

ORIGINAL RESEARCH

OPEN ACCESS

Full open access to this and thousands of other papers at <http://www.la-press.com>.

Fine-Grained Emotion Detection in Suicide Notes: A Thresholding Approach to Multi-Label Classification

Kim Luyckx, Frederik Vaassen, Claudia Peersman and Walter Daelemans

CLiPS Computational Linguistics Group, University of Antwerp, Belgium.

Corresponding author email: kim.luyckx@ua.ac.be

Abstract: We present a system to automatically identify emotion-carrying sentences in suicide notes and to detect the specific fine-grained emotion conveyed. With this system, we competed in Track 2 of the 2011 Medical NLP Challenge,¹⁴ where the task was to distinguish between fifteen emotion labels, from guilt, sorrow, and hopelessness to hopefulness and happiness.

Since a sentence can be annotated with multiple emotions, we designed a thresholding approach that enables assigning multiple labels to a single instance. We rely on the probability estimates returned by an SVM classifier and experimentally set thresholds on these probabilities. Emotion labels are assigned only if their probability exceeds a certain threshold and if the probability of the sentence being emotion-free is low enough. We show the advantages of this thresholding approach by comparing it to a naïve system that assigns only the most probable label to each test sentence, and to a system trained on emotion-carrying sentences only.

Keywords: emotion detection, multi-label classification, thresholds, probability estimates

Biomedical Informatics Insights 2012:5 (Suppl. 1) 61–69

doi: [10.4137/BII.S8966](https://doi.org/10.4137/BII.S8966)

This article is available from <http://www.la-press.com>.

© the author(s), publisher and licensee Libertas Academica Ltd.

This is an open access article. Unrestricted non-commercial use is permitted provided the original work is properly cited.



Introduction

According to the US National Institute of Mental Health (NIMH), suicide is one of the primary causes of death among adolescents and young adults. Consequently, a lot of resources and effort are being invested into the prevention of suicide—consider the multitude of initiatives such as the National Suicide Prevention Lifeline. Recently, the field of Natural Language Processing (NLP) has started to study suicide notes from a classification perspective in an attempt to understand, detect, and prevent suicidal behavior (e.g., Pestian et al, 2008).¹ Within this context, Track 2 of the 2011 Medical NLP Challenge was developed to kick-start innovation in emotion detection as well as in suicide prevention.

This paper presents a multi-label classification approach to emotion detection in suicide notes. In the field of computational linguistics, sentiment and subjectivity analysis are hot topics. Opinion mining is especially popular, as its practical applications are numerous. Emotion classification, on the other hand, is a more subtle and difficult task. Instead of simply classifying documents as positive or negative, it involves a more fine-grained classification scheme.

Emotions can be described using either dimensional or discrete emotion models. In dimensional models, documents can be placed in an emotion continuum defined by several dimensions. One such dimensional model is Russell's circumplex model of affect,² a two-dimensional space defined by "valence" and "arousal" dimensions. Discrete models, however, divide emotions into clearly delimited categories. The most famous discrete categorization scheme is Ekman's list of basic emotions, which include "fear", "anger", "sadness", "disgust", "surprise" and "happiness".³

While emotion classification is more fine-grained—and thus more difficult—than sentiment classification, it also allows for additional interesting applications. One such application is the detection of emotions in suicide notes, which is the focus of the task we describe in this paper. Being able to automatically annotate suicide notes for the emotions they contain would allow medical professionals to get a better understanding of the psychological processes that drive people to suicide. In addition, the knowledge of these emotional processes can help detect potential suicide cases and prevent them from happening.

Related Work

Approaches to emotion classification can be subdivided into two schools of thought: pattern-based approaches and machine learning approaches. Pattern-based approaches attempt to identify specific words or patterns that are characteristic of certain emotions. The AESOP system by Goyal et al (2010), for instance, is a rule-based system that attempts to analyze the affective state of characters in fables by identifying affective verbs and by projecting these verbs' influence on their patients.⁴ A considerable number of pattern-based experiments relies on sentiment or affect lexicons—lists of words annotated with their polarity or emotional contents. SentiWordNet and WordNet Affect are widely used examples of such emotion lexicons.^{5,6} More recently, Mohammad and Yang (2011) created a large emotion lexicon through crowd-sourcing,⁷ while Balahur et al (2011) built EmotiNet, a knowledge base of concepts with their associated affective value.⁸ These sentiment and emotion lexicons can be used to classify documents in pattern-based approaches, but they can also be used as a source of features in machine learning approaches.

