

APLICACIÓN DE LA COMPUTACIÓN EVOLUTIVA EN EL DIAGNÓSTICO DEL INFARTO AGUDO DEL MIOCARDIO

HOSPITAL DE SAN JOSÉ, BOGOTÁ DC, COLOMBIA 2012

John Jaime Sprockel MD*, Wilson Alzate**

Resumen

La evaluación de los pacientes con dolor torácico representa un reto para los profesionales de la salud, al ser el infarto cardíaco fuente importante de muertes en el mundo. Se presenta un algoritmo genético (AG) para seleccionar el mejor conjunto de reglas que puedan soportar su diagnóstico. Los individuos fueron representados como una combinación de 17 operaciones lógicas OR o AND (determinados como 1 ó 0) que relacionaban las 18 variables de la escala de Braunwald. Se seleccionó una población de 200 individuos y a partir de ellos se generaron 200 hijos por recombinación y mutación (95% y 5% de probabilidad respectiva), durante 200 iteraciones (generaciones). La función de correspondencia *fitness* fue calculada a partir de la evaluación del fenotipo de cada individuo en el conjunto de entrenamiento (119 pacientes). Tras validar las reglas resultantes en el conjunto de pruebas (40 pacientes) se alcanzó una precisión del 85% en el diagnóstico. Este resultado es parecido al desempeño de los médicos de urgencias y podría servir de apoyo en el diagnóstico diferencial del síndrome coronario agudo.

Palabras clave: inteligencia artificial, computación evolutiva, dolor torácico, síndrome coronario agudo, infarto miocárdico.

Abreviaturas: CE, computación evolutiva; SCA, síndrome coronario agudo; AG, algoritmo genético; PG, programación y genética.

USE OF EVOLUTIONARY COMPUTING FOR THE DIAGNOSIS OF ACUTE MYOCARDIAL INFARCTION - HOSPITAL DE SAN JOSÉ - BOGOTÁ DC - COLOMBIA

Abstract

Chest pain presents a diagnostic challenge for the healthcare professional given myocardial infarction is a leading cause of death worldwide. A genetic algorithm (AG) is herein presented for a better rule selection which can support diagnosis. Individuals were represented with a combination of 17 logical operations, OR or AND (determined as 1 or 0), which related the 18 variables of the de Braunwald scale. A population of 200 individuals was selected and 200 offsprings were generated by recombination and mutation (95% and 5% probability respectively), during 200 iterations (generations). The corresponding fitness function was calculated based on a phenotypic evaluation of individuals comprising the training set (119 patients). Eighty-five percent (85%) diagnostic accuracy was achieved, after validating the resulting rules in the testing set (40 patients). This result is similar to emergency doctors' performance, and has potential to serve as a support for differential diagnosis of acute coronary syndrome.

Key words: artificial intelligence, evolutionary computing, chest pain, acute coronary syndrome, myocardial infarction

Fecha recibido: abril 30 de 2013 - Fecha aceptado: agosto 23 de 2013

* Médico Internista. Candidato a Maestría en Ingeniería de Sistemas y Computación, Pontificia Universidad Javeriana. Instructor

Asociado. Fundación Universitaria de Ciencias de la Salud. Bogotá DC, Colombia.

** Ingeniero de Sistemas. Candidato a Maestría en Ingeniería de Sistemas y Computación, Pontificia Universidad Javeriana. Bogotá DC, Colombia.

Introducción

El síndrome coronario agudo (SCA), que incluye al infarto agudo del miocardio y la angina inestable, es una de las condiciones que se manifiesta por dolor torácico y pone en riesgo la vida de las personas que lo padecen, por lo tanto su diagnóstico apropiado y a tiempo determina la instauración de una serie de medidas que pueden cambiar su historia natural. Gran parte de los esfuerzos diagnósticos ante el dolor torácico en el servicio de urgencias se encaminan a esta entidad, pues sus síntomas son variables y percibidos de diferentes maneras por los pacientes.¹

La computación evolutiva (CE) representa una estrategia de simulación para resolver problemas complejos sobre la base de las teorías de la evolución natural y de la variación genética.² El algoritmo genético (AG) es uno de búsqueda de enjambre típico para optimización global, pero debido a que adopta la tecnología de búsqueda de probabilidad tiene un cierto grado de azar y algunos consideran que la convergencia local no es buena.

