Emotion Analysis of Internet Chat

Shashank and Pushpak Bhattacharyya
Computer Science and Engineering
IIT Bombay
Mumbai, India
(shashank,pb}@cse.iitb.ac.in

Abstract

We present a system for Emotion Analysis of Instant Messages (IM). Using Instance Based classifier we have shown that our system can outperform similar systems in the IM domain. Tagged instant messages and elaborate feature engineering can help a lot in increasing the performance of text classification of unstructured, ungrammatical text. The impact of class imbalance on classification has been studied and demonstration has been made of how undersampling can help mitigate this problem.

1 Introduction

Of late, *Instant Messaging* (IM) has been made popular by *Instant Messaging Clients* like AIM, MSN and more recently, GTalk. With this medium gaining vitality as a form of communication, a natural interest in a proper understanding of the peculiarities associated with Instant Messaging based communication in particular and *Computer Mediated Communication* (CMC) in general has increased too.

An important aspect of such CMC as IM is the *affective* or emotional content of the information involved. An ability to identify affective content and to classify the nature of the affect has myriad practical applications. While some interest has been reported from the point of view of administration and moderation of such communication (Holzman and Pottenger, 2003), a bigger motivation has come from the area of Affective User Interfaces (AUI) (Boucouvalas, Zhe and Xu, 2002; Liu, Lieberman and Selker, 2003). Such AUIs include improved chat clients that can provide real time feedback on the users' emotional state inferred from the conversation text. Another application of such ability can be automating facial expression of *Avatars* in online games, especially *Massively Multiplayer Online Role Playing Games* (MMORPGs).

2 Problem Domain

While much work has gone into emotion analysis of text from domains like news headlines and blog posts (Strapparava and Mihalcea, 2008; Wiebe and Cardie, 2005), relatively less attention has been given to similar analysis of instant messages. This can partly be explained by important differences in the nature of data that is typical of the two domains. The text in instant messages is far less structured than news headlines and blog posts, is less grammatical, has more unintended typographical errors, has special morphological conventions like vowel elongation and, last but not the least, is loaded with a colloquialism of its own, resulting in a large number of out-of-vocabulary (OOV) words. See Tagliamonte and Derek (2008) for a study of IM-speak from a linguistic point of view.

All these irregularities point towards an inherent difficulty in identification of emotion in instant messages as compared to more structured data from sources of formal text. For a detailed analysis of these difficulties see Schmidt and Stone (*http://www.trevorstone.org/school/ircsegmentation.pdf*).

The roadmap of the paper is as follows. Section 3 is on related work. Data preparation is described in section 4. Section 5 gives the experiments and results. Section 6 compares our work with existing one. Section 7 draws conclusions and points to future work.

Proceedings of ICON-2008: 6th International Conference on Natural Language Processing
Macmillan Publishers, India. Also accessible from http://ltrc.iiit.ac.in/proceedings/ICON-2008

3 Related Work

Wiebe and Cardie (2005) have made an extensive study of the discernibility of emotions in news headlines by human annotators. A reported inter-annotator disagreement of 20-30% clearly indicates yet another challenge in such a task, even for structured and formal text.

Strapparava and Mihalcea (2008) have reported results for emotion analysis of news headlines and blog posts using a range of techniques including keyword-spotting, Latent Semantic Analysis (LSA), Naïve Bayes, rule based analysis and Pointwise Mutual Information (PMI). Aman and Szpakowicz (2007) have reported results for similar dataset using Naïve Bayes and Support Vector Machines (SVM).

In the domain of instant messages, Boucouvalas *et al.* (2002) have reported the development of a user interface for real time feedback on emotions in a chat client. Holzman and Pottenger (2003), probably closest to our work, have reported very encouraging results on emotion analysis of internet chat using Text to Speech (TTS) conversion and subsequent learning based on phonetic features.

