Detección Anticipada de Riesgos en la Web

Leticia Cagnina[*], M. Paula Villegas[**], M. José Garciarena[***], Martín Loyola[****]
Sergio Burdisso, Darío Funez, Horacio Thomson[*****], Marcelo Errecalde

Proyecto: "Aprendizaje automático y toma de decisiones
en sistemas inteligentes para la Web" (PROICO 03-0620)
Laboratorio de Investigación y Desarrollo en Inteligencia Computacional (LIDIC)
Departamento de Informática, Universidad Nacional de San Luis
Ejército de los Andes 950 - (D5700HHW) San Luis - Argentina
e-mail de contacto: {merreca}@unsl.edu.ar

Resumen

Este artículo describe, brevemente, las tareas de investigación que nuestro grupo está llevando a cabo en el área de *Detección Anticipada de Riesgos* (DAR) *en la Web* . Esta línea de investigación comenzó en el año 2017 con la participación de nuestro grupo en la tarea *eRisk 2017: Pilot Task on Early Detection of Depression* donde se obtuvo el mejor desempeño (de acuerdo a la medida ERDE50) sobre un total de 30 contribuciones de 8 instituciones diferentes de Francia, Alemania, USA, México, Argentina, Canadá y Rusia. A patir de ese evento, se continuó participando en forma ininterrumpida en este evento en otras tareas de DAR vinculadas a depresión, anorexia, y auto-lesiones con distintos enfoques surgidos de los trabajos de postgrado de 5 tesistas de Maestría y Doctorado. En todas las participaciones del grupo, se han presentado propuestas que consituyen en la actualidad el estado del arte del área con más de 12 publicaciones en el tema.

Palabras claves: Minería de Textos, Procesamiento del Lenguaje Natural, Detección Anticipada de Riesgos, Sistemas Inteligentes para la Web

[*]Investigadora - CONICET
[**]Becaria CONICET - LIDIC
[***]Becaria UNSL
[****]Becario CONICET - IMASL
[*****]Becario CONICET - LIDIC

Contexto

La *Detección Anticipada de Riesgos en la Web* se está abordando dentro de la línea de investigación "Minería de Textos y de la Web" en el marco del Proyecto de Investigación Consolidado titulado "Aprendizaje automático y toma de decisiones en sistemas inteligentes para la Web" (PROICO 03-0620). El proyecto, aprobado por evaluadores externos a la UNSL, se desarrolla en el *Laboratorio de Investigación y Desarrollo en Inteligencia Computacional* (LIDIC) de la UNSL y ha sido financiado en forma directa por la UNSL y en forma indirecta por el Programa de Incentivos (22/F637), el CONICET, a través de un investigador adjunto y 4 becas de doctorado y una beca de postgrado de la UNSL.

1. Introducción

La Detección Anticipada de Riesgos (DAR) puede considerarse como un *problema multi-objetivo* en el que el desafío es encontrar un balance adecuado entre dos aspectos diferentes y relacionados: 1) la *precisión* en la iden-

tificación de usuarios de riesgo y, 2) el *tiempo mínimo* que requiere la detección de un usuario de riesgo para ser *confiable*. El primer aspecto es generalmente abordado como un problema de clasificación típico y evaluado con métricas de clasificación estándar como *precisión* (en inglés *precision*), *alcance/cobertura* (en inglés *recall*) y F_1. El segundo implica una política para decidir *cuándo* la información de un usuario catalogado como de riesgo es *suficiente* para dar la alarma/alerta y suele ser evaluado penalizando al retraso en tomar esa decisión. De hecho, las métricas de evaluación temporal utilizadas en DAR como $ERDE_\theta$ y $F_{latency}$ combinan ambos aspectos de diferentes maneras.

La DAR ha sido abordada en el contexto del Laboratorio de Predicción Temprana de Riesgos en Internet (eRisk) el cual se ocupa de la exploración de nuevos modelos de detección temprana de riesgos y metodologías de evaluación con impacto directo en aspectos sociales y de la salud [8]. El laboratorio comenzó en 2017 abordando el problema de la detección temprana de depresión en usuarios de un foro en línea (Reddit) [9]. En 2018, la detección temprana de signos de anorexia se agregó como un nuevo desafío para el laboratorio, junto con una versión ampliada de la tarea del año anterior [10]. Los datos de prueba se organizaron en 10 fragmentos (*chunks*) y fueron proporcionados a cada equipo fragmento a fragemento. Los modelos de los participantes se evaluaron utilizando la métrica de evaluación ERDE introducida por Losada y otros [8] para considerar tanto la corrección de la clasificación como la demora incurrida por el sistema para tomar la decisión. En 2019, la tarea de detección temprana de depresión fué reemplazada por dos nuevos desafíos: la detección temprana de signos de autolesiones (en inglés *self-harm*) y medir la gravedad de los signos de depresión [11]. Para esa edición del laboratorio, se consideraron nuevas medidas de desempeño. Primero, la medida de rendimiento $F_{latency}$ propuesta por Sadeque y otros [15] se incorporó como medida complementaria al ERDE. Por otro lado, se agregaron métricas de evaluación basadas en *ranking* para ayudar a los profesionales a tomar sus decisiones en problemas de la vida real. Ese año también marcó el final del procesamiento de datos basado en fragmentos. A partir de ese año, se utilizó un enfoque publicación por publicación (*post-by-post*) para las distintas tareas, que se asemeja a un escenario de la vida real donde los usuarios escriben sus mensajes de a uno por vez. En 2020 se eliminó la tarea de detección temprana de signos de anorexia pero las demás tareas se mantuvieron [12]. Finalmente, en 2021, se introdujo la tarea de detección temprana de signos de juego patológico mostrándose a continuación una breve descripción de las dos tareas en las que participó nuestro grupo de investigación:

