APPLICATION BASED ON FUZZY LOGIC TO DETECT AND PREVENT CYBERBULLYING THROUGH SMARTPHONES

José Á. Concepción-Sánchez, Pino Caballero-Gil and Jezabel Molina-Gil

Department of Computer Engineering and Systems,
University of La Laguna, Tenerife, Spain

ABSTRACT

Derived from bullying, cyberbullying is a new problem that is spreading because of the many advances in technology, like the Internet and smartphones, affecting especially to the world’s youth population. Currently, there are some studies to investigate their effects on victims or suggest different solutions to detect it. However, a definitive tool that can detect and prevent this harassment is no yet available. For this reason, this paper proposes a novel mobile application for smartphones that will allow detecting whether a person is being a victim of cyberbullying. For this, the application is based on a set of data processing techniques and fuzzy logic that make up a system for decision making, capable of detecting harassment effectively. In addition, this paper also includes some experimental results obtained by performing cyberbullying detection tests with the application.

KEYWORDS

Cyberbullying, Fuzzy Logic, Data Processing, Mobile Application, Security

1. INTRODUCTION

In recent years, when the Internet has expanded around the world and has become an indispensable tool in our everyday lives, cybersecurity is playing an increasingly fundamental role. That is why, in a society that lives constantly connected, the Internet is now an ideal space for cybercrimes, such as cyberbullying [1]. Unlike traditional bullying, where harassment of a person is usually verbal or physical, cyberbullying is characterised by harassment or intimidation produced through social networks or instant messaging systems. The consequences of suffering this problem are many and varied. Some of these are low self-esteem, emotional disorders, depression or anxiety among others [2].

Cyberbullying is a global problem (see Figure 1) [3]. It has increased by 88% in the last five years in countries like the UK [4]. Others countries, such as Singapore [5] or Argentina [6], have seen how more and more young people between the ages of 7-18 are being affected by this serious problem. There are countries that are starting to take action, as is the case of Germany, which has made a law proposal to penalise social networks that do not eliminate offensive and humiliating messages [7]. However, this is not enough since many of the messages are sent through instant messaging systems, which are much more difficult to control. In this way, all these data indicate that there is an urgent need to look for possible solutions that can detect, eradicate and prevent this problem.
This paper presents a novel mobile application that will allow the detection of possible cases of cyberbullying at an early stage in order to act accordingly, presenting itself as a real alternative to eliminate this problem. To do this, the proposed mobile application will analyse all the messages that are received on the victim's smartphone and will decide, using data processing and decision-making techniques, if cyberbullying can exist in the analysed content.

One of the techniques used in this paper is fuzzy logic [8], on which the system for decision making is based. Fuzzy logic is a computational intelligence technique that allows working with information with a high degree of imprecision. It is a multivalued logic that allows intermediate values to be able to define evaluations between yes and no. This is the main difference with the conventional logic that works with well-defined and precise information. This work is based on the use of this mathematical tool since it is not possible to use conventional logic to determine if a person is being a victim of cyberbullying. Throughout the paper, the operation of the system will be further detailed.

This paper is structured as follows. In section 2, some works related to this proposal are mentioned. In section 3, some mobile application features and operation are presented. Section 4 describes the proposed architecture for cyberbullying detection. In section 5 some experimental results of the cyberbullying detection are presented. Finally, the paper is closed with some brief conclusions and future works in section 6.

2. RELATED WORKS

There are numerous papers that study different ways of trying to combat against cyberbullying [9] [10]. For instance, in [11] police are described as an actor in addition to parents, students, schools and service providers on the Internet to combat this problem. According to that study, the police is one more means that can help in preventing cyberbullying by carrying out information tasks for students, parents and schools, creating online information systems (in addition to face-to-face channels), identifying perpetrators and helping the victims.

Research has been also carried out based on paradigms of text mining for topics related to the detection of cyberbullying such as the detection of online sexual harassers [12], the detection of vandalism [13] or the detection of cyberterrorism [14] [15]. However, few studies have been developed to find technical solutions that allow the detection of cyberbullying. Among these
studies, there are some focused on the detection of cyberbullying through patterns in the analysed texts. For example, the paper [16] proposes the use of machine learning to detect cyberbullying. According to it, through automatic learning, the proposed tool can detect language patterns used by bullies and victims, and develop rules to automatically detect bullying content. The data that were used to carry out the tests were extracted from the web Formspring.me, with a result of 78.5% accuracy in the detection of messages with harassment.

