Cyberbullying System Detection and Analysis

Yee Jang Foong
University of Birmingham, EECS,
Edgbaston, Birmingham B15 2TT, UK

Mourad Oussalah
Centre for Ubiquitous Computing,
University of Oulu, 90014 Oulu
Finland. Mourad.Oussalah@ee.oulu.fi

***Abstract*—Cyber-bullying has recently been reported as one that causes tremendous damage to society and economy. Advances in technology related to web-document annotation and the multiplicity of the online communities renders the detection and monitoring of such cases rather difficult and very challenging. This paper describes an online system for automatic detection and monitoring of Cyber-bullying cases from online forums and online communities. The system relies on the detection of three basic natural language components corresponding to Insults, Swears and Second Person. A classification system and ontology like reasoning have been employed to detect the occurrence of such entities in the forum / web documents, which would trigger a message to security in order to take appropriate action. The system has been tested on two distinct forums and achieves reasonable detection performances.**

***Index terms*—Corruption; string matching; Panama Papers**

I. INTRODUCTION

The phenomenon of cyberbullying, referred to as "willful and repeated harm inflicted through the use of computers, cell phones, and other electronic devices" [1-2] has drastically increased in recent years, especially in youth population, mainly due to advances in computerized technology, which can cause tremendous social and financial losses to click-and-mortar organizations in recent years. For instance, Hinduja and Patchin [3] reported that 10-40% of surveyed youth population admitted to have dealt with it either as a victim or as a by-stander where adolescents use technology to harass, threaten, humiliate, or otherwise hassle their peers. Teens have also created web pages, videos, and profiles on social media platforms making fun of others, using the distinguished abilities of camera-enabled devices, which violates universal privacy standards. ScanSafe's monthly "Global Threat Report" found that up to 80% of blogs contained offensive contents and 74% included porn in the format of image, video, or offensive languages. Besides, the open online chat systems and forums has significantly increased the spread of cyber bullying cases. This has negatively impacted organization and damaged economy as a whole. It also put extra pressure on security officers. The latter face increasing challenges for various reasons. First, cyber bullying can happen 24 hours a day, 7 days a week, and reach a kid even when he or she is alone. It can happen any time of the day or night. Second, cyber bullying messages and images can be posted anonymously and distributed quickly to a very wide audience. It can be difficult and sometimes impossible to trace the source. Third, deleting inappropriate or harassing messages, texts, and pictures is extremely difficult after they have been posted or sent.

Cyberbullies can have a mask by being anonymous on the chat rooms. Many forums and chat rooms don't require a real name to be registered as a user. This makes cyber bullies even more violent and brave. Anonymity and the lack of meaningful supervision in the electronic medium are the two factors that have aggravated this social menace. Besides, different from physical bullying, cyberbullying is "behind the scenes" as the messages, if posted on public forum, can stand for ages, creating a continuous frustration and harm to the victim, and potentially to many other users.

Negative consequences of cyberbullying victim are devastating. It can have a huge affect on the growing up process of a child. The child will lose confidence, feel depressed, become anti-social and many more negative consequences that will harm a child mentally and these often affect the victims until adulthood. Some serious cases might lead to a child committing suicide but more than often resulting in tragic outcomes [4]. Boyd [5] identified four aspects of the Web that can significantly magnify the impact and damage of bullying: persistence, searchability, replicability and invisible audiences.

Several attempts to deal with cyberbullying and offensive content have been reported, including many commercial products. For instance, few social networks have an "Online Safety Page" that leads to resources such as the anti-bullying sites of the government or other organizations, where the bullying issue is handled primary as a response to explicit complaints. However, the method soon becomes obsolete as the rate of daily received complaints overwhelms the ability of small groups of complaint handlers to deal with them. Other commercial solutions imitate and accommodate the spam-detection filter technology. In this respect, Appen and Internet Security Suites [6-10] have been endowed with moderate ability to detect and filter out online offensive contents by simply blocking webpages and paragraphs that contained dirty words. Nevertheless, such word-based

approaches fails to identify subtle offensive messages and affect web-site readability. For instance, the sentence "you are such a half-intelligent person" will not be identified as offensive content, because none of its words is included in general offensive lexicons. Besides, the approach often yields high false positive rate due to the word ambiguity problem where the word can have multiple meanings. Moreover existing methods treat each message as an independent instance without tracing the source of offensive contents.

