

Minería de sentimientos con detección de sarcasmo en un ambiente *streaming* de Big Data

Sentiment mining with sarcasm detection in a Big Data streaming environment

RAFAEL JIMÉNEZ CASTRO^{a*}, ROGELIO FLORENCIA JUÁREZ^a, VICENTE GARCÍA^a

^aDoctorado en Ciencias de la Ingeniería Avanzada, Departamento de Ingeniería Eléctrica y Computación, Universidad Autónoma de Ciudad Juárez, México

*Autor de correspondencia. Correo electrónico: jimenez.cstr@gmail.com

No. de resumen 5CP23-21	Formato Cartel
Evento 5.º Coloquio de Posgrados del IIT	Presentador Rafael Jiménez Castro
Tema Cómputo aplicado	Estatus Estudio en curso
Fecha de la presentación Mayo 24, 2023	

Resumen

El objetivo de la investigación es desarrollar modelos de aprendizaje automático en una arquitectura Big Data que integren información contextual para la detección en tiempo real del sarcasmo en tuits. Las características contextuales del tuit son difíciles de obtener, por lo cual se requiere de una forma novedosa de extracción del contexto de los tuits. La solución propuesta consiste en contextualizar las respuestas a los tuits, a través del análisis y detección de los temas/entidad contenidos en el tuit original y el tuit de respuesta. Para suplementar el contexto de la conversación, se implementará un análisis de los *hashtags* y emoticones utilizados en el hilo de la conversación con el fin de establecer un contexto de los autores de ambos tuits. La delimitación es para respuestas a tuits expuestas en Twitter y que contengan un mensaje de texto corto en español. La limitación de la cantidad 450 tuits que se pueden procesar en un intervalo de 15 minutos y el tamaño máximo de 280 caracteres por tuit son debido a restricciones impuestas por Twitter. La detección del sarcasmo en redes sociales es un tema importante, ya que implica que se podrá posiblemente indicar a los usuarios la presencia de sarcasmo en comentarios de forma automática para minimizar las reacciones negativas que se puedan generar en hilos de discusión y, además, su valor está en la mejora del rendimiento en el análisis de sentimientos donde la correcta detección puede mejorar la clasificación del sentimiento por un aproximado de 5.49 %.

Palabras clave: sarcasmo; minería de sentimientos; minería de redes sociales.

Abstract

The objective of the research is to develop machine learning models in a Big Data architecture that integrate contextual information for the real-time detection of sarcasm in tweets. The contextual characteristics of the tweet are difficult to obtain, which is why a novel way of extracting the context of the tweets is required. The proposed solution consists of contextualizing the responses to the tweets, through the analysis and detection of the topics/entities contained in the original tweet and the response tweet. To supplement the context of the conversation; An analysis of the hashtags and emoticons used in the thread of the conversation will be implemented to establish a context of the authors of both tweets. The delimitation is for responses to tweets posted on Twitter that contain a short text message in Spanish. The limitation of the number of 450 tweets that can be processed in a 15-minute interval and the maximum size of 280 characters per tweet are due to restrictions imposed by Twitter. The detection of sarcasm in social networks is an important issue since it implies that it will be possible to automatically indicate to users the presence of sarcasm in comments to minimize the negative



reactions that can be generated in discussion threads, in addition its value is in performance improvement in sentiment analysis where correct detection can improve sentiment rating by approximately 5.49%.

Keywords: sarcasm; sentiment mining; social media mining.

Entidad legal responsable del estudio

Universidad Autónoma de Ciudad Juárez.

Financiamiento

CONACYT - Beca Nacional Tradicional 2023-1 clave 2023-000002-01NACF-02639.

Conflictos de interés

Los autores declaran que no existe conflicto de intereses.