

# Aprendizaje de intervalos para Redes Bayesianas de Nodos Temporales

Pablo F. Hernández Leal

Instituto Nacional de Astrofísica, Óptica y Electrónica  
Coordinación de Ciencias Computacionales

**Resumen** En este trabajo se presenta un algoritmo para realizar el aprendizaje de intervalos en Redes Bayesianas de Nodos Temporales. Para ello se usa el algoritmo EM para obtener diferentes clusters que servirán de base para obtener los intervalos. Posteriormente se usa información de la red para refinar los intervalos. Se realizó la evaluación de los intervalos para una red bayesiana de 5 nodos usando el score relativo de Brier obteniendo buenos resultados para varios conjuntos de datos.

## 1. Introducción

Las Redes Bayesianas de Nodos Temporales (RBNT) son una extensión de las redes bayesianas y fueron creadas para representar procesos temporales de forma más sencilla y compacta que otras representaciones y son diferentes de las redes Bayesianas Dinámicas. En particular las RBNT se encuentran dentro de un grupo llamado Redes Bayesianas de Eventos[3]. En las RBNT cada nodo es llamado nodo temporal y puede representar intervalos de tiempo en donde la variable cambia de estado. Estos intervalos pueden ser diferentes en número y en tamaño, lo cual incrementa su granularidad. El trabajo se presenta como sigue, en la sección 2 se aborda el problema y las suposiciones iniciales, en la sección 3 se explica paso a paso el algoritmo propuesto, en la sección 4 se realizan las evaluaciones y pruebas y en la sección 5 se habla de las conclusiones y trabajo futuro.

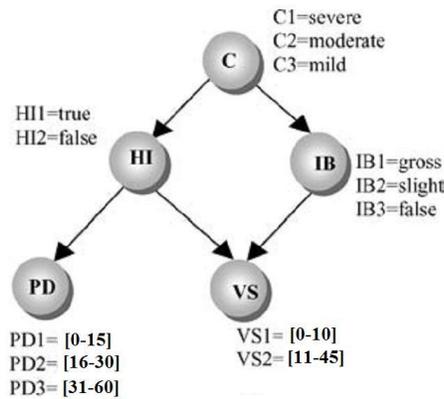
## 2. Problema

Las RBNT han sido probadas con éxito en los sistemas de plantas eléctricas[1], sin embargo, el problema que se tiene actualmente es que los intervalos asociados a cada nodo temporal son obtenidos de manera no automática, es decir con ayuda de un experto, lo que hace la tarea muy complicada, tardada y propensa a errores. Por ello se propone obtener un algoritmo de aprendizaje de intervalos para nodos temporales. Como una primera aproximación se tendrán las siguientes suposiciones:

- Se conoce la topología de la red.
- Se tienen datos suficientes para realizar el aprendizaje

### 3. Algoritmo Propuesto

Como suposición extra diremos que los eventos siguen alguna distribución conocida (en particular se realizaron pruebas para las distribuciones gaussiana y uniforme). Con esta suposición podemos usar la idea siguiente: usar un algoritmo de clustering, que nos divida ciertos grupos los cuales van a corresponder a ciertos intervalos. A continuación vamos a detallar un poco más esta idea. Para ejemplificar el proceso se hará uso de la red que se muestra en la figura 1



**Figura 1.** Un ejemplo de una Red Bayesiana de Nodos Temporales

#### 3.1. Encontrando intervalos iniciales

Con la suposición hecha antes, podemos usar el modelo de mezcla de gaussianas para aproximar los datos y por ello se propone usar el algoritmo EM (Expectation-Maximization)[2] el cual nos da como resultado los valores de las medias y las varianzas de los clusters. Por ahora supongamos que conocemos el número de clusters. Se propone que el intervalo se encuentre determinado por [Media menos Desv. Est. - Media más Desv. Est.], un ejemplo se muestra en la figura 2.

#### 3.2. Encontrando número de clusters

Ahora vamos a resolver el problema de seleccionar el número de clusters que se dará al algoritmo EM. Sabemos que la solución buscada debe de cumplir dos condiciones.

- Deben de ser pocos intervalos para que la complejidad de la red bayesiana no crezca demasiado.

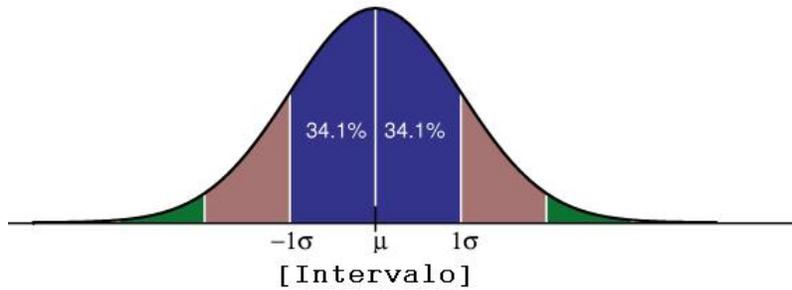


Figura 2. Definición del intervalo inicial

- Deben de ser intervalos que cuando se realicen procesos de diagnóstico o predicción sobre la red completa, los resultados sean buenos.

