Fast Learning for Sentiment Analysis on Bullying

Jun-Ming Xu, Xiaojin Zhu
Department of Computer Sciences
University of Wisconsin-Madison
Madison, WI 53706, USA
{xujm,jerryzhu}@cs.wisc.edu

Amy Bellmore
Department of Educational Psychology
University of Wisconsin-Madison
Madison, WI 53706, USA
abellmore@wisc.edu

ABSTRACT
Bullying is a serious national health issue among adolescents. Social media offers a new opportunity to study bullying in both physical and cyber worlds. Sentiment analysis has the potential to identify victims who pose high risk to themselves or others, and to enhance the scientific understanding of bullying overall. We identify seven emotions common in bullying. While some of the emotions are well-studied before, others are non-standard in the sentiment analysis literature. We propose a fast training procedure to recognize these emotions without explicitly producing a conventional labeled training dataset. We apply our procedure to social media posts on bullying and discuss our findings.

Categories and Subject Descriptors
H.2.8 [**DATABASE MANAGEMENT**]: Database Applications—*Data mining*; I.2.7 [**ARTIFICIAL INTELLIGENCE**]: Natural Language Processing—*Text analysis*

General Terms
Algorithms

Keywords
sentiment analysis, bullying, social media mining

1. INTRODUCTION TO BULLYING
Bullying, also called peer victimization, has been identified as a serious national health concern among adolescents [39, 38, 1]. One is being bullied or victimized when he or she is exposed repeatedly over time to negative actions on the part of others [30]. Far-reaching and insidious sequelae of bullying include intrapersonal problems [18, 17] and lethal school violence in the most extreme cases [26]. Youth who experience peer victimization report more symptoms of depression, anxiety, loneliness, and low self-worth compared to their nonvictimized counterparts [3, 4, 13, 15]. Bullying happens traditionally in the physical world and, recently, online as well; the latter is known as cyberbullying [7, 12, 42, 41].

Bullying takes multiple forms, most noticeably face-to-face physical (e.g., hitting), verbal (e.g., name-calling), and relational (e.g., exclusion) [2, 23, 29]. Cyberbullying reflects a venue (other than face to face contact) through which verbal and relational forms can occur. Bullying usually starts in primary school, peaks in middle school, and lasts well into high school and beyond [27, 36, 9]. Across a national sample of students in grades 4 through 12 in the United States, 38% of students reported being bullied by others and 32% reported bullying others [40].

Participants in a bullying episode take well-defined *roles* (see Figure 1). The traditional roles include the bully (or bullies), the victims, bystanders (who saw the event but did not intervene), defenders of the victim, assistants to the bully (who did not initiate but went along with the bully), and reinforcers (who did not directly join in with the bully but encouraged the bully by laughing, for example) [34]. We also defined two new roles that are common in social media: reporter (may not be present during the episode, unlike a bystander) and accuser (accusing someone as the bully) [43]. More than one person can have the same role in a bullying episode.

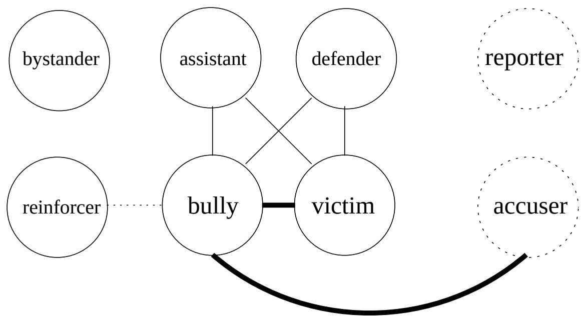


Figure 1: Participants' role in a bullying episode (Reproduced from [43])

Bullying has been studied from at least two angles. The social science study of bullying has a long history but is limited by small sample size and time-consuming data collection [28]. The computer science study of bullying is emerging with promising results [43, 22, 10, 33, 20, 5, 21]. The latter has a distinct focus on bullying in social media, where participants of a bullying episode post their bullying experiences. Such posts, called **bullying traces**, include but far exceed incidences of cyberbullying. In our previous work, we showed how computational techniques can help to answer several scientific questions by mining bullying traces in Twitter [43]. However, to our knowledge *emotions in bullying traces* have not received wide attention from the research community. It is the goal of the present paper to raise awareness of this fertile topic, and to provide initial results on the sentiment study of bullying.

