

Sesiones RICAPPS, 3 noviembre 2022

Proyecto RAICES-IMIENS-2022

UITeS-ISCIIII/Dpto. LSI-UNED

Ricardo Sánchez de Madariaga

El proyecto RAICES

- **Continuación de los proyectos:**

EXTRAE	IMIENS-2017
EXTRAE II	IMIENS-2019

- **REGLAS DE ASOCIACIÓN EN LA INVESTIGACIÓN DE ENFERMEDADES DE ESPECIAL INTERÉS (*RAICES*)**

- IMIENS-2021/2022

- UITeS-ISCI III *Ricardo Sánchez de Madariaga*
Adolfo Muñoz Carrero
Mario Pascual

- Dpto. LSI-UNED *Lourdes Araujo Serna/Juan Martínez Romo*
Andrés Duque

Proyectos EXTRAE

Algoritmo **EXTRA**E:

*Algoritmo semi-supervisado para etiquetar
Reglas de Asociación de conocimiento médico o
cualquier otro tipo de conocimiento*

Algoritmo EXTRAE

Publicado en:

Sánchez de Madariaga et al.

Semi-supervised incremental learning
with few examples for discovering medical
association rules

BMC Medical Informatics and Decision Making
(2022) 22:20

<https://doi.org/10.1186/s12911-022-01755-3>

Algoritmo EXTRAE

Algoritmo semi-supervisado

Supervisado/No supervisado

Mucho entrenamiento/ningún entrenamiento

Buenas prestaciones/Peores prestaciones

Algoritmo EXTRAE

Cuello de botella de la adquisición de conocimiento:

El entrenamiento es muy caro en términos económicos, de esfuerzo y de tiempo

Algoritmo EXTRAE

Solución:

Algoritmo con poco entrenamiento pero buenas prestaciones

Semi-supervisado

Algoritmo EXTRAE

Antecedentes:

Algoritmo de Yarowsky (1995)

WSD (Desambiguación del Sentido de las Palabras)

Algoritmo EXTRAE

Algoritmo de Yarowsky:

Tesis Doctoral 2008

Funciona peor en un texto genérico y mejor en un texto periodístico estereotipado

Optimización para cualquier tipo de texto genérico

Algoritmo EXTRAE

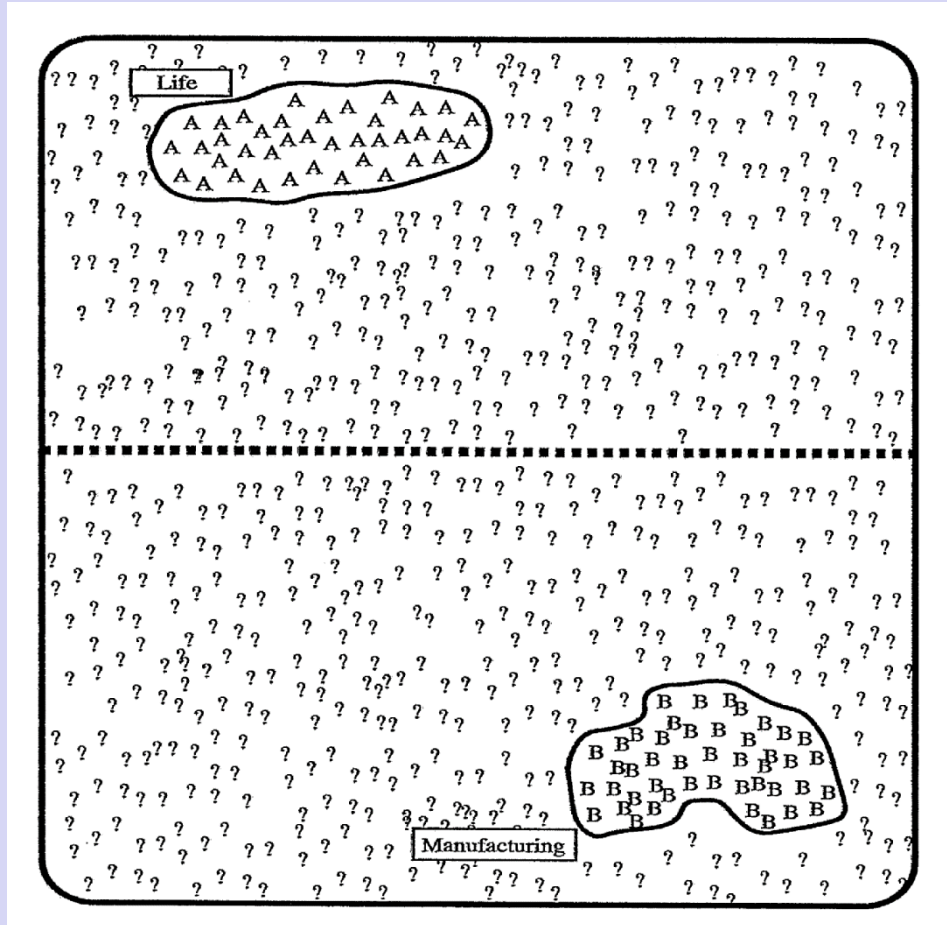
Algoritmo de Yarowsky:

Autoarranque (bootstrapping)

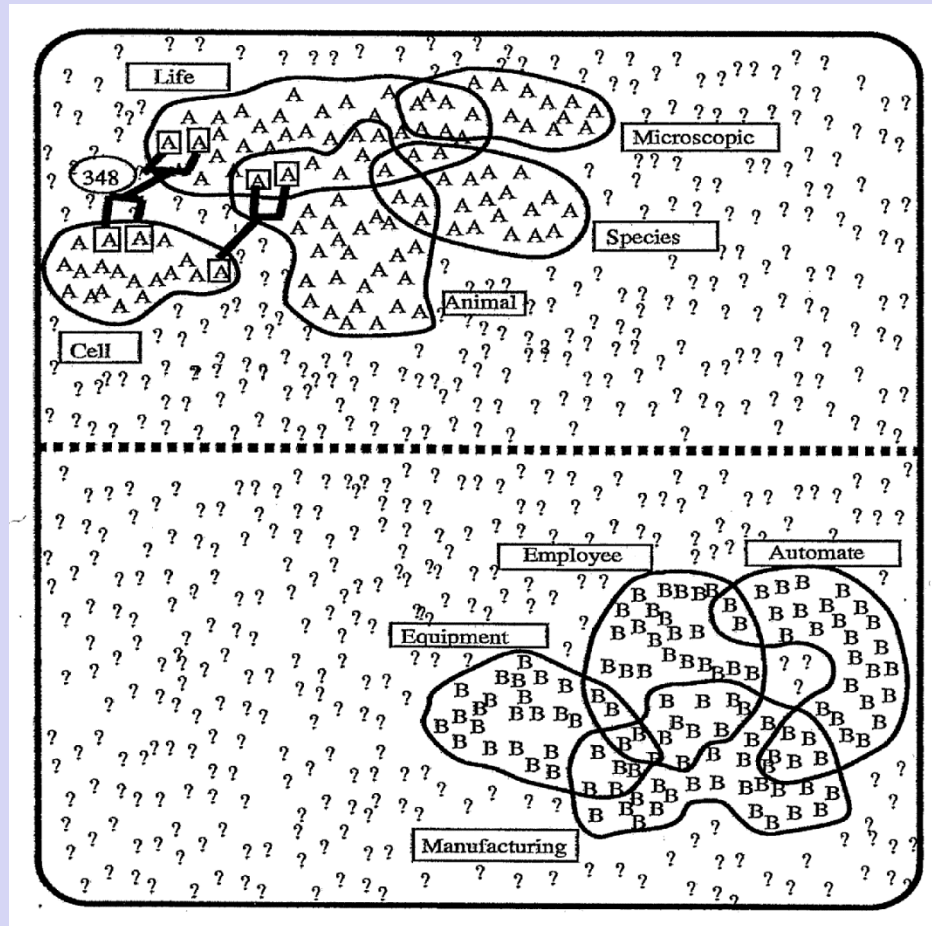
A partir de una semilla de pocos ejemplos etiquetados

