

14-04-2013

Redes Bayesianas

SISTEMAS DE APOYO A LA DECISION

JON ANDER FONTÁN, HELEN JOANA PAZ, EGOITZ PUERTA

Índice

Marco Teórico.....	2
1. Redes Bayesianas	2
2. Ejemplos de conexión entre nodos en redes bayesianas.....	2
2.1. Head-to-tail connection (Conexión Cabeza-a-cola).....	2
2.2. Tail-to-tail connection (Conexión Cola-a-Cola).....	3
2.3. Head-to-Head connection (Conexión cabeza-a-cabeza)	4
2.4. Combinaciones de las anteriores en un mismo grafo:	5
3. Ejemplos de modelos gráficos.....	5
3.1. Clasificador Naive Bayes.....	5
4. Diagramas de Influencia.	6
5. ¿Dónde podemos encontrar esto en Weka?	6
Marco Experimental.....	8
1. Descripción del Diseño e implementación.....	8
2. Ejemplo de ejecución.	8
3. Resultados Experimentales.	9
Bibliografía	10

Marco Teórico

1. Redes Bayesianas

Son grafos que se representan mediante nodos y arcos entre ellos. Los nodos representan variables aleatorias y los arcos son la influencia entre las dos variables aleatorias. Estos grafos son siempre acíclicos y dirigidos.

El valor de los nodos es la probabilidad de la variable aleatoria ($P(X)$) y el valor de los arcos refleja la probabilidad condicionada entre ellos, es decir, si hay un arco entre las variables X e Y este arco tendrá el valor correspondiente a la probabilidad de Y sabiendo que ha ocurrido X ($P(Y|X)$).

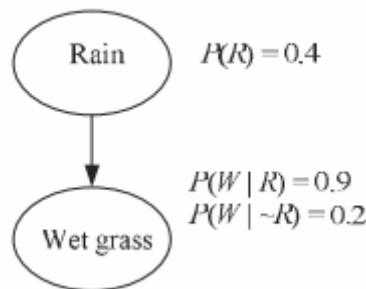


Ilustración 1

Se usan para hacer de una manera más gráfica y visible el cálculo de probabilidades. Tendremos que ir calculando probabilidades condicionadas de cada uno de ellos para hallar la probabilidad de un hecho en concreto. (Ilustración 1)

$$P(R|W) = \frac{P(W|R)P(R)}{P(W)}$$

2. Ejemplos de conexión entre nodos en redes bayesianas

2.1. Head-to-tail connection (Conexión Cabeza-a-cola)

Esta conexión se da en el caso en el que todos sus elementos están en serie. En la Ilustración 2 que se expone a continuación Z es independiente de X , es decir sabiendo el valor de X no podemos decir el valor de Z . Sin embargo, sabiendo el valor de Y ya podemos decir todo sobre Z .

$$P(Z|X, Y) = \frac{P(X)P(Y|X)P(Z|Y)}{P(X)P(Y|X)} = P(Z|Y)$$



Ilustración 2

En la Ilustración 3 para calcular la probabilidad de que la hierba este mojada antes tendremos que conocer previamente las probabilidades condicionadas a las hierba mojada sabiendo que ha llovido y no ha llovido, además es necesario conocer la probabilidad de que ha llovido. Esta probabilidad también es condicionada a uno o varios sucesos que han ocurrido

con anterioridad y tendrán que ser calculados de la misma forma. Esto tendrá que realizarse hasta llegar al nodo raíz el cual deberemos de conocer previamente su probabilidad.

