

# Hierarchical Approach to Emotion Recognition and Classification in Texts

Diman Ghazi<sup>1</sup>, Diana Inkpen<sup>1</sup>, and Stan Szpakowicz<sup>1,2</sup>

<sup>1</sup> School of Information Technology and Engineering, University of Ottawa

<sup>2</sup> Institute of Computer Science, Polish Academy of Sciences

{dghaz038,diana,szpak}@site.uottawa.ca

**Abstract.** We explore the task of automatic classification of texts by the emotions expressed. We consider how the presence of neutral instances affects the performance of distinguishing between emotions. Another facet of the evaluation concerns the relation between polarity and emotions. We apply a novel approach which arranges neutrality, polarity and emotions hierarchically. This method significantly outperforms the corresponding “flat” approach which does not take into account the hierarchical information. We also compare corpus-based and lexical-based feature sets and we choose the most appropriate set of features to be used in our hierarchical classification experiments.

**Keywords:** Sentiment analysis, emotion in text, emotion recognition, text classification, hierarchical classification.

## 1 Introduction

In recent years there has been a growing interest in automatic identification and extraction of opinions, emotions, and sentiment in text. Motivations for this task include the desire to provide tools for information analysts in government, commercial, and political domains, who want to automatically track attitudes and feelings in on-line forums [1]. Emotions, an important element of human nature, have also been widely studied in psychology and behavioral sciences. They have also attracted the attention of researchers in computer science and particularly in computational linguistics.

This paper looks at the categorization of a sentence into six basic emotions (defined as emotions with universally accepted distinctive facial expressions). Those are *happiness*, *sadness*, *fear*, *anger*, *disgust*, and *surprise* [2]. We added a class *non-emotional* for the sentences which bear no emotion. These seven classes are common in much of the previous work [3, 4, 5, 6].

There has been progress in research on polarity and sentiment analysis, but little work has been done in automatic recognition of emotion in text. We assume that emotions carried by a sentence are not independent of their polarity; therefore we try to find a link between them and we want to apply classification methods, which consider these connections.

Our proposed hierarchical classification is a novel method in emotional analysis, which considers the relation between polarity and emotion of a text. The main idea is to order these categories and their relations in a hierarchical form and perform classification based on this hierarchy.

We have two forms of hierarchy for classification. Using *two-level classification*, we explore how neutral instances affect the emotional analysis. In *three-level classification*, the first step is to define whether the instances are emotional. Next, the instances defined as emotional in the previous step are classified on their polarity. In the third step we assume that, among the six emotions, the instances of *happiness* have positive polarity, while the other five emotions are regarded as negative polarity. That is why we take the negative instances from the second step and classify them into the five negative emotion classes.

Our experiments on data annotated with emotions show that this approach significantly outperforms the corresponding “flat” approach. We note significantly improved precision, recall and F-measure of all emotional classes.

The remainder of this paper is organized as follows. Section 2 briefly presents previous work. Section 3 gives an overview of the dataset, lexicons and different feature sets which we compare. In Section 4 we describe the hierarchical classification method and we evaluate it by comparing it to previous flat classification results. We discuss future work and present a few conclusions in Section 5.

## 2 Previous Work

Computational approaches to emotional analysis have focused on various emotion modalities, but only limited work has been done in the direction of automatic recognition of emotion in text [3]. In SemEval 2007, one of the tasks was carried out in an unsupervised setting and the emphasis was on the study of emotion in lexical semantics [7-10].

The participants in the SemEval 2007 workshop took a linguistic approach, using enriched lexical resources such as SentiWordNet and WordNetAffect [8]. They also used statistics gathered from Web search engines [9] based on the hypothesis that groups of words which co-occur with a given emotion word are highly likely to express the same emotion. Mihalcea and Strapparava [7] exploited the co-occurrence of words in the text with the words which have explicit affective meaning. They used the WordNet affect list as the direct affective words, and implemented a variation of Latent Semantic Analysis to yield a vector space model which allows for a homogeneous representation of words [7]. These tasks were tested on the emotion classification of news headlines extracted from news Web sites.

The existence of an annotated corpus with rich information about opinions and emotions would support the development and evaluation of NLP systems which exploit such information. In particular, statistical and machine learning approaches have become the method of choice for constructing a wide variety of practical NLP applications [1]. Emotional classification is not exemplified.

