

MetaSelección: Un Marco Evolutivo para la Optimización de Características en Modelos Predictivos

Yoseline Torrealba

Maestría en Ingeniería Eléctrica, Universidad de Guanajuato
yw.torrealbachirinos@ugto.mx

Julio 2025

Resumen

La selección de características representa una etapa crucial en los procesos de análisis de datos. En este artículo se presenta un enfoque híbrido que combina métodos estadísticos y algoritmos de aprendizaje automático para optimizar la selección de atributos relevantes en conjuntos de datos multivariados. Los resultados muestran mejoras significativas en la precisión y eficiencia computacional de los modelos.

1. Introducción

LA selección de características es un paso esencial en el preprocesamiento de datos. Permite eliminar atributos irrelevantes o redundantes, lo cual mejora el desempeño de los algoritmos de aprendizaje supervisado. En esta investigación se exploran técnicas híbridas, combinando filtros estadísticos y métodos *wrapper*, evaluando su impacto sobre distintos conjuntos de datos.

2. Antecedentes

Se han propuesto múltiples métodos para la reducción de características. Guyon y Elisseeff (2003) clasificaron las técnicas en tres categorías: filtros, envoltentes y embebidas. En esta investigación se retoma dicha clasificación y se realiza una comparación práctica de varias estrategias.

3. Metodos y Materiales

A pesar de los avances, pocos trabajos comparan técnicas en diferentes dominios y condiciones experimentales. Este artículo propone una evaluación empírica estandarizada para identificar ventajas y desventajas de cada enfoque, proporcionando así una guía práctica.

4. Experimentos y resultados

Se emplearon tres conjuntos de datos públicos provenientes del repositorio UCI. Las técnicas evaluadas fueron: selección basada en varianza, *SelectK-Best*, RFE y LASSO. Los modelos utilizados incluyen árboles de decisión, SVM y regresión logística. La validación fue realizada con *k-fold cross-validation* ($k=5$).

5. Discución

Los resultados indicaron que LASSO obtiene el mayor rendimiento con menor número de características. En la Tabla 1 se muestra un resumen comparativo.

Cuadro 1: Comparación de precisión con distintas técnicas

Técnica	Dataset 1	Dataset 2	Dataset 3
Varianza	81.2 %	77.5 %	79.8 %
SelectKBest	84.0 %	80.1 %	83.2 %
RFE	85.7 %	82.4 %	85.5 %
LASSO	88.3 %	84.9 %	87.1 %

6. Conclusiones

Este estudio demuestra que la selección adecuada de características incide directamente en el desempeño de modelos de clasificación. Las técnicas híbridas ofrecen una ventaja considerable al combinar enfoques. Como trabajo futuro se plantea el uso de métodos basados en aprendizaje profundo y *autoencoders*.

Referencias

- [1] Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. *Journal of Machine Learning Research*, 3, 1157-1182.
- [2] Liu, H., & Motoda, H. (2005). *Feature Selection for Knowledge Discovery and Data Mining*. Springer.
- [3] Chandrashekar, G., & Sahin, F. (2014). A survey on feature selection methods. *Computers & Electrical Engineering*, 40(1), 16-28.