

Avances en el modelado del Cyberbullying desde la perspectiva de las Tecnologías del Lenguaje Humano

Progress in the modeling of Cyberbullying from the perspective of the human language technologies

Beatriz Botella Gil

Departamento de Lenguajes y
Sistemas Informáticos
Universidad de Alicante
E-03080 Alicante
Beatriz.botella@ua.es

Resumen: Las Teorías de la información y la comunicación (TIC) tienen un papel muy importante en nuestra sociedad. Uno de los usos negativos que se les da, es el acoso entre menores en la red "Cyberbullying". Hay un gran número de menores con acceso a las redes y que poseen teléfonos móviles con edades cada vez más tempranas. Este acoso genera una alerta social importante por las consecuencias que está derivado y hace que sea necesario su estudio para frenar el problema. Esta investigación tiene el fin de encontrar características presentes en los mensajes que puedan indicar junto a los sistemas computacionales actuales, la presencia de mensajes violentos entre menores y aportar posibles soluciones a un caso tan alarmante. **Palabras clave:** Cyberbullying, acoso en Internet, violencia en la red, tecnologías de la información y de la comunicación, TLH, modelado del comportamiento.

Abstract: Information and Communication Theories (ICT) play a very important role in our society. One of the negative uses given to them, is the harassment among minors in the network "Cyberbullying". There are a large number of minors with access to networks and who have mobile phones with younger and younger ages. This harassment generates an important social alert due to the consequences it has and makes it necessary to study it in order to stop the problem. The purpose of this research is to find characteristics present in the messages that can indicate, together with current computer systems, the presence of violent messages among minors and provide possible solutions to such an alarming case.

Keywords: cyberbullying, Internet harassment, Violence in the network, Technology of the information and communication, HLT, behavioral modeling.

1 Motivación de la investigación

Con la aparición de las tecnologías de la información y la comunicación, surge una nueva forma de relacionarnos en la sociedad. Muchas son las ventajas de esta nueva forma de conectarnos digitalmente, por ejemplo, ha facilitado una mayor relación entre personas desde cualquier lugar del mundo, la transferencia de conocimiento entre países o el acercamiento a la información de manera rápida. Pero con el uso indebido de las mismas, se ha detectado un incremento de actos negativos utilizando las herramientas digitales de comunicación, acoso en la red, suplantación

de identidades, acceso a información privada, etc.

Es una realidad que cada vez los menores con edades más tempranas tienen acceso a internet, ya sea mediante un ordenador en casa o en los colegios, como también la pertenencia de un móvil privado para ellos. Según el Instituto Nacional de Estadística (INE), en el caso de los jóvenes y menores cuya edad está comprendida entre los 16 y los 24 años encontramos que el 98,8% tienen teléfono móvil.

Uno de los actos negativos con lo que lucha la sociedad es el llamado Cyberbullying o acoso en la red entre menores. Los casos

cada año aumentan y muchos son los investigadores interesados en la materia debido a su gravedad. (Povedano et al., 2015; Sádaba Chalezquer, Charo and Bringué Sala, 2010; Álvarez-García et al., 2011).

Como punto de partida, debemos definir qué es el Ciberbullying de una manera veraz, (Sádaba Chalezquer, Charo and Bringué Sala, 2010) Lo definen como el uso malévolo y cruel de las tecnologías por parte de un agresor o agresores para humillar y vejar a una víctima, considerada como débil e indefensa, ante una audiencia amplia de espectadores.

Uno de los objetivos prioritarios a solucionar es la problemática del Ciberbullying o acoso en la red. Este problema que crece cada año en todo el mundo, es una evolución del acoso escolar tradicional, aquel acoso que se vive en los centros educativos. Un acoso que siempre ha estado presente, pero que, por la aparición de los sistemas digitales, ha evolucionado a un problema mayor, esto es debido a 3 factores:(Garaigordobil, 2011).