Stories sharing approach as in MTV's a thin line (http://www.athinline.org) is among educational solution where the individuals can learn from experts and peers experience /advices. Mishna et al. [11], among other social science researchers, explored the short and long term consequences of cyberbullying on school education, parenting and social workers.

Dinakar et al. [12] in their survey paper noticed that the current detection efforts of cyberbullying problems are largely absent or extremely naïve, while intervention efforts were largely offline and fail to provide specific actionable assessment and advice. Therefore, it becomes crucial to seek for advanced automatic cyberbullying systems. Nevertheless the potential effectiveness of such approach is still to be validated. Especially, will advanced linguistic analysis improve the accuracy and reduce false positives in detecting message-level offensiveness? Is the conceptual textual analysis efficient or secondary / third party information will be required to achieve sufficient classification result? This motivates the work highlighted in this paper where a new prototype for automatic cyberbullying detection using a combination of natural language processing technique and ad-hoc based approach. The feasibility and performance of the proposal have been assessed using some manually labelled corpus. Especially, WordNet lexical database is employed in order to identify semantically related words and evaluate the similarity with selected cyberbullying terms. On the other hand, a classification based approach is put forward in order to identify genuine cyberbullying cases.

II. Solution outline and methodology

A. Background

In the same spirit as natural language processing challenges tasks, e.g., misbehavior detection task of CAW 2.0 [13], the cyberbullying detection task is primarily focused on the content of the conversations (of the text written by the participants, both the victim and the bully), regardless the known features and characteristics of those involved.

Building on some social science and psychiatry studies (see, e.g., Mishnaa et al. [11], Hinduja and Patchin [3]), one hypothesizes that any cyberbullying case involves both Insult/Swear wording and Second person or Person name. We hypothesize when the association Insult/Swear wording and Person Name / Second person is validated then, the occurrence of cyberbullying case is enabled. Nevertheless such reasoning is not straight from natural language processing as it can see from the examples below.

"You are an idiot" is a typical example of cyberbullying as it contains both Insult/Swear word "Idiot" and Second person "You" as well as a clear association between the word and Second person.

"This computer is stupid" contains only an Insult/Swear and naturally it does not promote the sentence to a cyberbullying case.

"This computer is stupid despite you are hard-working person" contains both Insult/Swear word and Second person but it is not a cyberbullying case as the association between the two is not established.

"I know you are not stupid" contains both Second person, Insult/Swear word and there is an established connection between the two, but it is not a cyberbullying case.

In other words, the presence of the aforementioned conditions for cyberbullying case is only a necessary condition but it does not systematically entail cyberbullying because of the variety of natural language modifiers to express negation and opposition.

The paragraph "I found you nice today. Idiot" is a cyberbullying case despite the second sentence "Idiot" contains only an Insult/Swear wording and no Second person or Person entity, but since it refers to previous sentence, the link can easily established from an operator perspective.

The above few examples demonstrate the complexity of the task of identification of cyberbullying case using standard natural language processing tools, which requires investigating all the textual information of the phrase.

This motivates the ideas put forward in this paper where a combination of features will be employed in order to tackle the various forms of cyberbullying cases, which includes explicit evaluation of the association Swear/Insult word and Second name/Person entity. More formally, the design of cyberbullying system detection involves several steps: i) Crawling the content of website / online forum; ii) Parsing the textual content; iii) Extraction of key terms; iv) Use of linguistic and WordNet lexical databases to extract related terms to categories Insult and Swear; v) Use of Hash map for detecting the second person words; vi) Use a machine learning classifier in order to strengthen the detection ability; viii) Design of a graphical user interface (GUI) in order to interact with user.

