Decision support with text-based emotion recognition: Deep learning for affective computing

Bernhard Kratzwald[b], Suzana Ilic[a,*], Mathias Kraus[b], Stefan Feuerriegel[b], Helmut Prendinger[a]

[a] *National Institute of Informatics, 2-1-2 Hitotsubashi, 101-8430 Tokyo, Japan*
[b] *ETH Zurich, Weinbergstr. 56/58, 8092 Zurich, Switzerland*

Abstract

Emotions widely affect the decision-making of humans and, hence, affective computing takes emotional states into account with the goal of tailoring decision support to individuals. However, the accurate recognition of emotions within narrative materials presents a challenging undertaking due to the complexity and ambiguity of language. Even though deep learning has evolved as the state-of-the-art in various tasks from text mining, its benefits with regard to affective computing are not yet understood. We thus propose the following innovations: (1) we adapt recurrent neural networks from the field of deep learning to affective computing. (2) We extend these networks for predicting the score of different affective dimensions. (3) We implement transfer learning for pre-training word embeddings. Analyzing the results, we find that deep learning consistently outperforms traditional machine learning with improvements of up to 21 % in F1-score when labeling emotions and 6 % in forecast errors when rating the intensity of emotions. Altogether, the findings have considerable implications for the use of affective computing in providing decision support.

Keywords: Affective computing, Emotion recognition, Deep learning, Natural language processing, Text mining, Decision support

*Corresponding author.

Email addresses: bkratzwald@ethz.ch (Bernhard Kratzwald), suzana.ilic.nii@gmail.com (Suzana Ilic), mathiaskraus@ethz.ch (Mathias Kraus), sfeuerriegel@ethz.ch (Stefan Feuerriegel), helmut@nii.ac.jp (Helmut Prendinger)

1. Introduction

Emotions drive the ubiquitous decision-making of humans in their everyday lives [1, 2, 3]. Furthermore, emotional states can even implicitly affect human communication, attention and the personal ability to memorize information [4, 5]. While the recognition and interpretation of emotional states often comes naturally to humans, these tasks pose severe challenges to computational routines [e. g. 6, 7]. As such, the term *affective computing* refers to techniques that detect, recognize and predict human emotions (e. g. joy, anger, sadness, trust, surprise, anticipation) with the goal of adapting computational systems to them [8]. The resulting computer systems are not only capable of developing empathy [9] but can also provide decision support tailored to the emotional state of individuals.

Emotional information is conveyed by a multiplicity of physical and physiological characteristics. Examples of such indicators include vital signs, such as heart rate, muscle activity or sweat production on the skin surface [10]. Recent advances in neuroscience have also shown the ability to link neural oscillations from electroencephalography [11], EEG for short, or stimulation measured by NeuroIS to emotional experiences [12, 13]. A different stream of research tries to infer emotions from the content and its communication. These approaches to affective computing are primarily categorized by the modality of the message, i. e. whether it takes the form of speech, gesture or written information [14]. In this terminology, affective computing can comprise unimodal and multimodal analyses. For instance, videos allow for the recognition of facial expressions or vocal tone [15, 16, 17].

The focus of this work is on unimodal analysis of written material in English. This choice reflects the prominence of textual materials as a widespread basis for decision support [18, 19]. Illustrative example are as follows (a detailed review is given later in section 2.3). For instance, the use of affective language as a proxy for emotional closeness can be used to measure tie strength in social networks [20]. Similarly, marketing utilizes the recognition of emotional states in order to predict purchase intentions of customers [21], satisfaction with services [22], and even to measure the overall brand reputation [23]. In a related context, decision support can leverage affective signals in financial materials in order to suggest trading decisions [24], or forecast the economic climate [25]. Furthermore, affect can also improve processes and decision-making in the provision of health-care [26] or education [27]. Each of the above applications requires a different representation of emotions as defined by the underlying affect theory (see section 2.1 for a detailed overview).

Previous research on affective computing has merely utilized methods from traditional machine learning, while recent advances from the field of deep learning – namely, recurrent neural networks and transfer learning – have been overlooked. However, their use promises further improvements for affect-aware decision support. In fact, techniques from deep learning have become prominent in various decision support activities involving sequential data [e. g. 28] and especially linguistic materials [e. g. 29], where deep learning was able to enhance the performance when deriving decisions from unstructured data. One of the inherent advantages of deep learning is that it can successfully model highly non-linear relationships [30]. In addition, traditional machine learning largely relies upon bag-of-words and thus ignores the order of words in a text, whereas recurrent neural networks process documents word-by-word in order to compute a low-dimensional representation that can incorporate word order and long term dependencies [31].

This work contributes the following innovations to affect-aware decision support. First, we overcome the inherent limitations of bag-of-words from traditional machine learning and propose the use of recurrent neural networks from the field of deep learning for affective computing. More precisely, we utilize long short-term memory networks and gated recurrent units that can make predictions from running texts of varying lengths. Second, it is common in affective computing to measure not a single affective dimension (e. g. positive-negative) but rather infer a composition of multiple dimensions (e. g. anger, happiness, trust, surprise, anticipation). Accordingly, we change the network layout such that it supports multi-label outcomes. Third, we experiment with transfer learning as a strategy to further improve the accuracy of emotion recognition. This concept pre-trains word embeddings on a related but different dataset, thereby enlarges the dataset. This is known to help the training process in the context of deep learning, which often benefits from large datasets to fine-tune the vast number of degrees-of-freedom.

Even though affective computing has received great traction over the past years [32], there is a scarcity of widely-accepted datasets for text-based emotion recognition that can be used for benchmarking and that facilitate fair comparisons. A relatively small, but more common, dataset was provided by SemEval-2007 and consists of annotated news headlines [33]. A significantly larger, but underutilized, corpus is composed of affect-labeled literary tales [34]. Our literature review notes considerable differences across datasets that vary in their linguistic style, domain, affective dimensions and the structure of the outcome variable. With regard to the latter, the majority of datasets involves a classification task in which exactly one affective category is assigned to a document, while others request a numerical score across multiple dimensions, i. e. a regression task. Hence, it is a by-product of this research to contribute a holistic comparison that benchmarks different methods across datasets used in prior research. For this purpose, we conducted an extensive search for affect-labeled datasets that serves as the foundation for our computational experiments. As a result, we find that deep learning consistently outperforms the baselines from traditional machine learning at a statistically significant level. In fact, we observe performance improvements of up to 21 % in classification f1-score and 6 % in root mean squared forecast error.

The findings of this work have direct implications for management, practice and research. As such, various application areas - such as customer support, marketing or recommender systems - can be improved considerably through the use of affective computing. Similarly, all systems with human-computer interactions could further benefit from emotion recognition and a deeper understanding of empathy (e. g. chatbots and personal assistants). In fact, emotion detection could significantly impact and refine all use cases in which sentiment analysis (i. e. only positive/negative polarity) has already proved to be a valuable technology, since these lend themselves to a more fine-grained breakdown and decision-making beyond only one dimension. In academia, text-based emotion recognition supports the cognitive and social sciences as a new approach to measuring and interpreting individual and collective emotional states.

The rest of this paper is structured as follows. Section 2 reviews earlier works on text-based emotion recognition, including the underlying affect theories, datasets used for benchmarking and computational approaches. This reveals a research gap with regard to both deep neural networks and transfer learning within the field of affective computing. As a remedy, Section 3 introduces our methods rooted in deep learning, which are then evaluated in Section 4. Based on our findings, we detail implications for both research and management in Section 5, while Section 6 concludes.

2. Background

We specifically point out that the terms "sentiment analysis" and "affective computing" are often used interchangeably [35]. However, comprehensive surveys [36, 37] recognize clear differences that distinguish both concepts: sentiment analysis measures the subjective polarity towards entities in terms of only two dimensions, namely, positivity and negativity. Conversely, affective computing concerns the identification of explicit emotional states and, hence, this approach is also referred to as emotion recognition. For reasons of clarity, we strictly distinguish between the aforementioned concepts in our terminology. The choice of emotional dimensions depends on the underlying affect theory and involves a wide range of mental states such as, happiness, anger, sadness or fear. In this regard, the bivariate polarity of sentiment can refer to a subset of emotions in affective computing.

