Automatic Detection of Antisocial Behaviour in Texts

Myriam Munezero, Calkin Suero Montero, Tuomo Kakkonen and Erkki Sutinen
School of Computing, University of Eastern Finland
P.O.Box 111, FI-80101, Joensuu, Finland
E-mail: {firstname.lastname}@uef.fi

Maxim Mozgovoy and Vitaly Klyuev
The University of Aizu, Tsuruga, Ikki-machi
Aizu-Wakamatsu, Fukushima, 965-8580 Japan
E-mail: {mozgovoy, vkluev}@u-aizu.ac.jp

A considerable amount of effort has been made to reduce the physical manifestation of antisocial behaviour (ASB) in communities. However, the key to the early detection of ASB is, in many cases, in observing its manifestations in written language, which has not been studied in detail. In this work, we search for linguistic features that pertain to ASB in order to use those features for the automatic identification of ASB in texts. We use an ASB text corpus we have collected as a machine learning resource and approach the detection of ASB in texts as a binary classification problem where discriminating features are taken from the linguistic representation of texts in the form bag-of-words and ontology-based emotion descriptors. Results from preliminary experiments show that by exploiting the emotional information together with Bag-of-Words (BoW) over 90% accuracy in the classification of ASB in texts is reached. Our findings have positive implications in the early detection of potentially harmful behaviour.

Povzetek: Pri analizi asocialnih besedil v omrežjih dosežejo napredek v kvaliteti prepoznavanja z uporabo ontologij čustev.

1 Introduction

Text mining allows for the automatic assessment of linguistic features in texts. Based on the analysis results, it is possible to analyse, for instance the topics that the texts deal with, as well as linguistic styles used in the texts. Language syntax and semantics are tools that are used to express thoughts, opinions, beliefs and emotions through words. The words used can reveal important aspects of someone's social and psychological worlds [33]. Of interest to us, are words and linguistic features that express thoughts or feelings of harming another member of the community. In this paper, we analyse and discover the linguistic features that pertain to ASB based on *machine learning* (ML) and the *antisocial behaviour* (ASB) corpus we introduced in Munezero et al. [27]. Identifying these features will allow us to detect new instances of ASB

ASB is broadly defined as any unconsidered action taken against individuals or groups of individuals that may cause harm or distress to society [5]. Often

individuals involved in ASB have disclosed in advance their feelings and plans through oral or written language [30]. The Internet has been used as the outlet for the expression of such emotional states and / or plans of violent acts through the use of blogs or video sites [9]. Moreover, online communication is often used as a way of shouting out people's intentions before engaging in their acts of violence [21, 2].

The growth of the volume of harmful material on the Web has resulted in increased research for its automatic detection [8]. Being able to automatically detect negative material is beneficial, for instance, to managers of websites that allow users to post content or as part of an early warning system to authorities on possible threats to public safety. The automatic detection of ASB could also give rise to self-awareness systems for the individuals that are expressing thoughts or emotions related to ASB.

This paper investigates the linguistic features used in texts that relate to ASB. By employing ML algorithms we explore the linguistic features that can be used to reliably classify texts containing ASB. For our initial experiments, we explore the impact that BoW and emotions as linguistic features have on the classification of ASB.

This paper is based on: M. Munezero, M. Mozgovoy, T. Kakkonen, V. Klyuev and E. Sutinen, *Antisocial Behavior Corpus for Harmful Language Detection*, published in the Proceedings of the 3rd International Workshop on Advances in Semantic Information Retrieval (part of the FedCSIS' 2013 conference).

2   Related work

Much of the research work on ASB has been performed in the realm of social sciences and psychiatry. There have also been efforts towards detecting and preventing physical manifestations of ASB (such as violence) in communities (e.g. the Home Office in United Kingdom[1]). As such, this problem has not been particularly tackled from the perspective of computational linguistic analysis for early detection and intervention.

As no previous general models for detecting ASB from text exist, we provide an overview of the work done in the context of detecting cyberbullying, terrorism and criminal behaviour which all can be considered as specific forms of ASB.

