Can Word Embeddings Help Find Latent Emotions in Text? Preliminary Results

Armin Seyeditabari and Wlodek Zadrozny

Computer Science Dept. University of North Carolina at Charlotte
sseyedi1@uncc.edu wzadrozn@uncc.edu

Abstract

We report results of several experiments evaluating performance of word embeddings on semantic similarity of emotions. Our experiments suggest that the standard embeddings like GloVe and Word2Vec have very limited applicability in identifying emotions in text. Namely, using the standard arithmetic of emotions as a test, we show the mean reciprocal rank of a correct response is about 0.24, that is, combinations of word vectors are not a good proxy for expressed emotions. For example, the sum vector Joy+Fear, contrary to expectations, is not close to the vector representing Guilt. In addition, the opposite emotions, like Pessimism and Delight, have relatively high similarity to each other as word vectors (on average 0.2-0.44). Another experiment shows relatively low similarity (0.2-0.3) of word embeddings for similar emotions, such as Anger and Envy. Thus the standard methods for producing word embeddings are not adequate to represent relationships between emotion words. We conclude with a few hypotheses about improving the accuracy of embeddings in representing emotions.

Introduction

With the variety of places that humans can express themselves about any event or news, we have a great amount of data that can be used to analyze the emotional responses of a population. This could be to a political or social event, or to a new product introduced to the market. In our case, the impetus for the work presented here came from a project with social scientists on analyzing emotions in surveys about recent political protests (Seyeditabari et al. 2017).

There are fairly good methods to extract sentiment from a text, but in the fields such as social sciences or political science the need for extracting more fined grained emotions from different types of textual data such as open ended questions in surveys, comments on news articles, social media posts, etc. is apparent. Sentiment alone is not adequate to capture the intricacy of the human emotional state, and in particular, key-word based sentiment analysis.

In this paper we report preliminary results on an attempt to use distributional semantics to analyze expressions of emotions. If successful, this approach would allow us to reduce reliability on human analysis of surveys, allow surveys with larger number of participants, and potentially extend to the analysis of other social media. However, at this point results are mostly negative. Our experiment show that the standard word embeddings methods like GloVe (Pennington et al. 2014) and Word2Vec (Mikolov et al. 2013) do not predict the main emotion from component emotions. In contrast to the well-known (Dai et al. 2015; Pennington et al. 2014; Mikolov et al. 2013) approximate semantic arithmetic equalities such as

$$\text{King} - \text{Man} + \text{Woman} = \text{Queen} \quad (1)$$

we do not obtain

$$\text{Joy} + \text{Fear} = \text{Guilt} \quad (2)$$

In fact, the mean is reciprocal rank of a correct response for such sums of emotions is about 0.24, that is it combinations of word vectors are not a good proxy for expressed emotions.

Overview of prior work

Human emotions have been the subject of many studies in social and psychological disciplines e.g. (Kahneman et al. 2007; Strapparava et al. 2008; Vasilopoulos 2015; Evans et al. 2016). For example, Vasilopoulos shows that "fear stemming from a terrorist attack will increase motivation to seek out political information, yet will have a negative effect on actual participation. On the contrary, anger will hinder information-seeking but will mobilize participation in political action, even when such action entails an increased physical risk for the participant" (Vasilopoulos 2015). Clearly results such as these show the relevance of understanding emotions in political discourse and prediction of political events. Evans et al. provide a good overview of the use of NLP techniques in political sciences.