Machine learning methods rely on learning algorithms to automatically identify patterns in the data. Supervised systems are trained with previously annotated data from which they learn which features are most salient to distinguish between classes. The resulting model is then applied to previously unseen test documents in order to determine their class. Keshtkar and Inkpen (2009), for instance, classified LiveJournal posts according to their mood.⁹ Vaassen and Daelemans (2011) attempted to classify sentences into one of the eight emotions on Leary's interpersonal circumplex, a framework for interpersonal communication,¹⁰ while Van Zaanen and Kanters (2010) tried to determine the mood of songs based solely on their lyrics.¹¹ Unsupervised systems, on the other hand, do not require any pre-annotated data. Instead, they attempt to cluster documents into classes given their intrinsic properties. Kim and Valitutti (2010) evaluate a range of unsupervised methods while attempting to classify text into dimensional and categorical emotion models.¹²

Very little research has yet been published regarding the automatic classification of suicide nodes. Pestian et al (2008) attempted to distinguish simulated from



real suicide notes and found that machine learning algorithms were able to outperform human annotators in terms of accuracy.¹ Matykiewicz and Pestian (2009) used unsupervised machine learning to distinguish between suicide notes and newsgroup conversations.¹³ To our knowledge, no research had been done on the detection and identification of emotions in suicide notes before the 2011 Medical NLP Challenge. The work done in the context of Track 2 of that challenge will set the standard for future work.¹⁴

Data Set

As is described by Pestian et al (2011), the Computational Medicine Center—a division of the Cincinnati Children’s Hospital Medical Center—provided a training data set of 600 suicide notes for Track 2 of the 2011 Medical NLP Challenge.¹⁴ Annotators were asked to annotate these suicide notes on the token level and choose from the fifteen emotions listed in Figure 1. As a result, some sentences have been assigned multiple labels and it is also possible for sentences to be devoid of emotion. While annotators were asked to annotate on the token level, these annotations were extrapolated to the sentence level for the purpose of the Medical NLP Challenge. The data set thus contains the same emotion(s) for every token in a sentence. Figure 1 shows the distribution of the fifteen emotion labels across sentences.

The data set presents several challenges. First of all, 12% of the annotations are labeled with more than one emotion. This turns the task from a relatively straightforward single-label classification task where each sentence belongs to a single class, into a multi-label classification task in which more than one class

can be assigned to a sentence. As we will describe in more detail later, our solution to this problem is two-fold: to get cleaner training data, we split up the multi-labeled sentences into single-labeled fragments (when possible); for testing, we use probability thresholds to determine when multiple emotions can be assigned to a test sentence.

A second challenge is that half of the sentences in the data set had not been assigned an emotion. However, these sentences cannot simply be said to belong to a large “no emotion” class, since they can also be sentences for which no annotator agreement was reached, as was confirmed by the organizers.¹⁴ Therefore, including a large “no emotion” class during training could have a negative effect on performance, since that class would be inconsistent and noisy. As our experiments will show, however, it is still beneficial to include the “no emotion” class in training, as the resulting classifier outperforms the classifier trained on emotion-annotated sentences only.

A final challenge presented by the data set is that the notes are riddled with spelling and grammatical errors. We did not attempt to correct them and instead hypothesize that these errors might be clues to identify certain emotions.

Methodology

The development of our emotion classification system can be divided into a pre-processing step, followed by two distinct phases: the *calibration phase*, where the classifier is calibrated on single-labeled data, and the *thresholding phase*, where the classification schemes and thresholds for multi-label classification are determined. Both phases use training data exclusively and are evaluated using ten-fold cross-validation. Below, each step in the process—represented graphically in the block diagram in Figure 2—will be elaborated on.

