

Aprendiendo Redes Bayesianas para Clasificación: Buscando en el espacio de los Grafos Acíclicos Parcialmente Dirigidos

**Silvia Acid
Luis M. de Campos
Fco. Javier García**

**Uncertainty Treatment in Artificial Intelligence Research Group
Department of Computer Science and Artificial Intelligence**

Granada University (Spain)



Introducción



Clasificación supervisada



Aprender una red bayesiana y utilizarla como clasificador.



En un paradigma métrica+búsqueda, la optimización de la métrica no supone un mayor acierto en las predicciones.



Nuestro método se basa en la utilización de un algoritmo de búsqueda especializado



Nuestro espacio de búsqueda son los C-RPDAGS que combinan dos conceptos de equivalencia para dags:



Equivalencia de clasificación



Equivalencia de independencia

Introducción

Redes Bayesianas como Clasificadores

- Una Red Bayesiana es una representación gráfica de una distribución de probabilidad, con dos componentes:
 - Un grafo directo acíclico (DAG) $G=(U,E_g)$ donde $U=\{X_1, X_2, \dots, X_n\}$ es un conjunto de nodos que representa las variables del sistema y E_g es un conjunto de arcos que representa relaciones de dependencia directa entre las variables de U
 - Un conjunto de parámetros para cada variables de U , normalmente distribuciones de probabilidad condicionada.
- Una variable sólo se encuentra influenciado por el Markov Blanket de la misma (variables padres, hijos y padre de hijos)
- Para usar una red bayesiana como clasificador sólo nos hace falta saber el Markov Blanket de la variable a Clasificar

Redes Bayesianas como Clasificadores

Aprendizaje de Redes Bayesianas



Comos las redes tienen dos componentes, hay que realizar dos tareas:



Aprendizaje de la estructura gráfica



Aprendizaje de los parámetros



Para realizar el aprendizaje paramétrico es necesario conocer antes la estructura de la red.



Los algoritmos de aprendizaje estructural lo podemos dividir en:

- Métodos basados en detección de independencias
- Métodos basados en funciones de evaluación y técnicas de búsqueda (métrica+búsqueda)
- Métodos híbridos



En los métodos basados en el paradigma métrica+búsqueda, cada algoritmo está caracterizado por la función de evaluación y el procedimiento de búsqueda usado

Aprendizaje de Redes Bayesianas

Aprendizaje de Redes Bayesianas

- Los métodos basados en métrica+búsqueda buscan un grafo que maximice la métrica, dónde dicha métrica es una medida de adecuación del grafo a los datos.
- Estos métodos son eficientes principalmente gracias a la propiedad de descomponibilidad que muchas funciones de evaluación tienen.
- Una función de evaluación g se dice que es descomponible si la evaluación de cualquier estructura de una red bayesiana puede verse como un producto (o suma en el dominio logarítmico) de evaluaciones locales de un nodo y sus padres.

$$g(G : D) = \sum_{X \in U} g_D(X, Pa_G(X))$$

Aprendizaje de Redes Bayesianas

Clasificadores Bayesianos



Un método de métrica+búsqueda encuentra una red que maximice la función de evaluación, pero no necesariamente la capacidad predictora de la red.



Se han desarrollado una serie de algoritmos que realizan una búsqueda entre diferentes tipos de DAGs, en su mayoría extensiones del modelo Naive Bayes.:

- Naive Bayes (NB)
- Tree Augmented NB (TAN)
- Bayesian network augmented NB (BAN)
- Semi NB
- Selective NB
- K-dependent Bayesian Classifier (KDB)
- Hierarchical NB (HNB)



Además tenemos:

- Redes Bayesianas (Unrestricted Bayesian Networks –UBN-)
- Métodos basados en Markov Blanket (MB-GA, RMB-GA)
- Multiredes Bayesianas (Bayesian Multinets –BMN-)

Clasificadores Bayesianos

Class-Focused DAGs (C-DAGs)



Equivalencia de clasificación: Sean $G=(V,E_g)$ y $G'=(V,E_{g'})$ dos DAGs. Sea p cualquier distribución de probabilidad conjunta sobre V , sean p_G y $p_{G'}$ la distribución de probabilidad que factorizan G y G' respectivamente, que se definen como $p_G(X | Pa_G(X))=p_{G'}(X | Pa_G(X))$ y $p_{G'}(X | Pa_{G'}(X))=p_G(X | Pa_{G'}(X))$, $\forall X \in V$. Si $p_G(C | x_1, \dots, x_n)=p_{G'}(C | x_1, \dots, x_n)$, decimos que G y G' son equivalentes en clasificación.



Proposición: Dado cualquier DAG $G=(V,E_g)$, sea $G_c=(V,E_c)$ un subgrafo de G definido como sigue:

1. $Pa_{G_c}(C)=Pa_G(C)$,
2. $\forall X \in V, X \neq C$, si $C \in Pa_G(X)$ entonces $Pa_G(X) = Pa_{G_c}(X)$,
3. $\forall X \in V, X \neq C$, si $C \notin Pa_G(X)$ entonces $Pa_{G_c}(X)=\emptyset$

Entonces G y G_c son equivalentes en clasificación



Nótese que G_c no es igual al subgrafo de G inducido por el Markov Blanket de C . Este formalismo tiene menos arcos.

