

Diagnóstico de clusters de microcalcificaciones en Mamografías

Autor Ramón Gallardo Caballero

Última modificación jueves, 25 de noviembre de 2010

Esta línea de investigación es una aplicación de la clasificación de patrones y del Análisis de Imágenes al Diagnóstico Asistido por Computador (CADx) de clusters de Microcalcificaciones en Mamografías. Nuestra metodología se centra en el uso de Análisis en Componentes Independientes (ICA) como extractor de características, Algoritmos Genéticos como método de selección y Redes Neuronales como sistema de Clasificación. Como fuente de datos utilizamos la "Digital Database for Screening Mammography" (DDSM).

El cáncer de mama La incidencia del cáncer de mama en la mujer varía entre 40 y 75 de cada 100000, siendo el tumor maligno más frecuente en la población femenina española. Se estima que entorno a 6000 mujeres fallecen como consecuencia de esta enfermedad, por lo que el cáncer de mama constituye la primera causa de muerte por cáncer en mujeres, con una tasa de mortalidad de 28.2 por cada 100000 mujeres, lo que representa el 18.4% del total de muertes por cáncer en mujeres. La probabilidad actual de que una mujer española adquiera un cáncer de mama antes de cumplir 75 años se aproxima a un 5%. Esto significa que una de cada 20 mujeres desarrollará un cáncer de mama antes de esa edad.

Estas estadísticas tan poco prometedoras ilustran la magnitud del problema que representa el cáncer de mama. Un problema, en el que aunque se han identificado una serie de factores de riesgo, no se conocen medidas preventivas eficaces al cien por cien ni una cura específica y efectiva.

La ausencia de un factor de riesgo, distinto de la edad, que sea altamente significativo en la aparición de la enfermedad hace difícil establecer cualquier medida eficaz en la prevención del cáncer de mama. Por lo que en la actualidad, la medida más efectiva en la lucha contra el cáncer de mama es la detección precoz cuando el cáncer se encuentra en un estado temprano de desarrollo.

El método de detección más eficaz y extendido en estos momentos en los distintos sistemas sanitarios es el análisis mamográfico. Una técnica basada en rayos X que proporciona imágenes con un nivel de resolución espacial superior a la ecografía clásica. En el campo del diagnóstico basado en mamografía la mayor parte de los cánceres en desarrollo se detectan en base a unas estructuras denominadas clústeres de microcalcificaciones.

Una microcalcificación es una estructura muy pequeña (típicamente inferior a 1 milímetro), cuando aparecen agrupadas en determinadas configuraciones suelen indicar de forma indirecta la presencia de un cáncer en desarrollo.

La detección de este tipo de estructuras presenta en determinados casos algunas dificultades. Son relativamente pequeñas, en muchos casos se encuentran en zonas de bajo contraste y no podemos despreciar los efectos de estos parámetros sobre el nivel de cansancio/atención del experto. Este último es el motivo fundamental para que la utilización de sistemas de ayuda al diagnóstico (CAD) en este campo, se considere como uno de los métodos posibles que pueda ayudar a reducir las posibilidades de error diagnóstico. Nuestra propuesta Aunque hoy en día empiezan a popularizarse los mamógrafos digitales, hasta ahora la principal fuente de datos para las tareas de investigación ha sido el escaneado de mamogramas. Los escáneres mamográficos proporcionan un nivel de resolución elevado: tamaños de pixel del orden de decenas de micras y valores de profundidad de gris desde los 11 a los 16 bits (desde 2048 a 65536 niveles de gris). Esto nos puede dar una idea del nivel de precisión con el que podemos trabajar con los mamogramas.

Nuestra propuesta de trabajo se basa en la utilización de la técnica conocida como análisis en componentes independientes (ICA), como un sistema eficiente de extracción de características de la imagen; que posteriormente se utilizarán en un clasificador de tipo neuronal. El análisis en componentes independientes es un nuevo método de extracción de características que a diferencia de los métodos clásicos como pueden ser la varianza o la desviación típica utiliza estadísticos de orden superior. Además, en base a muestras del espacio de datos a representar, es capaz de inferir una base en la que nos permite representar con un determinado número de componentes cualquier imagen (señal) del tipo de datos a representar.