2. EMOTIONS IN BULLYING TRACES AS IDENTIFIED BY OUR SENTIMENT CLASSIFIER

Sentiment analysis on bullying traces is of significant importance. Victims usually experience negative emotions such as depression, anxiety and loneliness. In extreme cases such emotions are more violent or even suicidal, for example,

> "I'm tired of all this bullying. I could never stand up for myself & sometimes I just want to kill myself."

Detecting at-risk individuals via sentiment analysis enables potential interventions. In addition, social scientists are interested in sentiment analysis on bullying traces to understand participants' motivations.

There are a wide range of emotions expressed in bullying traces. After manually inspecting a number of bullying traces in Twitter, our domain experts identified seven most common emotions:

1. Anger: *"He is always laughing at me because he is a bully damnit! #Ashley"*
2. Embarrassment: *"@USER everyone is bullying me because I couldn't find the word peach in a crossword. It's 1am"*
3. Empathy: *"@USER I'm sorry you get bullied. I'm really surprised at how many people this has happened to. #bulliesSuck "*
4. Fear: *"i was being bullied and i didn't want to go to school really i would throw fits everymorning and i hope that michel sees this"*
5. Pride: *"Everyone on this earth is a bully , except me . Because I'm perfect. #jillism"*
6. Relief: *"@USER I was rambling and then... I cried. Like, CRIED. He was touched! APC helped me thru the teasing and bullying man..."*
7. Sadness: *"things were bad when I was younger I got bullied so much because of my disabilites I don't want the same thing happening to my brother."*

This list is by no means comprehensive. Other emotions or mixtures of several basic emotions may also appear in bullying traces. These seven categories are the most common ones based on our current study. Also note that due to the length limits (140 characters), an individual tweet may be only a sentence in a conversation thread. Therefore, the majority of bullying traces in Twitter cannot be associated with definite emotions.

3. A FAST MACHINE LEARNING PROCEDURE

As shown above, some emotions involved in bullying traces have not been well studied in sentiment analysis, for example, *embarrassment* and *relief*. To make the problem worse, manually labeling a large amount of training tweets is difficult and time consuming even for our domain experts.

Recognizing these challenges, we use a fast training procedure for sentiment analysis. Our goal is supervised learning, specifically classifying a tweet into one of the predefined emotion categories. However, we require no explicit labeled training data on tweets. Instead, we will rely on "distantly labeled data" (to be made clear next) that are much easier to obtain. We point out upfront that it will be difficult to assess the accuracy of the resulting classifier, since we do not have an in-domain labeled dataset. However, our observations point to a useful classifier. Coupled with the ease of training and its applicability to other emotions and domains, our procedure is still attractive.

3.1 Relations to Prior Work

Most sentiment analysis work focused on the overall polarity of a document: positive, negative or neutral [32, 24]. A few work considered several basic emotions at a finer level and created emotional lexicons for each category [37]. Recently, sentiment analysis on social media [44], especially Twitter [31], has been receiving increasing attention. Cambria *et al.* [6] proposed a sentiment analysis approach to identify malicious posts from social media. Our domain of bullying is fresh with very few existing resources. In addition, although bullying traces are abundant, only a small fraction of them are associated with strong emotions. It poses challenges to obtain enough training examples for all the emotion categories, especially the rare and non-standard ones.

Our approach is inspired by the "concept labeling" work of Chenthamarakshan *et al.* [8] to minimize the supervision effort in constructing text classification models. In their system, instead of labeling a set of training examples experts annotate how "concepts" are related to the target class. We push this idea further where neither labeled examples nor labeled concepts are necessary for building the emotion classifiers.

Our procedure consists of two steps. The first step is in the same spirit as the dictionary-based sentiment lexicon generation method [16], which exploits synonym structure of a dictionary to bootstrap the sentiment lexicon. Our second step is similar to the idea of corpus-based sentiment lexicon generation method [14, 19], which uses a domain corpus to extend sentiment lexicon by sentence structure or sentiment consistency assumption. As tweets are very short – usually a few sentences – the sentiment is usually consistent within a tweet.

3.2 Task Description

Following [43], we obtain bullying traces via Twitter streaming API such that: (i) each tweet contains at lease one the following keywords "bully, bullied, bullying", and (ii) the tweet passes a bullying-or-not text classifier. We want to recognize the emotion involved in each bullying trace identified by this procedure. We define 8 emotion classes: *anger*, *embarrassment*, *empathy*, *fear*, *pride*, *relief*, *sadness*, and *other*. The last class captures bullying traces without obvious emotion or not one of the seven emotions. Thus, our task is to build an eight-class text classifier with little supervision.