4 Data Preparation

4.1 Data Acquisition

We have used data from two sources. A bigger dataset of 10567 sentences is the NPS (Naval Post Graduate School) chat corpus[1]. Out of these only 7933 sentences, which were part of conversation, have been used (rest all were system messages). This corpus consists of chat logs gathered from different online chatting services converted to XML. Each statement from one of the participants of a chatting session is converted to one *Post* node. Each sentence is tokenized and tagged for Part-of-Speech (POS) information as well as dialogue act. A typical *Post* node from the corpus is shown in Figure 1. For detailed information about the corpus and dialogue acts refer to Forsyth and Martell (2007). Henceforth we will refer to this dataset as the NPS set.

```
<Post class="Statement" user="10-26-teensUser66">
    I have a problem with people PMing me to lol
    <terminals>
        <t pos="PRP" word="I"/>
        <t pos="VBP" word="have"/>
        <t pos="DT" word="a"/>
        <t pos="NN" word="problem"/>
        <t pos="IN" word="with"/>
        <t pos="NNS" word="people"/>
        <t pos="VBG" word="PMing"/>
        <t pos="PRP" word="me"/>
        <t pos="^RB" word="to"/>
        <t pos="UH" word="lol"/>
    </terminals>
</Post>
```

Figure 1: A typical *post node*

A smaller set of 2980 sentences was prepared from a set of logs[2]. Unlike the NPS set, this set was raw chat logs. We converted the smaller set called IRC[3] (Internet Relay Chat) set henceforth to XML with a schema close to that of NPS set. We first tokenized the sentences in raw IRC set using TreebankWordTokenizer from the NLTK library[4]. The tokens were subsequently POS tagged using a Hidden Markov Model (HMM) based tagger trained on the NPS set. An important difference between the two data sets is that the IRC set does not have any dialogue act tagging like the NPS chat. This avoided solving yet another classification problem (of dialog act tagging) on the IRC set.

[1] http://faculty.nps.edu/cmartell/NPSChat.htm

[2] http://www.demo.inty.net/Units/chatcorpus.htm

[3] http://en.wikipedia.org/wiki/Internet_Relay_Chat

[4] http://nltk.org

4.2 Data Normalization

Considering the similarity between the language used in IM and that in text messaging using Short Messaging Service (SMS) and taking a cue from the research in normalization of SMS text to formal text, we have explored the possibility of an improvement in classification performance due to such *translation*. To this end, we *translated* both the data sets using two independent web based SMS translation services *noslang*[5] and *transl8it*[6]. After translating raw sentences from the original corpora we re-tokenized and re-POS tagged the new corpora using TreebankWordTokenizer and a maximum entropy based Treebank POS tagger trained on the Brown Corpus respectively, both from the NLTK library.

4.3 Emotion Annotation

We created a web-based interface for annotating data with emotion values. The set of emotions used was the basic set of six emotions proposed by Ekman and Friesen (1996): *anger* (ANG), *disgust* (DIS), *fear* (FEA), *happiness* (HAP), *sadness* (SAD) and *surprise* (SUR). Besides these, a seventh emotion of *neutral* (NEU) was used for non-affective sentences. All 13547 sentences were annotated using a web-based interface. Sentences were presented to the annotator in-context (i.e., maintaining the order of the original conversation) and each sentence was assigned to *exactly* one of the six emotion categories, the one that was most prominent in the sentences (multiple emotion categories for a single sentence was avoided for simplicity). Sentences that were not affective were assigned the *neutral* category.

The complete dataset was annotated by one of the authors and hence we could not perform any inter-annotator agreement study on the resulting annotation.

The distribution of sentences among different categories has been shown in Tables 1 and 2. Typical of the domain, the distribution is highly skewed with the NEU and HAP classes taking up most of the share.

Emotion	**# of sentences**	**% of sentences**
ANG	281	3.54
DIS	274	3.45
FEA	153	1.92
HAP	2644	33.3
SAD	411	5.1
SUR	377	4.7
NEU	3793	47.8

Table 2: Distribution of sentences among emotion classes (NPS set)

Emotion	**# of sentences**	**% of sentences**
ANG	109	3.65
DIS	111	3.72
FEA	51	1.71
HAP	1014	34.0
SAD	119	3.99
SUR	132	4.42
NEU	1444	48.4

Table 3: Distribution of sentences among emotion classes (IRC set)

4.4 Feature Set

We have used a reasonably large ensemble of features. In all, there were 71 attributes for the NPS set and 70 for the IRC set (the IRC set did not have the dialogue act tagging, as mentioned before). The attributes can be classified into the following broad categories:

[5] http://www.noslang.com
[6] http://www.transl8it.com

4.4.1 Features obtained from the data (internal to the data)

- **Simple Counts:** number of words, length of longest word, length of shortest word and average world length
- **Part of Speech:** frequency count for different parts of speech
- **Emoticons:** presence and frequency of emoticons (a.k.a, Smiley)
- **Affective Morphology:** presence and strength of vowel elongation (e.g. *wohoooooo! I won!*) , consonant repetitions (e.g, *hahahaha*) and Capitalizations (e.g. *you are a BIG loser!)*
- **Punctuations:** presence and frequency of "?" and "!" marks
- **Dialogue Act:** this feature was available and hence used only for the NPS set.