1. Anonimato de los acosadores: Se puede acosar a una persona por internet, con insultos o publicando informaciones falsas o dando información de la vida privada de una persona con un perfil falso e incluso en algunas redes sociales desde el anonimato. Por lo tanto se tiene una imagen de impunidad ante estas circunstancias.(Chóliz y Chóliz, 2009)
2. No existe la empatía: Para los menores una forma de aprender las relaciones sociales es ser capaz de empatizar con los demás. En este caso no hay una visión física del sufrimiento que estamos causando a la víctima. No existe una culpabilidad por ver de manera explícita el sufrimiento que está generando.
3. Disponibilidad 24/7: Este es un factor importante para entender la gravedad del problema y porque crece el número de casos cada año. El acoso escolar "tradicional como hemos comentado, sucede sólo en el centro donde se encuentran los menores. Normalmente ocurrirá en el recreo o en la salida del colegio donde los adultos no estén presentes. En el caso de Ciberbullying puede ser las 24 horas del día, los 7 días de la semana, desde cualquier lugar del mundo y sin vigilancia adulta. Lo que además dificulta el seguimiento de los acosadores.

A través de los medios de los que disponemos hoy en día es muy fácil poder seguir acosando y persiguiendo al acosado en cualquier parte del mundo (Álvarez-García et al., 2011; Calvete et al., 2010).

El problema del Ciberbullying aumenta cada año, y desde 2014 se ha incrementado en un 240 %, según ha publicado la fundación ANAR en su II Estudio sobre el Ciberbullying, dato muy elevado que debemos investigar.

El Ciberbullying supone un nuevo reto para las escuelas por su peligrosa extensión (von Marées y Petermann, 2012). Y las formas en que sucede y las terribles consecuencias que afectan al bienestar psicológico de todos sus implicados, provocando desajustes psicosociales y psicopatológicos.

2 Antecedentes y trabajos relacionados

El interés por esta materia viene dado por el alarmante crecimiento de casos de Ciberbullying en la sociedad actual. En 2016 se ha incrementado este suceso en un 87,7% con 1.207 casos registrados, además de estas cifras, llama la atención que la edad de menores que lo sufren a disminuido considerablemente.

El término Ciberbullying nace con la aparición de las TIC (Tecnologías de la información y de la comunicación) en nuestra sociedad. El primer paso fue crear un área de investigación nueva y separarlo del llamado bullying o acoso escolar.

Mientras que el bullying se ciñe exclusivamente al ámbito escolar, el Ciberbullying es un problema de acoso entre menores a través de las TIC.

Cada año aumentan las cifras del acoso en la red, y es por ello que se ha incrementado el número de artículos en referencia a este tema. Podemos encontrar trabajos indicadores de porcentajes de casos reales existentes, trabajos donde abordan actividades y recursos para implantar en los colegios y trabajos que recogen la detección de mensajes a través de sistemas computacionales.

Algunos trabajos utilizados en esta investigación y que se han relacionado con las investigaciones de Ciberbullying son :

Trabajos Teoría

- *Libro Cyber Bullying , el acoso en la era digital. Robin Kowalski, Susan Limber,*

Patricia Agatston (Kowalski, Limber, y Agatston, 2008).

En este libro que trata de acoso cibernético podemos encontrar desde la definición del problema hasta posibles soluciones. También incluye la normativa legal y políticas específicas. Recoge datos de entrevistas reales con los estudiantes y con los padres. Se basan en cuestiones y preguntas para recabar información y reflejar los datos más relevantes.

- *III Estudio sobre acoso escolar y Cyberbullying según los afectados. Fundación Anar (?)*.

Con este informe quieren dar a conocer la evolución del acoso escolar y del Cyberbullying entre las víctimas y los cambios producidos en sus entornos detectados en el último año. Se basa en casos reales que conocen gracias a las llamadas que se han producido a esta fundación, tanto de víctimas como de sus familiares. Nos da información real de porcentajes y datos estadísticos de como se encuentra en la actualidad el problema.

- *Artículo Usos problemáticos y agresivos de las TIC por parte de adolescentes implicados en Cyberbullying (Giménez Gualdo, Maquilón Sánchez, y Arnaiz Sánchez, 2015)*.

