



PGM-SDA

R. Rumí,  
A. Salmerón

Personal

Objetivos

# Modelos gráficos probabilísticos para analítica escalable de datos

Rafael Rumí, Antonio Salmerón

Departamento de Matemáticas, Universidad de Almería



PGM-SDA Reunión de Coordinación. Granada, 4 de febrero de 2016



PGM-SDA

R. Rumí,  
A. Salmerón

Personal

Objetivos

## Equipo investigador

- Antonio Salmerón
- Rafael Rumí
- Irene Martínez
- Carmelo Rodríguez 😊
- Fernando Reche
- Pepe del Sagrado



# Personal

PGM-SDA

R. Rumí,  
A. Salmerón

Personal

Objetivos

## Equipo investigador

- Antonio Salmerón
- Rafael Rumí
- Irene Martínez
- Carmelo Rodríguez 😊
- Fernando Reche
- Pepe del Sagrado



PGM-SDA

R. Rumí,  
A. Salmerón

Personal

Objetivos

## Equipo de trabajo

- Helge Langseth
- Barry Cobb
- Antonio Fernández 😊😊
- Ana Devaki Maldonado (PhD student).
- Inma Pérez (PhD student) 😊😊 Octubre'15
- Maria Dolores Sánchez García 😊 Enero'16
- Ramón Sáez (CAJAMAR).



PGM-SDA

R. Rumí,  
A. Salmerón

Personal

Objetivos

## Equipo de trabajo

- Helge Langseth
- Barry Cobb
- Antonio Fernández 😊😊
- Ana Devaki Maldonado (PhD student).
- Inma Pérez (PhD student) 😊😊 Octubre'15
- Maria Dolores Sánchez García 😊 Enero'16
- Ramón Sáez (CAJAMAR).



# Objetivos en los que somos líderes

PGM-SDA

R. Rumí,  
A. Salmerón

Personal

Objetivos

## Modelado

- 1 Encapsulación de modelos en RB híbridas. (Antonio S.)** Rafa; Fernando; Helge; Ramón; Andrés C.; Cora; Julia; Ann
- 2 Dependencias funcionales en RB híbridas (Rafa)** Barry; Inma; Carmelo; Serafín; Thomas

## Software

- 7 Desarrollo de la plataforma (Antonio S.)** Irene; Pepe; Antonio F.; Luis de la Ossa; Pablo; Rafa Cabañas; F.J. García; Manuel Gómez

## Aplicaciones

- 11 Predicción del riesgo en instituciones financieras (Rafa)** Antonio S.; Helge; Ramón; Antonio F.



# Objetivos en los que somos líderes

PGM-SDA

R. Rumí,  
A. Salmerón

Personal

Objetivos

## Modelado

- 1 Encapsulación de modelos en RB híbridas. (Antonio S.)** Rafa; Fernando; Helge; Ramón; Andrés C.; Cora; Julia; Ann
- 2 Dependencias funcionales en RB híbridas (Rafa)** Barry; Inma; Carmelo; Serafín; Thomas

## Software

- 7 Desarrollo de la plataforma (Antonio S.)** Irene; Pepe; Antonio F.; Luis de la Ossa; Pablo; Rafa Cabañas; F.J. García; Manuel Gómez

## Aplicaciones

- 11 Predicción del riesgo en instituciones financieras (Rafa)** Antonio S.; Helge; Ramón; Antonio F.



# Objetivos en los que somos líderes

PGM-SDA

R. Rumí,  
A. Salmerón

Personal

Objetivos

## Modelado

- 1 Encapsulación de modelos en RB híbridas. (Antonio S.)** Rafa; Fernando; Helge; Ramón; Andrés C.; Cora; Julia; Ann
- 2 Dependencias funcionales en RB híbridas (Rafa)** Barry; Inma; Carmelo; Serafín; Thomas

## Software

- 7 Desarrollo de la plataforma (Antonio S.)** Irene; Pepe; Antonio F.; Luis de la Ossa; Pablo; Rafa Cabañas; F.J. García; Manuel Gómez

## Aplicaciones

- 11 Predicción del riesgo en instituciones financieras (Rafa)** Antonio S.; Helge; Ramón; Antonio F.



# Encapsulación de modelos en RB híbridas

PGM-SDA

R. Rumí,  
A. Salmerón

Personal

Objetivos

## Idea general

- RB orientadas a objetos (OOBN).
- Modelos que aparecen comúnmente en el análisis de datos → Pasarlos a partes de RB para encapsularlas.
- RB híbridas → MTEs y MoPs para representar modelos log-lineales y modelos aditivos generalizados.

## Beneficios

- Mejora de la precisión de la predicción escalables dado que los parámetros se aprenden de forma independiente
- Definir un proceso Knowledge Engineering (KE) específico para OOBN (Ann)



# Encapsulación de modelos en RB híbridas

PGM-SDA

R. Rumí,  
A. Salmerón

Personal

Objetivos

## Trabajos realizados

Barry R. Cobb, Alan W. Johnson, R. Rumí, A. Salmerón (2015). **Accurate Lead Time Demand Modeling and Optimal Inventory Policies in Continuous Review Systems.** International Journal of Production Economics. 163, 124-136.

