



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. Puerto Lumbreras, 8 de marzo de 2017



# Personal

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



# Personal

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 de los que somos responsables

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 de los que somos responsables

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 de los que somos responsables

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

- B.R. Cobb, A.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.
- A.D. Maldonado, P.A. Aguilera, A. Salmerón (2016) **Continuous Bayesian networks for probabilistic environmental risk mapping.** *Stochastic Environmental Research and Risk Assessment* 30, 1441-1455.
- R. Fernández-Roperero, M.J. Flores, R. Rumí, P.A. Aguilera (2017) **Applications of hybrid dynamic Bayesian networks to water reservoir management.** *Environmetrics* 28 (1), e2432.
- D. Ramos-López, A.R. Masegosa, A.M. Martínez, A. Salmerón, T.D. Nielsen, H. Langseth, A.L. Madsen (2017) **MAP inference in dynamic hybrid Bayesian networks.** *Progress in Artificial Intelligence*, In press.



# Encapsulación de modelos en RB híbridas

PGM-SDA

R. Rumí,  
A. Salmerón

Personal

Objetivos

- R. Fernández-Ropero, M.J. Flores, R. Rumí, P. Aguilera (2016) **Modelling time in species distribution models: a dynamic Bayesian network approach**. 5th International ECOSUMMIT.
- R. Fernández-Ropero (2016) **Hybrid Bayesian networks: A statistical tool in ecology and environmental sciences**. PhD Thesis. Directores: R. Rumí, P.A. Aguilera.

## A futuro

- A.D. Maldonado, A. Nicholson (2017) **Modeling sustainability using OOBNS**.



# Encapsulación de modelos en RB híbridas

PGM-SDA

R. Rumí,  
A. Salmerón

Personal

Objetivos

- R. Fernández-Ropero, M.J. Flores, R. Rumí, P. Aguilera (2016) **Modelling time in species distribution models: a dynamic Bayesian network approach**. 5th International ECOSUMMIT.
- R. Fernández-Ropero (2016) **Hybrid Bayesian networks: A statistical tool in ecology and environmental sciences**. PhD Thesis. Directores: R. Rumí, P.A. Aguilera.

## A futuro

- A.D. Maldonado, A. Nicholson (2017) **Modeling sustainability using OOBNS**.



# 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

## 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

P.P. Shenoy, R. Rumí, A. 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 a MTEs. Colaboración con Serafin. Actualmente implementado, pero probado solo con redes de juguete.
- D. Ramos-López, A. Salmerón, R. Rumí, A.R. Masegosa, H. Langseth, A.L. Madsen, T.D. Nielsen (2017) **Scaling up importance sampling in conditional linear Gaussian networks.** Para enviar al IJAR.



# Dependencias funcionales en RB híbridas

PGM-SDA

R. Rumí,  
A. Salmerón

Personal

Objetivos

## Temporización

Qué	Título	Cuándo
Hito (M1)	Desarrollo de los modelos completados	T7
Entregable (D3)	Estado del arte de dependencias funcionales en RB	T2
Entregable (D4)	Informe con la descripción de las soluciones 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.
- Por desgracia, a Odín ya le llegó su Ragnarok y ahora está en el servidor de la UAL.
- Plataforma: Osgi (Equinox) + JAVA
- Realizando pruebas sobre R (deal, bnlearn)
- Repositorio de software en funcionamiento.
- Paquete `ramidst` en el repositorio oficial de R.
- M.D. Sánchez-García, J. del Sagrado, A. Salmerón, R. Rumí (2016) **PGMs4SDA: a public repository for Probabilistic Graphical Models**. Procs of the 28th Benelux Conference On Artificial Intelligence (**BNAIC'16**), pp. 241 - 242.



# 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.
- Por desgracia, a Odín ya le llegó su Ragnarok y ahora está en el servidor de la UAL.
- Plataforma: Osgi (Equinox) + JAVA
- Realizando pruebas sobre R (deal, bnlearn)
- Repositorio de software en funcionamiento.
- Paquete `ramidst` en el repositorio oficial de R.
- M.D. Sánchez-García, J. del Sagrado, A. Salmerón, R. Rumí (2016) **PGMs4SDA: a public repository for Probabilistic Graphical Models**. Procs of the 28th Benelux Conference On Artificial Intelligence (**BNAIC'16**), pp. 241 - 242.