Human feelings (results of emotions)	Emotions	Opposite	Emotions
Love	Joy + Trust	Remorse	Sadness + Disgust
Guilt	Joy + Fear	Envy	Sadness + Anger
Delight	Joy + Surprise	Pessimism	Sadness + Anticipation
Submission	Trust + Fear	Contempt	Disgust + Anger
Curiosity	Trust + Surprise	Cynicism	Disgust + Anticipation
Sentimentality	Trust + Sadness	Morbidness	Disgust + Joy
Awe	Fear + Surprise	Aggression	Anger + Anticipation
Despair	Fear + Sadness	Pride	Anger + Joy
Shame	Fear + Disgust	Dominance	Anger + Trust
Disappointment	Surprise + Sadness	Optimism	Anticipation + Joy
Unbelief	Surprise + Disgust	Hope	Anticipation + Trust
Outrage	Surprise + Anger	Anxiety	Anticipation + Fear

Table 1*. Emotions arithmetic and opposite feelings. Our experiments show that word embedding of emotion words do not satisfy these equalities (see text for details). Table source: https://en.wikipedia.org/wiki/Contrasting_and_categorization_of_emotions*

Primary emotion	Secondary emotion	Tertiary emotion
Liking	Affection	Adoration · *Fondness · Liking · Attractiveness · Caring · Tenderness · Compassion · Sentimentality*
	Lust/Sexual desire	Desire · *Passion · Infatuation*
	Longing	Longing
Joy	Cheerfulness	Amusement · *Bliss · Gaiety · Glee · Jolliness · Joviality · Joy · Delight · Enjoyment · Gladness · Happiness · Jubilation · Elation · Satisfaction · Ecstasy · Euphoria*
	Zest	Enthusiasm · *Zeal · Excitement · Thrill · Exhilaration*
	Contentment	Pleasure
	Pride	Triumph
	Optimism	Eagerness · *Hope*
	Enthrallment	Enthrallment · Rapture
	Relief	Relief
Surprise	Surprise	Amazement · *Astonishment*
Anger	Irritability	Aggravation · *Agitation · Annoyance · Grouchy · Grumpy · Crosspatch*
	Exasperation	Frustration
	Rage	Anger · *Outrage · Fury · Wrath · Hostility · Ferocity · Bitter · Hatred · Scorn · Spite · Vengefulness · Dislike · Resentment*
	Disgust	Revulsion · *Contempt · Loathing*
	Envy	Jealousy
	Torment	Torment
Sadness	Suffering	Agony · *Anguish · Hurt*
	Sadness	Depression · *Despair · Gloom · Glumness · Unhappy · Grief · Sorrow · Woe · Misery · Melancholy*
	Disappointment	Dismay · *Displeasure*
	Shame	Guilt · *Regret · Remorse*
	Neglect	Alienation · *Defeatism · Dejection · Embarrassment · Homesickness · Humiliation · Insecurity · Insult · Isolation · Loneliness · Rejection*
	Sympathy	Pity · *Mono no aware · Sympathy*
Fear	Horror	Alarm · *Shock · Fear · Fright · Horror · Terror · Panic · Hysteria · Mortification*
	Nervousness	Anxiety · *Suspense · Uneasiness · Apprehension (fear) · Worry · Distress · Dread*

Table 2. *Three layered emotion classification by Parrott shows a categorization of emotions. Our experiments show that these categories are not respected by word embeddings (see text). Table source:https://en.wikipedia.org/wiki/Contrasting_and_categorization_of_emotions*

Techniques for emotion extraction from text are still developing, and they benefits from prior work on sentiment analysis; (Farzindar et al. 2015) discuss how the two are connected. Ambiguity and subjectivity in natural language are among many reasons that make it harder to extract emotion from text than, for example, named entities. Dimensionality reduction methods to identify emotions are often used; e.g. (Kim et al. 2010) use variants of LSA and non-negative matrix factorization to identify four emotions: Anger, Fear, Joy, and Sadness.

Word embeddings are a newer promising method of dimensionality reduction and building distributional semantic models (see e.g. (Bellegarda 2010) for an overview of its promises and open problems). Therefore, it makes sense to ask how well we can expect word embeddings to model the main classes of emotions and the relationships between them.

Experiments with embedding emotions

We experimented with distributional semantics in analyzing combinations of emotions. The motivation was to detect latent emotions in survey responses i.e. emotions not explicitly expressed. Since the responses use complicated language and often express multiple emotions it is natural to ask whether current methods of distributional semantic analysis could help in inferring latent emotions; the natural idea being that the sum of vectors representing words (or phrases) in a survey response should be close to the embedding of the vector representing the latent emotion. As a test of feasibility of the approach, we decided to check whether the distances between hypothetical responses consisting of two explicit emotion words and the vector of the representing the latent emotion that humans could reliably infer, e.g.

