

Detección Automática de Casos Urgentes en Foro de Salud Mental

Edgar Altszyler^{*†}, Ariel J. Berenstein[‡], David Milne[§], Rafael Calvo[§] and Diego Fernandez-Slezak ^{*}
^{*}Universidad de Buenos Aires, FCEyN, DC; Instituto de Investigación en Ciencias de la Computación-CONICET
[†]Fundación Sadosky
[‡]Hospital de Niños Ricardo Gutiérrez, Grupo de Informática Traslacional
[§]School of Electrical and Information Engineering, The University of Sydney, Australia

Introducción

Los foros orientados a temas de salud mental necesitan la supervisión de moderadores para brindar apoyo en casos delicados, como mensajes que expresan ideas suicidas. A medida que aumenta el tamaño del foro, la supervisión del moderador deja de ser factible sin la ayuda de sistemas automatizados de priorización [1]. En el presente artículo presentamos un sistema de aprendizaje automático supervisado para el triaje (clasificación según severidad) de mensajes de un foro de salud mental. Este sistema fue desarrollado en el contexto de la competencia CLPsych 2017 shared task [2] y sus resultados serán publicados en los *Proceedings of the CLPsych 2018 workshop* [3].

El triaje consiste en clasificar cada mensaje de un foro de salud mental según la necesidad de intervención. Los niveles del triaje son: *crisis*, *red*, *amber* y *green* reflejando una prioridad decreciente en la atención de los moderadores del foro. El dataset de la competencia contiene 146030 mensajes sin etiquetar y 1588 mensajes etiquetados por especialistas: 1188 mensajes como training set y 400 mensajes como test set. Esta competencia fue una extensión de la realizada en el 2016 [4].

La mayoría de los enfoques en la literatura se centran en el contenido de los mensajes, pero sólo unos pocos autores aprovechan las atributos contextuales. En nuestro trabajo aplicamos un enfoque novedoso teniendo en cuenta no sólo atributos capaces de captar el contenido del mensaje sino también el contexto en el que se producen, considerando el historial de mensajes y la red de interacciones.

Método

Para la clasificación de mensajes desarrollamos un sistema de aprendizaje automático supervisado con un enfoque orientado a la extracción de atributos capaces de captar el contenido semántico del mensaje, el estado mental del autor y el contexto en el que el mensaje fue escrito. Extrajimos un total de 575 atributos organizados en 7 categorías: (1) N-gramas pesados con una transformación TFIDF (2) representación vectorial de los mensajes usando un Skip-gram entrenado en el dataset de mensajes sin etiquetar y con los parámetros maximales encontrados en [5] (3) metadata de los mensajes (4) Lexicones psicolingüísticos y distancias

semánticas con palabras claves como, suicidio, depresión, ansiedad, etc. (5) atributos relacionados con la historia del autor (6) atributos extraídos de los mensajes inmediatamente anteriores (7) atributos extraídos de la red de interacciones.

Entrenamos un total de 18 modelos SVMs tanto lineales como RBF, alimentados con distintos conjuntos de atributos. También usamos 4 modelos de ensemble en los que distintos SVMs entrenados con distintos conjuntos de atributos se combinaron bajo un sistema de votación. Los hiperparámetros C fueron seleccionados en un esquema de 10-fold cross validation.

Resultados

Nuestras implementaciones estuvieron entre los primeros puestos en varias tareas [2], como la identificación de mensajes que requieren atención (macro-F1 de *green vs non-green*), la identificación de mensajes urgentes (macro-F1 de *crisis+red vs amber+green*) y la identificación de mensajes críticos (macro-F1 de *crisis vs non-crisis*). A su vez, realizamos un análisis de importancia de atributos que enfatizó la relevancia de los atributos que caracterizan las interacciones entre usuarios, y la relevancia de los léxicos psicolingüísticos PERMA asociados al bienestar [6], los cuales no habían sido utilizados en tareas de detección hasta ese momento.

References

- [1] Rafael A Calvo, M Sazzad Hussain, David Milne, Kjartan Nordbo, Ian Hickie, and Peter Danckwerts. Augmenting online mental health support services. In *Integrating Technology in Positive Psychology Practice*. 2016.
- [2] CLPsych Shared Task results: <http://clpsych.org/shared-task-2017/>.
- [3] Edgar Altszyler, Ariel J. Berenstein, David N Milne, Rafael A Calvo, and Diego Fernandez-Slezak. Using contextual information for automatic triage of posts in a peer-support forum. In *Proceedings of the Fifth Workshop on Computational Linguistics and Clinical Psychology*, 2018.
- [4] David N Milne, Glen Pink, Ben Hachey, and Rafael A Calvo. Clpsych 2016 shared task: Triageing content in online peer-support forums. In *Proceedings of the Third Workshop on Computational Linguistics and Clinical Psychology*, 2016.
- [5] Edgar Altszyler, Sidarta Ribeiro, Mariano Sigman, and Diego Fernández Slezak. The interpretation of dream meaning: Resolving ambiguity using latent semantic analysis in a small corpus of text. *Consciousness and cognition*, 2017.
- [6] H Andrew Schwartz and et al. Predicting individual well-being through the language of social media. In *Biocomputing 2016: Proceedings of the Pacific Symposium*, 2016.