

Clasificación de microcalcificaciones usando razonamiento basado en casos

Maria Salamó^(a)

Elisabet Golobardes^(a)

Josep M. Garrell^(a)

Joan Martí^(b)

^(a)Grupo de Sistemas Inteligents

^(b)Grupo de Visión por Computador y Robótica

Enginyeria i Arquitectura La Salle

Instituto de Informática y Aplicaciones

Universitat Ramon Llull

Universitat de Girona

Passeig Bonanova, 8 - 08022 Barcelona

Campus de Montivili - 17071 Girona

e-mail: {mariasal, elisabet, josepmg}@salleURL.edu

e-mail: joanm@eia.udg.es

Keywords: Case-based Reasoning, Machine Learning, Computer Vision, Human-Medicine and Healthcare

ABSTRACT. *Este artículo describe una aplicación para la detección de cáncer de mama a partir de mamografías. El principio básico de esta aplicación es conseguir una clasificación automática de los datos correspondientes a las microcalcificaciones encontradas en las mamografías, usando Aprendizaje Automático. Los datos de las microcalcificaciones de una mamografía han sido cedidos por la Universitat de Girona (UdG) y a su vez en colaboración con el Hospital Josep Trueta de Girona. La técnica de Aprendizaje Artificial que utilizamos para clasificar las mamografías es el Razonamiento Basado en Casos.*

1 Introducción

La incidencia del cáncer de mama varía dependiendo del país, pero las últimas estadísticas muestran que cada año se diagnostican 720.000 casos nuevos en todo el mundo, de los cuales un 20% suponen el desarrollo de tumores malignos.

Se ha comprobado que la implantación de programas masivos de *screening* ha repercutido en un incremento de la tasa de supervivencia de las mujeres ante esta enfermedad [Winfields94]; sin embargo, el porcentaje de mujeres que padecen cáncer y que es detectable mediante métodos clínicos o mamográficos constituye únicamente un reducido porcentaje de las mujeres que poseen una mama con alguna abnormalidad y que aún no es clínicamente visible. Consecuentemente, se hace necesario el desarrollo de nuevas estrategias para la identificación de forma prematura de aquellas pacientes sometidas a un elevado riesgo de contraer un cáncer invasivo de mama, con las consiguientes ventajas terapéuticas y sociales que ello puede representar.

En este sentido, es de común acuerdo entre la comunidad científica y médica que la incorporación de sistemas de tipo CAD (Computer Aided Diagnosis) en las etapas iniciales de diagnóstico y análisis de la enfermedad suelen traducirse en una mayor supervivencia de la población afectada por este tipo de neoplasia, de forma que la tendencia en la lucha contra el cáncer es la paulatina incorporación de nuevas técnicas y métodos encaminados a la detección precoz de elementos propios de la sintomatología cancerígena.

A tal efecto, la incorporación de técnicas basadas en Visión por Computador y Aprendizaje Artificial debe posibilitar a los especialistas médicos la posibilidad de disponer de nuevos datos para la emisión de un diagnóstico más fiable del riesgo de desarrollar cáncer.

Como técnica de Aprendizaje Automático se ha decidido utilizar el razonamiento basado en casos (CBR). Esta técnica de aprendizaje artificial fue usada con éxito en un problema similar aunque menos complejo [Garrell98, Garrell99].

El artículo se ha estructurado en las siguientes secciones. Inicialmente detallaremos el problema al que nos enfrentamos. A continuación describiremos las técnicas de inteligencia artificial que se han usado, Visión por Computador y Aprendizaje Artificial. En la siguiente sección mostraremos los resultados obtenidos. Finalmente, comentaremos las conclusiones y el trabajo actual y futuro.

2 Descripción del problema

El problema surge cuando un grupo de pacientes padece o puede llegar a padecer alguna anomalía en la mama, es a partir de este momento cuando en los hospitales se realiza una prueba mamográfica para realizar el estudio de la parte interna y llegar a detectar la presencia o ausencia de cáncer de mama. Las mamografías se utilizan debido a su alta sensibilidad en encontrar tumores pequeños. Entre las características detectables a partir de una mamografía están las distorsiones en la forma de la mama, el incremento de masa, de densidad y el incremento en el número de microcalcificaciones.

