

Escuela de Ingeniería en Computación

Electiva
Recuperación de la Información Textual

**Proyecto programado III
Clasificadores**

Profesor:

Jose Enrique Araya Monge

Estudiantes:

Natalia Vargas Reyes
Javier Rivera Madrigal

Fecha de entrega:

18/1/2021

II Semestre 2021

Introducción

La tarea consiste en aplicar y evaluar los clasificadores Rocchio y bayesianos ingenuos, donde se utilizan dos vectores, uno de entrenamiento y uno de prueba, que cada línea del vector es un documento con información importante el identificador, la clase a la que pertenece el documento, el total de términos en el mismo y estos términos que aparecen con su respectivo peso.

En el caso del clasificador Rocchio se van a usar tres pares de parámetros diferentes, variando el β y γ . Al finalizar cada clasificador dando como resultado el documento que pertenece al vector de prueba con sus clases(real y asignada por el clasificador) se evaluarán los resultados usando métricas tales como: precisión, recall, acierto y error.

Estado 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%	
Asignar clases a los documentos del test_set para cada uno de los tres casos	100%	
CLASIFICADOR BAYESIANO INGENUO		
Calcula los valores p_{ip} y q_{ip} 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)	100%	
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%	

Evaluación clasificador rocchio

Evaluación

----parámetros-----

a = 0,75 y b = 0,25

a = 0,85 y b = 0,15

coffee

a = 4

f = 5

c = 4

e = 26

b = 0

d = 1

precisión: 0.8

recall: 1

acierto: 0.967

error: 0.03

money-fx

a = 5

f = 6

c = 5

e = 25

b = 0

d = 1

precisión: 0.833

recall: 1

acierto: 0.96

error: 0.03

grain

a = 11

f = 11

c = 11

e = 20

b = 0

d = 0

precisión: 1

recall: 1

acierto: 1

error: 0

crude

a = 5

f = 5

c = 5

e = 26

b = 0

d = 0

precisión: 1

recall: 1

acierto: 1

error: 0

trade

a = 4

f = 4

c = 6

e = 25

b = 2

d = 0

precisión: 1

recall: 0.67

acierto: 0.935

error: 0.06

----parámetros-----

a = 0,95 y b = 0,05

coffee

a = 4

f = 6

c = 4

e = 25

b = 0

d = 2

precisión: 0.66

recall: 1

acierto: 0.935

error: 0.064

money-fx

a = 5

f = 6
c = 5
e = 25
b = 0
d = 1

precisión: 0.833
recall: 1
acierto: 0.96
error: 0.03

grain
a = 11
f = 11
c = 11
e = 20
b = 0
d = 0

precisión: 1
recall: 1
acierto: 1
error: 0

crude
a = 5
f = 5
c = 5
e = 26
b = 0
d = 0

precisión: 1
recall: 1
acierto: 1
error: 0

trade
a = 3
f = 3
c = 6
e = 25
b = 3
d = 0

precisión: 1
recall: 0.5

acierto: 0.9032

error: 0.0967

Evaluación clasificador bayesianos ingenuos

Para ver la evaluación manual del clasificador bayesianos, se debe ingresar en el siguiente link para acceder a la hoja de cálculo donde se muestran las tablas respectivas de cada clase con aciertos, error, recall y precisión.

https://docs.google.com/spreadsheets/d/1OUQZi5_3Ksd-w3hITqjPEYNdG7e9q_ZngJjz_D3zh1s/edit?usp=sharing

Comentarios finales

Para el estado final del programa, por la parte de lectura y conversión quedó excelente, ya que obtiene los datos de forma exitosa. Los clasificadores cumplen su función y devuelve el escalafón con las clases asignadas por el mismo.

Además para cada clasificador se llegó a la conclusión de que los valores estimados son frecuentemente acertados, lo que podría indicar que los cálculos están bien hechos.

En general, se lograron realizar todos los puntos solicitados, con un funcionamiento adecuado.