What Sparks Joy: The AffectVec Emotion Database

Shahab Raji
shahab.raji@cs.rutgers.edu
Rutgers University
Piscataway, NJ, USA

Gerard de Melo
gdm@demelo.org
Rutgers University
Piscataway, NJ, USA

ABSTRACT

Affective analysis of textual data is instrumental in understanding human communication in the modern era of social media. A number of resources have been proposed in attempts to characterize the emotions tied to words in a text. In this work, we show that we can obtain a database that goes beyond the common binary scores for emotion classification provided by past work. Instead, we harness the power of Big Data by using neural vector space models trained with large-scale supervision from co-occurrence patterns. We modify the vector space to better account for emotional associations, which then enables us to induce AffectVec, a new emotion database providing graded emotion intensity scores for English language words with regard to a fine-grained inventory of over 200 different emotion categories. Our experiments show that AffectVec outperforms existing emotion lexicons by substantial margins in intrinsic evaluations as well as for affective text classification.

CCS CONCEPTS

• **Computing methodologies** → **Language resources**; **Lexical semantics**; *Machine learning*.

KEYWORDS

emotion lexicon, affective computing, language resources

ACM Reference Format:
Shahab Raji and Gerard de Melo. 2020. What Sparks Joy: The AffectVec Emotion Database. In *Proceedings of The Web Conference 2020 (WWW '20), April 20–24, 2020, Taipei, Taiwan.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3366423.3380068

1 INTRODUCTION

Given the well-established role of emotion in human behavior and cognition, affective analysis methods hold promise as tools to assess different forms of data and shed light on the underlying affective associations. In the fast-growing field of emotion analysis for natural language, a number of emotion databases have been developed to reveal connections between words and emotions. These resources associate words with broad emotional categories such as *joy* or *anger*, based on emotion schemes such as those put forth by Plutchik [28] (discussed in further detail in Section 2.1).

In recent years, there has been substantial progress on vector space models drawing on large-scale supervision from distributional co-occurrence patterns [18, 26]. In their raw form, such dense vector representations of words do not appear to directly store information about emotional aspects of words.

In this work, we show that they nevertheless capture valuable emotional signals implicitly. We present a simple and elegant method to adapt the vector space and then extract emotional ties of words from it, revealing, for instance, what sparks joy as opposed to what is associated with fear or sadness. Based on this, we are able to induce a new database called *AffectVec*, which provides interpretable vectors that highlight such differences. In *AffectVec*, every word is mapped to a vector, in which a given dimension quantifies the degree of association of that word with a specific emotion.

Overall, this paper makes three main contributions. First, we show that previous emotion lexicons do not correlate well with human judgments of emotion–word associations. Second, we show that we can easily obtain a better emotion database by drawing on large-scale word vector representation learning coupled with affect-specific processing. Our *AffectVec* resource covers over 200 emotions and a substantially larger English vocabulary compared to existing human-annotated databases. Finally, we carry out a series of experiments evaluating *AffectVec* in multiple regards, including showing how it can outperform previous emotion databases in unsupervised text-level emotion classification.

2 BACKGROUND AND RELATED WORK

2.1 Emotion Models

The study of emotion has a long history in psychology and evolutionary theory. Plutchik [28] presented a model that classifies human emotions based on a hierarchy. This model, also known as Plutchik's wheel of emotions, theorizes that *anger*, *fear*, *sadness*, *disgust*, *surprise*, *anticipation*, *trust*, and *joy* are the eight basic human emotions, while other emotions arise from changes in the intensity or by combination of the eight basic emotions. Ekman [10] introduced his model with a focus on facial expressions, later dubbed micro-expressions, presenting findings in support of the hypothesis that the facial expressions for the basic emotions *anger*, *fear*, *sadness*, *disgust*, *surprise*, and *joy* are universal. Thus, he based his model on these six primary emotions. Russell and Mehrabian [29] proposed the PAD model, which describes each emotion with values along three dimensions, *pleasure*, *arousal*, and *dominance*. This representation, now more commonly known as the VAD model (where V stands for *valence*), is used in a considerable number of works on emotion analysis in different languages. The idea of mapping emotions along the four axes of *pleasantness*, *sensitivity*, *aptitude*, and *attention*, visualized in an hourglass-shaped diagram, was proposed by Cambria et al. [7]. Finally, biologically inspired models define emotions with respect to reactions in systems within the human body, for example the role of the production of monoamine neurotransmitters [17] or the rate of neurons firing [34].

This paper is published under the Creative Commons Attribution 4.0 International (CC-BY 4.0) license. Authors reserve their rights to disseminate the work on their personal and corporate Web sites with the appropriate attribution.
WWW '20, April 20–24, 2020, Taipei, Taiwan
© 2020 IW3C2 (International World Wide Web Conference Committee), published under Creative Commons CC-BY 4.0 License.
ACM ISBN 978-1-4503-7023-3/20/04.
https://doi.org/10.1145/3366423.3380068

Our work, in contrast, considers a large set of over 200 emotion labels. Our data thus directly enables very fine-grained distinctions.