Accordingly, this section first provides an overview of prevalent emotion models as specified by affect theories and, based on their dimensions, reviews computational methods for inferring affective information from natural language. This gives rise to a variety of use cases, which are detailed subsequently.

2.1. Affect theory

In the field of psychology, there is no consensus regarding a universal classification of emotions [38, 39], as physiological arousal in the proposed theories varies with causes, cognitive appraisal processes and context. Yet a conventional approach is to distinguish emotions based on how the underlying constructs are defined. On the one hand, emotions can be defined as a set of discrete states with mutually-exclusive meanings, while, on the other hand, emotions can also be specified by a combination of numerical dimensions, each associated with a rating of intensity. The categorization into either a discrete set or a combination of intensity labels yields later benefits with regard to computational implementations, as it directly helps in formalizing the different machine learning models.

Categorical emotion models involve a variety of prevalent examples, including the so-called basic emotions. These introduce a discrete set of emotions with innate and universal characteristics [40, 41]. One of the first attempts by Ekman and Friesen [42] to classify emotions led to the categorization of six discrete items labeled as basic: namely, anger, disgust, fear, happiness, sadness and surprise.The model was later extended by Averill et al. [43] to include trust and anticipation, resulting in eight basic emotions. An alternative categorization by Tomkins [40, 44] classifies nine primary affects into positive (enjoyment, interest), neutral (surprise) and negative (anger, disgust, dissmell, distress, fear, shame) expressions.

Dimensional models of emotion locate constructs in a two- or multi-dimensional space [6]. Here the assumption of disjunct categories is relaxed such that the magnitude along each dimension can be measured separately [45], yielding continuous intensity scores. Different variants have been proposed, out of which we summarize an illustrative subset in the following (cf. Figure 1). One of the earliest examples is Russell's circumplex model [45] consisting of bivariate classification into valence and arousal. Depending on the strength of both components, certain regions in the two-dimensional space are given explicit interpretations (such as tense, aroused, excited) according to 28 emotional states. The Wheel of Emotions is an extension of the circumplex model whereby eight primary emotion dimensions are represented as four pairs of opposites: joy versus sadness, anger versus fear, trust versus disgust, and surprise versus anticipation [46]. Recent approaches introduce complex hybrid emotion models, such as the Hourglass of Emotions [47], which represents affective states through both discrete categories and four independent, but concomitant, affective dimensions. However, neither the Wheel of Emotions nor the Hourglass of Emotions has not yet found its way into common datasets for training algorithms from affective computing.

(a) Cirumplex model (b) Cube of emotion

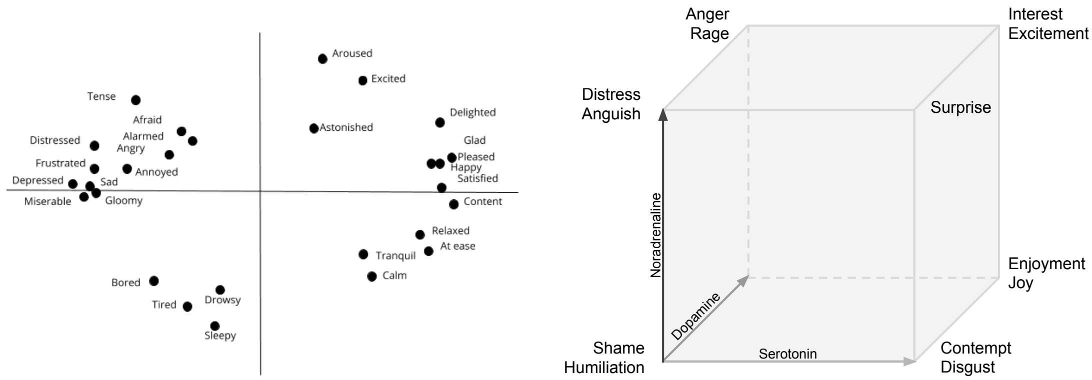

Figure 1: Schematic illustration of prevalent dimensional models of emotion. (a) Russell's circumplex model of affect [45], which introduces two dimensions, namely, valence and arousal. Certain regions in this space are then given labels according to 28 emotional states. (b) Lövheim cube of emotion.

2.2. Datasets for benchmarking

Table 1 provides a holistic overview of datasets used for text-based affective computing. These datasets exhibit fundamentally different characteristics and challenges, as they vary in size, domain, linguistic style and underyling affect theory. We summarize key observations in the following.

In terms of text source, the datasets refer to tasks that utilize narrative materials from classic literature [34], while others are based on traditional media [33] and even Twitter or Facebook posts [48]. Social media, in particular, tends to be informal and subject to variable levels of veracity, especially in comparison with more formal linguistic sources such as newspaper headlines. Similar variations become apparent in terms of where the annotations originate from. For instance, emotion labels can rely upon self-reports of emotional experiences [49] or stem from ex post labeling efforts by crowdsourcing [50].

The majority of datasets were annotated based on categorical emotion models, thereby defining a discrete set of labels. Even though the number and choice of emotions differs, one can identify four emotions that are especially common. More precisely, the following set appears in almost all categorical: namely, anger, joy (happiness), fear and sadness. The chosen emotions largely follow suggestions from the different affect theories and predominantly focus on basic emotions (or subsets thereof) due to their prevalence.

Some emotions occur more often than others in the usual routines of humans [46, 42] and one thus yields datasets, e. g. [33, 50], where the relative frequency of emotions is highly unbalanced. This imposes additional computational challenges as classifiers tend to overlook infrequent classes. As a consequence, we weight the classification loss with respect to the frequency of a label.

In contrast, dimensional models of emotions appear less prevalently. Only one dataset, composed of newspaper headlines [33], provides a score for each of the six emotion categories. From a methodological point of view, this categorization in dimension-based models facilitates the selection of a different computational model. While categorical models refer to machine learning with single-label classification tasks in the sense that we identify the appropriate item based on a discrete label dimensional models allow for multi-label classification in the sense that we predict more than one label for every item. Additionally, dimensional models of emotions reflect regression tasks where we construct emotion scores as continuous values.

Ref.	**Source**	**Samples**	**Emotions**				**Notes**
			Annotation	**Dimensions**	**Count**	**Affect theory**	
[34]	Literary tales	1,207	Categorical (m-out-of-n)	Anger, disgust, fear, happiness, sadness, surprise (pos.), surprise (neg.), neutral	8	Basic emotions from Ekman and Friesen [42]	Evaluations conventionally draw upon subset where all annotators agree
[50]	Election tweets	1,646	Categorical (1-out-of-n)	Anger, anticipation, disgust, fear, joy, sadness, surprise, trust	8	Basic emotions from Averill et al. [43]	
[49]	Self-report of experiences	7,666	Categorical (1-out-of-n)	Anger, disgust, fear, guilt, joy, sadness	7	Based on basic emotions from Ekman and Friesen [42]	Referred to as ISEAR dataset in related literature
[33]	Newspaper headlines	1,000	Numerical (for all dimensions)	Anger, disgust, fear, joy, sadness, surprise; each with valence score	6	Basic emotions from Ekman and Friesen [42] with valence score according to Russell [45]	SemEval-2007 (task 14); one numerical score per class
[51]	General tweets	7,902	Numerical (single dimension only)	Anger, fear, joy, sadness	4	n/a	SemEval-2018 (task 1); one class per instance with a numerical score
[48]	Facebook posts	2,894	Numerical	Valence, arousal	2	Circumplex model from Russell [45]	

Table 1: Overview of textual datasets used for affective computing in the literature grouped into classification and regression tasks for machine learning.

2.3. Applications of affective computing

Text-based affective computing drives decision support in a variety of application areas in which understanding the emotional state of individuals is crucial. Table 2 provides an overview of interesting examples from research, as well as actual use cases from businesses. Evidently, affective computing facilitates decision-making in all operational areas of businesses, such as management, marketing and finance. Beyond that, it also provides public decision support with respect to politics and even education, as well as healthcare for individuals.

