Middle-East Journal of Scientific Research
24 (Special Issue on Innovations in Information, Embedded and Communication Systems): 32-38, 2016
ISSN 1990-9233; © IDOSI Publications, 2016
DOI: 10.5829/idosi.mejsr.2016.24.IIECS.23136

A Survey of Emotion Analysis

S. Angel Deborah, R.S. Milton and S. Hannah

SSN College of Engineering, Chennai, India

Abstract: Emotions have been widely studied as they play an important role in human intelligence, rational decision making, social interaction, perception, memory, learning and creativity. An emotion is a strong feeling and is instinctive. It is triggered by situational cues and physiological responses and is multi-dimensional in nature. There is much progress in the field of sentiment analysis and opinion mining whereas the study on emotion analysis has just begun. The study of emotion attracts increasingly greater attention in the field of Natural Language Processing (NLP) due to its emerging wide applications. The automatic detection of emotions in texts is important for applications such as opinion mining and market analysis, effective computing, natural language interfaces and e-learning environments, including educational games. The ability to discern and understand human emotions is crucial for making interactive computer agents more human-like and makes it necessary to use machine learning approaches. In this paper, we are presenting a survey on the existing emotion detection techniques.

Key words: Emotion analysis · NLP · Machine Learning · Supervised Learning

INTRODUCTION

Recognizing user's emotions is a major challenge for both humans and machines. On the one hand, people may not be able to recognize or state their own emotions some times. On the other hand, machines need to have accurate ground truth for emotion modeling and also require advanced ma-chine learning algorithms for developing the emotion models. Emotion analysis is a computational study of how opinions, attitudes, emotions and perspectives are expressed in nature language [1]. It provides techniques for extracting useful information from natural language and summarizing it. It can thus be vital to service providers or production companies, allowing them to quickly assess how new products and features are being received. The applications of emotion analysis are broad and powerful. The ability to extract insights from social data is a practice that is being widely adopted by organizations across the world. Shifts in sentiment on social media have been shown to correlate with shifts in the stock market. Teaching a machine to analyze the various grammatical nuances, cultural variations, slang and misspellings that occur online is a difficult process. Humans are fairly intuitive when it comes to interpreting the tone of a piece of writing. Consider the following sentence: "My flight's been delayed. Brilliant!" Most humans would be able to quickly interpret that the person was being sarcastic. Without contextual understanding, a machine looking at the sentence above might see the word "brilliant" and categorize it as positive. It is indispensable to use machine learning techniques to train the machines for an effective emotion analysis.

Machine Learning Methodologies: Machine learning is concerned with learning an appropriate set of parameters within a model class from training data. The meta-level problems of determining appropriate model classes are referred to as model selection or model adaptation. Super-vised as well as unsupervised learning approaches have been used in emotion detection. Supervised learning approaches rely on labelled training data, a set of training examples. The supervised learning algorithm analyses the training data and infers a function, which we use for mapping new examples. In an unsupervised learning approaches, these algorithms try to find hidden structure in unlabeled data in order to build models for emotion classification [2]. From a theoretical point of view, supervised and unsupervised learning differ only in the causal structure of the model. In supervised learning, the model defines the effect one set of observations, called inputs, has on another set of observations,

Corresponding Author: S. Angel Deborah, SSN College of Engineering, Chennai, India.

called outputs. In other words, the inputs are assumed to be at the beginning and the outputs at the end of the causal chain. The models can include mediating variables between the inputs and the outputs. In unsupervised learning, all the observations are assumed to be caused by latent variables, that is, the observations are assumed to be at the end of the causal chain. In practice, models for supervised learning often leave the probability for inputs undefined. This model is not needed as long as the inputs are available, but if some of the input values are missing, it is not possible to infer anything about the outputs. If the inputs are also modelled, then missing inputs cause no problem since they can be considered latent variables as in unsupervised learning. There are various machine learning approaches to emotion analysis on text which are explained in the following sections.

