

Métodos Bayesianos II

García Prado, Sergio
sergio@garciparedes.me

1 de mayo de 2017

Resumen

[TODO]

1. INTRODUCCIÓN

[TODO]

2. REDES BAYESIANAS

[TODO]

2.1. ESTRUCTURA NAIVE BAYES

[TODO]

2.2. ESTRUCTURA TAN

[TODO]

2.3. ESTRUCTURA K2

[TODO]

3. EXPERIMENTOS

Tras haber descrito las *Redes Bayesianas* en la sección anterior, a continuación se presentan los resultados obtenidos tras realizar un conjunto de experimentos sobre el conjunto de datos **Credit**, que se describen en la sección ???. Los tests han consistido en la evaluación del comportamiento de los algoritmos de generación de *Redes Bayesianas* implementadas en *Weka* [too] variando el tamaño de los conjuntos de entrenamiento.

La metodología seguida ha sido la siguiente: Para cada conjunto de datos de entrenamiento se ha realizado un experimento de *Validación Cruzada de 10 particiones* a partir del cual se ha almacenado la tasa de error obtenida. Además, se han guardado los modelos generados en cada caso para después probarlos sobre un conjunto de datos independiente del de entrenamiento. Esta tarea se ha repetido con tres los conjuntos de datos de 3 tamaños diferentes sobre los algoritmos de generación de *Redes Bayesianas* descritos en la sección anterior: *Naive Bayes*, *K2* y *TAN*.

A continuación se describe la naturaleza del conjunto de datos utilizado para los experimentos así como los tamaños de sus particiones.

3.1. CONJUNTO DE DATOS

El conjunto de datos utilizado se denomina *Credit* y se puede acceder a él a través de <https://github.com/garciparedes/machine-learning-bayesian-2/tree/master/weka/datasets> [GP17]. Está formado por **11 atributos** de carácter nominal más la clase de destino formada por **2 valores**. Dichos atributos presentan un rango de valores reducidos, encontrándose todos ellos entre **2 y 4 valores distintos**.

En cuanto a los conjuntos de datos utilizados para los experimentos, todos ellos contienen los mismos atributos así como instancias representativas para todos los valores posibles de los atributos. Tampoco se dan atributos desconocidos. a continuación se describen los tamaños de cada conjunto de datos:

- **Datos_Credit_100:** Está formado por 100 instancias.
- **Datos_Credit_1000:** Está formado por 1000 instancias.
- **Datos_Credit_10000:** Está formado por 10000 instancias.
- **Test_Credit_1000:** Está formado por 1000 instancias.

Por último es necesario describir tanto el estimador como los parámetros de configuración utilizados para cada algoritmo. En cuanto al *método de estimación* de probabilidades, se ha escogido el *Estimador de Máxima Verosimilitud con corrección de Laplace* y $m = 0,5$. Dicho estimador se muestra en la ecuación (1), donde n_c representa el número de ejemplos de entrenamiento con la clase b_j , n el número de ejemplos de la de entrenamiento con la clase b_j y el atributo a_i . El valor p se obtiene mediante $p = Pr(A = a_i|B = b_j)$, es decir, la estimación a priori y m un determinado peso para la estimación a priori.

$$Pr'(A = a_i|B = b_j) = \frac{n_c + mp}{n + m} \quad (1)$$

En cuanto a los algoritmos de clasificación, *Naive Bayes* no tiene más parámetros adicionales. *K2* se ha configurado para que no comience de una red *Naive Bayes* y se ha permitido que cada nodo tenga un número arbitrario de padres. En el caso de *TAN*, se ha seleccionado el método de puntuación de *Máxima Verosimilitud con corrección de Laplace* descrito en la ecuación (1) con parámetro $m = 0,5$ al igual que en el caso del método de estimación de probabilidades.

En la sección 3.2 se presentan los resultados obtenidos así como una discusión acerca de los mismos.