Various examples can be identified whereby emotion-laden content facilitates decision support. For instance, firms can infer the perceived emotion of customers from online product reviews and base managerial implications on this data in order to support product development [52], as well as advertising [21]. In a financial context, emotional media content has been identified as a driver in the decision-making of investors [53], which can thus serve as a decision rule for stock investments [24]. Beyond that, affective computing can infer emotion concerning personal health conditions [54, 55, 22, 56] and during learning processes [27].

Domain	Application	Details	Reference
Management & marketing	Development strategy	Identification of perceived emotion towards products as a lever for product development	[52]
	Brand management	Applying emotion analysis to firm-related tweets for reputation management	[23]
	Churn prediction	Emotions customer responses to marketing content serve as a predictor of purchase intention	[21]
	Preference learning	Examining consumer behavior and emotional consequences related to product preferences	[57]
User interaction	Chabots	Regulation of emotion of stranded passengers through chatbots	[58]
	Social networks	Measuring tie strength in social networks with affective language as an indicator of emotional closeness	[20]
Finance	Investment decision	Predicting stock market movements based on emotionally-charged content	[24]
	Economic growth indicator	Excitement and anxiety in media articles as indicators of financial stability and economic shifts	[25]
Politics	Political participation	Emotion recognition for political participation and mobilization	[59]
	Public monitoring	Hate speech detection on Twitter	[60]
Health	Depression treatment	Analyzing emotional content for depressive symptoms in chat transcripts	[56]
	Suicide prevention	Detecting emotions in suicide notes	[55]
	Public health forecast	Predicting mortality from heart disease based on emotions expressed on Twitter	[61]
	Diagnosis	Emotional states are predicting the willingness to disclose personal health information	[54]
	Diagnosis	Social media emotion analysis for detecting poor health care conditions	[22]
Education	E-learning	Improving the learning experience by classifying and regulating e-learner emotions	[27]

Table 2: Application areas in research and industry where text-based emotion recognition facilitates decision support.

2.4. Computational methods

The automatic recognition of text-based emotions relies upon different computational techniques that comprise lexicon-based methods and machine learning. Due to wealth of approaches, we can only summarize the predominant streams of research in the following and refer to [14, 6] for detailed methodological surveys.

2.4.1. Lexicon-based methods

Lexicon-based approaches utilize pre-defined lists of terms that are categorized according to different affect dimensions [62]. On the one hand, these lexicons are often compiled manually, which can later be used for keyword matching. For instance, texts such as the Harvard IV dictionary, found within the General Inquirer software, or LIWC provide such lists with classification by domain experts [7]. These were not specifically designed for affective computing, but still include psychological dimensions (e. g. pleasure, arousal and emotion in the case of Harvard IV; anxiety, anger and sadness for LIWC). The NRC Word-Emotion Association lexicon was derived analogously but with the help of crowdsourcing rather than involving experts from psychology research [63]. The latter dictionary includes 10 granular categories, such as anticipation, trust, and anger.

In order to overcome the need for manual dictionary creation, heuristics have been proposed to construct affect-related wordlists. Common examples include the WordNet-Affect dictionary, which starts with a set of seed words labeled as affect and then assigns scores to all other words based on their proximity relative to the seed words [64]. Here the approach utilizes the graph structure that underlies the WordNet dictionary. However, the resulting affect dictionary includes only general categories of mood- or emotion-related words, rather than further distinguishing the type of emotion. More recent methods operate, for instance, via mixture models [65], fuzzy clustering [66] or by incorporating word embeddings [67]. The precision of dictionaries can further be improved by embedding these in linguistic rules that adjust for the surrounding context.

Dictionary-based approaches are generally known for their straightforward use and out-of-the-box functionality. However, manual labeling is error-prone, costly and inflexible as it impedes domain customization. Conversely, the vocabulary from the heuristics is limited to a narrow set of dimensions that were selected a priori and, as a result, this procedure has difficulties when generalizing to other emotions [cf. 68].

2.4.2. Machine learning

Machine learning can infer decision rules for recognizing emotions based on a corpus of training samples with explicit labels [69, 70]. This can overcome the aforementioned limitations of lexicon-based methods concerning scalability and domain customization. Moreover, it can also learn implicit signals of emotions, since findings from a comprehensive, comparative study suggest that affect is rarely communicated through emotionally-charged lexical cues but rather via implicit expressions [71].

Previous research has experimented with different models for inferring affect from narrative materials. Examples include methods that explicitly exploit the flexibility of machine learning, such as random forests [e. g. 72] and support vector machines [e. g. 73], both of which have commonly been deployed in literature. Studies have shown that random forests tends to compute faster, but support vector machines yield superior performance [73]. These classifiers are occasionally, but infrequently, fed with frequency data of affect cues from emotion lexicons [65]. However, the more common approach relies upon general linguistic features, i. e. bag-of-words with subsequent tf-idf weighting [74, 33]. Consistent with these works, we later draw upon machine learning models (i. e. random forest and support vector machine) together with tf-idf features as our baseline.

2.4.3. Deep learning

In the following, we discuss the few attempts applying deep learning to affective computing, but find that actual performance evaluations are scarce. The approach in [72] predicts aggression expressed in natural language using convolutional neural networks with a sliding window and subsequent max-pooling. However, this approach is subject to several limitations as the network is designed to handle only a single dimension (i. e. aggression) and it is thus unclear how it generalizes across multi-class predictions or even regression tasks that appear in dimensional emotion models. Even though the approach utilizes a "deep" network, its network architecture can only handle texts of predefined size, analogous to traditional machine learning. In this respect, it differs from recurrent networks, which iterate over sequences and thus can handle texts of arbitrary size.

Recurrent neural networks, such as long short-term memory (LSTM) networks, have recently achieved remarkable results in a variety of tasks in natural language processing, including sentiment analysis [75, 76]. Hence, this paper focuses on recurrent neural networks, as these are regarded as the state-of-the-art. Yet numerical experiments for affective computing are limited to the following exceptions. The work in [77] utilizes an LSTM that is pretrained with tweets based on the appearance of emoticons; however, this work does not report a comparison of their LSTM against a baseline from traditional machine learning. A different approach, [78], utilizes a custom LSTM architecture in order to assign emotion labels to complete conversations in social media. However, this approach is tailored to the specific characteristics and emotions of this type of conversational-style data. In addition, the conclusion from their numerical experiments cannot be generalized to affective computing, since the authors labeled their dataset through a heuristic procedure and then reconstructed this heuristic with their classifier.

Up to this point, the potential performance gains from using recurrent neural networks as the state-of-the-art in deep learning have not yet been studied in relation to text-based affective computing. This fact was also noted in a recent literature survey [6]. Hence, it is the objective of this paper to apply these deep neural networks, i. e. LSTMs, to affective computing and present a holistic evaluation.

3. Methods

This section presents our methods for inferring emotional states from narrative contents. We first summarize our baselines from traditional machine learning and then specify how we apply deep learning to affective computing. This is specifically grouped into classification tasks (where a set of emotions needs to be determined) and regression tasks (where the intensity of each affective dimension is represented by a numerical score). We finally detail transfer learning as a process that enables an inductive process of knowledge from the related task of sentiment analysis to this one. Figure 2 illustrates this pipeline.

3.1. Baselines from traditional machine learning

Traditional machine learning can only learn from a fixed-size vector of features and, for this purpose, features for machine learning are commonly built upon bag-of-words; i. e. instead of incorporating individual words, one draws upon tuples thereof (so-called n-grams) in order to preserve the contextual meaning of terms. The frequency of each

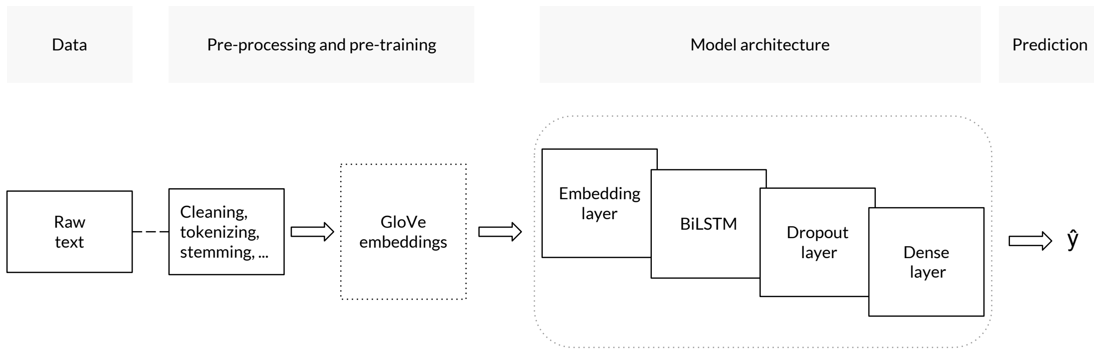

Figure 2: Illustrative pipeline for inferring affective states from narrative materials. This can either happen through (i) traditional machine learning with feature engineering or, as proposed in this work, (ii) deep recurrent neural networks, optionally in par with transfer learning.

n-gram is counted and further weighted by the tf-idf scheme in order to measure the relative importance of terms to a document within a corpus. Mathematically, this measure of term importance is obtained by computing the product of the term frequency and the inverse document frequency [79]. This approach serves as a widespread benchmark with which algorithms for natural language precessing are processed [29].

