Modelo para Analizar Mensajes y Detectar Actitudes Peligrosas a través de Análisis de Sentimientos con Algoritmos de Aprendizajes

Juan Calloni, Sergio Paez, Javier Saldarini, Juan Cuevas, Micaela Mulassano, Andrés Bianciotti, Eduardo Scarello, Leandro Banchio, Federico Degiovanni, Lucia Scharff

Grupo GARLAN / Secretaría de Ciencia y Tecnología
Facultad Regional San Francisco / Universidad Tecnológica Nacional
Av. de la Universidad 501, 03564-421147
facultad@sanfrancisco.utn.edu.ar

Resumen

Es de amplio conocimiento la utilización masiva de las distintas redes sociales. Éstas han cambiado los hábitos y características de la comunicación, tal como la facilidad de intercambio de información, la existencia de receptores globales y la accesibilidad de éstas tecnologías a todos los sectores sociales.

Las actitudes violentas y peligrosas en redes sociales constituyen un campo de estudio objeto de varias disciplinas. Detectar este tipo de actitudes de manera temprana, colaboran a la prevención de los efectos que éstas podrían causar.

Uno de los problemas que trata de resolver este proyecto, es la seguridad, en específico, las actitudes sospechosas o violentas, como violencia de género o bullying.

En definitiva, se trata de brindar un modelo que permita realizar análisis, extrayendo información de redes sociales de acceso público, para demostrar, con métodos de aprendizaje, tanto supervisados como no supervisados y con métodos de análisis de sentimientos y minería de opiniones, qué mensaje o grupo de mensajes se aproximan a ser peligrosos.

Palabras clave:

Modelo, Machine learning, Análisis de sentimientos, NoSQL, Minería de datos, Cloud computing.

Contexto

La línea de investigación presentada se encuentra en el marco de las áreas prioritarias para el desarrollo de las actividades de I+D que se formalizaron a través de la Resolución de Consejo Directivo N 353/2016 de la Universidad Tecnológica Nacional Facultad Regional San Francisco cuyas áreas prioritarias son:

- Gestión de procesos de negocios
- Ingeniería de software
- Gestión y tecnologías de las organizaciones
- Calidad de Software
- Seguridad de la información
- Bases de datos

El área principal es Ingeniería de Software relacionándose con el desarrollo en la Nube, métodos de aprendizajes, inteligencia artificial y bases de datos NoSQL. Éstas líneas se encuentran insertas en el Proyecto de Investigación y Desarrollo (PID) tutorado Nro EITUNCO0004317 “Modelo de un

Analizador de Mensajes en Redes sociales para la detección de usuarios con actitudes peligrosas o violentas a través de análisis de sentimientos con algoritmos de aprendizajes", el mismo se encuentra en desarrollo desde Abril de 2016. Dicho proyecto es ejecutado por el grupo de I+D GARLAN de la UTN Facultad Regional San Francisco.

El proyecto se encuentra homologado y financiado por la Secretaría de Ciencia y Tecnología de la Universidad Tecnológica Nacional, a través de la Disposición Nº 52/16 y se lo reconoce bajo el código TUN4317, el mismo está incluído en el Programa I&D + i de Electrónica, Informática y Comunicaciones de la Universidad Tecnológica Nacional.

Introducción

Este proyecto tiene como objetivos:

1. Determinar un modelo analizador de mensajes para la detección de actitudes peligrosas de usuarios a través de métodos de aprendizajes.
2. Desarrollar un aplicativo orientado a servicios (SaaS) accesible vía API RESTful que sirva para validar el modelo.
3. Implementar el prototipo con una muestra representativa de expresiones reales elegidas al azar para verificar el comportamiento de los algoritmos y parámetros del modelo.

En particular a través de estos objetivos las herramientas de inteligencia artificial como, Support Vector Machines (SVM)[1], que pertenece a un conjunto de algoritmos de aprendizaje supervisado que están propiamente relacionados con problemas de clasificación y regresión a partir de un conjunto de ejemplos de entrenamiento (de muestras), es posible etiquetar las clases y entrenar el sistema para construir un modelo que prediga la clase de una nueva muestra.

Por otro lado, K-means[2], es un método de agrupamiento y pertenece a un conjunto de algoritmos de aprendizaje no supervisado que tiene como objetivo la partición de un conjunto de 'n' observaciones en 'k' grupos en el que cada observación pertenece al grupo más cercano a la media. En el aprendizaje de máquina y la ciencia cognitiva, las redes neuronales artificiales (RNA) son una familia de modelos de aprendizaje estadísticos inspirados en las redes neuronales biológicas (los sistemas nervioso central de los animales, en particular, el cerebro) y pueden ser utilizadas como algoritmos de aprendizaje supervisados y no supervisados[3]. Con la ayuda de un experto en el dominio, se obtiene un conjunto de datos, con el cual entrenar los algoritmos de aprendizaje. Entonces, con una herramienta de análisis de minería de datos, como Weka, analizar qué métodos brindan mayor certeza. Una vez realizado, es posible sacar conclusiones, sobre qué algoritmo de aprendizaje utilizar; supervisado, no supervisado o varios de ellos. Él o los métodos seleccionados son los candidatos a implementar para realizar el modelo.