Naive Bayes Classification Algorithm: Probabilistic classifiers use the Bayes's theorem to calculate the probability P (c|d), that a document belongs to category c.

$$P(c|d) = \frac{P(d|c)P(c)}{P(d)}$$

In order to determine the probability P (d|c), it is assumed that the coordinates representing the document as a feature vector are independent. These classifiers are called as Naive Bayes (NB) classifiers. Some justification about the robustness of these classifiers can be found in Sentiment Classification using Machine Learning Techniques [3].

Maximum Entropy Classifier: The Maximum Entropy (MaxEnt) classifier is a discriminative classifier commonly used in Natural Language Processing. The MaxEnt classifier is a probabilistic classifier which belongs to the class of exponential models. Unlike the Naive Bayes classifier the MaxEnt does not assume that the features are conditionally independent of each other. Its estimate of P (c|d) takes the exponential form.

$$P_{ME}(c|d) = \frac{1}{Z(d)} \exp\left(\sum \lambda_{i,c} F_{i,c}(d,c)\right)$$

where Z(d) is a normalization function. Fi;c is a feature/class function for feature fi and class c. MaxEnt is based on the principle of maximum entropy and from all the models that fit our training data, selects the one which has the largest entropy. The MaxEnt classifier can be used to solve a large variety of text classification problems [3].

Support Vector Machines: The support vector machine (SVM) is the most widely used algorithm in the field of machine learning classification. A binary SVM is a hyperplane separating the feature space of positive instances from the feature space of negative instances. During the training phase, the hyperplane that can separate the positive feature space from the negative feature space with a maximal margin is chosen. They are large margin, rather than probabilistic classifiers, in contrast to Naive Bayes and MaxEnt classifiers [3]. The margin is the distance of the nearest point from the positive and negative sets to the hyperplane. Support vectors, the subset of the training instances, determine the hyperplane for a SVM. SVM classifiers perform extremely well irrespective of the dimensionality of the feature space [2,4].

Dirichlet Process: The Dirichlet process (DP) [2] is a stochastic process used in Bayesian nonparametric models of data, particularly in Dirichlet process mixture models whose sample paths are probability measures with probability one. It is a distribution over distributions, where each draw from a Dirichlet process is itself a distribution. For a random distribution to be distributed according to a DP, its marginal distributions have to be Dirichlet distributed. It is called a Dirichlet process because it has Dirichlet distributed finite dimensional marginal distributions. Distributions drawn from a Dirichlet process are discrete, but cannot be described using a finite number of parameters; thus the classification is a nonparametric model.

Latent Feature Models: Latent feature models represent a set of objects in terms of a set of latent features, each of which represents an independent degree of variation exhibited by the data. Such a representation of data is sometimes referred to as a distributed representation. In analogy to nonparametric mixture models with an unknown number of clusters, a Bayesian nonparametric approach to latent feature modeling allows for an unknown number of latent features. The stochastic processes involved here are known as the Indian buffet process (IBP) and the beta process (BP).

Hidden Markov Models: Hidden Markov models (HMMs) are popular models for sequential or temporal data, where each time step is associated with a state, with state transitions dependent on the previous state. An infinite HMM is a Bayesian nonparametric approach to HMMs, where the number of states is unbounded and allowed to

grow with the sequence length. It is defined using one DP prior for the transition probabilities going out from each state. To ensure that the set of states reachable from each outgoing state is the same, the base distributions of the DPs are shared and given a DP prior recursively.

Dependence Factor Graph Model: A factor graph consists of two layers of nodes, i.e., variable nodes and factor nodes, with links between them. The joint distribution over the whole set of variables can be factorized as a product of all factors. A popular solution [5] to multi-label classification is called binary relevance which constructs a binary classifier for each label, resulting in a set of independent binary classification problems.

