PGM-SDA Project

Probabilistic graphical models for scalable data analytics

UCLM subproject - third 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
    - Fernando Rubio
    - Enrique Martínez
- 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
- T8 Report describing the solutions designed
- Julia Flores, Ann E. Nicholson, Rosa F. Ropero (2016): Dynamic OOBNs applied to water management in dams. Proceedings of the Int. Conference on Knowledge Engineering and Applications (ICKEA 2016), 1-8, IEEE Press. DOI: 10.1109/ICKEA.2016.7803030

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

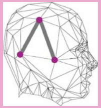
- Responsible: Puerta
- Participants: Gámez, de la Ossa, Nielsen, Cano, Cabañas, Gómez.
- Execution period: T1 - T12
- Milestones and Deliverables:
  - M3 - Learning developments successfully completed, T8
  - D8 - State of the art of current MapReduce-based approaches in PGMs research, T4
  - D9 - Report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

- Jacinto Arias, Jose A. Gámez, Jose M. Puerta (2015): Structural Learning of Bayesian Networks Via Constrained Hill Climbing Algorithms: Adjusting Trade-off between Efficiency and Accuracy. International Journal of Intelligent Systems 30(3), 292-325. DOI:10.1002/int.21701
- Jacinto Arias, Jose A. Gámez, Jose M. Puerta (2015) Scalable Learning of k-dependence Bayesian Classifiers under MapReduce. Proceedings of the Ninth IEEE International Conference on Big Data Science and Engineering, (TrustCom/BigDataSE/ISPA), Volume 2, 25-32. DOI: 10.1109/Trustcom.2015.558
- Jacinto Arias, Jose A. Gámez, Jose M. Puerta (2017): Learning distributed discrete Bayesian Network Classifiers under MapReduce with Apache Spark. Knowledge-Based Systems 117, 16-26. DOI: 10.1016/j.knosys.2016.06.013.



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

- Jacinto Arias, Jose A. Gámez, Thomas D. Nielsen, Jose M. Puerta (2016): A scalable pairwise class interaction framework for multidimensional classification. International Journal of Approximate Reasoning 68, 194-210. DOI: 10.1016/j.ijar.2015.07.007
- M. Julia Flores, Jose A. Gámez (2015): Impact on Bayesian Networks Classifiers When Learning from Imbalanced Datasets. Proceedings of the International Conference on Agents and Artificial Intelligence (ICAART), Volume 2, 382-389.
- Juan A. Aledo, Jose A. Gámez, David Molina, Alejandro Rosete (2016): FSS-OBOP: Feature subset selection guided by a bucket order consensus ranking. Proceedings of IEEE Symposium Series on Computational Intelligence (SSCI-CIDM 2016), 1-8, IEEE Press. DOI: 10.1109/SSCI.2016.7849924
- Juan A. Aledo, Jose A. Gámez, David Molina (2017) Tackling the supervised label ranking problem by bagging weak learners. Information Fusion 35, 38-50. DOI: 10.1016/j.inffus.2016.09.002
- Pablo Bermejo, Jose A. Gámez and Jose M. Puerta. Adapting the CMIM algorithm for multi-label feature selection. A comparison with existing methods. Summited to Expert Systems (2 review)
- Estudio Ensembles AODE y AnDE. Weak Classifier to Aesthetic Image Classification.

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 aesthetic 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 aesthetic classification.

Results

- Antonio Jesús Díaz-Honrubia, Johan De Praeter, Glenn Van Wallendael, J. L. Martinez, Pedro Cuenca, J. M. Puerta, J. A. Gámez: CTU splitting algorithm for H.264/AVC and HEVC simultaneous encoding. The Journal of Supercomputing 73(1): 190-202 (2017)
- 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. (2016)
- F. Rubio, J. Martínez-Gómez, M. J. Flores, J. M. Puerta: Comparison between Bayesian network classifiers and SVMs for semantic localization. Expert Syst. Appl. 64: 434-443 (2016)
- Jacinto Arias, Jesus Martínez-Gómez, Jose A. Gámez, Alba Garcia Seco de Herrera, Henning Muller (2016): Medical image modality classification using discrete Bayesian networks. Computer Vision and Image Understanding 151, 61-71. DOI: 10.1016/j.cviu.2016.04.002

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

- Sara Sáez-Atienzar, Jesus Martínez-Gómez, Juan I. Alonso-Barba, Jose M. Puerta, María F. Galindo, Joaquín Jordán, Luis de la Ossa (2015): Automatic quantification of the subcellular localization of chimeric GFP protein supported by a two-level Naive Bayes classifier. Expert Systems with Applications 42(3), 1531-1537. DOI: 10.1016/j.eswa.2014.09.052
- Pablo Bermejo, Jose A. Gámez, Jose M. Puerta, Marco A. Esquivias, Pedro J. Tárraga (2016): Construction of a Semi-Naive Model to Predict Early Readmission of COPD Patients by Using Quality Care Information. Proceedings of ICDM Workshop on Data Mining in Biomedical Informatics and Healthcare, 233-240, IEEE Press. DOI: 10.1109/ICDMW.2016.0040
- Francisco J. Pérez-Gil; Pablo Bermejo, Alicia Vivo, Pedro J. Tárraga. ProstaWeb: una herramienta online para la predicción de patologías prostáticas desde Atención Primaria. Journal of Negative and No Positive Results. 2(4), p. 144-151, 2017. DOI: 10.19230/jonnpr.1327