G_c es el mínimo subgrafo de cualquier DAG G que realiza la función de clasificador igual que G para la variable C

Class-Focused DAGs (C-DAGs)

Class-Focused DAGs (C-DAGs)



Class-focused DAG: Un DAG $G=(U,E_g)$ es un class-focused DAG (C-DAG) respecto de la variable C si y sólo si satisface la siguiente condición:

$\forall X,Y \in G$, Si $X \rightarrow Y \in E_g$ entonces $Y=C$ o $X=C$ o $C \rightarrow Y \in E_g$



De esta forma un C-DAG contiene sólo arcos entre la clase y otras variables (en cualquier dirección). Y arcos entre variables e hijos de la clase.



Un C-DAG puede verse como una representación canónica de una clase de DAGs, donde todos ellos son equivalentes como clasificadores para C

Class-Focused DAGs (C-DAGs)

PDAGs

Los objetos gráficos que vamos a usar para representar equivalencias en independencia de DAGs son los PDAGs (Acyclic Partically Directed Graphs): Un PDAG es un grafo que contiene enlaces dirigidos (arcos) y no dirigidos (aristas) pero no tiene ciclos dirigidos.

Definamos alguna notación:



El esqueleto de un DAG es el grafo no dirigido que resulta de ignorar la dirección de cada arco



Un patrón h-h (cabeza-cabeza – head-to-head-) en un DAG G es una tripleta ordenada de nodos (X, Y, Z) , donde G contiene los arcos $X \rightarrow Y$ y $Y \leftarrow Z$



Dado un PDAG $G=(V, E_g)$, se definen los siguientes conjuntos de nodos para cada nodo $X \in V$:

- $Ch_G(X) = \{ Y \in V \mid X \rightarrow Y \in E_g \}$, el conjunto de hijos de X .
- $Sib_G(X) = \{ Y \in V \mid X - Y \in E_g \}$, el conjunto de siblings de X .
- $Ad_G(X) = \{ Y \in V \mid X \rightarrow Y \in E_g \text{ o } X - Y \in E_g \}$, el conjunto de adyacentes de X . Obviamente $Ad_G(X) = Pa_G(X) \cup Ch_G(X) \cup Sib_G(X)$

Hay un tipo de PDAGs que pueden ser usados para representar canónicamente clases de DAGs que son equivalentes en independencias: se denominan Completed PDAGs (también conocidos por essential graphs o patterns). Este tipo reduce el espacio de búsqueda, pero el precio que tiene que pagar por esta reducción es que no aprovecha la propiedad de descomposición de las métricas, lo cual los hace poco eficientes.

Nuestro objetivo es aplicar la equivalencia de independencia a los C-DAGs, para ello utilizaremos los Restricted PDAGs (RPDAGs) que se diferencian ligeramente de los Completed PDAGs: dos diferentes RPDAGs pueden corresponder a la misma clase de equivalencia de DAGs pero que nos permiten aprovechar la propiedad de descomponibilidad de la métrica.

PDAGs

Restricted PDAGs (RPDAGs)



Restricted PDAG: Un PDAG $G=(V,E_G)$ es un RPDAG si y sólo si satisface las siguientes condiciones:

1. $\forall X \in V$, Si $Pa_G(X) \neq \emptyset$ entonces $Sib_G(X) = \emptyset$
2. G no contiene ningún ciclo dirigido
3. G no contiene ningún ciclo completamente no dirigido
4. Si $X \rightarrow Y \in E_G$ entonces $|Pa_G(Y)| \geq 2$ o $Pa_G(X) \neq \emptyset$



La primera condición prohíbe configuraciones del tipo $Y \rightarrow X \text{---} Z$ en el RPDAG G .

La condición 4 nos dice que $X \rightarrow Y$ existe en G sólo si es parte de un patrón h-h o existe otro arco (originado por un patrón h-h) que va a X .

Como los RPDAGs son representaciones de conjuntos de equivalencias de independencia de DAGs, definiremos que conjuntos de DAGs están representados por un RPDAG.



Extensión de un RPDAG: Dado un RPDAG $G=(V,E_G)$, decimos que un DAG $H=(V,E_H)$ es una extensión de G ($Ext(G)$) si y sólo si:

- G y H tienen el mismo esqueleto
- Si $X \rightarrow Y \in E_G$ entonces $X \rightarrow Y \in E_H$
- G y H tienen los mismos patrones h-h (el proceso de dirigir los enlaces de G para producir H no genera nuevos patrones h-h).

Restricted PDAGs (RPDAGs)

Class Focused RPDAG (C-RPDAG)



Class Focused RPDAG: Un class-focused RPDAG (C-RPDAG) es un RPDAG cuya extensiones son C-DAGs.



Caracterización de C-RPDAG: Un PDAG $G=(U,E_g)$ es un class-focused RPDAG (C-RPDAG) respecto de la variable C si y sólo si satisface las siguientes condiciones:

1. G no contiene ningún ciclo dirigido
2. Si $Pa_G(C) \neq \emptyset$ entonces $|Pa_G(C)| \geq 2$ y $Sib_G(C) = \emptyset$
3. $\forall X \in V, X \neq C$, si $Pa_G(X) \neq \emptyset$ entonces $C \in Pa_G(X)$ y además $|Pa_G(X)| \geq 2$ o $|Pa_G(C)| \geq 2$
4. $\forall X \in V, X \neq C$, si $Sib_G(X) \neq \emptyset$ entonces $Sib_G(X) = \{C\}$ y $Pa_G(X) = \emptyset$



Con la definición de C-RPDAG obtenemos una reducción del espacio de búsqueda al trabajar con elementos equivalentes en clasificación y equivalentes en independencias



