An Evaluation of Emotional Information in Semantic Word Embeddings

Anonymous ACL submission

Abstract

In recent years word embeddings learned as distributional representations of words have shown to capture semantic representation, often preseving relationships and seemingly capturing a notion of meaning in the vectors learned from the data. There exists a desire to construct emotionally intelligent systems, those that can understand emotion in the context of converstation, and for the generation of emotionally meaningful language. In this paper we briefly examine whether or not these distributionally learned representations preserve emotional values.

1 Introduction

Recent advances in language-modeling has conveyed the power of word-vectors' (*word-embeddings'*) ability to meaningfully preserve relationships between words, purely from observing data. The alure of these embeddings is that it suggests a computationally-discoverable structure within languages. Specifically, we wish to discover features of the structure enabling a language to convey concepts beyond regular sensory observation. For example, word-vectors have been demonstrated to succesfully capture relationships between entities such as countries and their capitals.

Another concept often conveyed in text, is human emotion. Works presented in Gibson et al. (2015, 2016) deal with enabling machines to extract emotion from language representation for improving counseling. Fitrianie and Rothkrantz (2008), Bautista et al. (2014), and Ghosh et al. (2017) all tackle methods for embedding emotion into emotional text generation. In essence they all require one fundamental component: that human language successfully captures human emotion.

The discussion presented focuses on trying to interpret the structure of emotion in language. It attempts to determine the effectiveness of word embedding to capture emotional content of words, and the shape of the space that emotion occupies within these embeddings. Experiments will attempt to examine different word-embeddings and specific linear and nonlinear structures which demonstrate the best accuracy for this task, and the possible conclusions reached.

2 Method

2.1 Emotion Data

To represent emotion we will use vectors that represent the valence, arousal, and dominance (VAD) of individual words. The specific data se being used contains just under 14,000 words collected by Warriner et al. (2013). The data represents responses from there 1,827 Amazon Mechanical Turk users that self-identified as residing inside the United States. Each responder was shown a set of of approximately 350 words and were asked to respond with with a score from a 1 to 9. Each participant responded with rating along a single emotional dimension, where 1 correlated to happy, excited, and controlled and 9 to unhappy, calm, and in control for valence, arousal, and dominance respectively.

In the following discussion when refering to emotion-vectors, we imply the vector representing the average VAD scores for each word.

2.2 Experiment 1

The first experiment examines how language models correspond to learning emotion related features. By comparing 2 different word-embeddings we aim to access whether or not small amounts of sentimental (i.e. emotionally charged) text contain stronger linguistic indicators for or whether a general word embedding trained a large corpus performs better.

The large corpus is the pretrained model `GoogleNews-vectors-negative300` (Google News) made available by McCormick (2016). It is an implementation of the *Word2Vec* model presented by initially by Mikolov et al. (2013) contructing a 300-dimensional word and phrase embedding for 3 million words and phrases. Approximately 100 billion words of text was used to train this model, which ultimately demonstrates an ability to preserve semantic relationships between words in English.

Buechel and Hahn (2017a,b) provide a dataset of 10k English sentences that were tagged with VAD data. This data set is used in conjunction with wine reviews provided by Thoutt (2017) to build a set of approximately 300k sentimental English sentences, that should all be sentimental or emotionally charged. This data set (referred to henceforth as *EmoBank*) is used to learn a 300-dimensional word embedding using the same Word2Vec methodology as described by Mikolov et al. (2013).

A linear projection will be evaluated, as it represents whether or not there exist fundamental components in the word-embeddings that correspond closely with emotional components.To accomplish this a single layer neural network with no activation and 3 neurons is trained. The training input data is a random selection of word-vectors from 80% of overlapping vocabulary (i.e. words that are in both the Google News and EmoBank vocabularies) and the corresponding VAD vectors using the mean-squared error (MSE) as the loss function we are attempting to minimize.

The evaluation of the learned projecttion is tested on the remaining 20%, and we compare the 2 models performance by looking at the MSE of each of the VAD compenents. Simarily we compare the results to the variance of the data set learned. The variance can be thought of the performance of a neural network that simply maps all inputs the averages of the data that it learned, and can indicate how strong the learned relationships are between the word-vectors and the emotion-vectors.