On the other hand, the papers [17] [18] [19] propose that for the detection it is necessary to take into account the context as well as the profile and characteristics of the users being studied. The obtained results reflect an improvement in the accuracy for the detection of cyberbullying. In our proposed system based on fuzzy logic, we have taken into account these studies, incorporating a series of input variables that will allow deducing if there are patterns where messages can be discarded depending on the context and the studied user.

Another interesting study [20] proposes a mechanism that allows recognising in social networks both content and highly potential users in terms of cyberbullying with a high degree of effectiveness. However, this proposal has been designed for its use in social networks, so it would not contemplate other forms such as instant messaging systems, which is one of the most used means for cyberbullying [21].

Generally, studies related to cyberbullying are focused on the search of patterns to detect it as well as on the effects it produces on its victims. In addition, a few proposed applications have focused on this specific area, being a clear disadvantage since today there are numerous different ways and tools that could be used to harass victims.

The system proposed in this paper does not focus on a specific area. The application installed on the victim's smartphone will detect messages and notifications from different media. Once detected, the app will proceed to process the content and, in case there are any word or expression that can be classified as a possible case of cyberbullying, fuzzy logic will be used as a mechanism to make the decision as to whether the victim may be actually being harassed.

3. MOBILE APPLICATION

According to statistical data, by the year 2020, there will be 6.1 trillion active smartphones around the world [22], which means that 70% of the world's population will use them [23]. In addition, the increase in the use of smartphones is growing faster among young people aged 16 to 24 [24]. All these data show that smartphones have become an essential part of our society, so it can be used in the worst case as a tool for cybercrime. Therefore, this work is focused on the development of an application for these mobile devices that serves as a tool to detect and prevent a global problem such as cyberbullying. Its main features are:

- Privacy: As in most cases the victims of harassment try to hide their situation [25], this application is designed so that the parents of the youth can install it on their mobile phone without he/she having to know it. To do this, the application will be installed on the smartphone in a hidden way, without an icon. Also, the bullies will also not know that the victim has the application installed on the mobile device. Thus, in case it is detected that the youth is being a victim of cyberbullying, only the parents will be notified so that they can take the appropriate measures (see Figure 2).

- Easy to use: At the usability level, the application is very intuitive and easy to use, since most of its operation is on the background. Parents will only have to configure the contact parameters and the code to access the application.
Once the application is installed on the smartphone of the possible victim, the number of the contact to be notified via SMS should be indicated in the event that a case of cyberbullying is detected. In addition, it is also necessary to enter a numeric code that will be used later to return the application to the foreground by making a call to that code. After setting these parameters, it will only be necessary to press a button that will leave the application running in the background until the call to the code provided in the initial configuration is made.

Note that even if the mobile reboots or goes off, the proposed application will start automatically and continue to run in the background when the smartphone turns on.

3.1. Notification Detection

The mobile application will be listening in the background so that every time a notification to the smartphone arrives, it can collect the information from it (see Figure 3). For this, the application parses the information that is in the notification to extract the texts and if possible, also the application and its issuer. Once obtained, if the content is sufficient, the app proceeds to the analysis to check if there is cyberbullying. Otherwise, it will be saved encrypted in a local database until there is more content to be able to analyse it.

Thus, any notification of instant messaging applications like WhatsApp or Telegram, email applications like Gmail or even social networks like Facebook as well as any other application...
that sends some type of text can be analysed to be able to detect if the user of the smartphone is a victim of cyberbullying.

3.2. Privacy and Security

The information that this application works with is very sensitive because most of this information is formed by conversations and personal notifications of the user of the application. Therefore, it is necessary that all this information is treated as carefully as possible because it could get to make a misuse of it and invade the privacy of the user.

In this way, no one will be able to obtain complete texts of notifications or conversations, neither the parents nor legal custodians who installed the application on the child. They will only get, in case cyberbullying is detected, the words and expressions that have triggered the alarm. No other content will be accessible to anyone.

In addition, the mobile application will only work locally with the data and will only save the collected information of the notifications in case it is insufficient to be able to analyse it, avoiding unnecessary analyses that could affect the battery of the mobile phone. In this case, the information will be encrypted with the cryptographic algorithm AES 256 CBC mode [26], in order to avoid that the information can be subtracted. Once there is enough content to analyse, the stored information will be removed from the database after analysis.

Finally, the local database of the application containing all the words and expressions related to cyberbullying is updated automatically by a call to a remote database that is continually updated with new words and expressions. In this case, no additional security measures are necessary since the call to update the database is done using HTTPS.