2.2 Emotion Lexicons

In the last decade, *affective computing*, a term introduced by Picard [27], has come to encompass a vast area of research; from analyzing and improving computational models of human emotion, creating emotional agents, to emotion analysis on different modalities of human communication.

Crowd-sourcing approaches have been used to compile databases to study the connection between words and emotions. Mohammad and Turney [23] labelled around 15,000 words with 8 basic emotions and two sentiments, providing binary tags for each word. Later work by Mohammad [22] developed a list of around 6,000 entries scored in the range of 0 to 1 for 4 emotions, reflecting the degree of emotion association. DepecheMood [32] and the improved DepecheMood++ [4] are based on simple statistical methods to create an emotion lexicon from emotionally tagged text crawled from specific Web sites. Their method starts from a crowd-based document labeling and then multiplies this data with a document–word cooccurrence matrix to acquire word-level emotion scores. WordNet-Affect [33] derives from the WordNet lexical database [12] but focuses on affective labels and their relationships. We compare against many of the aforementioned resources in our experimental evaluation.

Apart from these, there are a number of resources that provide valence, arousal, and sometimes dominance scores [21].

2.3 Vector Representations

Whilst the traditional Bag-of-Words model relies on a term–document matrix to model a corpus in terms of weighted term frequencies, in recent years, embedding models such as word2vec [18] and GloVe [26] have been used extensively to learn dense vector representations of words. The latter are mostly based on the distributional properties of words, and have proven useful in numerous semantic tasks. Unfortunately, the individual dimensions of such dense vectors do not bear any obvious inherent interpretation. Previous work has investigated embeddings in which each dimension corresponds to an emoji [31] or to a sentiment score in a specific domain [8, 9]. In *AffectVec*, we induce interpretable vectors in which each dimension captures the intensity of a specific emotional association.

Agrawal et al. [1] proposed pre-training an LSTM neural network on the task of emotion classification to learn word embeddings that better account for affective properties of words, but they do not aim at inducing an interpretable affective database. Finally, Khosla et al. [16] investigated incorporating valence, arousal, and dominance ratings into pre-existing dense word embeddings.

3 INDUCING *AFFECTVEC*

In the following, we first define our goal in Section 3.1. We then introduce our vector space modification technique in Section 3.2 and use it to induce *AffectVec* in Section 3.3. Finally, we also consider unsupervised affective text classification in Section 3.4.

3.1 Problem Definition

Our goal is to induce a new affective database that for a set of words $w \in \mathcal{V}$ from some vocabulary $\mathcal{V}$ provides affective vectors $\mathbf{v}_w \in [0, 1]^{|\Sigma|}$ for a set of affects Σ. Each such vector consists of a series of scores $\mu_{w,e} \in [0, 1]$ for affects $e \in \Sigma$ so as to quantify the strength of association of the word with the corresponding affect. For example, a word such as *party* is presumably more strongly linked to the emotion *joy* than a random word such as *screwdriver*, so the former connection should have a higher score than the latter.

Some existing resources such as EmoLex [23] provide only binary labels purporting to reveal whether or not a word is associated with an emotion. However, such decisions may be difficult to make in light of the gradedness of emotional ties. For instance, in EmoLex, *alertness* is marked as being associated with *fear*, which may hold true to a certain degree, but this association with fear presumably is weaker than those of the words *poison* or *war*. These difficulties can be overcome by instead aiming to quantify the strength of an association numerically using scores in $[0, 1]$.

Note that these affective scores may be regarded as priors, independent of specific contexts. While in general, the extent to which a word evokes a specific emotion depends on the particular context of occurrence of that word, resources that provide emotion associations at a general, decontextualized level have proven sufficiently useful to make this a worthwhile endeavour.

3.2 Sentiment-Aware Vector Space Modification

In order to obtain such an emotion database fully automatically, we draw on signals that can be extracted from massive amounts of textual data. Algorithms such as word2vec [18] and GloVe [26] are widely used to obtain dense word vectors from raw text. Cosine similarity scores between such vectors possess the property of correlating well with human judgments of semantic relatedness. Hence, we conjectured that they may also be able to reveal to what extent a given word is related to a particular emotion.

However, it turns out that the cosine similarity between a word and an emotion word in such vector spaces cannot be a particularly reliable metric to evaluate this connection. For instance, the two words *sadness* and *happiness* have a cosine similarity of 0.64 and 0.42 in the standard pretrained GloVe and Google word2vec models, respectively. While this rather high degree of relatedness intuitively makes sense, it would be wrong to conclude that the word *sadness* has happiness as one of its primary emotions.