Data pre-processing

As previously mentioned, the training set contains 302 sentences (or 12% of all sentences) with more than one emotion label. As the occurrence of the same sentence with different class labels might confuse the classifier and cause a drop in performance, we extract all multi-labeled sentences from the training data and examine whether we can manually attribute each emotion label to a specific sentence fragment. If so,

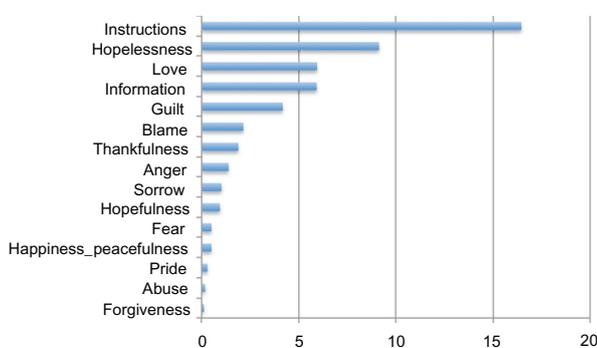


Figure 1. Emotion labels with an indication of their share in training as compared to the total number of emotion-annotated sentences ($N = 2,522$). Note that 49% of the training sentences (not in the figure) were not labeled.

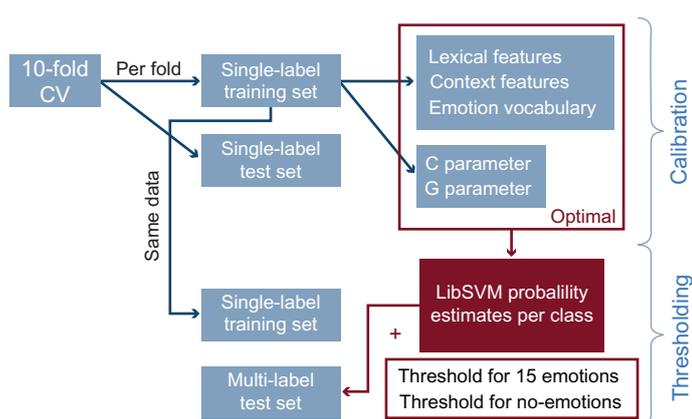


Figure 2. Block diagram showing the classification system and its calibration and thresholding phases.

the sentence is split into single-labeled fragments. Otherwise, the sentence is removed from the training set. The emotion labels in the training data are respected; labels are never changed or added. The resulting data set contains only single-labeled data. Table 1 shows some examples of re-annotated sentences.

Phase one: calibration

The first phase in the development of our emotion classification system involves learner parameter optimization and feature selection. This calibration step is performed in a ten-fold cross-validation setup using the re-annotated training set. Evaluation is done on single-label classification. As Support Vector Machines (SVM) have shown to be effective in the past, using an SVM classifier—in this case the LibSVM learning package—seems a logical choice.¹⁵ We use the default radial basis function kernel and optimize for the C (cost) and G (gamma parameter in the kernel function) parameters.

Three types of features are evaluated: lexical features, context features, and lexicon-based features. In the lexical feature set we include relative frequencies

of word unigrams, because the use of word bigrams, trigrams, character n -grams, or their combinations did not produce better results during preliminary experiments. Table 2 shows some of the most prominent lexical features (using chi-square for ranking) in the most frequently occurring emotions in the training data. While these features make intuitive sense, most other features are not unambiguously linked to a specific emotion.

After a careful review of the training data set, we noticed that some emotions seemed to co-occur more frequently (e.g., “information” and “instructions”, “forgiveness” and “love”). Moreover, when a sentence expressed a certain emotion, the emotion was likely to be continued over the following sentences. This was particularly the case for “no emotion”, “information”, “instructions”, and “hopelessness”, leading to the hypothesis that emotion-carrying sentences occur in clusters. We therefore decided to add context features as an extra feature type. These consist of the annotated labels of the n preceding and the n following sentences. Preliminary experiments showed that a left and right context of three emotions achieved the highest score.

