**Algoritmo de selección envolvente basado en principios de MRMR, y estrategias evolutivas en datos con series de tiempo.**

Liuben López Aparicio 1
1 Centro de Bioplantas. Universidad de Ciego de Ávila, liuben@bioplantas.cu .

**RESUMEN**

La selección de características es el proceso de selección de un subconjunto de características que mejoran el desempeño de los métodos de aprendizaje automático. Este proceso en datos con carácter temporal es más complejo debido a que es posible la pérdida de información temporal. Tomando en cuenta esta premisa la mayoría de enfoques de selección de características no pueden manejar datos temporales multivariados sin aplanamiento previo de los datos, lo que resulta en pérdida de información temporal o la desestimación de esta. En este trabajo se propone un método capaz de manejar datos temporales multivariados sin aplanamiento previo de los datos a partir de la fusión de dos técnicas, algoritmos genéticos y rMRM. El método propuesto toma como plataforma un algoritmo genético que calcula la relevancia de un atributo con su clase en el transcurso de los instantes de tiempo mediante la media de los coeficientes de correlación de Spearman y con los más relevantes conforma los cromosomas, y calcula mediante su función de calidad la redundancia entre los atributos utilizando la distancia de alineamiento temporal para puntuar los cromosomas y obtener los grupos de atributos menos redundantes entre sí. Resultando en la conservación del espacio de los datos y generando un grupo de soluciones fácilmente interpretables.

**Palabras clave:** Selección de características, Expresión de genes, Series de tiempo.

**Abstract**

Feature selection is the process of selecting a subset of features that improve the performance of machine learning methods. This process in temporary data is more complex because it is possible the loss of temporary information. Taking this premise into account, most feature selection approaches can not handle temporary multivariate data without prior flattening of the data, resulting in temporary information loss or rejection of the information. In this work we propose a method capable of handling multivariate temporal data without prior flattening of the data from the fusion of two techniques, genetic algorithms and MRM. The proposed method takes as a platform a genetic algorithm that calculates the relevance of an attribute with its class over the course of time instants by means of the Spearman correlation coefficients and with the most relevant forms the chromosomes, and calculates through its quality function the redundancy between the attributes using the temporal alignment distance to score the chromosomes and obtain the less redundant attribute groups. Resulting in the conservation of the data space and generating a group of easily interpretable solutions.

**Key words:** Feature selection, Gene expression, Time series.