'Fear' + 'Anticipation' = 'Anxiety' (3)

Our hypothesis was that word embeddings should follow the well-known calculus of emotions shown in Table 1, where we can see equations such as (2) or (3). Therefore, we performed several experiments using two the well know methods GloVe (Pennington et al. 2014) and Word2Vec (Mikolov et al. 2013). As state of the art methods for creating word vector representations, Word2Vec and GloVe have shown promising results in capturing the semantic similarity and fair results in semantic arithmetic (Mikolov et al. 2013). In this article, we test their performance on both emotion arithmetic and in computing similarities of emotions (see Table 2). We have performed several experiments evaluating the performance of the two canonical types of word embeddings on the task of computing the semantic similarity of emotions.

To our knowledge this is the first such evaluation, and given the importance of understanding emotions in text, our results should indicate the potential applicability of embeddings as a tool for identifying emotions in text.

Data preparation: Our embeddings vectors

For the corpus from which we create word embeddings we used the complete Wikipedia 2016. After converting it to plain text, we lower cased and tokenized the corpus and build 300-dimensional vector space models were for both Word2Vec and GloVe. For both methods a context window of 15 words where used to calculate the word-word co-occurrence matrices as the training data for the models. We used its source code distribution for training GloVe vectors and Gensim for Word2Vec model. We also used the original 300-dimensional GloVe for comparison.

Results

We have done several experiments to check the effectiveness of the two word embedding methods, Word2Vec and GloVe, in emotion similarity tasks.

Experiment 1: Arithmetic of emotions – checking for closeness.

For the first experiment, the arithmetic of emotions from Table 1 was the main focus. According to this table, we expected the vectors from both sides of each equation to be relatively similar to each other, i.e. "Love = Joy + Trust", i.e. the cosine similarity between "Love" and the result of "Joy + Trust" should be close to one. On the other hand, the similarity between "Love" and "Remorse", or between their respective sums "Joy + Trust" and "Sadness + Disgust" should be relatively low (closer to zero).

Our results do not conform to these expectations. We are finding low average cosine similarity of 0.40 for GloVe and 0.39 for Word2Vec. The original GloVe vectors also had the low average value of 0.34. We also calculated the correct answer rank for each of 24 equations of Table 1 to produce the mean reciprocal ranks (MRR) for the three models. The MRRs were 0.24, 0.23, and 0.24 for GloVe, Word2Vec and the original GloVe vectors, respectively. These low MRR values seem show the incapability of these vector spaces to capture the emotional content of words (irrespective of the corpus used to train the vectors).

Experiment 2. Arithmetic of emotions – the opposites

To check how these datasets, perform in distinguishing opposite emotions, we computed the cosine similarity between each pair of the opposites. The average cosine similarities for GloVe and Word2Vec were 0.23, and 0.28; and for the original GloVe vectors was 0.20. Relatively high values for cosine similarities for all vector spaces indicate that these models underperform in discrimination between opposite emotions.

	GloVe	Word2Vec	Original GloVe
Emotion arithmetic (high values are better)	0.40	0.39	0.36
Opposites: emotions (low values are better)	0.23	0.28	0.20
Opposites: equations (low values are better)	0.44	0.32	0.38
Similarity of sub-emotions (high values are better)	0.26	0.29	0.19

Table 3. *Results from our three experiments. The average cosine similarities of the experiments show that embeddings do not properly reflect semantic similarity of emotions words.*

We then replaced each emotion with their corresponding equations ("Joy + Trust" instead of "Love") and did the experiment again. The average cosine similarity for GloVe, Word2Vec, and the original GloVe vectors was 0.44, 0.32 and 0.38 respectively.