El presente artículo busca en primera instancia realizar una revisión del estado del arte en el uso de la CE en el diagnóstico del SCA, acto seguido se aplica un AG para descubrir un número de reglas lógicas que describan con mayor precisión a los pacientes que cursan con un SCA, a partir de las 18 variables contenidas en

la escala de valoración de la probabilidad de angina descrita por Braunwald y col. en 1994³ (**Tabla 1**).

Estado del arte de las herramientas evolutivas para el diagnóstico del SCA

Dentro del campo del diagnóstico del SCA la CE ha aportado varias posibilidades de abordaje, siendo la programación genética (GP) la estrategia que más representa una contribución directa al diagnóstico (**Tabla 2**). Bojarczuk y col.⁴ promueven su uso para el descubrimiento de reglas comprensibles de clasificación dentro del campo de la minería de datos, evalúan 165 atributos para la clasificación de 12 enfermedades que se presentan con dolor torácico, entre las cuales estaban las que constituyen el SCA. Obtuvo una precisión global del 97%. Zhou y col.⁵ desarrollaron un modelo de clasificación del riesgo del SCA para eventos cardiovasculares mayores, basado en un AG con búsqueda local flotante embebida (GA-LFE), tomando datos que incluyeron niveles de mieloperoxidasa y troponina. Los pacientes provenían de la Cleveland Clinic y fueron 60 casos y 60 controles. La precisión de la predicción promedio de 3, 5 y 7 biomarcadores seleccionados son 67,50, 72,92 y 77,08% respectivamente.

La mayoría de los artículos aprovechan las capacidades de la CE para el desarrollo de otras herramientas de los sistemas inteligentes. Nugent, López y col. en

Tabla 1. Probabilidad de enfermedad coronaria significativa ante síntomas sugestivos de angina inestable

Probabilidad	Alta	Intermedia	Baja
Historia	Dolor torácico o del miembro superior izquierdo como síntoma principal de consulta que se asemeja a un episodio anterior de angina. Historia conocida de enfermedad coronaria.	Dolor o malestar torácico o del miembro superior izquierdo como síntoma principal de consulta. Edad mayor de 75 años. Género masculino. Diabetes mellitus	Síntomas isquémicos probables en ausencia de características de probabilidad intermedia o alta. Uso reciente de cocaína.
Examen	Soplo mitral transitorio, hipotensión, diaforesis, edema pulmonar o estertores.	Enfermedad vascular extracardíaca.	Dolor torácico reproducible a la palpación.
Electrocardiograma	Desviación transitoria del ST nueva o presumible nueva (mayor de 1 mm) o inversiones de la onda T en múltiples derivaciones.	Ondas Q fijas. Depresión del ST entre 0,5 a 1 mm o inversión de la onda T mayor de 1 mm en una derivación.	Aplanamiento o inversión de la onda T menor de 1 mm en derivaciones con onda R dominante. ECG normal.
Marcadores cardíacos	Troponina I, T o CPK MB elevada.	Normal	Normal

Tabla 2. Computación evolutiva en el diagnóstico del SCA

Autor y año	Estrategia	Población	Resultado
Bojarczuk, 2000	PG para reglas comprensibles, clasificación. 165 atributos para clasificar 12 enfermedades.	138 pacientes (90 entrenamiento y 48 pruebas)	Precisión: 97%
Nugent, 2001	ANN, usando PG para elegir el número de generaciones para evitar el <i>overfitting</i> . 6 parámetros electrocardiográficos en diagnóstico de Infarto.	131 pacientes (97 entrenamiento y 44 pruebas)	Determinaron el mejor punto para el cese del entrenamiento.
Ming, 2010	ANN, usando AG para determinar la mejor distribución de la capa oculta. 13 características clínicas, electrocardiográficas y de laboratorio.	497 pacientes con SCA (300 entrenamiento y 197 pruebas)	Disminuyeron de 537 a 37 las iteraciones de aprendizaje. Precisión: 98.98%
Sepulveda, 2002	AG para seleccionar el mejor subconjunto de variables para ser aplicadas en el modelo de SVM.	726 pacientes con angina	El AG redujo de 75 a 14 las variables sensibilidad= 66,67% especificidad= 79,77% VPP = 79,12% VPN = 97,87%
Ha, 2007	Modelo evolutivo de hiper-redes, usaron CE para el reemplazo de los hiper-bordes, buscando reducir la cardinalidad (pesos).	135 pacientes	Seleccionaron 150 aptameros (proteínas) para dx, entre 3.000. No sirvió para infarto.
Zhou, 2009	AG con búsqueda local flotante embebida (GA-LFE) niveles de mieloperoxidasa y troponina en diagnóstico del infarto.	60 casos y 60 controles	La precisión de 3, 5 y 7 biomarcadores son 67.50, 72.92 y 77.08% respectivamente.