3.3 Fast Learning

Our learning procedure includes four steps: (1) collecting seed words, (2) collecting online documents, (3) creating feature extractors, and (4) building a text classifier. None of the steps requires explicit labeling a corpus.

Collecting Seed Words. We start by collecting seed words S_e which are related to each emotion e (except for the *other* category). Lexicons exist for certain emotions such as *anger* and *sadness* but not all [37]. As we want to handle the non-standard emotion categories, we create such lists from two general resources which are available for all emotions:

1. Many websites provide synonym dictionary service.[1] We look up the category name of emotion e such as "anger" and add all its synonyms to S_e^{SYN}.

2. We search for the category name of emotion e in WordNet [25, 11], and add all words appearing in the synsets to S_e^{WN}. In addition, we also include all words in synsets listed as their "derivationally related form" and their "similar to," "full troponym" or "full hyponyms" sets depending on the part of speech (adjectives, verbs, or nouns).

By doing so, we collect two seed word lists S_e^{SYN} and S_e^{WN} for each emotion e. This step took less than half an hour manually. Note that it does not require any human judgments and can be implemented automatically if preferred.

Collecting Online Documents. We can broaden the coverage of the keywords by collecting documents containing them. We invoked Twitter search API to query each keyword and retrieve up to 100 recent tweets per query. We queried each word in S_e^{SYN} and S_e^{WN} separately, and obtained two tweet corpora T_e^{SYN} and T_e^{WN}. Obviously, other search services can be employed, too. Given the seed word list, this step can be automated without any human intervention.

Creating Feature Extractors. We perform stopword removal and stemming on T_e^{SYN} and T_e^{WN} as in [43]. Our stopword list is based on the SMART system [35], augmented with domain specific stopwords such as "bully," "bullying," "bullied," "bullies," "@USER," "ref" and some punctuations. We then represent each tweet in T_e^{SYN}, T_e^{WN} by unigrams and bigrams features. We count the occurrences of each feature collectively within T_e^{SYN} or T_e^{WN} and remove features appearing less than five times. We define a vocabulary as the union of seed words in $S^{SYN} \cup S^{WN}$ and the remained features in $T^{SYN} \cup T^{WN}$.

[1] `http://www.synonyms.net/synonym`
`http://dico.isc.cnrs.fr/dico/en/search`
`http://dictionary.reference.com/`

With the vocabulary, we represent S_e^{SYN} as a feature vector v_e^{SYN} where the elements are the counts. We normalize the vector so that it has norm 1. Do the same for S_e^{WN}, T_e^{SYN}, and T_e^{WN} separately to obtain v_e^{WN}, v_e^{SYN}, and v_e^{WN}. Here we treat each of T_e^{SYN} and T_e^{WN} as a single large document. Furthermore, we treat the union $S_e^{SYN} \cup S_e^{WN} \cup T_e^{SYN} \cup T_e^{WN}$ as single document and compute its feature vector v_e^{all}. Thus, for each emotion e we have five feature vectors $\{v_e^{SYN}, v_e^{WN}, v_e^{SYN}, v_e^{WN}, v_e^{all}\}$. In total, we have 35 such feature vectors for the seven emotions.

We use these 35 vectors as feature extractors. Given a test document we apply the same text processing and represent it as feature vector d. We then compute the inner product

$$d^\top v$$

against each of the 35 feature extractors v above and obtain a 35-dimensional vector x. Clearly, no supervision from human is needed in this step either.

Building Text Classifiers. This is the step where traditionally labeled bullying tweets are needed. Instead, we use easy-to-obtain distantly-labeled data. Though our domain is tweets, we train a text classifier on Wikipedia pages. Wikipedia API supports downloading pages matching a title or category name query. For each word in $S_e^{SYN} \cup S_e^{WN}$, we collect the retrieved Wikipedia pages. [2] Each such page is automatically labeled as with emotion e. We therefore automatically constructed a labeled Wikipedia corpus with 964 pages.

