

Segmentación de Imágenes de Microarreglos de ADN usando Técnicas de Reconocimiento de Patrones



Juan Carlos Rojas Thomas

Informe de Tesis
para optar al grado académico de:

Magister en Ciencias de la Computación

Patrocinante: Luis Rueda

Resumen

Bioinformática es una nueva área del conocimiento producto de la utilización de técnicas computacionales para el tratamiento de datos proveniente de la biología. Una de las fuentes de datos tratadas son los genes. En algunas situaciones es deseable conocer el nivel de expresión de los genes de un cierto organismo en determinadas condiciones, para, por ejemplo, averiguar las causas de una enfermedad que le afecta. Los microarreglos de ADN son una técnica utilizada con este objetivo. Estos generan imágenes que corresponden a los niveles de activación de los genes en condiciones diferentes (por ejemplo, sano y enfermo). Cada zona en la cual se registra el nivel de activación del material genético se denomina "spot". Uno de los pasos fundamentales, una vez efectuado el experimento, es reconocer el área correspondiente a cada spot, separándola del fondo y del ruido presente. A este paso se le conoce como "segmentación" de la imagen.

El presente trabajo desarrolla un método de segmentación de imágenes de microarreglos de ADN basado en técnicas de reconocimiento de patrones. El esquema general del proyecto es el siguiente: en la primera etapa del método se realiza una clasificación no supervisada de los píxeles de spots usando técnicas de clustering. Luego, en la segunda etapa se realiza una clasificación supervisada de las regiones de spots, basada en la clasificación inicial de los píxeles. Es esta última clasificación la que debe entregar las zonas de la imagen correspondiente a cada spot. Junto con estas técnicas de reconocimiento de patrones se utilizaron otras herramientas, como el índice pearson para medir la correlación entre la intensidad de un píxel y la intensidad promedio de sus vecinos, de manera de eliminar imágenes que no contienen spots, un algoritmo para detectar bordes de regiones, y operadores morfológicos utilizados al final del proceso para eliminar ruido presente en la segmentación final.

Para la selección y configuración de cada clasificador se hizo un estudio, tanto de diversos clasificadores como de diversos conjuntos de características. Los resultados de estos se compararon entre si para determinar las mejores configuraciones. Para aumentar la precisión se dividió la etapa de la clasificación supervisada en dos pasos, uno dedicado a detectar el fondo de la imagen y otro dedicado a detectar el ruido, cada uno con su particular conjunto de características. Todas las técnicas mencionadas anteriormente dieron como resultado un método que, dada la imagen original de un spot, entrega la

segmentación de ella. Este método fue evaluado aplicándolo sobre un conjunto de spots seleccionados de diversos experimentos de microarreglos y se comparó visualmente la segmentación entregada con la imagen original.