$$P(C|W) = \frac{P(W|C)P(C)}{P(W)}$$

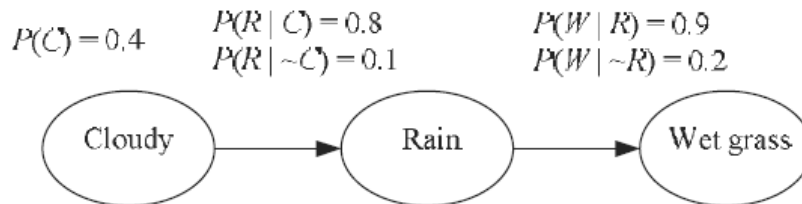


Ilustración 3

2.2. Tail-to-tail connection (Conexión Cola-a-Cola)

Es un ejemplo de red bayesiana en el cual los nodos siguen una jerarquía de un padre y varios hijos. (Ilustración 4)

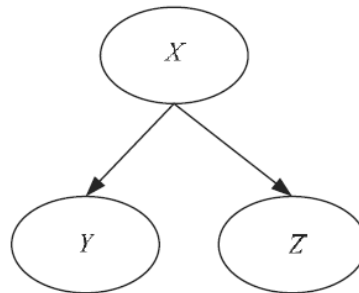


Ilustración 4

En la Ilustración 5 los dos nodos hijos que derivan de “Nublado” son independientes, por lo tanto sus probabilidades también lo son, para calcular la probabilidad de que llueva solo necesitaremos conocer si esta nublado o no. Lo mismo ocurre para “Aspersor”, solo necesitamos conocer si esta nublado o no. El hecho de que llueva o no, es un suceso totalmente independiente a que se active el aspersor.

$$P(Y, Z|X) = \frac{P(X)P(Y|X)P(Z|X)}{P(X)} = P(Y|X)P(Z|X)$$

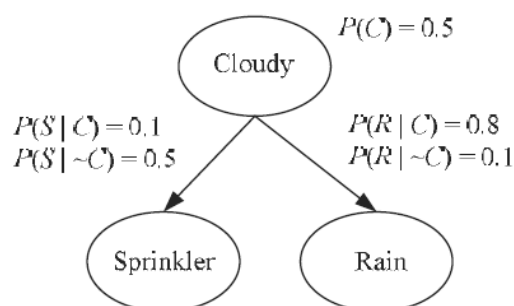


Ilustración 5

$$P(R|C) \frac{P(S|C)P(C)}{P(S)} + P(R|\sim C) \frac{P(S|\sim C)P(\sim C)}{P(\sim S)}$$

2.3. Head-to-Head connection (Conexión cabeza-a-cabeza)

Es un ejemplo de red bayesiana en la cual se sigue una jerarquía de varios padres con un mismo hijo, las probabilidades se relacionan. (Ilustración 6)

$$P(X, Y, Z) = P(X)P(Y)P(Z|X, Y)$$

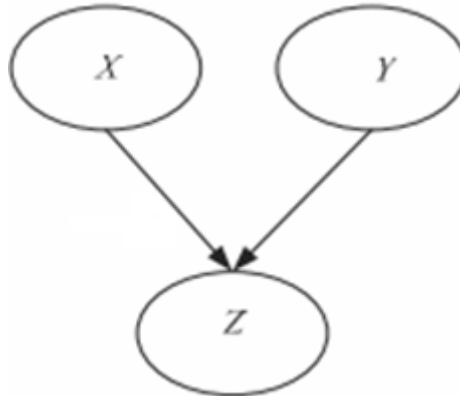


Ilustración 6

En la Ilustración 7 tenemos un caso en el cual el suceso “Hierba Mojada” se ve condicionada a dos variables aleatorias previas, “Aspersor” y “Llover”, por lo tanto la probabilidad de que el suelo este mojado depende directamente de las probabilidades de los dos nodos padres que la preceden.

