

USO DE REDES BAYESIANAS OBTENIDAS MEDIANTE OPTIMIZACIÓN DE ENJAMBRE DE PARTÍCULAS PARA EL DIAGNÓSTICO DE LA HIPERTENSIÓN ARTERIAL

María del Carmen Chávez, Gladys Casas, Jorge Moreira, Emilio González, Rafael Bello, Ricardo Grau
Centro de Estudios de Informática, Universidad Central "Martha Abreu" de Las Villas, Santa Clara,
Villa Clara, Cuba, CP 54830

RESUMEN

En el presente trabajo se combinan diferentes técnicas de Inteligencia Artificial para modelar el diagnóstico de la Hipertensión arterial. La base de datos utilizada es el resultado de un estudio realizado con personas supuestamente sanas, en cinco policlínicos de la ciudad de Santa Clara. Una forma de modelar las relaciones entre las variables es usando una red del bayesiana. El costo computacional del aprendizaje de la estructura de una red bayesiana desde datos, crece con el número de variables y el número de casos. Por lo que surge el problema de identificar una buena heurística para explorar el espacio de posibles redes. Los algoritmos evolutivos son métodos muy valiosos para encontrar buenas soluciones en problemas concretos, por lo que se usa el algoritmo de Optimización de Enjambre de Partícula (PSO) para la búsqueda de la estructura de la red bayesiana. Se ha realizado una extensión a la plataforma Weka (*Waikato for Environment Knowledge Analysis*) y se utilizan métricas de score global para evaluar las soluciones. El nuevo algoritmo forma parte de la clase *Bayesnet* de Weka. Los resultados obtenidos muestran buenos resultados en la clasificación de la Hipertensión Arterial.

ABSTRACT

In the present work, different Artificial Intelligence techniques are combined to model the diagnosis of hypertensive people. To develop the work a data base of Arterial Hypertension was used, which is result of a preliminary study made in five polyclinics of Santa Clara city, with supposedly healthy individuals. One of the ways to model the relations between variables is using a Bayesian network. The computational cost of the learning of a Bayesian network from data, grows with the number of variables and the number of cases, therefore, the problem of identifying a good heuristic to explore the space of possible networks arises. The evolutionary algorithms are being very valuable methods to find good solutions to concrete problems, that is why the Particle Swarm Optimization (PSO) algorithm is used for the network structure search. An extension to the Weak platform (*Waikato for Environment Knowledge Analysis*) was done, in which the new algorithm becomes part of the global score metrics implemented in the *Bayesnet* class in Weka. The obtained results show good classification of the Arterial Hypertension with Bayesian networks.

KEY WORDS: Bayesian networks, classification, PSO, Particle Swarm optimization, quality metric of optimization Bayesian networks, algorithms bio inspired, arterial high blood pressure.

MSC: 62P10

1. INTRODUCCIÓN

El corazón bombea sangre a través de las arterias a todo el cuerpo, la tensión que se genera en el interior arterial se denomina presión arterial. La hipertensión arterial (HTA) ó presión alta, es la elevación de esta presión arriba de un límite que se considera normal (*140/90 mmHg*). La HTA es la principal enfermedad crónica degenerativa y la más común causa de muerte, afecta aproximadamente al 20% de la población mundial. La elevación anormal de la presión constituye un importante factor de riesgo coronario.

Al medirse la presión arterial se anotan dos números, el mayor es la presión sistólica, y se corresponde con la presión del corazón al contraerse para bombear la sangre. El número menor es la presión diastólica, que es la presión de la sangre en las arterias en la fase de relajamiento del corazón. Para un correcto diagnóstico de hipertensión, el médico mide varias veces la presión arterial, en diferentes

condiciones de esfuerzo y en diferentes horas del día. En personas hipertensas, la variación es mayor y permanece alta la mayor parte del día, incluso en los periodos de descanso. El conocimiento actual de éste problema de salud pública a nivel mundial, obliga a buscar estrategias de detección, control y tratamiento [1], [2], [3].