A third type of features is based on a vocabulary of emotions divided in categories such as “depression”, “fear”, and “remorse”.¹⁶ For each of the adjectives in this list, we included relative frequencies in each sentence.

Phase two: thresholding

In the competition training set, 12% of the sentences were annotated with multiple emotions. We consider it safe to assume that this percentage will not be significantly different in the test set, as selection was done randomly. It is therefore essential to select a machine learning approach enabling multiple predictions per sentence.

LibSVM’s probability estimates offer one possible solution to this problem: as these estimates represent

Table 1. Examples of re-annotated sentences.

Original sentence	Labels	Fragments
The key is to my locker in the Bldg. where my tool kit is and I want him to have it.	Information Instructions	(i) The key is to my locker in the Bldg. where my tool kit is (ii) And I want him to have it.
Awfully sorry to take such a cowardly way, out, but I can see no other way. Everything I own I want you to have, please forgive me. I love you all.	Guilt Hopelessness Instructions Guilt Love	(i) Awfully sorry to take such a cowardly way out, (ii) But I can see no other way. (i) Everything I own I want you to have, (ii) Please forgive me. (iii) I love you all.

**Table 2.** Prominent lexical features in the most frequent emotions occurring in training.

Emotion	Prominent features
Instructions	<i>please, funeral, notify, give, pay, call, sell, forgive</i>
Hopelessness	<i>can't, longer, tired, stand, pain, sick, anymore, suffering</i>
Love	<i>love, darling, bless, dearly, !, heart, loving, affection</i>
Information	<i>\$, pocket, box, purse, closet, car, trunk, owe, bank</i>
Guilt	<i>forgive, sorry, caused, excuse, failed, guilty, fault, mercy</i>

the probability of a sentence belonging to each of the emotion classes, they allow us to identify the most probable emotions for each sentence. Using our calibrated classifier from Phase One, we experiment with various thresholds on these probability estimates. In order to ensure comparable results, we evaluate our system using ten-fold cross-validation using the same split as in the calibration phase, with the important difference that the test partitions now contain the original, multi-label training sentences. In other words, we train on single-label sentences and test on multi-label ones.

Apart from applying a threshold, we also explore the effect of the classification scheme. In a first type of experiments, we train our model on the fifteen emotions as well as on the “no emotion” sentences. This setup requires two thresholds: one for the emotions and one for the “no emotion” cases. A second type of experiments is only trained on emotion-carrying sentences, requiring only a single threshold. This setup is included because we suspect that the sentences annotated with “no emotion” contain a lot of noise. To evaluate how much better the thresholding approaches work compared to standard single-label classification, we also evaluate a naïve classifier trained on the full re-annotated data set (including the “no emotion” sentences), which assigns only the most probable label to each sentence. In short, the experiments for the second phase come in three flavors:

1. EMOTION/NO-EMOTION with thresholding: “no emotion” sentences included in training, two thresholds
2. EMOTION-ONLY with thresholding: trained on emotion-carrying sentences only, one threshold

3. EMOTION/NO-EMOTION without thresholding: “no emotion” sentences included in training, assigns only the most probable label

Results

In this section, we discuss the progress made by re-annotating the training data and evaluate the calibration and thresholding phases. We conclude this section by discussing the results obtained in the official test runs. We report on micro F_1 scores as well as precision and recall, as these are the main evaluation metrics used in Track 2 of the 2011 Medical NLP Challenge.

Data pre-processing

As a sanity check, we tested whether our decision to re-annotate complex sentences in the training set—i.e., sentences annotated with multiple emotions—was a good one. Table 3 shows classification performance before and after re-annotation, in an emotion detection experiment using token unigrams, using the EMOTION/NO-EMOTION classification scheme. Note that these scores were achieved using an optimized SVM classifier (see below for an account of our parameter optimization results). The results show that using the re-annotated data increased accuracy as well as micro F_1 , so we chose to use the re-annotated training set for all subsequent steps in the construction of our classification systems.

