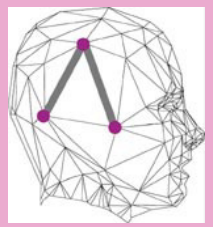
PGM-SDA Project

Probabilistic graphical models for scalable data analytics

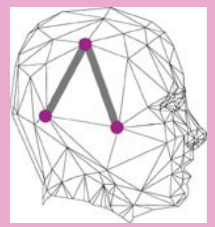
UCLM subproject - second report

J.M. Puerta
J.A. Gámez
Departamento de Sistemas Informáticos
UCLM - Albacete



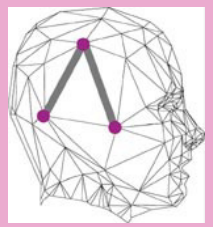
Contents

- Team
- (sub)Project objectives
  - Modeling
  - Learning
  - Software
  - Applications



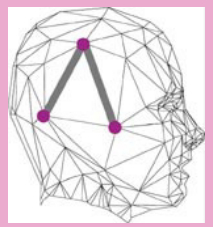
People at UCLM sub-project

- Research Team:
  - José M Puerta
  - José A. Gámez
  - Juan Angel Aledo
  - M. Julia Flores
  - Luis de la Ossa
  - Pablo Bermejo
  - Jacinto Arias
  - Javier Cózar
- Foreign collaborators
  - Thomas D. Nielsen (Aalborg)
  - Ann E. Nicholson (Monash)



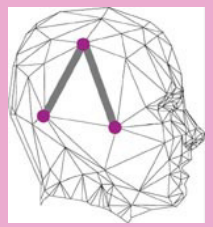
Participation of UCLM node

- Responsibles:
  - 4. Scalable algorithms for Probabilistic Graphical Models learning under MapReduce paradigm.
  - 5. Non-standard supervised classification using PGMs.
  - 10. Improving automation of multimedia complex tasks.
  - 12. Biomedicine applications
- Collaborators:
  - 1. Model encapsulation within hybrid Bayesian networks
  - 7. Software platform development.
- Marginal collaboration:
  - 2. Functional dependencies in hybrid Bayesian networks (Nielsen)
  - 3. Approximate inference and learning with recursive probability trees (Puerta)
  - 6. Learning Bayesian networks from data streams (Gámez)



Objective 1: Model encapsulation within hybrid BNs

- Responsible: Salmerón
- Participants: Rumí; Reche; Langseth; Sáez; Cano; Pérez-Ariza; Flores; Nicholson
- Execution period: T1 - T8
- Milestones and Deliverables:
  - M1 - Modeling developments successfully completed, T7
  - D1 - State of the art of model encapsulation in BN literature, T3
  - D2 - Report describing the solutions designed for hybrid BNs and OOBNs, T8

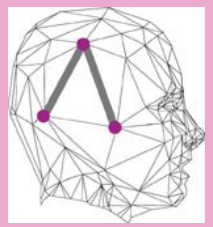


Objective 1: Model encapsulation within hybrid BNs

- Responsible: Salmerón
- Participants: Rumí; Reche; Langseth; Sáez; Cano; Pérez-Ariza; Flores; Nicholson
- Execution period: T1 - T8
- Milestones and Deliverables:
  - M1 - Modeling developments successfully completed, T7
  - D1 - State of the art of model encapsulation in BN literature, T3
  - D2 - Report describing the solutions designed for hybrid BNs and OOBNs, T8

Tasks

- Construction from expert knowledge. KE-procedure specific for OOBNs. Applications based on OOBNs.



Objective 1: Model encapsulation within hybrid BNs

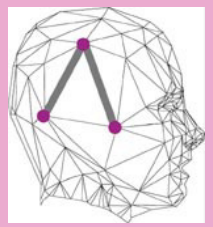
- Responsible: Salmerón
- Participants: Rumí; Reche; Langseth; Sáez; Cano; Pérez-Ariza; Flores; Nicholson
- Execution period: T1 - T8
- Milestones and Deliverables:
  - M1 - Modeling developments successfully completed, T7
  - D1 - State of the art of model encapsulation in BN literature, T3
  - D2 - Report describing the solutions designed for hybrid BNs and OOBNs, T8

Tasks

- Construction from expert knowledge. KE-procedure specific for OOBNs. Applications based on OOBNs.