The aforementioned features are then fed into the actual predictive models from traditional machine learning. Here we chose two approaches for both classification and regression as our baseline models, namely, random forest and support vector machine. These are consistently known for their superior performance in previous studies [e. g. 73]. Moreover, both approaches entail a high flexibility when modeling non-linear relationships and demonstrate a high accuracy even in settings where the number of potential features exceeds the number of observations.

3.2. Deep learning

3.2.1. Recurrent neural networks

Deep learning has triggered a paradigm change in machine learning, since it accomplished to yield unprecedented performance results on a various tasks from natural language processing [80]. The theoretical argument for this is that recurrent neural networks from deep learning can iterate over the individual words of a sequence with arbitrary length. Here the input directly consists of words $x_1, \ldots, x_N$ and thus circumvents the need for feature engineering (e. g. creating bag-of-words with tf-idf) as used in traditional machine learning. As a result, recurrent neural networks store a lower-dimensional representation of input sequence that encodes the whole document and can even maintain the actual word order and thus long-ranging semantics [81]. Because of this reason, recurrent neural networks differ from traditional machine learning, which can only adapt to short contexts due to the use of n-grams.

We utilize a specific variant, the long short-term memory model, which is known for being especially able to encode long dependency structures [82]. The overall architecture is arranged according to four layers, (a) an embedding layer that maps words in one-hot encoding onto low-dimensional vectors, (b) a recurrent layer to pass information on between words, (c) a dropout layer for preventing overfitting and (d) a final dense later for making the actual prediction. The latter varies according to whether an affective category or an emotional intensity is to be predicted. In the end, the weights in all neurons are estimated simultaneously during the training phase. The architecture of each layer is specified as follows:

(a) *Embedding layer*: Our first layer replaces the one-hot encoding of each word in the vocabulary by a numerical representation, in which close words in terms of semantic meaning are optimized to be neighboring words in terms of the distance between their word embeddings. For instance, the embedding of *"good"* will eventually be closer to the word embedding of *"great"* than to the word embedding of *"boring"*. This includes explicit semantics and, in addition, the dense (as opposed to sparse) facilitates the optimization routines for training the subsequent layers.

(b) *Recurrent layer*: The word embeddings are then passed on to a recurrent layer, i. e. an unidirectional LSTM or a bidirectional LSTM. The architecture of a recurrent layer is illustrated in Figure 3. Here recurrent layers draw upon a single feedforward neural network f, for which the connections between neurons form cycles. Thereby, recurrent layers can iterate over textual data word-by-word, thereby accumulating and memorizing information about the meaning of text in a hidden state vector.

    Formally, let e_i be the word embedding of the i-th word. Furthermore, f denote a simple feedforward network that serves as the recurrent layer, while h_i is a hidden state vector and o_i when processing the i-th element in the sequence. When moving from term i to $i{+}1$, the recurrent layer calculates the output o_{t+1} through the neural

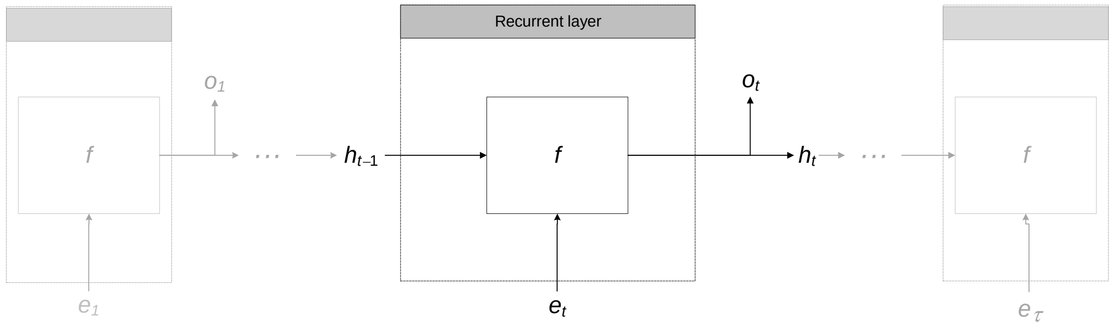

Figure 3: Schematic illustration of a recurrent layer that is unrolled over the input sequence. The i-th word is processed by feeding the embedding e_i into the neural network f. This computes an output vector o_i (that later links to the emotional state) and a hidden state h_i that can pass information to the next, thereby encoding the sequence $e_1, \ldots, e_t$ in this hidden state vector.

network f according to

$$o_{i+1} = f(h_i, e_{i+1}). \tag{1}$$

The recurrent layer is theoretically capable of accumulating text of arbitrary length, yet it requires a suitable design to overcome potential instabilities during optimization [83]. Therefore, this work follows common choices that advocate the use of long short-term memory networks. This architecture overcomes numerical instabilities by introducing an additional cell that stores the accumulated information with explicit update rules (see Figure 4). As an extension, we also experiment with a bidirectional variant (i. e. named BiLSTM) that duplicates the process in order to iterate over the word sequence in both directions.

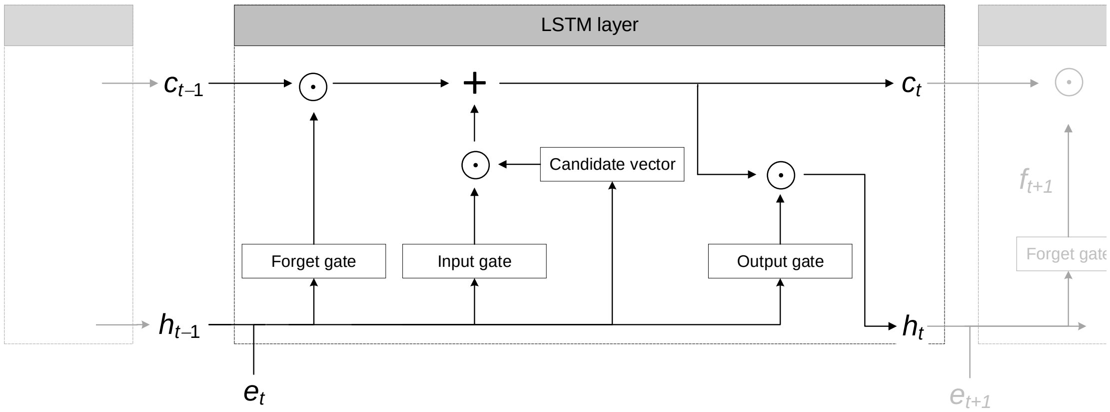

Figure 4: Schematic illustration of a long short-term memory that is again unrolled over the input sequence. The forget gate and the input gate are neural networks that update the cell based on the previous hidden state h_{i-1}, as well as the current input e_i. Furthermore, the output gate gives another neural network that computes the hidden state h_i. The hidden state h_n belonging to the final word then accumulates the complete document.

(c) *Dropout layer*: Deep neural networks can easily consist of up to millions of free parameters and, consequently, these models run the risk of overfitting. As a remedy, the weights in the network are regularized by dropping out a certain share of neurons in order to improve the generalizability of network.