Thus, we propose to mitigate this issue by modifying the vector space to better account for sentiment. Valence features prominently in the VAD emotion model [29]. Intuitively, we associate some emotions with positive and some emotions with negative sentiment. An et al. [3] investigated this premise in different cultures and provided a thorough analysis of the positivity and negativity of emotions. We observed a similar trend and found that there is a considerable correlation between the sentiment score of a word and the intensity of the association with an emotion. The first row in Table 4, discussed in more detail later in Section 4.3, shows a positive correlation between sentiment polarity scores (as provided by the SentiWordNet resource) and the intensity of emotions (as derived from human ratings).

RIGHTSLINK Copyright Clearance Center

Given these observations, we first obtain regular word vectors $\mathbf{v}_w \in V$ using any common word vector learning method, but then adapt the original vector space to better reflect emotional ties. This is achieved by minimizing the following objective to obtain new vectors $\mathbf{v}'_w \in V'$:

$$\begin{aligned}\ell(V, V') = &\sum_{(u,w)\in O} \max(\mathbf{0}, \varepsilon - d(\mathbf{v}'_u, \mathbf{v}'_w)) \\ + &\sum_{(u,w)\in S} d(\mathbf{v}'_u, \mathbf{v}'_w) \\ + &\sum_{(u,w)\in A} 1 - d(\mathbf{v}'_u, \mathbf{v}'_w) \\ + &\sum_{w\in\mathcal{V}} \sum_{u\in N(w)} \max(\mathbf{0}, d(\mathbf{v}'_u, \mathbf{v}'_w) - d(\mathbf{v}_u, \mathbf{v}_w)) \quad (1)\end{aligned}$$

Here, $d(\mathbf{v}, \mathbf{v}') = 1 - \cos(\mathbf{v}, \mathbf{v}')$ is the cosine distance and $\mathbf{0}$ denotes the null vector. Additionally, O is a set of word pairs with opposite sentiment polarity (e.g., *awesome* and *sadness*), A is a set of antonym word pairs with opposite meaning (e.g., *hot* and *cold*), and S is a set of word pairs with near-synonymous meaning (e.g., *party* and *celebration*). $N(w)$ is a function that yields a set of related words for w by retrieving nearest neighbours in the original vector space. Finally, ε is a hyperparameter.

Minimizing this objective function amounts to altering the vector space to better satisfy several sets of soft constraints. The first term ensures that words with opposite sentiment polarity are pulled apart in the new vector space, in terms of cosine distance. The remaining terms follow the work of Faruqui et al. [11] and Mrksic et al. [24], who proposed using synonymy and antonymy information to post-process word vectors. This can help us even further because the sentiment polarity information can be rather sparse. For example, we may pull apart *happy* and *sad*, but additional synonymy information can reveal that words such as *glad* and *elated* should behave similarly to *happy*.

The final term moderates the alteration of the vector space by ensuring that the modified vectors overall do not have radically different distances compared to the original vectors. For this, we define $N(w) = \{u \in \mathcal{V} \mid d(u, w) \leq \beta\}$ and adopt $\beta = 0.2$ following Mrksic et al. [24].

3.3 Interpretable Vector Induction

Based on the modified vector space with its sentiment-aware vectors $\mathbf{v}'_w$, we can then assess the connection between words and emotions more reliably. For this, we assume every particular emotion $e \in \Sigma$ is associated with a corresponding emotion word w_e.

For every word $w \in \mathcal{V}$, we create an interpretable *AffectVec* vector of the form

$$\mathbf{v}^{\Sigma}_w = (\mu_{w,e_1} \dots \mu_{w,e_n})^{\intercal} \in \mathbb{R}^{|\Sigma|}, \quad (2)$$

where each $\mu_{w,e_i} = \cos(\mathbf{v}'_w, \mathbf{v}'_{w_{e_i}})$.

Each dimension i in this vector space is associated with a particular emotion $e_i \in \Sigma$, and a corresponding entry in $\mathbf{v}^{\Sigma}_w$ quantifies to what extent the word w is associated with the emotion word for e_i, taking values in $[0, 1]$.

These vectors can then be used in a variety of ways. For instance, in order to obtain text-level emotion intensity predictions, one can feed such vectors into a deep neural network. We explore this in Section 4.4.

3.4 Unsupervised Affective Text Classification

Another option is to use the *AffectVec* database for unsupervised affective text classification. Most work on emotion classification is supervised, i.e., it assumes that one has a large in-domain training dataset that can be used to train a supervised learning model such as a convolutional neural network. Often, however, such training data is not available.

Using simple statistical techniques, we can instead make use of a database such as *AffectVec* to predict affective labels for a text without training data. To this end, we simply predict

$$\hat{e} = \operatorname*{arg\,max}_{e\in\Sigma_C} \sum_{w\in T} \lambda_w \, \mu_{w,e} \quad (3)$$

as the emotion label for an input text T. Here, $\Sigma_C \subseteq \Sigma$ is a relevant subset of emotion labels that are considered in the respective emotion classification task, λ_w are word-specific weights such as TF-IDF scores, and $\mu_{w,e}$ is the *AffectVec* score for word w with regard to emotion e. We thus compute a weighted sum of word emotion scores to obtain text-level predictions and predict the emotion attaining the highest such score. Experimental results on this approach are given in Section 4.5.