Perhaps the most notable related work has been carried out in a research project entitled "Intelligent information system supporting observation, searching and detection for security of citizens in urban environment" [41]. The project aimed at automatic detection of terrorist threats and recognition of serious criminal behaviour or violence based on multi-media content. Within the context of INDECT, criminal behaviour is defined as "behaviour related to terrorist acts, serious criminal activities or criminal activities in the Internet".

Our work differs from the one done in the INDECT project in the focus of the research. While INDECT aims at using the analysis of images, video, and text, our focus is on the analysis of text data.

In their cyberbullying study, Dinakar et al., [12] made use of YouTube comments that involved sensitive topics related to race and culture, sexuality and intelligence. Moreover, Yin et al. [45] made use of online forums for detecting online harassment. The cyber-pedophilia research by Bogdanova et al. [3] made use of perverted online journal texts based on which to learn models to discriminated pedophiles from non-pedophiles.

While the corpora used in the studies reported above contain some forms of negative behaviours, their focus is more than ours. We make use of a broader ASB corpus that contains text related to ASB ranging from suicide notes to terrorism and online threats.

2.1   Language expressivity in ASB

Fitzgerald [15] describes the language of ASB as being "deeply value laden, implying purposeful negative action and or behaviour harmful to others". In addition, some researchers have suggested that certain emotions are closely associated with ASB. Some of these emotions include anger, frustration, arrogance, shame, anxiety, depression, sadness, low levels of fear, and lack of guilt [7]. Based on these descriptions, it is reasonable to expect some distinguishing linguistic features in the ASB corpora that may include the use of words that are

---

deemed threatening, harmful or related to violence and emotions that are perceived as overly negative.

2.2   Detecting emotions in texts

Emotions have long been investigated in several studies ranging from social psychology to computational linguistics [19]. Lists of primary or "basic" emotions have been put forward in the psychological field prominently by Frijda [16], Ekman [13] and Plutchik [34] among others. The basic emotion categories used in these lists include: anger, sadness, joy, love, surprise, happiness, fear, and disgust (see [28] [37]), for a detailed compilation of primary emotion lists). Within the Natural Language Processing (NLP) research community, more often than not researchers use Ekman [13] six basic emotion categories: anger, disgust, fear, happiness, surprise and sadness [1] [39].

Performing emotion analysis on various types of text can help us understand and measure the emotions expressed in them. Broadly speaking, two main methods exist for the analysis of emotions within the NLP community: word lists-based and ML-based. Word list based methods use lexical resources such as lists of emotion-bearing words, lexicons or affective dictionaries [29] [14], and databases of commonsense knowledge [20], The *General Inquirer* (GI) [38], the *Affective Norms for English Words* (ANEW) [4], the *WordNet-Affect* [40] [42], and more recently the *NRC word-emotion association lexicon* [25] [24], are all well-known lexical resources.

Whereas ML-based methods cast the problem as a multi-class classification problem, for instance, the automatic emotion classification of news headlines into emotion categories [10]. A significant amount of annotated data is required that represents each of the emotions that are used as the classes. In this work, we use ML to classify texts as containing ASB or not. Our aim is to investigate which features are the best for indentifying instances of ASB in texts.

3   Experimental design

For an exploratory purpose, we conducted four experiments using the ASB corpus. We approached the classification task as a binary classification task, that is, a document is classified as either containing or not containing ASB. We compared the positive ASB texts first with each of the three negative sets of examples (Sect 3.1) and then all the corpora together. We approached it in this manner firstly because the corpora are written in different styles and we wanted to observe whether ASB texts show some distinct characteristics allowing for successful classification from each of the three negative sets, secondly because between the sets there was a balance in terms of the number of documents and average size in characters. We experimented with three supervised ML classifiers for the classification task (Sect 3.2) using three sets of features (Sect 3.3). Furthermore, with each experimental corpus, we used ten-cross validation, that is, the entire corpus was first

---

[1] http://www.homeoffice.gov.uk/crime/anti-social-behaviour/

partitioned into a training set and test set of 90% and 10% respectively, this process was performed ten times. The average results of the 10-cross validation are reported in Section 4.