Aman and Szpakowicz [3, 11] have shown that the best results on a dataset annotated with emotions are achieved by a combination of corpus-based unigram features and lexical-based features using Support Vector Machine (SVM). In general,

two supervised machine learning algorithms, SVM and Naïve Bayes, have long been a method of choice for sentiment recognition at the text level [3, 5, 12, 13].

So far most of the research has been concentrated on the feature selections and applying lexical semantics rather than focusing on different learning schemes. In this work, in addition to comparing two different sets of features, corpus-based and lexically-based, we would like to focus more on a new learning approach, namely hierarchical classification.

Koller and Sahami [14] carried out the first proper study of a hierarchical text categorization problem in 1997. Afterwards more work in hierarchical text categorization has been reported [13, 15]. The work in [13] is more related to this paper. The authors applied a hierarchical approach in mood classification – classifying blog posts into 132 moods. Even though moods and emotions may seem similar, their classification is quite different. In mood classification we talk about a large number of mood classes. Also, the hierarchical structure applied is based on a definition of moods and relations among them which is completely different than our hierarchical structure for considering polarity in classifying emotions.

For these reasons, we propose a new method, called hierarchical classification, in our machine-learning approach to classifying blog sentences into Ekman’s six emotion classes and one non-emotional class.

### 3 Resources and Feature Sets

In this section, we will first explain the resources, namely data set and lexicons, which have been used in our experiments. Next, we will compare three set of features to find the most proper one for hierarchical classification.

#### 3.1 Resources

The statistical methods typically require training and test corpora, manually annotated with respect to each language-processing task to be learned [1].

The main consideration in the selection of data for the emotional classification task is that the data should be rich in emotion expressions in order to contain numerous learning instances. Another consideration is that the data should comprise enough instances of all the emotion categories. That is why personal texts such as diaries and blogs have received more attention recently [3, 4, 13]. As a result, we chose for our experiments a corpus of blog sentences annotated with emotion labels, discussed in [3].

Each sentence is tagged by a dominant emotion in the sentence, or as *non-emotional* if it does not include any emotion. The dataset contains 173 weblog posts annotated by two judges. It is annotated based on Eckman’s six emotions [2] at the sentence level. The dataset contains 4090 annotated sentences, 68% of which were annotated as non-emotional. The highly unbalanced dataset with 68% of non-emotional sentences as the highest class and 3% of the *fear* and *surprise* classes prompted us to reduce the number of non-emotional sentences to 38% of all the sentences by removing 2000 of the non-emotional sentences, to reduce the unbalance. Table 1 shows the details of the chosen dataset.

**Table 1.** Data set specifications

	<i>Domain</i>	<i>Size</i>	<i># classes</i>	<i>Min-Max%</i>
<b>Data set 1</b>	Weblogs	2090	7	6-38 %

We now compare three sets of features. The first set is corpus-based, so we need no other external resources; the other two sets are lexically-based. The lexically-based approach requires lexical-semantic resources. In our experiments, we use three *emotional lexicons*: Prior-Polarity subjectivity lexicon<sup>1</sup> [16], WordNet Affect<sup>2</sup> [17], and an emotion lexicon derived from *Roget's Thesaurus*<sup>3</sup> [18].

The prior-polarity subjectivity lexicon contains over 8000 subjectivity clues collected from a number of sources. To create this lexicon, the authors began with the list of subjectivity clues extracted from [19]. The list was expanded by using a dictionary and a thesaurus, and added positive and negative word lists from the General Inquirer<sup>4</sup> [16]. The words are grouped into strong subjective and weak subjective clues; Table 2 presents the distribution of their polarity.

The WordNet Affect lexicon contains six lists of words corresponding to the six basic emotion categories. It is the result of assigning a variety of affect labels to each synset in WordNet [17]. Table 3 shows the distribution of words in WordNet Affect.

The emotional lexicon derived from *Roget's Thesaurus* was created automatically finding the emotion-related words [11]. The total number of words selected for inclusion in this lexicon is 2622. The distribution of words in different emotion classes is shown in Table 4.

