Possibilistic MDL: a New Possibilistic Likelihood Based Score Function for Imprecise Data

Maroua Haddad[1,2], Philippe Leray[2] and Nahla Ben Amor[1]

[1] LARODEC Laboratory ISG-Tunis, University of Tunis, Tunisia.
[2]LS2N CNRS 6004, University of Nantes, France.

Outline

Probabilistic likelihood

Recall on probabilistic likelihood function

$$\mathcal{D} = \begin{pmatrix} X_1 \\ x_{11} \\ x_{12} \end{pmatrix} \qquad L(\Theta, \mathcal{D}) = \theta_{x_{11}} * \theta_{x_{12}}$$

⇒ Infer different types of models

Probabilistic likelihood

Recall on probabilistic likelihood function

$$\mathcal{D} = \begin{pmatrix} X_1 \\ x_{11} \\ x_{12} \end{pmatrix} \qquad L(\Theta, \mathcal{D}) = \theta_{x_{11}} * \theta_{x_{12}}$$

⇒ Infer different types of models

Limitations of probabilistic likelihood

- Imprecise data ?
- Total ignorance : Probabilistic reasoning unsound
- Evidential adaptation of likelihood function ⟶ limited [Couso and Dubois, 2017]
  ⇒ Possibilistic likelihood

Possibility theory

- Introduced by Zadeh [Zadeh, 1978] and developed by Dubois and Prade [Dubois and Prade, 1988]
- Possibility distribution : $\pi : \Omega \rightarrow L = [0, 1]$
- Extreme Cases :
  - Complete Knowledge : $\exists \omega_0 \in \Omega$ s.t. $\pi(\omega_0) = 1$ and $\forall \omega \neq \omega_0, \pi(\omega) = 0$
  - Total Ignorance : $\forall \omega \in \Omega, \pi(\omega) = 1$.
- Normalization : $\exists \omega \in \Omega$ s.t. $\pi(\omega) = 1$

Possibility theory

- Introduced by Zadeh [Zadeh, 1978] and developed by Dubois and Prade [Dubois and Prade, 1988]
- Possibility distribution : $\pi : \Omega \rightarrow L = [0, 1]$
- Extreme Cases :
  - Complete Knowledge : $\exists \omega_0 \in \Omega$ s.t. $\pi(\omega_0) = 1$ and $\forall \omega \neq \omega_0, \pi(\omega) = 0$
  - Total Ignorance : $\forall \omega \in \Omega, \pi(\omega) = 1$.
- Normalization : $\exists \omega \in \Omega$ s.t. $\pi(\omega) = 1$

- Possibility measure Π : to what extent A is consistent with π

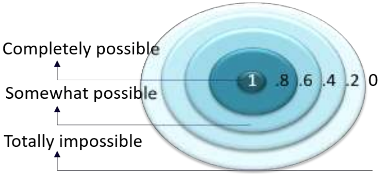
$$\Pi(A) = \max_{\omega \in A} \pi(\omega)$$

- Necessity measure N : to what extent A is implied by π

$$N(A) = 1 - \Pi(\neg A)$$

Possibility distribution

Numerical



Ordinal

	a_1	a_2	a_3
π_1	1	0.7	0.6
π_2	1	0.6	0.1

$\pi_1(a_1) < \pi_1(a_2) < \pi_1(a_3)$

$\pi_2(a_1) < \pi_2(a_2) < \pi_2(a_3)$

A= a_1 is more plausible than A= a_2

A= a_2 is more plausible than A= a_3

Possibilistic likelihood (1/2)

Random set likelihood function (Imprecise data)

- A random set $S =< A_{ik} \subseteq D_i, m(A_{ik}) >$ [Goodman and Nguyen, 1991]
- $m : 2^{card(D_i)} \mapsto [0, 1]$

$$\mathcal{D} = \begin{pmatrix} X_1 \\ x_{11}, x_{12} \\ x_{12} \end{pmatrix} \qquad mL(m, \mathcal{D}) = m_{x_{11}, x_{12}} * m_{x_{12}}$$

$\Rightarrow$ High complexity $\Rightarrow$ Approximation ?