Esta misma tarea es la que llevamos a cabo al descomponer una señal en sus componentes de Fourier o en el mundo del análisis multiresolución, al realizar una descomposición en wavelets. Pero existen diferencias significativas con los métodos citados. En primer lugar, tanto en una descomposición de Fourier como wavelet las funciones base son fijas. En una descomposición ICA la base se autogenera para adaptarse lo mejor posible a los datos a representar. Adicionalmente, a diferencia de las técnicas clásicas, un desarrollo ICA genera bases de forma que se maximice la no gaussianidad del espacio de datos de entrada, lo que puede traducirse en que las bases representan las características más interesantes del espacio a modelar. Y es precisamente este último hecho el que nos lleva a utilizar ICA como extractor de características, en vez de otras técnicas más extendidas como puede ser la mencionada transformada wavelet o un análisis en componentes principales.

El segundo elemento importante en nuestra arquitectura lo constituye el clasificador neuronal. No podemos negar que uno de los argumentos importantes a la hora de elegir este tipo de clasificador es la experiencia previa de la que

disponemos en el grupo de investigación con este tipo de sistemas, aspecto que supone una clara ventaja de partida. Pero adicionalmente este tipo de sistemas exhibe características que resultan especialmente interesantes para el problema a afrontar. Quizá la más conocida sea la capacidad de ajustar su funcionamiento en base a "ejemplos", coloquialmente diríamos que tiene capacidad de aprender. Pero otra no menos importante es lo que se conoce como capacidad de generalización. La generalización es la característica de proporcionar una respuesta correcta ante una entrada totalmente desconocida. Como puede deducirse, esta característica hace que un clasificador neuronal sea a priori un elemento muy adecuado para una tarea de clasificación como la que nos ocupa; donde la variabilidad de los datos de entrada es muy grande (variaciones de contraste, errores de toma de la mamografía, de escaneado, artefactos, ...). {mospagebreak title=Metodología empleada} Metodología empleada La base actual de nuestro trabajo es la base de datos mamográfica conocida como Digital Database for Screening Mammography (DDSM). Desarrollada por el grupo Island Vision de la Universidad del sur de Florida es quizá la base de datos de libre distribución más extensa y de mejor calidad. Comprende aproximadamente 2500 casos completos. Es decir, proporciona las cuatro vistas típicas de un estudio mamográfico (Craneo Caudal derecha e izquierda y Medio Lateral Oblicua derecha e izquierda). Adicionalmente suministra información del caso en cuestión como por ejemplo la edad, tipo de película mamográfica, escáner, etc. Pero lo que resulta más interesante a la hora de afrontar nuestro trabajo es el hecho de que proporciona lo que podemos denominar zona de certeza (ground truth) cuando en un pecho existe algún tipo de anomalía verificada histológicamente; especificando el tipo y la distribución de ésta en la nomenclatura internacional conocida como BIRADS de la ACR.