**Table 2.** Distribution of Prior-Polarity clues

<i>neutral</i>	<i>positive</i>	<i>negative</i>	<i>both</i>
6.9%	33.1%	59.7%	0.3%

**Table 3.** Distribution of the WordNet Affect emotional lexicon

<i>happiness</i>	<i>sadness</i>	<i>anger</i>	<i>disgust</i>	<i>surprise</i>	<i>fear</i>	<i>total</i>
398	201	252	53	71	141	1116

**Table 4.** Distribution of the *Roget's*-derived emotional lexicon

<i>happiness</i>	<i>sadness</i>	<i>anger</i>	<i>disgust</i>	<i>surprise</i>	<i>fear</i>	<i>total</i>
643	262	265	401	499	552	2622

<sup>1</sup> <http://www.cs.pitt.edu/mpqa>

<sup>2</sup> <http://www.cse.unt.edu/~rada/affectivetext/data/WordNetAffectEmotionLists.tar.gz>

<sup>3</sup> Several versions of *Roget's Thesaurus* are available. In the experiments reported here, the 1987 Penguin's *Roget's Thesaurus* was used [3].

<sup>4</sup> <http://www.wjh.harvard.edu/~inquirer/'p>

### 3.2 Feature Sets

We want to evaluate the performance of different lexicons by comparing them with unigram features. We also want to find the most appropriate set of features for our main experiments in the next section. Here, we perform three different experiments, all based on the flat classification approach.

The first experiment is a corpus-based classification which uses unigrams. The unigram models have been widely used in text classification, and shown to provide good results in sentiment classification tasks [12]. In this experiment, we use unigrams from the corpus, selected using feature selection methods from Weka<sup>5</sup>.

The second experiment consists in the classification with features derived from the polarity lexicon. The features here are the tokens that are common between the prior polarity lexicon and the chosen dataset – 796 tokens in total.

In the last experiment, we use a combination of the emotional lists of words from *Roget's Thesaurus* and WordNet Affect. All the common tokens between the combined lexicons and dataset's tokens comprise 1264 words.

Text classification quite often deals with high dimensionality of the feature space. Many learning algorithms do not scale to a high-dimensional feature space. SVM has been shown to give good performance in text classification experiments: it scales well to the large numbers of features [3, 12]. We also experimented with other classifiers: Naïve Bayes and Decision Tree. As expected, the results for our task were lower, so we do not present them here. Consequently, we use SVM as a machine-learning algorithm to be applied to the chosen features.

As a result of the emotional classification experiments using SVM (the SMO algorithm in Weka) [20] and setting “10-fold cross validation” as a testing option, we get the accuracy of 65.55% for the first set of features, 62.67% for the second set and 57.3% for the last feature set.

Based on the accuracy of the three experiments, unigram features outperform the other two types. Even though it is significantly better than the third one, its difference with the second feature set merits a discussion. Here, we would like to point out parameters – other than the accuracy of the results – which should be considered in choosing the list of features. The number of features is one of the main parameters. We are not interested in having better results by adding more features. What is more interesting is to find a list of fewer but more meaningful features which could contribute more to learning. In our experiments, the size of polarity features is quite smaller than the unigrams. Also, by checking the features manually we noticed that they appear to be more meaningful. For example, among the unigram features we have proper nouns such as names of people and countries. It is also possible to have misspelled tokens in unigrams, while the prior-polarity lexicon features are well-defined words usually considered as polar. Besides, lexical features are known to be more domain and corpus-independent. For these reasons, polar lexically-based features will be chosen for the main experiments in the next section.

---

<sup>5</sup> Machine-learning software available at <http://www.cs.waikato.ac.nz/ml/weka/>

## 4 Hierarchical Classification

Hierarchical categorization deals with categorization problems in which categories are organized in hierarchies. For most text categorization tasks the category hierarchies have been carefully composed by humans and represent our knowledge of the subject [15].

In this work, we use the hierarchical categories to impart an additional knowledge to our classification method. We will convey the information in two forms of hierarchy. The first one is a two-level hierarchy which represents the relation of emotion and neutrality in text. The second form is a three-level hierarchy which addresses the relation between polarity and emotions in addition to the relation between emotion and neutrality. That is based on the assumption that, among the six chosen emotions, *happiness* belongs to the positive polarity class, while the other five emotions are regarded as having negative polarity.

