Learning Tractable Multidimensional Bayesian Network Classifiers

Marco Benjumeda Concha Bielza
Pedro Larrañaga

Computational Intelligence Group
Universidad Politécnica de Madrid

Agenda

1. Background
2. Tractable MBCs
3. Experimental Results
4. Conclusions and Future Research

Background

Multidimensional Bayesian Network Classifiers

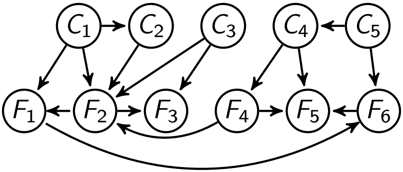


Figure 1: MBC structure

- **Multidimensional Bayesian network classifiers** (MBCs) (van der Gaag and de Waal, 2006) extend Bayesian network classifiers to the problem of multidimensional classification
- **Good performance** in multiple domains
- **Expressive** graphical representation

Multidimensional Bayesian Network Classifiers

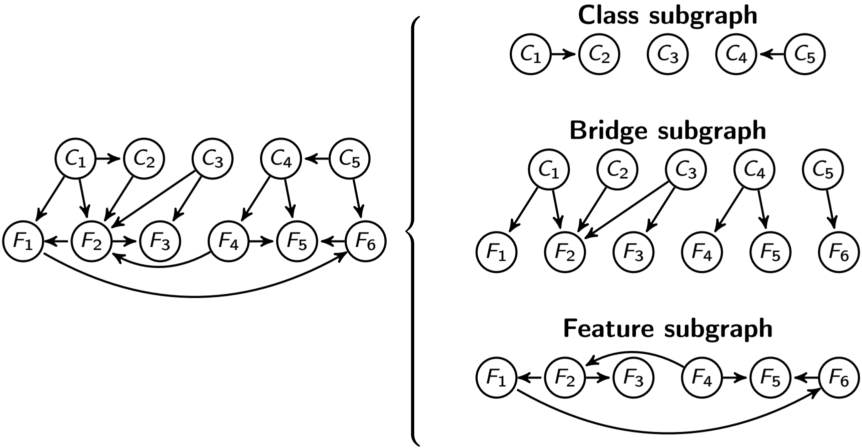


Figure 2: MBC subgraphs

Inference in MBCs I

- Performing multidimensional classification in an MBC is equivalent to **obtaining the MPE**

$$\max_{C_1,\ldots,C_5} \Pr(C_1,\ldots,C_5 | F_1,\ldots,F_6)$$

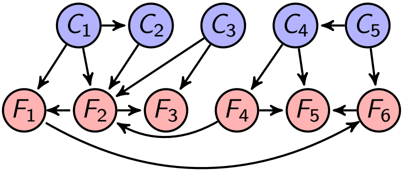


Figure 3: Multidimensional classification with an MBC

Inference in MBCs II

- If there are **unobserved feature variables**, multidimensional classification in MBCs is equivalent to **obtaining the MAP**. This can be intractable even if the treewidth is bounded (Park, 2002)

$$\max_{C_1,\dots,C_5} \Pr(C_1,\dots,C_5 | F_3,\dots,F_6)$$

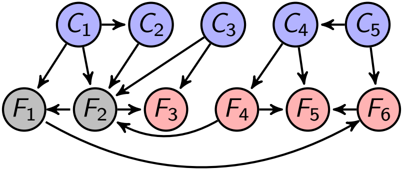


Figure 4: Multidimensional classification with an MBC

CB-Decomposable MBCs

The CB-decomposability of MBCs can **reduce the number of computations** required to perform multidimensional classification (Bielza et al., 2011)

$$\max_{\mathbf{c}\in\Omega\mathcal{C}} \Pr(\mathbf{c}|\mathbf{f}) \propto \prod_{i=1}^{r} \max_{\mathbf{c}_i\in\Omega\mathcal{C}_i} \prod_{C\in\mathcal{C}_i} \Pr(c|\mathbf{Pa}_{\mathcal{G}}(C)) \prod_{F\in\mathbf{Ch}_{\mathcal{G}}(\mathcal{C}_i)} \Pr(f|\mathbf{Pa}_{\mathcal{G}_B}(F), \mathbf{Pa}_{\mathcal{G}_F}(F))$$