El siguiente estudio examina la extensión y características del Cyberbullying en una muestra de escolares. Su objetivo es saber que situaciones ocurren en el ámbito escolar así como el uso problemático de Internet en menores y la posible adicción que los menores (ciberacosadores y cibervíctimas) hacen de las TIC.

- *Common Sense Reasoning for Detection, Prevention, and Mitigation of Cyberbullying KARTHIK (Dinakar, Picard, y Lieberman, 2015)*.

Trabajan en la detección de intimidación basado en el procesamiento del lenguaje natural. Analizan la apariencia, inteligencia, insultos raciales y étnicos, aceptación social y rechazo; y una vez analizado estos procesos crean una base de conocimiento de "sentido común" que codifica conocimientos particulares sobre

situaciones de intimidación. Además enseñan a las víctimas a hacer frente a este problema y los conecta entre sí en lo que llaman "apoyo emocional".

- *Artículo Cyberbullying in adolescents: Modalities and aggressors's profile (Calvete et al., 2010)*.

Realizan un cuestionario para evaluar la prevalencia de numerosas modalidades de Cyberbullying (formas de violencia, exposición a la violencia, aceptación y rechazo por los compañeros). En este cuestionario recogieron información de la justificación del uso de la violencia, las medidas de agresión y la exposición a la violencia.

Trabajos Sistemas computacionales

- *Automatic Detection of Cyberbullying on Social Networks based on Bullying Features (Zhao, Zhou, y Mao, 2016)*.

Proponen un programa para la detección del acoso cibernético. Usan una lista con palabras e insultos para determinar el nivel de intimidación existente en cada mensaje. Trabajan con la red social Twitter cuya finalidad es la creación de un sistema automático de detección cibernética para categorizar la información y detectar los casos de Cyberbullying.

- *Artículo Cyberbullying detection and prevention: Data mining and psychological perspective (Parime y Suri, 2014)*.

En este trabajo se centran en la minería de datos y técnicas de aprendizaje automático para detectar y prevenir el Cyberbullying. También se discute sobre los factores psicosociales relacionados con el acoso cibernético y como abordar este problema. Plantean algunas propuestas para los futuros algoritmos que puedan detectar y prevenir el Cyberbullying.

- *Methods for detection of Cyberbullying: A survey (Sugandhi et al., 2016)*.

Utilizan un corpus de 3 redes sociales con el objetivo de comprobar cual es el mejor método para detectar el Cyberbullying. Primero hay un pre procesamiento de los datos y a continuación experimentan con una máquina de soporte vectorial, algoritmo de machine learning B.48 y el clasificador "Naïve Bayes".

- *Cyberbullying Detection Based on Semantic-Enhanced Marginalized Denoising Auto-Encoder (Zhao y Mao, 2017)*

Esta investigación aborda el problema del Cyberbullying desde un modelo de aprendizaje distinto a los demás investigadores trabajando con Auto-encoder, donde han creado un sistema denominado Semantic-Enhanced Marginalized Denoising Auto-Encoder (smSDA). Trabajan con dos corpus distintos de redes sociales, Twitter y MySpace. Sus resultados han sido mejores que algunos de los otros estudios basados en el aprendizaje de textos.

- *Automatic detection of cyberbullying in social media text (Van Hee et al., 2018)*.

Este artículo se centra en la detección automática del acoso cibernético en el texto de los medios de comunicación sociales mediante el modelado de mensajes escritos por intimidadores, víctimas y espectadores del acoso cibernético. Describen la recopilación y la anotación detallada de un corpus de ciberacoso en inglés y holandés con lo que realizan una serie de experimentos de clasificación binaria para determinar la viabilidad de la detección automática del ciberacoso.

3 Actualidad

La investigación se inició con la búsqueda de artículos científicos y materiales relacionados con el Cyberbullying para recoger toda la información posible y poder trabajar con una base de conocimiento amplia.

