

Cyberbullying in the United States: A Zenti Use Case

Steven Cracknell (Zenti)

Antonia Czinger (Columbia University)

Approach

The main purpose of this evaluation is to understand the influences surrounding Cyberbullying and to determine a broader picture of suicide risk and the influences Cyberbullying has on individuals. This paper provides an outline that will be the starting point for developing a constellation of classifiers in Zenti to combat cyberbullying.

Research Question: *Can Cyberbullying (CB) be identified and prevented?*

Overview of Cyberbullying in the US

The paper will involve researching a variety of public sources with the goal of 'getting your head around' the concept of cyberbullying and understanding the magnitude of it.

- What is cyberbullying?
- What are the most common means of cyberbullying?
- What is the typical demographic (age group, gender, race)?
- What is the extent of it amongst different groups?
- How is it being tracked today?
- Is there any ground-breaking research around the prevention of CB?
- Are there any documented examples of CB that lead to someone's suicide?

Cyberbullying Strategy

Having a better understand of the extent and magnitude of CB – the goal will be to identify an area where CB appears to be most prominent and how we can potentially use Zenti to identify signs of cyberbullying (both bullying and being bullied)

This will require the collection of authentic text samples of:

- How people bully (what they say/don't say)
- Type of language use, behaviour they express, words, phrases, etc.
- How people that are being bullied cry out for help (what signals do they give out that they are being bullied?)

Creating specific CB classes in Zenti and fine tuning

This will involve the setup and training of specific CB classes that will be used to monitor communication with the aim of identifying instances of CB. The language used to train classes will be based on the research, with variations for abbreviations/misspellings, as well as samples pulled from the Twitter feed.

Analysing the results

Review and analyse results to determine internal consistency of the classifier (accuracy, precision, recall, false positives, false negatives) as well as significance and specificity.

Next Steps

Preliminary work on a CB classifier has shown that such language can be identified (see Figure 1 at end of document). Zenti would like to work with subject matter experts (SMEs) to continue development of additional classifiers. Zenti would like to partner with an organization that can use this actionable data in an intervention program, for both those bullying and those being bullied.

Background Research into CB

Background research serves as a starting point for identifying authentic text samples for classifier development in social media.

Cyberbullying: can it be identified and prevented?

Cyberbullying and cyber-harassment refers to bullying and harassment via email, texting, instant messenger, twitter, or other online sources. It may involve sending bullying or harassing messages, posting or sharing private content (such as nude pictures, drug use, or other potentially embarrassing personal information), impersonating the victim to manipulate harmful effects, or impersonating a third person to manipulate harmful effects.

Unlike other forms of bullying, it can happen anywhere at any time, even when the victim is at home, it is often anonymous, and it can cause lasting damage to the victim's reputation and search history if it involves the distribution of personal materials. The psychological distance afforded by cyberspace both in terms of anonymity and in terms of lack of physical presence may cause cyberbullies to detach from the consequences, feel less empathy, and be more vicious. Most of the research that has been done on cyber-bullying involves the bullying experienced by adolescents. We will take a look at this data before expanding our scope to include adults.

The Cyber Bullying Research Center is one of the premier sources for information in regards to cyberbullying of teenagers. Run by Dr. Sameer Hinduja (Florida Atlantic University) and Dr. Justin Patchin (University of Wisconsin-Eau Claire) it is dedicated to providing up-to-date information on the nature, extent, causes and consequences of cyber-bullying among teenagers. Over the last decade, the Cyber Bullying Research center has surveyed nearly 15,000 middle and high school students annually from over 97 different schools throughout the United States. The first two studies were voluntary online surveys and therefore yielded results that were disproportionately high. However, the eight most recent studies were conducted using random samples of school populations. These surveys found that about 25% of students were cyberbullied, 9% were cyberbullied within the last 30 days, and about 16% admitted to cyberbullying.

The statistics obtained by the Cyber Bullying Research Center seem to hold true across the board, or at least with the help of averaging. The Center has reviewed a number peer-reviewed articles, 51 of which included cyberbullying victimization rates and 42 of which included cyberbullying offending rates. The actual statistics in the studies ranged widely from 2.3% for victimization and from 1.2% to 44.1% for offending, but the average was remarkably similar to the numbers found by the Center. According the average of these studies about 21% of teens have been cyberbullied and about 15% have admitted

to cyberbullying. These numbers would seem to indicate that one out of four teens has experienced cyberbullying while one out of six has cyberbullied.