Existen otras bases de datos en este campo como puede ser MIAS o Nijmegen, pero o bien ya no son accesibles o no son gratuitas. La primera de ellas proporciona una versión reducida libre de cargo (miniMIAS), pero dada su escasa resolución espacial y espectral resulta muy poco útil a la hora de afrontar un problema de detección de microcalcificaciones. La fuente de datos Aunque la base de datos escogida (DDSM) parece muy completa, presenta una serie de características que consideramos conveniente modificar. La primera de ellas, aunque fácilmente solventable, es el hecho de que toda la información relativa a cada caso se proporciona en forma de archivos de texto. Dado el gran número de casos y mamogramas disponibles, este no parece ser el mejor formato. Por lo que uno de los primeros pasos fue la inserción de toda la información proporcionada por la DDSM en una base de datos relacional. La segunda de las características a modificar no resulta tan inmediata. Los casos que contiene DDSM proceden de varios centros médicos estadounidenses. La digitalización se realizó en los propios centros con el hardware disponible en estos, por lo que tenemos cuatro tipos distintos de casos en función del escáner utilizado. Los mamogramas se proporcionan en formato bruto (raw), es decir se proporciona el valor de gris capturado por el escáner. Pero cada escáner presenta un valor de gris distinto en función de su calibración. Además uno de los escáneres, a diferencia de los otros 3 lineales, presenta un comportamiento logarítmico; lo que conlleva una mayor precisión en una zona del espectro que en el resto. Debido a estas diferencias el primero de los procedimientos a realizar en este trabajo fue una conversión de los mamogramas a una magnitud común: densidad óptica. Éste es un parámetro con significación física ampliamente utilizado al que podemos convertir los distintos mamogramas gracias a las tablas de calibración de éstos. Generación del conjunto de prototipos Este fue quizá uno de los pasos más lentos del desarrollo dado que en nuestra propuesta como primera aproximación realizamos un diagnóstico a nivel de píxel. Por ello, con ayuda de un experto radiólogo realizamos un etiquetado de píxeles de una serie de clases predefinidas sobre regiones de los mamogramas, una vez expresados éstos en densidad óptica. El conjunto de clases a estudiar incluye no solo microcalcificaciones pertenecientes a un clúster (y por tanto indicativas de malignidad) sino también microcalcificaciones benignas, calcificaciones de centro transparente o uniformes, varillas, tejido saludable y varios tipos de artefactos detectados como por ejemplo los efectos del rodillo de la reveladora de mamogramas. En total generamos un conjunto de entrenamiento con más de 4600 prototipos de microcalcificación maligna, más de 6700 prototipos de microcalcificación benigna y más de 100000 prototipos de tejido saludable o estructuras benignas. Dada la gran cantidad de prototipos disponibles, y teniendo en cuenta el posterior proceso de entrenamiento del clasificador neuronal, decidimos generar varios conjuntos de entrenamientos en los que los porcentajes de prototipos de cada clase apareciesen distribuidos de distinta forma siempre primando la aparición de prototipos de la clase que presenta más interacciones (microcalcificación benigna). Estos conjuntos de prototipos se utilizarán en los distintos procesos de entrenamiento que se llevan a cabo a continuación. Obtención de las bases ICA El extractor de características ICA necesita generarse en base a muestras del espacio que debe modelar (mamogramas). Estas muestras son ventanas cuadradas cuyo lado marca la dimensión del análisis ICA. La estrategia típica de modelado consiste en tomar muestras aleatorias, pero en este caso después de varias pruebas y dada la extensión de las zonas normales respecto de las anomalías en los mamogramas, decidimos centrar las muestras en píxeles pertenecientes al conjunto de prototipos. Una vez se dispone de las matrices ICA la tarea de obtención de características se reduce a realizar operaciones de multiplicación matricial. Así por ejemplo en la imagen siguiente se muestra la descomposición ICA de un pequeño clúster de microcalcificaciones. <!--[if !vml]--><!--[endif]--> Entrenamiento de clasificadores neuronales El entrenamiento se lleva a cabo sobre conjuntos de prototipos preestablecidos que garantizan la posibilidad de obtener una muestra (patch) de al menos 51 píxeles de dimensión. Cada uno de estos conjuntos trata de modelar una característica específica, pero todos ellos incorporan la totalidad de los píxeles malignos. El entrenamiento se lleva a cabo en el clúster distribuido, lo que nos permite realizar unos 42 entrenamientos simultáneos. La estructura neuronal seleccionada es un perceptro multicapa con una única capa oculta cuyo tamaño oscila entre 50 y 200 neuronas. El procedimiento conlleva varias repeticiones a número de neuronas constante y crecimiento neuronal hasta llegar al máximo número de neuronas establecido. La red óptima junto con sus parámetros de rendimiento se almacena para uso posterior. Evaluación de rendimiento La evaluación de rendimiento se lleva a cabo a tres niveles distintos. El primero de ellos se realiza a nivel de prototipos, de forma que se