AG: algoritmo genético; PG: programación y genética; ANN: red neuronal artificial; SVM: máquina de soporte de vectores.

una publicación duplicada^{6,7} exponen el entrenamiento de una red neuronal para el diagnóstico del infarto miocárdico (con respecto a su localización) a partir de seis parámetros electrocardiográficos, utilizaron la PG para decidir el número de generaciones durante el entrenamiento para intentar evitar el *overfitting*. Ming y col.⁸ entrenaron una red con algoritmos genéticos para encontrar una solución superior, el resultado luego fue capacitado como los parámetros iniciales del algoritmo de red con la premisa de encontrar la mejor distribución de la capa oculta. Tomaron 497 pacientes con enfermedad coronaria (300 para entrenamiento y 197 para las pruebas), procedentes de la Cleveland Clinic y el Instituto Húngaro de Cardiología. Evaluaron trece características clínicas, electrocardiográficas y de laboratorio; lograron disminuir de 537 a 37 las iteraciones de aprendizaje con una precisión del 98,98% en el grupo de pruebas con respecto a la angiografía.

Sepulveda y col.⁹ en el Clinic de Barcelona describen la combinación de máquinas de soporte vectorial (*support vector machines*) con algoritmos genéticos para el diagnóstico de la angina inestable. El mecanismo evolutivo fue diseñado para seleccionar el mejor subconjunto de variables para ser aplicadas en el modelo, a partir de una cohorte de 726 pacientes que cumplieron con los criterios de clasificación de angina.

La sensibilidad fue 66,67%, especificidad 79,77%, valor predictivo negativo de 97,87% para el desarrollo de complicaciones cardiovasculares mayores de seis meses y el AG redujo de 75 a 14 las variables puestas a consideración. Ha y col.¹⁰ diseñaron un modelo evolutivo de hiper-redes, que es un nuevo modelo gráfico probabilístico aleatorio basado en gráficos unidireccionales. Se usó computación evolutiva para el reemplazo de los hiper-bordes, buscando reducir la cardinalidad de éstos, con lo cual podría coincidir con mayor cantidad de datos. Esta estrategia fue usada para identificar proteínas relacionadas con el SCA, usando un biochip de 3K (*aptamer-array*); a partir de 3.000 opcionadas, se seleccionaron 150 que tuvieron valores significativos ya que ayudaban a diferenciar con precisión 135 pacientes con SCA, quienes cursaban con angina estable, inestable o eran normales, pero no fue efectivo en el caso del infarto.

Metodología

Se tomaron los datos procedentes de un estudio de corte transversal realizado en el Hospital de San José de Bogotá DC, como parte de la línea de investigación de dolor torácico, en el grupo de medicina interna denominado “Estructuración e implementación de una ruta crítica para el manejo de pacientes con síndrome

coronario agudo en el servicio de urgencias del Hospital de San José de Bogotá DC”. La población estuvo compuesta por pacientes mayores de 18 años que consultaron al servicio de urgencias por dolor torácico, en el período comprendido entre el 20 febrero y el 30 de octubre de 2012. Este estudio captó pacientes con posible SCA a los cuales se le realizó una ruta crítica cuyo objetivo era identificar aquellos con SCA de alto riesgo y mejorar su calidad de atención.