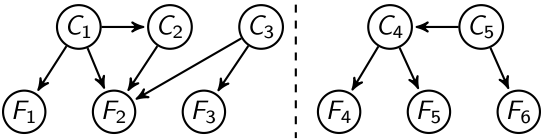


Figure 5: Connected components of an MBC

Inference Complexity in MBCs

- **MPE is generally NP-hard** (Kwisthout, 2011), but it can be computed in **polynomial time** in any BN if its **treewidth is bounded** (Sy, 1992)
- De Waal and van der Gaag (2007) demonstrated that treewidth($\mathcal{G}$) $\leq$ treewidth($\mathcal{G}_F$) $+ d$

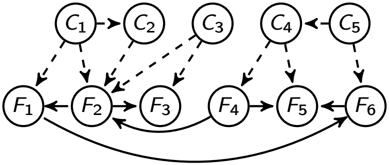


Figure 6: Class variables and feature subgraph of an MBC

Inference Complexity in MBCs

- Pastink and van der Gaag (2015) also suggested that the **complexity of an MBC** with an empty feature subgraph is given by the treewidth of the graph obtained by:
    1. **Moralizing** its structure
    2. **Removing all its feature** nodes from the moralized graph

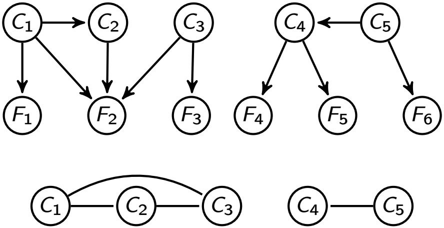


Figure 7: MBC structure (above) and moralized graph without feature nodes (below)

Learning MBCs

- Unconstrained complexity:
  - Tree–tree (van der Gaag and de Waal, 2006)
  - Polytree–polytree (de Waal and van der Gaag, 2007)
  - DAG–DAG (Bielza et al., 2011)

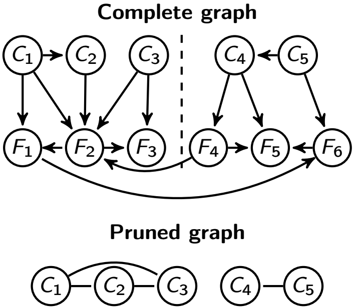
- Penalized complexity:
  - Polytree–empty (Corani et al., 2014)
  - Tree–empty (Pastink and van der Gaag, 2015)
  - CB–decomposable MBC (DAG–DAG) (Borchani et al., 2010)

Tractable MBCs

Motivation

- Previous research uses the **treewidth of the complete graph** to bound the complexity of multidimensional classification in MBCs
- We consider **query-dependent information**
- We provide a **new upper bound** for the complexity of multidimensional classification in MBCs
- Learn **DAG–DAG** MBCs

Tractable MBCs



	Before	After
ϕ_{C_1}	$\{C_1\}$	$\{C_1\}$
ϕ_{C_2}	$\{C_1, C_2\}$	$\{C_1, C_2\}$
ϕ_{C_3}	$\{C_3\}$	$\{C_3\}$
ϕ_{C_4}	$\{C_4, C_5\}$	$\{C_4, C_5\}$
ϕ_{C_5}	$\{C_5\}$	$\{C_5\}$
ϕ_{F_1}	$\{C_1, F_1, F_2\}$	$\{C_1\}$
ϕ_{F_2}	$\{C_1, C_2, C_3, F_2, F_4\}$	$\{C_1, C_2, C_3\}$
ϕ_{F_3}	$\{C_3, F_2, F_3\}$	$\{C_3\}$
ϕ_{F_4}	$\{C_4, F_4\}$	$\{C_4\}$
ϕ_{F_5}	$\{C_4, F_4, F_5, F_6\}$	$\{C_4\}$
ϕ_{F_6}	$\{C_5, F_1, F_6\}$	$\{C_5\}$