(d) *Dense layer*: The final dense layer ψ draws upon the output of the dropout layer with the aim of obtaining the final prediction output, i. e. a label in a classification or a continuous score in a regression. The structure is

3.2.2. Dense layer for affect prediction

The choice of the dense layer for making the final prediction depends on the desired type, i. e. whether we need to classify the document according to an emotional category or regress it against an intensity rating. Hence, the dense layer follows a linear operation in which every input neuron is connected to every output neuron through a coefficient that is optimized during training of the model. In general, dense layers are followed by activation functions which are non-linear functions that increase the flexibility of the model or, in the case of a classification task, map the vector output from the droput layer onto a categorical representation. The choice of the activation function is governed by the underlying task and we discuss both in the follow.

In the case of a classification, one commoly utilizes a softmax action function σ,

i. e. a generalization of the logistic function that squashes its input values $x_1, \ldots, x_k$ to values in the range $[0, 1]$. Mathematically, it computes

$$\sigma(x)_j = \frac{\exp x_j}{\sum_{k=1}^{n} \exp x_k}, \tag{2}$$

for output j with the additional property that $\sigma(x)_1, \ldots, \sigma(x)_k$ sums to one. This allows us predict the membership with regard to k different classes or categorical emotions by interpreting the estimate $\sigma(x)_j$ as a probability of x belonging to a specific class. When only one class is desired, then we compute $\arg\max_{\kappa \in \{1,\ldots,k\}} \sigma(x)_\kappa$ in order to identify the emotion with the highest probability.

In the case of regression task, we implement an affine transformation $\alpha x^T + \beta$. Thereby, the underlying representation in the form of a numerical values is aggregated onto a single numerical score that represents the intensity according to the desired affective dimension.

3.3. Transfer learning

Training deep neural networks is often associated with challenges due to the large number of degrees-of-freedom. In practice, this is encountered by large datasets in order to prevent overfitting and, hence, a different strategy is often applied when handling smaller datasets such as in our comparisons. Here the idea is to implement transfer learning, i. e. an inductive transfer of knowledge from a different yet related task to the problem under study. This often yields considerable improvements in predictive performance [29].

Formally, transfer learning optimizes the weights of a neural network based on a different, yet related, dataset $\mathcal{R}$. It then utilizes the estimated parameters as an initial value for further optimization with the help of the actual dataset $\mathcal{D}$ [84]. For this purpose, we suggest the use of sentiment analysis as a related tasks, since it shares the similarity in the sense that positive and negative polarity is inferred from linguistic materials. However, sentiment analysis differs from affective computing, as it does not address affective dimensions or emotional states.

In our experiments, we utilize a kaggle dataset[1] as a basis for knowledge induction. This dataset finds widespread application in sentiment analysis and includes about 100,000 samples labeled according to positive or negative sentiment. We then optimize the deep neural network with the goal of predicting the underlying sentiment scores. The resulting coefficients of the network are then further trained with actual dataset from affective computing. Here the differences in the data type of the prediction outcome (i. e. computing a positivity/negativity score versus affective dimensions) are handled by removing the dense layer and, instead, amending a new prediction layer that targets the new output. As a result, the majority of weights benefits from transfer learning, while only the neurons in the prediction layer are training starting from their initial value. The intuition of this approach is as follows: deep neural networks generally contain multiple layers, where layers towards the final prediction layers are supposed to encode the original input in a higher level of abstraction. Hence, the relatedness between both tasks enables the network to infer similar representation for both. The pseudocode of the overall process is stated in Algorithm 1.

Algorithm 1 Transfer learning

Input: Given training data $\mathcal{D}$ for affective computing and additional corpus $\mathcal{R}$

$m \leftarrow$ Initialize recurrent neural network (i. e. consisting of recurrent layer f, dense layer ψ, ...)
$m \leftarrow$ Estimate parameters using $\mathcal{R}$
$\psi \leftarrow$ Replace dense layer with randomly-initialized dense layer according to the dimensions of $\mathcal{D}$
$\psi \leftarrow$ Fine-tune ψ using $\mathcal{D}$
return Recurrent neural network m

3.4. Model estimation

Consistent with previous research [85], we tokenize each document, convert all characters to lower-case and remove punctuations, numbers, as well as stop words. Moreover, we perform stemming, which maps inflected words onto a base form; e. g. *"played"* and *"playing"* are both mapped onto *"play"*. We conducted all pre-processing operations by using the natural language tookit NLTK [86].

Our experimental setup involves four different architectures of deep neural networks.

[1] Kaggle: Twitter sentiment analysis: `https://www.kaggle.com/c/twitter-sentiment-analysis2`

In a first phase, we feed the pre-processed data to a unidirectional LSTM model, as well as to a second model, a bidirectional LSTM. Subsequently, we incorporate pre-trained word embeddings from GloVe[2].

For those dataset with no designated *test* set we conducted a random 80/20 split in training and test data. For the random forest classifier we manually optimized over the number of trees, number of maximum number of features for every split and the depth. For the support vector classifier we conducted an excessive grid-search over the hyperparamters following [87]. In detail, we experimented with linear, rbf and sigmoid kernels; optimizing C within $2^{-5}, 2^{-3}, \ldots, 2^{15}$ and $\gamma = 2^{-15}, 2^{-13}, \ldots, 2^{3}$.

For unbalanced datasets we weighted the loss function accordingly to prevent models from predicting the majority classes only. Additionally, we chose a suitable performance metric when reporting the results.

We used four different deep learning models. Two of them used pre-trained glove embeddings, whereas the other two used initially random embeddings and learned them conjointly with the classification task. Architecturally we used two types of models, one with an unidirectional recurrent layer and one with a bidirectional recurrent layer. We trained the models using the Adam optimizer and stopped training when we saw the test-set error increase. For reproducibility of the results we report the average F1-score over 10 independent runs.

4. Evaluation

This section reports our computational experiments that evaluate the improvements of using deep neural networks and especially transfer learning for affective computing. Here we draw upon all datasets from Table 1 and, according to the type of the underlying affect theory, we divide the results into classification and regression tasks.

4.1. Classification according to categorical emotion models

We being with classification tasks according to categorical emotion models, where the objective is to predict the predominant emotion(s). We follow previous literature

[2]The pre-trained word embeddings can be downloaded from `http://nlp.stanford.edu/data/glove.6B.zip`.

[e. g. 73] and choose analogously two prevalent baselines from traditional machine learning, namely, the random forest classifier and the support vector machine. Both are fed with bag-of-words with tf-idf weighting, whereas the proposed deep neural networks circumvent the need for feature engineering. Here we compare variants that extend the LSTM with bidirectional encodings and pretrained word embeddings. The resulting performance is listed in Table 3, where we account for unbalanced distributions of labels by using the $F1$-score. The $F1$-score is given by the harmonic mean of precision and recall,

$$F1 = 2\frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}. \tag{3}$$

Our results consistently identify a superior performance through the use of deep learning. We observe that, independent from the architecture, models with pre-trained Glove embeddings outperform their counterparts with randomly-initialized word embeddings. In fact, the use of pre-trained word embeddings yields performance improvements of up to 13.7 in f1-score over the best baseline. A explanation originates from the fact that the latter has to optimize these weights during the training process, thereby introducing considerably more degrees-of-freedoms and thus higher chance of overfitting. Furthermore, our initial expectations are confirmed as the bidrectional recurrent layers outperforms the variant with a unidirectional layer slightly. Here the improvements range between 4 % and 21 % across the datasets.

We find further patterns that link to the following implication. The performance gains from deep learning link to the size of the dataset; i. e. the larger the training set, the higher are the increases in the F1-score. For instance, the highest relative improvement over traditional machine learning is accomplished for ISEAR, which also represents the second largest dataset consisting of more than 7, 000 training samples.

Dataset	Traditional machine learning		Deep learning		Pre-trained word embeddings	
	Random forest	**SVM**	**LSTM**	**BiLSTM**	**LSTM**	**BiLSTM**
Literary tales [34]	63.2	64.7	62.9	60.9	67.4	**68.5**
Election tweets [50]	55.0	56.8	54.4	54.2	55.2	**57.8**
ISEAR [49]	44.5	55.5	54.5	56.6	**58.2**	57.0
Headlines [33]	35.6	35.4	39.4	39.6	41.6	**44.1**
General tweets [51]	52.8	54.3	56.2	55.6	**58.0**	57.5

Table 3: Holistic comparison of traditional machine learning and recurrent neural networks (with optional Glove word embeddings) for affective computing that is models as classification tasks. Here the outcome variable represents a single label according to predefined categorical emotion model. Accordingly, the performance is measured based on the F1-score. The best-performing model for each dataset is highlighted in bold.