# 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.
- Por desgracia, a Odín ya le llegó su Ragnarok y ahora está en el servidor de la UAL.
- Plataforma: Osgi (Equinox) + JAVA
- Realizando pruebas sobre R (deal, bnlearn)
- **Repositorio de software en funcionamiento.**
- **Paquete ramidst** en el repositorio oficial de R.
- M.D. Sánchez-García, J. del Sagrado, A. Salmerón, R. Rumí (2016) **PGMs4SDA: a public repository for Probabilistic Graphical Models**. Procs of the 28th Benelux Conference On Artificial Intelligence (**BNAIC'16**), pp. 241 - 242.



# 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.

## Trabajo en preparación

- A. Fernández, D. Ramos-López, A.R. Masegosa, R. Rumí, A. Salmerón (2017) **Credit risk modeling with the AMIDST toolbox: A real case study on the use of Bayesian networks to improve default prediction.** Para enviar a *Journal of Banking and Finance*.





# 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.

## Trabajo en preparación

- A. Fernández, D. Ramos-López, A.R. Masegosa, R. Rumí, A. Salmerón (2017) **Credit risk modeling with the AMIDST toolbox: A real case study on the use of Bayesian networks to improve default prediction.** Para enviar a [Journal of Banking and Finance](#).



# 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 en los que la UAL participa

PGM-SDA

R. Rumí,  
A. Salmerón

Personal

Objetivos

- **O3: Inferencia aproximada y aprendizaje con árboles de probabilidad recursivos.** (Manolo) Antonio S.
- **O5: Clasificación supervisada no-estándar con PGM.** (Jose A.) Rafa.
- **O6: Aprendizaje de RB a partir de streams de datos.** (Andrés C.) Antonio S.



## O5: Clasificación supervisada no-estándar con PGM

PGM-SDA

R. Rumí,  
A. Salmerón

Personal

Objetivos

- A.D. Maldonado, P.A. Aguilera, A. Salmerón (2016) **Modeling zero-inflated explanatory variables in hybrid Bayesian network classifiers for species occurrence prediction.** *Environmental Modelling & Software* 82, 31-43.
- A.D. Maldonado, R.F. Roper, P.A. Aguilera, R. Rumí, A. Salmerón (2015) **Continuous Bayesian networks for the estimation of species richness.** *Progress in Artificial Intelligence* 4, 49-57.
- A.D. Maldonado, R.F. Roper, P.A. Aguilera, R. Rumí, A. Salmerón (2015) **Estimation of species richness Using Bayesian networks.** *CAEPIA'2015*. Lecture Notes in Artificial Intelligence 9422, 153-163.



## O5: Clasificación supervisada no-estándar con PGM

PGM-SDA

R. Rumí,  
A. Salmerón

Personal

Objetivos

- A.D. Maldonado, R.F. Roperro, P.A. Aguilera, R. Rumí, A. Salmerón (2015) **Continuous Bayesian networks vs. other methods for regression in environmental modelling**. *Procedia Environmental Sciences* 26, 70-73.
- J.C. Luengo, R. Rumí (2016) **Mixtures of Polynomials for Regression Problems**. *DMIN'16 - WORLDCOMP'16* Proceedings of the 2016 International Conference on Data Mining, pp. 208-214
- J. del Sagrado, A. Fernández, R. Rumí, A. Salmerón (2017) **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

PGM-SDA

R. Rumí,  
A. Salmerón

Personal

Objetivos

- I. Pérez-Bernabé, A. Fernández, R. Rumí, A. Salmerón (2016) **Parameter learning in hybrid Bayesian networks using prior knowledge**. *Data Mining and Knowledge Discovery* 30, 576-604.
- I. Pérez-Bernabé, A. Salmerón, H. Langseth (2015) **Learning conditional distributions using Mixtures of Truncated Basis functions**. *ECSQARU'2015*. Lecture Notes in Artificial Intelligence 9161, 397-406.
- I. Pérez-Bernabé (2015) **Learning hybrid Bayesian networks using mixtures of truncated basis functions**. PhD thesis. Directores: *A. Salmerón, H. Langseth*.
- A.R. Masegosa, D. Ramos-López, T.D. Nielsen, H. Langseth, A. Salmerón, A.L. Madsen (2017) **Bayesian models of data streams with Hierarchical Power Priors**. Enviado a *ICML 2017*.



# Resumen de publicaciones

PGM-SDA

R. Rumí,  
A. Salmerón

Personal

Objetivos

	<b>Q1</b>	<b>&lt; Q1</b>	<b>No JCR</b>	<b>Congresos</b>	<b>Tesis</b>
<b>Publicados</b>	5	1	3	5	2
<b>En preparación</b>	5	-	-	1	-