El estado de hiperreactividad vascular se consiguió mediante una ergometría isométrica denominada Prueba del Peso Sostenido (PPS) [4]. Esta prueba basa su principio en introducir al método clásico de la medición de la tensión arterial la condición de que los pacientes realicen, en posición sentada, un ejercicio físico isométrico, que consiste en mantener un peso de 500 gramos con el brazo izquierdo extendido en ángulo recto al cuerpo durante 2 minutos. La presión arterial se toma en el brazo contrario antes del ejercicio y a partir del segundo 50 del segundo minuto.

Las redes bayesianas (RB) son una herramienta poderosa de representación del conocimiento. Una RB es un grafo acíclico dirigido (GAD) con una distribución de probabilidad asociada a cada nodo. Los nodos en la red representan las variables, atributos o rasgos del dominio de aplicación, y los arcos entre los nodos representan las relaciones de dependencia entre las variables [6].

Encontrar un modelo de RB consta de dos partes fundamentales, determinar la estructura de la red y obtener las tablas de probabilidades asociadas a cada nodo.

La búsqueda de una estructura adecuada puede interpretarse como un problema de optimización, se trata de hallar la red de mejor calidad en el espacio de posibles redes, donde la calidad puede medirse por una métrica que evalúa la red de acuerdo a los datos de partida. Existen varias métricas que evalúan la calidad de las redes, específicamente con enfoque bayesiano, K2, basados en criterios de información o entropía, Akaike Information Criterion, Minimum Description Length [5]. En el trabajo se han realizado las pruebas con la métrica de Bayes, basado en los análisis realizados por Bouckaert en [5].

El alto costo en tiempo y recursos inherentes a los algoritmos exactos de búsqueda, ha conllevado al auge y desarrollo de las heurísticas y metaheurísticas cuyo uso ha arrojado resultados muy alentadores. Dentro de ellas, los algoritmos bioinspirados y en particular la Inteligencia de Enjambre (Swarm Intelligence, SI) ha sido objeto de estudio, investigación y aplicación por su simplicidad y robustez [14], [16], [17], [15]. Nos centraremos ahora en hallar la estructura de la red, en dicha búsqueda se empleará el modelo computacional: Optimización en Enjambre de Partículas (Particle Swarm Optimization, PSO) [14], [16], [17], [15]. En [8] se hizo un primer intento usando métricas de score local y los resultados para problemas de clasificación inicialmente no fueron satisfactorios.

En el presente trabajo se mejora la eficiencia de los modelos obtenidos pues se utilizan métricas de score global, o sea se mejora el desempeño de la red teniendo en cuenta la muestra y la exactitud de la clasificación. La validación de los resultados se mide haciendo validaciones cruzadas con diez subconjuntos, que es el estándar para esta tarea.

2. REDES BAYESIANAS

Una red bayesiana (RB) es un par (D,P), donde D es un grafo acíclico dirigido (GAD), $P = \{p(x_1|\tau_1), \dots, p(x_n|\tau_n)\}$ es un conjunto de n distribuciones de probabilidad condicionales, una por cada variable x_i (nodos del grafo), y τ_i es el conjunto de padres del nodo x_i en D. El conjunto P define la distribución de probabilidad conjunta asociada, como muestra la ecuación (1):

$$p(x) = \prod_{i=1}^n p(x_i|\tau_i) \quad x = (x_1, x_2, \dots, x_n) \quad (1)$$

Uno de los problemas que debe resolverse cuando se utilizan las RB, es la búsqueda de la estructura que mejor se ajuste a los datos. Esta tarea tiene implícito dos pasos:

- Aprendizaje estructural: obtener la estructura de la RB, es decir, las relaciones de dependencia e independencia entre las variables involucradas
- Aprendizaje paramétrico: dada una estructura de RB, obtener las probabilidades a priori y condicionales requeridas

En este trabajo se propone un algoritmo para la obtención desde datos de la estructura de la RB, mediante la utilización de un algoritmo de inteligencia colectiva.

Las RB son un tipo especial de sistema basado en el conocimiento, por lo que es posible hacer inferencia a partir de conocimiento a priori. A este proceso se le llama propagación de evidencias [6].