Evaluation of calibration phase

During the calibration phase, our aim was to find optimal feature types and to select the best performing parameters for our SVM classifier. Note that during this phase, we accept the label returned by LibSVM—i.e., the one with highest class probability estimate. Only during the second phase will we apply thresholding on the probability estimates and—potentially—assign multiple labels.

As described in the Methodology section, we experimented with token unigrams, but also consulted

Table 3. The effect of re-annotating data on emotion detection performance on training data (optimized scores).

	Original data	Re-annotated data
Accuracy	0.6096	0.6372
Micro F_1	0.5656	0.6004

lists of emotion-related keywords (+ EMOWORDS) and tried including context features (+ CONTEXT) in order to check whether these high-level feature types were helpful for emotion detection. Table 4 shows micro F_1 scores with each of these feature types. Although lists of emotion-related keywords are heavily used in emotion detection (see for instance Kim et al, 2010 and Ghazi et al, 2010),^{12,17} adding them to the token unigrams did not help improve classification performance. Including the emotions occurring in the context of the current sentence did not lead to satisfying results, even though there was some evidence to suggest clusters of same-type emotions (cf. Methodology section).

For each of the feature types mentioned above, we performed a grid search on LibSVM's C (cost) and G (gamma parameter in the radial basis function) parameters in order to optimize the classifier's performance. The overall best scores were obtained after optimization on token unigrams, as shown in Table 4. We thus decided to continue working with token unigrams in the thresholding phase.

Visualized in Figure 3 is the grid search applied to token unigrams. While varying the C (on the y-axis) and G (on the x-axis) parameters, we checked whether micro F_1 scores increased. The darker colored regions indicate low performance, while the lighter colored regions indicate higher performance. The best performing combination is $C = 8$ and $G = 8$, shown on the right-hand side of Figure 3.

Evaluation of thresholding phase

After having determined the best learner parameters and features for single-label classification, we experimented with ways to assign multiple labels to a single sentence. LibSVM provides probability estimates for each of the classes present in training, so instead of simply assigning the most probable emotion label to the test sentence, we can use probability thresholds: any emotion with a probability that exceeds

Table 4. Classification performance (in micro F_1) on training data by feature type, before and after optimization.

	Default parameters	Optimized
Token unigrams	0.4169	0.6004
Tokens + EMOWORDS	0.4192	0.5618
Tokens + CONTEXT	0.3337	0.5925

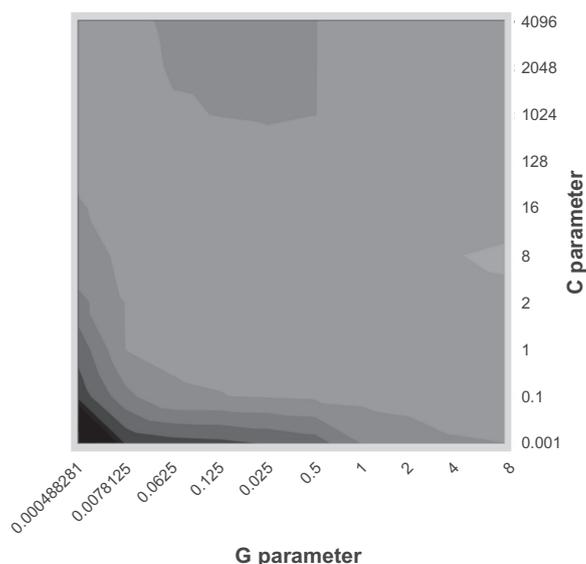


Figure 3. Visualization of the effect of C and G parameter tuning on micro F_1 , using token unigrams.

a given threshold will be assigned to the sentence. Where these thresholds lie, however, is dependent on the classification scheme (cf. Methodology section).

The EMOTION/NO-EMOTION scheme requires two thresholds: one for the emotion classes, and one for the “no emotion” class. Figure 4 shows a surface plot of the micro F_1 score for different values of *emotion* and *no-emotion* thresholds. The optimal values were 0.19 for the *emotion* threshold and 0.80 for the *no-emotion* threshold.

