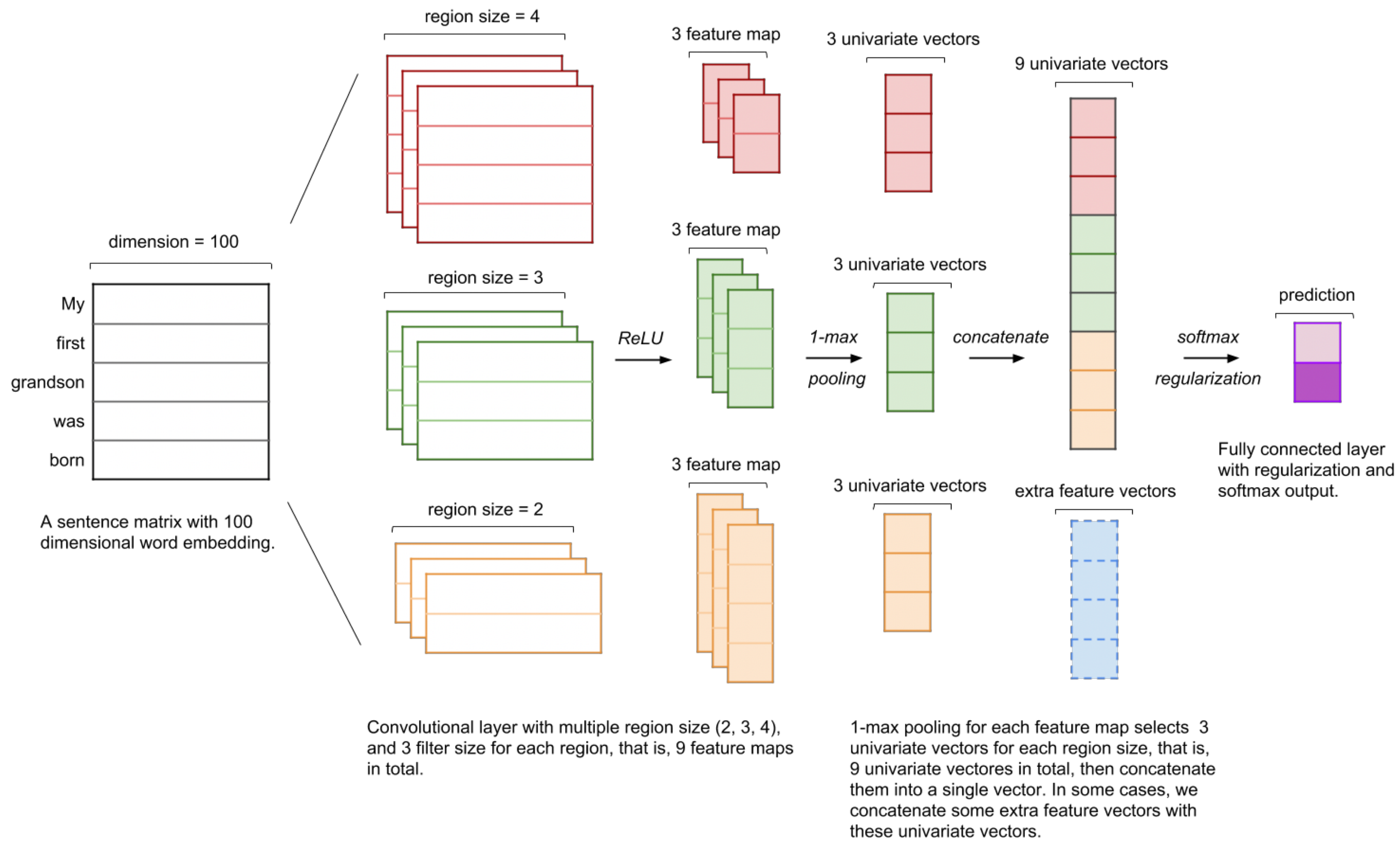

Figure 2. A Diagram for the CNN model with region size (2, 3, 4) and filter size 3 for a single sentence.