Descripción de la estrategia evolutiva

- *Representación de los individuos:* se diseñó un conjunto de reglas que asociaban los resultados de los datos de cada uno de los ítems evaluados con los conectores lógicos AND y OR, con ello la representación de los individuos estaba dada por un cromosoma de 17 bits, de manera que **1** representaba los AND y **0** los OR.
- *Tamaño de la población:* 200 individuos.
- *Iniciación de la población:* se realizó de manera aleatoria especificando un punto de corte de alelos positivos de 7, es decir, dándole una probabilidad mayor (7/10) para los OR y una menor (3/10) para los AND. Ésta decisión se tomó tras el análisis hecho de los ejemplos, al ver que en los pacientes se expresa con más frecuencia la ausencia de datos positivos.
- *Mecanismo de evolución:* se realizó mediante mutación y recombinación, entregando una menor probabilidad para la primera (5%).
- *Selección:* para la selección de individuos se utilizó una estrategia elitista.
- *Función de correspondencia (fitness):* consistió en evaluar con el fenotipo (ANDs y ORs) los ejemplos (119 de los 159 que se tenían en la base de datos), combinando las variables según el fenotipo y evaluando el resultado final con el del ejemplo. Cada vez que se encontró una coincidencia entre el valor arrojado por el programa y el de los ejemplos se sumaba uno. El mejor individuo era aquel con la mayor correspondencia.

- *Criterio de terminación:* número de iteraciones, inicialmente fue de 1.000. Vale la pena resaltar que luego de cualquier evolución, con el fin de evitar redundancia, los individuos repetidos fueron eliminados de la población.

Resultados

La base de datos estuvo compuesta por 81 pacientes en quienes se descartó el SCA y 78 confirmados (38 de ellos con infarto). 48,4% pertenecían al género masculino y 20,8% eran diabéticos. El 8,8% tuvo historia de enfermedad coronaria conocida y 27,7% presentó elevación de los biomarcadores cardíacos.

Tras correr el conjunto de entrenamiento, se halló que el mejor individuo en todas las iteraciones tuvo un *fitness* (correspondencia) de 92/119, cuyo promedio en la población final fue también de 91/119. El número de individuos de esta población final fue 184, luego de 87 mutaciones y 913 recombinaciones. Aplicando tanto el mejor individuo como la población final a los casos de prueba (40/119), se obtuvo un resultado idéntico con respecto al del ejemplo en 33 de 40 casos para la población final y en 34 de 40 casos para el mejor individuo de todas las poblaciones. Por último se obtuvo un porcentaje de éxito de 82,5% para la población final y 85% para el mejor individuo.

Se realizaron también algunas pruebas alternando el número de individuos en la población inicial, el número de iteraciones, la probabilidad de alelos positivos y de mutación, obteniendo diferentes resultados que se muestran en la **Tabla 3**.

Discusión

El SCA es un problema complejo, su diagnóstico implica tener en cuenta una amplia cantidad de datos con presentación variada o que puede ser interpretada de distintas maneras por varios profesionales de la medicina.

Si bien la mejor solución podría haber sido modelada con un árbol de decisiones mediante el uso de programación genética, nos topamos con varias dificultades.

Tabla 3. Resultados tras realizar pruebas con diferentes condiciones iniciales

Población	Alteraciones	Alelos positivos	Mutación	Mejor fitness	Población final	Éxito población final	Éxito mejor individuo
50	500	8/10	10/100	92	44	82	85
100	1500	5/10	20/100	92	97	82	85
150	2000	4/10	50/100	92	121	86.47	85
20	100	9/10	1/100	91	1	85	82
300	600	2/10	40/100	92	184	82	85

Al no estar familiarizados con el uso del lenguaje LISP buscamos *frameworks* que permitieran el manejo de la GP en Java. Exploramos la posibilidad del uso de JGAP, EpochX y finalmente JCLEC. La librería JGAP es el que mayor difusión tiene para la creación de aplicaciones de CE y GP, pero la encontramos confusa y difícil para cada uno de los pasos que debían darse como era la definición de las funciones y terminales, la representación de los individuos, la evolución y la correspondencia (*fitness*). Al evaluar EpochX logramos bastantes avances, pero se fracasó al intentar modelar la función de correspondencia. Al buscar ejemplos o documentación acerca de su uso, fueron muy escasos y con falta de pertinencia. Por último, JCLEC parecía prometedor pero encontramos dificultades en su ejecución al tener inmerso un desarrollo con XML y ser problemático el uso de los ejemplos para determinar la correspondencia. Ante estos obstáculos concluimos que las ayudas existentes para la PG son poco amigables, poco o nada intuitivas, con escasa documentación o ejemplificación en la red, haciéndolas complicadas para su uso sin contar con personas familiarizadas con ellas.