Se creó un informe que recopilaba datos pertenecientes a los casos de Cyberbullying, donde se encuentra:

1. La edad y el género de la víctima.
2. Su entorno y los lugares donde ocurre el Cyberbullying.
3. Cuáles son los medios y el motivo de que ocurra este acoso.
4. Cuándo empieza y su duración y frecuencia.
5. Qué lenguaje se utiliza en el acoso y si se siguen las mismas pautas en la mayoría de los casos.

6. Cuáles son las diferentes fases del proceso.
7. Cómo se relacionan las fases con su lenguaje violento o acción.

Una vez se finalizó el informe, se creó una guía de anotación inicial sobre el Cyberbullying y la técnica de anotación mediante el programa "Gate" para anotar los mensajes pertenecientes a mensajes reales entre menores por la red. En este proceso se selecciona si el mensaje era violento o no, su nivel de alerta, tipo de violencia encontrado (amenazas, anuncio de agresión, ataques a sus aficiones, contra su familia, contra su forma de ser, contra su físico, su situación familiar, su género, su raza y otros. Además añadimos la etiqueta de elegir el rol, ejecutor, incitador, pasivo y víctima. Y un apartado de notas por si los anotadores necesitaban explicar o apuntar alguna corrección en algún texto.

Una vez obtenida la guía de anotación necesitábamos un Corpus. La recolección de Corpus de mensajes reales entre menores no ha sido posible debido a la dificultad legal de poder obtener estos mensajes. Es por eso que decidimos hacer una descarga de mensajes en las redes sociales para poder avanzar en nuestra investigación y formar un modelo de mensajes violentos genérico. Para ello, el Corpus existente, surgió de la descarga de 633 mensajes de la red social Twitter. Las semillas utilizadas tenían relación con el caso de "la manadaz .^{el} pequeño Gabriel".

El Corpus ha sido etiquetado según una nueva guía de anotación que se ha realizado. Y se encuentra en proceso de modificaciones por algunos datos que se consideró debían estar recogidos y en la primera guía no se encontraban y la modificación de los niveles de alerta.

Antes de proceder a trabajar con el corpus, estamos realizando investigación exhaustiva de cuales son los sistemas computacionales utilizados en relación a los casos de Cyberbullying, para poder comparar sus resultados con los nuestros y utilizar técnicas nuevas o implantar modificaciones que nos permitan obtener unos buenos resultados.

3.1 Objetivos

1. Guía de Cyberbullying.
2. Ampliación de nuestro corpus violento
3. Creación de un corpus de mensajes reales entre menores en la red.

4. Identificación de mensajes violentos, relacionando palabras claves o comportamientos que nos ayuden a etiquetarlo en el Ciberbullying.
5. Definir cuales son los elementos necesarios en un mensaje para etiquetarlo como Ciberbullying.
6. Creación de un programa de detección automática del Ciberbullying.

3.2 Hipótesis

1. El Ciberbullying es el acoso más frecuente entre los jóvenes.
2. Cada año incrementa el número de casos de Ciberbullying en menores y con edades mas pequeñas.
3. Se pueden detectar un mensaje como Ciberbullying siguiendo unas normas de este acoso específico.
4. Posibilidad de que pueda crearse un programa que detecte automáticamente la presencia de Ciberbullying.

3.3 Experimentos propuestos

Como experimentos para esta investigación se encuentran los siguientes:

- Fijar las reglas presentes en los mensajes de Ciberbullying.
- Ampliar nuestro Corpus y probarlo en sistemas computacionales.
- Determinar que sistema es el más adecuado.
- Conseguir poner en marcha el programa y que éste sea capaz de detectar y señalar la presencia de Ciberbullying en las redes sociales.

4 Cuestiones específicas de investigación

1. Es posible la detección del Ciberbullying a través del *Machine Learning* para frenar la problemática del Ciberbullying.
2. Se puede definir distintas fases que tiene el acoso entre menores en la red para saber cual es el nivel de gravedad que encontramos en el Ciberbullying.
3. Se podría crear una alerta automática capaz de avisarnos de que se está generando Ciberbullying a un menor.