B. Dataset

We used the social media platform ASKfm [14] where users ask questions publicly on other users' pages. It also provides possibility to ask questions anonymously as well as possibility to view samples of users' profiles. Public proxy servers located in UK and USA were employed in order to retrieve English written

posts. The Scrapy crawler[1] has been used for this purpose. We deliberately select questions / answers that contain at least one word that belong to Insult or Swear category using a simple string matching function. This was motivated by the desire to gather dataset that will likely contain cyberbullying cases. We separately store the usernames of the questioners. An SQL server database was employed in order to store and index all the dataset attributes (usernames, questions, posts, date, links), which ultimately boost the indexing and retrieving functions. A total of around 10,000 questions and answers were gathered from the site. Around 17% of the total collected dataset is found to entail genuine cyberbullying cases, after an initial brief scrutiny of the dataset. The entire data set was then split into a training set and a test set. 70 percent of both the negative and positive examples were used as a training set while the remaining 30 percent were used as the test set. Fig 1 shows example of comments found in each category (through manual labelling)

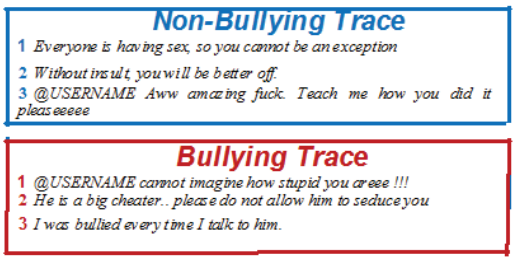

Fig 1. Example of Bullying-Nonbullying traces from corpus

C. Preprocessing

Prior to subsequent analysis, automatic pre-processing procedure assembles the comments for each user and chunks them into sentences. Next, web links and unknown characters were removed. For each sentence, the incorrect wording are corrected in the following way. The word is first mapped to WordNet lexical database. If an entry is not found, we seek whether it has an entry in the list of saved usernames, Named-entities (using Illinois Named-entity tagger), SMS dictionary / abbreviations (using SMS dictionary Netlingo (www.netlingo.com/acronyms.php). If no entry is found at any of the linguistic dictionaries, we check for the presence of character duplication that will be removed. If neither the original nor the transformed word is recognized, the word is inputted to Norvig spell-correcting algorithm (http://norvig.com/spell-correct.html), the unknown word is therefore substituted by the suggested correct wording only if its Edit distance with respect to the original is one. Although, we agree that such restriction would discard potential genuine corrections, we also want to diminish the impact of false negative by avoiding deleting deliberate user's incorrect wording. Furthermore, the lexicons found in the text such as smiley faces, brushing faces, among others, are replaced by their textual equivalent expressions. This will ensure that such symbols are also taken into account in the feature space that will detailed later on.

D. Features

Central to the methodology is the choice of the (textual) features that will be employed for the classifier. In this context, consistent with the "gestalt" principle (the whole is greater than the sum of its parts) [15], we hypothesize that a combination of modestly accurate features coming from heterogeneous data modalities can outperform methods that employ a single modality. More specifically, our approach utilizes the following features:

- *Tf-Idf* (term-frequency times inverse document frequency). Although, this is very standard and commonly employed feature in information retrieval and text mining literature [16], our implementation introduces two key novelties. First, WordNet lexical database [17] as well as some SMS repositories were used in order to convey a rich vocabulary of Insult and Swear related words. More specifically, the external linguistic resource (www.noswearubg.com/ dictionary) contains a detailed vocabulary of insulting/swearing words. Besides, each entry of the above wording is mapped to WordNet lexical database, so that if a link is found, then the direct hyponym and direct hypernym are also added to the list. This constitutes an extended set of (Insult / Swear) Vocabulary, referred to as V_0. Second, the weights in *tf-idf* matrix of words found in V_0 will be boosted by a constant factor (while the overall weight is upper bounded by one). This approach is also in light with Nahar et al. [18] who magnified the weights corresponding to bullying words by a factor of two. The reason for this is that bullying comments often contain bad words and scaling these features can make it easier to find a good separation in vector space for the classifier. Finally, we selected the top 100 words that yield the highest *tf-idf* score as *tf-idf* features.
- Linguistic Inquiry and Word Count (*LIWC*[2]) features. Especially, *LIWC* provides more than 90 descriptive variables which includes word counts, summary variables, various word categories, personal concern categories, details and frequency of punctuations and the informal languages used. This includes, Psychological processes, Use of sexual words, Personnel concerns, Third person singular pronouns. We restricted the LIWC features to only categories that can convey bullying attitude. This concerns the categories: Second person, Total number of pronouns, Swear words, Negative emotion, Anxiety, Anger, Sadness and Sexual. This yields a total of 8 features.

[1] https://scrapy.org/

[2] LIWC 2015 was used. http://liwc.wpengine.com/

- Unusual capitalization. Interestingly, in view of social observation, words, excluding Named-entities and sentence starting letters, with capitalization may convey strong relationship to cyberbullying. Therefore, one counts the total number of occurrences of such capitalization in the underlying document (blog) as one additional feature which concatenates *tf-idf* and *LIWC* features.
- Dependency features. Especially, we experiment with Stanford Dependency Parser [19] whenever the occurrence of Insult/Swear word is found. We report all dependencies where the Insult/Swear word is related to a pronoun or to a Person named-entity or username. Dependency types (extracted by Stanford dependency parser) and relevant to cyberbullying detection are summarized in Table 1. This allows us to quantify effectively the association of Insult/Swear word to Second Name / Person entity. Therefore for 16 dependency types of Table 1, and given a word *w* of V_0 and a pronoun *pr* or person-entity / username *pu*, we have for each dependency type 4 features, constituted of rel(w,pr), rel(pr, w), rel(w,pu), rel(pu, w). This makes the total number of (relational) features 64. Therefore, the total number of features becomes 100 (tf-idf) + 8 (LIWC) + 1 (Capitalization) + 64 (Dependency) = 173

Table 1. List of relevant Parser dependency types

Dependency Type	Meaning
abbrev	abbreviation modifier
acomp	adjectival complement
amod	adjectival modifier
appos	noun compound modifier
nn	noun compound modifier
partmod	participial modifier
iobj	direct object
iobj	indirect object
nsubj	nominal subject
nsubjpass	passive nominal subject
xsubj	controlling subject
agent	passive verb's subject
conj	conjunct
parataxis	parataxis
poss	holds between the user and its possessive determiner
rcmod	relative clause modifier

E. Classifier

Due to its proven efficiency in binary classification and theoretical soundness, we used support vector machines (SVM) classifier. Basically, SVM algorithm attempts to find the line separating the two classes so that the margin between the closest positive example and the closest negative example is as large as possible. Basically, it uses training data to learn a classifying function with which it can classify a new data that has not been previously seen in one of the (two) categories, in case of binary classification.

We adopted Joachims' implementation SVMlight [20]. The SVM was trained with a linear kernel on the training data. Besides, since the training data is imbalanced, containing mostly negative examples we have to adjust the cost-factor parameter J, a factor representing how much the cost of an error on a positive example should outweigh an error on a negative example. Tuning the parameters was done with a parameter sweep where C assumed the values [0.001, 0.01, 0.1, 1, 10, 100] and J assumed the values [10, 50, 100]. The parameter setup achieving the highest F2-measure was recorded and used for evaluation. Before feeding the features to the SVM all features were normalized with the length of the feature vector (L2 norm). A 10-fold cross validation was conducted in this experiment. After training the models, they were applied to our test dataset. When comments have been converted to document vectors they are written to a file which is then fed to the SVM^{light} implementation along with the trained model file, and the result (whether the test post is cyberbullying or not) is returned to the client.