3.1 Corpora

The following subsections describe each of the four text corpora used in the experimental study. As we are firstly concerned with the binary classification analysis, we compared both positive (those with ASB) (Sect 3.1.1) and negative (Sect 3.1.2 − 3.1.4) examples (non-ASB texts). In order to obtain the negative examples, we used two popular sentiment corpora, movie reviews [31], the emotion annotated corpus (ISEAR) [36] and factual Wikipedia article extracts [44]. Table 1 summarises the documents in the four corpora.

3.1.1 ASB corpus

The ASB corpus is a collection of aggressive, violent, and hostile texts. The texts were collected from various blog posts and news-websites which Munezero et al. [27] could conclusively identify as being ASB. In total 148 documents were identified as ASB. The collection is all English texts, having topics such as: serial killer manifestos, antisocial texts, terrorism, violence-based texts, and suicide notes.

Important to us, the messages in these documents are reflective of the author's thoughts and emotions. The corpus was collected specifically for the purpose of detecting ASB, conflict, crime and violence behaviour from text documents. The collection is based on the research on ASB that has shown that aggression, violence, hostility, and lack of empathy are among the traits that are most directly associated with ASB [6] [32].

3.1.2 International Survey on Emotion Antecedents and Reactions

The ISEAR corpus is a collection of student reports on situations in which the respondents felt any of the seven major emotions: joy, fear, anger, sadness, disgust, shame, and guilt. The responses include descriptions of how they appraised the situation and how they reacted [36].

3.1.3 Movie reviews

This collection consists of 2000 movie reviews. They are labelled in respect to their polarity: negative and positive. The corpus was first used in [31], and now is often applied in sentiment analysis and opinion mining research as a standard development and test set.

3.1.4 Wikipedia text extracts

We searched and collected Wikipedia articles by using similar concepts such as those we found to be characteristic ASB: killing, terror, violence, aggression, and frustration. The aim of including these texts was to observe how well our classification algorithms could distinguish between ASB texts and informative texts containing similar keywords.

3.2 Classifiers

For classifying the documents into the two classes, we experimented with three supervised ML classifiers: Multinomial Naïve Bayes, SMO for the implementation of Support Vector Machines, and J48 for Decision Trees. The three selected algorithms have shown to be effective in various text classification studies. We made use of the WEKA tool [17] to implement the classifiers used in our study.

Multinomial Näive Bayes (MNB). The NB classifier is a probabilistic model that assumes independence of the attributes used in the classification. The classifier has shown good performance even when the sample size is small [11]. We used the MNB classifier implemented in WEKA, which uses a multinomial distribution for each of the features.

Support Vector Machine (SVM) is based on the maximum margin hyperplane rather than probabilities as the Naïve Bayes [23]. The SVM classifier aims to find a hyperplane, represented by a vector that maximally separates the document vectors in one class from those in the other [31].

J48 Decision Tree (J48) is an implementation of the C4.5 decision tree in WEKA. Decision trees are predictive models that are used for classification tasks by starting at the root of tree and moving through it until a leaf is encountered [35]. The decision tree is built from the input training data using the property of information gain or entropy to build and divide nodes of the decision tree in a manner that best represents the training data and the feature vector [12].

3.3 Classification features

3.3.1 Bag-of-Words

As a first experiment with the ASB corpus, we used the Vector Space Model approach so as to consider the words as independent entities. The model makes an implicit assumption that the order of words in document does not matter, which is also referred to as the Bag-of-Words (BoW) assumption. The approach is sufficient for many classification tasks, as the collection of words appearing in the document (in any order) is usually sufficient to differentiate between semantic concepts [23]. Each document in the corpora was represented as a feature vector composed of binary attributes for each word that occurs in the file.