3. PSO PRINCIPIOS BÁSICOS

PSO es una metaheurística de optimización estocástica basada en una población. Un enjambre (*swarm*) se define como una colección estructurada de organismos (agentes) que interactúan. La inteligencia no está en los individuos sino en la acción de todo el colectivo. Cada organismo (partícula) se trata como un punto en un espacio N dimensional el cual ajusta su propio “vuelo” de acuerdo a su propia experiencia y la experiencia del resto de la banda. La banda (*swarm*) “vuela” por el espacio de búsqueda localizando regiones o partículas prometedoras [17].

3.1 PSO en el aprendizaje de redes bayesianas

La búsqueda de la estructura de la red puede formularse como un problema de optimización en el espacio de las posibles redes Ω , en otras palabras, determinar $X_{\text{ópt}} \in \Omega$:

$H(X_{\text{ópt}}) \geq H(X_i), \forall X_i \in \Omega$, donde la función objetivo H considerada es una métrica que evalúa la RB en cada partícula del espacio de búsqueda. Se pueden usar cualesquiera de las métricas de score local descritas en el capítulo 4 de la tesis de Bouckaert [5], las cuales se encuentran ya implementadas en la plataforma Weka, en el software que se han obtenido los resultados.

En la modelación del problema de búsqueda a partir del algoritmo PSO se define cada partícula como una red bayesiana la cual se representa como una matriz de adyacencias $B = [B_{ij}]$ donde $b_{ij} = 1$ si el atributo i es padre del atributo j, (si existe un arco de i a j) y $b_{ij} = 0$ en otro caso. Por tanto el espacio

de búsqueda Ω tiene cardinal 2^{n^2} , de hecho se puede trabajar con dicho espacio, pero habría que chequear que no existan ciclos. Esto se puede lograr, por ejemplo, eliminando de forma aleatoria arcos que formen parte de ciclos existentes [8]. Se propone entonces una forma de generar el espacio de búsqueda garantizando que no existan ciclos, o sea, partiendo de que un grafo dirigido representa un ordenamiento topológico, si y sólo si este no presenta ciclos, es posible a partir de una permutación formar un grafo acíclico dirigido [8].

El problema de optimización que se propone es binario, por lo que el algoritmo PSO original [19], [18] debe ser adaptado.

X_i : Partícula (matriz del espacio Ω), $\{X_1, X_2, \dots\}$: Bandada (conjunto de partículas), $\{V_1, V_2, \dots\}$: Velocidades (matrices del espacio Ω asociadas a cada partícula que indican su movimiento), $\{XpBest_1, XpBest_2, \dots\}$: Mejores puntos del espacio localizados por cada partícula, $XgBest$: Mejor punto localizado por la bandada.

En el algoritmo *Inicializar valores* asigna aleatoriamente valores a la población de X_i y V_i , $XpBest_i$ de cada partícula como copia de X_i y $XgBest$ con el mejor valor, *CantGeneraciones* es la cantidad de “generaciones” en que van a interactuar las partículas y *cantPart* es la cantidad de partículas que van a existir en cada generación. La variable t se utiliza como contador de generaciones. La subrutina *Generar red acíclica* G_π genera una red acíclica como se vio en [8] a partir de una permutación aleatoria π de $(1, 2, \dots, n)$ con distribución uniforme, dicha red se representa como una matriz de adyacencia. La actualización de las partículas se logra añadiendo la velocidad a cada partícula obtenida en la iteración t , la velocidad se obtiene utilizando las expresiones que se muestran en la ecuación 2.

$$V_i = w V_i + c_1 \text{rand} (XpBest_i - X_i) + c_2 \text{rand} (XgBest - X_i) \quad (2)$$

$$S(V_{ijk}) = \frac{1}{1 + e^{(-V_{ijk})}}$$

En la expresión de V_i , el primer término es la memoria de la partícula, el segundo la parte cognitiva o conocimiento privado, el tercero la parte social que permite la colaboración, w es el peso de inercia, c_1 y c_2 son los llamados factores de aprendizaje cognitivo y social respectivamente y rand es un número aleatorio entre 0 y 1.