Nos centramos sólo en aquellas estructuras útiles para clasificación



No se impone una estructura NB, aunque se puede obtener



Podemos movernos por estructuras más expresivas



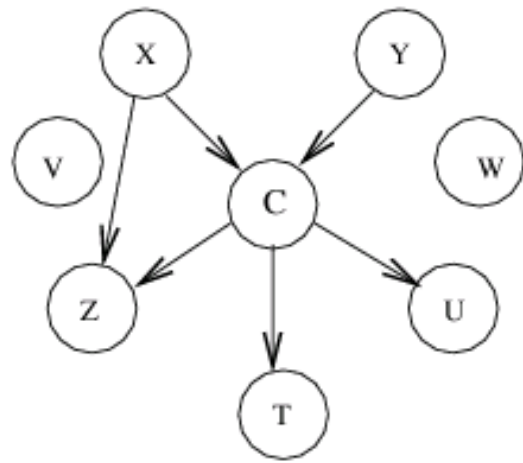
La selección de características es inherente a nuestro enfoque



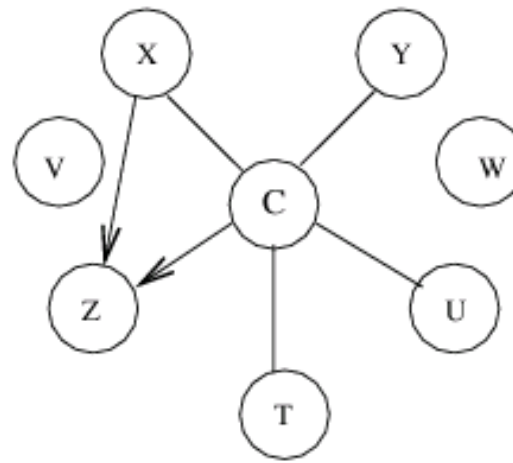
Téngase en cuenta que C-RPDAG no es una representación canónica de clases equivalentes de clasificación de DAGs.

Class Focused RPDAG (C-RPDAG)

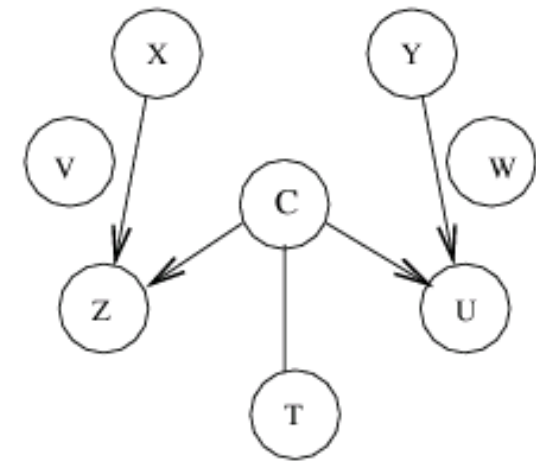
C-RPDAG Search Space



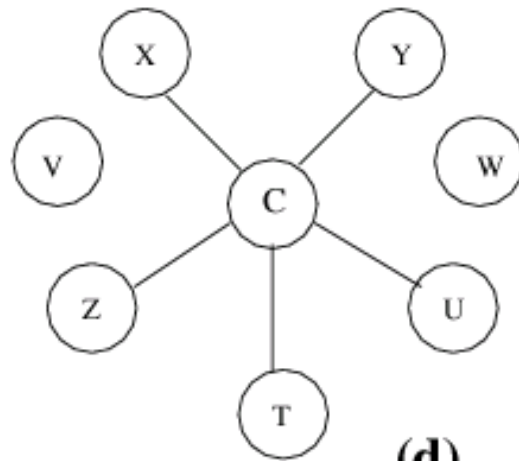
(a)



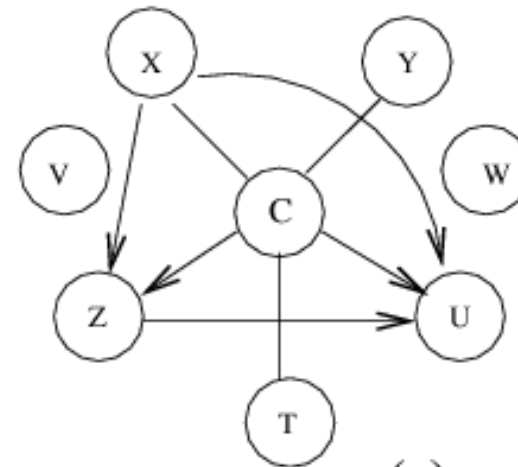
(b)



(c)



(d)



(e)

C-RPDAG Search Space

El Método de Búsqueda

- Usamos un método de búsqueda local en el espacio de los C-RPDAGs
- Comenzamos con un grafo vacío.
- Necesitamos definir una serie de operadores para movernos de una configuración a otra configuración vecina que siga estando en el espacio de los C-RPDAGs
- Los operadores son básicamente: la inclusión de un enlace (orientado o no) entre dos nodos no adyacentes y el borrado de un enlace entre dos nodos adyacentes.
- Los operadores deberán cumplir una serie de condiciones para obtener un vecino válido.

