Minería de datos Práctica 1:Clustering knn-means

Jose Ignacio Sánchez
Josu Rodríguez

30 de septiembre de 2014

ÍNDICE DE CONTENIDO

ÍNDICE DE TABLAS

ÍNDICE DE FIGURAS

1. Introducción

El objetivo principal de esta práctica es obtener la capacidad de formular un algoritmo de aprendizaje automático de clasificación ***No-Supervisada***. Por otra parte, se trabajarán la capacidad de sintetizar uns técnica de aprendizaje automático no-supervisado, conocer su coste computacional así como sus limitaciones de representación y de inteligibilidad

2. Recursos

- PC con aplicación Weka.
- Bibliografía.
- Librerías de Weka.
- Manual de Weka.
- Guía de la práctica.
- Ficheros para los datos de la práctica: food.arff, colon.arff.
- Otros ficheros que no están en formato *.arff*:
  - En formato *.txt*: ClusterData.atributos.txt (este fichero si tiene la clase asociada para evaluar la calidad del *clustering* en ClusterData.clase.txt).
  - E formato *.csv* bank-data.csvclustering

3. Clasificación NO-supervisada o *Clustering*

(*Definición) 3.1 Se considera* ***clasificación no-supervisada*** *cuando el conjunto de entrenamiento no están las instancias etiquetadas con el valor de la clase. Es un experimento exploratorio, que trata de agrupar las instancias en grupos definidos por similitud entre las características de las instancias que pertenecen al mismo grupo y disimilitud entre las que pertenecen a grupos distintos. Tecnicamente estos grupos son llamados Clusters.*

3.1. Clustering *k-means*

3.1.1. Algoritmo en pseudocódigo

4. Diseño

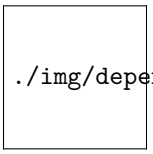

Figura 1: Dependencias del sistema[1].

5. Implementación

5.0.2. Problemas encontrados

5.0.3. Soluciones adoptadas

6. Validación del *software*

6.1. Diseño del banco de pruebas

7. Análisis de resultados

7.1. Modificando inicializaciones

7.2. Modificando distancia Minkowski

7.3. Criterios de convergencia

7.3.1. Número fijo de iteraciones

7.3.2. Disimilitud entre *codebooks*

7.4. Distintas métricas

7.4.1. Manhattan

7.4.2. Euclídea

7.4.3. Minkowski

8. Clasificación supervisada respecto de

9. Conclusiones

- Breve descripción de las motivaciones para llevar a cabo técnicas de clustering.
- Conclusiones a la vista de los resultados más relevantes.
- Conclusiones generales.(Análisis de fortalezas del sw y reflexiones sobre la tarea.
- Análisis de puntos débiles y propuestas de mejoras.

10. Valoración subjetiva

1. ¿Has alcanzado los objetivos que se plantean?
2. ¿Te ha resultado de utilidad la tarea planteada?
3. ¿Qué dificultades has encontrado?Valora el grado de dificultad de la tarea.
4. ¿Cuánto tiempo has trabajado en esta tarea? Desglosado:

Coste temporal	
Diseño de software	
Implementación de software	
Tiempo trabajando con Weka	
Búsqueda bibliográfica	
Informe	1

5. Sugerencias para mejorar la tarea. Sugerencias para que se consiga despertar mayor interés y motivación en los alumnos.

6. Críticas(constructivas).