In the remainder of this section, we will give a detailed explanation of both proposed hierarchical forms. The experiments and results of applying them to our chosen data set will be presented.

### 4.1 Two-Level Classification

The main goal of this task is to find out how the presence of neutral instances affects the performance of features for distinguishing between emotional classes. This was motivated by a similar work in polarity classification [16].

In the two-level classifications, the first level, emotional versus non-emotional classification, tries to determine whether an instance is neutral or emotional. The second step takes all instances which level 1 classified as emotional, and tries to classify them into one of the six emotions. The two levels of this hierarchy are as follows:

- **Emotional versus non-emotional:** We change the annotated data set by keeping all the non-emotional instances as is, and changing the class of all the other six emotional classes to “Emo” (for *emotional*). The result of this experiment and its comparison with the flat classification shows that the recall of non-emotional instances decreases while the precision increases. This happens to non-emotion class that used to be the dominant class in flat classification but it no longer dominates in hierarchical classification. Classifiers tends to give priority to a dominant class, so more instances are placed in this class; thus, classification achieves low precision and high recall. Hierarchical method tends to produce higher precision in this case.
- **Six-class classification:** After classifying the instances based on whether they are emotional, now we would like to classify the instances which level 1 classified as emotional. That will be done in two ways. In the first experiment, we assume that all the non-emotional instances are correctly classified, so we would only need to be concerned about the mistakes in level 2 (this is called *gold standard*). The second experiment considers the misclassified non-emotional instances of the first level as well. Both sets of results and the flat classification results are shown in Table 5.

**Table 5.** Two-level emotional classification (the highest precision, recall, and F-measure values for each class are shown in bold)

		Two-level classification			Flat classification		
		Precision	Recall	F-measure	Precision	Recall	F-measure
1 <sup>st</sup> level	<i>Emo</i>	0.88	0.85	0.86	--	--	--
	<i>Non-emo</i>	<b>0.88</b>	0.81	<b>0.84</b>	0.54	<b>0.87</b>	0.67
2 <sup>nd</sup> level gold standard	<i>happiness</i>	0.59	<b>0.95</b>	<b>0.71</b>	<b>0.74</b>	0.60	0.66
	<i>sadness</i>	<b>0.77</b>	<b>0.49</b>	<b>0.60</b>	0.69	0.42	0.52
	<i>fear</i>	<b>0.91</b>	<b>0.49</b>	<b>0.63</b>	0.82	<b>0.49</b>	0.62
	<i>surprise</i>	<b>0.75</b>	<b>0.32</b>	<b>0.45</b>	0.64	0.27	0.37
	<i>disgust</i>	0.66	<b>0.35</b>	<b>0.45</b>	<b>0.68</b>	0.31	0.43
	<i>anger</i>	<b>0.72</b>	<b>0.33</b>	<b>0.46</b>	0.67	0.26	0.38
2 <sup>nd</sup> level based on the 1 <sup>st</sup> level result	<i>happiness</i>	0.59	<b>0.96</b>	<b>0.73</b>	<b>0.74</b>	0.60	0.66
	<i>sadness</i>	<b>0.79</b>	<b>0.48</b>	<b>0.60</b>	0.69	0.42	0.52
	<i>fear</i>	<b>0.92</b>	<b>0.54</b>	<b>0.68</b>	0.82	0.49	0.62
	<i>surprise</i>	<b>0.66</b>	0.25	<b>0.37</b>	0.64	<b>0.27</b>	<b>0.37</b>
	<i>disgust</i>	0.65	<b>0.33</b>	<b>0.44</b>	<b>0.68</b>	0.31	0.43
	<i>anger</i>	0.65	<b>0.30</b>	<b>0.41</b>	<b>0.67</b>	0.26	0.38

By comparing the results of the second level of *gold standard* classification with the flat classification, we can see that the F-measure of all the emotional classes in the two-level experiment is higher than the F-measure of the emotional classes in the flat classification. In two emotion classes, however, the precision of the flat approach is higher. In *disgust* the difference is insignificant but the difference between precision and recall of the *happiness* class in the flat approach and the two-level approach cannot be ignored. This can be explained by the fact that at the second level of the two-level classification we do not have the non-emotional instances any more. The *happiness* class is the dominant class, with 42% of all the instances. This makes the classifier consider most all the instances it is not sure about as *happiness*. This results in high recall and low precision for the *happiness* class. We hope to address this big gap between precision and recall of the *happiness* class in the next experiments, three-level classification, which separates *happiness* from the other five emotions. So, it makes the number of instances of each level more balanced.