- π is a contour function of a random set [Shafer, 1976] :

$$\mathrm{CF}_{m \to \pi}(x_{ik}) = \pi(x_{ik}) = \sum_{A_{ik} | x_{ik} \in A_{ik}} m(A_{ik})$$

Possibilistic likelihood (1/2)

Random set likelihood function (Imprecise data)

- A random set $S =< A_{ik} \subseteq D_i, m(A_{ik}) >$ [Goodman and Nguyen, 1991]
- $m : 2^{card(D_i)} \mapsto [0,1]$

$$\mathcal{D} = \begin{pmatrix} X_1 \\ x_{11}, x_{12} \\ x_{12} \end{pmatrix} \qquad mL(m,\mathcal{D}) = m_{x_{11},x_{12}} * m_{x_{12}}$$

$\Rightarrow$ High complexity $\Rightarrow$ Approximation ?

- π is a contour function of a random set [Shafer, 1976] :

$$\mathrm{CF}_{m\to\pi}(x_{ik}) = \pi(x_{ik}) = \sum_{A_{ik} | x_{ik} \in A_{ik}} m(A_{ik})$$

Possibilistic likelihood function (Imprecise data)

$$\mathcal{D} = \begin{pmatrix} X_1 \\ x_{11}, x_{12} \\ x_{12} \end{pmatrix} \qquad \pi L(\pi,\mathcal{D}) = \pi_{x_{11}} * \pi_{x_{12}} * \pi_{x_{12}}$$

Possibilistic likelihood (2/2)

Maximizing random sets likelihood

$$\hat{m}_{ik} = argmax(mLL(m_{ik}, \mathcal{D})) = \frac{N_{A_{ik}}}{N}$$

Possibilistic likelihood (2/2)

Maximizing random sets likelihood

$$\hat{m}_{ik} = argmax(mLL(m_{ik}, \mathcal{D})) = \frac{N_{A_{ik}}}{N}$$

Maximizing possibilistic likelihood

- Under constraint : $\sum_{k=1}^{|D_i|} \pi_{ik} = S_i$: imprecision degree of X_i

$$\hat{\pi}_{ik} = argmax(\pi LL(\pi_{ik}, \mathcal{D})) = \frac{N_{ik}}{N} * S_i$$

Possibilistic likelihood (2/2)

Maximizing random sets likelihood

$$\hat{m}_{ik} = argmax(mLL(m_{ik}, \mathcal{D})) = \frac{N_{A_{ik}}}{N}$$

Maximizing possibilistic likelihood

- Under constraint : $\sum_{k=1}^{|D_i|} \pi_{ik} = S_i$: imprecision degree of X_i

$$\hat{\pi}_{ik} = argmax(\pi LL(\pi_{ik}, \mathcal{D})) = \frac{N_{ik}}{N} * S_i$$

$$argmax(mLL(m_{ik}, \mathcal{D})) = \hat{m}_{ik} \quad \xrightarrow{CF_{m \to \pi}} \quad \pi^{*}_{ik}$$

$$argmax(\pi LL(\pi_{ik}, \mathcal{D})) = \hat{\pi}_{ik}$$

Possibilistic likelihood (2/2)

Maximizing random sets likelihood

$$\hat{m}_{ik} = argmax(mLL(m_{ik}, \mathcal{D})) = \frac{N_{A_{ik}}}{N}$$

Maximizing possibilistic likelihood

- Under constraint : $\sum_{k=1}^{|D_i|} \pi_{ik} = S_i$: imprecision degree of X_i

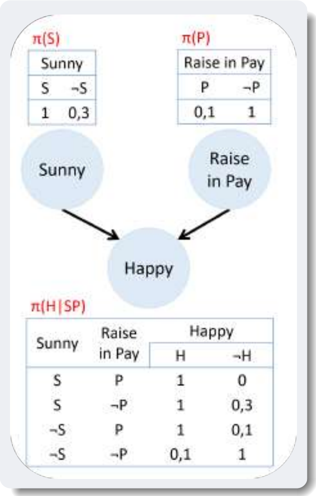
$$\hat{\pi}_{ik} = argmax(\pi LL(\pi_{ik}, \mathcal{D})) = \frac{N_{ik}}{N} * S_i$$

$$argmax(mLL(m_{ik}, \mathcal{D})) = \hat{m}_{ik} \quad \xrightarrow{CF_{m \to \pi}} \quad \pi^*_{ik}$$

$$argmax(\pi LL(\pi_{ik}, \mathcal{D})) = \hat{\pi}_{ik} = \pi^*_{ik}$$

⇒ Infer different types of possibilistic models from *imprecise* data : Case of possibilistic networks

