

TAREA PROGRAMADA 3
DOCUMENTACIÓN

1) Introducción

La tarea consiste en aplicar y evaluar dos clasificadores (Rocchio y bayesianos ingenuos) para una colección dada, la cuál se extrajo en por medio de dos archivos, los cuales contienen vectores de documentos extraídos de la colección de noticias Reuters21578. Dichos vectores se dan por medio de archivos tipo CSV con un formato especificado. El objetivo es leer de estos archivos los documentos correspondientes y aplicar los algoritmos de clasificación por medio del programa a crear. Sumado a esto, se realiza manualmente la evaluación de los resultados obtenidos por medio de las métricas de precisión, recall, acierto y error.

2) Tabla de estado final del proyecto

ETAPA	% DE COMPLET.	COMENTARIO O ACLARACIÓN
Lectura y conversión		
Puede leer y convertir datos de test_set	100	
Puede leer y convertir datos de train_set	100	
CLASIFICADOR ROCCHIO		
Calcula centroides	100	
Calcula similitudes centroides-vectores para los tres casos	100	
Produce un escalafón para cada documento del test_set para cada uno de los tres casos	100	
Asigna clases a los documentos del test_set para cada uno de los tres casos	100	
CLASIFICADOR BAYESIANO INGENUO		
Calcula los valores p_{ij} y q_{ij} para los términos	100	
Calcula similitudes para los documentos del test_set	100	
Produce un escalafón para cada documento del test_set	100	
Asigna clases a los documentos del test_set	100	
Evaluación (puede ser manual)		
Elabora la tabla de contingencia para cada clase	100	
Calcula las métricas: acierto, error, precisión y exhaustividad	100	
EVALUACIÓN (puede ser manual)		
Elabora la tabla de contingencia para cada clase	100	
Calcula las métricas: acierto, error, precisión y exhaustividad	100	

3) Comentarios finales

En general, se da por exitosa la implementación de los dos clasificadores, así como la lectura de las colecciones de entrenamiento y prueba brindadas como material de pruebas para estos. No se encontraron mayores limitaciones respecto a la funcionalidad del programa, y se considera que este realiza lo esperado de manera razonablemente eficiente. Por otro lado, en cuanto a la evaluación

manual realizada, se obtuvo cabalmente los valores de precisión, recall, acierto y error para las tres corridas de Rocchio y la corrida de bayesianos ingenuos efectuadas. Resulta de particular interés el hecho de que las corridas de Rocchio tuvieran métricas mucho mejores si se le compara con el clasificador de bayesianos ingenuos.