Figure 8: MBC structure and pruned graph (left), and domain of the potential of each node before and after they are updated with evidence $\mathbf{f} = (f_1, \ldots, f_6)$ (right)

Tractable MBCs

	This work	Pastink and van der Gaag (2015)
Feature subgraph	DAG	Empty
Bound	Pruned graph	Complete graph
Structure family	DAG–DAG	Tree–empty

Table 1: Differences between Pastink and van der Gaag (2015) and this work

Tractable MBCs

Theorem 1

Let $\mathcal{G} = (\mathcal{C} \cup \mathcal{F}, \mathcal{A}_C \cup \mathcal{A}_B \cup \mathcal{A}_F)$ be the structure of an MBC $\mathcal{B}$. If the ***treewidth of its pruned graph*** *$\mathcal{G}'$ and the number of parents of each node that belongs to $\mathcal{F}$ are* ***bounded***, *$\mathcal{B}$ can perform classification in* ***polynomial time***.

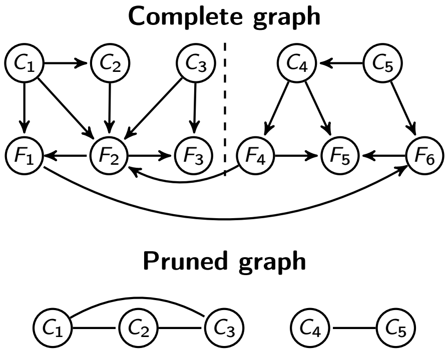


Figure 9: MBC structure and pruned graph

Tractable MBCs

Corollary 2

Let $\mathcal{G} = (\mathcal{C} \cup \mathcal{F}, \mathcal{A}_C \cup \mathcal{A}_B \cup \mathcal{A}_F)$ be the structure of an MBC $\mathcal{B}$. If the ***number of class variables*** *d and the number of parents of each node in $\mathcal{F}$ are* ***bounded****, $\mathcal{B}$ can perform classification in* ***polynomial time****.*

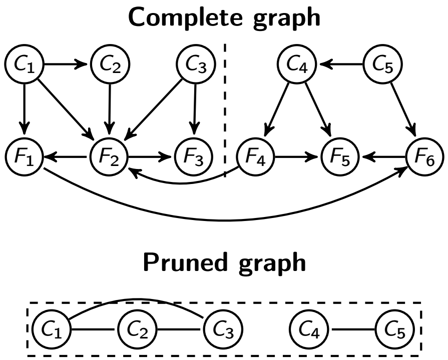


Figure 10: MBC structure and pruned graph

Tractable MBCs

Corollary 3

Let $\mathcal{G} = (\mathcal{C} \cup \mathcal{F}, \mathcal{A}_C \cup \mathcal{A}_B \cup \mathcal{A}_B)$ be the structure of a ***CB-decomposable*** *MBC $\mathcal{B}$. If the* ***number of class variables in each component*** *of $\mathcal{G}$ and the number of parents of each node in $\mathcal{F}$ are bounded, $\mathcal{B}$ can perform classification in* ***polynomial time****.*

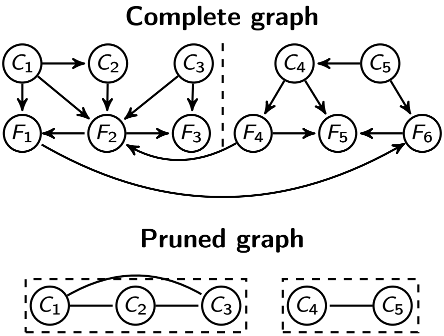


Figure 11: MBC structure and pruned graph

Learning Method

- We adapt **order-based search** (OBS) (Bouckaert, 1992) to learn tractable MBCs
- We learn CB-decomposable MBCs, **bounding** the maximum number of **class variables per component**
- Rejecting candidates that exceed the bound of class variables per component bounds the complexity of multidimensional classification
- To search in the space of orderings we use the procedure described by Teyssier and Koller (2005)