Even though good semantic models should distinguish opposite emotions, our results confirm that co-occurrence based vector space models "cannot discriminate antonyms from synonyms" (Santus et al. 2014). Altogether, we see the inability of the standard embeddings to perform well in emotion arithmetic and in distinguishing the opposites.

Experiment 3. Similarities of in-category emotions

In the third experiment, the similarity of all secondary and tertiary emotions for each primary emotion in Parrott's three layered categorization (Table 2) was the focus. We wanted to see if the embedding of emotions in the same categories were close. As before, we calculated the pairwise cosine similarity of the embedding of the words in subcategories, for each primary emotion. As the emotions under each primary emotion are similar to each other, we expected that the average cosine similarity between each group should be a relatively high value (the closer to 1, the better). However, the average value for GloVe vectors was 0.26, and for Word2Vec was 0.29. It was even lower (0.19) for original GloVe vectors. Again the low values for the average cosine similarity show that the expected relationship between emotions are not reflected in these distributional models.

Conclusions and Future Work

In this paper, we have shown mostly negative results about word embeddings produced by the standard methods (Mikolov et al. 2013; Pennington et al. 2014): they do not follow the rules of emotions arithmetic; and they do not preserve the similarities of emotions -- in particular, we obtain higher similarity with opposite emotions than with secondary, closely related, emotions (see Table 3), which makes the models unusable for analyzing survey data.

These results contradict the expectations of the semantic arithmetic results of (Mikolov et al. 2013; Pennington et al. 2014; Dai et al. 2015), and repeated in (Evans et al. 2016).

However, we plan to continue to experiment with distributional methods in identifying emotions. One idea is to try a combination of embeddings and dictionary methods along the lines of (Changeux et al. 2006). In parallel, we are creating an annotated corpus of the surveys texts, and will be trying to use the annotations to guide semantic representations. Additionally, we are experimenting with parsing-based methods for emotions extraction. Our other options include different methods for producing embeddings, e.g. paragraph vectors (Dai et al. 2015) and paragraph vectors with additional semantic annotations (e.g. from dependency parsing), possibly using a large corpus with high frequency of emotion words.

References

Bellegarda, J.R., 2010. Emotion analysis using latent affective folding and embedding.: Proc. NAACL HLT workshop on comp. approaches to analysis and generation of emotion in text.

Changeux, J.P.P et al. 2006. *Neurobiology of human values*.: Springer.

Dai, A.M. et al. 2015. Document embedding with paragraph vectors.: *arXiv preprint arXiv:1507.07998*.

Evans, J.A. and Aceves, P., 2016. Machine Translation: Mining Text for Social Theory.: *Annual Review of Sociology*, (0).

Farzindar, A. and Inkpen, D., 2015. Natural Language Processing for Social Media. *Synthesis Lectures on Human Lang. Techn.*

Kahneman, D. and Sunstein, C.R., 2007. Indignation: psychology, politics, law. *U of Chicago Law & Economics, Olin Working Paper*: No. 346.

Kim, S.M. et al. 2010. Evaluation of unsupervised emotion models to textual affect recognition. Proc. NAACL HLT Workshop Comp. Approaches Analysis and Generation of Emotion in Text.

Mikolov, T., et al. 2013. Distributed representations of words and phrases and their compositionality. *Advances in neural information processing systems*: (pp. 3111-3119).

Pennington, J. et al. 2014. Glove: Global Vectors for Word Representation. *EMNLP*: (Vol. 14, pp. 1532-43).

Santus, E., Lu, Q., Lenci, A. and Huang, C.R., 2014. Taking Antonymy Mask off in Vector Space. In *PACLIC* (pp. 135-144).

Strapparava, C. and Mihalcea, R., 2008. Learning to identify emotions in text. Proc. ACM symposium on Applied Computing.

Seyeditabari A., Levens S., Maestas C., Walsh J., Danis C., Zadrozny W. 2017. "Cross Corpus Emotion Classification Using Survey Data". Proc. AISB Symposium on Computational Modelling of Emotion.

Vasilopoulos, P., 2015. Terrorist Events, Emotional Reactions and Political Participation. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2761740