We run each Wikipedia page in this corpus through our feature extractors to represent the page as a 35-dimensional vector. We train a standard 7-class SVM (note we do not model the "other" class yet) on the Wikipedia corpus. We compared linear and RBF kernels, tuned SVM regularization parameter C and γ in the RBF kernel function ($\exp(-\gamma\|x-y\|^2)$) in the grid $\{10^{-3}, 10^{-2}, \ldots, 10^{3}\}$ with 10-fold cross validation. The best model is obtained with RBF kernel, $C = 1000$ and $\gamma = 0.1$.On the Wikipedia corpus, it achieves a CV error of 15%. Its confusion matrix is shown in Table 1.

3.4 Model Evaluation and Usage

To understand the performance of the trained SVM, we compare it against three baseline methods. Note the comparison is based on the Wikipedia corpus, not the Twitter domain where we have no labeled data. Using the SVM on Twitter will be discussed at the end of this section.

The three baseline methods are:

1. S^{SYN}. For test document d, we compute the inner product $d^\top v_e^{SYN}$ for each emotion e and predict the class with the maximum value:
$$e^* = \arg\max_e d^\top v_e^{SYN}.$$
Ties are randomly broken.

[2] It is important to note that the nature of the Wikipedia API means that the pages do not necessarily contain the query keywords, which enables us to learn something more than keyword matching.

Table 1: Confusion Matrix of the 7-class SVM on the Wikipedia corpus

	predicted as						
	ang.	emb.	emp.	fear	pri.	rel.	sad.
ang.	112	0	0	9	0	2	3
emb.	0	21	0	3	0	2	1
emp.	1	0	7	7	0	1	2
fear	3	1	0	381	1	23	4
pri.	0	0	0	4	23	0	1
rel.	2	1	0	42	0	198	3
sad.	4	0	2	14	0	4	82

Table 2: Cross validation error of different methods

Fast Training SVM	S^{SYN}	S^{WN}	Majority
0.15	0.31	0.43	0.42

2. S^{WN}. Same as above but use the WordNet keywords:
$$e^* = \arg\max_e d^\top v_e^{WN}.$$
Both baselines are related to simple keyword matching.

3. Majority. All five feature extractors make their own predictions as above, and there is a majority vote among the five for the final decision. Again, ties are randomly broken.

Table 2 shows the cross validation error of these methods. The proposed fast training SVM achieves the lowest error.

However, the above results were all on the Wikipedia corpus. Recall that our test domain is Twitter, for which we do not have labeled data. When a test tweet comes, we first convert it into the 35-dimensional vector via the feature extractors and apply the trained SVM. We set a threshold τ on the margin output from SVM, whenever the largest margin is lower than τ, we predict it as *other*. Otherwise, we predict the label with the largest margin. The threshold τ is set manually by controlling the positive rate at 5% on a separate random tweet data set.

4. EMOTION IN BULLYING TRACES

We apply the SVM to 3,001,427 bullying traces from August 5, 2011 to April 12, 2012 (about eight months). Figure 2 shows the number of daily bullying traces in each emotion categories. Overall, the number of bullying traces is increasing because of growing social media usages. All emotion curves have the similar shape but different offset (note the y-axis is in log scale), indicating that the fraction of different emotions remain stable in the study period. The curves show a weekly (7-day) pattern, which we hypothesize is due to fewer direct interactions among students during the weekends. The few spikes are caused by celebrity events related to bullying which generated a large number of tweets. In what follows, we remove the few spike days since they are outliers.

We aggregate the counts over the study period for each category and compute their fractions over all bullying traces. Figure 3(a) shows that most (94%) bullying traces are not associated with obvious emotions, which matches our observations from manual inspection. Figure 3(b) presents a break down of the 6% emotional bullying traces. Half of them contain *fear*, followed by *sadness*, *anger* and *relief*. *Embarrassment*, *empathy* and *pride* are virtually absent. This also highlights the data skewness issue if the human annotators were to manually label bullying traces.

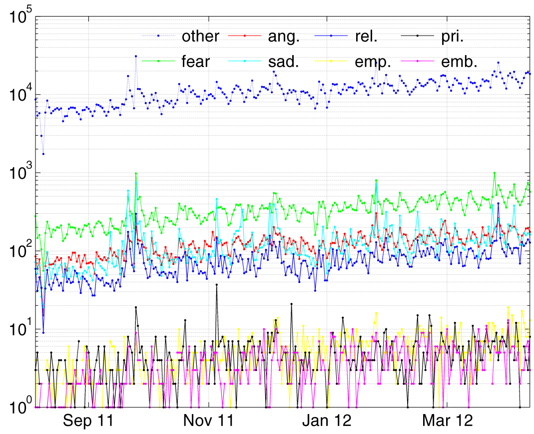
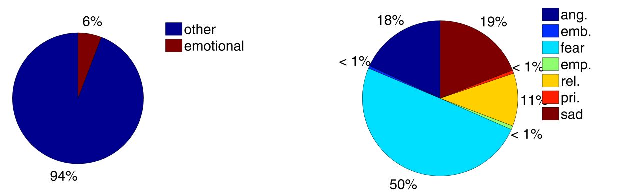


Figure 2: The daily counts of bullying traces in different emotion categories from August 5, 2011 to April 12, 2012.



