

UNIVERSIDAD CARLOS III DE MADRID

APRENDIZAJE AUTOMÁTICO

COMPUTER SCIENCE ENGINEERING

Tutorial 1: Plataforma PacMan

Authors:

Daniel MEDINA GARCÍA

Alejandro RODRÍGUEZ SALAMANCA

February 18, 2016

Contents

1	Los ficheros de datos	3
1.1	¿Cuántos atributos de entrada tiene el fichero de datos? ¿De qué tipo son? .	3
1.2	¿Podría un algoritmo de aprendizaje automático identificar esa función con los datos que hay en ese fichero? ¿Por qué?	3
2	Clasificar con ZeroR	4
2.1	¿Qué resultado en términos de instancias correctas ofrece el algoritmo ZeroR?	4
2.2	¿Qué ocurre si se selecciona otro algoritmo de clasificación permitido para ese conjunto de datos?	4
2.3	¿Cuáles son las diferencias al repetir los pasos anteriores con el otro fichero <code>badges_plain.arff</code> ?	4
2.4	¿Qué ocurre si seleccionamos el algoritmo <code>trees</code> / ID3 en el segundo fichero?	4
3	Generando nuevos atributos	5
3.1	Propón 6 nuevos atributos y explica por qué los has elegido.	5
3.2	¿Cuántos atributos tiene el fichero <code>badges1.arff</code> y de qué tipo son?	5
3.3	¿Qué otro tipo de información estadística se muestra sobre los atributos? Tras pulsar el botón “Visualize all” indica qué se muestra y si hay algún atributo que no se visualice.	5
3.4	Genera un clasificador con ZeroR, ¿qué ocurre? Compara los resultados con los obtenidos en el ejercicio anterior.	5
3.5	Genera un clasificador con <code>trees</code> / ID3, ¿qué ocurre? ¿Qué se podría hacer para solucionar este problema?	5
4	Clasificar con ID3: resolviendo problemas	6
4.1	¿Qué información aparece en el desplegable tras abrir la pestaña <i>Classify</i> / <i>ZeroR</i> / <i>Capabilities</i> ?	6
4.2	¿Qué información proporciona “more”?	6
4.3	¿Qué efecto tiene el filtro <i>Filter</i> / <i>unsupervised</i> / <i>attribute</i> / <i>Discretize</i> sobre el conjunto de datos con <i>bins</i> igual a 5?	6
4.4	¿Cuántas instancias clasifica bien cuando marcamos <i>Use training set</i> en <i>Test options</i> ? ¿Qué porcentaje representa?	6
4.5	¿Qué crees que indica la “matriz de confusión”?	6
4.6	¿Cuántas instancias de cada tipo se han clasificado mal?	6
4.7	¿Cuál es la primera instancia del conjunto de entrenamiento que se clasifica mal? ¿Por qué?	6
4.8	¿Cómo se clasificaría la instancia “Eloisa Figueroa”? ¿Cuáles son los atributos de este nombre? ¿Qué ocurre con los valores de esta instancia si utilizas el filtro usado anteriormente?	6
4.9	¿Cómo se clasifica la instancia nueva?	7
5	Clasificar con J48 (C4.5)	8
5.1	Pregunta	8

6	Pregunta 1	9
6.1	Pregunta	9
7	Pregunta 1	10
7.1	Pregunta	10

1 Los ficheros de datos

1.1 ¿Cuántos atributos de entrada tiene el fichero de datos? ¿De qué tipo son?

Uno. Decir tipo y demás y por qué el de clase no es de entrada.

1.2 ¿Podría un algoritmo de aprendizaje automático identificar esa función con los datos que hay en ese fichero? ¿Por qué?

No porque mi respuesta es de libro.

2 Clasificar con ZeroR

2.1 ¿Qué resultado en términos de instancias correctas ofrece el algoritmo ZeroR?

Respuesta

2.2 ¿Qué ocurre si se selecciona otro algoritmo de clasificación permitido para ese conjunto de datos?

Respuesta

2.3 ¿Cuáles son las diferencias al repetir los pasos anteriores con el otro fichero `badges_plain.arff`?

Respuesta

2.4 ¿Qué ocurre si seleccionamos el algoritmo `trees` / `ID3` en el segundo fichero?

Respuesta

3 Generando nuevos atributos

3.1 Propón 6 nuevos atributos y explica por qué los has elegido.

Respuesta

3.2 ¿Cuántos atributos tiene el fichero badges1.arff y de qué tipo son?

Respuesta

3.3 ¿Qué otro tipo de información estadística se muestra sobre los atributos? Tras pulsar el botón “Visualize all” indica qué se muestra y si hay algún atributo que no se visualice.

Respuesta

3.4 Genera un clasificador con ZeroR, ¿qué ocurre? Compara los resultados con los obtenidos en el ejercicio anterior.

Respuesta

3.5 Genera un clasificador con trees / ID3, ¿qué ocurre? ¿Qué se podría hacer para solucionar este problema?

No te deja ejecutarlo, por qué? Creo que tenía que ver con el tipo de datos

4 Clasificar con ID3: resolviendo problemas

4.1 ¿Qué información aparece en el desplegable tras abrir la pestaña *Classify / ZeroR / Capabilities*?

Respuesta

4.2 ¿Qué información proporciona “more”?

Respuesta

Los atributos de entrada pueden modificarse a través de tareas de preprocesamiento. En los siguientes pasos vamos a modificar ciertos atributos de `badges1.arff` para que pueda clasificarse con ID3.

4.3 ¿Qué efecto tiene el filtro *Filter / unsupervised / attribute / Discretize* sobre el conjunto de datos con *bins* igual a 5?

Respuesta

4.4 ¿Cuántas instancias clasifica bien cuando marcamos *Use training set* en *Test options*? ¿Qué porcentaje representa?

Respuesta

4.5 ¿Qué crees que indica la “matriz de confusión”?

Respuesta

4.6 ¿Cuántas instancias de cada tipo se han clasificado mal?

Respuesta

Ahora pulsamos *More Options* y seleccionamos la opción *Output predictions*

4.7 ¿Cuál es la primera instancia del conjunto de entrenamiento que se clasifica mal? ¿Por qué?

Respuesta

4.8 ¿Cómo se clasificaría la instancia “Eloisa Figueroa”? ¿Cuáles son los atributos de este nombre? ¿Qué ocurre con los valores de esta instancia si utilizas el filtro usado anteriormente?

Respuesta

A continuación modificamos el fichero original introduciendo el nombre anterior y cambiamos la clase a “positiva”, teniendo en cuenta que si contiene enumerados y se introduce un nuevo valor hay que especificarlo también en la definición de los valores posibles del enumerado. Después volvemos a generar el clasificador con ZeroR y training set seleccionado.

4.9 ¿Cómo se clasifica la instancia nueva?

Respuesta

5 Clasificar con J48 (C4.5)

Volvemos a la pestaña de preproceso para cargar `badges1.arff` , usando el training set para volver a generar el clasificador.

5.1 Pregunta

Respuesta

6 Pregunta 1

6.1 Pregunta

Respuesta

7 Pregunta 1

7.1 Pregunta

Respuesta