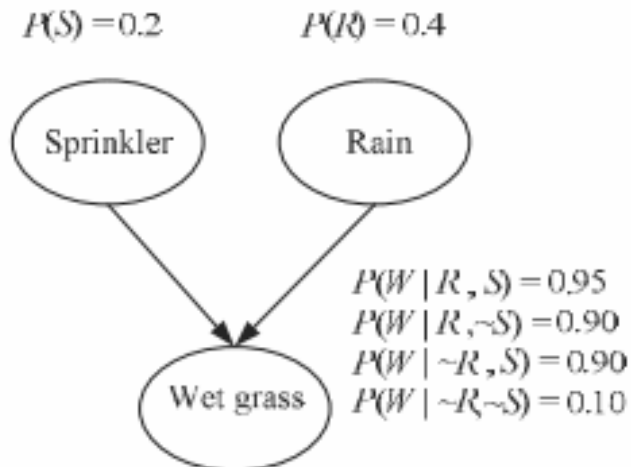


Ilustración 7

$$P(S|W) = \frac{P(W|S)P(S)}{P(W)}$$

2.4. Combinaciones de las anteriores en un mismo grafo:

En un mismo grafo se pueden dar todas las anteriores conexiones ocurriendo simultáneamente, siempre y cuando no generen ningún ciclo entre ellas.

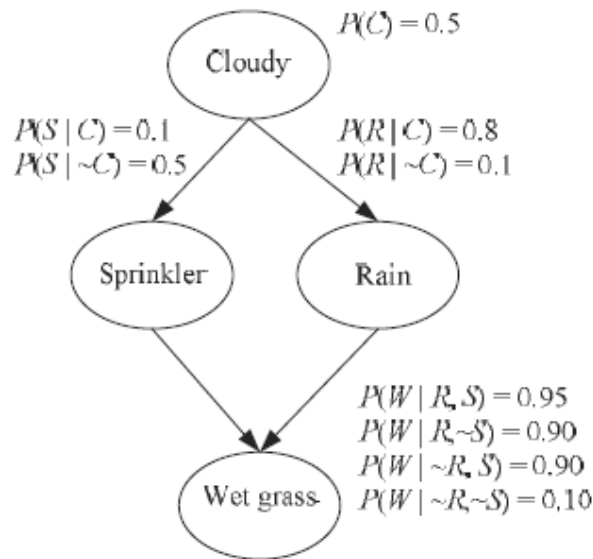


Ilustración 8

En la Ilustración 8 para calcular las probabilidades tendremos que ir calculándolas de forma descendente, ya que todas las probabilidades dependen de la variable aleatoria “Nublado”, para calcular el resto aplicamos los patrones de los puntos anteriormente citados (Apartados 2.2 y 2.3)

3. Ejemplos de modelos gráficos

3.1. Clasificador Naive Bayes

El clasificador Naive Bayes nos permite simplificar la red bayesiana que tenemos en los casos en los que un nodo tenga varios nodos hijo independientes entre ellos, se podrán juntar en un nodo conectado al nodo padre de cuyo valor del arco será el multiplicatorio de todos los valores de cada uno de los arcos de los nodos hijo del grafo original. (Ilustración 9)

