**System Integration and Implementation II:**

**Second Written Report**

**Tsinghua University, Taiwan, ROC**

**Spring 2018**

**TOPIC:** Cyberbullying Detection in Social Network platforms

**DESCRIPTION:** Multilingual tool to detect abusive or sarcastic language signaling cyberbullying

**ADVISOR NAME:** 陳宜欣 (Chen Yi-Shin)

**GROUP MEMBER AND INFORMATION:**

Eva Virginia Arevalo Rivara (艾怡華)

Cel: 0903200702

E-mail: [93.evar@gmail.com](mailto:93.evar@gmail.com)

1. **INTRODUCTION:**

Bullying is a phenomenon that continuously targets a subject with repeated abuse. The pervasiveness of social media platforms have facilitated communication within and across communities. However, it has also provided a platform for this behavior. This is what we refer to as “Cyberbullying”; ie. the use of violent/abusive speech against a specific target.

Social platforms have indeed provided a way for abusers to infiltrate the personal lives of the targets in a way that wasn’t possible before. They have contributed to alienation of interlocutors by removing the context of social interaction from conversations, and thus eliminating the burden of responsibility of one’s speech and rendering empathic interactions that generally rule in the realm of personal interaction impossible. A further analysis of this behavior is better suited to the realm of Social Sciences.

However, according to the Anti-Defamation League, although only 28% of students aged 10-18 experienced cyberbullying in 2014, as much as 87% of them have at some point witnessed other people being harassed[[1]](#footnote-1). 90% of social media-using teens who have witnessed online cruelty say they have ignored mean behavior on social media, and more than a third (35%) have done this frequently.[[2]](#footnote-2) The abused themselves reported cyberbullied at a rate of 23%.

This prompts a motivation to find a way to automatize this process of detection of cyberbullying, which has been the target of the present project.

1. **BUILDING THE DATASETS:**
   1. **DATA CRAWLER:**
   2. **PREPROCESSING:**
   3. **ENCODING:**
2. **CLASSIFIER:**
   1. **ARCHITECTURE OF THE MODEL:**
   2. **SETTING THE PARAMETERS:**
   3. **10 FOLDING CROSS VALIDATIONS:**
3. **PROBLEMS/DIFFICULTIES:**
4. **COMPARISON WITH ORIGINAL PROPOSAL:**
5. **FUTURE WORK:**
6. **CONTRIBUTION:**

Previous work in Cyberbullying[[3]](#footnote-3) focus on using Ortony lexicons (eg. Ortony Lexicon[[4]](#footnote-4)), Lists of Profane words or Topic specific Bigrams for text categorization. Other approaches work with Part-of-speech tags that are tightly related to the grammar of a language. This approach, while effective, works only with precompiled lists of word per topic, sentiment and/or language.

1. **WORKFLOW AND ORGANIZATION:**

.

1. **PAPERS REFERENCED FOR THE RESEARCH PHASE:**

This semester, I continued my goal of reading a paper biweekly. I managed to surpass this goal and read a little bit more than initially expected on this topic and get acquainted with a topic I enjoy. Below are the names of the paper I read and referenced:

1. **C. Hutto and Eric Gilbert, VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text, International AAAI Conference on Web and Social Media, 2014**
2. **Martin Wöllmer, Moritz Kaiser, Florian Eyben, Björn Schuller, Gerhard Rigoll, LSTM-Modeling of continuous emotions in an audiovisual affect recognition framework, Image and Vision Computing, Volume 31, Issue 2, 2013, Pages 153-163**
3. **Sundermeyer, Martin & Schlüter, Ralf & Ney, Hermann. LSTM Neural Networks for Language Modeling, 2012**
4. **Z. C. Lipton, D. C. Kale, C. Elkan, and R. C. Wetzel, Learning To Diagnose With LSTM Recurrent Neural Networks, Computing Research Repository, 2015**
5. **Caiming Xiong, Stephen Merity, Richard Socher, Dynamic Memory Networks for Visual and Textual Question Answering, Computing Research Repository, 2016**
6. **Ankit Kumar, Ozan Irsoy, Jonathan Su, James Bradbury, Robert English, Brian Pierce, Peter Ondruska, Ishaan Gulrajani, Richard Socher, Ask Me Anything: Dynamic Memory Networks for Natural Language Processing, Computing Research Repository, 2015**
7. **Sainbayar Sukhbaatar, Arthur Szlam, Jason Weston, Rob Fergus, End-To-End Memory Networks, Computing Research Repository, 2015**
8. **SCHEDULE FOR THE SECOND SEMESTER:**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Task | 02.25 –03.10 | 03.12 – 03.24 | 03.25 –  04.07 | 04.09 –04.21 | 04.22 – 05.05 | 05.06 –05.20 |
| Research | **✓** | **✓** | **✓** | **✓** | **✓** |  |
| Data Collection | **✓** |  |  |  |  |  |
| Classifier with Graph approach | **✓** | **✓** | **✓** |  |  |  |
| Building the classifier model |  |  | **✓** | **✓** |  |  |
| Training the model and getting results |  |  |  | **✓** | **✓** |  |
| Applying 10-fold cross validation |  |  |  |  | **✓** |  |
| Final report |  |  |  |  |  | **✓** |

1. **CONCLUSION**

During this process, I have gotten more acquainted with the whole process of doing research on my field and extract knowledge to apply it into other projects. I expect to delve deeper into areas related to Data Science, Natural Language Processing and Artificial Intelligence/Deep Learning in the future, and use the knowledge acquired through this project in my future academic research and personal/professional projects.

1. **ACKNOWLEDGMENTS**

Special thanks to PhD candidate Elvis Saravia and my advisor Yi-Shin Chen for helping me with guidance in this project and providing me with the opportunity and the means to carry it.

1. Statistics on Bullying, Anti-Defamation League, 2013 [↑](#footnote-ref-1)
2. Teens, Kindness and Cruelty on Social Network Sites, by Amanda Lenhart et al.,2011 [↑](#footnote-ref-2)
3. Modeling the Detection of Textual Cyberbullying by Karthik Dinakar, Roi Reichart, and Henry Lieberman, 2011 [↑](#footnote-ref-3)
4. [↑](#footnote-ref-4)