4.2. Regression according to dimensional affect models

Predicting affect intensity is differs form classification. Instead of predicting the emotion for a given input, we are interested in the score for an given emotion and input. We evaluate all our models by the mean squared error, given by

$$\text{MSE} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2. \tag{4}$$

Where y_i is the ground truth score and $\hat{y}_i$ the predicted output. It is important to mention that the mean squared error is a relative measure and dependent on the scale of y. Hence we can only compare values relative to each other within a dataset.

We evaluated a deep learning model to predict the emotion intensity against two methods from classical machine learning. For the two baseline models we chose again the aforementioned hyper-parameter search. As deep learning model we used a bidirectional architecture with pre-initialized word embeddings.

As shown in Table 4 our deep learning models outperform the baseline in all 7 tasks. For fear score prediction the margin was very small, for all other datasets the margin was considerably large. Therefore, we show that deep learning is not only state-of-the-art for emotion prediction but also for predicting the affect intensity.

Dataset	Traditional ML		Deep learning
	Random forest	SVM	BiLSTM
Headlines - Valence [33]	1906.0	1927.3	**1799.7**
Facebook - Valence [48]	1.030	0.951	**0.904**
Facebook - Arousal [48]	3.960	3.616	**3.358**
General Tweets - Anger [51]	0.0314	0.0323	**0.0286**
General Tweets - Fear [51]	0.0245	**0.0226**	**0.0222**
General Tweets - Joy [51]	0.0339	0.0294	**0.0271**
General Tweets - Sadness [51]	0.0294	0.0274	**0.0242**

Table 4: Holistic comparison of traditional machine learning and recurrent neural networks (with optional Glove word embeddings) for affective computing that is models as regression tasks. Here the outcome variable represents a score according to a given emotion. Accordingly, the performance is measured based on the mean squared error. The best-performing model for each dataset is highlighted in bold.

4.3. Transfer learning

We picked two challenging datasets of twitter messages to demonstrate our transfer-learning technique. Twitter is composed of a very noisy language including slang, hash-tags, typos etc. This poses additional difficulties on affect classification. Furthermore, the election tweets dataset is with $1,646$ samples very small.

To overcome these limitations we propose to use techniques from transfer learning. We train a bidirectional lstm network on the different but related taks of sentiment analysis. By using a dataset of $100,000$ samples we can efficiently train parameters for word embeddings and the recurrent layer. Finally, we remove the prediction layer and replace it with a new prediction layer for our set of emotions and fine-tune the parameters for the task at hand.

Dataset	BiLSTM (Plain)	BiLSTM + Glove	BiLSTM + TL
Election Tweets [50]	54.2	57.8	**58.4**
General Tweets [51]	55.6	57.5	**58.5**

Table 5: Results of Experiments on transfer learning compared to two deep neural networks without transfer learning.

We compare three variants of bidirectional LSTM networks in Table 5. First, a network with randomly initialized word embeddings; second, the same network with pre-initialized glovew word embeddings; third, the network trained with our aforementioned transfer-learning method. As we can see transfer learning gives additional improvement of up to 5.5 % over the plain network with random embeddings and up to 2.2 % over the network with pre-initialized glove word embeddings.

5. Discussion

5.1. Comparison

Our series of experiments unfolds considerable and consistent performance improvements from using deep learning over traditional machine learning. Here our holistic analysis demonstrates that deep learning models constantly outperform the classic baseline. Interestingly, the deep neural networks were even able to learn the underlying relationships from the rather small datasets of merely 1,000 observations. However, we observe an overall pattern that the performance improvements tend to be higher the larger the dataset is. In addition, we observe further improvement from using word embeddings as these reduce the high-dimensional vectors with terms as one-hot encoding to lower-dimensional spaces. Here the improvements range between 4 % and 21 %.

In the majority of experiments, the superior results stem from using a bi-directional LSTM as compared to a simple LSTM. This architecture can process sequential input, such as as sequences of words, of arbitrary length and, consistent with earlier findings in other domains, appears also beneficial for affective computing. We finally note that not only traditional machine learning but all network architectures required extensive training such that embeddings and dropout layer functioned well together.

Finally, the task of emotion recognition in affective computing is related to sentiment analysis which infers a positive/negative polarity from linguistic materials. Hence, it is interesting to study whether one can further improve performance through an inductive transfer of knowledge – despite the different objective, linguistic style and annotation scheme. As a result, our implementation of transfer learning yields additional 5.5 % improvements over plain networks.

5.2. Implications for management and practice

Better predictive analytics can spark improvements for decision-making and decision support. As a consequence, competition on analytics has been a widely accepted theme, as even little improvements in prediction accuracy can escalate revenues due to the accumulation over the whole customer base. However, competition is at an edge with the recent advent of deep learning, since, for transfer learning and word embeddings to be effective, large datasets are required that are often only available to multinational companies. Hence, the instrument of transfer learning and embeddings benefits practitioners unequally. A potential, yet constrained, remedy is the use of public data sources.

Affective computing for linguistic materials opens new opportunities for business models and consumer-centered services. Detecting and subsequently responding to emotional states of users, customers, patients and employees has the potential to significantly accelerate and improve management processes and optimize human-computer-interaction systems. Here text remains a critical form of communication, while attempts have been made to apply affective computing to speech or other multi-modal input [14], including visual data and speech [15, 16, 17]. Management should critically assess potential use cases in critical areas of operations from their own organization. Our overview in Section 2 (see Table 2) provides illustrative examples, while further applications are likely to arise with recent methodological innovations.

5.3. Implications for research

Deep learning promises to create additional value for firms, organizations and individuals in a variety of business units and domains, yet its use in the actual practice of decision support remains scarce. Hence, it should be the goal of future research in the realm of decision support to identify precious use cases, outline potential value gains and derive recommendations concerning combinations of network architectures and training routines that were found to be effective. Deep learning (as well as all other forms of predictive modeling) merely offer predictive hindsight, but rarely prescribe actual management decisions to reach the desired outcome. As a remedy to this, our discipline is well-equipped with the means to study how predictions can actually be translated into effective decision-making as another compelling road for future research.

Further improving the performance of affective computing would considerably benefit from a rigorous suite of baseline datasets. In the status quo, a variety of datasets with distinct goals and purposes is commonly used for benchmarking methodological innovations for affective computing. For instance, our literature survey identified four different strategies for annotating, including simple labels, multi-class labels and numerical scores. Moreover, the set of affective dimensions varied between two (i. e. valence, arousal without explicitly naming emotions) to a set of 8 emotions (e. g. anger, disgust, surprise). However, this directly links to challenges concerning comparability and generalizability. In this sense, a network architecture that has been found effective for one annotation scheme might not work out for other datasets. On top of that, different labels prohibit transfer learning and thus impede performance. We thus suggest a standardized approach to annotations.

According to our literature review, datasets for affective computing vary in size from 1,000 instances to 7,902, yet all of them are remain fairly small. This is known to limit the performance of bidirectional LSTMs and other deep neural network architectures, which generally require large-scale datasets. Future research should thus aim at creating larger datasets in order to enable the effective exploitation of deep learning.

6. Conclusion

Affective computing allows to infer individual and collective emotional states from textual data and thus offers an anthropomorphic path for the provision of decision support. It use promises benefits in a wide range of application areas, ranging from human-computer interactions to managerial decision-making and even public decision support. In the status quo, affective computing is almost exclusively implemented via lexicon-based methods and traditional machine learning. However, the complexity and ambiguity of emotion-laden language cannot be accurately reflected by the simplicity of dictionaries or frequency-based machine learning, since emotional content is often conveyed by linguistic expressions implicitly. Conversely, recurrent neural networks from the field of deep learning can provide a remedy as these map narrative materials onto a lower-dimensional representation while maintaining the order (and thus semantic context) of the whole sequence of words.

