

Práctica 2

Análisis de las reseñas de Tripadvisor

INTELIGENCIA ARTIFICIAL EN LAS ORGANIZACIONES

GRUPO 83-1

Miguel Gutiérrez Pérez
100383537@alumnos.uc3m.es

Mario Lozano Cortés
100383511@alumnos.uc3m.es

Alba Reinders Sánchez
100383444@alumnos.uc3m.es

Alejandro Valverde Mahou
100383383@alumnos.uc3m.es

GitHub: *InteligenciaArtificialOrganizaciones*

31 de octubre de 2020

Índice

1. Introducción	2
2. Parte 1: Clasificación	2
2.1. Análisis y preprocesado de datos	2
2.2. Experimentación	2
2.2.1. Experimentación básica	2
2.2.2. Experimentación avanzada	2
2.3. Comentario de los resultados obtenidos	2
3. Parte 2: Clustering	2
3.1. Experimentación	2
3.2. Mejor resultado	2
4. Conclusiones	3
5. Referencias	4
6. Anexos	5

1. Introducción

La siguiente sección incluye

2. Parte 1: Clasificación

2.1. Análisis y preprocesado de datos

Explicación de los pasos previos de lo que vimos en clase y división en cat

2.2. Experimentación

Generalidades de todos los experimentos. hacer hincapié en lo de las stopwords. Caso base

2.2.1. Experimentación básica

Tablas de cada uno de lo que hicimos esa tarde

2.2.2. Experimentación avanzada

Combinación de los mejores resultados básicos

2.3. Comentario de los resultados obtenidos

Mucho text

3. Parte 2: Clustering

En esta segunda parte se realiza un problema de *clustering*. Para ello se parte de los datos obtenidos en la primera parte y se les aplica el algoritmo ***K-Medias*** modificando sus parámetros tal y como se explica en la sección de *Experimentación*.

El proceso que se sigue es el siguiente:

- Cargar el archivo *.arff* con los datos generado en la parte anterior
- Generar modelos a partir de estos datos con *K-Medias* y compararlos
- Analizar el mejor modelo
- Ejecutar algoritmos de generación de reglas y árboles de decisión con el mejor modelo

3.1. Experimentación

Los experimentos que se llevan a cabo son los que se muestran en la en la Figura X. Los parámetros que se modifican son la *seed* (10, 20 y 30), el *números de clústers* (2, 3, 4, 5 y 6) y el *tipo de distancia* (Euclidea y Manhattan).

3.2. Mejor resultado

10 modelos de K Medias variando n Clústeres, semillas y funciones de distancia
Analizar al menos 1 (en detalle)
clasificación sobre el modelo y sacar el árbol

ID Experimento	Error	Seed	Número de clusters	Tipo de distancia
0	25436,31171	10	5	Euclidean
1	25382,55604	20	5	Euclidean
2	25218,09711	30	5	Euclidean
3	25703,79957	10	2	Euclidean
4	25704,73815	20	2	Euclidean
5	25704,73815	30	2	Euclidean
6	25545,30221	10	3	Euclidean
7	25684,13735	20	3	Euclidean
8	25643,51841	30	3	Euclidean
9	56512,35917	10	5	Manhattan
10	56666,77425	20	5	Manhattan
11	56680,27896	30	5	Manhattan
12	25509,57983	10	4	Euclidean
13	25670,38789	20	4	Euclidean
14	25449,76246	30	4	Euclidean
15	25250,01674	10	6	Euclidean
16	25342,88323	20	6	Euclidean
17	25160,2531	30	6	Euclidean

Tabla de experimentos con *K-Medias*

4. Conclusiones

5. Referencias

1. Introduction to Neurons in Neural Networks. Medium. Consultado en Octubre 2020. Url: <https://medium.com/artificial-neural-networks>

6. Anexos

1. Perceptron Multicapa usando '*K Fold*'
perceptron_kfold.py
2. Perceptron Multicapa usando '*split percentage*'
perceptron_split.py
3. Programa para realizar la predicción de los modelos
predict.py
4. Tabla de resultados de los experimentos de la primera parte
valores_reales_vs_predicciones_ℰ_errores_absolutos_parte1.xlsx
5. Tabla de resultados de los experimentos de la segunda parte
valores_reales_vs_predicciones_ℰ_errores_absolutos_parte2.xlsx