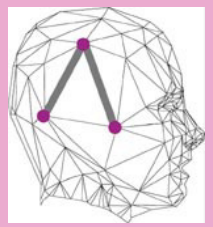
Results

- T3 ? State of the art of the encapsulation...



Objective 4: Scalable Algorithms for Probabilistic Graphical Models Learning under MapReduce Paradigm

- Responsible: Puerta
- Participants: Gámez, de la Ossa, Nielsen, Cano, Cabañas, Gómez.
- Execution period: T1 - T12
- Milestones and Deliverables:
  - M3 - Learning developments successfully completed, T8
  - D8 - State of the art of current MapReduce-based approaches in PGMs research, T4
  - D9 - Report describing the solutions designed for parameter learning, T8
  - D10 - Report describing the solutions designed for structural learning, T12

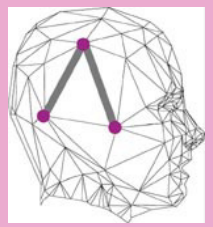


Objective 4: Scalable Algorithms for Probabilistic Graphical Models Learning under MapReduce Paradigm

- Responsible: Puerta
- Participants: Gámez, de la Ossa, Nielsen, Cano, Cabañas, Gómez.
- Execution period: T1 - T12
- Milestones and Deliverables:
  - M3 - Learning developments successfully completed, T8
  - D8 - State of the art of current MapReduce-based approaches in PGMs research, T4
  - D9 - Report describing the solutions designed for parameter learning, T8
  - D10 - Report describing the solutions designed for structural learning, T12

Tasks

- Design of PGM learning algorithms under the MapReduce paradigm. Structural Learning and Parameter Learning. Big Data



Objective 4: Scalable Algorithms for Probabilistic Graphical Models Learning under MapReduce Paradigm

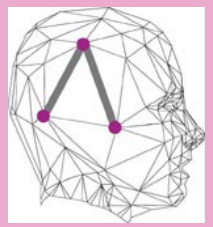
- Responsible: Puerta
- Participants: Gámez, de la Ossa, Nielsen, Cano, Cabañas, Gómez.
- Execution period: T1 - T12
- Milestones and Deliverables:
  - M3 - Learning developments successfully completed, T8
  - D8 - State of the art of current MapReduce-based approaches in PGMs research, T4
  - D9 - Report describing the solutions designed for parameter learning, T8
  - D10 - Report describing the solutions designed for structural learning, T12

Tasks

- Design of PGM learning algorithms under the MapReduce paradigm. Structural Learning and Parameter Learning. Big Data

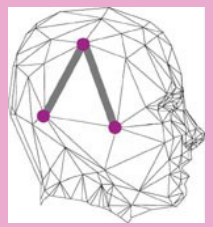
Results

- J. Arias, J. A. Gámez, J. M. Puerta: Structural Learning of Bayesian Networks Via Constrained H.C. Algorithms: Adjusting Trade-off between Efficiency and Accuracy. IJIS 30(3): 292-325 (2015)
- J. Arias, J. A. Gámez, J. M. Puerta: Scalable Learning of k-dependence Bayesian Classifiers under MapReduce. TrustCom/BigDataSE/ISPA (2) 2015: 25-32
- J. Arias, J. A. Gámez, J. M. Puerta: Distributed Discrete Bayesian Network Classifiers under MapReduce with Apache Spark. Submitted to KNOSYS.