Los avances tecnológicos han aumentado la calidad en la digitalización de mamografías. Estas nuevas tecnologías permiten detectar y caracterizar perfectamente las microcalcificaciones y el volumen de masa muscular. Normalmente las microcalcificaciones (Ca^{++}) y el crecimiento de la masa muscular son un signo preciso de la existencia de cáncer de mama. Las microcalcificaciones aparecen como puntos pequeños y brillantes situados arbitrariamente en la mama y con formas muy diversas. Las mamografías digitalizadas de estos pacientes han sido cedidas por el Hospital Josep Trueta de Girona a la Universitat de Girona.

La Universitat de Girona mediante técnicas de Visión Artificial [Marti98] realizaron la segunda etapa de nuestro problema. En esta etapa se procedió a la extracción de los datos más característicos de cada una de las microcalcificaciones encontradas en una mamografía teniendo en cuenta descriptores de forma.

Nuestro propósito es centrar este artículo en una tercera etapa, dedicada al Aprendizaje Automático a partir de los resultados obtenidos en la etapa anterior, dónde se utilizan las técnicas de Visión por Computador. Ver la figura 1.



Figura 1: Etapas del problema.

Una vez estudiadas las etapas de división de nuestro problema, nos centraremos en la naturaleza del mismo. Nuestro propósito es clasificar las mamografías en tres clases: mamografías benignas, malignas o sin clasificar. Es importante remarcar que la clasificación se realiza utilizando únicamente la información de las Ca^{++} encontradas en una mamografía.

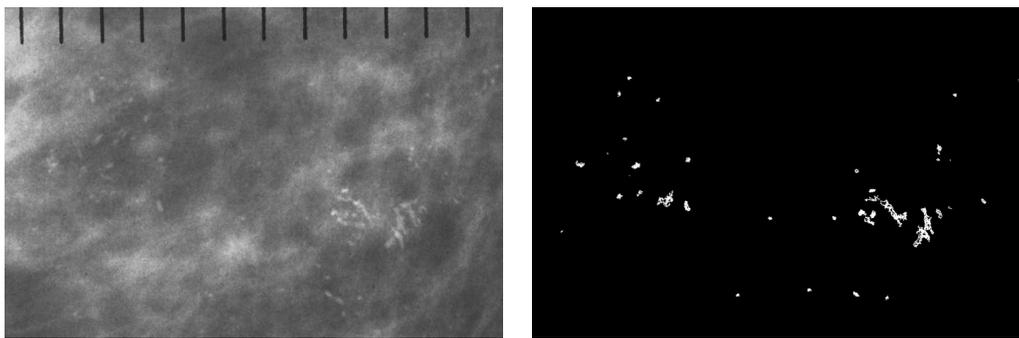
3 Técnicas de Inteligencia Artificial utilizadas

En primer lugar, detallaremos cómo se han extraído los datos de las mamografías y, posteriormente, detallaremos el proceso de tratamiento de los datos para conseguir la clasificación automática.

3.1 Segmentación de las microcalcificaciones

Con el objetivo de aislar las microcalcificaciones del resto de la mamografía se ha utilizado el algoritmo de segmentación propuesto por Shen [Shen94] a partir del crecimiento de regiones con diferentes valores de tolerancia, según la textura de la región analizada. Ello ha permitido la caracterización, a partir de parámetros de forma, de las microcalcificaciones dudosas (marcadas por los expertos radiólogos). El proceso de segmentación ha permitido la determinación de los parámetros de forma que mejor caracterizan las microcalcificaciones: Área, Perímetro, Compacidad, Número de agujeros, Rugosidad, Elongación, Diámetro mínimo, máximo y medio de Feret, Perímetro convexo, Centroide y Ejes principal y secundario.

La figura 2 muestra dos imágenes mamográficas, la primera de ellas muestra una mamografía original y la segunda muestra una mamografía segmentada.