- **Tarea 1**: Detección anticipada de signos de juego patológico. Para esta tarea, el objetivo era detectar, tan pronto como sea posible, a los usuarios que fueran jugadores compulsivos o que tuvieran patrones tempranos de juego patológico. Los datos de la tarea consistían en una serie de escritos de usuarios de medios sociales recopilados en orden cronológico. No se proporcionaron datos de entrenamiento, por lo que cada equipo tuvo que construir su propio corpus para entrenar sus modelos.

- **Tarea 2**: Detección temprana de signos de autolesión. Para esta tarea, el objetivo era el mismo que con las ediciones de eRisk 2019 y 2020, es decir, procesar secuencialmente las evidencias y detectar rastros tempranos de autolesión tan pronto como sea posible. Ese año, los datos de entrenamiento fueron la combinación de los datos de entrenamiento y prueba de la edición 2020.

El desempeño en ambas tareas se evaluó utilizando medidas de clasificación estándar (precisión, cobertura, y F_1), medidas que penalizan el retraso en la respuesta (ERDE y $F_{latency}$), y métricas de evaluación basadas en rankings. Las valores de F_1 y $F_{latency}$ se calcularon con respecto a la clase positiva. Para

calcular estas medidas, para cada publicación de cada usuario, a los modelos participantes se les pidió que proporcionaran una decisión, que indicaba si el usuario estaba en riesgo (indicado con un uno) o no (indicado con un cero), y una puntuación, que representaba el nivel de riesgo del usuario (estimado de la evidencia vista hasta ese momento). En ese contexto, si un usuario fue clasificado como de riesgo, las decisiones posteriores no fueron consideradas.

2. Líneas de Investigación y Desarrollo

Nuestros trabajos en la DAR, se relacionan con trabajos previos que abordaban el problema de procesar datos en forma secuencial y clasificarlos lo antes posible [14]. En el área específica de DAR, los principales enfoques utilizados fueron:

- Enfoque *Temporal Variation of Terms* (TVT) [5]
- Enfoque *Flexible Temporal Variation of Terms* (FTVT) [6]
- Enfoque *Sequential-Incremental Classification* (SIC) [6]
- Enfoque *k-TVT* [4]
- Enfoque *SS3*[1, 2, 3]
- Enfoque *EarlyModel*[13]
- Enfoque *EARLIEST*[13, 7]

2.1. Resultados Obtenidos

En 2017, en el marco de la Conference and Labs of the Evaluation Forum (CLEF 2017), nuestro grupo participa en el eRisk 2017 (https://early.irlab.org/2017/index.html) con un método diseñado específicamente para este tipo de tarea (*TVT*) obteniendo el *mejor valor* en la medida $ERDE_{50}$[16, 5]. En 2018, se utilizan los enfoques *FTVT* y *SIC* obteniéndose los mejores valores de $ERDE_5$ tanto en *anorexia* como *depresión* y el valor más alto de precisión en *anorexia* **0.91** [6]. En los años 2019 y 2020 se participó en ambos casos con el enfoque SS3 [1, 2, 3] obteniéndose resultados del estado del arte en el área. En 2021 [13], y a diferencia de las participaciones anteriores en los eRisk Labs, el énfasis se puso en las políticas de alerta temprana que deciden si un usuario catalogado como de riesgo debe ser efectivamente reportado como tal. Allí, propusimos tres políticas diferentes de alerta temprana para la detección temprana del riesgo de juego patológico y detección temprana del riesgo de autolesiones. El primer enfoque utiliza modelos de clasificación estándar para identificar a los usuarios de riesgo y una política de alerta temprana simple (manual) basada en reglas. El segundo enfoque es un modelo de aprendizaje profundo entrenado de extremo a extremo (end-to-end) que aprende simultáneamente a identificar a los usuarios de riesgo y la política de alerta temprana a través de un enfoque de aprendizaje por refuerzo. Finalmente, el último enfoque consiste en un modelo simple e interpretable que identifica a los usuarios de riesgo, integrado con una política global de alerta temprana. Esa política, basada en el nivel de riesgo estimado (global) para todos los usuarios procesados, decide qué usuarios deben informarse como riesgosos. Con respecto a los resultados alcanzados, nuestros modelos obtuvieron el mejor rendimiento en términos de métricas de rendimiento basadas en decisiones (F_1, $ERDE_{50}$, $F_{latency}$) así como en términos de medidas de rendimiento basadas en ranking, para ambas tareas. Además, en términos de la medida de $F_{latency}$, el rendimiento obtenido en la primera tarea fue el doble que el segundo mejor equipo.