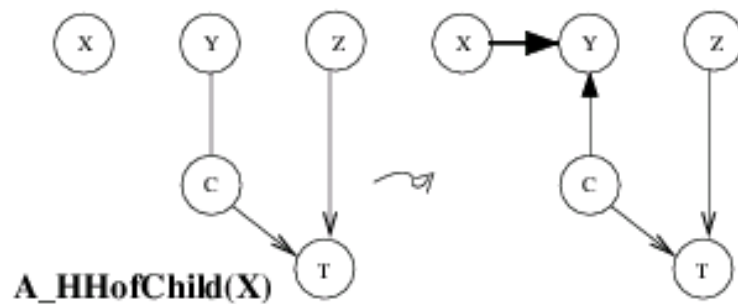
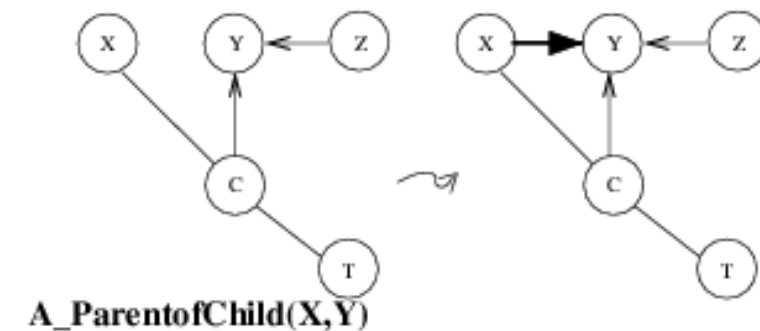
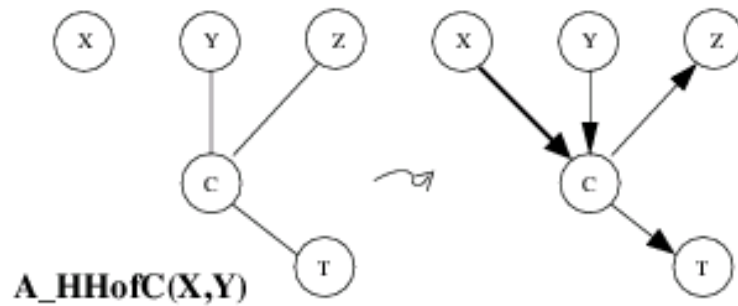
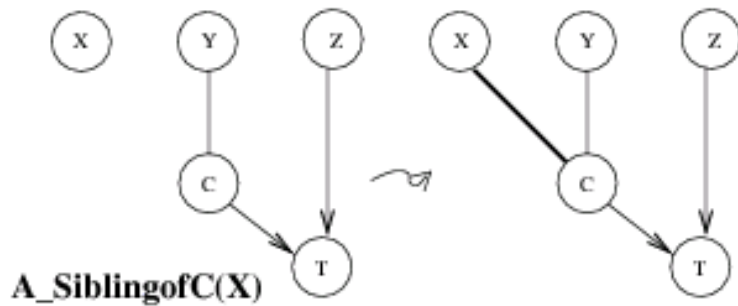
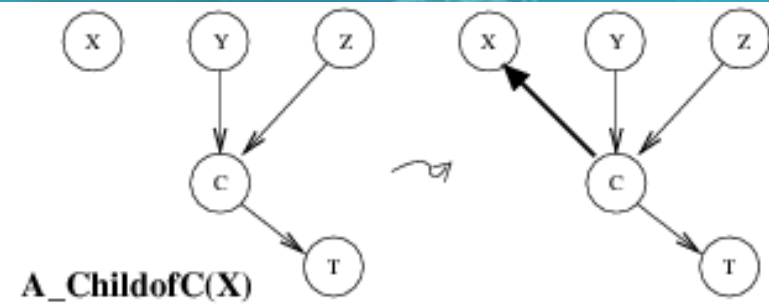
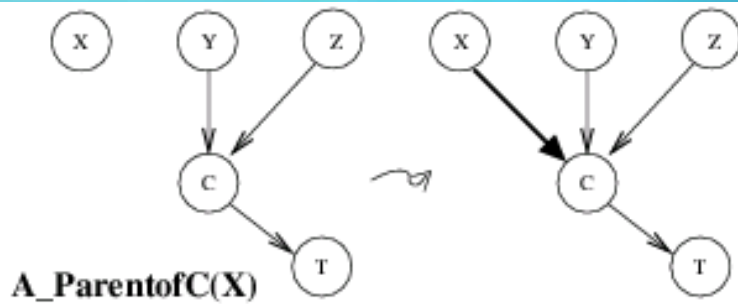
El Método de Búsqueda

El Método de Búsqueda

| Operator | Conditions | Actions |
|---------------------------|--|---|
| A_ParentOfC(X) | $X \notin Ad_G(C); Pa_G(C) \neq \emptyset$ | insert($X \rightarrow C, G$) |
| A_ChildOfC(X) | $X \notin Ad_G(C); Pa_G(C) \neq \emptyset$ | insert($C \rightarrow X, G$) |
| A_SiblingOfC(X) | $X \notin Ad_G(C); Pa_G(C) = \emptyset$ | insert($X - C, G$) |
| A_HHOfC(X, Y) | $X \notin Ad_G(C); Pa_G(C) = \emptyset; Y \in Sib_G(C)$ | insert($X \rightarrow C, G$) delete($Y - C, G$) insert($Y \rightarrow C, G$) $\forall Z \in Sib_G(C) \setminus \{Y\} \{$ delete($C - Z, G$) insert($C \rightarrow Z, G$) $\}$ |
| A_ParentOfChild(X, Y) | $X \notin Ad_G(Y); Y \in Ch_G(C)$ There is no directed path from Y to X in G | insert($X \rightarrow Y, G$) |
| A_HHOfChild(X, Y) | $X \notin Ad_G(Y); Y \in Sib_G(C)$ There is no directed path from Y to X in G | insert($X \rightarrow Y, G$) delete($C - Y, G$) insert($C \rightarrow Y, G$) |

