

Título: MODELO PARA LA COMBINACIÓN DE CLASIFICADORES E INTEGRACIÓN DE CONOCIMIENTO APLICADO A DATOS DE TIPO GENÓMICO

Nombre: REBOIRO JATO, MIGUEL

Universidad: Universidad de Vigo

Departamento: Informática

Fecha de lectura: 02/02/2011

Programa de doctorado: Tecnologías Avanzadas para el Desarrollo de Software Inteligente

Dirección:

> **Director:** Daniel González Peña

> **Director:** FLORENTINO FERNÁNDEZ RIVEROLA

Tribunal:

> **presidente:** FERNANDO DÍAZ GÓMEZ

> **secretario:** ROSALÍA LAZA FIDALGO

> **vocal:** PAULO JORGE FREITAS DE OLIVEIRA NOVAIS

> **vocal:** JOSÉ LUIS CAPELO MARTÍNEZ

> **vocal:** MARÍA ELENA BAJO PÉREZ

Descriptor:

> INTELIGENCIA ARTIFICIAL

El fichero de tesis no ha sido incorporado al sistema.

Resumen: En este trabajo se presenta genEnsemble, un modelo para la combinación de clasificadores e integración de conocimiento aplicado a datos de tipo genómico. En este contexto, se ha constatado la necesidad de mejorar los resultados obtenidos en tareas de clasificación automática de muestras, de cara a una aplicación de la tecnología de microrrayas en el diagnóstico clínico de enfermedades complejas como el cáncer. En este sentido, la presente investigación busca aunar los beneficios obtenidos por la integración del conocimiento biológico en técnicas de minería de datos, con las ventajas proporcionadas por la fusión de clasificadores de cara a una posible mejora en la precisión y robustez de los modelos.

El modelo propuesto en la presente investigación es capaz de integrar varios conjuntos de genes durante el proceso de clasificación, utilizando cada uno de ellos como una selección características a aplicar sobre los datos de entrada. La división del espacio generada a partir de los conjuntos de genes es la base sobre la que se asienta geneEnsemble. Utilizando esta división, el modelo entrena varios clasificadores que son posteriormente evaluados para obtener una estimación acerca de su rendimiento esperado. Esta evaluación determina los clasificadores que formarán parte del modelo final. La arquitectura de geneEnsemble contempla la posibilidad de llevar a cabo una sustitución de los tipos de clasificadores base y de las heurísticas utilizadas para la fusión final de clasificadores, por lo que se consigue una alta flexibilidad posibilitando su configuración y adaptación a diferentes contextos.

La justificación de la hipótesis defendida en este trabajo se lleva a cabo de forma experimental, empleando distintos conjuntos de datos reales accesibles a través de Internet. Los resultados obtenidos a partir de los experimentos realizados con el modelo propuesto se comparan con los generados mediante la utilización de distintos algoritmos de clasificación simple y varios esquemas de estándar de combinación de clasificadores, lo que permite la realización de un análisis cuantitativo y cualitativo de la eficacia del sistema desarrollado. Por último, y a la vista de los resultados obtenidos tras la experimentación realizada, se concluye que la integración de conocimiento biológico y la combinación adecuada de clasificadores permite mantener y/o superar la precisión en las clasificaciones, aumentando la robustez al combinar datos experimentales procedentes de distintos laboratorios.