Possibilistic networks [Fonck, 1992]



Possibilistic conditioning

- Product-based ΠN_* :
  - Product-based conditioning

$$\pi(\omega|_* A) = \begin{cases} \frac{\pi(\omega)}{\Pi(A)} & \text{if } \omega \in A \\ 0 & \textit{otherwise.} \end{cases}$$

- Min-based ΠN_{min}
  - Min-based conditioning

$$\pi(\omega|_{min} A) = \begin{cases} 1 & \textit{if } \pi(\omega) = \Pi(A) \text{ and } \omega \in A \\ \pi(\omega) & \textit{if } \pi(\omega) < \Pi(A) \text{ and } \omega \in A \\ 0 & \textit{otherwise.} \end{cases}$$

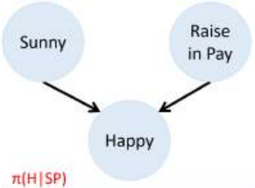
Possibilistic networks [Fonck, 1992]

π(S)

Sunny	
S	¬S
1	0,3

π(P)

Raise in Pay	
P	¬P
0,1	1



π(H|SP)

Sunny	Raise in Pay	Happy	
		H	¬H
S	P	1	0
S	¬P	1	0,3
¬S	P	1	0,1
¬S	¬P	0,1	1

Possibilistic chain rule

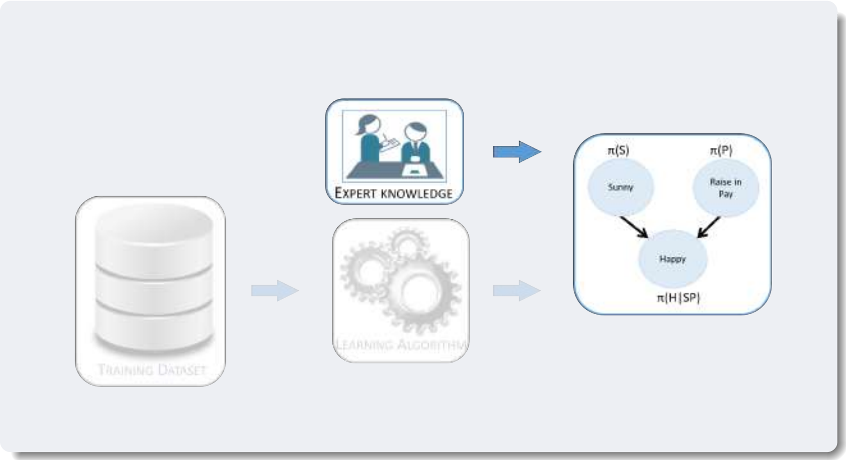
$$\pi(X_1, .., X_n) = \otimes_{i=1..n} \pi(X_i | Pa(X_i))$$

ΠN_* : $\otimes = *$: / ΠN_{min} : $\otimes = min$:

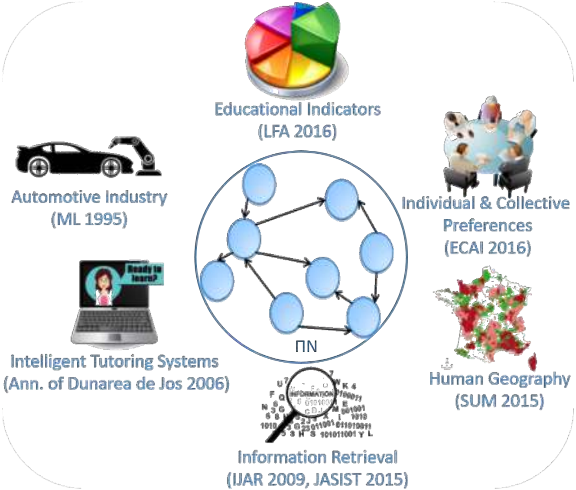
Joint possibility distribution

Sunny	Raise in Pay	Happy	*	min
S	P	H	0,1	0,1
S	P	¬H	0	0
S	¬P	H	1	1
S	¬P	¬H	0,3	0,3
¬S	P	H	0,03	0,1
¬S	P	¬H	0,03	0,1
¬S	¬P	H	0,03	0,1
¬S	¬P	¬H	0,3	0,3