Gaussian Process: Gaussian Process models [6] as they are applied in machine learning are an attractive way of doing non parametric Bayesian modelling in a supervised learning problem. A Gaussian process is a collection of random variables, any finite number of which have a joint Gaussian distribution. A key advantage of GP based modelling is its ability to learn hyper-parameters directly from data by maximizing the marginal likelihood. Like other kernel methods, the Gaussian Process can be optimized exactly, given the values of their hyper-parameters and this often allows a fine and precise trade-off between fitting the data and smoothing.

Survey of Existing Systems

Learning Word Vectors for Sentiment Analysis: Unsupervised vector-based approaches to semantics can model rich lexical meanings, but they largely fail to capture sentiment information that is central to many word meanings and important for a wide range of Natural Language Processing (NLP) tasks. This model [4] uses a mix of unsupervised and supervised techniques to learn word vectors, capturing semantic term document information as well as rich sentiment content. The model is evaluated using small, widely used sentiment and subjectivity corpora, a large dataset of movie reviews to serve as a more robust benchmark. The dataset of Pang and Lee [7], which contains subjective sentences from movie review summaries and objective sentences from movie plot summaries was used to compare the results of the Sup-port Vector Machine (SVM) model with the Latent Dirichlet Allocation (LDA) model. The SVM model performed better than LDA, which models latent topics directly.

Thumbs up? Sentiment Classification using Machine Learning Techniques: The problem of classifying documents not by topic, but by overall sentiment, is considered in [3] which determines whether a review is positive or negative. Using movie reviews as data and employing three machine learning methods, namely the Naive Bayes, maximum entropy classification and support vector machines, it was found that standard machine learning techniques definitively outperform human-produced baselines. A challenging aspect of this problem that seems to distinguish it from traditional topic-based classification is that while topics are often identifiable by keywords alone, sentiment can be expressed in a more subtle manner. For example, the sentence "How could anyone sit through this movie?" contains no single word that is obviously negative. Good indicator words for the sentiments (seven positive and negative) in movie reviews were chosen and converted into simple decision procedures that essentially count the number of the proposed positive and negative words in a given document. The study mainly focused on features based on unigrams (with negation tagging) and bigrams. The main drawback is that a human would easily detect the true sentiment of the review, but bag-of-features classifiers would presumably find these instances difficult, as there are many words indicative of the opposite sentiment to that of the entire review. Fundamentally, it seems that some form of discourse analysis is necessary using more sophisticated techniques than mere positional features. The results produced via machine learning techniques were found to be good in comparison to human generated baselines. In terms of relative performance, Naive Bayes tends to do the worst and SVMs tends to do the best.

Emotions from Text: Machine Learning for Text-based Emotion Prediction: In addition to information, text contains attitudinal and more specifically, emotional content. [8] explores text-based emotion prediction problem empirically, using supervised ma-chine learning with the SNoW learning architecture. The goal is to classify the emotional affinity of sentences in the narrative domain of children's fairy tales, for subsequent usage in appropriate expressive rendering of text-to-speech synthesis. Initial experiments on a preliminary dataset of 22 fairy tales show encouraging results over a naive baseline and BoW approach for classification of emotional versus non-emotional contents, with some dependency on parameter tuning.

The text based emotion prediction task (TEP) addresses what emotion or emotions most appropriately describe a certain text passage and determine the valance. The observations point to three issues: first, the current dataset is too small. Second, the data is not easily separable given the subjective nature of the task and the rather low inter annotator agreement. Third and finally, the EMOTION class is combined by basic emotion labels, rather than an original annotated label. The main drawback is that emotions are poorly understood and it is especially unclear which features may be important for their recognition from text.

