



Clasificación de Multi Etiqueta de Documentos usando Embedding Contextual de Palabras y Word Storm

Kelvin Paul Pucho Zevallos

Orientador: Prof. Yuber Elmer Velazco Paredes

**UNSA - Universidad Nacional de San Agustín de Arequipa
Mayo de 2023**

Índice general

1. Introducción	3
1.1. Problema	3
1.2. Objetivos	3
1.2.1. Objetivo General	3
1.2.2. Objetivos Específicos	4

Capítulo 1

Introducción

1.1. Problema

El proceso de clasificación manual de documentos de diferentes áreas es una tarea laboriosa y propensa a errores, especialmente cuando se trata de grandes volúmenes de documentos. Además, la gran variedad de subdisciplinas y temas dentro de un tema en específico por ejemplo Ciencia de la Computación dificulta aún más la tarea de identificar la temática principal de cada paper. Además la clasificación de documentos en múltiples etiquetas implica asignar múltiples etiquetas o categorías a un documento, lo que añade complejidad al proceso de clasificación. Los enfoques tradicionales basados en características estáticas pueden no capturar adecuadamente la semántica contextual y las relaciones entre palabras en un texto. Por lo tanto, se necesita una técnica más avanzada que pueda capturar la información contextual de las palabras y aprovechar las relaciones semánticas entre ellas para mejorar la precisión de la clasificación de documentos en múltiples etiquetas.

1.2. Objetivos

1.2.1. Objetivo General

Evaluar y comparar la efectividad del embedding contextual de palabras y Word Storm en la clasificación de documentos en múltiples etiquetas, mejorando el rendimiento de la clasificación de documentos en términos de precisión, cobertura y eficiencia utilizando técnicas avanzadas de procesamiento de lenguaje natural.

1.2.2. Objetivos Específicos

- Investigar y seleccionar modelos de embedding contextual de palabras adecuados para la clasificación de documentos en múltiples etiquetas.
- Explorar y desarrollar una representación de Word Storm para capturar y aprovechar las relaciones semánticas entre las etiquetas utilizadas en la clasificación de documentos.
- Validar experimentalmente los enfoques propuestos utilizando conjuntos de datos de referencia y casos de estudio relevantes en el dominio de interés.