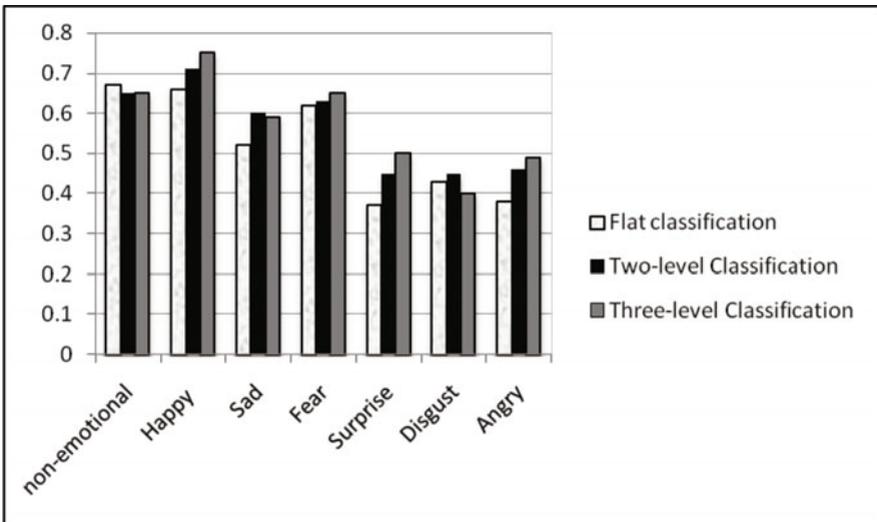
At the second level, when we consider the mistakes of the first level, we have to ignore both the false-negative emotional instances and false-positive non-emotional instances in order to obtain the set of instances which we use in the second level; therefore the number of instances will drop to 1101 from 1290. Because of the different number of instances, in this case the results are not quite comparable with the results of the original one-level task. Despite having to deal with the noise from the first level, the hierarchical approach still gets higher F-measures for five of the emotion classes and equal F-measure for the *surprise* class.

## 4.2 Three-Level Classification

In this approach, we go even further: we break the seven-class classification task into three levels. We add an assumption that the *happiness* class is positive and the remaining five emotions are negative.

**Table 6.** Three-level emotional classification (the highest precision, recall, and F-measure values for each class are shown in bold)

		Three-level Classification			Flat Classification		
		Precision	Recall	F-measure	Precision	Recall	F-measure
1 <sup>st</sup> level	<i>Emo</i>	0.88	0.85	0.86	--	--	--
	<i>Non-emo</i>	<b>0.88</b>	0.81	<b>0.84</b>	0.54	<b>0.87</b>	0.67
2 <sup>nd</sup> level gold standard	<i>positive</i>	<b>0.89</b>	<b>0.65</b>	<b>0.75</b>	0.74	0.60	0.66
	<i>negative</i>	0.79	0.94	0.86	--	--	--
3 <sup>rd</sup> level gold standard	<i>sadness</i>	0.63	<b>0.54</b>	<b>0.59</b>	<b>0.69</b>	0.42	0.52
	<i>fear</i>	<b>0.88</b>	<b>0.52</b>	<b>0.65</b>	0.82	0.49	0.62
	<i>surprise</i>	<b>0.79</b>	<b>0.37</b>	<b>0.50</b>	0.64	0.27	0.38
	<i>disgust</i>	0.42	<b>0.38</b>	0.40	<b>0.68</b>	0.31	<b>0.43</b>
	<i>anger</i>	0.38	<b>0.71</b>	<b>0.49</b>	<b>0.67</b>	0.26	0.38
2 <sup>nd</sup> level based on the 1 <sup>st</sup> level result	<i>positive</i>	<b>0.90</b>	<b>0.66</b>	<b>0.76</b>	0.74	0.60	0.66
	<i>negative</i>	0.78	0.94	0.85	--	--	--
3 <sup>rd</sup> level based on the 2 <sup>nd</sup> level result	<i>sadness</i>	0.64	<b>0.64</b>	<b>0.59</b>	<b>0.69</b>	0.42	0.52
	<i>fear</i>	<b>0.90</b>	<b>0.53</b>	<b>0.67</b>	0.82	0.49	0.62
	<i>surprise</i>	<b>0.83</b>	<b>0.35</b>	<b>0.50</b>	0.64	0.27	0.37
	<i>disgust</i>	0.37	<b>0.45</b>	0.40	<b>0.68</b>	0.31	<b>0.43</b>
	<i>anger</i>	0.35	<b>0.56</b>	<b>0.43</b>	<b>0.67</b>	0.26	0.38