F. Data labelling

Central to the supervised classification task of SVM is the training phase. This involves manual labelling of individual posts whether they correspond to cyberbullying or not. In order to minimize the internal resources and benefit from existing platforms elsewhere, we used Amazon's Mechanical Turk service to determine the labels for our training corpus. Mechanical Turk is an online marketplace that allows requestors to post tasks (called HITs) which are then completed by workers that are cheaply paid by the requestors per HIT completed. The process is anonymous and very quick. Due to the subjectivity nature of the comprehending task, and without providing any further guidelines or exemplification, we asked three workers to label each post as either cyberbullying or not, and, then a majority voting is used to infer the correct label.

The general skeleton of the approach is highlighted in Fig 2. Especially, Stanford parser was employed for dependency features whose discourse is extended to include two consecutive sentences (current sentence and previous sentence). Besides, our implementation also introduces co-referencing and semantic role labelling [21] in order to strengthen the word relationship discovery. Illinois named-entity recognition [22] was employed in order to identify Person entities.

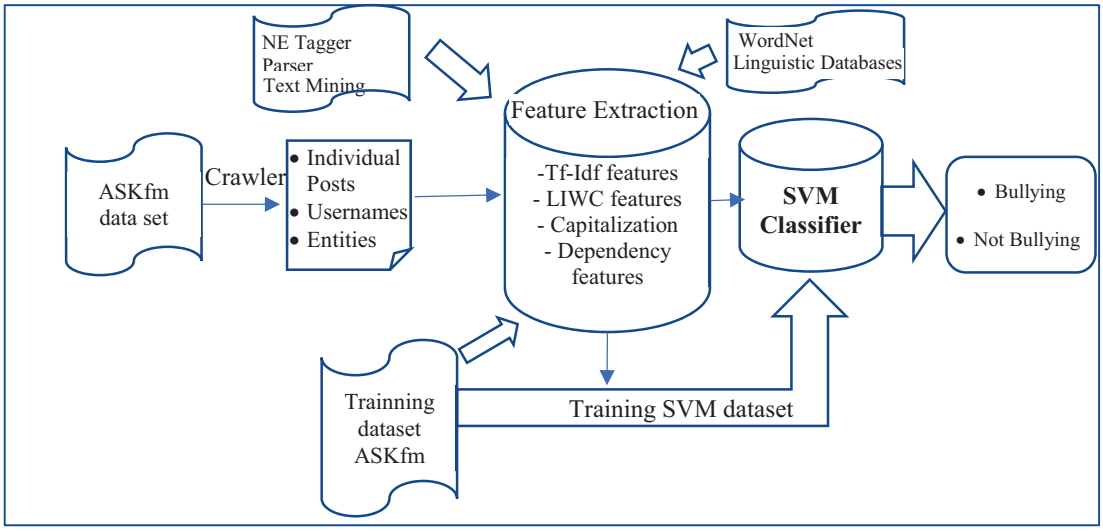

Fig 2. General synoptic of the Cyberbullying detection system

G. User interface (GUI)

The GUI is built with the aim of simplicity and user-friendliness. In this design, the user can choose to use a manual input text on the GUI as an input or choose a text document to be the input. To manually type in text as input, the user can just type it in the text pane provided at the input text area, see Fig. 3. A web interface was also implemented as C# ASP.NET MVC application using AngularJS.