Let $\{f_1, …, f_m\}$ be a predefined set of m features that can appear in a document. Let $n_i(d)$ be the number of times f_i occurs in a document d. Then each document d is represented by the document vector $d := (n_1(d), n_2(d), …, n_m(d))$ [31]. If a word appears in a given document, its corresponding attribute is set to 1; otherwise it is set to 0. Generally, the BoW approach works well for text classification. However, it does not take into consideration any semantic and contextual information.

Moreover, in order to reduce the number of words in the BOW representation we used the LovinsStemmer [22] in order to replace each word by its stem.

Table 1: Corpora description with source, number of documents and average document size.

Corpus	Source	No. of Documents	Avg. Document Size (characters)
ASB	[27]	148	680
ISEAR	[36]	265	110
Movie reviews	[31]	178	390
Wikipedia extracts	[44]	212	680
Total		803	1860

3.3.2   Emotions

Emotions reveal connections of individuals to values in the social world and hence, are the triggers of many social psychological phenomena, such as altruism, antisocial behavior and aggression [32]. In our experiments, we analyse in particular, emotions that might be present in the ASB corpus and analyse whether they are reliable classification features.

To identify the emotions presented in the corpora, we made use of an emotion ontology introduced in [26]. It is an ontology of emotion categories whereby each category contains a set of emotion classes and emotion words. Figure 1, demonstrates the negative emotion with samples from the disliking class.

The emotion ontology is based on WordNetAffect and it contains 85 classes and 1,499 words. On average, the ontology contains 17.6 words per emotion class which gives a relatively wide coverage of emotion classes and emotion words. This together with the fact that the ontology was not fitted on to any particular

dataset or text corpus makes it suitable to be used in our experiments as a basis for ASB classification.

For the classification, we made use of two types of emotion-based features: ontology-dependent and ontology-independent emotion features. The ontology-dependent features are collected through a tagging process using the emotion ontology. Through the tagging process, we collected tags such as the sum of all the relative frequencies of the emotion classes that belong to the emotion categories represented in the ontology. While the ontology-independent emotion features were obtained by using the SentiStrength system [42] to calculate the emotion strength of a text.
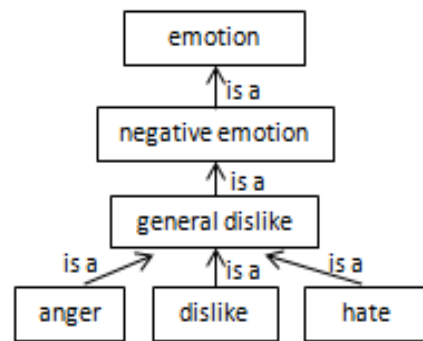

Figure1: Sample from the negative-emotion category section of the ontology.

4   Results

For an exploratory purpose, we conducted four experiments using the ASB corpus and three corpora as negative examples of ASB (Subsection 3.1.2 - 3.2.4). We explored the impact that BoW and emotions as classification features have on the detection of ASB texts. In the first experiment, binary classifiers using the three classifiers described in Subsection 3.2 were trained on ASB+ISEAR, in the second on ASB+Movie reviews, and in the third on ASB+Wikipedia extracts. Finally, all the corpora were combined into a single data set.
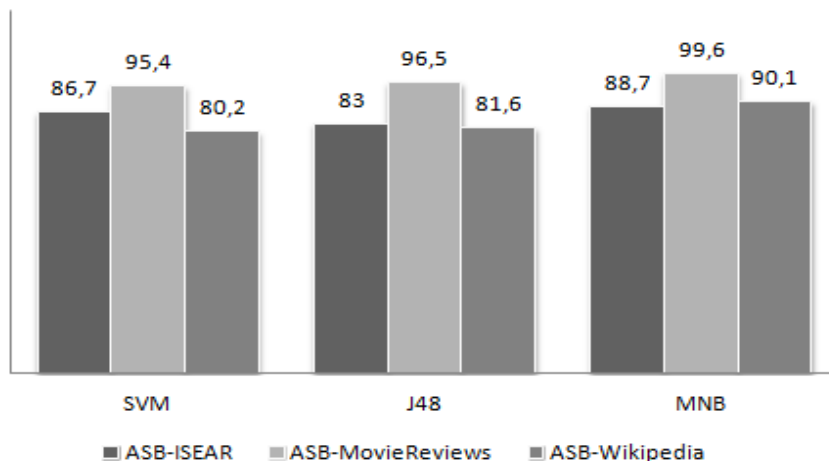