Los siguientes pasos describen el proceso a realizar para la extracción de la información que será usada para el análisis posterior.

Son utilizadas las API's de las redes sociales para obtener información, y poder realizar estudios con datos reales. En específico, en este trabajo se obtendrán mensajes de la red social Twitter, y dejando a futuros proyectos la implementación con otras API's de redes sociales y sistemas de participación. La información obtenida es ingresada a una base de datos NoSQL (orientada a documentos, para almacenar metadatos) y así analizarlo de forma local. Es usado este tipo de base de datos ya que el prototipo deberá funcionar en la nube y una de sus características es la alta escalabilidad[4]. Al final, para validar el modelo, el proyecto planea construir un aplicativo orientado a servicio (SaaS) como prototipo

utilizando un método heurístico probarlo y validar el modelo propuesto. Se trabaja con algoritmos de aprendizaje no supervisado de clustering para agrupar mensajes con características necesarias del dominio de aplicación, y a partir de éste crear un modelo basado en técnicas de aprendizaje supervisado, lo cual requiere una tarea de etiquetado y validación por un experto en el dominio. En el caso de los modelos no supervisados, estamos trabajando con el algoritmo K-means. Los resultados positivos del primer agrupamiento son clasificados manualmente.

Luego, con métodos de regresión logística y RNA es posible determinar a qué conjunto pertenece un nuevo elemento de prueba. Durante las primeras etapas del proyecto fue contemplada y planificada la posibilidad de analizar documentos públicos provenientes de las filtraciones de cables que publicó el portal Wikileaks, sin embargo; los análisis posteriores determinaron que la profundidad semántica de la temática expuesta escapaba a los alcances este proyecto. El Idioma elegido es el castellano utilizando algoritmos de aprendizaje enfocados para determinar el modelo que mejor resuelve la clasificación. Resumiendo, los pasos de la investigación son:

Obtención del conjunto de datos y agrupamiento de los mensajes

Generación del dataset supervisado

Implementación del algoritmo que clasificará nuevos elementos de prueba. La cantidad de iteraciones depende del número de algoritmos a investigar hasta encontrar el que mejor se adapte

Análisis de los resultados

Despliegue y evaluación del prototipo SaaS en un entorno de cloud computing.

El proyecto busca corroborar nuestra hipótesis: Es posible clasificar mensajes de forma automatizada de acuerdo a diversos intereses sociales, centrándonos en actitudes violentas y ciberbullying.

Líneas de Investigación, Desarrollo e Innovación

El desarrollo e investigación se basa en encontrar técnicas de aprendizaje de máquina que puedan aportar resultados predictivos en la materia de seguridad social a través de aplicaciones que un usuario pueda utilizar. Con esto, nos referimos a implementar algoritmos de aprendizaje que detecten mensajes en español y tengan consecuencias negativas en la vida social, creando herramientas útiles que sean implementadas utilizando tecnologías de última generación.

Resultados Obtenidos/Esperados

Dentro de los principales avances que hemos obtenido hasta el momento se encuentra el desarrollo de modelos para realizar un SaaS que permita detectar mensajes de texto. En un principio vinculados a actitudes sospechosas de usuarios y actualmente enfocándonos en mensajes que puedan relacionarse con el ciberbullying[5, 6, 7, 8,9].

Para complementar, fue creada una interfaz de comunicación con las redes sociales para obtener el conjunto de datos a utilizar en el desarrollo del sistema. En la experimentación realizamos las pruebas con la API Search de Twitter.

Para poder almacenar y consultar el conjunto de datos obtenidos ha sido creada una interfaz para guardar los mensajes a estudiar en una base de datos y ejecutar el proceso offline. Esta interfaz nos permite experimentar con distintas estrategias de almacenamiento reutilizando la mayor cantidad de código.

Para no tener que programar todos los algoritmos de aprendizaje de máquina y optimizar los tiempos de análisis y clasificación de los datos se creó una interfaz para conectarse con una API que provea los algoritmos de aprendizaje necesarios en la investigación. Actualmente se centró específicamente en la arquitectura del sistema y se está

trabajando en el análisis y clasificación del conjunto de datos.