3. Formación de Recursos Humanos

Trabajos de tesis vinculados con las temáticas descritas previamente:

- 1 tesis de Maestría en ejecución con beca de postgrado de la UNSL.
- 3 tesis de Doctorado en ejecución con becas de CONICET.
- 1 tesis de Doctorado finalizada con beca de CONICET.

Referencias

[1] S. G. Burdisso, M. Errecalde, and M. Montes-y Gómez. A text classification framework for simple and effective early depression detection over social media streams. *Expert Systems with Applications*, 133:182 – 197, 2019.

[2] S. G. Burdisso, M. Errecalde, and M. Montes-y Gómez. UNSL at eRisk 2019: a unified approach for anorexia, self-harm and depression detection in social media. In *Working Notes of CLEF 2019*, Lugano, Switzerland, 2019. CEUR Workshop Proceedings.

[3] S. G. Burdisso, M. Errecalde, and M. Montes-y Gómez. τ-SS3: A text classifier with dynamic n-grams for early risk detection over text streams. *Pattern Recognition Letters*, 138:130 – 137, 2020.

[4] L. C. Cagnina, M. L. Errecalde, M. J. Garciarena Ucelay, D. G. Funez, and M. P. Villegas. k-tvt: a flexible and effective method for early depression detection. In *XXV Congreso Argentino de Ciencias de la Computacion. CACIC 2019. Libro de actas*, pages 547–556, 2019.

[5] M. L. Errecalde, M. P. Villegas, D. G. Funez, M. J. Garciarena Ucelay, and L. C. Cagnina. Temporal variation of terms as concept space for early risk prediction. In *Working Notes of the Conference and Labs of the Evaluation Forum - CEUR Workshop Proceedings*, volume 1866, 2017.

[6] D. G. Funez, M. J. Garciarena Ucelay, M. P. Villegas, S. G. Burdisso, L. C. Cagnina, M. Montes y Gomez, and M. L. Errecalde. Unsl's participation at erisk 2018 lab. In *Working Notes of the Conference and Labs of the Evaluation Forum - CEUR Workshop Proceedings*, volume 2125, 2018.

[7] T. Hartvigsen, C. Sen, X. Kong, and E. Rundensteiner. Adaptive-halting policy network for early classification. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 101–110, 2019.

[8] D. E. Losada and F. Crestani. A test collection for research on depression and language use. In *Proc. of Conference and Labs of the Evaluation Forum (CLEF 2016)*, pages 28–39, Evora, Portugal, 2016.

[9] D. E. Losada, F. Crestani, and J. Parapar. erisk 2017: Clef lab on early risk prediction on the internet: experimental foundations. In *International Conference of the Cross-Language Evaluation Forum for European Languages*, pages 346–360. Springer, 2017.

[10] D. E. Losada, F. Crestani, and J. Parapar. Overview of erisk: early risk prediction on the internet. In *International Conference of the Cross-Language Evaluation Forum for European Languages*, pages 343–361. Springer, 2018.

[11] D. E. Losada, F. Crestani, and J. Parapar. Overview of erisk 2019 early risk prediction on the internet. In *International Conference of the Cross-Language Evaluation Forum for European Languages*, pages 340–357. Springer, 2019.

[12] D. E. Losada, F. Crestani, and J. Parapar. Overview of erisk at clef 2020: Early risk prediction on the internet (extended overview). 2020.

[13] J. M. Loyola, S. G. Burdisso, H. Thompson, and M. L. Errecalde. Unsl at erisk2021: A comparison of three early alert policies for early risk detection. In *Working Notes of the Conference and Labs of the Evaluation Forum - CEUR Workshop Proceedings*, volume 2936, 2021.

[14] J. M. Loyola, M. L. Errecalde, H. J. Escalante, and M. M. Gomez. Learning when to classify for early text classification. In *Argentine Congress of Computer Science*, pages 24–34. Springer, 2017.

[15] F. Sadeque, D. Xu, and S. Bethard. Measuring the latency of depression detection in social media. In *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*, pages 495–503, 2018.

[16] M. P. Villegas, D. G. Funez, M. J. Garciarena Ucelay, L. C. Cagnina, and M. L. Errecalde. Lidic - unsl's participation at erisk 2017: Pilot task on early detection of depression. In *Working Notes of CLEF 2017 - CEUR Workshop Proceedings*, volume 1866, 2017.