Figure 2: Accuracy results of SVM, J48 and MNB classifier with emotions+BoW as classification features (%).

Table 2: Results from SVM classifier (%).

Corpora	Features	Accuracy	Precision	Recall	F-measure
ASB + ISEAR	Emotions	84.9	85.8	84.9	84.9
	BoW	86.2	88.7	86.2	86.0
	Emotions+BoW	**86.7**	**89.1**	**86.8**	**86.7**
ASB + MovieReview	Emotions	81.1	82.5	81.2	79.1
	BoW	**95.8**	**95.8**	**95.8**	**95.7**
	Emotions+BoW	95.4	95.5	95.4	95.3
ASB + Wikipedia	Emotions	69.7	70.3	69.7	69.0
	BoW	79.6	80.2	79.6	79.6
	Emotions+BoW	**80.2**	**80.3**	**80.3**	**80.3**

Table 3: Results from J48 classifier (%).

Corpora	Features	Accuracy	Precision	Recall	F-measure
ASB + ISEAR	Emotions	**84.9**	**84.9**	**84.9**	**84.9**
	BoW	79.2	79.5	79.2	79.2
	Emotions+BoW	83.0	83.2	83.0	83.0
ASB + MovieReview	Emotions	84.6	84.6	84.6	84.6
	BoW	**98.1**	**98.1**	**98.1**	**98.1**
	Emotions+BoW	96.5	96.6	96.5	96.5
ASB + Wikipedia	Emotions	62.5	62.3	62.5	62.3
	BoW	**82.9**	**83.6**	**82.9**	**82.9**
	Emotions+BoW	81.6	82.1	81.6	81.6

Table 4: Results from MNB classifier (%).

Corpora	Features	Accuracy	Precision	Recall	F-measure
ASB + ISEAR	Emotions	71.7	72.0	71.7	71.5
	BoW	**88.7**	**90.0**	**88.7**	**88.6**
	Emotions+BoW	**88.7**	**89.6**	**88.7**	**88.6**
ASB + MovieReview	Emotions	76.9	77.0	76.9	74.2
	BoW	98.4	98.4	98.4	98.4
	Emotions+BoW	**99.6**	**99.6**	**99.6**	**99.6**
ASB + Wikipedia	Emotions	58.5	59.4	58.6	58.5
	BoW	89.5	90.4	89.5	89.3
	Emotions+BoW	**90.1**	**90.8**	**90.1**	**90.1**

The performances of the classifiers were then compared in terms of accuracy, precision, recall and F-measure. We made use of ten-fold cross validation whereby samples of data are randomly drawn for analysis and the classification algorithm then computes predicted values [23]. Table 2, 3 and 4 show the average of the ten-fold cross validation results on the corpora for each of the ML classifiers with a) BoW, b) emotions, and c) emotions + BoW, as features.

Using SVM classifier, the emotion + BoW features performed better in two of the experiments (ASB+ISEAR and ASB+wikipedia). With J48, the emotions were the better discriminator for the ASB+ISEAR. However the BoW model performed better for the other sets. In looking at the MNB classifier, the emotions + BoW feature set performed better in all the three sets. Hence, in majority of the cases, the addition of emotions to BoW provided better accuracy results. We note that our experiments are preliminary, especially as there is no standard ASB corpus available and the number of documents in the ASB corpora is relatively small.

Figure 2 summarizes the accuracy results of the three classifiers using both the emotions + BoW as features.