**Fig. 1.** The comparison of F-measure result of the flat classification with the hierarchical classification approaches for seven classes

In the three-level classifications, the first level is the same as in the previous approach. After determining the non-emotional instances of the first level, in the second level we classify into *positive* and *negative* the instances which level 1 classified as emotional. We only consider *happiness* as positive. Finally, we classify the negative instances into five negative emotion classes. The results of this classification are shown in Table 6.

As we can see in the results of the second level, we increased the precision of the *happiness* class. In the polarity classification level (the second level) the data are almost balanced, with 42% of positive instances. That makes the instances defined as *happiness* more precise.

At the third level, except in the class *disgust*, we see an increase in the F-measure of all classes in comparison to both the two-level and flat classification. It is hard to compare the result of all three applied approaches from the above tables. We have combined all the results in Figure 1, which displays the F-measure of each class for the three approaches.

## 5 Conclusions and Future Work

The focus of this study was an emotional analysis and classification of emotions in sentences. We first defined two different sets of lexicon-based features to compare with the bag-of-words classification method. As a result, we noticed that unigram features were not much better than polarity lexicon features. Besides, the polarity lexicon features had some benefits over unigrams. In particular, they are supposed to be less corpus-dependent. They also have fewer features than unigram sets; therefore the rest of our experiments were based on the polarity lexicon features.

In the emotional classification we noticed that having non-emotional instances in the dataset degrades the results significantly; therefore we applied a two-level classification which defines the non-emotional instances in one step and considers the rest as emotional. As a result we saw a considerable improvement in the classification results.

The second part of the hierarchical classification experiments considers the polarity of emotions, that is, their positivity or negativity. Here we assume that *happiness* is a positive emotion, while *sadness*, *surprise*, *fear*, *disgust*, and *anger* are negative. We added one more step into our hierarchy to convey the information of the relation between polarity and emotions in our system.

On the other hand, a classifier trained on unbalanced data gives biased results for the classes with more instances. In this case most of the learning algorithms behave just as people do. They mainly learn the dominant class and if they see an instance they do not have information about, they will classify it as an instance of the bigger class, since it is more probable. The hierarchical classification approach was better at dealing with the highly unbalanced data.

In the future, we plan to expand our work by testing on other available emotion-annotated data sets. There are three other available datasets annotated with emotions. The first one is a data set of 700 sentences extracted from blogs, which are annotated with nine emotions and one neutral class [4]. The second available dataset is composed of sentences in the narrative domain of nineteen children's fairy tales

which have affect labels from two annotators. The affects are the same as the six emotions we considered, with the difference that the *surprise* class is broken into two subclasses of positive surprise and negative surprise [5]. The last set is the TextAffect dataset available for SemEval 2007, Task 14 [6]. We are interested in seeing how the emotional classification of these data sets will be affected by our proposed classification approach.

Another interesting future task would be to consider different levels of our hierarchy as different tasks which could be handled differently. In the last experiment, the three-level classification, we had three different tasks, namely emotional versus non-emotional, polarity, and emotional classification. Each of these tasks has its own specification; therefore we can definitely benefit from analyzing each task separately and defining different sets of features and classification methods for each task rather than using the same method for every task. One of the first steps for our future work will be using the emotional lexicon and their corresponding features in the third level of our hierarchy – classifying five negative emotions. We believe that using the features derived from the two emotion lexicons will improve the results of the third level.