Objective 5: Non-standard supervised classification using PGMs

- Responsible: Gámez
- Participants: Puerta, Flores, Bermejo, Nielsen, Rumí, García-Castellano, Masegosa.
- Execution period: T1 - T10
- Milestones and Deliverables:
  - M3 - Learning developments successfully completed, T8
  - D11 - State of the art: non-standard supervised classification with PGMs, T2
  - D12 - Report: designed preprocessing algorithms and classifiers, T6
  - D13 - Report : designed classifiers to deal with combined problems, T10

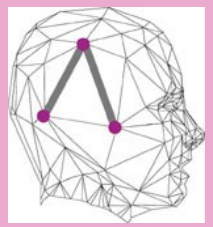


Objective 5: Non-standard supervised classification using PGMs

- Responsible: Gámez
- Participants: Puerta, Flores, Bermejo, Nielsen, Rumí, García-Castellano, Masegosa.
- Execution period: T1 - T10
- Milestones and Deliverables:
  - M3 - Learning developments successfully completed, T8
  - D11 - State of the art: non-standard supervised classification with PGMs, T2
  - D12 - Report: designed preprocessing algorithms and classifiers, T6
  - D13 - Report : designed classifiers to deal with combined problems, T10

Tasks

- Development of scalable algorithms based on PGMs for non-standard supervised classification problems.
- Development of pre-processing algorithms specific for non-standard classification



Objective 5: Non-standard supervised classification using PGMs

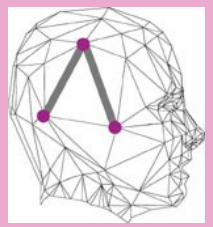
- Responsible: Gámez
- Participants: Puerta, Flores, Bermejo, Nielsen, Rumí, García-Castellano, Masegosa.
- Execution period: T1 - T10
- Milestones and Deliverables:
  - M3 - Learning developments successfully completed, T8
  - D11 - State of the art: non-standard supervised classification with PGMs, T2
  - D12 - Report: designed preprocessing algorithms and classifiers, T6
  - D13 - Report : designed classifiers to deal with combined problems, T10

Tasks

- Development of scalable algorithms based on PGMs for non-standard supervised classification problems.
- Development of pre-processing algorithms specific for non-standard classification

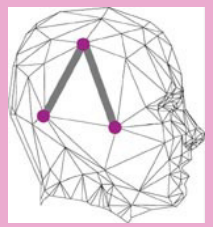
Results

- J. Arias, J. A. GÃ¡mez, T. D. Nielsen, J. M. Puerta: A scalable pairwise class interaction framework for multidimensional classification. Int. J. Approx. Reasoning 68: 194-210 (2016)
- J. Flores and J.A. Gámez: Impact on Bayesian Networks Classifiers When Learning from Imbalanced Datasets. ICCART 2015.



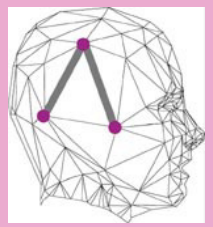
Objective 7: Software platform development

- Responsible: Salmerón
- Participants: Martínez, del Sagrado, Fernández, de la Ossa, Bermejo, Cabañas, García-Castellano, Gómez.
- Execution period: T1 - T12
- Milestones and Deliverables:
  - M4 - A running prototype of the software platform, T5
  - D16 - Software requirements specification, T2
  - D17 - Design document of the prototype software platform, T5
  - D18 - Documentation of the software platform and user's handbook, T12



Objective 10: Improving automation of multimedia complex tasks

- Responsible: Puerta
- Participants: Gámez, de la Ossa, Flores
- Execution period: T5 - T12
- Milestones and Deliverables:
  - M5 - Requirement analysis completed, T9
  - D25 - Description of the PGM-based Transcoder, T8
  - D26 - Description of the algorithms for image evaluation, T12

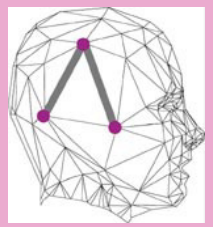


Objective 10: Improving automation of multimedia complex tasks

- Responsible: Puerta
- Participants: Gámez, de la Ossa, Flores
- Execution period: T5 - T12
- Milestones and Deliverables:
  - M5 - Requirement analysis completed, T9
  - D25 - Description of the PGM-based Transcoder, T8
  - D26 - Description of the algorithms for image evaluation, T12

Tasks

- Video.- Transcoding video streams from current standard H.264/AVC to the coming one HEVC
- Images.- PGMs-based supervised classification algorithms for automatic annotation and aesthethic classification.