evalúa la tasa de acierto global y sobre la clase objetivo (microcalcificaciones). El segundo nivel evalúa el rendimiento del sistema en base al diagnóstico de overlays completos, comparando resultados con los diagnósticos reales almacenados en la base de datos. En este caso así como en el tercer nivel necesitamos una herramienta básica que se encarga de auto generar lo que se conoce como regiones de interés. Una región de interés (ROI) es una marca rectangular, circular o libre que delimita la zona en la que se extiende la anomalía indicada. En nuestro caso la generación de regiones de interés presenta diversas fases: Individualización de calcificaciones. Generación de mapa de densidad. Filtrado del mapa de densidad. Individualización de ROIs. Para la evaluación de rendimiento de ROIs hemos propuesto dos criterios distintos. El primero de ellos más grosero es considerar el diagnóstico acertado si el overlay presenta al menos una región de interés marcada por el especialista y nuestro sistema diagnostica al menos una ROI. El segundo exige además que exista solapamiento entre la ROI indicada por el especialista y la generada para considerar acertado el diagnóstico. Siendo este último criterio más restrictivo hemos decidido utilizarlo en lugar del primero. En cuanto al diagnóstico de mamogramas completos, el procedimiento de generación de ROIs es idéntico al utilizado en el diagnóstico de overlays, variando únicamente los criterios de diagnóstico. Nuevamente hemos establecido dos criterios distintos. En el primero de ellos consideramos que una ROI generada es correcta si se solapa con la ROI marcada por el especialista (si ésta existe) y en el segundo exigimos que la ROI generada incluya completamente a la ROI establecida por el especialista. El segundo de ellos presenta dificultades de aplicación, dado que en ocasiones la ROI generada es lo suficientemente pequeña como para no incluir la ROI marcada por el especialista y sin embargo visualmente se detecta una ubicación totalmente correcta. {mospagebreak title=Implementación práctica} Implementación práctica Como puede deducirse del contenido anterior, el sistema que utilizamos es muy completo y algo complejo. En este apartado describimos en profundidad los elementos que lo componen y sus particularidades. Servidor DDSM La base de datos que proporciona el Island Vision Group está accesible mediante FTP, por lo que realizamos un volcado de esta sobre un servidor montado al efecto con una capacidad de almacenamiento de 290 gigabytes de los cuales 240 están dedicados exclusivamente a almacenar los archivos de imagen y de configuración que constituyen la base de datos. Base de datos PostgreSQL Como ya se ha mencionado, la base de datos DDSM proporciona la información de cada uno de los casos en formato archivo. Este formato no es el más adecuado para trabajar de forma intensiva con ellos como es nuestro caso, por lo que decidimos integrar toda esta información en un servidor de bases de datos SQL que corre en uno de nuestros servidores Linux. Al mismo tiempo este motor de base de datos nos sirve para almacenar la información resultante de las distintas fases del proyecto, por lo que disponemos de forma centralizada de todos los datos que podemos necesitar tanto de entrada como de salida. Infraestructura de cálculo Algunas