Cyberbullying is correlated with traditional bullying, which despite the growth of modern technology remains more common. The Urban Institute's study on bullying showed that 17% of students had been victims of cyberbullying, 41% of students had been victims of physical bullying, and 45% of students had been victims of psychological bullying. There are some gender differences when it comes to the type of bullying experienced. 50% of female students have been subjected to psychological bullying compared to 39% of male students. 45% of male students have been subjected to physical bullying, compared to 37% of female students. 23% of females reported cyberbullying victimization, compared to 11% of males. Additionally of the 17% of victims who sought help, females comprised two thirds. It is noted that 71% of victims turned to their parents, 56% turned to friends, 38% turned to school counselors, and 35% turned to teachers in a study by Zweig, Dank, Lachman & Yahner in 2013.

There are some factors that place certain students more at risk for bullying than others. People are often bullied due to race, weight, sexuality, or disability. According to one study by Russell, Sinclair, Poteat, and Koenig one third of adolescents report witnessing racially motivated bullying (Russell, Sinclair, Poteat, & Koenig, 2012). One fourth of boys and one third of girls report weight-based teasing from peers, but prevalence rates increase to approximately 60% among the heaviest students (Puhl, Luedicke, & Heuer, 2011). 84% of surveyed students reported witnessing a bully teasing or harassing overweight peers during physical activities (Puhl, Luedicke, & Heuer, 2011). LGBT students are at higher risk for bullying with 61% of LGBT students feeling "unsafe or uncomfortable as a result of their sexual orientation."

Bullying and cyberbullying are related to low self-esteem, suicidal ideation, anger, frustration, and a whole host of emotional and psychological problems. Bullied adolescents suffer higher rates of suicidal ideation, suicide attempts, and suicide; cyberbullying in particular seems to be correlated highly with this type of suicidal thinking (JAMA Pediatrics, 2014). Victims tend to show higher levels of insecurity, depression, anxiety, and loneliness while perpetrators tend to show higher levels of conduct issues and a dislike for school (American Medical Association, 2001). Oddly enough, students who both bully and are bullied are at greater risk for both mental health and behavior problems than those who are merely perpetrators or merely victims (Center for Disease Control, 2012). 22% of frequent perpetrators, 29% of frequent victims, and 38% of frequent bully/victims reported suicidal thinking over the past year (Adolescent Health, 2013). However, it is important to note that the association between bullying and suicide-related behaviors is also usually combined with other factors, such as a depressive temperament or delinquency (Hertz, Donato, & Wright, 2013).

Currently, a vast majority of the research concerns adolescents; the literature on adult cyberbullying is still sadly lacking. Despite focusing on adolescents, the Cyber Bullying Research Centers reports receiving more emails, phone calls, and comments from adults who have been cyberbullied than from teens. In fact, when cyberbullying does happen to adults it can be even more nefarious and far reaching. Often, serious steps are taken to damage the person's public reputation. For instance, revenge porn websites, which

involve the non consensual public distribution of sexual images are a largely adult phenomenon. Adult cyberbullying is also often tied to long-term abusive relationships such as the relationship to a lover or a family member.

There are several ways of attempting to curb cyberbullying. There are certain common sense solutions to cyberbullying such as not engaging with the bully, refusing to reply to bullying messages, and reaching out to a trusted person for help. Bullying and cyberbullying laws vary from state to state. State and local lawmakers make laws (in their state and education codes) and use model policies (that provide guidance to districts and schools) to address bullying. There may be a single law or multiple laws to deal with the phenomenon. In some cases, bullying even appears in the criminal code of a state that may apply to juvenile offenders. Zenti may be used by schools, lawmakers, or employers to find people who may be violating their policies or laws.

Cyberbullying Strategy

According to the U.S. Legal Definitions, Cyber-bullying could be limited to posting rumors or gossips about a person in the internet bringing about hatred in others minds; or it may go to the extent of personally identifying victims and publishing materials severely defaming and humiliating them.

In the context of Zenti we are looking at identifying tweets that are bullying. Tweets that are bullying must have one or more of the following characteristics: 1) be insulting 2) be exclusionary, 3) be threatening, or 4) be a way of calling on others to also bully the subject. Cyberbullying must involve a victim and a perpetrator.