From Figure 2, we see that accuracy results in all three classifiers are over 85% which indicates a relatively high accuracy. The best accuracy (99,6%) was reached on the ASB + Movie reviews set with the MNB classifier.

We further noted that when all the four corpora (ASB + ISEAR + MovieReviews + Wikipedia) were combined, the classifiers were not learning. This was due to the imbalance in the class distribution, i.e. the majority of the texts were from the negative (not ASB) class, which then causes ML algorithms to perform poorly on the minority class [18]. However, with the MNB classifier, we observed that it was able to learn in spite of the imbalance.

A closer look at the most predictive features revealed emotion classes such as 'general-dislike', 'hate', 'anxiety', and 'sadness' as expected based on the known connection between ASB and negative emotions. Surprisingly, however the emotion class 'affection' also appeared as a contributing attribute.

5   Conclusion and future work

In this paper, we applied text classification techniques for the analysis and detection of ASB. We reported on experiments where the linguistic features, BoW and emotions were used for the classification of ASB. Our experimental results illustrated that linguistic features such as BoW and emotions can be used successfully to classify ASB in text. We found that the performance of MNB was consistently better than that of J48 and SVM when using the emotions + BoW features. In comparison, when using emotion features alone, the J48 and SVM had the highest accuracy on the ASB+ISEAR (84,9%) and with the BoW features alone, J48 had the highest accuracy with the ASB+MovieReview (98,1%). Thus both features are essentially for achieving high classification accuracy.

Deeper analysis of the features further revealed subsets of emotion features that most contributed to the classification accuracy.

ASB is a growing concern to the society, and in some instances to the government and law enforcement agencies around the world. In line with creating a safer community, identifying the individuals who pose a danger to a community involves analysing the information they put forward. Thus future work involves exploiting the identified linguistic features to build a model to classify new instances of ASB in text as part of an early detection system. Using the features, we would also like to explore the categorizations between different types of ASB, for example physical manifestations of ASB such as violence to other individuals, and non-physical acts such as cyberbullying.

Additionally, with the identified features, we would like to extend the corpus. A larger corpus would allow us to have a larger training set for ML algorithms allowing for learning of new features for building a classification model. In this case, fewer than 200 records were used that could be confidently identified as ASB, and due to this amount, we observed that the SVM and J48 classification models were not learning due to the imbalance in the data.

These experiments were our first attempt at automatically detecting ASB in texts. The results we demonstrated are promising, but experiments on large-scale date are necessary to confirm the robustness of our approach.

Moreover, in this paper, we investigate BoW and emotions as features, but in future we plan to include semantic analysis which could additionally reveal features for ASB identification.

Regardless, in our work, we have found that NLP techniques have potential for the early detection of ASB while the harmful behaviour might still be at its planning stage. Our results have direct applications for national and local security.

Acknowledgement

References

[1] Alm, C. O., & Sproat, R. (2005). Emotional sequencing and development in fairy tales. *Springer*, (pp. 668–674).

[2] Böckler, N., Seeger, T., Sitzer, P., & Heitmeyer, W. (2013). *School Shootings: International Research, Case Studies, and Concepts for Prevention*. New York, USA: Springer.

[3] Bogdanova, D., Rosso, P., & Solorio, T. (2012). Modelling fixated discourse in chats with cyberpedophiles. *Proceedings of the Workshop on Computational Approaches to Deception Detection* (pp. 86–90). Association for Computational Linguistics.

[4] Bradley, M. M., & Lang, P. J. (1999). *Affective norms for English words (ANEW). Instruction Manual and Affective Ratings*. Technical report, University of Florida, The Center for Research in Psychophysiology.

[5] Card, R., & Ward, R. (1998). *The Crime and Disorder Act*. Retrieved 3 28, 2013, from legislation.gov.uk:
http://www.legislation.gov.uk/ukpga/1998/37/contents

[6] Clarke, D. (Abingdon, UK). *Pro-Social and Anti-Social Behaviour*. 2003: Taylor & Francis.