4 EXPERIMENTS

4.1 Vector Space Modification

For our sentiment-aware vector space modification, we first collect the relevant constraints based on sentiment polarity information as well as resources providing semantic relationships.

Input Word Vectors. We use the Paragram-SL999 vectors [35] as our input vectors, as these were trained to have a high quality in terms of word similarity ratings.

Sentiment Polarity Constraints. In order to create the set of word pairs O in Eq. 1 to serve as soft constraints, we consider words with opposite polarity in the NRC Valence, Arousal, Dominance Lexicon [21]. In our later experiments, we also compare using SentiWordNet [6] and VADER [15] to obtain such constraints. These resources all provide real-valued sentiment scores, but the respective scales differ, so we first rescale them to the range $[-1, 1]$. Moreover, in the case of SentiWordNet, we take the average of all word sense-specific scores that it provides for a given word. Subsequently, to identify pairs of words with opposite sentiment for inclusion in O, we do not simply consider the midpoint 0, as two words might both have scores with a very low magnitude. Rather, we apply a threshold of 2ε on the difference between the sentiment scores of each pair of words to eliminate a large number of neutral words and ensure a sufficiently polarized set of opposites (recall that ε is the hyperparameter from Eq. 1).

SentiWordNet, with 117,600 words and their respective sentiment scores, provides the largest number of opposite pairs. However, as we later find in the results in Table 4, VADER and the NRC Valence dataset, despite having a much smaller number of constraints, give rise to a higher-quality vector space.

Synonymy and Antonymy Information. We consult PPDB [25] and WordNet [12] to compile sets of antonyms and synonyms

RIGHTSLINK

Table 1: Spearman correlation with SimLex-999 and Mturk-771 word pair ratings

Vectors	SimLex-999	Mturk-771
word2vec	0.442	0.671
GloVe	0.408	0.715
Paragram-SL999 [35]	0.685	0.719
Counterfitted vectors [24]	0.735	0.708
Our Approach	0.745	0.696

for each of these resources that can then be used as additional constraints in S and A from Eq. 1.

Optimization. For optimization, we rely on stochastic gradient descent. We set $\varepsilon = 0.2$, which empirically appeared to work well in our comparison with human judgments. For a fair comparison, this value was not adjusted any further for the downstream unsupervised text classification task.

Evaluating Semantic Similarity. Although the purpose of our method ultimately is to capture the emotional ties of words, we first assessed to what extent the resulting distortions of the vector space might have altered the semantic similarity relationships. Two particularly prominent large resources used for word-level relatedness evaluation are SimLex-999 [14] and Mturk-771 [13]. The former measures semantic similarity, while the latter measures arbitrary semantic relatedness. The results are given in Table 1. Our method shows a better Spearman correlation with the similarity ratings in SimLex-999 than alternative word vectors. On the Mturk-771 dataset, in contrast, the results are roughly on a par with other word vectors. Hence, our method maintains a similar level of semantic relatedness, while yielding improved semantic similarity ratings.

4.2 *AffectVec* Emotion Vectors

Subsequently, we use our method to produce vectors providing affect intensities for a large set of fine-grained emotions, with much finer-grained distinctions than considered in previous work. This database can be used directly to find the emotional association between words and a large number of emotions.

Fine-Grained Emotion Set. To obtain a comprehensive set Σ of target affects, we take the union of sets of emotions drawn from the resources WordNet-Affect [33], the HUMAINE model [5], and WordNet [12]. WordNet-Affect and HUMAIN both present a hierarchy of emotional words, from general coarse-grained emotions to more fine-grained ones at lower levels. However, there are several entries for a single emotion, referring to the different contexts and situations in which the emotion arises, e.g., positive-fear, negative-fear. We merge such instances and keep only the main emotion word. To identify emotion words within WordNet, a large lexical database, we consider the hyponym tree of the word *emotion*, which results in a set of candidate affect words. The union of the three sets consists of 239 target affects.

Coverage. Statistics for the coverage of the resulting *AffectVec* database are given in Table 2. Our resource has a significantly larger vocabulary than current human-annotated affect lexicons, though automatically induced ones may be larger. However, consider that *AffectVec* covers by far more emotions than previous work, which typically adopts a set of 4–7 emotions.

Table 2: Coverage of *AffectVec*

Metric	Count
Number of emotions	239
Number of words	76,427
Number of word – emotion pairs	18,266,053

4.3 Comparison with Human Judgments

To ascertain the quality of *AffectVec*, we compare the intensity scores that it provides for word–emotion pairs against real-valued scores stemming from human annotation.