## Trabajos anteriores

Barry R. Cobb, Rafael Rumí, Antonio Salmerón (2013). **Inventory management with Log-normal demand per unit time.** Computers and Operations Research 40, 1842-1851.



# Encapsulación de modelos en RB híbridas

PGM-SDA

R. Rumí,  
A. Salmerón

Personal

Objetivos

## A futuro

- Estancia de Ana con Ann en Australia en Septiembre'16, que versará sobre reusabilidad de RB.
- Artículo 'Modeling time in environmental sciences: An approach based on dynamic bayesian networks' a enviar próximamente a Environmetrics, junto con Julia.
- 'Modelling time in species distribution models: a dynamic Bayesian network approach' Enviado al congreso ECOSUBMIT.
- Redacción de un artículo junto con Ann sobre la estancia de Rosa Roperó con Ann, sobre redes dinámicas.



# Encapsulación de modelos en RB híbridas

PGM-SDA

R. Rumí,  
A. Salmerón

Personal  
Objetivos

## Temporización

Qué	Título	Cuándo
Hito (M1)	Desarrollo de los modelos completos	T7
Entregable (D1)	Estado del arte de encapsulación en RB	T3
Entregable (D2)	Informe con las soluciones diseñadas para RB híbridas y OOBN	T8

Temporización Total: T1 - T8



# Dependencias funcionales en RB híbridas

PGM-SDA

R. Rumí,  
A. Salmerón

Personal

Objetivos

## Idea general

- RB son válidas para datos híbridos y para dependencias funcionales, pero independientemente.
- Queremos fusionar ambos problemas.
- Algunas soluciones, sobre todo para relaciones lineales y casos sencillos.
- Queremos ampliarlo a casos complejos y funciones no lineales, como max, min, etc ...
- Así como poder realizar inferencia sobre estos modelos.



# Dependencias funcionales en RB híbridas

PGM-SDA

R. Rumí,  
A. Salmerón

Personal

Objetivos

## Beneficios

- Poder analizar problemas mas complejos.
- La inferencia se puede tratar de dos formas  
a) Reaproximando b) Simulando
- Esencial en el contexto de bigdata, por lo que la opción de IS parece adecuada dado su escalabilidad.



# Dependencias funcionales en RB híbridas

PGM-SDA

R. Rumí,  
A. Salmerón

Personal

Objetivos

## Trabajos anteriores

Prakash P. Shenoy, Rafael Rumí, Antonio Salmerón (2015).  
**Practical aspects of solving hybrid Bayesian networks containing deterministic conditionals.** International Journal of Intelligent Systems 30, 265-291.

## A futuro

Aplicar muestreo por importancia. Colaboración con Serafin. Actualmente implementado, pero probado solo con redes de juguete.



# Dependencias funcionales en RB híbridas

PGM-SDA

R. Rumí,  
A. Salmerón

Personal

Objetivos

## Temporización

Qué	Título	Cuándo
Hito (M1)	Desarrollo de los modelos completa- dos	T7
Entregable (D3)	Estado del arte de dependencias fun- cionales en RB	T2
Entregable (D4)	Informe con la descripción de las solu- ciones para dependencias funcionales en RBs híbridas	T7

Temporización Total: T1 - T7



# Desarrollo de la plataforma

PGM-SDA

R. Rumí,  
A. Salmerón

Personal

Objetivos

## Idea general

- Diferentes grupos desarrollan métodos y algoritmos cada uno en su propio software y sistema.
- Construiremos una SOA (Service Oriented Architecture) para incorporar los algoritmos desarrollados en el proyecto.
- Estos algoritmos se podrán ejecutar a través de una interfaz de servicios web , que incluirá una API para que se pueda usar en móviles Android.
- Esto expandirá la aplicabilidad de los desarrollos metodológicos logrados en el proyecto.
- [Página web del proyecto funcionando.](#)



# Desarrollo de la plataforma

PGM-SDA

R. Rumí,  
A. Salmerón

Personal

Objetivos

## Idea general

- Diferentes grupos desarrollan métodos y algoritmos cada uno en su propio software y sistema.
- Construiremos una SOA (Service Oriented Architecture) para incorporar los algoritmos desarrollados en el proyecto.
- Estos algoritmos se podrán ejecutar a través de una interfaz de servicios web , que incluirá una API para que se pueda usar en móviles Android.
- Esto expandirá la aplicabilidad de los desarrollos metodológicos logrados en el proyecto.
- Página web del proyecto funcionando.



