

# Aplicación de algoritmos metaheurísticos para la selección de características en el problema de minería de opinión en Twitter

Suilan Estévez Velarde  
Alejandro Piad Morffis  
Yudivián Almeida Cruz

Curso 2014-2015

## Encontrar opiniones

- Extraer si se da una opinión en los mensajes.
- Extraer si la opinión es negativa, positiva o neutra.

## Encontrar opiniones

- Extraer si se da una opinión en los mensajes.
- Extraer si la opinión es negativa, positiva o neutra.

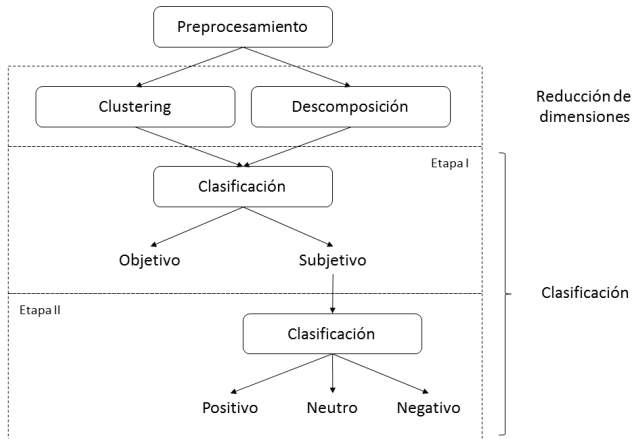
## Problemas

- El volumen de información almacenado en las redes sociales no es procesable por seres humanos.
- La valencia de una opinión es un factor importante para realizar tareas con impacto social, económico o político.
- No existe una propuesta universalmente aceptada que de solución al problema.

## Etapas

- Preprocesamiento o normalización de los datos.
- Reducción de dimensiones.
  - Descomposición.
  - Agrupamiento de los términos.
- Clasificación supervisada.

# Propuesta



# Diseño experimental

## Preprocesamiento

7 fases independientes ( $2^7 = 128$  posibles combinaciones).

## Reducción de dimensiones

11 algoritmos (37 instancias).

## Clasificación

14 algoritmos (74 instancias).

## Total

$\sim 26 \cdot 10^6$  posibles combinaciones ( $\times 30$  corridas).

## Definición

$$\max_{\theta} \{P_{\theta}(X_t, y_t, X_p, y_p) \in [0, 1]\}$$

s.a.

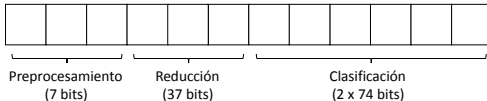
$$\theta \in 2^{[0,1]^K}$$

## Algoritmo Genético

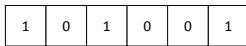
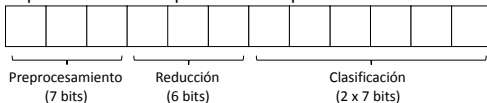
- Representación binaria
- Selección por torneo
- Mutación uniforme
- Cruzamiento por bloques

# Representación

## Representación clásica: 192 componentes



## Representación compacta: 27 componentes



$$= 0.101001_2$$
$$= 0.64025_{10}$$

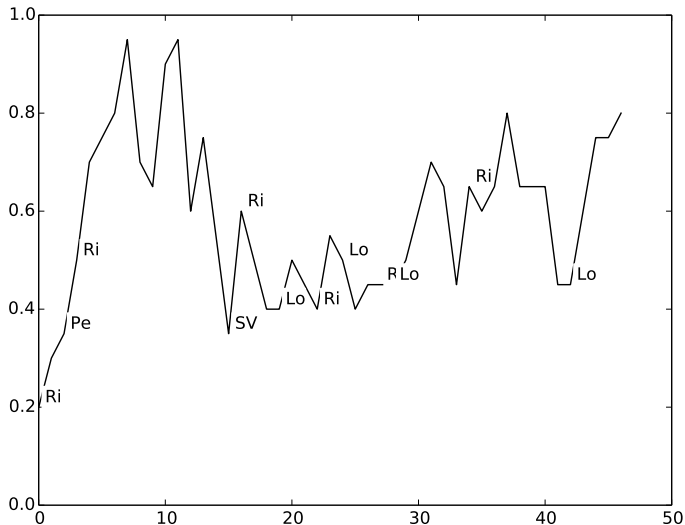
$$[0.64025_{10} \cdot 37] = 23$$

Reducción  
(6 bits)



Comparación		
Parametrización	No	Si
Preprocesamiento	-	-
Reducción	ICA	PCA
Clasificación O-S	Ridge	Log. Reg.
Clasificación P-Ne-N	Ridge	Log. Reg.
Evaluación	0.526	0.560
Precisión	0.464	0.518
Desviación	0.017	0.017

# Experimentación



## Conclusiones

- Con un número mucho menor de experimentos se obtienen resultados satisfactorios.
- Aprende rápidamente a no utilizar *clusters*.
- El Preprocesamiento no es influyente.

## Conclusiones

- Con un número mucho menor de experimentos se obtienen resultados satisfactorios.
- Aprende rápidamente a no utilizar *clusters*.
- El Preprocesamiento no es influyente.

## Recomendaciones

- Mejorar el criterio de evaluación.
- Utilizar multclasificación.
- Crecer el conjunto de entrenamiento.

# Aplicación de algoritmos metaheurísticos para la selección de características en el problema de minería de opinión en Twitter

Suilan Estévez Velarde  
Alejandro Piad Morffis  
Yudivián Almeida Cruz

Curso 2014-2015