

Objetivos

Dadas las autopsias verbales (sólo el texto) conseguir el modelo que mayor precisión nos preste.

RQ1: Cual de los dos embeddings nos brinda mejor precisión/rendimiento?

RQ2: Pueden dos tipos de embeddings funcionar bien entrelazados?

Tarea y Datos

Debemos identificar la causa de muerte de entre 12 agrupaciones diferentes dada la autopsia verbal.

Contamos con 5209 autopsias verbales de una calidad aceptable ya que cada una cuenta con alrededor de 50 palabras en su promedio.

Cada clase representa un grupo de diferentes causas de muerte y su correlación con la autopsia verbal se basa en palabras clave que indiquen síntomas comunes, por ejemplo, la palabra “tos” si la clase es Enfermedades relacionadas con el sistema respiratorio.

Representación textual

Para el pre-proceso, hemos limpiado caracteres que aportan ruido y eliminado las stop-words
Para representación de los embeddings, hemos usado FlagEmbedding y SentenceEmbedding de HuggingFace

Model1

Para el primer modelo hemos usado FlagEmbeddings y BGE Reranker
Librerías: sentence_transformer y FlagEmbedding

Model2

El segundo modelo calcula la similitud BGE entre los SentenceEmbeddings de las clases y el FlagEmbedding de la frase a evaluar

Model4

El cuarto y último modelo usa los FlagEmbeddings y el reranker BGE de similitud

LowerThreshold

El punto mas bajo de nuestro clasificador sería uno aleatorio, entonces 0,0833 para 12 clases

Model3

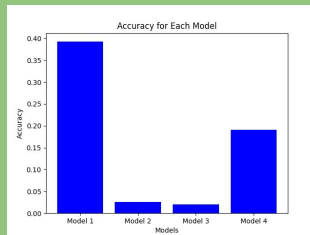
El tercer modelo calcula la similitud BGE entre los FlagEmbeddings de clase y SentenceEmbedding de la frase

Conclusiones

En definitiva, cruzar embeddings diferentes de misma dimensión no es conveniente en absoluto, y en nuestro caso los FlagEmbeddings valen la pena a costo del tiempo de procesamiento. Haber utilizado modelos complejos, nos ha logrado una precisión complaciente y competitiva. En contraparte, ha mermado el tiempo de desarrollo y reducido de sobremanera el ámbito de experimentación y análisis de resultados. Dicho ésto, queda claro que no hemos sido capaces de gestionar de manera correcta el tiempo del que disponíamos, y queda como objeto de mejora realizar un análisis de resultados más visual y profundo.

Referencias: [BAAI/bge-reranker-base · Hugging Face](#)
[distilroberta-base · Hugging Face](#)

Resultados experimentales



El largo tiempo de procesamiento de nuestro modelo nos ha impedido realizar un análisis exhaustivo de los resultados pero hemos comprobado que el modelo sufre una gran pérdida de precisión al entrecruzar embeddings y que es bastante consistente a pesar de utilizar un porcentaje reducido de datos.

Tras probar 4 combinaciones diferentes 1:FlagEmbeddings y BGE, 2 FlagEmbeddings + SentenceEmbeddings 3. SentenceEmbedding + FlagEmbeddings y BGE, 4. SentenceEmbeddings y BGE, vemos que el modelo 1 presta una mayor confianza a costo de rendimiento que es hasta 6 veces peor que en SentenceEmbedding.