III. Experiments and Results

A. Implementation and metrics

Standard evaluation metrics employed in information retrieval field, which include precision, recall and f-score [16], were used in this study. In particular, precision presents the percent of identified posts that are truly offensive messages. Recall measures the overall classification correctness, which represents the percent of actual offensive messages posts that are correctly identified. False positive (FP) rate represents the percent of identified posts that are not truly offensive messages. False negative (FN) rate represents the percent of actual offensive messages posts that are unidentified. F-score represents the weighted harmonic mean of precision and recall, which is defined as:

$$F_1 = 2\frac{\mathrm{Pr}ecision.\mathrm{Re}call}{\mathrm{Pr}ecision+\mathrm{Re}call}$$

$$F_2 = 5\frac{\mathrm{Pr}ecision.\mathrm{Re}call}{4\mathrm{Pr}ecision+\mathrm{Re}call}$$

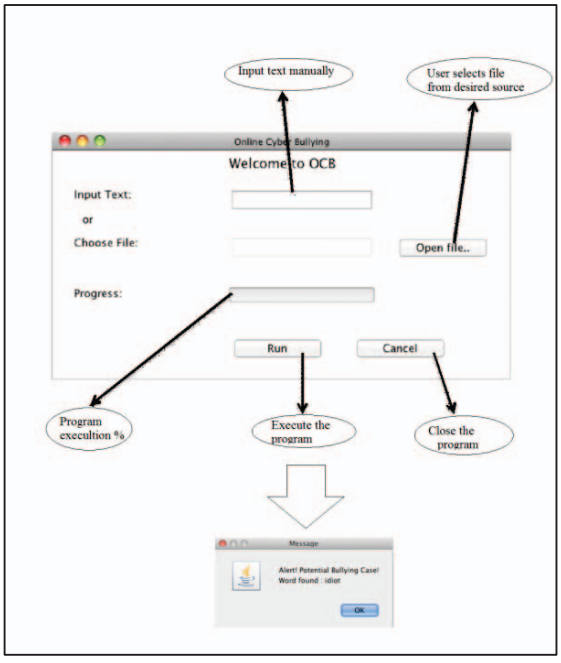

Fig 3. GUI design and example of output

B. Experimentation

The algorithm was applied to the test the collected ASKfm dataset. The testing dataset includes randomly selected 30% of the total dataset, while the remaining 70% were used for training purpose. We tested the contribution of each set of the features to the classifier. The results were evaluated in terms of accuracy, precision, recall, F1-measure and F2-measure. Table 2 summarizes the results.

Table 2. Overall result on test dataset using various features

Feature	Acc.	Prec.	Reca.	F1-mea	F2-mea
tf-Idf	97,3%	31,2%	68,4%	42,85%	55.23%
LIWC	76,4%	28,4%	57,1%	32,56%	41,97%
Depen	67,5%	27,3%	60,6%	37,64%	48,72%
tf-Idf+LIWC	97,8%	42,4%	75,1%	54,20%	65,01%
LIWC + Depen	82,1%	38,4%	69,5%	49,47%	59,81%
tf-Idf+Depen	97,9%	58,9%	78,4%	67,26%	73,53%
All features	**99,4%**	**69,0%**	**84,9%**	**76,13%**	**81,15%**

The result highlighted in Table 2 indicates the following:

- The relatively high accuracy shown in Table 2 of all most all features is partially explained by the selection procedure employed in the initial dataset collection where the use of bullying string matching allowed us to gather relatively high relevant data.
- The use of all features (*Idf-idf* + *LIWC* + *Capitalization* + *Dependency*) provides as expected the best result in terms of accuracy, precision, recall, and, thereby, *F1-measure* and *F2-measure* as well.
- *tf-Idf* still perform individually much better than other individual features; namely, *LIWC* and *Dependency* features taken alone.
- The combination of *Dependency* and *tf-idf* features outperforms that of dependency and LIWC feature set. This demonstrates the relevance of the dependency feature to capture cyberbullying cases.
- The preceding shows that almost all features-like approach performed relatively well in terms of recall but they exhibited relatively moderate to low performance in terms of precision, although the –All features based approach achieves acceptable rate of 69%. This testifies of the task difficulty in discriminating bullying from non-bullying cases given that the post actually contains potential bullying terms.
- The relatively low precision results in cyberbullying have also been pointed out in other research findings. For instance, Kontostathis et al. [23] reported an order of magnitude of the precision from 18% to 84% according to various queries. Bigelow et al. [24] reported a precision of less than 50 % in almost all considered cases.
- In summary, although it is fairly simple to achieve high recall rates by using large dictionaries of bad words and word vectors to determine offensive language, achieving high precision through good discrimination power between comments containing bad language from actual cyberbullying comments is still an open issue.
- It should be noted that the reasoning advocated in our approach takes into account only the textual information of the post regardless the characteristics of the user. However, it is worth reporting alternative results that look beyond the single post and explore the available information regarding the sender, especially the profile of the user. The latter contains valuable information about the user, including his/her location, age, gender, hobbies and possibly some past activities. Such information has been explored by Bigelow et al. [24], although the overall results are still not fully satisfactory because of various reasons, including the often incomplete profile information and unstructured data. Nevertheless, we believe that expanding our reasoning to include profile feature would be a promising future research. Other research direction includes the use of Paragraph Vectors [25] to aid in bullying classification where it will be possible to find posts of similar semantic meaning to known bullying posts. On the other hand, Google has also opened up SyntaxNet [26], a syntactic language parser of higher accuracy than standard Stanford Parser. The investigation of such directions is worth exploring as part of our future research.
- Following [27-28], the set of *tf-idf* features can further be optimized and reduced using Latent Semantic Analysis, Principal Component Analysis or any other dimension reduction techniques in order to restrict to select the most relevant features with respect to bullying detection.

IV. Conclusion

This paper describes an automated cyberbullying system detection. The proposal uses natural language processing techniques, text mining and machine learning in order to infer whether a post belongs to bullying category or not. A combination of features have been employed. This includes standard tf-idf whose weights are boosted for those terms that belong to Insult / Swear category, LIWC selected features (those who likely convey bullying / offense meaning), Unusual capitalization count, and Dependency parser in order to relate the offensive word with corresponding entity. Finally Support Vector Machine with special setup in order to deal with largely unbalanced dataset was employed for classification task. In our work, a case study from ASKfm social media dataset has been investigated, where

Amazon Mechanical Turk Service was used to label training posts. The experiment investigated the performance of the classifier when using various combination of the aforementioned features. It turns out that the augmented feature set constituted of a concatenation of *tf-Idf*, *LIWC*, *Capitalization*, *Dependency*, yields the highest performance in terms accuracy, precision, recall, F1 and F2 scores.

As described earlier the classification of individual posts is more or less limited in precision where distinction between bullying posts and regular posts including bad language. Another limitation lies in the ability to retrieve comments from ASKfm. A question does not show up on a profile page unless it has been answered, which means that if the victim does not answer a bullying question there is no way for this particular scanner to find it.

On the other hand, the use of Amazon Mechanical Turk is also not fully risk free as it brings extra subjectivity and possible uncertainty of the results as it is not excluded that the some of the participants were fully ignorant or with a limited linguistic ability, which substantially decreases the quality of the results in overall.

This work opens up new direction for future research through using advanced parser, dimension reduction and taking into account user's profile in order to strengthen the detection capabilities. In terms of significance of the results, further statistical testing employing second order statistics can be employed in order to strengthen the observations noticed in Table 2 for instance.

REFERENCES

[1] T. Johnson, R. Shapiro, and R. Tourangeau, "National survey of American attitudes on substance abuse XVI: Teens and parents.," in The National Center on Addiction and Substance Abuse. vol. 2011, 2011

[2] Ditch the Label. The annual cyberbullying survey. http://ditchthelabel.org/downloads/cyberbullying2013.pdf. [Online; accessed January-2016].

[3] S. Hinduja and J. W. Patchin, "Bullies Move Beyond the Schoolyard: A Preliminary Look at Cyberbullying," Youth Violence And Juvenile Justice, vol. 4, 2006, pp. 148–169.

[4] H. Cowie. Cyberbullying and its impact on young people's emotional health and well-being. The Psychiatrist, 37(5):167-170, 2013.