This work applies recurrent neural networks from the recent wave of deep learning to affective computing and contributes a holistic comparison across multiple affect-labeled datasets. Our computational experiments span categorical and dimensional emotion models, which require tailored algorithmic implementations involving, e. g., multi-class classification, as well as regression tasks. Our results show that pre-trained long short-term memory models consistently outperform the baseline models from traditional machine learning. The performance improvements can even reach up 21 % in f1-score for classification and 6 % in RMSE for regression. We further propose the use of transfer learning, which is responsible for additional performance improvements between 2 and 5%. As a direct recommendation for use cases of affective computing, we propose a shift from traditional machine learning to recurrent neural networks, even for fairly small datasets of around 1,000 samples as in our case.

References

References

[1] K. Oatley, W. G. Parrott, C. Smith, F. Watts, K. Oatley, W. G. Parrott, C. Smith, F. Watts, K. Oatley, W. G. Parrott, C. Smith, F. Watts, Cognition and Emotion over twenty-five years 9931 (2016) 1341–1349.

[2] J. Greene, H. J, How (and where) does moral judgement work?, Trends in Cognitive Sciences 6 (2002) 517–523.

[3] N. Schwarz, Emotion, cognition, and decision making, Cognition & Emotion 14 (2000) 433–440.

[4] N. Derakshan, M. W. Eysenck, Introduction to the special issue: Emotional states, attention, and working memory, Cognition and Emotion 24 (2010) 189–199.

[5] R. J. Dolan, Emotion, Cognition, and Behavior, Science 298 (2002) 1191–1194.

[6] S. Poria, E. Cambria, R. Bajpai, A. Hussain, A review of affective computing: From unimodal analysis to multimodal fusion, Information Fusion 37 (2017) 98–125.

[7] Y. R. Tausczik, J. W. Pennebaker, The psychological meaning of words: LIWC and computerized text analysis methods, Journal of Language and Social Psychology 29 (2010) 24–54.

[8] R. W. Picard, Affective Computing: Challenges, Affective Computing (1997) 1–17.

[9] R. W. Picard, Affective Computing, MIT press (1995) 1–16.

[10] J. Tao, T. Tan, R. W. Picard, LNCS 3784 - Affective Computing and Intelligent Interaction, ???? URL: `https://link.springer.com/content/pdf/10.1007/11573548.pdf{#}page=884`.

[11] S. M. Alarcao, M. J. Fonseca, Emotions Recognition Using EEG Signals: A Survey, IEEE Transactions on Affective Computing 3045 (2017) 1–20.

[12] P. J. Astor, M. T. P. Adam, P. Jerčić, K. Schaaff, C. Weinhardt, Integrating Biosignals into Information Systems: A NeuroIS Tool for Improving Emotion Regulation, Journal of Management Information Systems 30 (2013) 247–278.

[13] E. Lux, F. Hawlitschek, M. T. P. Adam, J. Pfeiffer, Using Live Biofeedback for Decision Support: Investigating Influences of Emotion Regulation in Financial Decision Making, ECIS 2015 Proceedings (2015) 0–12.

[14] R. A. Calvo, S. D'Mello, Affect detection: An interdisciplinary review of models, methods, and their applications, IEEE Transactions on Affective Computing 1 (2010) 18–37.

[15] Y.-L. Chen, C.-L. Chang, C.-S. Yeh, Emotion classification of YouTube videos, Decision Support Systems 101 (2017) 40–50.

[16] M. El Ayadi, M. S. Kamel, F. Karray, Survey on speech emotion recognition: Features, classification schemes, and databases, Pattern Recognition 44 (2011) 572–587.

[17] C. Shan, S. Gong, P. W. McOwan, Facial expression recognition based on Local Binary Patterns: A comprehensive study, Image and Vision Computing 27 (2009) 803–816.

[18] F. Hogenboom, F. Frasincar, U. Kaymak, F. De Jong, E. Caron, A Survey of event extraction methods from text for decision support systems, Decision Support Systems 85 (2016) 12–22.

[19] V. Dhar, Data science and prediction, Communications of the ACM 56 (2013) 64–73.

[20] P. V. Marsden, K. E. Campbell, Reflections on conceptualizing and measuring tie strength, Social Forces 91 (2012) 17–23.

[21] S. H. Ang, S. Y. M. Low, Exploring the Dimensions of Ad Creativity, Psychology & Marketing 17 (2000) 835–854.

[22] F. Greaves, D. Ramirez-Cano, C. Millett, A. Darzi, L. Donaldson, Harnessing the cloud of patient experience: using social media to detect poor quality healthcare: Table1, BMJ Quality & Safety 22 (2013) 251–255.

[23] D. Al-Hajjar, A. Z. Syed, Applying sentiment and emotion analysis on brand tweets for digital marketing, 2015 IEEE Jordan Conference on Applied Electrical Engineering and Computing Technologies, AEECT 2015 (2015).

[24] E. Gilbert, K. Karahalios, Widespread Worry and the Stock Market, Proceedings of the 4th International AAAI Conference on Weblogs and Social Media (2010) 58–65.

[25] R. Nyman, P. Ormerod, D. Tuckett, Measuring Financial Sentiment to Predict Financial Instability : A New Approach based on Text Analysis Centre for the Study of Decision Making Uncertainty , University College , London February 2015 (2015) 1–19.

[26] L. Medeiros, T. Bosse, Social Informatics 10046 (2016) 439–453.

[27] P. Rodriguez, A. Ortigosa, R. M. Carro, Extracting emotions from texts in E-learning Environments, Proceedings - 2012 6th International Conference on Complex, Intelligent, and Software Intensive Systems, CISIS 2012 (2012) 887–892.

[28] J. Evermann, J. R. Rehse, P. Fettke, Predicting process behaviour using deep learning, Decision Support Systems 100 (2017) 129–140.

[29] M. Kraus, S. Feuerriegel, Decision support from financial disclosures with deep neural networks and transfer learning, Decision Support Systems 104 (2017) 38–48.

[30] Y. Lecun, Y. Bengio, G. Hinton, Deep learning, Nature 521 (2015) 436–444.

[31] A. C. Ian Goodfellow, Yoshua Bengio, Deep Learning, volume 521, MIT Press, 2017. doi:10.1016/B978-0-12-391420-0.09987-X. arXiv:arXiv:1011.1669v3.

[32] F. N. Ribeiro, M. Araújo, P. Gonçalves, M. A. Gonçalves, F. Benevenuto, Sentibench-a benchmark comparison of state-of-the-practice sentiment analysis methods, EPJ Data Science 5 (2016) 23.

[33] C. Strapparava, R. Mihalcea, Semeval-2007 task 14: Affective text, Proc. of SemEval-2007 (2007) 70–74.

[34] C. O. Alm, Affect in text and speech, Ph.D. thesis, University of Illinois at Urbana-Champaign, 2008. URL: http://lrc.cornell.edu/swedish/dataset/affectdata/COAlm{_}Dissertation.pdf.

[35] M. Munezero, C. S. Montero, E. Sutinen, J. Pajunen, Are they different? affect, feeling, emotion, sentiment, and opinion detection in text, IEEE Transactions on Affective Computing 5 (2014) 101–111.

[36] B. Pang, L. Lee, Opinion Mining and Sentiment Analysis, Foundations and Trends® in InformatioPang, B., & Lee, L. (2006). Opinion Mining and Sentiment Analysis. Foundations and Trends® in Information Retrieval, 1(2), 91231. doi:10.1561/1500000001n Retrieval 1 (2006) 91–231.

[37] A. Yadollahi, A. G. Shahraki, O. R. Zaiane, Current State of Text Sentiment Analysis from Opinion to Emotion Mining, ACM Computing Surveys 50 (2017) 1–33.

[38] N. H. Frijda, The Laws of Emotion, American Psychologist 43 (1988) 349–358.

[39] C. E. Izard, Emotion Theory and Research: Highlights, Unanswered Questions, and Emerging Issues, Annual Review of Psychology 60 (2009) 1–25.

[40] S. Tomkins, Affect imagery consciousness: Volume I: The positive affects, Springer publishing company, 1962.

[41] C. E. Izard, Basic emotions, relations among emotions, and emotion-cognition relations., Psychological review 99 (1992) 561–5.