(a) Original

(b) Segmentada

Figura 2: Mamografías digitalizadas

3.2 Razonamiento basado en casos

Una vez finalizada la etapa de extracción de los datos, se inicia la etapa de clasificación de estos datos. Concretamente, se realiza esta clasificación utilizando CBR. Los datos extraídos en la etapa anterior están representados por un conjunto variable de microcalcificaciones, donde cada Ca^{++} dispone de un conjunto de 25 atributos que la caracterizan. El tipo de cada atributo puede ser o bien real, o bien entero, o bien binario, y con un rango distinto de posibles valores. Una de las técnicas que mejor se adapta al problema de las mamografías y, teniendo en cuenta un trabajo previo [Golobardes97, Garrell98, Garrell99], es el Razonamiento Basado en Casos.

A continuación expondremos las ideas primordiales del CBR.

El razonamiento basado en casos [Kolodner93] integra el aprendizaje artificial y la resolución de problemas. Esta metodología utiliza una filosofía similar a la que usamos los humanos: intenta resolver los casos nuevos de un problema a partir de casos ya resueltos anteriormente. La resolución de estos casos nuevos aporta información y conocimiento con el propósito de resolver otros casos en el futuro.

Clásicamente, este método viene caracterizado por cuatro fases [Aamodt94]: una primera fase de **recuperación** de casos similares al caso nuevo que ya hayan sido resueltos anteriormente; una segunda fase de **adaptación**, en la cual se intenta adaptar la solución de los casos anteriores recuperados para resolver el caso nuevo; una tercera fase de **revisión**, en la cual se revisa la solución propuesta; y, finalmente, una cuarta fase de **almacenamiento**, en la cual se almacena la información relevante obtenida de esta nueva resolución.

3.3 Aplicación de CBR sobre nuestro problema

Dado que el problema que queremos resolver es un problema de clasificación, esto simplifica alguna de las fases del ciclo del razonamiento basado en casos, como la fase de adaptación. Este hecho se produce porque consideramos que el diagnóstico para el caso nuevo es el mismo que el del caso recuperado y ya resuelto anteriormente. De hecho, la fase de recuperación es la que se considera el núcleo de todo el ciclo de razonamiento basado en casos porque debe encontrar los casos más similares para dar solución al caso que se desea resolver.

En nuestro problema, la clave para un diagnóstico correcto consiste en recuperar el caso (mamografía) más similar al que queremos resolver. Pero, ¿qué características nos permiten evaluar la similitud entre dos casos?, ya que, difícilmente dos casos vienen representados por el mismo número de Ca++.

La respuesta, tampoco es evidente para los expertos. Así pues, podríamos evaluar la similitud desde dos perspectivas diferentes:

- Suponer que el conjunto de Ca++ que describen una mamografía, es el que determina si la mamografía es de diagnóstico cancerígeno o no. Es decir, aplicar una función de evaluación global del caso.
- O bien, suponer que una única Ca++, o un conjunto muy reducido de Ca++, es la que nos permite establecer el diagnóstico. Con esta perspectiva sólo necesitamos encontrar la Ca++ que nos representa cada caso.

Debido a que no disponemos de información certera sobre cómo escoger la Ca++ que representa un caso, hemos decidido realizar un primer análisis de la capacidad del CBR a partir de una política global.

A partir de los factores definidos anteriormente, centramos la fase de recuperación en los siguientes pasos:

El primer paso es **normalizar** o aplanar el caso a resolver en un caso estándar (un vector de características). Al reducir el problema estamos consiguiendo la aplicabilidad de las funciones de similitud empleadas normalmente en CBR. Esta fase, a su vez, se divide básicamente en dos etapas:

(1) Transformación del problema en un vector de características. (2) Unificación de criterios de similitud a alto nivel para conseguir evaluar globalmente un caso.

