

CC66Q: Diplomado Inteligencia Artificial

NLP - Evaluación 2 v1.0

Profs. Mauricio Cerda, Fabián Villena

7 de diciembre 2021

En esta actividad se buscará crear un *chatbot* en base a la identificación de preguntas tipo. Se busca que Ud. entregue un jupiter notebook con sus resultados, o en su defecto un pdf.

La siguiente actividad se puede realizar de manera individual o grupal. Los criterios de evaluación son 20 % legibilidad de su entrega (texto, código, o ambos), 20 % resolver el problema solicitado, 60 % utilizar técnicas estudiadas en el curso.

1. Creacion y limpieza de la base de datos (2 pts)

Se trabajará con el idioma castellano, y se creará primero una base de datos de tipos de mensajes.

- Elija una organización que cuenta con una lista de preguntas y respuestas. Por ejemplo https://www.sii.cl/preguntas_frecuentes/preguntasfrecuentes.tm
- Genere una tabla (tablaQA.xls) con las preguntas y respuestas de la organización seleccionada.
- Genere una tabla (tiposmensajes.xls) con ejemplos de las siguientes clases de mensaje: “saludo”, “despedida”, “nombre”, “informacion”. Por ejemplo, “buenos dias robot” sería un mensaje de clase “saludo”. Para la clase “información” utilice las preguntas de tablaQA.xls.
- Genere respuestas predeterminadas (respuestasDefecto.xls) para los mensajes de tipo “saludo”, “despedida”, “nombre”, (más de una). Por ejemplo para el tipo “saludo” podría generar “Buenos dias humano”, “hola hola!”, “Hola en este hermoso dia!”.

- Describa en términos generales las tablas que construyó.

2. Análisis de distancia (2 pts)

Para el primer *chatbot* utilizará una métrica de su elección para responder a cada pregunta/texto del usuario.

- Utilice algún embeddings utilizado en el curso para codificar el texto de entrada.
- Con el texto de entrada codificado y usando embeddings proponga una manera de identificar la clase del texto de entrada.
- Si el texto de entrada es del tipo “información”, busque ahora la pregunta mas similar y retorne la respuesta asociada.
- Reporte el resultado con textos de prueba.

3. Análisis transformer (2 pts)

Para el segundo *chatbot* utilice un transformer (por ejemplo BERT).

- Utilice el modelo transformer para clasificar el texto de entrada, y para extraer la respuesta de la tabla de preguntas y respuestas cuando el mensaje sea del tipo “información”.
- Reporte el tipo de red, y las métricas de entrenamiento usadas. Si requiere generar más ejemplos puede generar variaciones de sus tablas usando el tutorial del www.github.com/makcedward/nlpaug/blob/master/example/textual_augmenter.ipynb
- Reporte el resultado con los textos de prueba.