4. PROPOSED ARCHITECTURE

The proposed system architecture to detect possible cases of cyberbullying (see Figure 4) is composed of three main steps:

- Get data from notifications: Messages that the possible victim of cyberbullying receives in his smartphone.
- Data processing: Analyse the messages to eliminate unnecessary words (prepositions, articles, ...) and look for matches with the local database fed by words and expressions that can refer to bullying.
- Fuzzy logic system: Verify, from a series of entries, if there really is a possible case of cyberbullying.

![Figure 4. Structure of the decision-making system.](image)
In the following subsections, the operations of each of the steps will be described in more depth.

4.1. Get Data From Notifications

The objective of this first step is to obtain the text that is going to be analysed in search of possible cyberbullying. To do this, once the application detects a push notification [27], it will check if it has enough content to analyse it at the moment, or if there are messages saved from previous notifications that, concatenated with the content of the current notification, give a set of reasonable information for analysis. If it is sufficient, the next step will be executed. Otherwise, the contents of the notification will be stored in the database for future analysis.

In addition, the mobile application will attempt to obtain the name of the contact or phone number that generated the push notification by parsing the content of the notification. In this way, if there is cyberbullying, the stalker will be detected too.

It is worth mentioning that it has been decided to store and concatenate several messages when they do not contain a minimum amount of text because the battery in the smartphones is limited and running the analysis process for each message can consume a lot of battery. So, we guarantee that it will only run when there is a reasonable amount of content for analysis.

4.2. Data Processing

Once the messages that will be analysed are obtained, in this step they are processed to simplify the chains and to check if there are possible words or expressions identified with cyberbullying.

4.2.1. Pre-Processing

Data pre-processing [28] is a very important step because most of the content that is sent in instant messaging systems or social networks is usually negligible. In this way, a data cleaning method has been implemented to allow rejecting words that lack information such as articles, prepositions or other predefined words that are not to be evaluated later. Using this mechanism, the processing times in the subsequent steps are reduced.

Once the negligible content is removed, the resulting string is divided into words and word sets that will be used in the next step. The result would be classified as shown in Table 1.

<table>
<thead>
<tr>
<th>Id</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>wordm</td>
</tr>
<tr>
<td>m</td>
<td>…</td>
</tr>
<tr>
<td>m+1</td>
<td>expressionm+n</td>
</tr>
<tr>
<td>m+n</td>
<td>…</td>
</tr>
</tbody>
</table>

4.2.2. Analysis of Words and Expressions

As mentioned in Subsection 3.2 about privacy and security, there is a local database in the mobile application that contains a set of predefined words and expressions that may be related to cyberbullying. This database will be used to check if there is any match between its content and the set of words and expressions obtained in the previous step. Its operation is detailed in Figure 5.
If no match is detected, the possibility of cyberbullying will be discarded in the analysed messages and the use of the fuzzy logic mechanism will not be necessary. On the other hand, if any coincidence is detected with the database, the inputs will be prepared for entry into the fuzzy logic system.

![Figure 5. Checking matches.](image)

### 4.3. Fuzzy Logic System

The fuzzy logic system will be responsible for making the final decision about whether the potential victim is receiving cyberbullying. This mechanism is used because it allows obtaining a greater range of decision making, where a deterministic system would not be able to solve the problem. Therefore, the outputs of this system are not only based on yes/no, but also a third option is considered where an incidence is generated. When the system considers that there are not sufficient data to indicate that the user is receiving cyberbullying but the possibility cannot be ruled out, a new incidence will be saved in the database of the application that will be taken into account for future decision making.

In this way, if the process of analysis of words and expressions finds some match between the database and the analysed chains, the input values composed of the following linguistic variables will be initialized:

- **Different Detected (DD):** Number of different words and expressions that have been detected in the analysed chain.

- **Total Detected (TD):** The total number of words and expressions detected in the analysed chain, including repetitions of the same.

- **Last Incident (LI):** Days since the last time an incident was generated.

- **Total Incidents (TI):** Total number of generated incidents in the last thirty days.

In turn, each of the linguistic variables is composed of the linguistic terms HIGH, MEDIUM and LOW, which indicate the possibility that the possible victim is suffering cyberbullying. Figure 6 shows the graphs where they are represented. The X coordinate corresponds to the values that can be taken by the linguistic variables and the Y coordinate with the probabilities corresponding to the linguistic terms.

In the case of DD, it is considered that up to two different expressions detected in the strings have LOW-MEDIUM probability that it is cyberbullying since often young people use a colloquial vocabulary to talk to each other. However, from the four expressions, it is considered that there could be a high degree of probability that the individual is a victim of cyberbullying.
On the other hand, in TD the value ranges of the linguistic terms increase because the repetitions of the expressions or words are taken into account. Thus, the linguistic term HIGH takes its maximum probability from the eight total coincidences.