El segundo paso es aplicar una **función de matching** o similitud sobre el caso reducido o normalizado. A su vez, los casos recuperados también se normalizan para que sean equitativas las comparativas sobre las funciones de similitud. En este punto hemos decidido aplicar

diferentes funciones de similitud para analizar la viabilidad del CBR en este problema. Las funciones de similitud que se han testado en nuestro problema son:

Distancia Minkowski. Esta función está catalogada dentro del grupo de funciones de similitud global porque recupera el caso que globalmente sea más similar (o que se encuentre más próximo) al caso que se desea resolver, la fórmula es la siguiente:

$$Similitud(caso_X, caso_Y) = \sqrt[r]{\sum_{i=1}^F |x_i - y_i|^r}$$

donde F es el conjunto de características o de atributos y; r es el valor de las dos variantes que utilizamos de esta función: distancia de Hamming $r = 1$, distancia Euclidiana $r = 2$; y los valores x_i , y_i representan el valor del atributo i para el $caso_X$ y para el $caso_Y$ respectivamente.

Distancia de Clark. Viene descrita por la fórmula siguiente:

$$Similitud(caso_X, caso_Y) = \sqrt[r]{\sum_{i=1}^F \left(\frac{1}{(|x_i + y_i|^r)} (|x_i - y_i|^r) \right)}$$

donde F es el conjunto de características y r muestra las dos variantes a ejecutar, típicamente este valor es el de distancia Euclidiana $r = 2$, aunque también da buenos resultados el valor de distancia de Hamming; y los valores x_i , y_i representan el valor del atributo i para el $caso_X$ y para el $caso_Y$ respectivamente. Esta función, tal y como se puede apreciar, es muy similar a la función anterior. La diferencia entre ambas es la ponderación que realiza la distancia de Clark a partir del valor de los mismos atributos que se están evaluando. Este factor de ponderación o normalización permite que comparativas numéricas entre atributos muy distantes se equilibren.

Distancia del Coseno. Esta función se basa en las propiedades de vectores en un espacio euclídeo. Mide el coseno del ángulo de dos vectores en un espacio N -dimensional. Su expresión es la siguiente, siendo F el número de propiedades; y representando x_i , y_i el valor del atributo i para el $caso_X$ y para el $caso_Y$:

$$Similitud(caso_X, caso_Y) = \frac{\sum_{i=1}^F (x_i \cdot y_i)}{\sqrt{\sum_{i=1}^F x_i^2 \sum_{i=1}^F y_i^2}}$$

Cuando todas las propiedades son binarias se le puede dar una interpretación no geométrica. Tomando el sumatorio del denominador como el número de atributos comunes entre las dos distancias y el denominador como la media geométrica del número de atributos que posee x e y , entonces la medida se podría interpretar como la relación de atributos comunes que poseen ambas instancias.

El último paso es **seleccionar** el caso (o casos) que mayor grado de similitud tengan con el caso que deseamos resolver. Este es el caso (o casos) a partir del cual diagnosticaremos el caso actual (fase de adaptación del CBR).

4 Resultados de la experimentación

A continuación mostraremos los resultados de las pruebas realizadas sobre los datos. La experimentación realizada se basa en un conjunto de 216 mamografías digitalizadas ¹. Este conjunto se ha dividido en un conjunto de entrenamiento y en un conjunto de test ². El primer conjunto servirá para entrenar el sistema, y el segundo conjunto para evaluar la capacidad de diagnosticar correctamente.

4.1 Resultados mediante clasificación estadística

Manteniendo inalterados los procesos clínicos y de segmentación mediante técnicas de Visión por Computador que condujeron a la obtención de los datos referentes a los parámetros de forma de las microcalcificaciones, se realizó un estudio estadístico que permitiera evaluar la importancia de los diferentes parámetros extraídos, así como su incidencia en la construcción de una función de riesgo para el desarrollo del cáncer de mama [Marti98]. Utilizando una función logit como función de enlace para el modelo logístico regresivo, se concluyó que los parámetros significativos eran "número de *clusters*, número de agujeros, área, elongación de Feret, *roughness*, y elongación". Asimismo, el comportamiento del modelo predictivo que se obtuvo no difería significativamente del que mostraban los expertos médicos, aunque sí ofrecía un comportamiento mucho más conservador (mayor número de casos sin clasificar, con un índice de aciertos prácticamente idéntico).

