

# A new feature subset selection using bottom up clustering

David Quesada López

Computational Intelligence Group, Departamento de Inteligencia Artificial, Universidad Politécnica de Madrid, Spain

## Objetivos

Queremos seleccionar un subconjunto de variables con un método de filtrado basado en cluster jerárquico para:

- Reducir la dimensionalidad de nuestro dataset.
- Conseguir un subconjunto de variables relevantes y no redundantes que sean igual o más eficientes y eficaces que el conjunto inicial.
- Aprovechar tanto la velocidad del filtrado como la capacidad para agrupar variables del clustering.

## Introducción

La selección del subconjunto de variables es una parte esencial en el preprocesado de los datos para reducir la dimensionalidad. Existen varios métodos para esto:

- Wrapper [1], que seleccionan subconjuntos de variables en base a su precisión pero son muy costosos computacionalmente.
- Filter [3], que miden la relevancia de las variables sin mucho coste pero sin asegurar buena precisión final.

En nuestro algoritmo CFSS usaremos un método de filtrado basado en clustering jerárquico para agrupar las variables redundantes en clusters y elegir las relevantes para nuestro subconjunto. Otros investigadores han usado diferentes tipos de clustering para resolver este problema [2] [4]



Figure 1: Figure caption

## Materiales

Para completar esta investigación se han usado los siguientes materiales:

- 11 datasets válidos del repositorio UCI para probar el funcionamiento de CFSS.
- Los métodos mRMR, ReliefF y L1-LSMI para comparar el rendimiento de CFSS.
- El algoritmo GACH para acotar el número de clusters de CFSS.
- Un clasificador kNN en el que probar los subconjuntos devueltos por los métodos anteriores en sus respectivos datasets.

## Methods

Lorem ipsum dolor **sit amet**, consectetur adipiscing elit. Sed laoreet accumsan mattis. Integer sapien tellus, auctor ac blandit eget, sollicitudin vitae lorem. Praesent dictum tempor pulvinar. Suspendisse potenti. Sed tincidunt varius ipsum, et porta nulla suscipit et. Etiam congue bibendum felis, ac dictum augue cursus a. **Donec** magna eros, iaculis sit amet placerat quis, laoreet id est. In ut orci purus, interdum ornare nibh. Pellentesque pulvinar, nibh ac malesuada accumsan, urna nunc convallis tortor, ac vehicula nulla tellus eget nulla. Nullam lectus tortor, *consequat tempor hendrerit* quis, vestibulum in diam. Maecenas sed diam augue.

## Conclusion

Nunc tempus venenatis facilisis. **Curabitur suscipit** consequat eros non porttitor. Sed a massa dolor, id ornare enim. Fusce quis massa dictum tortor **tincidunt mattis**. Donec quam est, lobortis quis pretium at, laoreet scelerisque lacus. Nam quis odio enim, in molestie libero. Vivamus cursus mi at *nulla elementum sollicitudin*.

## Artículo real

Dehghan Z., Mansoori E. G. (2016). A new feature subset selection using bottom-up clustering. *Pattern Analysis and Applications*, 1-10.

## Referencias

- [1] Kohavi R, John GH (1997) Wrapper for feature subset selection. *Artif Intell* 97(1-2):273-324
- [2] Mitra P, Murthy C, Pal SK (2002) Unsupervised feature selection using feature similarity. *IEEE Trans Pattern Anal Mach Intell* 24(3):301-312
- [3] Peng H, Long F, Ding C (2005) Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundance. *IEEE Trans Pattern Anal Mach Intell* 27(8):1226-1238
- [4] Song Q, Ni J, Wang G (2013) A fast clustering-based feature subset selection algorithm for high-dimensional data. *IEEE Trans Knowl Data Eng* 25(1):1-14



POLITÉCNICA

## Important Result

Lorem ipsum dolor **sit amet**, consectetur adipiscing elit. Sed commodo molestie porta. Sed ultrices scelerisque sapien ac commodo. Donec ut volutpat elit.

## Mathematical Section

Nam quis odio enim, in molestie libero. Vivamus cursus mi at nulla elementum sollicitudin. Nam quis odio enim, in molestie libero. Vivamus cursus mi at nulla elementum sollicitudin.

$$E = mc^2 \quad (1)$$

Nam quis odio enim, in molestie libero. Vivamus cursus mi at nulla elementum sollicitudin. Nam quis odio enim, in molestie libero. Vivamus cursus mi at nulla elementum sollicitudin.

$$\cos^3 \theta = \frac{1}{4} \cos \theta + \frac{3}{4} \cos 3\theta \quad (2)$$

Nam quis odio enim, in molestie libero. Vivamus cursus mi at nulla elementum sollicitudin. Nam quis odio enim, in molestie libero. Vivamus cursus mi at nulla elementum sollicitudin.

$$\kappa = \frac{\xi}{E_{\max}} \quad (3)$$

## Results



Figure 2: Figure caption

Nunc tempus venenatis facilisis. Curabitur suscipit consequat eros non porttitor. Sed a massa dolor, id ornare enim:

Treatments	Response 1	Response 2
Treatment 1	0.0003262	0.562
Treatment 2	0.0015681	0.910
Treatment 3	0.0009271	0.296

Table 1: Table caption