Besides these attributes a set of other attributes was used that were obtained with the help of external resources.

4.4.2 External Features given by other tools and resources

We obtained a set of affective keywords belonging to categories like *Negative, Positive,* and *Pleasur* from the General Inquirer corpus[7]. Frequency of keywords from these categories was used as a feature.

We also used as feature the similarity of words in a sentence with root emotion words (*joy, anger, sadness etc.*) in Wordnet[8]. WNSimilarity API in Perl[9] was used for the same. The similarity measure used was *lesk*.

An important step in this process was to measure the similarity for *slang* words. As is suggested in Tagliamonte and Derek (2008), a non-trivial amount of IM-speak consists of non-standard words (e.g., abbreviations like *LOL* for ***L**aughing **O**ut **L**oud* and terms like *biatch* for *bitch*). Words and terms from this *language* have almost always a very high affective value and hence very crucial for our purpose. Since these words are not present in Wordnet, we used another resource for this purpose. For each word in the corpus that was not present in Wordnet, we looked for its definition in the *urbandictionary*[10]. *Urbandictionary* is a collaborative online dictionary for slang. We used rankings determined by user voting on the website for each definition to determine whether a definition was reliable enough. This was necessary, given that this is a community generated content and not as reliable as a resource like Wordnet. For each term for which at least one definition could be found, its similarity to root emotion words was determined as the average of the similarities of all words in all the (reliable) definitions for the term.

Another feature set was generated using ConceptNet[11]. This is another community generated resource that intends to collect *common sense* relationships among day to day concepts using assertions like *Lizards have a tail* and *Flowers are fragrant* to gain ability to connect concepts and generate meanings as human beings do. Using the Python API for ConceptNet, for each sentence we assigned scores for similarity between the concepts in the sentence and the roots concepts for emotion categories like *happiness* and *sadness*. We also calculated emotion score for each sentence by measuring its similarity with sentences labeled with a particular emotion

Yet another feature set was generated by using Latent Semantic Analysis. We used Python gensim library[12] to perform Latent Semantic Indexing using each sentence as a separate document and measuring similarity between these *documents*. Once the similarity was calculated we used the same method as for ConceptNet, described above, to measure emotion score for each sentence.

During generation of all the features scores mentioned above, a stop list was used to remove the most frequent stop words. The stop list was generated by counting frequency of all words and selecting most frequent words without affective worth.

5 Experiments and Results

We have conducted four sets of classification experiments: *neutral vs, emotional, positive vs. negative, only emotional (no-neutral)* and *all classes (emotional + neutral)*. For each of these experiments we have conducted four sub-experiments, one each for 10 fold cross validation within the two datasets and one each for cross testing with one dataset as the training set and the other as the

[7] http://www.wjh.harvard.edu/~inquirer/
[8] http://wordnet.princeton.edu/
[9] http://wn-similarity.sourceforge.net/
[10] http://www.urbandictionary.com/
[11] http://conceptnet.media.mit.edu/
[12] http://nlp.fi.muni.cz/projekty/gensim/

test set. We have duplicated the experiments for the translated data sets. We have also done some preliminary experiments on mitigation of the class imbalance problem.

Although we used a set of classifiers for our experiments from the weka[13] library including Naïve Bayes, J48, Decision Tree, IBk and SVM, similar to the observation in Holzman and Pottenger (2003), Instance Base classifier (or IBk) performs at least as good as any other classifier in the set. The results, hence, have been reported only for IBk. For all the experiments, we varied the parameter *K* (the number of nearest neighbors) for the classifier over a range in steps of 5. In all cases, for all the datasets, the range [1, 50] has turned out to be enough to capture all variations.

5.1 Neutral vs. Emotional

To assess the ability of our system to differentiate between neutral and emotional sentences we relabeled the sentences with a label other than *NEU*, as *EMO*. The resulting distribution is shown in Table 3.

	Class	#	%	Total
NPS	*EMO*	4140	52.18	7933
	NEU	3793	47.81	
IRC	*EMO*	1536	51.54	2980
	NEU	1444	48.45	

Table 3: Distribution of sentences between the classes (emotional vs. neutral)

The results for classification using IBk in terms of Precision, Recall and F-1 measures are shown in Tables 4 (*cv* means *cross validation* and *ct* means *cross testing*).

