

Aprendizaje Automático

GRADO EN INGENIERÍA INFORMÁTICA

Tutorial 3

Curso 2020/2021

Jorge Rodríguez Fraile, 100405951, Grupo 83, 100405951@alumnos.uc3m.es
Carlos Rubio Olivares, 100405834, Grupo 83, 100405834@alumnos.uc3m.es

Índice

Introducción	3
Ejercicio 1: KnowledgeFlow	3
Punto 1.11	3
Punto 1.13	3
Ejercicio 2: Experimenter	3
Punto 2.3.5	4
Punto 2.3.6	4
Punto 2.3.7	4
Punto 2.3.9	4
Conclusiones y dificultades encontradas	4

Introducción

En este documento se describirán diversos ejercicios realizados por Weka, el primero de ellos tendrá relación con el uso de la herramienta KnowledgeFlow. En el segundo ejercicio utilizaremos el código del primer tutorial para generar un fichero .arff y poder estudiar diversos resultados mediante la herramienta Experimenter.

Ejercicio 1: KnowledgeFlow

Punto 1.11

El TextViewer asociado al nodo J48 muestra la ejecución del árbol con todas sus instancias y número de nodos terminales, además del tamaño del árbol.

Por otro lado, el otro TextViewer muestra los resultados de la predicción, incluyendo accuracy y la matriz de confusión. Básicamente muestra los resultados con los que trabajábamos en el tutorial anterior.

El porcentaje de instancias predichas correctamente es de 86.2105 %.

Punto 1.13

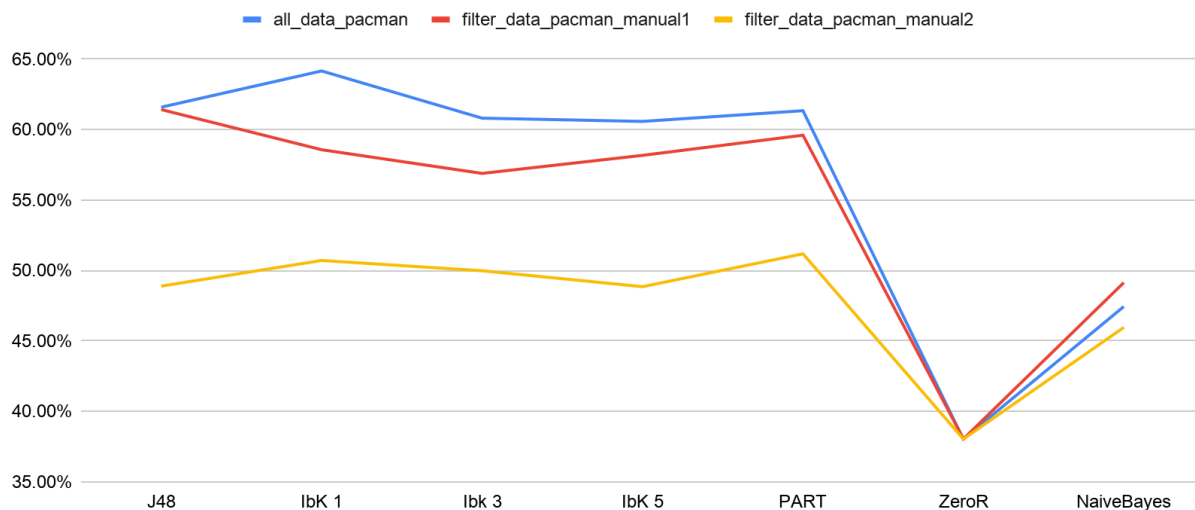
La utilidad de esta funcionalidad es poder elegir cómo se van a tratar las instancias de una manera secuencial, indicando por todas las fases que debe pasar y qué representaciones son las que nos interesa del set de datos.

Ejercicio 2: Experimenter

Los atributos de los distintos conjuntos de ejemplos (523 instancias) son:

- all_data_pacman: pacmanx,pacmany,fantasmasVivos,distanciaMinAFantasma.
- filter_data_pacman_manual1: pacmanx,pacmany,distanciaMinAFantasma
- filter_data_pacman_manual2: fantasmasVivos,distanciaMinAFantasma

La clase para los tres es la indicada en el enunciado, la dirección de Pac-man en ese estado.



Punto 2.3.5

El mejor conjunto es el que contiene todos los datos sin filtrar, ya que ha conseguido un 64.17% de instancias predichas con el algoritmo IbK con $K=1$.

Punto 2.3.6

Creemos que el mejor algoritmo es el J48, ya que, aunque no es el que brinda el mejor resultado individual, es el más constante y con mejor media.

Punto 2.3.7

La diferencia entre los algoritmos para el mismo conjunto no varía mucho a excepción de ZeroR, $\approx 38\%$, que es considerablemente peor que el resto de los algoritmos, sobre todo comparado con el mencionado en el Punto 2.3.5. Se puede apreciar en la gráfica que es el mínimo.

Punto 2.3.9

El Experimenter es una herramienta muy útil para evaluar sobre un mismo conjunto de datos distintos algoritmos de clasificación simultáneamente, además no solo nos permite evaluar un conjunto de datos, si no que podemos introducir varios conjunto para analizar más fácilmente los resultados. También evalúa como de buena ha sido con respecto al esquema base.

Conclusiones y dificultades encontradas

En definitiva, este tutorial nos ha servido para terminar de conocer la herramienta de Weka y lo que podemos conseguir con la herramienta Experimenter y KnowledgeFlow. No nos hemos encontrado muchas dificultades a la hora de realizar los ejercicios ya que todo está bastante guiado y a la hora de cambiar el código de python para generar el archivo arff ya conocíamos lo suficientemente bien la estructura del código para que no nos generará ningún problema.