

Capítulo 4

Resultados y Conclusiones

En el marco del problema de selección de características podemos decir que ambos métodos son exitosos. En nuestro caso de prueba, el HPV, logramos reducir el número de características iniciales de 205 a 14 para el caso de la selección con PCA y de 205 a 4 en el caso de los algoritmos genéticos, representando una reducción en el tamaño del conjunto original de 93.1% y 98% respectivamente. La tabla 4.1 nos muestra la comparación de los mejores resultados de ámbos métodos.

De igual manera, ambos algoritmos necesitan ajustes.

En el caso de la selección usando PCA y RNA's. Fue fácil observar que es necesario el remplazo de las RNA's como clasificadores para las etapas de verificación debido al tiempo requerido para su entrenamiento. En muchos casos éste llegó a ser de días enteros, haciendo que el método sea imposible de aplicar de manera práctica. El remplazo ideal, en el caso de un problema de clasificación simple, es una comparación procedural de los valores de los vectores de características, es decir, verificar algorítmicamente la

Tabla 4.1: Resultados de selección de enzimas usando PCA-RNA

Método	Tiempo de ejecución	$f_{count}(\bar{x})$	$f_{qty}(\bar{x})$	$f_{sep}(\bar{x})$
PCA-RNA	2.5hrs	15	8.136	0
GA	15mins	4	33.702	0

separabilidad de las clases del problema. Una implementación sencilla de dicho método puede ejecutarse en tiempo $o(n^2)$ y este tiempo puede mejorarse si se buscan estructuras de datos que aprovechen las características del conjunto de datos que se esté trabajando.

Otra mejora indispensable en el caso de la selección con PCA es el uso de una técnica de recorrido más compleja después de la selección de los k componentes. Actualmente se recorren las características con base en su coeficiente en la matriz de transformación *TRANS*, y solamente con base en éste. Esto se encontró insuficiente para alcanzar resultados óptimos. Una técnica más compleja, como un recorrido de un árbol o algún tipo de implementación *greedy*, apoyada con la información de la matriz *TRANS* resultaría más efectivo.

En el caso de la selección usando algoritmos genéticos, observamos que estos se comportan satisfactoriamente. Y existen varios factores cruciales en el desempeño del algoritmo, éstos no son difíciles de ajustar.

Los ajustes que hicimos al proceso de creación de la población inicial y mutación resultaron muy exitosos. El primero reduciendo efectivamente el tiempo necesario para la ejecución de cada corrida y el segundo aumentando la presión para la producción de soluciones más pequeñas.

Una observación muy positiva fue el hecho de que el algoritmo probado fue muy similar al algoritmo canónico y existe mucha documentación relativa a la forma de mejorar el algoritmo canónico. Esto nos indica que si aplicamos técnicas más complejas en las diferentes etapas de éste posiblemente obtendremos mejores resultados.

La etapa en la que más impacto causaría un cambio es la etapa de selección de padres y selección de individuos para inserción en la siguiente población. Pudimos observar que lo que más fuertemente limitó la efectividad del algoritmo fue la dominación de las poblaciones por unas cuantas soluciones *fuertes*. Esto causa que el espacio de búsqueda sea limitado y ésta pierda efectividad. Así pues, una técnica de selección que garantice

la diversidad de individuos sin comprometer la presión de la selección sería ideal para el problema de selección de características.

Como trabajo a futuro debemos agregar la necesidad de aplicar ambos procedimientos a otro caso de prueba. Aunque deseamos pensar que los procedimientos propuestos son genéricos y pudieran aplicarse a cualquier caso de selección de características, resulta necesario verificar esta afirmación en al menos otro caso de prueba.