Entre los valores recomendados en [8], se tomaron los valores $c_1 = c_2 = 2$, pero en realidad se recomienda en el trabajo que c_1 y c_2 no tomen necesariamente el mismo valor sino, que se generen aleatoriamente con distribución en el intervalo [0,2], $w = 0.5 + \text{rand} / 2$. Otras funciones para la transformación, $S : \mathfrak{R} \rightarrow [0,1]$ aparecen en [19], [12], [13].

3.2 Algoritmo PSO

```

Inicializar valores;
t = 0;
Repeat
  Generar red acíclica  $G_\pi$ 
  For each  $i = \overline{1, cantPart}$ 
    Calcular  $V_i(t+1)$  y limitarla a  $[-Vmax, +Vmax]$ 
    Calcular  $S(V_i)$ 
    Actualizar  $X_i$ :
    For all j, k: If  $\text{rand}() < S(V_{ijk})$  then  $X_{ijk}(t+1) = (G_\pi)_{jk}$  else  $X_{ijk}(t+1) = 0$ 
  end For
  For each  $i = \overline{1, cantPart}$ 
    Evaluar  $X_i$  //Aplicar la métrica
    Actualizar  $XpBest_i$ 
  endFor
  Actualizar  $XgBest$ 
Incrementar t
Until  $t > CantGeneraciones$ 

```

Es necesario estimar la calidad del modelo de red obtenido en la fase de entrenamiento, o sea, es necesario tener en cuenta la información que brinda la función fitness u objetivo. Para lograr mejor funcionamiento del algoritmo PSO en el trabajo se muestran los resultados usando las medidas de calidad global implementadas en Weka [20], estas medidas evitan el sobre ajuste (*overfitting*) de los datos mediante el método de validaciones cruzadas (*crossvalidation*), las métricas se basan en resultados de la RB como clasificador, por lo que se extiende Weka con este nuevo algoritmo para buscar la estructura de una RB, los resultados de la clasificación son buenos, pero se aprecia que si se realiza una selección de atributos óptima con el propio método, según se aprecia en [20], los resultados de la clasificación de la HTA son mejores y se obtienen redes menos complejas.

La implementación se realizó haciendo una extensión al paquete *Weka* con el objetivo de minimizar el tiempo de implementación y rehusar código libre. Para los detalles consultar [20].

Otras medidas de evaluación de la RB, son por ejemplo: una basada en la sensibilidad y la especificidad, y la otra en la sensibilidad y precisión, estas se muestran en la ecuación 3 y 4, donde TP son los casos positivos que se predicen como positivos, TN son los casos negativos clasificados como negativos, y FP y FN son los casos mal clasificados en cada una de las clases.

La función de la ecuación 4 se debe utilizar para el caso donde se tengan clases minoritarias, y además se recomienda usar la precisión de Laplace, ver ecuación 5, donde k es la cantidad de clases [13], [9] y [11].

Estas medidas han sido utilizadas por algunos autores como se aprecia en [9] y [10].

$$\text{sensitividad} \times \text{especificidad} = \frac{TP}{TP + FN} \cdot \frac{TN}{TN + FP} \quad (3)$$

$$\text{sensitividad} \times \text{precisión} = \frac{TP}{TP + FN} \cdot \frac{TP}{TN + FP} \quad (4)$$

$$\text{precisión} = \frac{1 + TP}{1 + k + TN + FP} \quad (5)$$

4. REDES BAYESIANAS EN LA CLASIFICACIÓN DE LA HTA

En nuestro estudio la muestra estuvo constituida por un total de 849 individuos supuestamente sanos entre 18 a 78 años de edad, de ambos sexos, pertenecientes a 5 policlínicos de la ciudad de Santa Clara. Se confeccionó una historia clínica con información del paciente contenida en las siguientes variables: edad, sexo, raza, índice de masa corporal, bebe, fuma, diabetes mellitus, dislipidemia, número de padres con HTA, número de abuelos con HTA, tensión arterial sistólica y diastólica basal, al primer y segundo minuto, presión arterial media, glicemia, triglicéridos, colesterol total, hdl y ldl y perfil psíquico. A partir del análisis de las mismas, los pacientes fueron clasificados en normotensos, hiperreactivos e hipertensos.

La red obtenida con el algoritmo PSO de búsqueda local se muestra en la figura 1, en la que se hace el diagnóstico del tipo de HTA basándose solamente en los rasgos: *disamin1* y *2*, *diagexp*, *sistmin2*, edad glicemia e *imc*.