El Método de Búsqueda

El Método de Búsqueda



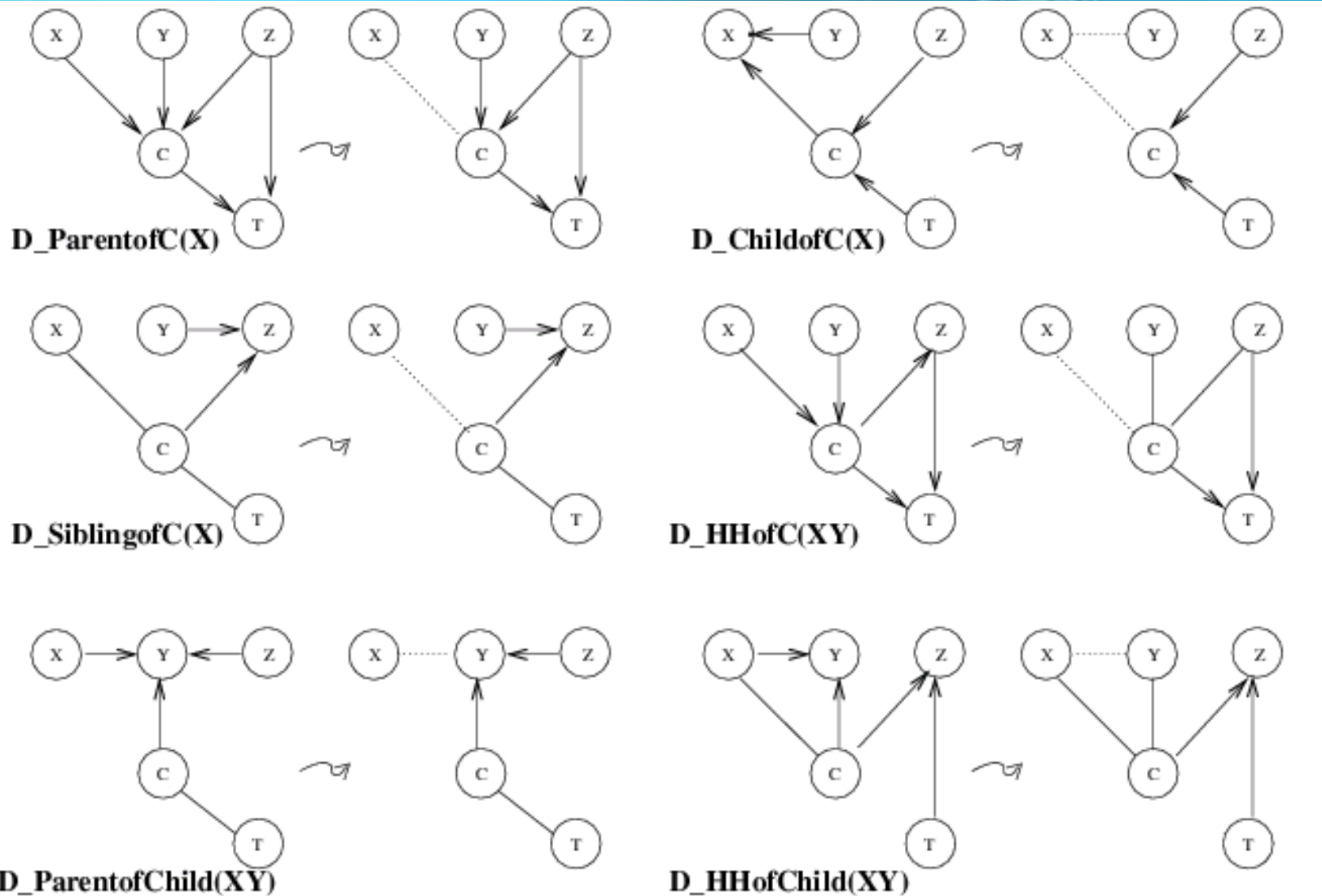
El Método de Búsqueda

El Método de Búsqueda

| Operator | Conditions | Actions |
|---------------------------|--|---|
| D_ParentOfC(X) | $X \in Pa_G(C); Pa_G(C) \geq 3$ | delete($X \rightarrow C, G$) |
| D_ChildOfC(X) | $X \in Ch_G(C)$ | $\forall Z \in Pa_G(X)$ delete($Z \rightarrow X, G$) |
| D_SiblingOfC(X) | $X \in Sib_G(C)$ | delete($X - C, G$) |
| D_HHOfC(X, Y) | $X \in Pa_G(C); Y \in Pa_G(C); Pa_G(C) = 2$ | delete($X \rightarrow C, G$) delete($Y \rightarrow C, G$) insert($Y - C, G$) $\forall Z \in Ch_G(C)$ if $Pa_G(Z) = \{C\}$ { delete($Z \rightarrow C, G$) insert($Z - C, G$) } |
| D_ParentOfChild(X, Y) | $X \in Pa_G(Y); Y \in Ch_G(C)$ $ Pa_G(Y) \geq 3$ or $Pa_G(C) \neq \emptyset$ | delete($X \rightarrow Y, G$) |
| D_HHOfChild(X, Y) | $X \in Pa_G(Y); Y \in Ch_G(C)$ $ Pa_G(Y) < 3; Pa_G(C) = \emptyset$ | delete($X \rightarrow Y, G$) delete($C \rightarrow Y, G$) insert($C - Y, G$) |