	Class	P	R	F	K
NPS (cv)	*EMO*	0.817	0.7181	**0.7646**	11
	NEU	0.7271	0.8257	**0.7737**	11
IRC (cv)	*EMO*	0.858	0.7356	**0.7921**	16
	NEU	0.752	0.8774	**0.8103**	26
NPS (ct)	*EMO*	0.7050	0.5182	**0.5973**	1
	NEU	0.5921	0.8677	**0.7039**	16
IRC (ct)	*EMO*	0.6732	0.6705	**0.6708**	16
	NEU	0.6413	0.6580	**0.6495**	21

Table 4: Accuracy of *emotional vs. neutral* classification with IBk

5.2 Positive Emotion vs. Negative Emotion

For this experiment we removed all sentences with labels *NEU* (neutral) and relabeled *HAP* (happy) as *POS* (positive). We also removed all sentences with the label *SUR* (surprise) as it can belong to either of the target classes, rest of the classes like *anger, sad, disgust etc.* were relabeled as NEG (negative). The resulting distribution of sentences is shown in table 5. The accuracy figures appear in table 6.

	Class	#	%	Total
NPS	*POS*	2644	70.26	3763
	NEG	1119	29.73	
IRC	*POS*	1014	72.22	1404
	NEG	390	27.77	

Table 5: Distribution of sentences between the classes (positive emotion vs. negative emotion)

	Class	P	R	F	K
NPS (cv)	*NEG*	0.6958	0.7033	**0.6995**	16
	POS	0.873	0.8698	**0.8718**	16
IRC (cv)	*NEG*	0.7526	0.7333	**0.7428**	11
	POS	0.8984	0.9072	**0.9028**	11
NPS (ct)	*NEG*	0.4912	0.7179	**0.5833**	11
	POS	0.8627	0.7140	**0.7835**	11
IRC (ct)	*NEG*	0.5772	0.5942	**0.5856**	6
	POS	0.8184	0.8456	**0.8318**	11

Table 6: Accuracy of *positive vs. negative* emotions classification with IBk

5.3 Within Emotions

To study our system's ability to tell one emotion from the other in absence of noise in the form of neutral data, we tested with all sentences marked with the label *NEU* removed.

	Class	#	%	Total
NPS	*ANG*	281	6.78	4140
	FEA	153	3.69	
	SAD	411	9.92	
	SUR	377	9.10	
	HAP	2644	63.86	
	DIS	274	6.61	
IRC	*ANG*	109	7.09	1536
	FEA	51	3.32	
	SAD	119	7.74	
	SUR	132	8.59	
	HAP	1014	66.01	
	DIS	111	7.22	

Table 7: Distribution of sentences amongst various emotions

[13] http://www.cs.waikato.ac.nz/ml/weka/

The resulting class distribution and results are shown in Tables 7 and 8, respectively. For brevity only the best results are shown (IRC, 10 fold cross validation and cross testing with NPS set as training data and IRC set as testing data).

	Class	P	R	F	K
IRC (cv)	*ANG*	0.4313	0.4036	**0.4170**	1
	FEA	0.7222	0.2549	**0.3768**	16
	SAD	0.5324	0.3445	**0.4183**	16
	SUR	0.6534	0.5075	**0.5851**	21
	HAP	0.8284	0.9289	**0.8758**	6
	DIS	0.4782	0.2972	**0.3667**	16
NPS (ct)	*ANG*	0.1929	0.2018	**0.1973**	6
	FEA	0.0406	0.0980	**0.0574**	1
	SAD	0.2061	0.3949	**0.2708**	21
	SUR	0.7142	0.2272	**0.3448**	11
	HAP	0.7566	0.8984	**0.8214**	21
	DIS	0.0977	0.1171	**0.1065**	1

Table 8: Accuracy of classification of various emotions with IBk

5.4 Effect of Translation

Contrary to our expectations, the results for translated data sets have been at best as good as those for the raw datasets. As an example, the effect of translation on performance in terms of F-1 score for 10 fold cross validation on the IRC set for all classes has been shown in Table 9.