Basándonos en estas dos razones obtenemos nuestra solución propuesta; ya que buscamos pocos intervalos se propone usar el algoritmo EM con valores de  $k$  de 1 hasta un valor máximo, para este trabajo se uso  $k$  de 1 a 3 . Además para buscar el mejor conjunto de intervalos se hará uso de una evaluación sobre la red aplicando el score relativo de Brier, el cual mide el error de los valores marginales de cada nodo dada cierta evidencia.

### 3.3. Usando información de la red

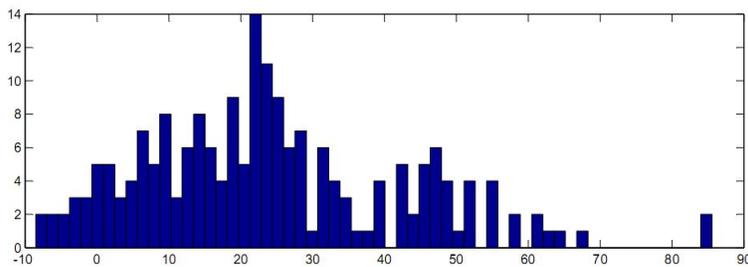
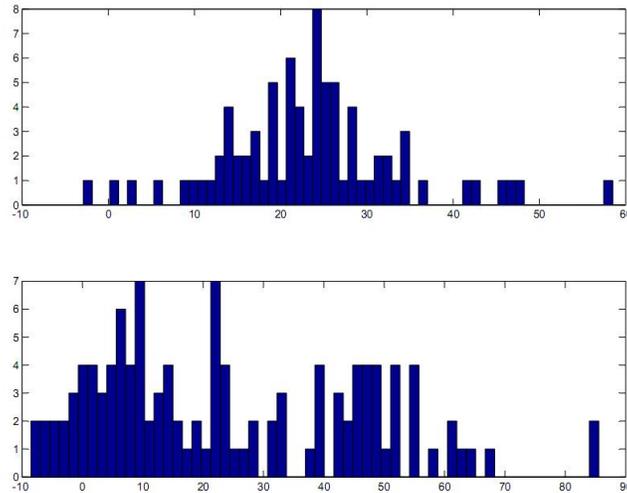


Figura 3. Histograma de eventos para un nodo temporal

En la figura 3 se muestra el histograma de eventos para el nodo PD de la red de la figura 1, de forma subjetiva se puede decir que existen 2 clusters definidos en el histograma. Sin embargo, este histograma muestra eventos que sucedieron en 3 intervalos. Por ello se propone usar información de la red bayesiana para mejorar nuestros intervalos obtenidos. Ya que conocemos la topología de la red

y conocemos los valores del padre de cada nodo temporal, podemos dividir los datos de los eventos de cada nodo dependiendo del valor del padre. Para el nodo PD de la red de ejemplo, sabemos que el padre(HI) tiene los valores `true` y `false`, por lo que tendremos dos conjuntos de datos para el nodo temporal PD, en la figura 4 se muestran los histogramas para cada conjunto de datos. En el histograma superior se puede apreciar un solo cluster definido, y en el histograma inferior, dos. Lo cual nos da los 3 intervalos originales de la red.



**Figura 4.** Histogramas de eventos para un nodo temporal, cada histograma representa un valor del padre del nodo

### 3.4. Uniendo intervalos

Ya que vamos a aplicar el algoritmo EM por cada valor del padre del nodo temporal, tendremos varios intervalos que posiblemente pueden traslaparse o quedar distanciados. La solución que se propone es obtener todas las combinaciones de los intervalos obtenidos por cada valor del padre. Un ejemplo se muestra en la figura 5. Posteriormente, para cada combinación se ordenan los intervalos de acuerdo a sus valores iniciales. Finalmente se reajustan los valores, tomando la media del valor final de cada intervalo más el valor inicial del siguiente. Este proceso se muestra de forma gráfica en la figura 6