Ground Truth Data. The NRC Affect Intensity lexicon [22] provides intensity scores for the emotions *anger*, *fear*, *sadness*, and *joy*, based on annotations by human judges. This lexicon covers around 1,500 word n-grams for each of the aforementioned emotions, with scores representing the strength of association between the n-gram and emotion. Since we study word representations, we remove bigrams, resulting in a total of 5,616 word–emotion pairs with scores.

Evaluation Metric. Following SemEval 2018: Task 1 [20], we use the Pearson correlation to evaluate to what extent the various emotion databases correlate with the ground truth human assessments of emotion intensity.

Baselines. We consider the following sentiment and emotion resources for comparison (see Table 3 for statistics).

SentiWordNet [6] is a sentiment lexicon based on WordNet providing real-valued polarity scores for specific word senses of words. Since a word can have multiple senses, we calculate the sentiment score by taking the average of the sentiment scores for all of its senses listed in SentiWordNet.

DepecheMood [32] is a prominent automatically derived emotion lexicon that provides eight real-valued scores for each word, corresponding to *anger*, *fear*, *sadness*, *joy*, *amused*, *annoyed*, *inspired*, and *neutral*. Among these, we consider only the first four, as our ground truth data does not cover the remaining emotions.

DepecheMood++ [4] is an extension of DepecheMood that draws on more advanced statistical techniques and a larger corpus.

EmoLex [23] is a widely used emotion lexicon obtained using crowdsourcing. It covers 14k unigrams and their binary association with eight emotions, as well as positive and negative sentiment.

Pre-trained Word Vectors as well serve as a baseline, following the discussion in Section 3.2. We consider the cosine similarity of word–emotion pairs in word2vec trained on the Google News corpus [18], GloVe [26] trained on Twitter (200-dim.) and CommonCrawl (840B, 300-dim.), as well as the counterfitted vectors by Mrksic et al. [24].

Results. Our principal results are given in Table 4. The first row shows that there is a positive correlation between sentiment polarity scores and the ground truth data for a number of emotions, though most pronounced for joy.

Next, we consider DepecheMood, DepecheMood++, and EmoLex as prominent emotion lexicons. Surprisingly, one can substantially

Table 3: Details of datasets

Dataset	Granularity	Size	Values	Categories
SentiWordNet [6]	words	117K	graded	sentiment polarity
VADER [15]	words	7,500	graded	sentiment polarity
NRC VAD dataset [21]	words	20K	graded	valence, arousal, dominance
DepecheMood [32]	words	37K	graded	7 emotions+neutral
DepecheMood++ [4]	words	187K	graded	7 emotions+neutral
EmoLex [23]	words	14K	binary	8 emotions+polarity scores
AffectData [2]	sentences	1,207	binary	5 emotions
ISEAR [30]	sentences	7,666	binary	7 emotions
WASSA 2017 Text Emotion Intensity Task [19]	tweets	7,102	graded	4 emotions

Table 4: Results of comparison with human judgments (Pearson correlation coefficients)

Method	Anger	Fear	Sadness	Joy	Overall
SentiWordNet [6]	0.164	0.157	0.227	0.404	0.225
DepecheMood [32]	0.120	0.094	0.210	0.010	0.110
DepecheMood++ [4]	0.173	0.168	0.308	0.149	0.191
EmoLex [23]	0.009	0.154	0.060	0.048	0.066
Our Method (word2vec)	0.309	0.232	0.314	0.568	0.334
Our Method (GloVe Twitter 200-dim.)	0.135	0.071	0.377	0.181	0.166
Our Method (GloVe CC 840B 300-dim.)	0.343	0.242	0.317	0.531	0.333
Our Method (Counterfitted Vectors)	0.362	0.293	0.398	0.566	0.390
Our Method (SentiWN constraints)	0.365	0.292	0.427	0.522	0.387
Our Method (VADER constraints)	0.434	0.377	0.464	0.642	0.463
Our Method (NRC valence constraints)	**0.480**	**0.524**	**0.568**	**0.692**	**0.551**

outperform such state-of-the-art emotion resources by applying our proposed method of exploiting cosine similarity scores even with commonly available pre-trained word vectors such as Google's word2vec and GloVe and obtain word–emotion scores. The same technique with counterfitted vectors [24] shows a fairly consistent improvement across all emotions and the union of emotions.

However, invoking our method of considering cosine similarities in conjunction with our vector space modification technique yields the best results, as we obtain substantially higher Pearson correlations with the human ground truth ratings. This is particularly true with the default setup of using NRC valence constraints, but also holds true when using the small VADER lexicon for constraints. Merely the SentiWordNet resource does not seem to provide sufficiently high-quality valence information, perhaps because it was derived semi-automatically.