El Método de Búsqueda

El Método de Búsqueda



El Método de Búsqueda

Función de Evaluación

-  El método de búsqueda descrito se tiene que aplicar con una función de evaluación
-  Usamos una métrica no especializada en clasificación usada comúnmente en el aprendizaje de Redes Bayesianas (por ejemplo BIC, BDe, K2, etc.)
-  Por eficiencia la métrica usada debe ser descomponible. Así sólo tendremos que evaluar la parte que el operador modifica y no el grafo entero.
-  Sea G un C-RPDAG y G' cualquier C-RPDAG obtenido de aplicar los operadores descritos anteriormente. Sea g una métrica descomponible:

Función de Evaluación

Función de Evaluación

(a) If the operator is A_ParentOfC(X) then

$$g(G' : D) = g(G : D) - g_D(C, Pa_G(C)) + g_D(C, Pa_G(C) \cup \{X\})$$

(b) If the operator is A_ChildOfC(X) then

$$g(G' : D) = g(G : D) - g_D(X, \emptyset) + g_D(X, \{C\})$$

(c) If the operator is A_SiblingOfC(X) then

$$g(G' : D) = g(G : D) - g_D(C, \emptyset) + g_D(C, \{X\})$$

(d) If the operator is A_HHOfC(X, Y) then

$$g(G' : D) = g(G : D) - g_D(C, \{Y\}) + g_D(C, \{X, Y\})$$

(e) If the operator is A_ParentOfChild(X, Y) then

$$g(G' : D) = g(G : D) - g_D(Y, Pa_G(Y)) + g_D(Y, Pa_G(Y) \cup \{X\})$$

(f) If the operator is A_HHOfChild(X, Y) then

$$g(G' : D) = g(G : D) - g_D(Y, \{C\}) + g_D(Y, \{C, X\})$$

Función de Evaluación

Función de Evaluación

(g) If the operator is $D_ParentOfC(X)$ then

$$g(G' : D) = g(G : D) - g_D(C, Pa_G(C)) + g_D(C, Pa_G(C) \setminus \{X\})$$

(h) If the operator is $D_ChildOfC(X)$ then

$$g(G' : D) = g(G : D) - g_D(X, Pa_G(X)) + g_D(X, \emptyset)$$

(i) If the operator is $D_SiblingOfC(X)$ then

$$g(G' : D) = g(G : D) - g_D(C, \{X\}) + g_D(C, \emptyset)$$

(j) If the operator is $D_HHOfC(X, Y)$ then

$$g(G' : D) = g(G : D) - g_D(C, Pa_G(C)) + g_D(C, Pa_G(C) \setminus \{X\})$$

(k) If the operator is $D_ParentOfChild(X, Y)$ then

$$g(G' : D) = g(G : D) - g_D(Y, Pa_G(Y)) + g_D(Y, Pa_G(Y) \setminus \{X\})$$

(l) If the operator is $D_HHOfChild(X, Y)$ then

$$g(G' : D) = g(G : D) - g_D(Y, \{C, X\}) + g_D(Y, \{C\})$$

Función de Evaluación

Resultados Experimentales

- 31 Conjuntos de Datos de UCI (excepto mofn-3-7-10 y corral)
- Las variables continuas han sido discretizadas y los casos con valores perdidos han sido eliminados
- Se ha estudiado nuestro método contra NB, TAN, BAN y búsqueda local en un Red Bayesiana sin restricciones (UBN)
- BAN y UBN usan búsqueda local en el espacio de DAGs
- La métrica usada es BDeu
- Suavizado en el aprendizaje paramétrico

| Data Set | Instances | Attributes | Classes |
|---------------|-----------|------------|---------|
| adult | 45222 | 14 | 2 |
| australian | 690 | 14 | 2 |
| breast | 682 | 10 | 2 |
| car | 1728 | 6 | 4 |
| chess | 3196 | 36 | 2 |
| cleve | 296 | 13 | 2 |
| corral | 128 | 6 | 2 |
| crx | 653 | 15 | 2 |
| diabetes | 768 | 8 | 2 |
| DNA-nominal | 3186 | 60 | 3 |
| flare | 1066 | 10 | 2 |
| german | 1000 | 20 | 2 |
| glass | 214 | 9 | 7 |
| glass2 | 163 | 9 | 2 |
| heart | 270 | 13 | 2 |
| hepatitis | 80 | 19 | 2 |
| iris | 150 | 4 | 3 |
| letter | 20000 | 16 | 26 |
| lymphography | 148 | 18 | 4 |
| mofn-3-7-10 | 1324 | 10 | 2 |
| mushroom | 8124 | 22 | 2 |
| nursery | 12960 | 8 | 5 |
| pima | 768 | 8 | 2 |
| satimage | 6435 | 36 | 6 |
| segment | 2310 | 19 | 7 |
| shuttle-small | 5800 | 9 | 7 |
| soybean-large | 562 | 35 | 19 |
| splice | 3190 | 60 | 3 |
| vehicle | 846 | 18 | 4 |
| vote | 435 | 16 | 2 |
| waveform-21 | 5000 | 21 | 3 |

Resultados Experimentales

Resultados Experi

Media de 3 evaluaciones, en cada evaluación se ha realizado un 10-fold Cross Validation.