(a) emotional or not (b) fraction within emotional

Figure 3: Fraction of emotion categories

Recall that participants in a bullying episode take several well-defined roles. We hypothesize that different roles may express different emotions. We apply author-role classifier [43] to the bullying traces, therefore, each tweet is associated with an author role label by this classifier. Figure 4 shows the fraction of emotions for each author's role. We assume that authors of one role generate tweets in one emotion with probability p. The bars show the MLE estimations of p and the error bars indicate the Binomial 95% confidence intervals. Compared to other roles, accusers seem to express more fear but less anger. Reporter and victims seem to experience more sadness and relief than other roles. However, these observations should be taken with a grain of salt: The emotion in a bullying trace may not be the author's own feelings. It is possible that the authors sometimes discuss other participants' emotions. Our emotion classifier is not capable of distinguishing emotions of the author vs. of other people. A detailed analysis with deeper natural language processing remains future work. In addition, we have noticed that accusers often express fear jokingly (i.e., teasing; see below). For example, *"@USER lol really?! I'm so scared!! I hope I am not verbally beaten. You cyber bully ;),"* *"@USER you are such a bully!!!haha & im sooooo scared if him.lol."* This might help explain why accusers seems to have more fear.

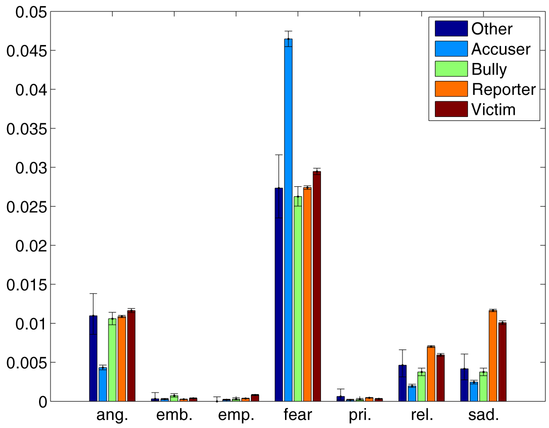


Figure 4: The fraction of emotions by author's role.

In our previous work, we found that many bullying traces were written jokingly (i.e., *teasing*), which may indicate the lack of severity of a bullying episode [43]. It is interesting to see if there is any differences in terms of emotion between teasing and non-teasing bullying traces. In Figure 5, we observe that teasing bullying traces contain less *sadness* and *relief*. This seems reasonable, as in general these emotions are expressed more seriously rather than jokingly. On the other hand, teasing bullying traces contain more *fear*. We speculate that people may pretend to be afraid of a bully even though in reality they are not. For example,*"@USER I'm so scared haha there's like ten girls then like 30 lads! I'm gonna get so bullied#boohoo,"* *"@USER eh ya!!!! sometimes i very scared to approach them, like i want to bully them like that LOL HAHAHA.."*

5. CONCLUSIONS

We identified a wide range of emotions in bullying traces and proposed a fast learning procedure to train a model to automatically recognizing them. We applied the trained model to Twitter posts on bullying and report several interesting findings. We hope this study encourages the community to devote more effort to sentiment analysis on bullying, with the goal of reliably identifying individuals at-risk.

Acknowledgments

We thank Kwang-Sung Jun, Hillary Breidel, and Maria Hughes for their help on the data and the paper.

6. REFERENCES

[1] American Psychological Association. APA resolution on bullying among children and youth, 2004. `http://www.apa.org/about/governance/council/policy/bullying.pdf`.