La arquitectura del sistema se creó utilizando la tecnologías de contenedores Docker[10] para montar la aplicación. La cual está formada básicamente por los siguientes contenedores:

- Contenedor web que aloja la lógica del dominio desarrollada con el lenguaje de programación Java.
- Contenedor de base de datos para alojar el conjunto de datos extraído de las redes sociales, utilizando MongoDB como sistema de almacenamiento.
- Un sistema de contenedores que implementa sharding para almacenar el conjunto de datos 1-Gram de Google[11], utilizando MongoDB como sistema de almacenamiento.

Nuestros objetivos a corto plazo involucran:

- Implementar un modelo similar al planteado en el paper TweetNorm[12] para realizar un preprocesamiento y normalización de los mensajes en el conjunto de datos.
- Terminar de definir el conjunto de datos que utilizaremos como entrenamiento para los algoritmos de aprendizaje de máquina.
- Analizar distintos tipos de algoritmos de aprendizaje, K-means, SVM, RNA, para evaluar el conjunto de datos y obtener el algoritmo que mejor se adapte a nuestro dominio

El objetivo final es evaluar cómo funcionan los distintos módulos del prototipo en un ambiente de cloud computing.

Formación de Recursos Humanos

La formación de recursos humanos en el área de AI (Artificial Intelligence) y Machine Learning ha ganado campo en los últimos años en la ciencias de computación, además de que año a año los algoritmos y modelos se vuelven más complejos, por lo que tener recursos humanos enfocados en estas disciplinas es de vital importancia.

La investigación en la disciplina de NLP (Natural Language Processing), detección de imágenes, analizadores de video, etc., también ha tenido un crecimiento importante con las nuevas tecnologías de procesamiento de datos, por lo que conceptos de escalabilidad y desempeño también deben ser tenidos en cuenta a la hora de investigar, diseñar e implementar.

El proyecto plantea dar conocimientos sobre AI, Machine Learning, NLP, bases de datos NoSQL, así como también aplicar conceptos de arquitecturas en la nube para una escalabilidad horizontal y un buen desempeño.

Dentro del proyecto se contempla la formación de alumnos de las cátedras Inteligencia Artificial, Gestión de Datos, Diseño de Sistemas y Redes de Información en los aspectos referentes a métodos de aprendizaje supervisados o no supervisados, en bases de datos NoSQL, aspectos de cómo aplicar conceptos de arquitecturas en la nube y el desarrollo de aplicaciones SaaS.

Se pretende capacitar en estas tecnologías a docentes y alumnos avanzados de la carrera Ingeniería en Sistemas de Información mediante talleres que se dictarán a medida que avance el proyecto. Esto permitirá contar con personal de apoyo, captar nuevas ideas y opiniones para mejorar el proyecto.

Se llevará a cabo una serie de conferencias informativas en la Universidad Tecnológica Nacional Facultad Regional San Francisco para exponer el proyecto conceptualizado con las principales ventajas ofrecidas desde un marco teórico (previo al inicio, presentación oficial para la puesta en marcha y transición del proyecto). ■

Referencias

[1] Cortes, Corinna; Vapnik, Vladimir. Support-vector networks. Machine learning, 1995, vol. 20, no 3, p. 273-297.

[2] Kanungo, Tapas, et al. An efficient k-means clustering algorithm: Analysis and implementation. IEEE transactions on pattern analysis and machine intelligence, 2002, vol. 24, no 7, p.881-892.

[3] Russell, Stuart J and Norvig, Peter, 2004, Artificial intelligence. Englewood Cliffs, N.J. : Prentice Hall.

[4] Sadalage, Pramod J.; Fowler, Martin. NoSQL distilled: a brief guide to the emerging world of polyglot persistence. Pearson Education, 2012.

[5] Morales, J. and Arias Orduña, A. (2007). Psicología social. Madrid: McGraw-Hill Interamericana de España. p.457, p.459, p.489.

[6] Anderson, Janna; RAINIE, Lee. Millennials will benefit and suffer due to their hyperconnected lives. Washington DC, Pew Research Center, 2012.

[7] Hernández Prados, María Ángeles and Solano Fernández, Isabel María, 2007, Ciberbullying, un problema de acoso escolar. Revista Iberoamericana de Educación a Distancia. 2007.

[8] Cervantes Benavides, L. (2013). Una propuesta para identificar, clasificar y tipificar el Bullying (Acoso Escolar). Revista Iberoamericana para la Investigación y el Desarrollo Educativo.

[9] Smith, P. (1999). The nature of school bullying. London: Routledge.

[10] "Docker". Docker. N.p., 2016. Web. 13 Julio 2016.

[11] Google Ngram Viewer, 2016. http://storage.googleapis.com/books/ngrams/books/datasetsv2.html [online]

[12] Tweet Norm. http://www.lrec-conf.org/proceedings/lrec2014/pdf/442_Paper.pdf Febrero de 2016.