[7] Cohen, L. J. (2005). Neurobiology of Antisociality. In C. Stough, *Neurobiology of Exceptionality* (pp. 107-124). New York, USA: Kluver Academic/Plenum Publishers.

[8] Correa, D., & Sureka, A. (2013). *Solutions to Detect and Analyze Online Radicalization: A Survey*. Delhi, India: Indraprastha Institute of Information Technology.

[9] Crowley, S. (2007, 11 7). *Finland Shocked at Fatal Shooting*. Retrieved 3 28, 2013, from BBC News: http://news.bbc.co.uk/1/hi/world/europe/7084045.stm

[10] Danisman, T., & Alpkocak, A. (2008). Feeler: Emotion Classification of Text Using Vector Space Model. *AISB 2008 Convention, Communication, Interaction and Social Intelligence. 2*, pp. 53-59. Aberdeen, UK. Affective Language in Human and Machine.

[11] De Ferrari, L., & Struart, A. (2006). Mining housekeeping genes with a Naive Bayes classifier. *BMC Genomics, 7*(277).

[12] Dinakar, K., Reichart, R., & Lieberman, H. (2011). Modeling the detection of textual cyberbullying. *International Conference on Weblog and Social Media-Social Mobile Web Workshop.*

[13] Ekman, P. (1993). Facial Expression and Emotion. *American Psychologist, 8*(4), 376-379.

[14] Elliot, C. (1992). *The affective reasoner: A process model of emotions in a multi-agent system.* Ph.D. thesis,, Northwestern University, Institute for the Learning Sciences.

[15] Fitzgerald, M. (2011). *Submission to the Department of Human Services on behalf of Public Housing Tenants in relation to Human Rights concerns raised by the Anti-Social Behavior Pilot.* Fitzroy Legal Service.

[16] Frijda, N. H. (1986). Emotional Behavior. In *The Emotions. Studies in Emotion and Social Interaction.* Cambridge University Press.

[17] Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. H. (2009). The WEKA Data Mining Software: An Update;. *SIGKDD Explorations, 11*(1).

[18] Hulse, J. V., Khoshgoftaar, T. M., & Napolitano, A. (2007). Experimental Perspectives on Learning from Imbalanced Data. *Proceedings of the 24th International Conference on Machine Learning.* Corvallis, OR.

[19] Liu, B. (2010). Sentiment analysis and subjectivity. In N. Indurkhya, & F. J. Damerau (Eds.), *Handbook of Natural Language Processing, (2nd ed.).* Boca Raton, Florida, USA: CRC Press, Taylor and Francis Group.

[20] Liu, H., Lieberman, H., & Selker, T. (2003). A Model of Textual Affect Sensing using Real-World Knowledge. *Proceedings of the 2003 IUI*, (pp. 125-132).

[21] Logan, M. (2012, July). *Case Study: No More Bagpipes*. Retrieved March 15, 2013, from The Threat of the Psychopath: http://www.fbi.gov/stats-services/publications/law-enforcement-bulletin/july-2012/case-study

[22] Lovins, J. B. (1968). Development of a stemming algorithm. *Mechanical Translation and Computational Linguistics., 11*, 22-31.

[23] Miner, G., Elder, J., Hill, T., Nisbet, R., Delen, D., & Fast, A. (2012). *Practical text mining and statistical analysis for non-structured text data applications (1st ed).* Waltham, MA: Academic Press.

[24] Mohammad, S. M. (2012). Portable Features for Classifying Emotional Text. *Proceedings of the 2012 NAACL HLT*, (pp. 587–591).

[25] Mohammad, S. M., & Turney, P. D. (2010). Emotions Evoked by Common Words and Phrases: Using Mechanical Turk to Create an Emotion Lexicon. *Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text*, (pp. 26-34).