Insulting tweets are likely to involve name calling and profanity i.e. “you are a disgusting slut bitch,” but may also be more subtle i.e. “no one wants to listen to your tone def singing; please don’t ever perform at the talent show again.” Exclusionary tweets are tweets that seek to socially isolate the subject by publicly asking him or her not to partake in an activity for reasons that appear to be personal as opposed to practical i.e. “you are too ugly to pose for prom pictures with us” not “since you are out of state, you can’t go on the NY trip.” Threatening tweets may threaten anything from physical violence, to harassment, to exclusion i.e. “If you come to prom, we will be pouring pigs blood over you.” Finally, tweets that call on others to perpetuate the bullying will use words that call for an action i.e. “everyone should throw stuff at Jane.”

Identifying tweets that are cyberbullying is complicated by the fact that there are some tweets that may look like cyberbullying, but actually involve something more innocuous. These confusing tweets may involve: 1) joking between friends, 2) calling out a group of people for damaging behavior. Luckily, there are usually some markers that can help us identify the false positives. While joking jibes between friends might sound like bullying, they are often tempered with words like “lol” or “jk” or a smiley face. Tweets that call out a group will usually mention that group i.e. “people who drive hummers, suck.” If a broad group is mentioned, it is not cyberbullying because the target is not easily identifiable—remember, we are looking for a personal attack. Notably, a tweet can still be racist or prejudiced without being cyberbullying (i.e. “blacks shouldn’t be allowed to vote”).

| First | Latest | Analyse | Subject |
|--------------|--------------|--------------------------------|---|
| Oct 29 00:43 | Oct 29 00:43 | Cyber Bullying | @aubreestapleyy You are a piece of shit. And I hate your fucking guts. |
| Oct 29 00:43 | Oct 29 00:43 | Cyber Bullying | @emeryemeryii Goodnight @facebook! And by the way, #SuckMyBalls! Yours in Jesus Christ you guys suck, Emery (My Fucking Name) Emery |
| Oct 29 00:41 | Oct 29 00:41 | Cyber Bullying | @daniig55 @FloggnawGeneral EW AFTER THIS TWITTER FIGHT IM BLOCKING YOU JUST CAUSE IM DONE SEEING YOUR UGLY BLACK ASS CAST A SHADOW OVER MY MENTIONS |
| Oct 29 00:41 | Oct 29 00:41 | Cyber Bullying | @darthdanaa @vRoyalxshot You fucked a horse. Shut up you degenerate fucking gypsy bitch. |
| Oct 29 00:36 | Oct 29 00:36 | Cyber Bullying | @saxjetson Annnnnnd you're a sorry ass bitch but you know...your shit to deal with lol not mine |
| Oct 29 00:36 | Oct 29 00:36 | Cyber Bullying | @daniellefriberg Like go fuck yourself douche bag. I fucking hate you. Can't believe you piece of shit cheated on me |
| Oct 29 00:34 | Oct 29 00:34 | Cyber Bullying | @letiffanyloyy I . Dont. Fuck. With. You. You Lil' Stupid ass Bitch I A'int Fucking With You |
| Oct 29 00:32 | Oct 29 00:32 | Cyber Bullying | @heightsides GO AWAY FLY NO ONE LIKES U |
| Oct 29 00:29 | Oct 29 00:29 | Cyber Bullying | @_merkerler @Rycroscopic Really... Oh wait I forgot, we're "FUCKING DISGUSTING." #sorrynotsorry |
| Oct 29 00:27 | Oct 29 00:27 | Cyber Bullying | @imdopendurnopee @1zzay #LiesToldOnTwitter take your lying ass out of my mentions please |
| Oct 29 00:26 | Oct 29 00:26 | Cyber Bullying | @_tweetdonem get dick. You fucking asshole. "@rusik_05_05: @ZeeMaraj17 @_TWEETDonEM hi girls" |

Figure 1. The Zenti team created a broad classifier that would identify CB language as a proof of concept. Zenti would like to work with subject matter experts to refine the language training for such classifiers.

Works Cited

"Cyberbullying Facts - Cyberbullying Research Center." Cyberbullying Research Center Cyberbullying Facts Comments. N.p., n.d. Web. 12 Oct. 2014.

"Student Facts on Bullying & Cyberbullying." Megan Meier Foundation. N.p., n.d. Web. 12 Oct. 2014.