Results

Model	3-layer CNN	ResNet-18	ResNet-34	ResNet-50	ResNet-101	ResNet-152
Agency-Classification w/ Bert	76.24%	82.60%	82.72%	83.04%	83.35%	83.08%

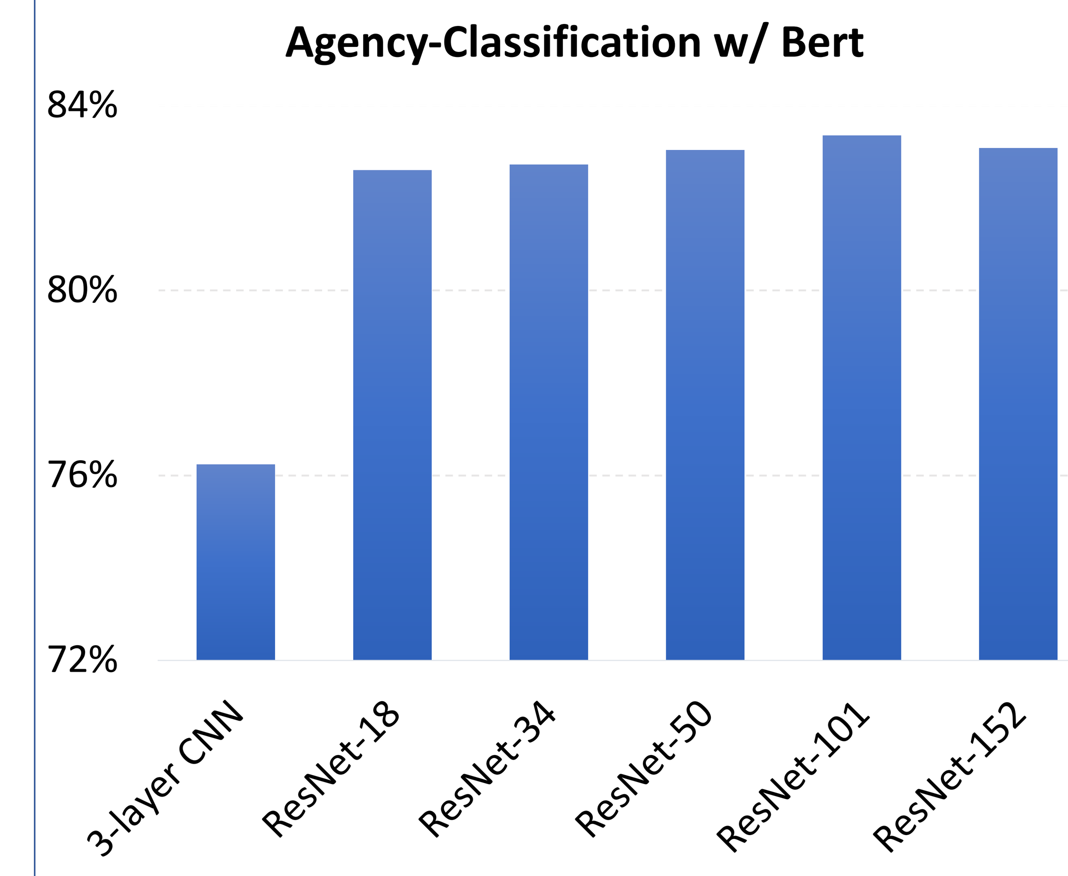

Table 1 & Figure 3. Comparison of Agency-Classification w/ Bert of different models

Discussion & Conclusions

This study demonstrates the possibility of characterizing happiness computationally, which can shed light on the understanding of what behaviours make people happy. Through the experiments, we show that Bert can provide better semantic embedding compared to GloVe in terms of encoding happiness information. Our experiment results also indicate that ResNet with different number of layers while deeper networks tend to perform slightly better.

References

- J. Wu et al., "HappyDB: A Corpus of 100,000 Crowdsourced Happy Moments," pp. 1–11, 2018.
- J. Wu, R. Compton, G. Rakshit, M. Walker, P. Anand, and S. Whittaker, "CruzAffect at AffCon 2019 Shared Task: A feature-rich approach to characterize happiness," pp. 1–11, 2019.