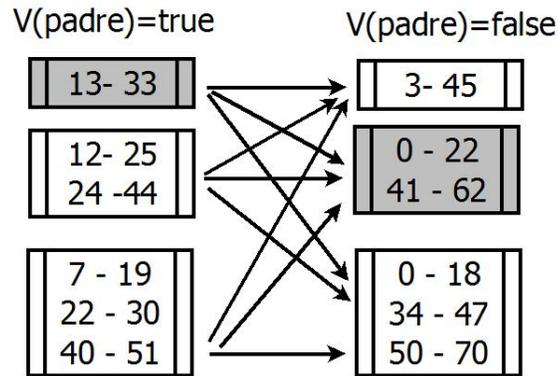


Figura 5. Combinar intervalos dados por cada valor del padre del nodo temporal

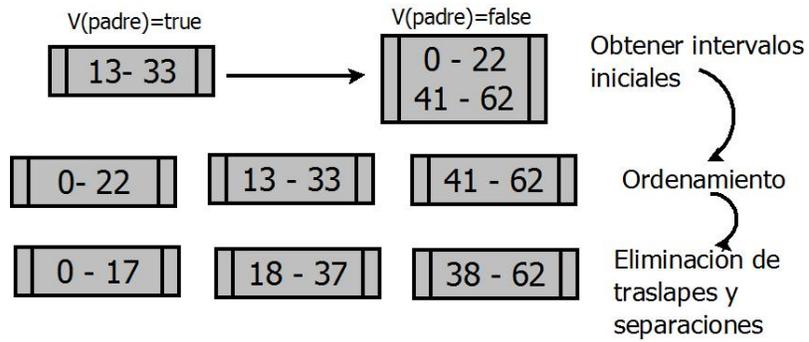


Figura 6. Proceso para obtener intervalos finales

#### 4. Evaluación

El score de Brier se define como

$$BS = \sum_{i=1}^n (1 - P_i)^2$$

Donde  $P_i$  es la probabilidad posterior marginal del valor correcto de cada nodo. El valor máximo del score de brier es  $BS_{max} = \sum^n 1^2$ , con estas dos fórmulas se define un score relativo

$$RBS = 1 - (BS/BS_{max}) \times 100$$

donde un valor de 100 correspondería a cero errores.

Se realizaron 3 tipos de pruebas para la red de ejemplo:

- Diagnóstico. Se instancia el nodo C y se predice el valor del resto.
- Predicción. Se instancian los nodos PD y VS y se predice el valor del resto.
- Predicción y diagnóstico. Se instancian los nodos HI e IB y se predice el valor del resto.

Intervalos reales Nodo PD	Intervalos reales Nodo VS	Evaluación usando RBS		
Intervalos encontrados	Intervalos encontrados	Pred.	Diag.	Pred. y Diag.
[0 – 15][16 – 30][30 – 60]	[0 – 10][11 – 45]			<b>61.68</b>
[2 – 17] [18 – 43] [44 – 64]	[0 – 13] [14 – 37] [38 – 53]	<b>62.21</b>	<b>71</b>	<b>61.68</b>
[2 – 17] [18 – 36] [37 – 47] [48 – 68]	[0 – 13] [14 – 37] [38 – 53]	60.78	69.53	58.44
[2 – 17] [18 – 36] [37 – 47] [48 – 68]	[1 – 13] [14 – 38] [39 – 48]	61.13	71.72	61.56
[2 – 14] [15 – 33] [34 – 51] [52 – 64]	[1 – 13] [14 – 38] [39 – 48]	60.44	71.32	60.22
[2 – 15] [16 – 25] [26 – 39] [40 – 45] [46 – 52] [53 – 68]	[0 – 2] [3 – 9] [10 – 25] [26 – 32] [33 – 43] [44 – 54]	55.32	70.86	53.09

**Tabla 1.** Algunos intervalos obtenidos para 200 datos totales y su evaluación usando el score de Brier

La primera evaluación fue con 200 datos, 70% para entrenamiento y el resto para prueba, algunos resultados se muestran en la tabla 1. Es importante observar cómo los mejores resultados se obtuvieron con intervalos que son muy parecidos a los reales. Además cuando el número de intervalos aumenta, los resultados fueron disminuyendo en calidad.

La siguiente prueba fue similar a la anterior sólo que con 300 datos. Estos resultados se muestran en la tabla 2. De lo anterior se puede comprobar que cuantos más datos, más certeros son los intervalos generados.