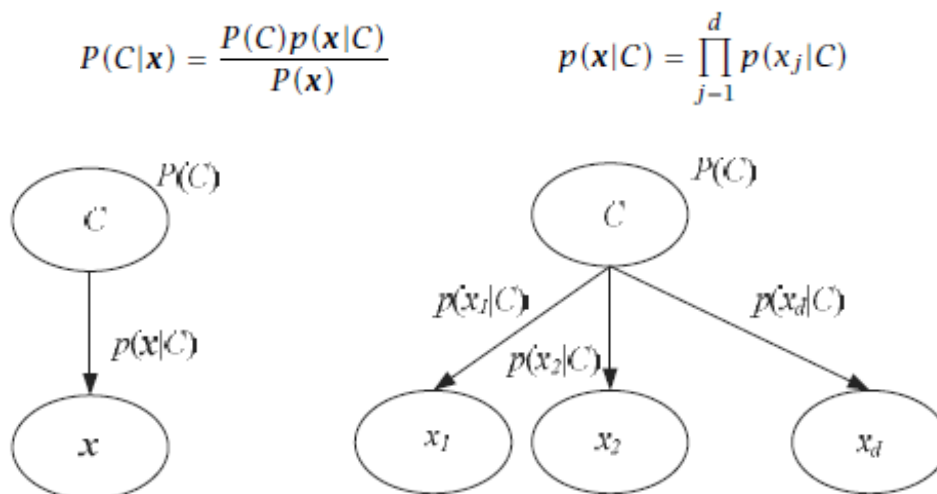


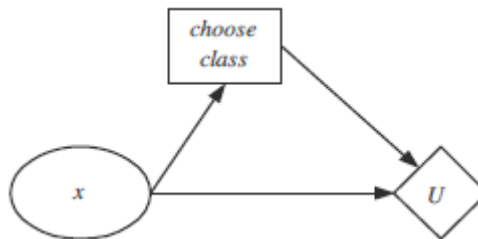
Ilustración 9

4. Diagramas de Influencia.

Son modelos gráficos que permiten la generalización de los modelos gráficos para incluir las decisiones y utilidades. Un diagrama de influencia contienen: nodos de probabilidad que representan variables aleatorias que utilizamos en los modelos gráficos, nodos de decisión y un nodo de utilidad. El nodo de decisión representa una elección de acciones y el nodo de utilidad es el que calcula la utilidad. Las decisiones pueden estar basadas en nodos de probabilidad y puede afectar a otros nodos de azar y al nodo de utilidad.

La inferencia en un diagrama de influencia es la propagación de probabilidad en un modelo gráfico. Para cada posible decisión, se calcula la utilidad y se toma la decisión en función de la utilidad más alta.

El diagrama de influencia para la clasificación de una entrada dada, se muestra en la imagen. En función de la entrada x , se escoge una clase que incurre en una cierta utilidad (riesgo).



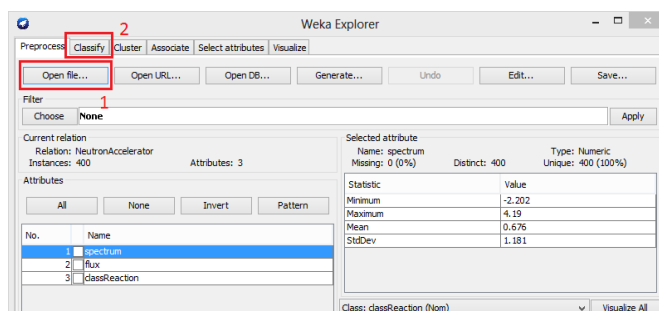
5. ¿Dónde podemos encontrar esto en Weka?

Weka nos da la posibilidad de configurar el clasificador correspondiente a las redes bayesianas (BayesNet). Para poder utilizar este clasificador en Weka y poder ver todas las configuraciones que nos permite tendremos que hacer lo siguiente.

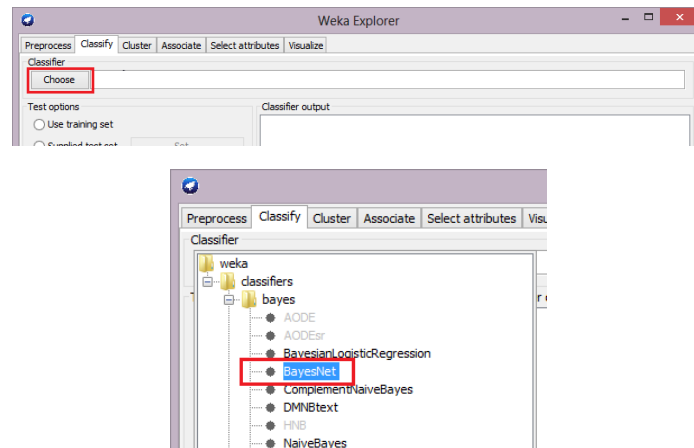
Accedemos a La interface gráfica de Weka y hacemos clic en el botón “Explorer”.



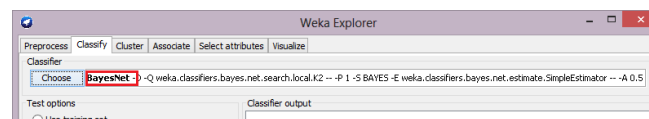
Abrimos el fichero de origen de datos que queramos analizar y le damos a la pestaña “Classify”.