The EMOTION-ONLY experiment—where the classifier was trained on emotion-carrying sentences only—requires only one threshold. Figure 5 shows the effect of different emotion probability thresholds on precision, recall, and micro F_1 scores using this

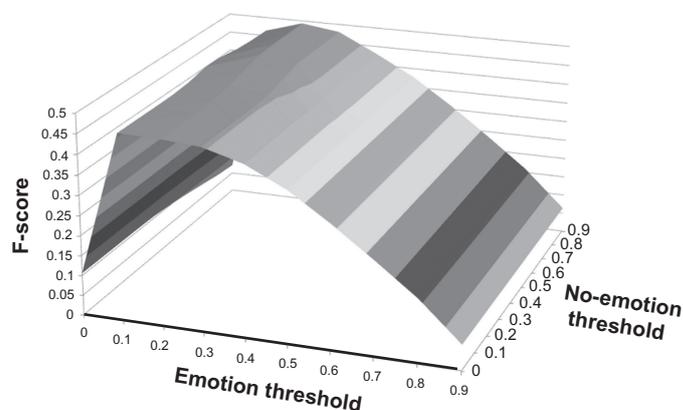


Figure 4. Effect of emotion and no-emotion probability thresholds on training performance (micro F_1).

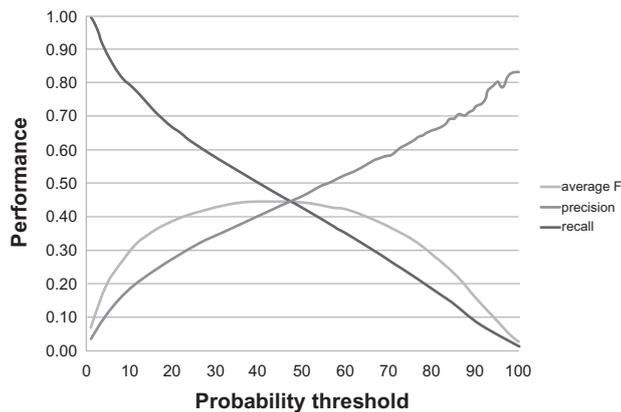


Figure 5. Effect of emotion probability thresholds on training performance.

classification scheme. The optimal value for the *emotion* threshold was 0.49.

Table 5 shows precision, recall, and micro F_1 scores for the two classification schemes with thresholds, as well as for a naïve variation on the EMOTION/NO-EMOTION experiment where only the most probable label was assigned (*no thresholding*). The EMOTION/NO-EMOTION system trained on both emotion-carrying and “no emotion” sentences performs best, with a micro F_1 score of 0.4954.

Evaluation on test data

For the official test run, we submitted the three systems developed during the second phase. The test set contained 300 suicide notes, annotated in the same fashion as the original training set, but previously unseen by our systems. Table 6 shows performance of the three systems in terms of precision, recall, and micro F_1 score. The n value in the last column indicates the number of emotions predicted by the system.

The results shown in Table 6 are in line with results during development, even to the point where they improve over results obtained during training (cf. Table 5). It is clear that the second system—trained only on the fifteen emotion classes, enhanced with a single emotion threshold—detects more emotions than the

last one and considerably more than our first, naïve system. In terms of micro F_1 , however, the last system—trained on emotions as well as “no emotion” cases—clearly outperformed the other two. This result is higher than the mean performance (0.4875) of the groups participating in the challenge, and only just below the median of 0.5027. The top scoring team’s result was 0.6139.

All of the systems shown in Table 6 started from training data we re-annotated. With hindsight, if we had continued to work with the original data—with the option of the same sentence having various emotion labels—performance on test data would have been better than the systems we actually submitted. The micro F_1 score of 0.5230 (precision = 0.5373; recall=0.5094; $n=1206$) achieved by not re-annotating is an improvement over our best system. Our decision to continue working with re-annotated sentences was supported by results during development, but worked to our disadvantage during the test phase.