En la figura 2 se muestra la estructura de RB obtenida usando un paso previo de selección de atributos y el algoritmo de aprendizaje PSO global. La clasificación se basa en 11 rasgos. Se obtienen buenos resultados para la clasificación de la HTA en normotensos (0.926), hiperreactivos (0.925) e hipertensos (0.976) El porcentaje de bien clasificados es de 94.96.

Tabla 1. Resumen de las medidas de validación para los modelos de redes bayesianas obtenidas utilizando distintas implementaciones del algoritmo PSO para clasificar la HTA.

PSO (Parámetros)	Exactitud	Verdaderos Positivos	Área bajo la curva ROC
Búsqueda Local 2 padres, 40 partículas y 1000 iteraciones	89.21	Normotensos (0.864) hiperreactivos (0.888) hipertensos (0.911)	normotensos (0.976) hiperreactivos (0.942) hipertensos (0.988)
Búsqueda Global 2 padres, 40 partículas y 1000 iteraciones	93.72	normotensos (0.895) hiperreactivos (0.918) hipertensos (0.972)	normotensos (0.993) hiperreactivos (0.983) hipertensos (0.998)
Búsqueda Global 2 padres, 40 partículas y 1000 iteraciones (haciendo selección de atributos)	94.96	normotensos (0.926) hiperreactivos (0.925) hipertensos (0.976)	normotensos (0.996) hiperreactivos (0.985) hipertensos (0.997)

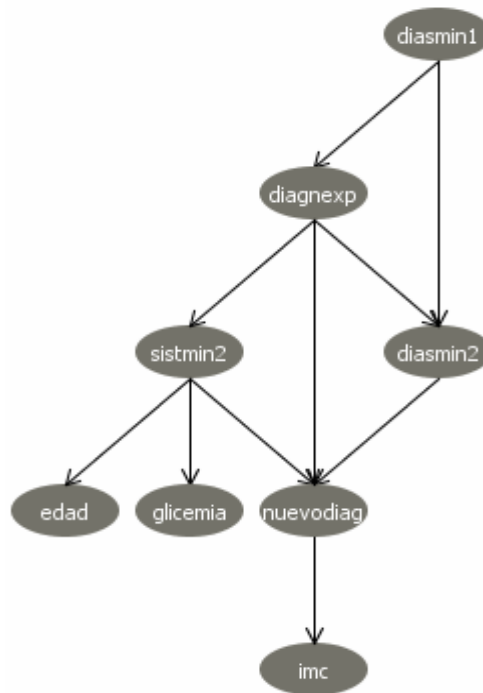


Figura 1. RB que muestra la posibilidad de realizar el nuevo diagnóstico teniendo en cuenta siete rasgos de los inicialmente tomados en la muestra. Con esta RB se obtienen resultados muy favorables para las tres clases normotensos (0.927), hiperreactivos (0.87) e hipertensos (0.97), la de los hipereactivos es la más confusa y por ello el porcentaje de bien clasificados más bajo.

En la tabla 2 se muestran los resultados de validación con otros enfoques alternativos reportados en la literatura para la tarea del aprendizaje de la estructura de una RB, los que no se diferencian significativamente de los que obtiene el algoritmo que se propone, pero se aprecia que cuando los parámetros del algoritmo propuesto son: Búsqueda Global, 2 padres, 40 partículas y 1000 iteraciones (haciendo selección de atributos previamente) los resultados son mejores relativamente a los obtenidos por el resto de los que se utilizaron para comparar.

5. CONCLUSIONES

El aprendizaje estructural de RB con el algoritmo de búsqueda PSO, ha mostrado ser un algoritmo eficiente, cuando se usan métricas de calidad global, cuando la red se evalúa con métricas de calidad local, los resultados suelen ser peores que en el caso global haciendo uso de validaciones cruzadas. Al hacer extensiones a Weka se logra validar cada uno de los modelos obtenidos mediante el uso de medidas implementadas en dicha plataforma y además es posible comparar los resultados con otros clásicos para la misma tarea reportados en la literatura. Se obtienen buenos modelos de RB para la clasificación de la HTA, con una exactitud alrededor del 95% de buena clasificación.