Experimental Results

Experimental Framework

- Learning Methods:
  - **CB-OBS**
  - Tree–tree (van der Gaag and de Waal, 2006)
  - Polytree–polytree (de Waal and van der Gaag, 2007)
  - Filter (DAG–DAG) (Bielza et al., 2011)
  - Small–tw (tree–empty) (Pastink and van der Gaag, 2015)
- Performance measures:
  - Treewidth (τ)
  - Treewidth of the pruned graph (τ_p)
  - Size
  - Mean accuracy (acc_M)

Datasets

Dataset	Classes	Features	Instances
ANDES	74	82	5000
MUNIN1	62	86	5000
DIABETES	138	284	5000
MEDICAL	45	110	978
ENRON	53	124	1702

Table 2: Basic properties of the datasets

Experimental Results

Method	τ	τ_p	size	acc_M
CB–OBS	7.8±0.7	5.2±1.0	679±157	0.778±0.001*
Tree–tree	18.4±0.8	7.8±0.4	3710±724	**0.779±0.002***
Polytree–polytree	21.8±2.4	8.2±0.4	5221±342	0.777±0.003
Filter (DAG–DAG)	25.2±1.2	9.2±1.2	8756±3542	0.776±0.003
Small–tw	5.0±0.0	5.0±0.0	1382±171	0.764±0.004

Method	τ	τ_p	size	acc_M
CB–OBS	5.8±0.4	3.6±0.5	391±27	0.757±0.001
Tree–tree	14.2±1.9	7.6±1.2	2604±1226	**0.758±0.001***
Polytree–polytree	19.8±2.9	9.0±1.4	6071±4146	0.756±0.001
Filter (DAG–DAG)	21.2±1.7	8.6±1.0	5861±3918	0.756±0.001
Small–tw	5.0±0.0	5.0±0.0	1012±27	0.750±0.002

Method	τ	τ_p	size	acc_M
CB–OBS	72.2±2.9	5.4±0.5	2200±367	**0.934±0.015***
Tree–tree	66.6±10.9	37.4±2.6	$(5.183 \pm 8.999) \times 10^{12}$	–
Polytree–polytree	100.6±8.0	54.4±5.7	$(3.216 \pm 3.950) \times 10^{18}$	–
Filter (DAG–DAG)	93.0±3.2	55.8±2.5	$(1.154 \pm 1.551) \times 10^{18}$	–
Small–tw	5.0±0.0	5.0±0.0	2802±179	0.931±0.016

Table 3: Performance in ANDES (above), MUNIN1 (center), and DIABETES (below)

Experimental Results

Method	τ	τ_p	size	acc$_M$
CB–OBS	6.0±0.0	3.2±0.7	179±23	**0.988±0.002**∗
Tree–tree	13.0±1.4	5.8±1.9	802±735	0.987±0.002
Polytree–polytree	20.6±1.0	14.2±1.7	173039±182991	–
Filter (DAG–DAG)	10.0±1.3	4.0±0.9	286±52	0.987±0.002
Small–tw	3.0±0.0	2.8±0.4	250±7	0.986±0.002

Method	τ	τ_p	size	acc$_M$
CB–OBS	31.4±1.9	3.8±0.4	273±35	0.946±0.008
Tree–tree	12.0±1.1	5.4±0.5	743±85	**0.946±0.007**∗
Polytree–polytree	15.0±1.7	9.6±0.8	5025±2572	0.946±0.007
Filter (DAG–DAG)	32.8±1.8	4.6±0.5	494±61	0.945±0.010
Small–tw	5.0±0.0	5.0±0.0	671±38	0.944±0.009