Error analysis on the multi-label predictions made by the EMOTION/NO-EMOTION system in Table 6 reveals that multiple labels were predicted for 45 sentences (15% of all sentences in the test set). 80% of these decisions were partially correct, meaning that one of the predicted labels was the same as (one of) the gold standard labels. The system achieved a correct multi-label decision in 16% of these cases. The EMOTION-ONLY system hardly produced any multi-label annotations.

Some of the single labels most confused by the EMOTION/NO-EMOTION system are “instructions” and “information”, and “instructions” and a “no emotion” prediction. The feeling of “hopelessness” is often given the “no emotion” label, which is unfortunate considering the possible applications of emotion detection, more specifically the prevention of (repeated) suicide attempts. The other two systems, however, perform worse in this respect: they fail to identify any emotion in the test sentences annotated with the “hopelessness” label.

Table 5. Multi-label classification performance (in micro F_1) on training data, using token unigrams.

	Precision	Recall	Micro F_1	Std. dev.
EMO/NO-EMO (NO THRESHOLDING)	0.6732	0.3442	0.4536	0.0512
EMOTION-ONLY	0.4292	0.4730	0.4482	0.0595
EMOTION/NO-EMOTION	0.5013	0.4936	0.4954	0.0480

**Table 6.** Multi-label classification performance on test data.

	Precision	Recall	Micro F_1	N
EMO/NO-EMO (NO THRESHOLDING)	0.6763	0.3483	0.4598	655
EMOTION-ONLY	0.4674	0.4788	0.4730	1303
EMOTION/NO-EMOTION	0.5190	0.4859	0.5018	1191

Conclusion and Future Research

We presented experiments in fine-grained emotion detection applied to suicide notes. During development, we first re-annotated part of the data that was ambiguous and noisy, and then worked in two phases. The first phase of experiments focused on single-label classification experiments, analyzing the effect of the feature type used and the values of two classifier parameters. During the second phase, we shifted focus towards multi-label classification. Because the sentence-level annotations provided can contain multiple emotions per sentence, we devised a thresholding approach. By setting thresholds on the probability estimates per class, we allowed for multi-label predictions.

Test run results were in line with training results, with best performance achieved by a system trained on all fifteen emotion labels as well as on the “no emotion” class. In comparison with other teams participating in the challenge, we scored above the mean and only just below the median. Error analysis revealed that there was a substantial amount of confusion between “hopelessness” and a “no emotion” prediction. Considering the potential applications of emotion detection, this is unfortunate.

In order to achieve better results, an alternative technique could be to restrict to a small selection of emotions or to develop separate classifiers for each emotion, in order to particularly focus on emotions relevant for prevention purposes. The systems we developed were however challenged by the inconsistently annotated data. A considerable amount of noise was caused by a decision made by the challenge organizers to include sentences where no agreement was reached in a large group of “no emotion” cases. While making a distinction between “no-emotion” and “no-agreement” bears no clinical relevance, it would have had a considerable effect on classification results.

Acknowledgements

The research of Luyckx and Daelemans is partially funded through the IWT project AMiCA: Automatic

Monitoring for Cyberspace Applications. Frederik Vaassen’s research is funded through the EWI project TTNWW. Claudia Peersman’s research is funded through the IOF project DAPHNE.

Disclosures

Author(s) have provided signed confirmations to the publisher of their compliance with all applicable legal and ethical obligations in respect to declaration of conflicts of interest, funding, authorship and contributorship, and compliance with ethical requirements in respect to treatment of human and animal test subjects. If this article contains identifiable human subject(s) author(s) were required to supply signed patient consent prior to publication. Author(s) have confirmed that the published article is unique and not under consideration nor published by any other publication and that they have consent to reproduce any copyrighted material. The peer reviewers declared no conflicts of interest.