Objective 10: Improving automation of multimedia complex tasks

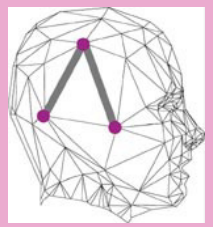
- Responsible: Puerta
- Participants: Gámez, de la Ossa, Flores
- Execution period: T5 - T12
- Milestones and Deliverables:
  - M5 - Requirement analysis completed, T9
  - D25 - Description of the PGM-based Transcoder, T8
  - D26 - Description of the algorithms for image evaluation, T12

Tasks

- Video.- Transcoding video streams from current standard H.264/AVC to the coming one HEVC
- Images.- PGMs-based supervised classification algorithms for automatic annotation and aesthethic classification.

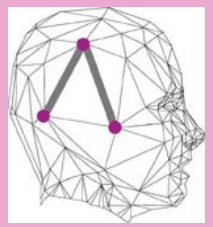
Results

- Adaptive Fast Quadtree Level Decision Algorithm for H.264 to HEVC Video Transcoding. IEEE Transactions on Circuits and Systems for Video Technology. 26(1): 154-168 (2016)
- Low Complexity Heterogeneous Architecture for H.264/HEVC Video Transcoding. J Real-Time Image Proc. DOI 10.1007/s11554-014-0477-z. (2015)
- Fast QuadTree Level Decision Algorithm For H.264/HEVC Transcoder. ICIP 2014.
- A Data-Driven Probabilistic CTU Splitting Algorithm for Fast H.264/HEVC Video Transcoding. DCC 2015



Objective 12: BioMedicine applications

- Responsible: Gámez
- Participants: Puerta, de la Ossa, Bermejo, Nicholson
- Execution period: T4 - T11
- Milestones and Deliverables:
  - M5 - Requirement analysis completed, T9
  - D29 - Description of the method and software system created for COPD readmission prediction, T7
  - D30 - Description of the method and software for cell identification, T11

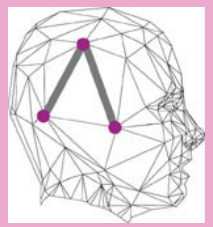


Objective 12: BioMedicine applications

- Responsible: Gámez
- Participants: Puerta, de la Ossa, Bermejo, Nicholson
- Execution period: T4 - T11
- Milestones and Deliverables:
  - M5 - Requirement analysis completed, T9
  - D29 - Description of the method and software system created for COPD readmission prediction, T7
  - D30 - Description of the method and software for cell identification, T11

Tasks

- Chronic Obstructive Pulmonary Disease (COPD) is one of the diseases that cause more re-admissions in our hospitals. To predict readmission of patients after 2 and 4 weeks.
- Cell identification from microscopic images.



Objective 12: BioMedicine applications

- Responsible: Gámez
- Participants: Puerta, de la Ossa, Bermejo, Nicholson
- Execution period: T4 - T11
- Milestones and Deliverables:
  - M5 - Requirement analysis completed, T9
  - D29 - Description of the method and software system created for COPD readmission prediction, T7
  - D30 - Description of the method and software for cell identification, T11

Tasks

- Chronic Obstructive Pulmonary Disease (COPD) is one of the diseases that cause more re-admissions in our hospitals. To predict readmission of patients after 2 and 4 weeks.
- Cell identification from microscopic images.

Results

- Automatic quantization of the subcellular localization of chimeric GFP protein supported by a two-level Naive Bayes Classifier. ESWA. 42 (2015) 1531?1537. http://dx.doi.org/10.1016/j.eswa.2014.09.052.