Mejor clasificador en negrita y peor en cursiva

C-RPDAG nunca es el peor y es el mejor mayor número de veces

| Data Set | NB | TAN | BAN | UBN | CRPDAG |
|---------------|----------------------|---------------------|----------------------|----------------------|----------------------|
| adult | <i>83.13 ± 0.52</i> | 85.25 ± 0.57 | 85.54 ± 0.55 | 85.43 ± 0.51 | 85.29 ± 0.45 |
| australian | 85.22 ± 3.60 | <i>84.93 ± 3.53</i> | 85.94 ± 3.96 | 85.75 ± 4.63 | 86.04 ± 4.25 |
| breast | 97.56 ± 1.50 | <i>95.55 ± 2.30</i> | 97.56 ± 1.71 | 97.65 ± 1.75 | 97.56 ± 1.71 |
| car | <i>85.40 ± 2.58</i> | 93.87 ± 1.96 | 93.09 ± 2.03 | 92.98 ± 2.10 | 93.07 ± 1.97 |
| chess | <i>87.84 ± 1.79</i> | 92.32 ± 1.59 | 96.88 ± 1.23 | 97.02 ± 1.03 | 96.53 ± 1.19 |
| cleve | 83.24 ± 6.87 | <i>80.88 ± 8.52</i> | 82.57 ± 7.52 | 83.57 ± 7.61 | 81.78 ± 7.56 |
| corral | <i>86.82 ± 9.09</i> | 99.49 ± 1.54 | 100.00 ± 0.00 | 100.00 ± 0.00 | 100.00 ± 0.00 |
| crx | 86.97 ± 3.94 | <i>86.26 ± 5.19</i> | 86.56 ± 4.36 | 86.71 ± 4.31 | 87.02 ± 3.70 |
| diabetes | <i>78.17 ± 3.75</i> | 78.91 ± 4.02 | 78.65 ± 3.70 | 79.09 ± 3.34 | 78.78 ± 3.74 |
| DNA-nominal | 95.40 ± 0.84 | <i>94.84 ± 0.86</i> | 95.31 ± 1.05 | 95.95 ± 1.14 | 96.15 ± 1.10 |
| flare | <i>80.40 ± 3.59</i> | 82.99 ± 3.16 | 82.93 ± 2.99 | 82.27 ± 2.99 | 83.27 ± 2.98 |
| german | 75.27 ± 3.43 | <i>72.70 ± 3.68</i> | 74.67 ± 3.41 | 73.80 ± 4.62 | 74.30 ± 4.07 |
| glass | 72.78 ± 8.02 | <i>68.59 ± 7.38</i> | 71.83 ± 8.58 | 69.06 ± 9.59 | 73.10 ± 7.51 |
| glass2 | <i>83.38 ± 8.79</i> | 85.04 ± 7.91 | 85.04 ± 7.99 | 84.82 ± 8.44 | 85.86 ± 7.81 |
| heart | 83.46 ± 8.45 | 82.84 ± 8.33 | <i>82.10 ± 8.41</i> | 83.33 ± 8.06 | 82.22 ± 8.16 |
| hepatitis | <i>85.00 ± 10.38</i> | 90.00 ± 8.64 | 88.75 ± 8.76 | 89.58 ± 11.06 | 88.75 ± 9.47 |
| iris | <i>94.22 ± 4.77</i> | <i>93.33 ± 5.43</i> | 94.44 ± 4.55 | 94.89 ± 5.27 | 94.67 ± 4.64 |
| letter | <i>74.01 ± 1.00</i> | 85.86 ± 0.79 | 84.35 ± 0.70 | 84.43 ± 0.85 | 85.08 ± 0.75 |
| lymphography | 84.41 ± 10.20 | 81.95 ± 10.60 | 84.43 ± 10.29 | <i>79.08 ± 11.85</i> | 81.54 ± 11.07 |
| mofn-3-7-10 | <i>85.40 ± 3.50</i> | 90.69 ± 2.26 | 87.29 ± 3.55 | 99.17 ± 2.48 | 99.50 ± 1.15 |
| mushroom | <i>95.38 ± 0.84</i> | 99.98 ± 0.06 | 100.00 ± 0.00 | 100.00 ± 0.00 | 100.00 ± 0.00 |
| nursery | <i>90.26 ± 0.80</i> | 92.28 ± 0.85 | 91.82 ± 0.81 | 93.12 ± 0.81 | 93.52 ± 0.61 |
| pima | <i>77.91 ± 4.75</i> | 78.78 ± 4.56 | 78.03 ± 4.56 | 78.78 ± 4.64 | 78.03 ± 4.54 |
| satimage | <i>82.43 ± 1.03</i> | 88.52 ± 0.96 | 88.29 ± 0.98 | 83.49 ± 1.72 | 88.19 ± 1.12 |
| segment | <i>92.16 ± 1.69</i> | 95.02 ± 1.26 | 94.34 ± 1.28 | 94.56 ± 1.29 | 94.34 ± 1.29 |
| shuttle-small | <i>99.01 ± 0.37</i> | 99.67 ± 0.25 | 99.72 ± 0.18 | 99.69 ± 0.17 | 99.60 ± 0.23 |
| soybean-large | 91.70 ± 3.56 | <i>86.11 ± 4.16</i> | 92.53 ± 3.37 | 88.55 ± 3.96 | 90.33 ± 3.25 |
| splice | 95.45 ± 1.08 | <i>94.90 ± 0.98</i> | 95.28 ± 1.08 | 95.73 ± 1.11 | 96.27 ± 1.13 |
| vehicle | <i>62.85 ± 4.79</i> | 70.73 ± 3.99 | 71.56 ± 3.69 | 65.13 ± 4.99 | 71.91 ± 3.91 |
| vote | <i>90.04 ± 5.10</i> | 94.57 ± 3.02 | 94.34 ± 3.16 | 94.57 ± 3.05 | 93.49 ± 3.22 |
| waveform-21 | 81.82 ± 1.32 | 82.89 ± 1.26 | 82.36 ± 1.67 | <i>79.81 ± 1.76</i> | 82.45 ± 1.67 |
| Times best | 2 | 8 | 6 | 9 | 12 |
| Times worst | 18 | 10 | 1 | 2 | 0 |

Resu

Resultados Experimentales

En esta tabla comparamos cada método con los demás.