Learning to Identify Emotions in Text: [2] describes experiments concerned with the automatic analysis of emotions in text. A large dataset of news titles is annotated for six basic emotions: anger, disgust, fear, joy, sadness and surprise and evaluated [9]. This paper describes experiments concerned with the emotion analysis of news headlines. In Knowledge-based Emotion Annotation, the task of emotion recognition is performed by exploiting the use of words in a text and in particular their co-occurrence with words that have explicit affective meaning. The Corpus based Emotion Annotation experiments were performed relying on blog entries from LiveJournal.com. The average length of the blog posts in the final corpus was 60 words per entry, 100-400 characters a length within a range comparable to one of the headlines. The implementations of five different systems for emotion analysis by using the knowledge-based and corpus based approaches are:

- WN-Affect (WordNet Affect) presence, which is used as a baseline system and which annotates the emotions in a text simply based on the presence of words from the WordNet Affect lexicon.
- Latent Semantic Analysis (LSA) single word, which calculates the LSA similarity between the given text and each emotion, where an emotion is represented as the vector of the specific word denoting the emotion (e.g. joy).
- LSA emotion synset, where in addition to the word denoting an emotion, its synonyms from the WordNet synset are also used.
- LSA all emotion words, which augments the previous set by adding the words in all the synsets labeled with a given emotion, as found in WordNet Affect.
- Naive Bayes (NB) trained on blogs, which is a Naive Bayes classifier trained on the blog data annotated for emotions.

As expected, different systems have different strengths. The system based exclusively on the presence of words from the WordNet Affect lexicon has the highest precision at the cost of low recall. Instead, the LSA system using all the emotion words has by far the largest recall, although the precision is significantly lower. In terms of performance for individual emotions, the system based on blogs gives the best results for joy and anger, which correlates with the size of the training data set. For all the other emotions, the best performance is obtained with the LSA models. The main drawback here is that the correlation between the emotions is not discussed, as only the emotion feature extraction is mainly dealt with here. Also, the inductive transfer of emotion and multi-task learning is not augmented.

Joint Emotion – Topic Modelling for Social Affective Text Mining: [10] is concerned with the problem of social affective text mining, which aims to discover the connections between social emotions and affective terms based on user-generated emotion labels. The model proposed is a joint emotion-topic model by augmenting latent Dirichlet allocation with an additional layer for emotion modeling. It first generates a set of latent topics from emotions, followed by generating affective terms from each topic. The model follows a three step generation process for affective terms, which first generates an emotion from a document-specific Dirichlet distribution, then generates a latent topic from a multi nominal distribution conditioned on emotions and finally generates document terms from another multi nominal distribution based on latent topics. Experimental results on an online news collection containing 2,858 articles from the Sina society channel show that the proposed model can effectively identify meaningful latent topics for each emotion. To achieve Social Affective Text Mining, first a baseline emotion-term model that uses Naive Bayes to model social emotion and affective terms via their co-occurrences is used. Then the emotion-topic model that can jointly estimate the latent document topics and emotion distributions in a unified probabilistic graphical model is applied. One of the key advantages for the proposed emotion topic model is its ability to uncover hidden topics that exhibit strong emotions. The latent topics can be categorized into three types.

- Empty topic which has little opportunity to be generated from any emotion.
- Single-emotion topic which is dominated by a specific kind of emotion like amusement, surprise, anger, boredom and warmness.

- Multi-emotion topic which shares a variety of emotions at the same time.

Evaluation on emotion prediction further verifies the effective-ness of the model. Emotional modelling is the key task that is performed here.

Joint Emotion Analysis via Multi-task Gaussian Process: The Intrinsic Coregionalisation Model (ICM), a low rank approach matrix which combines a vector-valued Gaussian Process is used in [11]. By extending the GP regression framework to vector valued outputs, the so called coregionalisation models are obtained. Specifically, a separable vector-valued kernel known as ICM is employed. This process is usually performed to learn the noise variance and kernel hyper parameters, including the coregionalisation matrix [12], [13]. Important reasons to employ the model is three-fold:

- Datasets for this task are scarce and small so it is hypothesized that a multi-task approach will result in better models by allowing a task to borrow statistical strength from other tasks.
- The annotation scheme is subjective and very fine-grained and is therefore heavily prone to bias and noise, both of which can be modeled easily using GPs.
- Finally, the goal is to learn a model that shows sound and interpretable correlations between emotions.