References

1. John P. Pestian, Pawel Matykiewicz, Jacqueline G. Phelan. Using natural language processing to classify suicide notes. In: *Proceedings of the Workshop on Current Trends in Biomedical Natural Language Processing*, BioNLP '08, Stroudsburg, PA, USA. Association for Computational Linguistics. 2008;96–7.
2. Russell JA. A circumplex model of affect. *Journal of Personality and Social Psychology*. 1980;39:1161–78.
3. Paul Ekman. Universals and cultural differences in facial expressions of emotion. In: *Nebraska Symposium on Motivation*. University of Nebraska Press, 1971.
4. Amit Goyal, Ellen Riloff, Hal Daumé, Nathan Gilbert. Toward Plot Units: Automatic Affect State Analysis. In: *Workshop on Computational Approaches to Analysis and Generation of Emotion in Text*, 2010.
5. Andrea E. Baccianella, Fabrizio Sebastiani. SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining. In: *Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10)*, Valletta, Malta, May 2010. European Language Resources Association (ELRA).
6. Strapparava C, Valitutti A. WordNet-Affect: an affective extension of WordNet. In: *Proceedings of LREC*. 2004;4:1083–6.
7. Saif Mohammad, Tony Yang. Tracking Sentiment in Mail: How Genders Differ on Emotional Axes. In: *Proceedings of the 2nd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis (WASSA 2.011)*, Portland, Oregon. Association for Computational Linguistics. June 2011:70–9.
8. Alexandra Balahur, Jesús M. Hermida, Andrés Montoyo. Detecting Implicit Expressions of Sentiment in Text Based on Commonsense Knowledge. In: *Proceedings of the 2nd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis (WASSA 2.011)*, Portland, Oregon. Association for Computational Linguistics. June 2011:53–60.



9. Fazel Keshtkar, Diana Inkpen. Using Sentiment Orientation Features for Mood Classification in Blogs. In: *IEEE International Conference on Natural Language Processing and Knowledge Engineering (IEEE NLP-KE 2009)*, Dalian, China, 2009.
10. Frederik Vaassen, Walter Daelemans. Automatic Emotion Classification for Interpersonal Communication. In: *Proceedings of the 2nd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis (WASSA 2011)*, Portland, Oregon, Association for Computational Linguistics. June 2011:104–10.
11. Menno van Zaanen, Pieter Kanters. Automatic Mood Classification Using TF*IDF Based on Lyrics. In: J. Stephen Downie and Remco C. Veltkamp, editors, *ISMIR*. International Society for Music Information Retrieval, 2010;75–80.
12. Sunghwan M. Kim, Alessandro Valitutti, Rafael A. Calvo. Evaluation of Unsupervised Emotion Models to Textual Affect Recognition. In: *Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text*, Los Angeles, CA. Association for Computational Linguistics. June 2010:62–70.
13. Matykiewicz P, Duch W, Pestian J. Clustering semantic spaces of suicide notes and newsgroups articles. In: *Proceedings of the Workshop on Current Trends in Biomedical Natural Language Processing*, BioNLP '09, Stroudsburg, PA, USA. Association for Computational Linguistics. 2009;179–84.
14. Pestian J, Matykiewicz P, Linn-Gust M, et al. Sentiment analysis of suicide notes: A shared task. *Biomedical Informatics Insights*: 2012;5 (Suppl. 1):3–16.
15. Chih-Chung Chang, Chih-Jen Lin. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*. 27 2011;2:27:1–27.
16. Drummond T. Vocabulary of emotions, s.d. URL <http://www.sba.pdx.edu/faculty/mblake/448/FeelingsList.pdf>.
17. Diman Ghazi, Diana Inkpen, Stan Szpakowicz. Hierarchical versus Flat Classification of Emotions in Text. In: *Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text*, Los Angeles, CA. Association for Computational Linguistics. June 2010;140–6.

Publish with Libertas Academica and every scientist working in your field can read your article

“I would like to say that this is the most author-friendly editing process I have experienced in over 150 publications. Thank you most sincerely.”

“The communication between your staff and me has been terrific. Whenever progress is made with the manuscript, I receive notice. Quite honestly, I’ve never had such complete communication with a journal.”

“LA is different, and hopefully represents a kind of scientific publication machinery that removes the hurdles from free flow of scientific thought.”

Your paper will be:

- Available to your entire community free of charge
- Fairly and quickly peer reviewed
- Yours! You retain copyright

<http://www.la-press.com>