2.3 Experiment 2

The second experiment is designed to compare the performances of some non-linear embeddings against the linear embedding. This experiment is designed to asses whether or not emotion exists as a linear subspace of the semantic word-embeddings learned by Word2Vec.

To accomplish this, the Google News Word2Vec model will be trained on 80% of all of the words that appear both in its vocabularry and the VAD data. The different neural networks used were:

- *Linear* - single layer densely connected neural network with no activation function
- *ReLU* - 2 layer neural network with one hidden layer activated by Rectified Linear Unit (ReLU) and a densily connected output layer with no activation
- *Sigmoid* - 2 layer neural network with one hidden layer activated by Sigmoid and a densily connected output layer with no activation
- *Leaky ReLU* - 3 layer neural network with one hidden layer with no activation, a second with a Leaky ReLU activation and and output layer with no activation.

Similarly to Experiment 1 the loss and comparison criteria were the MSE of the outputs on the remaining 20% of words. Once again the MSE was also compared against the variance of the data.

3 Results

3.1 Experiment 1

Component	Data Variance	MSE	
		Google	EmoBank
Valence	1.43920	0.58749	1.29711
Arousal	0.81877	0.49058	0.76640
Dominance	0.79727	0.48412	0.78720

Table 1: Training Examples comparison between EmoBank and Google Word2Vec models

Component	Data Variance	MSE	
		Google	EmoBank
Valence	1.44463	0.68653	1.44066
Arousal	0.81829	0.58126	0.87611
Dominance	0.77392	0.51093	0.87313

Table 2: Testing Examples comparison between EmoBank and Google Word2Vec models

Indicated by the smaller MSE in both train and test sets the Google News Word2Vec embedding outperforms the EmoBank Word2Vec model.

Source	Mean	Variance and MSE
Data	5.05335	1.62523
Linear Valence	5.02238	0.63053
ReLU Valence	5.07406	0.60693
Sigmoid Valence	5.07471	0.52348
Leaky ReLU Valence	5.01538	0.59535

Table 3: Valence Training Examples

Source	Mean	Variance and MSE
Data	4.21099	0.80281
Linear Arousal	4.21468	0.50102
ReLU Arousal	4.20469	0.48095
Sigmoid Arousal	4.18126	0.47872
Leaky ReLU Arousal	4.21087	0.48020

Table 4: Arousal Training Examples

These results indicate that the features discovered by the Google News model are stronger indicators of emotion than those discovered in the EmoBank model.

While EmoBank Word2Vec seems to perform better than just guessing the mean on training, it does worse than the variance on the test set, indicating it overfit itself to the training examples. This is indicitive of relying heavily on bias factors for the final result, futher indicating that it failed to succesfully learn a meaningful linear projection between the word space and the emotional space.

3.2 Experiment 2

Intuitively, using a larger training and test sets improves the accuracy of the word-vectors to predict the emotion-vectors. While performing compari-

Source	Mean	Variance and MSE
Data	5.18298	0.88463
Linear Dominance	5.14642	0.47937
ReLU Dominance	5.20327	0.48086
Sigmoid Dominance	5.19022	0.44858
Leaky ReLU Dominance	5.12235	0.44444

Table 5: Training Dominance Examples

Source	Mean	Variance and MSE
Data	5.05335	1.62010
Linear Valence	5.04103	0.66228
ReLU Valence	5.09063	0.62524
Sigmoid Valence	5.08294	0.57025
Leaky ReLU Valence	5.03105	0.62010

Table 6: Valence Test Examples

Source	Mean	Variance and MSE
Data	4.21099	0.80255
Linear Arousal	4.21117	0.51866
ReLU Arousal	4.19784	0.49404
Sigmoid Arousal	4.17127	0.49552
Leaky ReLU Arousal	4.20352	0.49322

Table 7: Arousal Test Examples

Source	Mean	Variance and MSE
Data	5.18298	0.86395
Linear Dominance	5.16406	0.48195
ReLU Dominance	5.21421	0.47356
Sigmoid Dominance	5.19382	0.45596
Leaky ReLU Dominance	5.13655	0.44599

Table 8: Dominance Test Examples

bly, nonlinear activations seem to improve the general performance of the system, implying that that there is non-linearity in the projection between word-vectors and emotion-vectors.