The Radial Basis Function (RBF) data kernel over a bag-of-words feature representation was used for the experiments. Words were down cased and lemmatized using the WordNetlemmatizer in the NLTK toolkit and then used the GPy toolkit to combine the kernel with a coregionalisation model over the six emotions, comparing a number of low-rank approximations. The results shows that the learned coregionalisation matrix, reorders the emotions to emphasize the learned structure. The resulting matrix follows a block structure, clustering some of the emotions. The two interesting behaviors are noted.

- Sadness and fear are highly correlated. Anger and disgust also correlate with them, although to a lesser extent and could be considered as belonging to the same cluster. There is a correlation between surprise and joy. These are intuitively sound clusters based on the polarity of these emotions.
- The model also learns anti-correlations, especially between joy/surprise and the other emotions.

The multi-task GP models perform better for smaller datasets, when compared to single-task models.

A Joint Model of Feature Mining and Sentiment Analysis for Product Review Rating: The information in customer reviews is of great interest to both companies and consumers. This information is usually presented as non-structured free-text and therefore automatically extracting and rating user opinions about a product is a challenging task. Moreover, this opinion highly depends on the product features on which the user judgments and impressions are expressed. The main aim in [14] is to predict the overall rating of a product review based on the user opinion about the different product features that are evaluated in the review. First, the system identifies the features that are relevant to consumers when evaluating a certain type of product, as well as the relative importance or salience of such features. The system then extracts from the review the user opinions about the different product features and quantifies such opinions. The salience of the different product features and the values that quantify the user opinions about them are used to construct a vector of feature intensities which represents the review that will be the input to a machine learning model that classifies the review into different rating categories. Over 1000 hotel reviews from booking.com are evaluated and the results compare favorably with those achieved by other systems addressing similar evaluations.

Sentence-level Emotion Classification with Label and Con-text Dependence: Predicting emotion categories such as anger, joy and anxiety, expressed by a sentence is challenging due to its inherent multi-label classification difficulty and data sparseness. [5] addresses these two challenges by incorporating the label dependence among the emotion labels and the context dependence among the contextual instances into a factor graph model. Specifically, we recast sentence level emotion classification as a factor graph inferring problem in which the label and context dependence are modeled as various factor functions. Empirical evaluation demonstrates the great potential and effectiveness of our proposed approach to sentence level emotion classification. Dependence Factor Graph (DFG) is used to model the label and context dependence in sentence-level emotion classification. In the DFG approach, both the label and context dependence are modeled as various factor functions and the learning task aims to maximize the joint probability of all these factor functions.

Empirical evaluation demonstrates the effectiveness of DFG approach to capturing the inherent label and context dependence. There are two basic observations.

Label Dependency: One sentence is more likely to take some pair of emotion labels (hate and angry) than some other pair of emotion labels (hate and happy).

Context Dependency: Two instances from the same con-text are more likely to share the same emotion label than those from a random selection.

Each instance is treated as a bag-of-words and transformed into a binary vector encoding the presence or absence of word unigrams. Three evaluation metrics to measure the performances of different approaches to sentence-level emotion classification are employed. First, the Hamming loss evaluates how many times an instance-label pair is misclassified considering the predicted set of labels and the ground truth set of labels. Second, the accuracy gives an average degree of the similarity between the predicted and the ground truth label sets of all test examples. Third, the F1-measure is the harmonic mean between precision and recall. It can be calculated from true positives, true negatives, false positives and false negatives, based on the predictions and the corresponding actual values. The DFG approach with both label and context dependence further improves the performance with a large margin, irrespective of the amount of training data available.