Varias Iteraciones

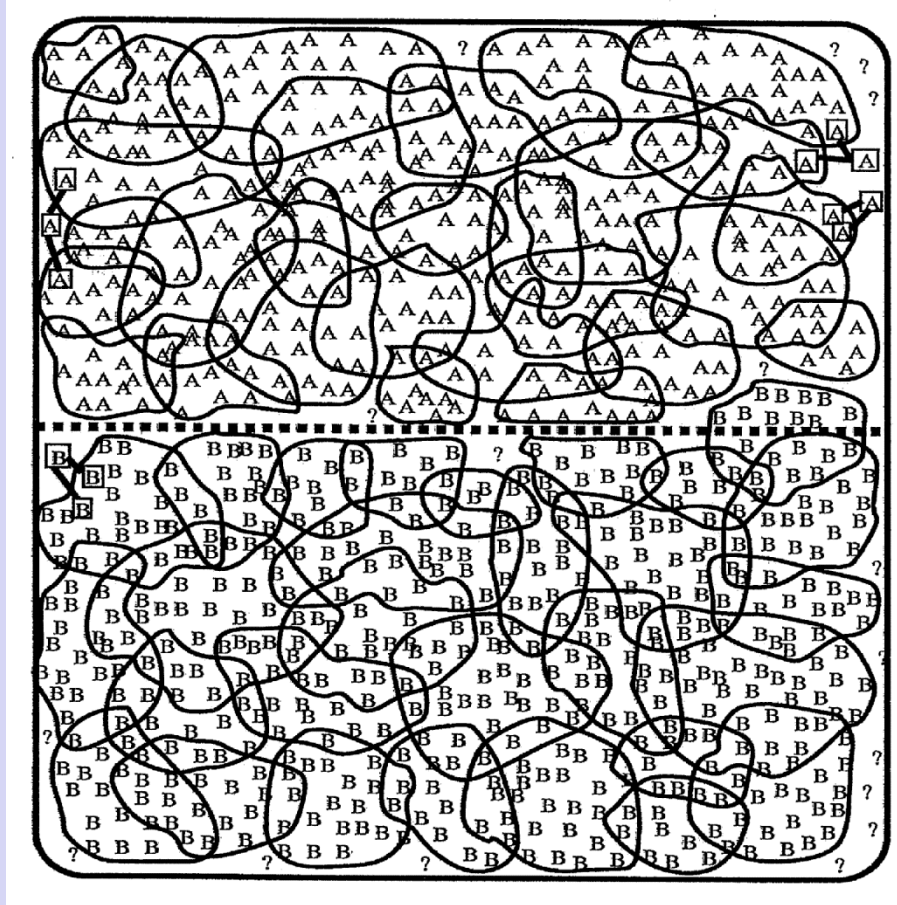
Algoritmo EXTRAE



Algoritmo EXTRAE



Algoritmo EXTRAE



Algoritmo EXTRAE

Base de Datos

Pacientes con Problemas (enfermedades)

Algoritmo FPGrowth genera todas las RA de la forma

$$A B C \Rightarrow D$$

Se puede regular su Precisión y su Exhaustividad mínimas por frecuencias, pero la mayoría de RAs son aleatorias y sin interés

Algoritmo EXTRAE

*EXTRA*E, en cada iteración,
a partir de una semilla mínima de RAs
etiquetadas

Módulo supervisado: algoritmo ML estándar

Módulo no supervisado: test de Fisher

Algoritmo EXTRAE

Test estadístico de Fisher:

p-value de cada RA (estructura de las reglas)

Ordenación: UMBRAL (V ó F)

RA V si V según ML estándar Y según Fisher
 F si F según ML estándar Y según Fisher

Algoritmo EXTRAE

Resultados:

	F-measure	ROC-Area
<i>MEJOR</i>		
<i>Supervisado:</i>	0.71	0.73
<i>EXTRAE</i>		
<i>Semi-supervisado:</i>	0.75	0.80

Proyecto RAICES

RAICES

Reglas de Asociación en la Investigación de Enfermedades de Especial Interés

Aplicamos EXTRAE a el VIH extrayendo los datos de la cohorte CoRIS (Cohorte de la Red de Investigación en SIDA)

Proyecto RAICES

RAICES

Reglas de Asociación en la Investigación de Enfermedades de Especial Interés

Colaboración con el Hospital Universitario 12 de Octubre de Madrid (Dra Otilia Bisbal)

Etiquetado de Dataset CoRIS

Criterios de etiquetado

Proyecto RAICES

Publicado en:

Lourdes Araujo et al.

Discovering HIV related information by means
of association rules and machine learning

Scientific Reports (2022)

Proyecto RAICES

Resultados:

	F-measure	ROC-Area
<i>MEJOR</i> <i>Supervisado:</i>	0.81	0.86
<i>EXTRAE</i> <i>Semi-supervisado:</i>	0.84	0.88

Proyecto RAICES

Se ha llamado la atención sobre relaciones entre enfermedades que ya se conocían en la literatura pero estaban poco estudiadas o se han descubierto nuevas relaciones que merecen ser estudiadas por expertos médicos debidamente cualificados

Proyecto RAICES

Aplicación a Datasets de otras enfermedades de especial interés:

HIV, COVID-19, neoplasia, etc

Relación entre ellas

Proyecto RAICES

Adquisición de nuevo conocimiento médico

Sistematización de conocimiento ya existente

Proyecto RAICES

Muchas gracias por vuestra atención