



Aplicación de técnicas avanzadas de aprendizaje automático para identificar emociones en textos

Carolina Cardoso, Lorena Talamé, Matías Amor, Agustina Monge
Universidad Católica de Salta

RESUMEN

En este proyecto se pretende aprender a clasificar textos cortos (opiniones) generados en la red social Twitter según el sentimiento que expresan, aplicando técnicas avanzadas de aprendizaje automático como redes neuronales. El proyecto se encuentra en desarrollo. En la primera etapa se exploró en la clasificación de textos usando una red neuronal LSTM y en la etapa actual, se está analizando algunas formas de representar los textos para crear un corpus de palabras embebidas que serán utilizadas en otros experimentos

LINEA DE INVESTIGACION

Este proyecto de investigación propone detectar sentimientos expresados en opiniones textuales emitidas en una red social. El proyecto se desarrolla en las siguientes etapas:

- Revisión de la literatura relevante al problema de minería de opiniones y sentimientos.
- Evaluación y comparación las técnicas de aprendizaje profundo para la clasificación de textos.
- Evaluación y comparación de otras técnicas de aprendizaje automático para clasificación de opiniones textuales.
- Desarrollo de un prototipo para la clasificación de opiniones.

Se utilizan los tweets capturados para el proyecto anterior y otros obtenidos durante el año 2020, totalizando más de 150000 tweets.

Generalmente los usuarios de redes sociales suelen usar emojis para destacar lo que se quiere expresar, como una forma de entonación de la voz o de expresión corporal. En el proyecto anterior, se demostró que mantener emojis y hashtags resulta relevante, por lo cual, deben ser transformados a texto.

Se usa Python, uno de los lenguajes de programación más aceptados por la comunidad científica. Es potente y se caracteriza por su simplicidad, distribución open source y la posibilidad de integración con múltiples librerías.

CONTEXTO

Se continúa la línea de investigación en minería de textos. En los proyectos anteriores se exploraron diversos problemas relacionados a minería de textos: clasificación de documentos, reconocimiento de entidades con nombres, generación automática de resúmenes, búsquedas de respuestas y análisis de sentimientos.



En el último proyecto se identificó cuatro emociones en textos extraídos de la red social Twitter con dos enfoques: aplicando aprendizaje automático supervisado y semi supervisado y a partir de lexicones de dimensiones emocionales.

RESULTADOS

Se iniciaron los experimentos con redes neuronales LSTM (Long Short-Term Memory) para el aprendizaje de clasificación de textos. La característica principal de estas redes es que la información puede persistir introduciendo bucles en el diagrama de la red, para decidir cuál será el siguiente. Se utilizó el dataset clasificado y creado en el proyecto anterior, un conjunto de tweets en lenguaje español y generado en nuestro país.

Los próximos objetivos a alcanzar son:

- Crear un corpus de palabras embebidas a partir de la colección de tweets disponibles
- Experimentar el aprendizaje con el corpus de palabras embebidas y con otros disponibles para investigaciones.
- Seleccionar los algoritmos que mejor clasifiquen las opiniones.

RECURSOS HUMANOS

El equipo de trabajo está integrado por tres docentes y una reciente egresada de la carrera de Ingeniería en Informática. Se espera incorporar al proyecto alumnos interesados en la temática.