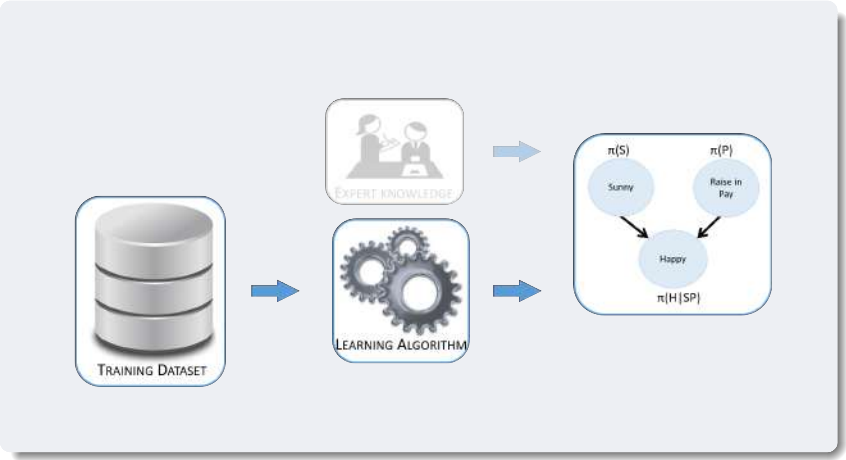
How to build a possibilistic network ?



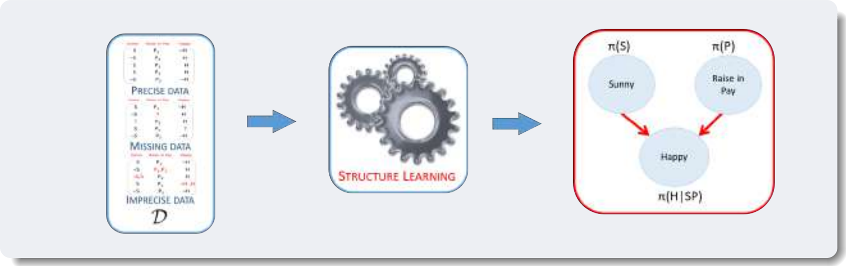
Applications



How to build a possibilistic network ?



Structure learning of ΠN



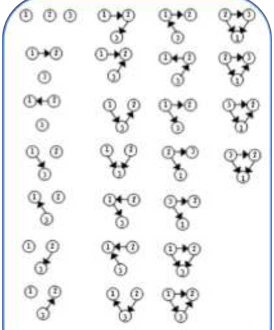
Structure learning of BN

Score-based approach

- Search space (DAGs)

$$NS(n) = \begin{cases} 1 \text{ if } n = 0 \text{ or } n = 1 \\ \sum_{i=1}^{n} (-1)^{i+1} C_i^n 2^i (n-1) NS(n-i) \text{ if } n > 1 \end{cases}$$

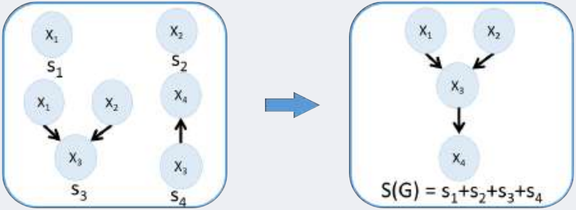
- Exhaustive search is impossible
- Heuristics to traverse DAGs space :
    - Reducing search space : Search sub-networks with high scores and combine them
    - Performing greedy search : Search in networks space and pick the one with the highest score



25 possibe DAGs with 3 nodes

Score properties

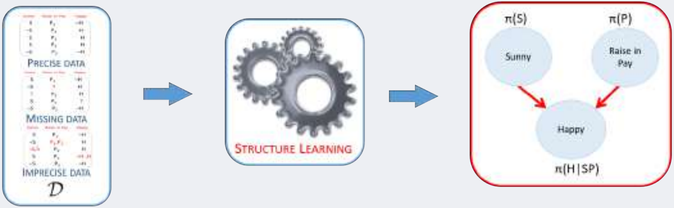
Decomposability



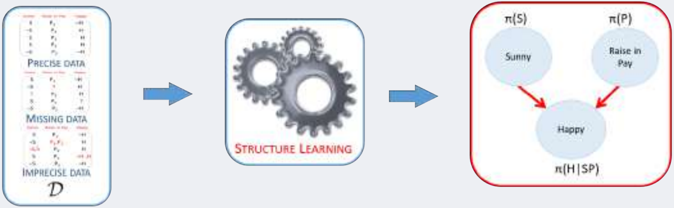
Likelihood equivalence

- Two equivalent structures have the same score

Structure learning of ΠN



Structure learning of ΠN



- Score-based approaches : πMWST, πK2 [Borgelt and Kruse, 2003]
- Hybrid method : [Sangüesa et al., 1998]