En cada fila tenemos cuantas veces es mejor un algoritmo respecto a los demás

En cada columna tenemos cuantas veces es peor un algoritmo

| | | NB | TAN | BAN | UBN | CRPDAG | |
|--------|--------|----|-----|-----|-----|--------|----|
| Better | NB | – | 12 | 7 | 7 | 5 | 31 |
| | TAN | 19 | – | 13 | 10 | 12 | 54 |
| | BAN | 23 | 17 | – | 13 | 10 | 63 |
| | UBN | 24 | 19 | 16 | – | 12 | 71 |
| | CRPDAG | 25 | 19 | 15 | 17 | – | 76 |
| | | 91 | 67 | 51 | 47 | 39 | |

Resultados Experimentales

Resultados Experimentales

- En esta tabla tenemos cuando nuestro método es mejor o peor que el resto y, además, este resultado es significativo.
- Para ello se ha utilizado el Test de rangos con signos de Wilcoxon para muestras pareadas.
- El número de veces que es mejor significativamente es 36 frente al número de veces que es peor es 9

| | NB | TAN | BAN | UBN | Total |
|------------------|----|-----|-----|-----|-------|
| CRPDAG is better | 16 | 9 | 5 | 6 | 36 |
| CRPDAG is worse | 2 | 3 | 3 | 1 | 9 |

Resultados Experimentales

Resultados Experimentales

- Hemos observado que UBN obtiene mejores resultados que NB y TAN
- Para determinar la influencia de la métrica se ha hecho una comparativa de nuestro método usando las mátricas BDeu y BIC
- BIC es peor

| Data Set | BIC | BDeu |
|------------------------|----------------------|----------------------|
| adult | 85.42 ± 0.48* | 85.29 ± 0.45 |
| australian | 86.28 ± 3.57 | 86.04 ± 4.25 |
| breast | 97.56 ± 1.71 | 97.56 ± 1.71 |
| car | 85.63 ± 2.69 | 93.07 ± 1.97* |
| chess | 95.81 ± 1.24 | 96.53 ± 1.19* |
| cleve | 82.46 ± 7.61 | 81.78 ± 7.56 |
| corral | 100.00 ± 0.00 | 100.00 ± 0.00 |
| crx | 86.62 ± 3.67 | 87.02 ± 3.70 |
| diabetes | 78.57 ± 3.44 | 78.78 ± 3.74 |
| DNA-nominal | 96.33 ± 1.04* | 96.15 ± 1.10 |
| flare | 82.77 ± 2.83 | 83.27 ± 2.98 |
| german | 74.40 ± 4.46 | 74.30 ± 4.07 |
| glass | 70.12 ± 7.26 | 73.10 ± 7.51 |
| glass2 | 84.83 ± 8.42 | 85.86 ± 7.81 |
| heart | 82.59 ± 8.25 | 82.22 ± 8.16 |
| hepatitis | 87.50 ± 10.11 | 88.75 ± 9.47 |
| iris | 94.22 ± 4.77 | 94.67 ± 4.64 |
| letter | 76.73 ± 0.95 | 85.08 ± 0.75* |
| lymphography | 81.78 ± 10.56 | 81.54 ± 11.07 |
| mofn-3-7-10 | 93.56 ± 1.88 | 99.50 ± 1.15* |
| mushroom | 100.00 ± 0.00 | 100.00 ± 0.00 |
| nursery | 91.30 ± 0.76 | 93.52 ± 0.61* |
| pima | 78.51 ± 4.51 | 78.03 ± 4.54 |
| satimage | 84.57 ± 1.12 | 88.19 ± 1.12* |
| segment | 92.17 ± 1.69 | 94.34 ± 1.29* |
| shuttle-small | 99.79 ± 0.17* | 99.60 ± 0.23 |
| soybean-large | 91.70 ± 3.56 | 90.33 ± 3.25 |
| splice | 96.30 ± 1.05 | 96.27 ± 1.13 |
| vehicle | 71.75 ± 3.80 | 71.91 ± 3.91 |
| vote | 92.96 ± 4.17 | 93.49 ± 3.22 |
| waveform-21 | 82.47 ± 1.58 | 82.45 ± 1.67 |
| Times better | 12 | 16 |
| Times signific. better | 3 | 7 |

Resultados

Conclusiones



Especializando el algoritmo de búsqueda en un esquema de métrica+búsqueda conseguimos buenos resultados sin usar una métrica especializada.



La búsqueda basada en C-RPDAGs nos permiten trabajar con un espacio de búsqueda asequible y obtenemos redes no restringidas.



La utilización de una métrica u otra influye en la construcción de la red como clasificador.



Sería interesante utilizar un algoritmo de búsqueda más potente en el espacio de C-RPDAGs

Conclusiones

Clases de Elvira



Paquete

elvira.learning.classification.supervised.discrete



Clases:

- **MarkovBlanketLearning.java**: Abstracta, hereda de Learning.java e implementa el interfaz Classifier.java
- **LocalSearchLearning.java**
- **BANLearning.java**
- **CRPDAGLearninig.java**

Clases de Elvira