4.2 Resultados usando razonamiento basado en casos

A continuación mostraremos los resultados obtenidos en el sistema de clasificación automático usando CBR. Concretamente, evaluamos los resultados a partir de contrastar las diferentes funciones de similitud aplicadas.

Función de Similitud	% Correctos	% Incorrectos	%No clasificables	% $\frac{\text{Aciertos}}{\text{decididos}}$
Hamming	65.2%	33.3%	1.4%	66.1%
Euclidiana	52.1%	44.9%	2.9%	53.7%
Clark	63.7%	30.4%	5.7%	67.6%
Coseno	55.0%	31.0%	13.0%	63.3%

Tabla 1: Resultados con datos normalizados

La tabla 1 muestra los resultados obtenidos con las diferentes funciones de similitud que se han estudiado. Es necesario mencionar que los datos se han normalizado para tener intervalos con el mismo rango de valores para cada atributo, así las funciones de similitud no benefician a un atributo con rango de valores muy alto respecto a un atributo con un rango de valores menor.

¹Estas mamografías están divididas en lateral, cráneo-caudal, oblicua y biopsia. Cada una de ellas simboliza el área enfocada para realizar la mamografía. No se han considerado las mamografías extraídas mediante biopsia en este conjunto de experimentación.

²Estos dos conjuntos han sido elegidos por los médicos especialistas.

La comparativa de nuestros resultados se realiza utilizando como referencia un estudio que se realizó con el mismo conjunto de mamografías sobre un equipo de especialistas en radiología. Los médicos especialistas fueron capaces de evaluar un **51%** de mamografías correctamente. A partir de esta referencia, podemos mencionar que los resultados obtenidos son prometedores debido a su buen funcionamiento inicial y teniendo en cuenta que es la primera aproximación sobre el problema, utilizando únicamente factores globales de evaluación para cada mamografía.

Al mismo tiempo, podemos observar que nuestros resultados ya mejoran el resultado de los médicos especialistas y también detectamos la importancia en la elección de una función de similitud que se adapte lo mejor posible a nuestro problema. Cualquiera de las tres funciones de similitud testadas sobre los datos mejora la previsión de los médicos especialistas, pero es muy importante destacar que el resultado varía ampliamente de una a otra, destacándose entre las cuatro estudiadas la distancia Clark y la de Hamming. Este hecho es debido a que las dos funciones estabilizan los resultados mediante una ponderación, en el caso de la distancia de Hamming se busca la distancia real y, en el caso de la distancia de Clark, se estabiliza el resultado mediante el factor de normalización lo que permite tener casi una distancia real normalizada por los propios atributos.

En el caso de los médicos especialistas el resultado fue un 51% de casos acertados, un 4% de casos incorrectos y un 45% de casos indecisos. En nuestros resultados se puede ver que el porcentaje de indecisos es mucho menor porque es difícil dar un criterio de aceptación de un resultado. Este factor nos indica que las variaciones mediante un criterio u otro se deben evaluar cuidadosamente. Por ejemplo, en la distancia del Coseno podemos obtener un resultado de aciertos del 40,5%, de incorrectos del 20,2%, de no clasificables del 39.3% y de $\frac{\text{aciertos}}{\text{decididos}}$ del 66.6%. Este resultado mejora el porcentaje de aciertos respecto los decididos a clasificar y aumenta el porcentaje de no clasificables, ésta sería una política más conservadora. Esta comparativa y nuestros resultados apuntan a que tengamos esperanzas en mejorar los resultados previos, teniendo en cuenta que no se ha realizado un estudio totalmente cualitativo sobre el problema y que permanecen muchos puntos por explorar y evaluar.

5 Conclusiones y líneas futuras

Este artículo ha introducido una aplicación para la clasificación automática de mamografías digitalizadas. Mediante un estudio previo [Golobardes97] que se realizó en un problema similar pero de menor complejidad, se ha decidido realizar la clasificación automática usando como técnica el CBR.