For its part, the linguistic variable LI takes the linguistic term HIGH with its maximum probability until seven days since the last incidence occurred. A set of incidents in a short period of time can be a clear indicator that the possible victim is receiving cyberbullying.

Finally, the higher the TI value, the greater the likelihood that the individual is being harassed, with the linguistic term HIGH being most likely from the fourteen incidences.

Each of these linguistic terms is represented by a membership function [29] that defines them. For our case, the triangular type membership functions have been used. In Figure 7, the membership functions corresponding to the linguistic terms of the variable LI are shown as an example.

\[
\begin{align*}
    f_{HIGH}(x) &= \begin{cases} 
        1, & x \leq 7 \\
        \frac{15 - x}{15 - 7}, & 7 < x < 15 \\
        0, & x \geq 15
    \end{cases} \\
    f_{MEDIUM}(x) &= \begin{cases} 
        \frac{x - 7}{15 - 7}, & 7 < x \leq 15 \\
        \frac{21 - x}{21 - 15}, & 15 < x < 21 \\
        0, & x \geq 21
    \end{cases} \\
    f_{LOW}(x) &= \begin{cases} 
        \frac{X - 15}{27 - 15}, & 15 < x \leq 27 \\
        0, & x \geq 27
    \end{cases}
\]

Figure 7. Membership functions of LI's linguistic terms.
The next step after fusing is the formulation of specific rules for expressing the combination of influences. As an example, Figure 8 shows a simple rule structure, where DECISION is an output linguistic variable defined by three linguistic terms: YES, INCIDENCE and NO. These linguistic terms are associated with their corresponding membership functions, where the output will depend on which linguistic term has the maximum probability, using the max-membership defuzzification method. Besides, there may be more than one value assignment rule for the DECISION. In this case, the assignments to the DECISION are combined by an implicit AND, so the probability corresponding to the DECISION corresponds to the minimum value between all the input linguistic variable probabilities.

As an illustrative example using the rules shown in Figure 8, if DD has one match as value, its linguistic terms will be fuzzified with 0.5 as LOW and 0.5 as MEDIUM. On the other hand, if TD has a value of four matches, its linguistic term MEDIUM will be fuzzified with a probability of 1. Once fuzzified the linguistic terms, of the three rules established in the example, the third is the one that would be fulfilled, reason why DECISION would have NO as a result. Defuzzifying this linguistic term would indicate that the text analysed does not contain evidence of cyberbullying. If more rules were fulfilled, they would be combined by AND, and the linguistic term with the highest probability will be chosen, as mentioned above.

Finally, once the system returns the output, the mobile application will proceed to perform the corresponding actions depending on the value obtained.

5. EXPERIMENTAL RESULTS

Along with the application and proposed fuzzy-logic-based architecture, a series of experimental tests have been carried out to verify the degree of efficiency in the detection of possible cases of cyberbullying through smartphones. For this, tests have been performed simulating different environments that could be given in a real case. These are described in Table 2.

These environments have been selected for the study because they are the most likely to detect cyberbullying. For example, a case has not been added in which texts have no content related to cyberbullying because the application will directly discard it using data processing. However, in the first case, the application could detect cyberbullying because young people often use, to communicate with each other, words that could be found within the application database.
Table 2. Description of the cases studied.

<table>
<thead>
<tr>
<th>Case</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Use of slang or informal language in conversations between friends. In this case, the application should not detect cyberbullying.</td>
</tr>
<tr>
<td>2</td>
<td>Isolated cases where there could be numerous coincidences with words and expressions related to cyberbullying. An incidence should be generated for future analysis.</td>
</tr>
<tr>
<td>3</td>
<td>Intermediate harassment. It is the most difficult case to classify because there is evidence of possible cyberbullying but it is difficult to locate where the boundary is between harassment or not.</td>
</tr>
<tr>
<td>4</td>
<td>The person is a victim of cyberbullying. It should be detected.</td>
</tr>
</tbody>
</table>

For each of these environments, notifications have been generated with texts that simulate each one to check the degree of effectiveness of the application when it comes to decision making and detection of cyberbullying. The obtained results are classified in Table 3, where:

- **No-Cyberbullying** represents the percentage of samples that were analysed without finding evidence of cyberbullying. In these cases, the application will not take any action.

- **Incidences** are the messages that generated some incidence. The application will save the incident in the database so that it can then be taken into account for the next analysis.