CONCLUSION

Machine learning approaches are a better option for detecting emotions in texts, even when there is only an indirect reference to it. The supervised learning approach is more often used in emotion detection because it usually leads to better results than unsupervised learning. The Naive Bayes is the primary classification algorithm used by many authors as their baseline. The Maximum Entropy and the Latent Dirichlet Process both outperforms the Naive Bayes classification algorithm [8] in terms of mean classification accuracy. The Support Vector Machines outperforms the NB, MaxEnt [5], DP, LDA [4] and Dependence Factor Graph Models [9] in terms of accuracy. The main advantages of using the SVM include:

- Ease to come up with a decision boundary if we have a non-linear kernel.
- SVM maximizes margin, so the model is slightly more robust.
- SVM supports kernels, so we can model even non- linear relations.
- In a high dimensional space, SVMs have been reported to work better for text classification.

SVMs are often prone to unbiased classification datasets, Gaussian Process does not usually suffer from this problem. The GP works best when the dataset is small and outperforms the SVM model. The GP-based modelling can learn hyper parameters directly from data by maximizing the marginal likelihood [6]. Like other kernel methods, the Gaussian Process can be optimized exactly, given the values of their hyper-parameters and this often allows a fine and precise trade-off between fitting the data and smoothing. These are the practical advantages of using a Gaussian Process:

- The kernel hyper-parameters can be learnt via evidence maximization.
- GPs provide full probabilistic prediction and an estimate of uncertainty in the prediction.
- GPs can be easily extended and incorporated into hierarchical Bayesian mode.

Hence we can conclude that, in case of large datasets, SVM performs better than all the available methodologies and , when we have a minimal dataset, GP outperforms SVM.

REFERENCES

1. Rich Caruana, 1997. Multitask Learning, M.S. thesis, University of Haifa.
2. Carlo Strapparava and Rada Mihalcea, 2008. Learning to identify emotions in text. In Proceedings of the ACM Symposium on Applied Computing.
3. Bo Pang and Lillian Lee, 2002. Thumbs up? Sentiment Classification using Machine Learning Techniques. In Proceedings of 2002 Conference on Empirical Methods in Natural Language Processing (EMNLP).
4. Andrew, L. Maas, Raymond E. Daly, Peter T. Pham, Dan Huang, Andrew Y. Ng and Christopher Potts, 2011. Learning Word Vectors for Sentiment Analysis. The 49th Annual Meeting of the Association for Computational Linguistics.

5. Shoushan Li, Lei Huang, Rong Wang and Guodong Zhou, 2015. Sentence-level Emotion Classification with Label and Context Dependence. In the Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing, pp: 10451053.
6. Carl Edward Rasmussen and Christopher K.I. Williams, 2006. Gaussian processes for machine learning, volume 1. MIT Press Cambridge.
7. Bo Pang and Lillian Lee, 2008. Opinion Mining and Sentiment Analysis. Foundations and Trends in Information Retrieval.
8. Cecilia Ovesdotter Alm, Dan Roth and Richard Sproat, 2005. Emotions from text: machine learning for text-based emotion prediction. In the Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing (HLT/EMNLP), pp: 579-586.
9. Carlo Strapparava and Rada Mihalcea, 2007. SemEval-2007 Task 14: Affective Text. In Proceedings of SEMEVAL.
10. Shengliang Xu, Li Zhang, Rong Yan, Zhong Su, Dingyi Han and Yong Yu, 2009. Joint Emotion-Topic Modeling for Social Affective Text Mining. Ninth IEEE International Conference on Data Mining, pp: 699-704.
11. Trevor Cohn, Daniel Beck and Lucia Specia, 2014. Joint Emotion Analysis via Multi-task Gaussian Processes. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp: 1798-1803.
12. Edwin, V. Bonilla, Kian Ming A. Chai and Christopher K.I. Williams, 2008. Multi-task Gaussian Process Prediction. Advances in Neural Information Processing Systems.
13. Ebden, M., 2008. Gaussian Processes for Regression: A Quick Introduction.
14. Jorge Carrillo de Albornoz, Laura Plaza, Pablo Gervas and Alberto Daz, 2011. A Joint Model of Feature Mining and Sentiment Analysis for Product Review Rating. ECIR 2011, LNCS 6611, pp: 5566.