$\Rightarrow$ Not based on likelihood function

Possibilistic score

Possibilistic MDL

- Minimum description length (MDL) principle [Rissanen, 1978]
- Compromise between likelihood and complexity

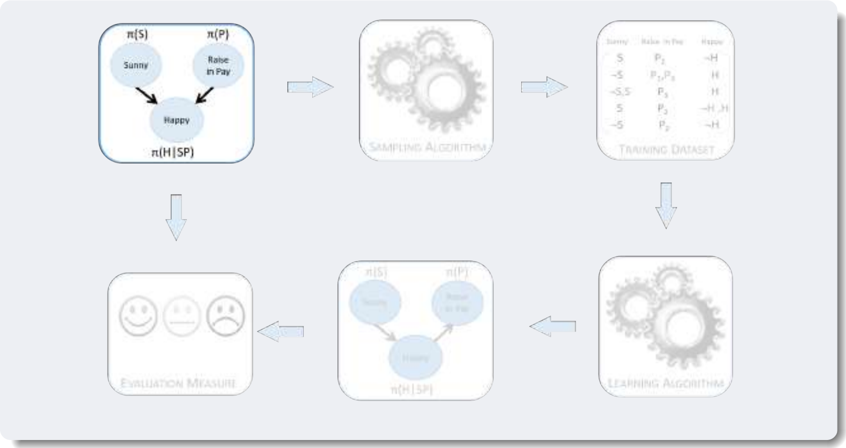
$$\pi MDL(G|\mathcal{D}) = \pi LL(\pi, G, \mathcal{D}) - dim(G)$$

$$\pi MDL(G|\mathcal{D}) = \sum_{i=1}^{n}\sum_{j=1}^{q_i}\sum_{k=1}^{r_i} N_{ijk} \log \hat{\pi}(X = x_{ik}|Pa(X_i) = x_j) - \sum_{i=1}^{n} |D_i| * \prod_{X_j \in Pa(X_i)} |D_j|$$

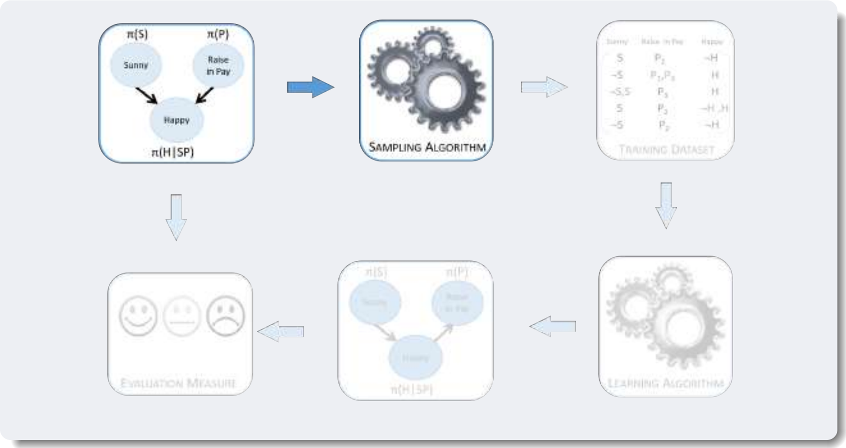
Score properties :

- Decomposability ✓
- Likelihood equivalence ?