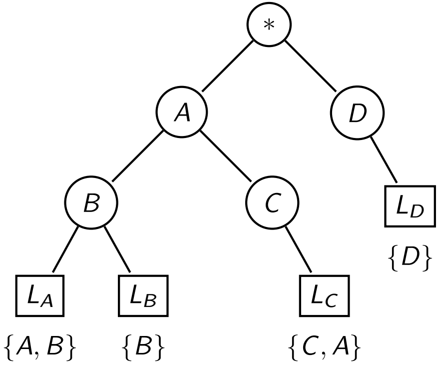
Table 4: Performance in MEDICAL (above) and ENRON (below)

Conclusions and Future Research

Conclusions

- We provided **new upper bounds** for the complexity of MBCs
- We showed that **CB-decomposability** can be used to **efficiently guarantee the tractability** of MBCs
- We proposed a learning method that learns tractable CB-decomposable MBCs
- Experimental results showed that the proposed method is **competitive** with other state-of-the-art methods in terms of **accuracy**, also ensuring that the learned models can perform multidimensional classification efficiently
- We observed that some models remain tractable even with a large treewidth

Future Research: Learning Elimination Orders



Equivalent elimination orders:

{B,C,D,A}
{B,C,A,D}
{C,B,D,A}
...

- Not restricted to **topological orders**
- Each tree represents a **set of elimination orders**
- **Reduces the redundancy** of the space of elimination orders and DAGs

Bibliography I

Concha Bielza, Guangdi Li, and Pedro Larrañaga. Multi-dimensional classification with Bayesian networks. *International Journal of Approximate Reasoning*, 52 (6):705–727, 2011.

Hannen Borchani, Concha Bielza, and Pedro Larrañaga. Learning CB-decomposable multi-dimensional Bayesian network classifiers. *Proceedings of the 5th European Workshop on Probabilistic Graphical Models*, pages 25–32, 2010.

Remco R Bouckaert. Optimizing causal orderings for generating dags from data. In *Proceedings of the 8th International Conference on Uncertainty in Artificial Intelligence*, pages 9–16. Morgan Kaufmann Publishers Inc., 1992.

Giorgio Corani, Alessandro Antonucci, Denis D Mauá, and Sandra Gabaglio. Trading off speed and accuracy in multilabel classification. In *Proceedings of the 7th European Workshop on Probabilistic Graphical Models*, pages 145–159, 2014.

Bibliography II

Peter R de Waal and Linda C van der Gaag. Inference and learning in multi-dimensional bayesian network classifiers. In *Proceedings of the 9th European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty*, pages 501–511. Springer, 2007.

Johan Kwisthout. Most probable explanations in Bayesian networks: Complexity and tractability. *International Journal of Approximate Reasoning*, 52(9): 1452–1469, 2011.

James D Park. MAP complexity results and approximation methods. In *Proceedings of the 18th Conference on Uncertainty in Artificial Intelligence*, pages 388–396. Morgan Kaufmann Publishers Inc., 2002.

Arnoud Pastink and Linda C van der Gaag. Multi-classifiers of small treewidth. In *Proceedings of the 13th European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty*, pages 199–209. 2015.

Bon K Sy. Reasoning MPE to multiply connected belief networks using message passing. In *Proceedings of the 10th National Conference on Artificial intelligence*, pages 570–576. AAAI Press, 1992.

Bibliography III

Marc Teyssier and Daphne Koller. Ordering-based search: A simple and effective algorithm for learning Bayesian networks. In *Proceedings of the 21st Conference on Uncertainty in Artificial Intelligence*, pages 584–590. AUAI Press, 2005.

Linda C van der Gaag and Peter R de Waal. Multi-dimensional Bayesian network classifiers. In *Proceedings of the 3rd European Workshop on Probabilistic Graphical Models*, pages 107–114, 2006.