4.4 Text-Level Emotion Intensity Prediction

As a downstream task, we consider how our word-level emotion database helps for text-level emotion intensity prediction. Specifically, we train a deep neural model for tweet emotion intensity prediction. The model consists of a CNN layer with 64 filters and a kernel size of 5, a pooling layer of size 4, and an LSTM layer with 100 units. The LSTM layer is finally connected to a dense layer with mean squared error as loss function to predict the intensity score. We train the model on the tweet emotion intensity scores provided for SemEval 2018 [20]. In the dataset, for each of the four target emotions, 1,500 to 2,200 sentences are given along with an emotion intensity rating. We rely on a train–validation–test split ratio of 70%/10%/20%.

The Pearson correlation results are given in Table 5. We compare providing just regular word vectors (word2vec or GloVe) as an embedding layer for the model (Reg), against augmented word vectors, for which we concatenate *AffectVec* to the regular word vectors (Reg⊕Aff). We observe that the latter option improves the results significantly.

4.5 Unsupervised Text Classification

Next, we assess how different emotion databases fare on the task of unsupervised text classification, considering several affective classification benchmarks. Although most work on emotion classification is supervised, i.e., it is assumed that one has a large in-domain training dataset that can be used to train a supervised learning model such as a convolutional neural network, *AffectVec* allows for unsupervised affective text classification.

Method. For this, we rely on the method from Section 3.4 to obtain a text-level emotion label. In other words, for all words in the text document, we consult the emotion database to obtain their respective scores for a given emotion, and compute weighted sums of such scores over all words in the text. Finally, the emotion with the highest score is selected as the classification. In case of a tie,

Table 5: Text-level emotion intensity prediction using regular word vectors (Reg) and after concatenation with *AffectVec* word vectors (Reg⊕Aff), evaluated using Pearson Correlation, with relative improvements given in percent

Method	Anger		Fear		Sadness		Joy		Overall	
	Reg	Reg⊕Aff	Reg	Reg⊕Aff	Reg	Reg⊕Aff	Reg	Reg⊕Aff	Reg	Reg⊕Aff
word2vec	0.669	+7%	0.693	+8%	0.682	+8%	0.677	+9%	0.681	+7%
GloVe	0.629	+9%	0.617	+15%	0.556	+22%	0.574	+19%	0.593	+13%

an emotion was chosen randomly in our experiments. As preprocessing, we lower-cased all words. Out-of-vocabulary words are assumed to have an emotion score of 0.

Datasets. We consider several benchmark datasets for this (see also Table 3).

Affect Data: Alm [2] provides a corpus of about 1,200 sentences from three story collections, labeled with five emotions.

ISEAR: From a large study of the emotional responses to different situations [30], there is a corpus of around 7,600 sentences (situations) and the corresponding emotions. This dataset is labelled with seven emotions.

WASSA Shared Task for Emotion Intensity: Mohammad and Bravo-Marquez [19] created a dataset of tweets labelled with the strength of emotional association for four emotions. We convert their intensity scores in $[0, 1]$ to a classification task by considering the emotion with the highest score as the label if said score is ≥ 0.5. Tweets not assigned a label are eliminated.

Results. The results in terms of the classification accuracy are given in Table 6. The random baseline chooses an emotion class entirely randomly. All other results are for our unsupervised method from Section 3.4, either with uniform word weights λ_w, or with TF-IDF weights as λ_w. We compare different word-level emotion scores as input to the method, including DepecheMood (DM), DepecheMood++ (DM++), EmoLex, and cosine similarity computed on different pre-trained word vectors. *AffectVec* outperforms the baselines, achieving superior results compared to the state-of-the-art emotion databases as well as compared to the cosine similarity on regular word vectors.

Table 6: Unsupervised text classification accuracy

Method		Affect Data	ISEAR	WASSA
Random		0.200	0.143	0.250
Unweighted	DM	0.401	0.180	0.328
	DM++	0.405	0.238	0.369
	EmoLex	0.448	0.255	0.514
	word2vec	0.331	0.239	0.450
	GloVe	0.136	0.185	0.383
	AffectVec	0.473	0.283	0.510
TF-IDF	DM	0.395	0.185	0.327
	DM++	0.408	0.237	0.352
	EmoLex	0.445	0.261	0.496
	word2vec	0.400	0.255	0.465
	GloVe	0.162	0.204	0.400
	AffectVec	0.483	0.313	0.522

4.6 Qualitative Analysis

Generalization for out of vocabulary words. *AffectVec*'s coverage is fairly large compared to previous emotion databases. Still, there are out-of-vocabulary words. To address these cases, we can consider the neighbourhood of such words using supplementary vector similarities and obtain a score via those neighbors. As an example, the word *annihilated* has a high association with the emotion *anger*, 0.898, in the ground truth. This is not a genuine out-of-vocabulary word, as *AffectVec* provides a predicted score of 0.702. However, if it were an out-of-vocabulary word, we could still consider top-10 neighbours in some alternative vector space that includes this word. If we use our modified vectors for this and average the *anger* scores for the top ten most similar words to "annihilated", we obtain an average of 0.736.