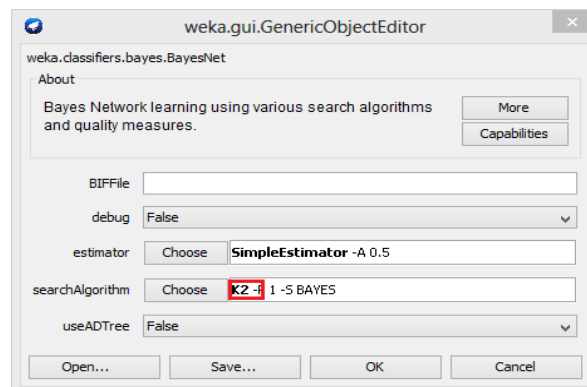
Hacemos clic en el botón Choose para seleccionar el clasificador y seleccionamos en clasificador BayesNet que se encuentra en la siguiente ruta: *weka -> classifiers -> bayes*.



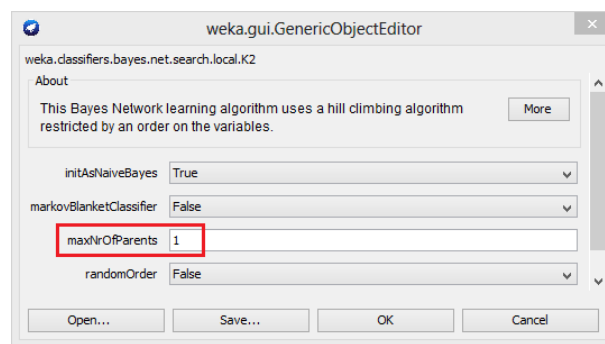
A continuación para poder definir los parámetros que influyen a BayesNet tenemos que hacer clic en su nombre.



Nos aparecerán las distintas opciones que nos permite configurar el clasificador BayesNet, pero la que nos interesa a nosotros es poder definir el número de parientes que puede tener un nodo. Para configurarlo tenemos que hacerlo en el apartado del algoritmo de búsqueda, para ello tenemos que hacer clic en el que nos aparece (K2).



En la ventana que nos aparece tenemos que modificar el parámetro de “maxNrOfParents” al que se requiera en cada momento.



Marco Experimental

1. Descripción del Diseño e implementación.

Para esta práctica en grupo hemos utilizado el algoritmo BayesNet y para determinar su eficacia, hemos usado como baseline el algoritmo OneR.

Antes de empezar a usar los datos propiamente dichos hemos realizado un preproceso de estos. Para el baseline simplemente hemos desordenado los datos con un randomize. Mientras que para BayesNet se han normalizado, dándole el mismo peso a ambos atributos; discretizado, ya que el algoritmo solo trabaja con valores discretos y por último los hemos mezclado.

Tras este preproceso, hemos hecho un barrido de parámetros para encontrar la mejor configuración de los métodos para los datos. Para ello se ha escaneado el minBucketSize en el baseline mientras que el parámetro usado en BayesNet ha sido el máximo número de padres. Ambos escaneos se han realizado desde uno hasta el número de instancias totales de los datos.

Para realizar la estimación utilizamos el método de k-Fold Cross validation con valor de 10, lo que hace este método es dividir los datos en 10 partes iguales y utiliza 9 de esas partes para entrenar el modelo y 1 para probarlo, obtiene los porcentajes de aciertos y fallos y vuelve a repetir cambiando la parte que usa para probar y entrenar.

Para comparar los resultados hemos usado la figura de mérito f-measure, que relaciona la cobertura de acierto con la precisión, haciendo de este valor una buena forma de comparación.

Tras obtener la mejor configuración de parámetros, se ha procedido a crear un estimador basado en el método correspondiente, para predecir los datos reservados para el test.