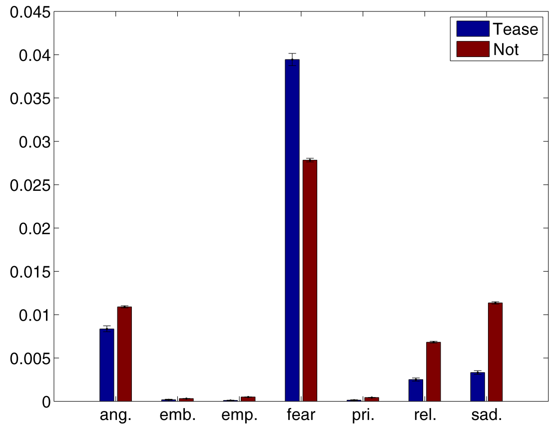


Figure 5: The fractions of emotions by teasing or not.

[2] J. Archer and S. M. Coyne. An integrated review of indirect, relational, and social aggression. *Personality and Social Psychology Review*, 9:212–230, 2005.

[3] A. D. Bellmore, M. R. Witkow, S. Graham, and J. Juvonen. Beyond the individual: The impact of ethnic context and classroom behavioral norms on victims' adjustment. *Developmental Psychology*, 40:1159–1172, 2004.

[4] B. K. Biggs, J. M. Nelson, and M. L. Sampilo. Peer relations in the anxiety-depression link: Test of a mediation model. *Anxiety, Stress & Coping*, 23(4):431–447, 2010.

[5] T. Bosse and S. Stam. A normative agent system to prevent cyberbullying. In *WI-IAT 2011*, pages 425–430, 2011.

[6] E. Cambria, P. Chandra, A. Sharma, and A. Hussain. Do not feel the trolls. In *the 9th ISWC*, 2010.

[7] W. Cassidy, M. Jackson, and K. N. Brown. Sticks and stones can break my bones, but how can pixels hurt me? students' experiences with cyber-bullying. *School Psychology Int'l*, 30(4):383–402, 2009.

[8] V. Chenthamarakshan, P. Melville, V. Sindhwani, and R. D. Lawrence. Concept labeling: Building text classifiers with minimal supervision. In *the 22nd IJCAI*, pages 1225–1230, 2011.

[9] C. R. Cook, K. R. Williams, N. G. Guerra, T. E. Kim, and S. Sadek. Predictors of bullying and victimization in childhood and adolescence: A meta-analytic investigation. *School Psychology Quarterly*, 25(2):65–83, 2010.

[10] K. Dinakar, R. Reichart, and H. Lieberman. Modeling the detection of textual cyberbullying. In *the 5th ICWSM Workshop on Social Mobile Web*, 2011.

[11] C. Fellbaum. *WordNet: An Electronic Lexical Database*. MIT Press, Cambridge, MA, 1988.

[12] B. K. Fredstrom, R. E. Adams, and R. Gilman. Electronic and school-based victimization: Unique contexts for adjustment difficulties during adolescence. *J. Youth and Adolescence*, 40(4):405–415, 2011.

[13] S. Graham, A. Bellmore, and J. Juvonen. Peer victimization in middle school: When self- and peer views diverge. In J. E. Zins, M. J. Elias, and C. A. Maher, editors, *Bullying, victimization, and peer harassment: A handbook of prevention and intervention*, pages 121–141. Haworth Press, New York, NY, 2007.

[14] V. Hatzivassiloglou and K. McKeown. Predicting the semantic orientation of adjectives. In *the 35th ACL*, pages 174–181, 1997.

[15] D. S. J. Hawker and M. J. Boulton. Twenty years' research on peer victimization and psychosocial maladjustment: A meta-analytic review of cross-sectional studies. *J. of Child Psychology And Psychiatry*, 41(4):441–455, 2000.

[16] M. Hu and B. Liu. Mining and summarizing customer reviews. In *the 10th ACM SIGKDD*, pages 168–177, 2004.

[17] S. R. Jimerson, S. M. Swearer, and D. L. Espelage. *Handbook of Bullying in Schools: An international perspective*. Routledge/Taylor & Francis Group, New York, NY, 2010.

[18] J. Juvonen and S. Graham. *Peer harassment in school: The plight of the vulnerable and victimized*. Guilford Press, New York, NY, 2001.

[19] H. Kanayama and T. Nasukawa. Fully automatic lexicon expansion for domain-oriented sentiment analysis. In *the EMNLP 2006*, pages 355–363, 2006.

[20] A. Kontostathis, L. Edwards, and A. Leatherman. Text mining and cybercrime. In M. W. Berry and J. Kogan, editors, *Text Mining: Applications and Theory*. John Wiley & Sons, Ltd, Chichester, UK, 2010.

