

Practica 13

En esta práctica se usará varias herramientas para análisis de sentimientos

PRIMERA PARTE

En la primera parte la meta es recopilar y analizar tweets utilizando el "Análisis de texto de AYLIEN" para RapidMiner.

Para ejecutar este proceso es necesario tener una cuenta en AYLIEN. Para crear una cuenta por favor acceda al siguiente enlace

<https://developer.aylien.com/signup?source=rapidminer>

Las tareas a ejecutar son las siguientes

- Recoger tweets usando el operador de búsqueda de Twitter
- Analizar Sentimiento utilizando el Operador Analizar Sentimiento
- Asignar los tweets a diferentes categorías usando el operador de categorizar
- Visualizar los resultados y hacer los datos consumibles y comprensibles

Paso 1. Recopilación de tweets

Usando el operador Search Twitter cree un proceso de búsqueda como lo haría con la API de búsqueda de Twitter. Usando los parámetros ejecute un proceso de búsqueda, por ejemplo buscando tweets que contengan la palabra clave "Samsung". Elimine algunas búsquedas quitando retweets (-rt) y enlaces (-http). Para facilitar el proceso restrinja el número de tweets y coloque el idioma en el que desea recibir las respuestas.

Recuerde además que le parámetro `Result type` permite devolver solo tweets recientes o populares.

Paso 2. Analizando los tweets

Con la colección de tweets almacenados en un `ExampleSet` el objetivo ahora es verificar el sentimiento de cada tweet, es decir, si son Positivos, Negativos o Neutros.

Una opción para ejecutar este proceso es usar el operador `Analyze Sentiment Operator`. Los parámetros necesarios son seleccionar como atributo de entrada, la opción `texto`. Con esta configuración se tiene un proceso de análisis de opinión de Twitter relativamente simple que recopila tweets sobre el tema especificado y los analiza para determinar la polaridad (es decir, positiva, neutra o negativa) y la subjetividad (es decir, subjetiva u objetiva) de cada tweet.

Si se ejecuta el proceso, entonces podemos visualizar el conjunto de datos que contiene los resultados no solo con los tweets que se extrajeron, sino también su polaridad y subjetividad correspondientes, así como un puntaje de confianza para ambos:

SEGUNDA PARTE

En la segunda parte se busca analizar el sentimiento en las reseñas en línea de la película Star Wars usando el método basado en léxico.

En este taller se describe dos técnicas de análisis de sentimientos disponibles en RapidMiner:

- basado en léxico y
- basado en modelos, que utiliza aprendizaje automático.

Técnica basada en léxico

Para ejecutar el proceso de análisis de opinión a los Tweets relacionados con la película de Star Wars se recomienda ejecutar una búsqueda en la red social twitter

Averigüe el operador dentro de rapidminer que permite ejecutar búsquedas. Primero se debe configurar una conexión a Twitter. Busque los 1000 tweets más recientes o populares para el hashtag de consulta "#starwars".

Luego de probar el funcionamiento de este operador seleccione el atributo "Text", que incluye el contenido de los tweets recuperados (es decir, ignoramos el resto de los metadatos).

Procese cada uno de los documentos (tweets) usando alguna representación como TF-IDF para la creación de un espacio vectorial. Previo a ello se recomienda ejecutar algunas tareas de preprocesamiento como:

Tokenización para extraer las palabras individuales de cada tweet y, así, representar un tweet con una "bolsa de palabras". También se debería filtrar las palabras vacías (stop words). Se pide investigar la forma de extraer la raíz disponible de una palabra usando la extensión Wordnet, que encuentra el lema de cada palabra. Finalmente, se solicita aplicar el operador "Extract Sentiment" (nuevamente desde la extensión de Wordnet). Se recomienda usar todas las entidades disponibles de voz parcial, como sustantivos, verbos, adjetivos y adverbios.

Como tarea adicional de esta práctica se pide que la salida del proceso se pueda representar gráficamente en RapidMiner con un histograma de los valores de sentimiento calculados (atributo "sentimiento"). Un ejemplo de resultado que se espera se muestra a continuación. El histograma ofrece ideas interesantes; por ejemplo, podemos detectar en la parte derecha de la distribución la existencia de un subgrupo de usuarios de Twitter que expresa un sentimiento positivo relativamente alto, que probablemente serán fieles seguidores de la serie Star Wars.