La solución que se expuso maneja un AG para la creación, evolución y selección de la población; comparte con el sistema clasificador de aprendizaje tipo Michigan que se evalúa por separado, cada una de un conjunto de reglas, pero no se utilizó refuerzo ni penalización en la función de correspondencia. Se encontró que un único individuo puede tener un rendimiento que iguala al de la población completa de reglas encontradas. La precisión presenta un rango adecuado para el apoyo del diagnóstico.

Los resultados obtenidos aunque ofrecen una buena tasa de precisión, valdría la pena intentar una solución que pudiese tener en cuenta otros operadores lógicos como XOR, NAND, NOT, NOR, así como signos de agrupación considerando relaciones de precedencia, para lo cual necesariamente se tendría que usar la GP, representando el programa como un árbol. Para esto, es importante considerar el apoyo de la programación funcional (LISP) o el uso de un *framework* específico para GP como JGAP, EpochX, JCLEC u otro.

Conclusiones

El presente es un estudio exploratorio que busca hacer una propuesta para el uso de un algoritmo genético como apoyo al diagnóstico del síndrome coronario agudo; la precisión del 85% demuestra que dichas reglas tienen el potencial de convertirse en una ayuda en la toma de decisiones.

Referencias

1. Erhardt L, Hertz J, Bossaert L, Halinen M, Keltai M, Koster R, et al. Task force on the management of chest pain. *Eur Heart J*. 2002 Aug; 23(15):1153-76.
2. Evolution strategies and evolutionary programming. In: Dumitrescu D. *Evolutionary computation*. Boca Raton, Florida: CRC Press; 2000. p. 261-79.
3. Braunwald E, Jones RH, Mark DB, Brown J, Brown L, Cheitlin MD, et al. Diagnosing and managing unstable angina. *Agency for Health Care Policy and Research. Circulation*. 1994, 90(1): 613-22.
4. Bojarczuk CC, Lopes HS, Freitas AA. Genetic programming for knowledge discovery in chest-pain diagnosis. *IEEE Eng Med Biol Mag*. 2000 Jul-Aug; 19(4):38-44.
5. Zhou X, Wang H, Wang J, Wang Y, Hoehn G, Azok J, et al. Identification of biomarkers for risk stratification of cardiovascular events using genetic algorithm with recursive local floating search. *Proteomics*. 2009 Apr; 9(8):2286-94.
6. Lopez JA, Nugent C, Black ND, Smith AE. Automated synthesis of prediction models for neural network based myocardial infarction classifiers. *Proceedings of the 23rd Annual International Conference of the IEEE*; 2001 Oct. 25-28; Istanbul, Turkey. p. 3803-06.
7. Nugent CD, Lopez JA, Smith AE, Black ND. Prediction models in the design of neural network based ECG classifiers: a neural network and genetic programming approach. *BMC Med Inform Decis Mak*. 2002; 2:1.
8. Ming L, Yan-chun L, Xin-min W, Xiao-gang D. The application of GA-BP algorithm to intelligent diagnosis of coronary heart disease. *Proceedings of the International Conference on Computer, Mechatronics, Control and Electronic Engineering (CMCE)*, IEEE. 2010 Aug. 24-26: Changchun, China. p. 367-70
9. Sepulveda-Sanchis J, Camps-Valls G, Soria-Olivas E, Salcedo-Sanz S, Bousono-Calzon C, Sanz-Romero G, Marrugat de la Iglesia J. Support vector machines and genetic algorithms for detecting unstable angina. *Comput Cardiol*. 2002; 29: 413-16.
10. Ha Jw, Eom J, Kim S, Zhang B. Evolutionary hypernetwork models for aptamer-based cardiovascular disease diagnosis. *Proceedings of the GECCO conference companion on Genetic and evolutionary computation*. 2007 July 7-11: London, England. p. 2709-16.