[26] Montero, C. S., Kakkonen, T., & Munezero, M. (2014). Investigating the Role of Emotion-based Features in Author Gender Classification of Informal Text. A. Gelbukh (Ed.): *CICLing 2014, Part II, Lecture Notes in Computer Science 8404*, (pp. 98–114), Springer-Verlag Berlin Heidelberg 2014.

[27] Munezero, M., Mozgovoy, M., Kakkonen, T., Klyuev, V., & Sutinen, E. (2013). Antisocial behavior corpus for harmful language detection. *Federate Conference in Computer Science.* Krakow, Poland.

[28] Ortony, A., Clore, G. L., & Collins, A. (1994). The Structure of the Theory. In *Chapter 2 in The Cognitive Structure of Emotions* (pp. 15-33). Cambridge University Press.

[29] Ortony, A., Clore, G. L., & Foss, M. A. (1987). The Referential Structure of the Affective Lexicon. *Cognitive Science, 11*, 341-364.

[30] O'Toole, M. E. (2000). *School Shooter: A Threat Assessment Perspective. National Center for the Analysis of Violent Crime.* Quantico, Virginia, USA.: Federal Bureau of Investigation.

[31] Pang, B., Lee, L., & Vaithyanatha, S. (2002). Thumbs up?: sentiment classification using machine learning techniques. *Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10:* (pp. 79-86). Association for Computational Linguistics.

[32] Parrot, G. W. (2001). *Emotions in Social Psychology.* Philadelphia, Pennsylvania, USA: Taylor & Francis.

[33] Pennebaker, J. W., Mehl, M. R., & Niederhoffer, K. G. (2003). Psychological Aspects of Natural Language Use: Our Words, Our Selves. *Annual Review of Psychology, 54*, 547-577.

[34] Plutchik, R. (2001). The Nature of Emotions. *American Scientist, 89*(4), 344-350.

[35] Quinlan, J. R. (1993). *C4.5: Programs for machine learning.* San Mateo, Calif: Morgan Kaufmann Publishers.

[36] Scherer, K. R., & Wallbott, H. G. (1994). Evidence for universality and cultural variation of differential emotion response patternin. *Journal of personality and social psychology, 66*, 310.

[37] Shaver, P., Schwartz, J., Kirson, D., & O'Connor, C. (1987). Emotion knowledge: Further exploration of a prototype approach. *Journal of Personality and Social Psychology, 52*, 1061-1086.

[38] Stone, P. J., Dunphy, D. C., Smith, M. S., & Ogilvie, D. M. (1966). *The General Inquirer: A Computer Approach to Content Analysis.* Cambridge, Massachusetts, USA: The MIT Press.

[39] Strapparava, C., & Mihalcea, R. (2008). Learning to Identify Emotions in Text. *Proceedings of the ACM SAC'08*, (pp. 1556-1560).

[40] Strapparava, C., & Valitutti, A. (2004). WordNet-Affect: an Affective Extension of WordNet. *Proceedings of the 4th LRE*, (pp. 1083-1086).

[41] The INDECT Consortium. (n.d.). *XML Data Corpus: Report on Methodology for Collection, Cleaning and Unified Representation of Large Textual Data from Various Sources: News Reports Weblogs Chat.* Retrieved 10 10, 2010, from http://www.indect-pro-ject.eu/files/deliverables/public/INDECT_Deliverable_4.1_v20090630a.pdf (2010, Dec. 10).

[42] Thelwall, M., Bucley, K., Paltoglou, G., & Cai, D. (2010). Sentiment Strength Detection in Short Informal Text. *Journal Of The American Society for Information Science And Technology., 61*(12), 2544–2558.

[43] Valitutti, A., Strapparava, C., & Stock, O. (2004). Developing Affective Lexical Resources. *PsychNology, 2*(1), 61-83.

[44] Wikimedia Foundation. (2013, May 08). Wikipedia: The Free Enceclopedia.

[45] Yin, D., Xue, Z., Hong, L., Davison, B. D., Kontostathis, A., & Edwards, L. (2009). Detection of harassment on web 2.0. *Proceedings of the Content Analysis in the WEB*, *2*.