Se ha puesto de manifiesto que el razonamiento basado en casos es una buena técnica, una vez comprobados los resultados, para tratar el problema en cuestión, aunque queda pendiente evaluar otros métodos. Entre estos resultados es importante destacar dos características: la primera, es la importancia de la correcta elección de la función de similitud sobre la que trabaja el sistema; la segunda, es el buen funcionamiento del sistema utilizando una evaluación global de las mamografías.

El problema permanece abierto, y tal como se ha definido en todo este artículo quedan muchos puntos por considerar, y analizar la viabilidad de los mismos. El handicap de nuestro trabajo actual consiste en evaluar los siguientes factores:

(1) La función de similitud es un eje principal dentro del ciclo del CBR, de tal manera

que influye sustancialmente en los resultados. Por este motivo, queremos encontrar otras funciones que se adapten al problema. (2) Estudiar factores como la ponderación de los atributos en las funciones de similitud que encontramos, evaluado ya en algunos trabajos previos [Golobardes97, Golobardes98]. (3) También se ha considerado importante el estudio de los criterios de aceptación (en la fase de revisión) de los resultados generados por el CBR. Es posible que se deban ajustar los criterios de aceptación del sistema en el momento de decidir si una mamografía es benigna, maligna, o simplemente no es posible clasificarla. Así pues, de momento, parece que las técnicas de Aprendizaje Artificial nos permitirán mejorar el diagnóstico de las mamografías en cancerígenas, no cancerígenas, o no clasificables, aumentando al mismo tiempo su fiabilidad.

6 Agradecimientos

Queremos dar las gracias al Instituto de Informática y Aplicaciones de la Universitat de Girona por su colaboración. También queremos agradecer a Enginyeria i Arquitectura La Salle, Universitat Ramon Llull, su apoyo a nuestro grupo de investigación en Sistemas Inteligentes.

Referencias

- [Aamodt94] A. Aamodt, and E. Plaza. "Case-based reasoning: foundational issues, methodological variations, and system approaches". *AI Communication Press*, volume 7, pages 39–59, 1994.
- [Garrell98] J.M. Garrell, and E. Golobardes, and E. Bernadó, and X. Llorà. "Automatic Classification of Mammary Biopsy Images with Machine Learning Techniques". *Engineering of Intelligent Systems (EIS'98)*, Volume 3: Artificial Intelligence, pages 411-418, 11-13 February 1998
- [Garrell99] J.M. Garrell, and E. Golobardes, and E. Bernadó, and X. Llorà. "Automatic diagnosis with Genetic Algorithms and Case-Based Reasoning". *To appear in Artificial Intelligence in Engineering*, Elsevier Science Ltd., 1999
- [Golobardes97] E. Golobardes, and J.M. Garrell. "Avaluació d'un sistema classificador basat en casos per a la diagnosi de biòpsies de teixit de glàndules mamàries". *Jornades d'Intel·ligència Artificial: Noves tendències (JIA'97)*, pages 41-49, 23-35 Octubre 1997.
- [Golobardes98] E. Golobardes "Aportacions al raonament basat en casos per resoldre problemes de classificació". *PhD Thesis, Departamento de Informática, Enginyeria i Arquitectura La Salle, Universitat Ramon Llull, Julio 1998.*
- [Kolodner93] J. Kolodner. "Case-Based Reasoning". Morgan Kaufmann Publishers, 1993
- [Marti98] J. Martí, and X. Cufí, and J. Regincós, et al. "Shape-based features selection for microcalcification evaluation". *Proceedings of the SPIE Medical Imaging Conference on Image Processing*, Vol. 3338, pp. 1215–1224, San Diego (USA), 1998.
- [Shen94] L. Shen, and R.M. Rangayyan, and J.L. Leo Desautels. "Detection and classification of mammographic calcifications". *K.W. Bowyer, and S. Astley, editors, State of the Art in Digital Mammographic Image Analysis*, pp. 198-212. World Scientific Publishing, 1994.
- [Winfields94] D. Winfields, and M. Silbiger, and G.S. Brown, et al. "Technology transfer in digital mammography". *Report of the Joint National Cancer Institute*. Workshop of May 19-20, Invest. Radiol. April, 29(4): pp. 507-515, 1994.