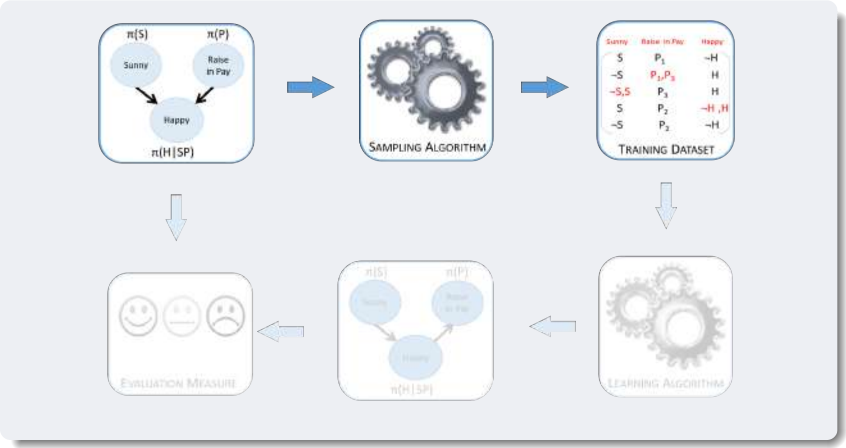
Evaluation strategy [Haddad et al., 2015]



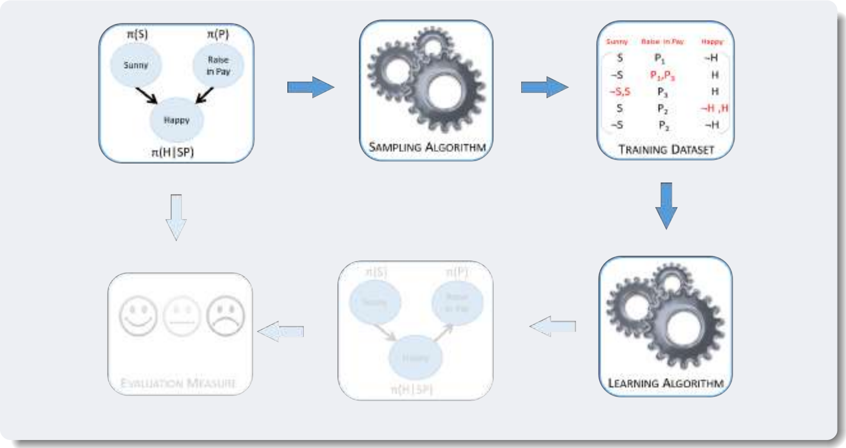
Evaluation strategy [Haddad et al., 2015]



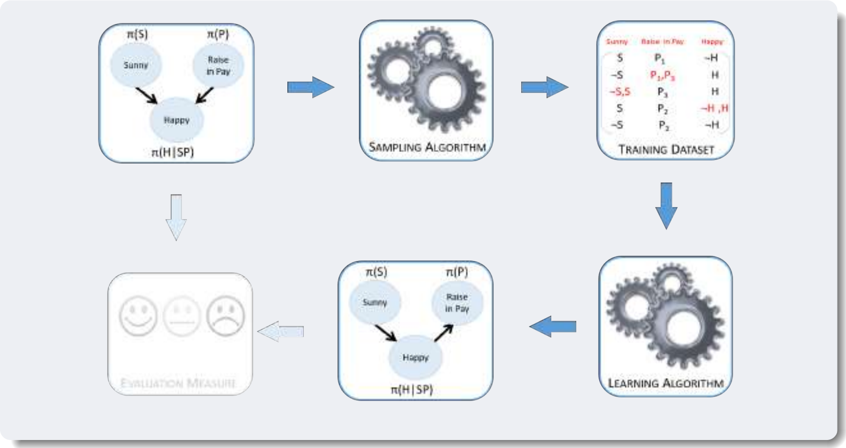
Evaluation strategy [Haddad et al., 2015]



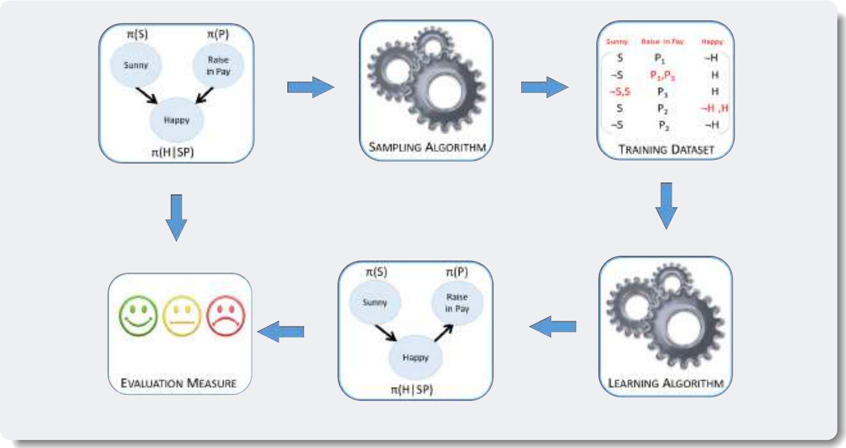
Evaluation strategy [Haddad et al., 2015]