Posteriormente se realizaron otros experimentos. Se probó al modelo para ver como afectaba que los datos siguieran una distribución uniforme. Los resultados son buenos, ya que las evaluaciones tienen casi los mismos resultados que

Intervalos reales Nodo PD	Intervalos reales Nodo VS	Evaluación usando RBS		
Intervalos encontrados	Intervalos encontrados	Pred.	Diag.	Pred. y Diag.
[0 – 15][16 – 30][30 – 60]	[0 – 10][11 – 45]			
[0 – 17][18 – 37][38 – 63]	[0 – 19][20 – 40]	<b>64.21</b>	<b>71.33</b>	<b>62.99</b>
[0 – 17][18 – 35][36 – 48][49 – 71]	[0 – 19][20 – 40]	63.66	71.88	61.04
[0 – 14][15 – 26][27 – 40][41 – 63]	[0 – 19][20 – 40]	62.68	72.28	60.84
[0 – 10][11 – 20][21 – 34][35 – 45][46 – 63]	[0 – 19][20 – 40]	62.61	72.97	61.17
[0 – 13][14 – 20][21 – 34][35 – 42][43 – 48][49 – 71]	[0 – 9][10 – 22][23 – 35][36 – 53]	57.34	67.10	52.15

**Tabla 2.** Algunos intervalos obtenidos con 300 datos totales y su evaluación usando el score de Brier

Intervalos reales Nodo PD	Intervalos reales Nodo VS	Evaluación usando RBS		
Intervalos encontrados	Intervalos encontrados	Pred.	Diag.	Pred. y Diag.
[0 – 15][16 – 30][30 – 60]	[0 – 10][11 – 45]			
[2 – 11][12 – 37][38 – 54]	[6 – 18][19 – 38]	<b>62.97</b>	<b>71.22</b>	<b>63.20</b>
[2 – 11][12 – 37][38 – 54]	[4 – 11][12 – 22][23 – 40]	62.60	73.79	62.04
[2 – 12][13 – 31][32 – 45][46 – 55]	[6 – 18][19 – 38]	62.10	70.98	62.53
[2 – 10][11 – 22][23 – 30][31 – 38][39 – 48][49 – 55]	[2 – 8][9 – 17][18 – 25][26 – 30][31 – 41]	57.72	73.70	55.44

**Tabla 3.** Algunos intervalos obtenidos con 200 datos totales con distribución uniforme y su evaluación usando el score de Brier

Intervalos reales Nodo PD	Intervalos reales Nodo VS	Evaluación usando RBS		
Intervalos encontrados	Intervalos encontrados	Pred.	Diag.	Pred. y Diag.
[0 – 15][16 – 30][30 – 60]	[0 – 10][11 – 45]			
[3 – 13][14 – 34][35 – 53]	[1 – 17][18 – 35]	<b>63.56</b>	<b>74.67</b>	<b>64.34</b>
[4 – 13][14 – 31][32 – 38][39 – 53]	[1 – 17][18 – 35]	64.99	76.20	65.21
[3 – 13][14 – 30][31 – 46][47 – 55]	[1 – 17][18 – 35]	65.01	76.95	64.90
[3 – 12][13 – 23][24 – 30][31 – 44][45 – 56]	[2 – 7][8 – 15][16 – 26][27 – 32][33 – 39]	56.74	71.62	50.94

**Tabla 4.** Algunos intervalos obtenidos con 300 datos totales con distribución uniforme y su evaluación usando el score de Brier

cuando los datos seguían una distribución normal. En la tabla 3 se muestran los resultados con 200 datos, el 70 % se uso para entrenamiento y el resto para prueba. Finalmente en la tabla 4, se presentan los resultados para 300 datos.

Para esta red que se usó como ejemplo, se puede observar que los intervalos encontrados tuvieron mejores resultados cuando se realizaba Diagnóstico en la red, es decir cuando se instanciaban los nodos hoja y se observaban los nodos superiores.

## 5. Resultados y Alcances

Se presentó un algoritmo para obtener intervalos para Redes Bayesianas de Nodos Temporales. El algoritmo se basa en ideas de clustering y además usa información de la topología de la red. Se probó el algoritmo con una red simple de 5 nodos y se obtuvieron resultados en base al score de Brier con valores entre 63 y 73 de calidad.

Como trabajo futuro se plantea usar este algoritmo en un dominio de plantas eléctricas o un dominio médico y evaluar los resultados. Además se tiene que resolver como realizar el aprendizaje cuando existe dependencia entre dos nodos que tengan intervalos. Finalmente se buscará una medida tipo MDL para manejar el error y los intervalos y la complejidad de ellos (su número) para elegir el mejor conjunto de intervalos.

## Referencias

1. G. Arroyo-Figueroa and L.E. Sucar. A temporal Bayesian network for diagnosis and prediction. In *Proceedings of the 15th Annual Conference on Uncertainty in Artificial Intelligence (UAI-99)*, pages 13–22, 1999.
2. A.P. Dempster, N.M. Laird, D.B. Rubin, et al. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39(1):1–38, 1977.
3. S. F. Galán, G. Arroyo-Figueroa, F. J. Díez, and L. E. Sucar. Comparison of two types of event bayesian networks: A case study. *Appl. Artif. Intell.*, 21(3):185–209, 2007.