4 Conclusion

The results of §3.1 indicate that the smaller more domain specific Word2Vec model does not capture strong indicators of emotion in its embedding. From this we can conclude that a more general training corpus will allow the model to embed emotional features into the semantic relationships between words. This is likely explainable by the fact tha language is rarely devoid of emotion. Writers will use vocabulary that they feel will best elicit the desired reaction from a reader, strongly impacting the relationships that words have amongst themselves, and with emotions.

§3.2 examines the structure of the subspace occupied by the emotional features. The results demonstrate that the non-linear activation functions improved the prediction of VAD values. From this it follows that either the non-linear activation allows the data to over fit to noise in the data, or because the space itself is actually highly nonlinear. However, looking at the differences between variance and means from the test and training sets suggests that the vocabulary is fairly homogenous in distribution, implying that based on this data set is likelier that the projection must contain nonlinearity.

In general, though the results indicate that emotion does exist as a subcomponent of the distributional representation of semantic meaning.

5 Future Work

In the future these methods and results can be used to motivate architectures for better emotional text synthesis and processing. Extending this work to phrase level emotion prediction would potentially enable an improvement in text based dialogue systems, as they would now have a better indication of emotional context.

References

Susana Bautista, Dpto Ingeniería, Pablo Gervás, Alberto Díaz, and Dpto Ingeniería. 2014. Adaptive text generation based on emotional lexical choice https://doi.org/10.1145/2662253.2662273.

Sven Buechel and Udo Hahn. 2017a. EmoBank: Studying the impact of annotation perspective and representation format on dimensional emotion analysis. *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics (EACL-2017)* 2:578–585.

Sven Buechel and Udo Hahn. 2017b. Coping with Different Perspectives of Text Understanding in Emotion Annotation. *Proceedings of the 11th Linguistic Annotation Workshop* pages 1–12. http://www.julielab.de.

Siska Fitrianie and Leon J.M. Rothkrantz. 2008. The generation of emotional expressions for a text-based dialogue agent. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 5246 LNAI:569–576. https://doi.org/10.1007/978-3-540-87391-4_72.

Sayan Ghosh, Mathieu Chollet, Eugene Laksana, Louis-Philippe Morency, and Stefan Scherer. 2017. Affect-LM: A Neural Language Model for Customizable Affective Text Generation https://doi.org/10.18653/v1/P17-1059.

James Gibson, Doan Can, Bo Xiao, Zac E. Imel, David C. Atkins, Panayiotis Georgiou, and Shrikanth Narayanan. 2016. A deep learning approach to modeling empathy in addiction counseling. *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH* 08-12-Sept(76):1447–1451. https://doi.org/10.21437/Interspeech.2016-554.

James Gibson, Athanasios Katsamanis, Francisco Romero, Bo Xiao, Panayiotis Georgiou, and Shrikanth Narayanan. 2015. Multiple instance learning for behavioral coding. *IEEE Transactions on Affective Computing* PP(99):81–94. https://doi.org/10.1109/TAFFC.2015.2510625.

Chris McCormick. 2016. Google's trained word2vec model in python. http://mccormickml.com/2016/04/12/googles-pretrained-word2vec-model-in-python/.

Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Distributed Representations of Words and Phrases and their Compositionality pages 1–9. https://doi.org/10.1162/jmlr.2003.3.4-5.951.

Zack Thoutt. 2017. Wine reviews: 130k wine reviews with variety, location, winery, price, and description. https://www.kaggle.com/zynicide/wine-reviews.

Amy Beth Warriner, Victor Kuperman, and Marc Brysbaert. 2013. Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behavior Research Methods* 45(4):1191–1207. https://doi.org/10.3758/s13428-012-0314-x.