- **Cyberbullying** are the messages that generated alarms. An SMS will be sent with the words and expressions that generated the alarm to the number phone entered in the application settings. If possible, the name of the application will also be sent from where the cyberbullying was done as well as the name of the stalker.

Table 3. Final results.

<table>
<thead>
<tr>
<th>Case</th>
<th>No-Cyberbullying (%)</th>
<th>Incidences (%)</th>
<th>Cyberbullying (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>94%</td>
<td>6%</td>
<td>0%</td>
</tr>
<tr>
<td>2</td>
<td>79%</td>
<td>18%</td>
<td>3%</td>
</tr>
<tr>
<td>3</td>
<td>39%</td>
<td>48%</td>
<td>13%</td>
</tr>
<tr>
<td>4</td>
<td>9%</td>
<td>69%</td>
<td>22%</td>
</tr>
</tbody>
</table>

As can be seen in the first case, most of the texts were classified without cyberbullying content. This is due to the data processing mechanism detailed in the previous section and to the combination of parameters used in the fuzzy logic system, thus avoiding that slang or informal conversations among youth, are identified as cyberbullying. On the other hand, in the second and third cases, a greater number of incidents were generated, something normal due to the detection of a greater number of words and expressions identified with cyberbullying. In these situations, if the data are not very certain, the system will assimilate it by default as cyberbullying and notify the parents of the possible victim, preventing a future problem. Finally, in the latter case, the percentage of alarms increased because there were a big number of incidents, which is normal because the messages that were used in this case always had expressions or words related to cyberbullying.

The obtained results have been satisfactory, since the detection of cyberbullying through the notifications has been in accordance with the environment studied. However, there is still room for improvement. For example, in the first case, although no alarms were generated, 0% of incidents should have been obtained because the analysed texts simulated a conversation between friends. In this way, it may be necessary to add new mechanisms and systems that allow to refine...
and improve the decision-making process for the detection of this problem and to prevent false positives.

6. CONCLUSIONS

Cyberbullying is a problem that affects many young people around the world, and every year the numbers of affected continue to grow very quickly. In the present, there are some studies and small contributions that propose methods to solve this problem but only in specific situations. This paper proposes the use of a mobile application that will be installed on the smartphone of the possible victim to analyse the received messages and notifications, allowing the detection at an early stage of a possible case of harassment or cyberbullying. For this, the application has been enriched with a set of data processing methods and a system based on fuzzy logic that will be in charge of determining if the user is being a victim of cyberbullying. The obtained results showed that the proposal promises to be an effective tool to combat against this problem.

Among the main advantages of this system are the use of smartphones as a means to detect cyberbullying due to its widespread use, as well as the analysis of texts of notifications that any application can generate, being different from other proposals that focus on a specific application or system. In addition, it is very easy to use and totally invisible to save the privacy of the possible victim.

Finally, as future works, it is necessary to study the inclusion of new complementary models that allow improving the accuracy of the proposed system to decide if an individual is being harassed, as mentioned in the previous section. For instance, to include the use of artificial neural networks could help the application to acquire its own knowledge and that it then serves in decision making. In addition, the application should also be improved with voice recognition systems so that it can analyse the notifications that contain this type of messages. Finally, it is also intended to develop the application for the rest of mobile operating systems, so that the application can be accessible to all users regardless of the smartphone.

ACKNOWLEDGEMENTS

Research supported by the Spanish Ministry of Economy and Competitiveness, the European FEDER Fund, and the CajaCanarias Foundation, under Projects TEC2014-54110-R, RTC-2014-1648-8, MTM2015-69138-REDT and DIG02-INSITU.

REFERENCES


AUTHORS

José Á. Concepción-Sánchez is a student of a Master in Mobile Application Development. He graduated as Computer Science Engineer in 2016 from the University of La Laguna and belongs to the CryptULL Research group. During his research period, he has participated in various national and international conferences and also participates in the research projects of the group to which he belongs.

Pino Caballero-Gil is a Full Professor of Computer Science and Artificial Intelligence at the University of La Laguna, Spain, where she leads the CryptULL research group on Cryptology. Her major research interests are in secure mobile applications, stream ciphers, strong identification, cryptographic protocols, vehicular networks and security in wireless networks.

Jezabel Molina-Gil received her Computer Science Engineering Degree from the University of Las Palmas de Gran Canaria (España) in 2007 and her Ph.D. from the University of La Laguna in 2011. Her research is on VANET security, specially in cooperation and data aggregation. She belongs to the CryptULL research group devoted to the development of projects on cryptology since 2007, and is involved in several projects and publications related to this area. She has authored several conference and journal papers.