# Desarrollo de la plataforma

PGM-SDA

R. Rumí,  
A. Salmerón

Personal

Objetivos

## Trabajos realizados

- Servidor funcionando (Odin) con SSH, SVN y Apache para gestionar el análisis de requisitos, listo para que todos los grupos accedan a él.
- Plataforma: Osgi (Equinox) + JAVA
- Realizando pruebas sobre R (deal, bnlearn)



# Desarrollo de la plataforma

PGM-SDA

R. Rumí,  
A. Salmerón

Personal  
Objetivos

## Temporización

Qué	Título	Cuándo
Hito (M4)	Prototipo de la plataforma de software	T5
Entregable (D16)	Especificación de requisitos del software	T2
Entregable (D17)	Documento con el diseño del prototipo de la plataforma	T5
Entregable (D18)	Documentación de la plataforma del software y manual de usuario	T12

Temporización Total: T1 - T12



# Predicción del riesgo en instituciones financieras

PGM-SDA

R. Rumí,  
A. Salmerón

Personal

Objetivos

## Idea general

- Predicción del riesgo en operaciones de crédito.
- En particular, tasa de morosidad (default rate) y su relación con operaciones pasadas, de forma que se pueda predecir con antelación.
- Actualmente dichas predicciones se basan en regresión logística y support vector machine fundamentalmente.
- Aplicaremos los algoritmos del objetivo 1 para mejorar la predicción, así como también se hará uso de la plataforma desarrollada en el objetivo 7.



# Predicción del riesgo en instituciones financieras

PGM-SDA

R. Rumí,  
A. Salmerón

Personal

Objetivos

## Beneficios

- Si se identifica con antelación y precisión a un potencial moroso, el banco evita pérdidas.
- Si se detectan buenos clientes, se evita que la entidad rechace operaciones de crédito que finalmente resultarían en beneficios.

## Trabajos realizados

Proyecto AMIDST: El 2016 es el año de finalización, en el que se desarrolla con mayor profundidad la aplicación con CAJAMAR.



# Predicción del riesgo en instituciones financieras

PGM-SDA

R. Rumí,  
A. Salmerón

Personal

Objetivos

## Beneficios

- Si se identifica con antelación y precisión a un potencial moroso, el banco evita pérdidas.
- Si se detectan buenos clientes, se evita que la entidad rechace operaciones de crédito que finalmente resultarían en beneficios.

## Trabajos realizados

Proyecto AMIDST: El 2016 es el año de finalización, en el que se desarrolla con mayor profundidad la aplicación con CAJAMAR.



# Predicción del riesgo en instituciones financieras

PGM-SDA

R. Rumí,  
A. Salmerón

Personal

Objetivos

## Temporización

Qué	Título	Cuándo
Hito (M5)	Análisis de requisitos completado	T9
Entregable (D27)	Estado del arte e informe de requisitos	T9
Entregable (D28)	Prototipo de la solución	T12

Temporización Total: T7 - T12



# Otros objetivos

PGM-SDA

R. Rumí,  
A. Salmerón

Personal

Objetivos

Otros objetivos en los que la UAL participa sin ser responsable:

- O3: Inferencia aproximada y aprendizaje con árboles de probabilidad recursivos. (Manolo) Antonio S.
- O5: Clasificación supervisada no-standar con PGM. (Jose A.) Rafa. Tenemos algunas publicaciones:
  - 'Modeling zero-inflated explanatory variables in hybrid Bayesian network classifiers for species occurrence prediction' Enviado a Environmental Modelling & Software en Diciembre'15
  - 'Continuous Bayesian networks for probabilistic environmental risk mapping', Stochastic Environmental Research and Risk Assessment, in press.



## Otros objetivos

PGM-SDA

R. Rumí,  
A. Salmerón

Personal

Objetivos

Otros objetivos en los que la UAL participa sin ser responsable:

- O5: Clasificación supervisada no-standar con PGM. (Jose A.) Rafa. Tenemos algunas publicaciones:
  - 'Estimation of Species Richness Using Bayesian Networks' CAEPIA'15 y 'Continuous Bayesian networks for the estimation of species richness' Progress in Artificial Intelligence (2015) 4:49 - 57
  - 'Continuous Bayesian networks vs. other methods for regression in environmental modelling' Procedia Environmental Sciences 26 (2015) 70 - 73. Conference Spatial Statistics 2015: Emerging Patterns
  - 'Supervised Classification using Hybrid Probabilistic Decision Graphs' a enviar a Journal of Statistical Software próximamente.
- O6: Aprendizaje de RB a partir de streams de datos. (Andrés C.) Antonio S.
  - Colaborar con AMIDST



# Interacción entre objetivos

PGM-SDA

R. Rumí,  
A. Salmerón

Personal

Objetivos

## Interacción entre los diferentes objetivos

1 → 7,11

2 → 7,11

3 → 6,7,8,9

4 → 5,7

5 → 7,8,9,10,12

6 → 4,5,7,10

7 → 8,9,10,11,12



# Interacción entre objetivos

PGM-SDA

R. Rumí,  
A. Salmerón

Personal

Objetivos

## Interacción entre los diferentes objetivos

1 → 7,11

2 → 7,11

3 → 6,7,8,9

4 → 5,7

5 → 7,8,9,10,12

6 → 4,5,7,10

7 → 8,9,10,11,12