fases del proyecto requieren gran capacidad de cálculo y almacenamiento volátil (memoria). Afortunadamente nuestro grupo de investigación dispone de varios servidores de cálculo multiprocesador con importantes cantidades de memoria que nos han permitido llevar a cabo el proyecto y sin los cuales hubiese sido imposible plantearse afrontarlo. Entre ellos podemos destacar un servidor Opteron biprocesador con 16 Gigabytes de memoria RAM, otro servidor Athlon64 monoprocesador con 4 Gigabytes de memoria y varios servidores biprocesador de 32 bits que sustentan la base de datos y el sistema de ficheros y copia de seguridad. Menciona aparte merece el recurso de cálculo más importante del que disponemos, un cluster tipo Beowulf con 45 nodos AthlonXP-2000 cada uno de ellos con 512 Megabytes de memoria física donde se realizan todas las operaciones de cálculo masivo. Este elemento nos permite reducir de forma dramática los tiempos de cálculo, dado que permite reducir de 40 a 45 veces el tiempo necesario para completar una batería de pruebas determinada. La interfaz de gestión: MAMOPROT Todo trabajo serio con imágenes requiere la creación de una herramienta que permita visualizar éstas y realizar operaciones de marcado y corrección. La herramienta que hemos desarrollado en nuestro caso se denomina Mamoprot y nos ha permitido solventar distintos problemas surgidos en las diferentes etapas de desarrollo de este proyecto. Mamoprot es una aplicación modular escrita en C++ y desarrollada con la librería de widgets QT que corre en entorno Linux. Está completamente integrada con el servidor de base de datos relacional que gestiona el proyecto. Permite múltiples operaciones, de forma que además de las operaciones básicas de visualización (ajuste de contraste, brillo, filtrado, etc.); trabajando sobre overlays permite visualizarlos y marcar prototipos sobre ellos, visualizar o realizar el diagnóstico de un overlay con una configuración especificada, observar la descomposición ICA de éste y recomponerlo en base a algunas componentes, etc. Este primer módulo es el que se utiliza con los expertos a la hora de establecer los prototipos iniciales del sistema. Trabajando sobre mamogramas permite además de visualizarlos con las distintas zonas de interés marcadas por los expertos; obtener el diagnóstico del mamograma completo visualizando las regiones de interés generadas por el sistema de diagnóstico especificado, insertar nuevos overlays en el sistema a partir del mamograma visualizado. Y finalmente, permite añadir nuevas configuraciones de diagnóstico de forma sencilla seleccionando las distintas características del sistema a insertar. Todos estos puntos hacen del mamoprot uno de los pilares de este proyecto. En la siguiente figura puede observarse un ejemplo de utilización del programa en la tarea de visualización del resultado de diagnóstico de un mamograma completo. Donde los rectángulos discontinuos en azul marcan overlays auxiliares no malignos, el rectángulo discontinuo en rojo marca el overlay maligno definido en la base de datos DDSM, el rectángulo en verde marca el overlay correctamente generado por nuestro sistema de diagnóstico y el overlay de borde continuo en azul marca un falso positivo que ha generado el sistema. Puede hacer click en la imagen para observarla con mayor detalle.