Se trabaja en una implementación paralelizada, para realizar aplicaciones con mayor volumen de datos, por ejemplo en aplicaciones Bioinformáticas.

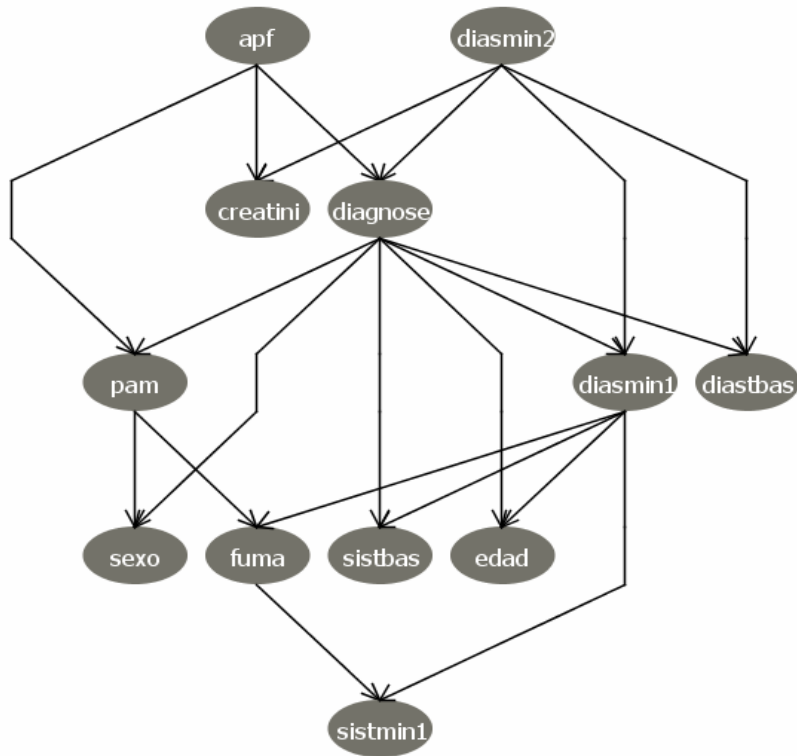


Figura 2. Los exactitud de la red por clases es: normotensos (0.926), hiperreactivos (0.925) e hipertensos (0.976). En este caso se logra la clasificación con 11 rasgos incluidos en el modelo.

Tabla 2. Resumen de las medidas de validación cuando se utilizan distintos algoritmos de aprendizaje estructural de RB.

Algoritmos	Exactitud	Verdaderos Positivos	Área bajo la curva ROC
K2	92.74	Normotensos (0.907) hiperreactivos (0.869) hipertensos (0.968)	normotensos (0.995) hiperreactivos (0.975) hipertensos (0.995)
HC	93.36	Normotensos (0.929) hiperreactivos (0.869) hipertensos (0.968)	Normotensos (0.997) hiperreactivos (0.98) hipertensos (0.995)
TAN	93.81	Normotensos (0.935) hiperreactivos (0.881) hipertensos (0.968)	Normotensos (0.995) hiperreactivos (0.982) hipertensos (0.996)

Received May 2008
Revised November 2008

REFERENCIAS

- [1] ANDERSON, K.M., ODELL, P.M., WILSON, P.W.F. y KANNEL, W.B. (1991): Cardiovascular disease risk profiles. **American Heart Journal**, 121. 293–298.
- [2] ARMARIO, P., HERNÁNDEZ DEL REY, R. y MARTIN, M. (2002): Estrés, enfermedad cardiovascular e hipertensión arterial. **Med. Clin(Barc)**. 119, 23- 29.