Emotion	Raw F1	Translated F1
ANG	0.3461	0.3240
FEA	0.4117	0.2916
SAD	0.4268	0.3176
SUR	0.4300	0.4079
HAP	0.8039	0.7834
NEU	0.8183	0.7967
DIS	0.4662	0.3896

Table 9: Accuracy with and without *translation, i.e.,* data normalization

5.5 Mitigating Class Imbalance

Highly skewed distribution of sample data among target classes has been known to cause degradation in performance for many widely used classifiers (Japkowicz and Stephen, 2002; Guo *et al.*, 2008). Three broad approaches to solving this problem, *undersampling*, *oversampling* and *cost based classification* have been reported. We have conducted experiments to show how *undersampling* can be effective in mitigating the adverse effect of class imbalance.

Emotion	Full NPS Set	UnderSampled NPS Set
ANG	0.2112	0.3544
FEA	0.0992	0.1732
SAD	0.2469	0.3130
SUR	0.3166	0.5380
HAP	0.7703	0.6133
NEU	0.7765	0.5958
DIS	0.1113	0.4024

Table 10: Accuracy improvement with undersampling

Table 10 compares the F-1 scores for all classes for 10 fold cross validation on the NPS set (with a spread of 1.0, i.e. equal numbers from all classes) with and without applying undersampling.

6 Comparison with existing work

We have chosen Holzman and Pottenger (2003; *call it HP03*) and Danisman and Alpkocak (2008; *call it DA08*) for comparing the performance of our system.

While DA08 reports the performance of vector space models and a few other common classifiers on formal text, HP03- closer to our work- reports performance of IBk classifier on internet chat data.

Comparison of relevant results for the three systems is done experiment wise in the following sections. Table 11 shows the distribution among classes for the data in HP03 and DA08.

	HP03		DA08	
Emotion	#	%	#	%
Angry	59	5.0	1072	16.7
Disgust	11	0.9	1066	16.6
Fear	0	0.0	1080	16.8
Happiness	124	10.5	1077	16.8
Sadness	14	1.2	1067	16.6
Surprise	28	2.4	1052	16.4
Neutral	942	79.9	0	0.0

Table 11: Distribution of sentences amongst various emotions in HP03 and DA08

6.1 Neutral vs. Emotional

A comparison of results for this classification reported by HP03 and our system is shown in Table 12 and Table 13 (best results for 10 fold CV and cross testing respectively).

Class	Evaluation criterion	HP03	Our system
Neutral	Precision	0.851	0.752
	Recall	0.977	0.8774
	F-1	**0.9091**	**0.8123**
Emotional	Precision	0.815	0.858
	Recall	0.375	0.7356
	F-1	**0.5137**	**0.7921**

Table 12: Comparing our accuracy with other work (cross validation)

Class	Evaluation Criterion	HP03	Our System
Neutral	Precision	0.854	0.5921
	Recall	0.993	0.8677
	F-1	**0.9183**	**0.7039**
Emotional	Precision	0.750	0.7050
	Recall	0.474	0.5182
	F-1	**0.5809**	**0.5973**

Table 13: Comparing our accuracy with other work (cross testing)

As is evident from the tables above, our results are comparable to those reported in HP03. Our results, however, are more balanced between classes than those reported in HP03.

6.2 Positive vs. Negative

Table 14 compares the performance (in terms of F-1 scores) of our system with the results reported for formal text in DA08. Here again, the results are more balanced for us and as seen in the previous section has class imbalance as an important factor (2.33 for us vs. 4.0 for DA08).

Class	Our system	DA08
POS	0.7428	69.0
NEG	0.9028	93.8

Table 14: Comparing our accuracy with other work (positive and negative emotions)

6.3 Only Emotions

We compare our results for classification into only the emotion classes with similar experiments in [3] and DA08. The comparison is shown in Table 15. Our system performs the best of the three in three out of 5 classes. For the *FEA* we are a close second. Although for the class *ANG* we are far behind HP03, we outperform DA08 which worked on formal text with a balanced dataset in four out of five classes by margins as high as 37%.

Emotion	HP03	DA08	Our System (CV)	Our System (CT)
Anger	**0.745**	0.242	0.4170	0.1772
Disgust	-	0.098	**0.3667**	0.1513
Fear	-	**0.427**	0.3768	0.0432
Joy	0.581	0.496	**0.8758**	0.7795
Sad	0.333	0.322	**0.4183**	0.2504

Table 15: Comparing our accuracy with other work (all and only emotions)

6.4 All Classes

Comparable results for this experiment were available only for HP03. Table 16 shows the results for our system compared against those reported in HP03.