[42] P. Ekman, W. V. Friesen, Ekman.pdf, Journal of Personality and Social Psychology 53 (1987) 712–717.

[43] J. Averill, R. Plutchik, H. Kellerman, Theories of emotion, 1980.

[44] S. S. Tomkins, Affect imagery consciousness, 2: The negative affects. (1963).

[45] J. A. Russell, A Circumplex Model of Affect, Journal of Personality and Social Psychology 39 (1980) 1161–1178.

[46] R. Plutchik, The nature of emotions: Human emotions have deep evolutionary roots, 2001. doi:10.1511/2001.4.344.

[47] E. Cambria, A. Livingstone, A. Hussain, The hourglass of emotions, Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 7403 LNCS (2012) 144–157.

[48] D. Preotiuc-Pietro, H. Schwartz, G. Park, Modelling Valence and Arousal in Facebook posts, of NAACL-HLT (2016) 9–15.

[49] H. Scherer, K. R. Wallbott, How universal and specific is emotional experience?, Social Science Information 26 (1986) 763–795.

[50] S. M. Mohammad, X. Zhu, S. Kiritchenko, J. Martin, Sentiment, emotion, purpose, and style in electoral tweets, Information Processing and Management 51 (2015) 480–499.

[51] S. M. Mohammad, F. Bravo-Marquez, M. Salameh, S. Kiritchenko, Semeval-2018 Task 1: Affect in tweets, in: Proceedings of International Workshop on Semantic Evaluation (SemEval-2018), New Orleans, LA, USA, 2018.

[52] R. Ullah, N. Amblee, W. Kim, H. Lee, From valence to emotions: Exploring the distribution of emotions in online product reviews, Decision Support Systems 81 (2016) 41–53.

[53] N. Pröllochs, S. Feuerriegel, D. Neumann, IS HUMAN INFORMATION PROCESSING AFFECTED BY EMOTIONAL CONTENT? UNDERSTANDING THE ROLE OF FACTS AND EMOTIONS IN THE STOCK MARKET Research-in-Progress, Icis (2016) 1–11.

[54] C. L. Anderson, R. Agarwal, The digitization of healthcare: Boundary risks, emotion, and consumer willingness to disclose personal health information, Information Systems Research 22 (2011) 469–490.

[55] B. Desmet, V. Hoste, Emotion detection in suicide notes, Expert Systems with Applications 40 (2013) 6351–6358.

[56] R. Van Der Zanden, K. Curie, M. Van Londen, J. Kramer, G. Steen, P. Cuijpers, Web-based depression treatment: Associations of clients word use with adherence and outcome, Journal of Affective Disorders 160 (2014) 10–13.

[57] R. Chitturi, R. Raghunathan, V. Mahajan, Form Versus Function:How the Intensities of Specific Emotions Evoked in Functional Versus Hedonic Trade-Offs Mediate Product Preferences, Journal of Marketing Research 44 (2007) 702–714.

[58] C. Paper, L. Medeiros, V. Universiteit, W. Vrije, U. Amsterdam, C. Natalie, Progress in Artificial Intelligence 8154 (2013).

[59] N. A. Valentino, T. Brader, E. W. Groenendyk, K. Gregorowicz, V. L. Hutchings, Election night's alright for fighting: The role of emotions in political participation, Journal of Politics 73 (2011) 156–170.

[60] P. Burnap, M. L. Williams, Cyber Hate Speech on Twitter: An Application of Machine Classification and Statistical Modeling for Policy and Decision Making, Policy & Internet 7 (2015) 223–242.

[61] J. C. Eichstaedt, H. A. Schwartz, M. L. Kern, G. Park, D. R. Labarthe, R. M. Merchant, S. Jha, M. Agrawal, L. A. Dziurzynski, M. Sap, C. Weeg, E. E. Larson, L. H. Ungar, M. E. P. Seligman, Psychological Language on Twitter Predicts County-Level Heart Disease Mortality, Psychological Science 26 (2015) 159–169.

[62] S. M. Mohammad, From once upon a time to happily ever after: Tracking emotions in mail and books, Decision Support Systems 53 (2012) 730–741.

[63] S. M. Mohammad, P. D. Turney, Crowdsourcing a word-emotion association lexicon, Computational Intelligence 29 (2013) 436–465.

[64] C. Strapparava, A. Valitutti, WordNet-Affect: an affective extension of WordNet, Proceedings of the 4th International Conference on Language Resources and Evaluation (2004) 1083–1086.

[65] A. Bandhakavi, N. Wiratunga, D. Padmanabhan, S. Massie, Lexicon based feature extraction for emotion text classification, Pattern Recognition Letters 93 (2017) 133–142.

[66] S. Poria, A. Gelbukh, E. Cambria, A. Hussain, G. B. Huang, EmoSenticSpace: A novel framework for affective common-sense reasoning, Knowledge-Based Systems 69 (2014) 108–123.

[67] M. Li, Q. Lu, Y. Long, L. Gui, Inferring Affective Meanings of Words from Word Embedding, IEEE Transactions on Affective Computing 3045 (2017) 1–14.

[68] A. Agrawal, A. An, Unsupervised Emotion Detection from Text Using Semantic

and Syntactic Relations, 2012 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology (2012) 346–353.

[69] T. Danisman, A. Alpkocak, Feeler: Emotion Classification of Text Using Vector Space Model vol. 2 (2008).

[70] S. Chaffar, D. Inkpen, Using a heterogeneous dataset for emotion analysis in text, Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 6657 LNAI (2011) 62–67.

[71] A. Balahur, J. M. Hermida, A. Montoyo, Detecting implicit expressions of emotion in text: A comparative analysis, Decision Support Systems 53 (2012) 742–753.

[72] D. Gordeev, Detecting State of Aggression in Sentences Using CNN, in: 18th International Conference, SPECOM 2016, volume 9811, Springer International Publishing, Budapest, 2016, pp. 240–245. URL: http://link.springer.com/10.1007/978-3-319-43958-7. doi:10.1007/978-3-319-43958-728.

[73] D. Chatzakou, A. Vakali, K. Kafetsios, Detecting variation of emotions in online activities, Expert Systems with Applications 89 (2017) 318–332.

[74] C. O. Alm, D. Roth, R. Sproat, Emotions from text, Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing - HLT '05 (2005) 579–586.

[75] R. Socher, A. Perelygin, J. Wu, Recursive deep models for semantic compositionality over a sentiment treebank, Proceedings of the ... (2013) 1631–1642.

[76] J. Barnes, R. Klinger, S. S. im Walde, Assessing State-of-the-Art Sentiment Models on State-of-the-Art Sentiment Datasets (2017) 2–12.

[77] B. Felbo, A. Mislove, A. Søgaard, I. Rahwan, S. Lehmann, Using millions of emoji occurrences to learn any-domain representations for detecting sentiment, emotion and sarcasm (2017).

[78] U. Gupta, A. Chatterjee, R. Srikanth, P. Agrawal, A Sentiment-and-Semantics-Based Approach for Emotion Detection in Textual Conversations (2017).

[79] C. Ludwig, Text Retrieval 24 (2007) 1–21.

[80] J. Hirschberg, C. D. Manning, Advances in natural language processing, Science 349 (2015) 261–266.

[81] I. Goodfellow, Y. Bengio, A. Courville, Y. Bengio, Deep learning, volume 1, MIT press Cambridge, 2016.

[82] S. Hochreiter, J. Schmidhuber, Long short-term memory, Neural computation 9 (1997) 1735–1780.

[83] Y. Bengio, P. Simard, P. Frasconi, Learning long-term dependencies with gradient descent is difficult, IEEE transactions on neural networks 5 (1994) 157–166.

[84] S. J. Pan, Q. Yang, A survey on transfer learning, IEEE Transactions on Knowledge and Data Engineering 22 (2010) 1345–1359.

[85] C. D. Manning, H. Schütze, Foundations of statistical natural language processing, MIT press, 1999.

[86] E. Loper, S. Bird, NLTK: The Natural Language Toolkit (2002).

[87] C.-W. Hsu, C.-C. Chang, C.-J. Lin, et al., A practical guide to support vector classification (2003).