Greater Coverage of Emotions. While previous works on computational emotion analysis mainly focused on four or eight emotions, *AffectVec* covers a large number of emotions that in some cases only slightly differ with respect each other (e.g., *wrath*, *anger*, and *fury*). However, these words nonetheless appear in rather different contexts with different connotations. Thus, it can be of value to provide separate scores for different fine-grained emotions.

In addition, comparing to other distributed representation vectors, *AffectVec*'s vectors are interpretable on each dimension.

5 CONCLUSION

This work introduces *AffectVec*, a high-coverage resource for fine-grained emotion analysis. *AffectVec* provides graded assessments of emotion ties for words showing a greater correlation with human judgments than previous state-of-the-art resources. Indeed, our experiments reveal that current lexicons show a surprisingly low correlation with human-solicited ground truth ratings. In addition, we show why simple semantic relatedness techniques do not always capture the emotional associations of words. To address this, we present a sentiment-aware vector space modification technique that subsequently allows us to induce an emotion database covering a large set of 239 emotions, which is orders of magnitude more than previous work. Finally, we show how *AffectVec* can successfully be used for text-level emotion intensity prediction and for unsupervised affective text classification. AffectVec is available for download online[1].

REFERENCES

[1] Ameeta Agrawal, Aijun An, and Manos Papagelis. 2018. Learning Emotion-enriched Word Representations. In *Proceedings of the 27th International Conference*

[1] http://emotion.nlproc.org

RIGHTSLINK

on Computational Linguistics. Association for Computational Linguistics, Santa Fe, New Mexico, USA, 950–961. https://www.aclweb.org/anthology/C18-1081

[2] Ebba Cecilia Ovesdotter Alm. 2008. AFFECT IN TEXT AND SPEECH.

[3] Sieun An, Li-Jun Ji, Michael Marks, and Zhiyong Zhang. 2017. Two sides of emotion: exploring positivity and negativity in six basic emotions across cultures. *Frontiers in psychology* 8 (2017), 610.

[4] Oscar Araque, Lorenzo Gatti, Jacopo Staiano, and Marco Guerini. 2018. DepecheMood++: a Bilingual Emotion Lexicon Built Through Simple Yet Powerful Techniques. *arXiv preprint arXiv:1810.03660* (2018).

[5] HUMAINE Association et al. 2006. Humaine emotion annotation and representation language (earl): Proposal.

[6] Stefano Baccianella, Andrea Esuli, and Fabrizio Sebastiani. 2010. SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining. In *in Proc. of LREC*.

[7] Erik Cambria, Andrew Livingstone, and Amir Hussain. 2012. The Hourglass of Emotions. In *Proceedings of the 2011 International Conference on Cognitive Behavioural Systems* (Dresden, Germany) *(COST'11)*. Springer-Verlag, Berlin, Heidelberg, 144–157.

[8] Xin Dong and Gerard de Melo. 2018. Cross-Lingual Propagation for Deep Sentiment Analysis. In *Proceedings of the 32nd AAAI Conference on Artificial Intelligence (AAAI 2018)* (New Orleans, LA, USA). AAAI Press, 5771–5778.

[9] Xin Dong and Gerard de Melo. 2018. A Helping Hand: Transfer Learning for Deep Sentiment Analysis. In *Proceedings of ACL 2018* (Melbourne). 2524–2534. https://doi.org/10.18653/v1/P18-1235

[10] Paul Ekman. 1992. An argument for basic emotions. *Cognition and Emotion* 6, 3-4 (1992), 169–200. https://doi.org/10.1080/02699939208411068 arXiv:https://doi.org/10.1080/02699939208411068

[11] Manaal Faruqui, Jesse Dodge, Sujay Kumar Jauhar, Chris Dyer, Eduard H. Hovy, and Noah A. Smith. 2014. Retrofitting Word Vectors to Semantic Lexicons. *CoRR* abs/1411.4166 (2014). arXiv:1411.4166 http://arxiv.org/abs/1411.4166

[12] Christiane Fellbaum (Ed.). 1998. *WordNet: an electronic lexical database.* MIT Press.

[13] Guy Halawi, Gideon Dror, Evgeniy Gabrilovich, and Yehuda Koren. 2012. Large-scale Learning of Word Relatedness with Constraints. In *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (Beijing, China) *(KDD '12)*. ACM, New York, NY, USA, 1406–1414. https://doi.org/10.1145/2339530.2339751

[14] Felix Hill, Roi Reichart, and Anna Korhonen. 2014. SimLex-999: Evaluating Semantic Models with (Genuine) Similarity Estimation. *CoRR* abs/1408.3456 (2014). arXiv:1408.3456 http://arxiv.org/abs/1408.3456

[15] Clayton J Hutto and Eric Gilbert. 2014. Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Eighth international AAAI conference on weblogs and social media*.