Emotion	HP03 (CV)	HP03 (CT)	Our System (CV)	Our System (CT)
Anger	-	-	**0.3461**	0.0953
Disgust	-	-	**0.4662**	0.0742
Fear	-	-	**0.4117**	0.0440
Joy	0.581	0.496	**0.8039**	0.6133
Sad	0.333	0.322	**0.4268**	0.1311
Surprise	-	-	**0.4300**	0.2182
Neutral	0.9101	**0.9218**	0.8183	0.6698

Table 16: Comparing our accuracy with other work (all emotions and neutral)

Again, we outperform HP03 in six out of seven classes by significant margins. The only class where we perform poorly *NEU*, which as seen in previous section can be attributed to a large extent on the skew in distribution in the data set in Liu et al. (2003) which causes it to show very high results for the most frequent class (*NEU)* in contrast to our more balanced performance.

7 Conclusion and future work

In this paper we have shown how we can improve the performance of emotion classification of instant messages using elaborate annotation and feature engineering. We have also shown how the problem of class imbalance in this domain can adversely affect performance and how measures like undersampling can help. We have also made a preliminary study on the use of SMS translators for normalization of instant message text and subsequent classification of the *formal* text thus obtained.

We have been able to improve upon the currently reported results both in terms of the balance in class-wise performance as well as the absolute performance for different emotion classes.

Following are our future work. While we have started with the (rather intuitive) assumption that internet chats should be more difficult to analyze than formal text, it remains to be shown that it indeed is so. Although we have shown how undersampling can help deal with class imbalance, there are more sophisticated approaches in this area that need to be tested and compared against. As mentioned earlier in view of the poor performance with web based translators, we would also like to test if better translators can better the classification performance.

References

Boucouvalas, A. C., Zhe, Xu. 2002. Real Time Text-To-Emotion Engine for Expressive Internet Communications. *International Symposium on CSNDDSP*.

Liu, H., Lieberman, H. and Selker, T. 2003. A model of textual affect sensing using real-world knowledge. In *Proceedings of the 8th international Conference on intelligent User interfaces*, Miami, Florida, USA.

Holzman L. and Pottenger W. 2003. Classification of emotions in internet chat: An application of machine learning using speech phonemes, *Technical Report LU-CSE-03-002, Lehigh University*.

Strapparava Carlo and Mihalcea Rada. 2008. Learning to identify emotions in text, In *SAC '08: Proceedings of the 2008 ACM Symposium on Applied Computing*.

Wiebe J., Wilson T. and C. Cardie C. 2005. Annotating expressions of opinions and emotions in language. *Language Resources and Evaluation*.

Tagliamonte Sali A. and Derek Denis. 2008. **LINGUISTIC RUIN? LOL! INSTANT MESSAGING AND TEEN LANGUAGE,** *American Speech*.

Schmidt A.P. and Stone T.K.M. Detection of topic change in IRC chat logs.http://www.trevorstone.org/school/ircsegmentation.pdf

Saima Aman and Szpakowicz Stan. 2007. Identifying Expressions of Emotion in Text. In *Proceedings of 10th International Conference on Text, Speech and Dialogue*, *Lecture Notes in Computer Science 4629, Springer,*.

Forsyth Eric N. and Martell Craig H. 2007. Lexical and Discourse Analysis of Online Chat Dialog, *Proceedings of the First IEEE International Conference on Semantic Computing (ICSC 2007)*.

Ekman, P. and Friesen, W. V. 1996. The repertoire of nonverbal behavior: Categories, origins, usage and coding. *Semiotica*.

Japkowicz N. and Stephen S. 2002. The class imbalance problem: A systematic study. *Intell. Data Anal.* 6, 5.

Xinjian Guo, Yilong Yin Cailing Dong, Gongping Yang, Guangtong Zhou. 2008. On the Class Imbalance Problem, in *Proceedings of the 4th International Conference on Natural Computation (ICNC2008),vol.4, pp.192-201*.

Danisman T. and Alpkocak A., Feeler: Emotion classification of text using vector space model, in *AISB 2008 Convention Communication, Interaction and Social Intelligence*, Aberdeen, UK.

Raghunathan Karthik and Krawczyk Stefan, 2009. Investigating SMS Text Normalization using Statistical Machine Transltion, http://nlp.stanford.edu/courses/cs224n/2009/fp/27.pdf