Para evitar dependencias a archivos se ha creado una función que pide un dato por consola y lo recoge lo que el usuario introduce. En caso de que el usuario no introduzca los datos correctamente se pedirá que lo introduzca de nuevo hasta que sea correcto.

El barrido de parámetros se ha realizado mediante el método de fuerza bruta, almacenando únicamente el valor actual y el máximo obtenido hasta el momento.

2. Ejemplo de ejecución.

Para ejecutar nuestras pruebas se ha de lanzar los dos métodos por separado. Por ejemplo desde el entorno de desarrollo Eclipse, tras haber importado las clases correspondientes. El funcionamiento de ambos métodos es similar y por ello se explicará solo uno de ellos.

Tras lanzar el método, lo primero que ocurre es que nos pide un archivo con datos de entrenamiento. Este archivo ha de ser de tipo *.arff. Se le introducirá la ruta donde se encuentren el archivo, o el nombre del archivo únicamente si este se encuentra en la misma carpeta donde se ejecuta el programa.

Windows: "C:\\workspace\\DSS\\neutrons.arff"

Linux: "/home/user/workspace/DSS/neutrons.arff"

Si el archivo está en la misma carpeta: "neutrons.arff"

Tras unos segundos, el programa mostrara el resultado del barrido de parámetros, mostrando el máximo de ellos y las figuras de mérito asociadas.

Tras esto nos pedirá otro nuevo archivo con los datos de prueba. Estos datos han de tener el mismo esquema que los anteriores, con el mismo número de atributos y estar en formato *.arff. La única diferencia es que en el atributo clase ha de estar sin informar (indicado como "?"). La ruta se introducirá de igual manera que la anterior.

A continuación nos pedirá una nueva ruta, para el archivo de salida. Como las veces anteriores, se ha de introducir la ruta completa para fijar el path donde dejar el archivo con los resultado, o bien solo especificar el nombre si se quiere guardar en la misma carpeta que el programa ejecutado. Este archivo puede existir, en cuyo caso quedara totalmente sobrescrito, o no existir, con lo que se procederá a su creación. Esta ruta solo se considerará valida si el archivo tiene una extensión *.txt.

Windows: "C:\\workspace\\DSS\\resultados.txt"

Linux: "/home/user/workspace/DSS/ resultados.txt"

Si el archivo está en la misma carpeta: "resultados.txt"

Este archivo de salida dispondrá una línea con la clase predicha relativa a la instancia correspondiente en el archivo con datos de testeo. El resultado se dará con los valores correspondientes a posibles valores de la clase.

3. Resultados Experimentales.

A continuación observaremos los resultados obtenidos con los modelos OneR y BayesNet estudiado con un marco de evaluación 10-fold cross validation. Empleando las figuras de mérito: Accuracy, Precision, Recall, F-measure, Area under the ROC curve para cuantificar los dos modelos.

Figura de Merito	Definición	Valor en Baseline (OneR)	Valor en Bayes Net
Accuracy	Es el porcentaje de aciertos.	79,75	86
Precision	Es el porcentaje de las instancias correctamente clasificadas como positivos, entre el total de <i>predicciones</i> positivas.	0,755364807	0,89130435
Recall	Es el porcentaje de las instancias correctamente clasificadas como positivas entre el total de positivos.	0,88	0,82
F-Measure	Es un valor que indica que tan bueno es el modelo que estamos usando, es dos veces la Precision por el Recall entre la Precision mas el Recall.	0,812933025	0,85416667
Area under the ROC	Es la probabilidad de que un clasificador predijera una instancia positiva (elegida aleatoriamente) más alta que una negativa.	0,7975	0.9128625

Podemos observar que dependiendo del tipo de figura de mérito que usemos es mejor un modelo u otro, sin embargo observando los resultados de forma global, podríamos decir que el BayesNet parece ser más adecuado para realizar predicciones.

Bibliografía

- Alpaydin Machine Learning 2010. Cap 16 (Graphical Models)
- Documentación online de weka.