3.2. RESULTADOS

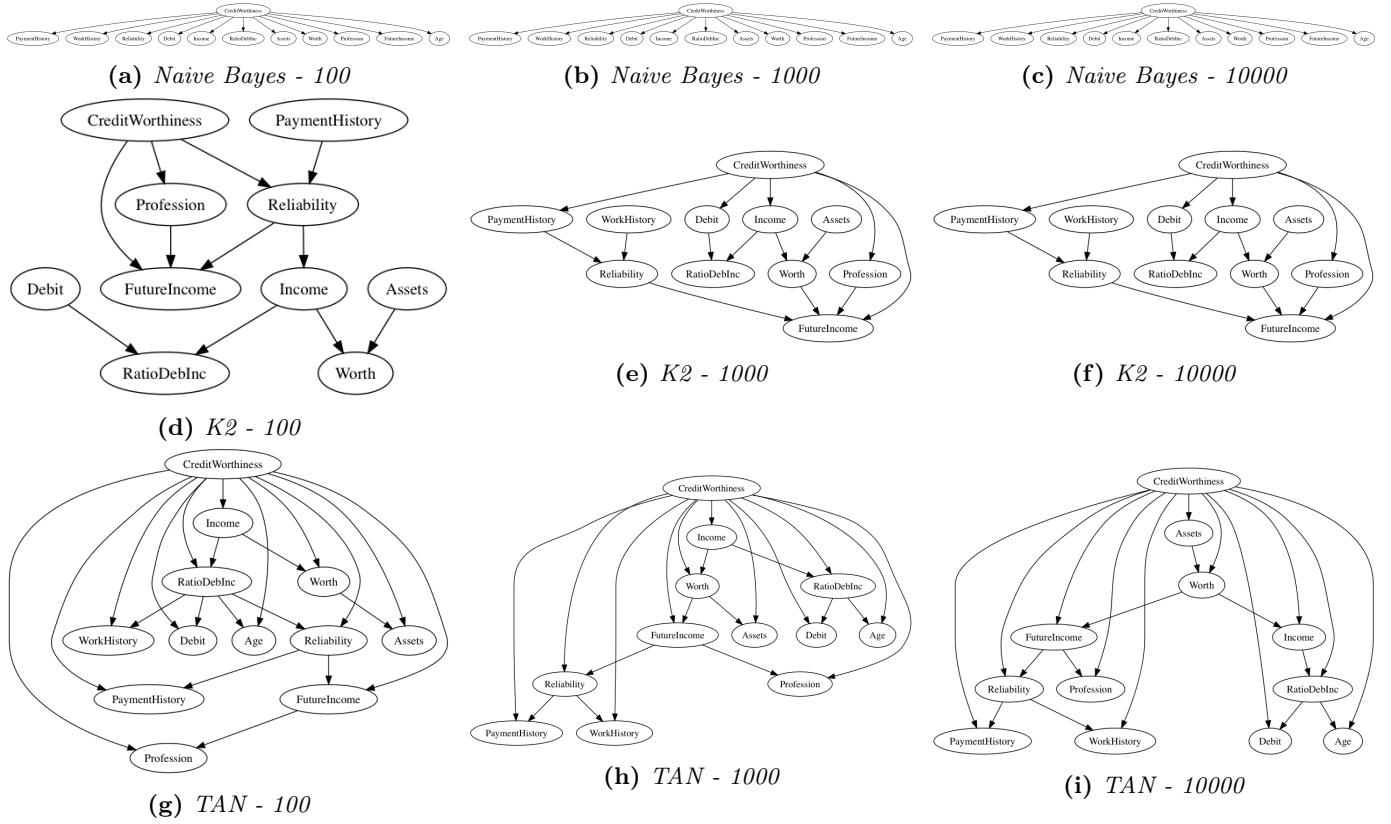
En la figura 1 se muestra la estructura de las distintas *Redes Bayesianas* generadas por cada algoritmos dependiendo del tamaño del conjunto de datos de entrenamiento. Nótese que estas redes han sido generadas a partir de un experimento de *Validación Cruzada de 10 particiones* sobre conjuntos de datos de 100, 1000 y 10000 instancias respectivamente.

En cuanto al algoritmo *Naive Bayes*, la estructura de la red generada es trivial en ambos casos, suponiendo dependencia directa de la clase sobre todos los atributos, por lo que se genera un árbol de profundidad 1 donde el nodo raíz es la clase y los nodos hoja se corresponden con los 11 atributos. Nótese que el tamaño del conjunto de datos de entrenamiento no genera variaciones tal y como se muestra en las figuras 1a, 1b y 1c.

Las redes generadas por el algoritmo *K2* presentan una característica diferenciadora respecto del resto. En este caso no todos los atributos tienen como padre al nodo referido a la *clase* (algunos a pesar de no ser la clase de destino no tienen padre). Esto se traduce de manera práctica en que dichos atributos no serán utilizados durante las fases de clasificación de nuevas instancias. La razón es la estrategia de construcción de redes seguida por este algoritmo, que añade una arista tan solo cuando la puntuación obtenida mediante el indicador de verosimilitud se mejora. Las estructuras generadas a partir de cada conjunto de datos de distinto tamaño se muestran en las figuras 1d, 1e y 1f.

[TODO descripción TAN]

[TODO]

**Figura 1:** Redes Bayesianas generadas a partir de DatosCredit

Clasificación mediante Redes Bayesianas									
Datos	Tasa de Error								
	Naive Bayes			K2			TAN		
	100	1000	10000	100	1000	10000	100	1000	10000
Entrenamiento	36,000 %	31,700 %	29,690 %	28,000 %	36,100 %	28,500 %	35,000 %	29,500 %	28,270 %
Datos Test	30,000 %	27,800 %	27,500 %	33,300 %	30,400 %	25,400 %	33,300 %	24,900 %	25,000 %

Tabla 1: Tasas de error obtenida a partir de distintas configuraciones de Redes Bayesianas

REFERENCIAS

- [CCAG17] Teodoro Calonge Cano and Carlos Javier Alonso González. Técnicas de Aprendizaje Automático, 2016/17.
- [GP17] Sergio García Prado. Métodos bayesianos 2. <https://github.com/garciparedes/machine-learning-bayesian-2>, 2017.
- [too] Weka. <http://www.cs.waikato.ac.nz/ml/weka/>.