Agradecimientos

Esta investigación ha sido financiada parcialmente por la Universidad de Alicante, la Generalitat Valenciana, el Gobierno Español (Ministerio de Educación, Cultura y Deporte) y las Ayudas Fundación BBVA a equipos de investigación científica 2016 mediante los proyectos: TIN2015-65100-R, TIN2015-65136-C2-2-R, GRE16-01 y Análisis de Sentimientos Aplicado a la Prevención del Suicidio en las Redes Sociales (ASAP).

Especial agradecimiento a Patricio Martínez Barco y al Departamento de Lenguajes y Sistemas Informáticos de la Universidad de Alicante (DLSI).

Bibliografía

- Álvarez-García, D., J. C. Núñez, L. Álvarez, A. Dobarro, C. Rodríguez, y P. González-Castro. 2011. Violencia a través de las tecnologías de la información y la comunicación en estudiantes de secundaria. *Anales de psicología*, 27(1):221–231.
- Calvete, E., I. Orue, A. Estevez, L. Villard, y P. Padilla. 2010. Cyberbullying in adolescents: Modalities and aggressors' profile. *Computers in Human Behavior*, 26(5):1128–1135.
- Chóliz, M. y M. C. Chóliz. 2009. Ellas, ellos y su móvil: Uso, abuso (¿y dependencia?) del telefono móvil en la adolescencia. *Originales*, 34(1):74–88.
- Dinakar, K., R. Picard, y H. Lieberman. 2015. Common sense reasoning for detection, prevention, and mitigation of cyberbullying. En *IJCAI International Joint Conference on Artificial Intelligence*, volumen 2015-Janua, páginas 4168–4172.
- Garaigordobil, M. 2011. Prevalencia y consecuencias del cyberbullying: una revisión. *Journal of Psychology and Psychological Therapy*, 11(2):233–254.
- Giménez Gualdo, A. M., J. J. Maquilón Sánchez, y P. Arnaiz Sánchez. 2015. Usos problemáticos y agresivos de las TIC por parte de adolescentes implicados en cyberbullying. *Revista de Investigación Educativa*, 33(2):335.
- Kowalski, R. M., S. P. Limber, y P. W. Agaston. 2008. *Cyber Bullying: The New Moral Frontier*.

- Parime, S. y V. Suri. 2014. Cyberbullying detection and prevention: Data mining and psychological perspective. En *2014 International Conference on Circuits, Power and Computing Technologies, ICCPCT 2014*, páginas 1541–1547.
- Povedano, A., M.-J. Cava, M.-C. Monreal, R. Varela, y G. Musitu. 2015. International Journal of Clinical and Health Psychology Victimization, loneliness, overt and relational violence at the school from a gender perspective. *International Journal of Clinical and Health Psychology*, 15:44–51.
- Sádaba Chalezquer, Charo and Bringué Sala, X. 2010. Niños y adolescentes españoles ante las pantallas: rasgos configuradores de una generación interactiva. *CEE Participación Educativa*, 15(June 2009):86–104.
- Sugandhi, R., A. Pande, S. Chawla, A. Agrawal, y H. Bhagat. 2016. Methods for detection of cyberbullying: A survey. En *International Conference on Intelligent Systems Design and Applications, ISDA*, volumen 2016-June, páginas 173–177.
- Van Hee, C., G. Jacobs, C. Emmery, B. DeSmet, E. Lefever, B. Verhoeven, G. De Pauw, W. Daelemans, y V. Hoste. 2018. Automatic detection of cyberbullying in social media text. *PLoS ONE*, 13(10).
- von Marées, N. y F. Petermann. 2012. Cyberbullying: An increasing challenge for schools. *School Psychology International*, 33(5):467–476, oct.
- Zhao, R. y K. Mao. 2017. Cyberbullying Detection Based on Semantic-Enhanced Marginalized Denoising Auto-Encoder. *IEEE Transactions on Affective Computing*, 8(3):328–339.
- Zhao, R., A. Zhou, y K. Mao. 2016. Automatic detection of cyberbullying on social networks based on bullying features. En *Proceedings of the 17th International Conference on Distributed Computing and Networking - ICDCN '16*, páginas 1–6.