[5] D. Boyd, *Why Youth (Heart) Social Network Sites: The Role of Networked Publics in Teenage Social Life*. MacArthur Foundation Series on Digital Learning, Youth, Identity, and Digital Media, David Buckingham, Ed., MIT Press, Cambridge, MA, 2007

[6] Cybersafetysolutions.com. What can i do if i am cyberbullied. http://www.cybersafetysolutions.com.au-/fact-what-to-do-if-i-am-bullied.shtml. [Online; accessed January-2016]

[7] Bsecure. Available http://www.safesearchkids.com/BSecure.html

[8] Cyber Patrol. Available: http://www.cyberpatrol.com/cpparentalcontrols.asp

[9] eBlaster. Available: http://www.eblaster.com/

[10] IamBigBrother. Available: http://www.iambigbrother.com

[11] F. Mishna, M. Saini, and S. Solomon, Ongoing and online: Children and youth's perceptions of cyber bullying. *Children Youth Services Rev. 31*, 12, 1222–1228, 2009

[12] K. Dinakar, R. Reichart, and H. Lieberman. Modeling the detection of textual cyberbullying. In The Social Mobile Web, 2011

[13] http://www.ra.ethz.ch/CDstore/www2009/caw2.barcelonamedia.org/index.html

[14] Nobullying.com. Understanding the reasons behind ask.fm bullying. http://nobullying.com/ask-fm-cyber-bullying/. [Online; accessed February-2017]

[15] D. Schultz. *A History of Modern Psychology*. Burlington: Elsevier Science, 2013

[16] B. Ribeiro-Neto and R. Baeza-Yates, Modern Information Retrieval, ACM Press, 1999

[17] C. Felbaum and G. Miller, WordNet, An Electronic Lexical Database, 1998

[18] V. Nahar, X. Li, and C. Pang. An effective approach for cyberbullying detection. Communications in Information Science and Management Engineering, 3(5):238, 2013

[19] Stanford Natural Language Processing Group. The stanford parser. http://nlp.stanford.edu/software/lex-parser.shtml . [Online; accessed 24-February-2016].

[20] T. Joachims. Making large-scale svm learning practical. LS8-Report 24, University of Dortmund, LS VIII-Report, 1998.

[21] D. Jurafsky and J. H. Martin, Speech and Language Processing, Chapter 22: Semantic Role Labelling, Technical Report, Stanford University, 2015.

[22] Illinois Named Entity Tagger, available at http://cogcomp.cs.illinois.edu/page/software_view/NETagger

[23] A. Kontostathis, L. Edwards, and A. Leatherman, "Chatcoder: Toward the tracking and categorization of internet predators," In Proc. Text Mining Workshop 2009 held in conjunction with the Ninth SIAM International Conference on Data Mining, 2009

[24] J. L. Bigelow, A. Edwards, L. Edwards, detecting Cyberbullying using Latent semantic indexing, Proceedings of ACM CyberSafety Conference 2016, USA, p. 11-14

[25] Q. V. Le and Tomas Mikolov. Distributed representations of sentences and documents. arXiv preprint arXiv: 1405- 4053, 2014.

[26] S. Petrov. Announcing syntaxnet: The world's most accurate parser goes open source. http://googleresearch.blogspot.se/2016/05/ announcing-syntaxnet-worldsmost. html. [Online; accessed 16-May-2016

[27] Y.Chen, Y. Zhou, S. Zhu, and H. Xu. Detecting offensive language in social media to protect adolescent online safety. In Privacy, Security, Risk and Trust (PASSAT), 2012 International Conference on and 2012 International Confernece on Social Computing (SocialCom), pages 71{80. IEEE, 2012

Y. Chen, L. Zhang, A. Michelony, and Y. Zhang. 4is of social bully fillteing: identity, inference, in influence, and intervention. In Proceedings of the 21st ACM international conference on Information and knowledge management, pages 2677-2679. ACM, 2012