## References

1. Wiebe, J., Wilson, T., Cardie, C.: Annotating Expressions of Opinions and Emotions in Language. *Language Resources and Evaluation* 39, 165–210 (2005)
2. Ekman, P.: An Argument for Basic Emotions. *Cognition and Emotion* 6, 169–200 (1992)
3. Aman, S.: Identifying Expressions of Emotion in Text. Master's thesis, University of Ottawa, Ottawa, Canada (2007)
4. Neviarouskaya, A., Prendinger, H., Ishizuka, M.: Compositionality Principle in Recognition of Fine-Grained Emotions from Text. In: *Proc. Third International ICWSM Conference*, pp. 278–281 (2009)
5. Alm, C.O., Roth, D., Sproat, R.: Emotions from text: machine learning for text- based emotion prediction. In: *Proc. Joint Conference on Human Language Technology / Empirical Methods in Natural Language Processing (HLT/EMNLP 2005)*, Vancouver, Canada, pp. 579–586 (2005)
6. Strapparava, C., Mihalcea, R.: SemEval-2007 Task 14: Affective Text (2007)
7. Strapparava, C., Mihalcea, R.: Learning to Identify Emotions in Text. In: *Proc. ACM Symposium on Applied computing, Fortaleza, Brazil*, pp. 1556–1560 (2008)
8. Chaumartin, F.: Upar7: A knowledge-based system for headline sentiment tagging. In: *Proc. SemEval 2007, Prague, Czech Republic (June 2007)*
9. Kozareva, Z., Navarro, B., Vazquez, S., Montoyo, A.: Ua-zbsa: A headline emotion classification through web information. In: *Proc. SemEval 2007, Prague, Czech Republic (June 2007)*
10. Katz, P., Singleton, M., Wicentowski, R.: Swat-mp: the semeval-2007 systems for task 5 and task 14. In: *Proc. SemEval 2007, Prague, Czech Republic (June 2007)*
11. Aman, S., Szpakowicz, S.: Using Roget's Thesaurus for Fine-grained Emotion Recognition. In: *Proc. Third International Joint Conf. on Natural Language Processing (IJCNLP)*, Hyderabad, India, pp. 296–302 (2008)
12. Kennedy, A., Inkpen, D.: Sentiment classification of movie reviews using contextual valence shifter. *Computational Intelligence* 22, 110–125 (2006)

13. Keshtkar, F., Inkpen, D.: Using Sentiment Orientation Features for Mood Classification in Blog Corpus. In: IEEE International Conf. on Natural Language Processing and Knowledge Engineering, Dalian, China, September 24-27 (2009)
14. Koller, D., Sahami, M.: Hierarchically Classifying Documents Using Very Few Words. In: Proc. International Conference on Machine Learning, pp. 170–178 (1997)
15. Kiritchenko, S., Matwin, S., Nock, R., Fazel Famili, A.: Learning and Evaluation in the Presence of Class Hierarchies: Application to Text Categorization. LNCS, pp. 395–406. Springer, Heidelberg (2006)
16. Wilson, T., Wiebe, J., Hoffmann, P.: Recognizing contextual polarity: an exploration of features for phrase-level sentiment analysis. *Computational Linguistics* 35(3), 399–433 (2009)
17. Strapparava, C., Valitutti, A.: WordNet-Affect: an affective extension of WordNet. In: Proc. 4th International Conference on Language Resources and Evaluation (LREC 2004), Lisbon, Portugal, pp. 1083–1086 (2004)
18. Jarmasz, M., Szpakowicz, S.: Roget's Thesaurus and Semantic Similarity. In: Nicolov, N., Bontcheva, K., Angelova, G., Mitkov, R. (eds.) Recent Advances in Natural Language Processing III: Selected Papers from RANLP, John Benjamins, Current Issues in Linguistic Theory, vol. 260, pp. 111–120 (2003)
19. Riloff, E., Wiebe, J.: Learning extraction patterns for subjective expressions. In: Proc. Conference on Empirical Methods in Natural Language Processing (EMNLP 2003), Sapporo, pp. 105–112 (2003)
20. Witten, I.H., Frank, E.: *Data Mining: Practical Machine Learning Tools and Techniques*, 2nd edn. Morgan Kaufmann, San Francisco (2005)