{mospagebreak title=Resultados} Resultados En los distintos trabajos encontrados, podemos encontrar formas muy distintas de proporcionar resultados. En primer lugar muchos de los trabajos se limitan únicamente a detectar microcalcificaciones, otros muchos utilizan fuentes de datos en los que realmente es difícil encontrarlas (como por ejemplo usando bases de datos con poca resolución o escasa profundidad de muestreo como por ejemplo 8bits) debido a la baja calidad de la fuente de datos. Otros proponen aproximaciones mucho más radicales diagnosticando ventanas de

cierto tamaño sin descender a nivel de pixel ni preocuparse de identificar lo que se consideran rasgos característicos de malignidad. Y finalmente una característica muy extendida es la utilización de bases de datos ad-hoc o la selección de algunos casos muy específicos de bases de datos más extensas. Este último punto hace muy difícil la comparación de los resultados con otros grupos de investigación, pero lamentablemente es la técnica dominante en este campo. Los trabajos más detallados, suelen presentar un enfoque en el que se realiza detección de microcalcificaciones, agrupación en regiones de interés y evaluación de la malignidad de estas regiones de interés. Esa es la estrategia que decidimos adoptar en nuestro trabajo y respecto de la que presentamos resultados preliminares en este apartado. Análisis ROC En este campo los resultados más completos suelen especificarse mediante curvas ROC (Receiver Operating Characteristic), específicamente midiendo el área encerrada bajo esta curva. Para un clasificador binario dado una curva ROC representa la Sensitividad frente a (1 - Especificidad). Una segunda forma de construir una curva ROC consiste en representar la fracción de True Positives (TP, aciertos) frente a la fracción de False Positives (FP, falso positivo). Esta técnica se puede aplicar tanto a expertos humanos como a clasificadores informáticos (con algunos reparos, suelen usarse curvas FROC) y permite evaluar los resultados de un diagnóstico. Las primeras aplicaciones provienen de la segunda guerra mundial donde se aplicó a la evaluación de las capacidades de detección de los operadores de radar en entornos ruidosos de elementos hostiles. Como se ha especificado, esta técnica utiliza el recuento de TPs y FPs; no existe un criterio uniforme sobre lo que se considera un TP o un FP a la hora de diagnosticar un mamograma. Hay criterios más relajados en los que consideran que una ROI generada es TP si el mamograma efectivamente tiene una ROI positiva y otros más restrictivos exigen solapamiento entre la ROI generada y la existente para catalogar ésta como TP. Debido a la utilización de una arquitectura de generación automatizada, generalmente no se tiene control sobre la forma de crecimiento de las regiones de interés, por lo que en determinadas situaciones una ROI puede crecer de manera excesiva lo que hará que prácticamente siempre incluyese la ROI marcada por el experto y será considerada un TP lo mismo ocurrirá en el caso de obtener un FP agravándose la situación en este caso dado que el mamograma sólo tendrá un FP enmascarando el pernicioso efecto de crecimiento desmedido. Para dar cuenta de este efecto hemos decidido utilizar en vez de el número de falsos positivos, un nuevo parámetro que denominamos Falso Positivo Ponderado por imagen (WFPi). Este parámetro da cuenta tanto del efecto de crecimiento de un posible FP como del posible crecimiento de un TP. De forma que si una ROI diagnosticada como FP presenta un área superior a un área especificada (establecida como el área de la mayor ROI marcada dentro del subconjunto BCRP_CALC en DDSM) contribuirá a este parámetro tantas veces como exceda este valor más uno. Análogamente un TP excesivamente grande contribuirá con el exceso de área a este parámetro, de forma que WFPi se incrementará tantas veces como supere el área máxima especificada para una ROI. Matemáticamente podremos expresarlo como: Somos conscientes que esta forma de proporcionar resultados empeora nuestras estadísticas, pero consideramos firmemente que es la forma más correcta. Resumen de resultados A continuación presentamos algunos resultados relativos al trabajo que hemos realizado hasta el momento, al ser un trabajo en desarrollo estos debe considerarse como resultados provisionales. Así mismo se proporciona un enlace a un subportal en el que proporcionamos una versión extendida de los resultados para las distintas configuraciones establecidas. En primer lugar pueden apreciarse los resultados obtenidos para una de las mejores configuraciones probadas. WFPi WFPi Sens.(%) 0.400.40

24.5 2.182.2054.12.792.8363.83.334.1784.12.784.2088.52.654.3590.43.576.7497.6

A continuación para este mismo sistema mostramos una curva ROC para los resultados obtenidos sobre el conjunto de overlays. Y finalmente para el conjunto de mamogramas puede apreciarse la curva FROC obtenida.

{mospagebreak title=Futuros desarrollos} Futuros desarrollos A corto plazo el trabajo en esta línea se va a centrar en la adición de nuevos prototipos que ayuden a mejorar el rendimiento de los clasificadores. Estos nuevos prototipos se van a insertar en zonas que en las que hemos observado una menor eficiencia de los clasificadores. Una de las fases siguientes va a consistir en optimizar el rendimiento del sistema en distintas facetas del mismo. En primer lugar utilizando algoritmos genéticos vamos a tratar de optimizar algunos clasificadores para evaluar las posibilidades de mejora mediante la selección de características ICA. Finalmente si los resultados obtenidos son buenos evaluaremos las posibilidades de implementación en entorno clínico.