Evaluation strategy [Haddad et al., 2015]



Evaluation strategy [Haddad et al., 2015]



Structure learning algorithm evaluation (1/2)

Experimental protocol

1. Generate 20 random ΠN_0 possibilistic networks ({10, 20} variables)
2. Sample $\Pi N_0 \Rightarrow$ data sets of 1000 observations using Consonant_sampling, Imp_control_sampling, Cons_control_sampling algorithms [Haddad et al., 2015]
3. Learn possibilistic networks ΠN_l using greedy search πGS combined with πMDL and networks structures using existing methods $\pi K_2, \pi MWST$ combined with d_{χ^2} and d_{mi} [Borgelt and Kruse, 2003]
4. Compute editing distance between ΠN_0 and ΠN_l : number of operations required to transform ΠN_0 DAG into ΠN_l DAG

Structure learning algorithm evaluation (2/2)

Method \ n	Editing distance 10	Editing distance 20
πGS + πMDL	**19.77 +/- 1.5**	**31.55 +/- 2.92**
πGS + $\sum d_{\chi^2}$	28.83 +/- 2.32	51.66 +/- 1.33
πGS + $\sum d_{mi}$	35.66 +/- 2.06	49.55 +/- 1.41-
πMWST + d_{χ^2}	23.44 +/- 1.63	47.33 +/- 0.88
πMWST + d_{mi}	22.77 +/- 1.6	47.55 +/- 1.41
π K2 + d_{χ^2}	27.44 +/- 2.95	42.22 +/- 6.87
πK2 + d_{mi}	28.38 +/- 4.53	42.77 +/- 5.66

- πMDL outperforms d_{χ^2} and d_{mi} when used by GS

Conclusion & perspectives

Conclusion

- Two likelihood functions : random set likelihood function and possibilistic likelihood function
- Infer different types of random set/possibilistic models : Case of possibilistic networks
  $\Rightarrow$ Learn possibilistic network structure from imprecise data : experimentally validated

Perspectives

- A comparative study on a large number of benchmarks and problems
- Use numerical evaluation measures e.g. distance measure between joint and local distributions
- Evaluate the impact of non-satisfaction of Markov likelihood property on the learned possibilistic network structure quality

References I

Borgelt, C. and Kruse, R. (2003).
Operations and evaluation measures for learning possibilistic graphical models.
Artificial Intelligence, 148(1) :385–418.

Couso, I. and Dubois, D. (2017).
Maximum likelihood under incomplete information : Toward a comparison of criteria.
In *Soft Methods for Data Science*, pages 141–148. Springer.

Dubois, D. and Prade, H. (1988).
Possibility theory.
Springer.

Fonck, P. (1992).
Propagating uncertainty in a directed acyclic graph.
In *Proceedings of the fourth Information Processing and Management of Uncertainty Conference*, volume 92, pages 17–20.

Goodman, I. R. and Nguyen, H. T. (1991).
Uncertainty models for knowledge-based systems.
Technical report, DTIC Document.

Haddad, M., Leray, P., and Amor, N. B. (2015).
Evaluating product-based possibilistic networks learning algorithms.
In *Proceedings of Symbolic and Quantitative Approaches to Reasoning with Uncertainty*, pages 312–321.

References II

Rissanen, J. (1978).
Modeling by shortest data description.
Automatica, 14(5) :465–471.

Sangüesa, R., Cabós, J., and Cortes, U. (1998).
Possibilistic conditional independence : A similarity-based measure and its application to causal network learning.
International Journal of Approximate Reasoning, 18(1) :145–167.

Shafer, G. (1976).
A mathematical theory of evidence, volume 1.
Princeton university press Princeton.

Zadeh, L. A. (1978).
Fuzzy sets as a basis for a theory of possibility.
Fuzzy Sets and Systems, 100 :9–34.