Corpus for Cyberbullying Prevention

Asunción Moreno, Antonio Bonafonte, Igor Jauk, Laia Tarrés, Victor Pereira

Innovation and Technology Center of the Universitat Politècnica de Catalunya

asuncion.moreno@upc.edu, antonio.bonafonte@upc.com, ij.artium@gmail.com, laiatarresb@gmail.com, vjp1994@msn.com

Abstract

Cyberbullying is the use of digital media to harass a person or group of people, through personal attacks, disclosure of confidential or false information, among other means. That is to say, it is considered cyberbullying, or cyber-aggression to everything that is done through electronic communication devices with the intended purpose of harming or attacking a person or a group.

In this paper we present a starting project to prevent cyberbullying between kids and teenagers. The idea is to create a prevention system. A system which is installed in the mobile of a kid and, if a harassment is detected, some advice is given to the child. In case of serious or repeated behavior the parents are alerted.

The focus of this paper is to describe the characteristics of the database to be used to train the system

Index Terms: bullying, cyberbullying, databases, children

1. Introduction

The fight against bullying and cyberbullying among children and adolescents is becoming a priority. A study by Save the Children published by EL PAÏS[1] reveals that 1 out of 10 children during ESO has suffered harassment or cyberbullying in Spain.

The most frequent types of harassment are direct or indirect insults, spreading rumors, damage to property, physical damage, exclusion and threats.

According to the Ministry of Health, INE and OMS, in 2015, the number of suicides of children under 15 years was 12 and in the range of 15 to 29 years was 247. On the other hand, the number of suicide attempts of children under 15 years old was 273, of whom the majority, 234, were girls.

d-LAB is a program of Mobile World Capital Barcelona whose objective is to carry out Calls for Proposals that serve to promote responses to social problems through the implementation of collaborative pilot projects between private companies and public entities.

Earlier this year, d-LAB launched a call entitled "Fighting cyberbullying through mobile technologies"[2]. The company SafeToNet (STN) was the winner through the "Safeguarding children online" project.

STN proposed a service to get ahead of a child harassing another person via mobile and alert their parents. The service includes warning about sexually improper photographs, and identification of bullying.

The Pilot Project began in July 2018. The Project partners are: d-LAB, SafeToNet (London, UK), Orange and the Innovation and Technology Center of the Universitat Politècnica de Catalunya (CIT-UPC).

The Project consists of several phases:

- Generation of a database in Spanish and Catalan for the training of cyberbullying prevention systems.
- Adaptation of the System currently available in English to Catalan and Spanish.
- Test in several Catalan schools with children between 12 and 14 years old chosen in a way that provides a diverse socio-economic environment to cover cultural varieties.
- Validation and improvement based on the system test on Orange employees and their family members.

This paper deals with the generation of a database in Spanish and Catalan for the training of cyberbullying prevention systems.

2. Database specifications

Two databases have been created, one in Catalan and another in Spanish spoken in Spain. Each database consists of 140,000 posts, of which 100,000 are labeled by two people and 40,000 by a single person.

The posts are labeled according to their content in 7 categories: aggression, anxiety, depression, distress, sexuality, use of substances, and violence. For each category, 5 levels of concern have been established. 1: post nothing worrying, 5: post extremely worrying.

The annotation has been made through a platform available in STN in which there is an automatic system to present the posts to be scored and an efficient annotation system based on keyboard or mouse interchangeably.

[1] https://elpais.com/elpais/2016/02/18/media/1455822566_899475.html

[2] https://d-lab.tech/challenge-3/

3. Recruiting of labelers and technical support

17 people were selected and hired for the labelling task. Most of the annotators are students. They have different backgrounds. Following recommendations from SafeToNet, most of the labelers are psychology students. There are 4 male and 13 female annotators, all of them between 20 and 30 years old. Table 1 shows the background summary of the annotators:

CODE	BACKGROUND
F1	Philosophy student
F2	Teacher degree
F3	Psychology student
F4	Criminology student
F5	Biomedicine student
F6	Psychology student
F7	Social Communication student
F8	Psychologist degree
F9	Biology and Neuroscience student
F10	Statistics and economy student
F11	Children's education
F12	Psychology student
F13	Engineering student
M1	Engineering student
M2	Psychology student
M3	Physical activity and sports degree
M4	Engineering student

Table 1. Code and Background of the annotators

Annotators were contracted on a part time basis (20 hours/week). They work from home. They have some flexibility to manage their dedication to the project during each week. However, their daily dedication to the project can never be higher than 5 hours. This dedication was set to avoid tiredness and lack of concentration during the annotation procedure.