[21] A. Latham, K. Crockett, and Z. Bandar. A conversational expert system supporting bullying and harassment policies. In *the 2nd ICAART*, pages 163–168, 2010.

[22] H. Lieberman, K. Dinakar, and B. Jones. Let's gang up on cyberbullying. *Computer*, 44:93–96, 2011.

[23] T. D. Little, C. C. Henrich, S. M. Jones, and P. H. Hawley. Disentangling the "whys" from the "whats" of aggressive behavior. *Int'l J. of Behavioral Development*, 27:122–133, 2003.

[24] B. Liu and L. Zhang. A survey of opinion mining and sentiment analysis. In C. C. Aggarwal and C. Zhai, editors, *Mining Text Data*, pages 415–463. Springer US, 2012.

[25] G. Miller. Wordnet: a lexical database for english. *Communications of the ACM*, 38(11):39–41, 1995.

[26] M. H. Moore, C. V. Petrie, A. A. Braga, and B. L. McLaughlin. *Deadly lessons: Understanding lethal school violence*. The National Academies Press, Washington, DC, 2003.

[27] T. R. Nansel, M. Overpeck, R. S. Pilla, W. J. Ruan, B. Simons-Morton, and P. Scheidt. Bullying behaviors among US youth: prevalence and association with psychosocial adjustment. *J. Amer. Medical Assoc.*, 285(16):2094–2100, 2001.

[28] A. Nishina and A. D. Bellmore. When might aggression, victimization, and conflict matter most?: Contextual considerations. *J. of Early Adolescence*, pages 5–26, 2010.

[29] K. Nylund, A. Bellmore, A. Nishina, and S. Graham. Subtypes, severity, and structural stability of peer victimization: What does latent class analysis say? *Child Development*, 78:1706–1722, 2007.

[30] D. Olweus. *Bullying at school: What we know and what we can do*. Blackwell, Oxford, UK, 1993.

[31] A. Pak and P. Paroubek. Twitter as a corpus for sentiment analysis and opinion mining. *Computer*, pages 1320–1326, 2010.

[32] B. Pang and L. Lee. Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2(1-2):1–135, 2008.

[33] M. Ptaszynski, P. Dybala, T. Matsuba, F. Masui, R. Rzepka, and K. Araki. Machine learning and affect analysis against cyber-bullying. In *the 36th AISB*, pages 7–16, 2010.

[34] C. Salmivalli. Participant role approach to school bullying: Implications for intervention. *J. of Adolescence*, 22(4):453–459, 1999.

[35] G. Salton. *The SMART Retrieval System*. Prentice-Hall, Englewood Cliffs, NJ, 1971.

[36] P. K. Smith, K. C. Madsen, and J. C. Moody. What causes the age decline in reports of being bullied at school? Towards a developmental analysis of risks of being bullied. *Educational Research*, 41(3):267–285, 1999.

[37] C. Strapparava and A. Valitutti. Wordnet-affect: an affective extension of wordnet. In *the 4th LREC*, volume 4, pages 1083–1086, 2004.

[38] The American Academy of Pediatrics. Policy statement–role of the pediatrician in youth violence prevention. *Pediatrics*, 124(1):393–402, 2009.

[39] The White House. Background on White House conference on bullying prevention, 2011. http://www.whitehouse.gov/the-press-office/2011/03/10/background-white-house-conference-bullying-prevention.

[40] T. Vaillancourt, V. Trinh, P. McDougall, E. Duku, L. Cunningham, C. Cunningham, S. Hymel, and K. Short. Optimizing population screening of bullying in school-aged children. *J. of School Violence*, 9:233–250, 2010.

[41] H. Vandebosch and K. V. Cleemput. Cyberbullying among youngsters: profiles of bullies and victims. *New media & society*, 11(8):1349–1371, 2009.

[42] J. Wang, R. J. Iannotti, and T. R. Nansel. School bullying among adolescents in the united states: Physical, verbal, relational, and cyber. *J. Adolescent Health*, 45(4):368–375, 2009.

[43] J.-M. Xu, K.-S. Jun, X. Zhu, and A. Bellmore. Learning from bullying traces in social media. In *NAACL-HLT*, 2012.

[44] C. Yang, K. H.-Y. Lin, and H.-H. Chen. Emotion classification using web blog corpora. In *WI 2007*, pages 275–278, 2007.