[16] Sopan Khosla, Niyati Chhaya, and Kushal Chawla. 2018. Aff2Vec: Affect-Enriched Distributional Word Representations. *CoRR* abs/1805.07966 (2018). arXiv:1805.07966 http://arxiv.org/abs/1805.07966

[17] Hugo Lövheim. 2012. A new three-dimensional model for emotions and monoamine neurotransmitters. *Medical Hypotheses* 78, 2 (2012), 341 – 348. https://doi.org/10.1016/j.mehy.2011.11.016

[18] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013. Distributed Representations of Words and Phrases and their Compositionality. In *Advances in Neural Information Processing Systems 26*, C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger (Eds.). Curran Associates, Inc., 3111–3119. http://papers.nips.cc/paper/5021-distributed-representations-of-words-and-phrases-and-their-compositionality.pdf

[19] Saif Mohammad and Felipe Bravo-Marquez. 2017. WASSA-2017 Shared Task on Emotion Intensity. In *Proceedings of the 8th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*. Association for Computational Linguistics, Copenhagen, Denmark, 34–49. https://doi.org/10.18653/v1/W17-5205

[20] Saif Mohammad, Felipe Bravo-Marquez, Mohammad Salameh, and Svetlana Kiritchenko. 2018. Semeval-2018 task 1: Affect in tweets. In *Proceedings of The 12th International Workshop on Semantic Evaluation*. 1–17.

[21] Saif M. Mohammad. 2018. Obtaining Reliable Human Ratings of Valence, Arousal, and Dominance for 20,000 English Words. In *Proceedings of The Annual Conference of the Association for Computational Linguistics (ACL)*. Melbourne, Australia.

[22] Saif M. Mohammad. 2018. Word Affect Intensities. In *Proceedings of the 11th Edition of the Language Resources and Evaluation Conference (LREC-2018)*. Miyazaki, Japan.

[23] Saif M. Mohammad and Peter D. Turney. 2013. Crowdsourcing a Word–Emotion Association Lexicon. 29, 3 (2013), 436–465.

[24] Nikola Mrksic, Diarmuid Ó Séaghdha, Blaise Thomson, Milica Gasic, Lina Maria Rojas-Barahona, Pei-Hao Su, David Vandyke, Tsung-Hsien Wen, and Steve J. Young. 2016. Counter-fitting Word Vectors to Linguistic Constraints. *CoRR* abs/1603.00892 (2016). arXiv:1603.00892 http://arxiv.org/abs/1603.00892

[25] Ellie Pavlick, Pushpendre Rastogi, Juri Ganitkevitch, Benjamin Van Durme, and Chris Callison-Burch. 2015. PPDB 2.0: Better paraphrase ranking, fine-grained entailment relations, word embeddings, and style classification. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)* (Beijing, China). Association for Computational Linguistics, 425–430. https://doi.org/10.3115/v1/P15-2070

[26] Jeffrey Pennington, Richard Socher, and Christopher D. Manning. 2014. GloVe: Global Vectors for Word Representation. In *Empirical Methods in Natural Language Processing (EMNLP)*. 1532–1543. http://www.aclweb.org/anthology/D14-1162

[27] Rosalind W. Picard. 1997. *Affective Computing*. MIT Press, Cambridge, MA, USA.

[28] Robert Plutchik. 1980. a General Psychoevolutionary Theory of Emotion. In *Theories of Emotion*. Elsevier, 3–33. https://doi.org/10.1016/B978-0-12-558701-3.50007-7

[29] James A Russell and Albert Mehrabian. 1977. Evidence for a three-factor theory of emotions. *Journal of Research in Personality* 11, 3 (1977), 273 – 294. https://doi.org/10.1016/0092-6566(77)90037-X

[30] Klaus R Scherer and Harald G Wallbott. 1994. Evidence for universality and cultural variation of differential emotion response patterning. *Journal of personality and social psychology* 66, 2 (1994), 310.

[31] Abu Awal Md Shoeb, Shahab Raji, and Gerard de Melo. 2019. EmoTag – Towards an Emotion-Based Analysis of Emojis. In *Proceedings of RANLP 2019* (Varna, Bulgaria). 1094–1103. https://doi.org/10.26615/978-954-452-056-4_126

[32] Jacopo Staiano and Marco Guerini. 2014. Depeche Mood: a Lexicon for Emotion Analysis from Crowd Annotated News. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, Vol. 2. 427–433.

[33] Carlo Strapparava and Alessandro Valitutti. 2004. WordNet-Affect: an Affective Extension of WordNet. *Vol 4*. 4 (01 2004).

[34] Silvan S Tomkins and Robert McCarter. 1964. What and where are the primary affects? Some evidence for a theory. *Perceptual and motor skills* 18, 1 (1964), 119–158.

[35] John Wieting, Mohit Bansal, Kevin Gimpel, and Karen Livescu. 2015. From paraphrase database to compositional paraphrase model and back. *Transactions of the Association for Computational Linguistics* 3 (2015), 345–358.