A training week (16th-20th July) was organized at UPC premises. A person from SafeToNet was in charge of the training during the first three days. Annotators were instructed, one category at a time, with several examples chosen from real posts. The last two days, students had a simulation of real work using the annotation platform reaching the labeling of 400 posts each day. Those posts were carefully chosen to show simultaneously several categories and several concern levels. Another re-training week was necessary at the mid part of the project to consensus inter-annotator agreement.

4. Compilation and selection of posts

Posts were selected from several sources such as twitter, teenagers' chats, blogs, forums, medical consultation web sites, etc. Data was manually or automatically downloaded, cleaned, formatted and selected. Posts were chosen to have a minimum number of characters (without counting –not discarding- @names and internet addresses) of 50 and a total maximum of 280. Spanish as spoken in Latin America posts were discarded when possible. As expected, Catalan data was harder to collect. The number of information in internet in Spanish is huge compared against Catalan websites, blogs and consulters. In addition, Catalan speakers are bilingual, and it is very common to find Spanish posts in Catalan sites. Spanish and Catalan can be naturally mixed even in the same post.

5. Quality Assessment

To calculate the consistency between the annotators we used three indices: Cohen's kappa index, Accuracy, and Cronbach's alpha index.

5.1. Cohen's kappa index

The Cohen's kappa index [1] between two annotators labeling the same data measures the consistency between the annotations and compares them with the case of the annotation being random. To calculate the consistency between two annotators, the common messages labeled by both of them are searched and the results of the annotation are compared for each characteristic. The Cohen's kappa index is calculated as:

$$\kappa = \frac{p_0 - p_e}{1 - p_e}$$

Where p_0 is the relative agreement between raters (accuracy), and p_e Is the probability of chance agreement. Cohen's kappa index was calculated on binarized categories: (category A: level of concern 1 or 2; category B: level of concern 3, 4, or 5)

5.2. Cronbach's alpha coefficient

The Cronbach alpha coefficient has been calculated per each category *(y)* as

$$\alpha_y = \frac{K}{K-1}\left[1 - \frac{\sum_{i=1}^{2}\sigma_{yi}^2}{\sigma_t^2}\right]$$

Where K: number of items (K=2: 2 annotators)
σ_{yi}^2 variance of each item (i.e vector of length nposts of category y of tagger i)
σ_t^2 variance of the total (i.e. vector: $t=y_1+y_2$ of length nposts)

6. Results

index	aggression	anxiety	depression	distress	sexuality	substance	violence
Accuracy	0,90	0,98	0,96	0,84	0,97	0,99	0,97
Cohen's kappa	0,70	0,62	0,63	0,58	0,84	0,80	0,70
Cronbach's alpha	0,76	0,53	0,61	0,59	0,80	0,84	0,71

Table 2. Mean indexes for all the categories

The table shows a lower inter agreement in anxiety and distress. This is a consequence of no direct translation of distress within Spanish and Catalan. The annotators had difficulties to distinguish both categories.

7. Discussion

The data collection will be finished on October 15th so that the first prototype will be ready one month later. The results of the complete project will be presented in Mobile World Congress 2019.

8. Acknowledgements

We want to thank d-LAB and STN for the trust placed in us to carry out the project.

9. References

[1] J. Cohen. A Coefficient of Agreement for Nominal Scales. https://doi.org/10.1177/001316446002000104
[2] L.J. Cronbach Coefficient alpha and the internal structure of tests. Psychometrika, Sep 1951, vol 16, Issue 3, pp 297-334