- [3] BENET, M., PENNINI, A., TORRES, J. y PERAZA, S. (2003): Cardiovascular Risk Factors among Individuals under Age 40 with Normal Blood Pressure. **Rev. Esp Salud Pública**, 77. 143-150.
- [4] BENET, M., YANES, N., GONZÁLEZ, J., PENNINI, A. y GARCÍA, J. (2001): Criterios diagnósticos de la prueba del peso sostenido en la detección de pacientes con hipertensión arterial. **Med. Clin (Barc)**, 116, 645-649.
- [5] BOUCKAERT, R.R. (1995): Bayesian Belief Networks: From Construction to Inference, **PhD Thesis**, Promotor: Prof. Dr. J. Van Leeuwen, Co-promotor: Dr. L.C. Van der Gaag, Faculteit Wiskunde en Informatica, Utrecht University
- [6] CASTILLO, E., GUTIÉRREZ, J.M. y HADI, A.S. (1997): **Expert Systems and Probabilistic Network Models**. Springer-Verlag, New York.
- [7] CHÁVEZ, M.C., CASAS, G., FALCÓN, R., MOREIRA, B. y R., G. (2007). Building Fine Bayesian Networks Aided by PSO-based Feature Selection. MICAI 2007: **Advances in Artificial Intelligence, LNCS, Springer Berlin / Heidelberg** 4827: 441- 451.
- [8] CHÁVEZ, M.C., SILVEIRA, P., CASAS, G., GRAU, R. y BELLO, R. (2007). Aprendizaje estructural de redes bayesianas utilizando PSO. Memorias en Boletín de la Sociedad Cubana de Matemática, Vol 5, Trabajo IA7, Número Especial en CD de COMPUMAT, ISSN: 1728-6042, Holguín, Cuba
- [9] CLARK, P. Y BOSWELL, R. (1991). Rule induction with CN2: Some recent improvements. **Machine Learning**, EWLS 91, 151- 163.
- [10] CORREA, E.S., FREITAS, A.A. Y JOHNSON, C.G. (2007). Particle Swarm and Bayesian Networks. Applied to Attribute Selection for Protein Functional Classification. **Proceedings of the GECCO: Conference companion on Genetic and evolutionary computation, ACM Special Interest Group on Genetic and Evolutionary Computation, NY, USA**, 2651-2658.
- [11] FALCO, D., CIOPPA, A.D. Y TARANTINO, E. (2007): Facing classification problems with Particle Swarm Optimization. **Applied Soft Computing**. 7, 652–658.
- [12] FERAT, S., TILLET, J., RAGHUVEER, R. y RAO, T.M. (2004): An evolutionary algorithmic approach to learning a Bayesian network from complete data. **Data Mining and Knowledge Discovery: Theory, Tools, and Technology VI, Orlando, FL, USA**, 5433, 88-99.
- [13] FERAT, S., YAVUZ, M.C., ARNAVUT, Z. y ULUYOL, O. (2007): Fault diagnosis for airplane engines using Bayesian networks and distributed particle swarm optimization. **Parallel Computing, Elsevier** 33: 124–143.
- [14] KENNEDY, J. (1997): The particle swarm: social adaptation of knowledge. **IEEE International Conference on Evolutionary Computation**, April 13–16, 303–308.
- [15] KENNEDY, J. y EBERHART, R.C. (1995): A new optimizer using particle swarm theory. **In: Sixth International Symposium on Micro Machine and Human Science. Nagoya**: 39–43.
- [16] KENNEDY, J. y EBERHART, R.C. (1995): Particle swarm optimization. **In: Proceedings of IEEE International Conference on Neural Networks, Perth**, 1942–1948.
- [17] KENNEDY, J. Y SPEARS, W.M. (1998): Matching algorithms to problems: an experimental test of the particle swarm and some genetic algorithms on the multimodal problem generator. **Proceedings of the IEEE International Conference on Evolutionary Computation**, 39- 43.
- [18] MAHAMED, G.H.O., ANDRIES, P.E. y AYED, S. (2005): Dynamic Clustering using PSO with Application in Unsupervised Image Classification. **Transactions on Engineering, computing and Technology**, 9: V9-36.

[19] WANG, X., YANG, J., TENG, X., XIA, W. y JENSEN, R. (2006). Feature Selection Based on Rough Sets and Particle Swarm Optimization. **Pattern Recognition Letter**, 28, 459-471

[20] WITTEN, I.H. y FRANK, E. (2005). **Data Mining Practical Machine Learning Tools and Techniques**, Morgan Kaufman, San Francisco, USA.