Making Set-valued Predictions in Evidential Classification: A Comparison of Different Approaches

Liyao Ma & Thierry Denœux

Introduction

- Classification : label predictions

  $$\Omega = \{\omega_1, \cdots, \omega_n\}$$

- Uncertainty $\rightarrow$ set-valued predictions
- Dempster-Shafer theory

Decision making view of classification

Precise assignments $\mathcal{F} = \{f_{\omega_1}, \cdots, f_{\omega_n}\}$

- Precise assignments + complete preorder : Maximum Expected Utility principle

Decision making view of classification

Precise assignments $\mathcal{F} = \{f_{\omega_1}, \cdots, f_{\omega_n}\}$

- Precise assignments + complete preorder : Maximum Expected Utility principle
- The uncertain case
  - Precise assignments + partial preorder
  - Partial assignments + complete preorder

Decision making view of classification

Precise assignments $\mathcal{F} = \{f_{\omega_1}, \cdots, f_{\omega_n}\}$

- Precise assignments + complete preorder : Maximum Expected Utility principle
- The uncertain case
  - Precise assignments + partial preorder
  - Partial assignments + complete preorder

Partial assignments $\mathcal{F} = \{f_A, A \in 2^\Omega \setminus \{\emptyset\}\}$

Two families of decision strategies

- Precise assignments + partial preorder
  - $\mathcal{F} = \{f_{\omega_1}, \cdots, f_{\omega_n}\}$
  - Interval dominance, maximality, weak dominance...
  - Lack of information $\rightarrow [\underline{\mathbb{E}}_m(f_i), \overline{\mathbb{E}}_m(f_i)]$
  - Set of non-dominated acts $\mathcal{F}^* = \{f_{\omega_1}, f_{\omega_2}\}$
- Partial assignments + complete preorder
  - $\mathcal{F} = \{f_A, A \in 2^\Omega \setminus \{\emptyset\}\}$
  - Generalized maximin, maximax, Hurwicz, minimax regret...
  - The optimal act $\mathcal{F}^* = \{f_{\{\omega_1, \omega_2\}}\}$

Defining the utility of set-valued predictions

acts	states of nature		
	ω_1	ω_2	ω_3
$f_{\{\omega_1\}}$	1.0000	0.2000	0.1000
$f_{\{\omega_2\}}$	0.2000	1.0000	0.2000
$f_{\{\omega_3\}}$	0.1000	0.2000	1.0000

Defining the utility of set-valued predictions

acts	states of nature		
	ω_1	ω_2	ω_3
$f_{\{\omega_1\}}$	1.0000	0.2000	0.1000
$f_{\{\omega_2\}}$	0.2000	1.0000	0.2000
$f_{\{\omega_3\}}$	0.1000	0.2000	1.0000
$f_{\{\omega_1,\omega_2\}}$	?	?	?
$f_{\{\omega_1,\omega_3\}}$	?	?	?
$f_{\{\omega_2,\omega_3\}}$	?	?	?
$f_{\{\omega_1,\omega_2,\omega_3\}}$	?	?	?

Defining the utility of set-valued predictions

- Ordered Weighted Average (OWA) operator

$$\hat{u}_{A,j} = F\left(\{u_{ij} \mid \omega_i \in A\}\right) = \sum_{k=1}^{|A|} w_k u_{(k)j}^{A}$$

  - Tolerance degree of imprecision

$$TOL(\boldsymbol{w}) = \sum_{k=1}^{|A|} \frac{|A|-k}{|A|-1} w_k$$

  - weights calculation

$$\max_{\boldsymbol{w}} ENT(\boldsymbol{w}) := -\sum_{k=1}^{|A|} w_k \log w_k$$

$$\text{s.t. } TOL(\boldsymbol{w}) = \gamma$$

$$\sum_{k=1}^{|A|} w_k = 1$$

Defining the utility of set-valued predictions

acts	states of nature		
	ω_1	ω_2	ω_3
$f_{\{\omega_1\}}$	1.0000	0.2000	0.1000
$f_{\{\omega_2\}}$	0.2000	1.0000	0.2000
$f_{\{\omega_3\}}$	0.1000	0.2000	1.0000
$f_{\{\omega_1,\omega_2\}}$	0.8400	0.8400	0.1800
$f_{\{\omega_1,\omega_3\}}$	0.8200	0.2000	0.8200
$f_{\{\omega_2,\omega_3\}}$	0.1800	0.8400	0.8400
$f_{\{\omega_1,\omega_2,\omega_3\}}$	0.7373	0.7455	0.7373

Experimental Comparisons

- UCI and artificial Gaussian data sets
- Classification performances with varying γ
- Performances with noised test sets
- Performances with increasing training set size

Conclusions

- Two approaches are contrasted
  - partial preorder among precise assignments
  - complete preorder among partial assignments
- the utility of set-valued prediction : OWA
- experimental comparisons
  - set-valued predictions perform better
  - cautious rules preferred

Thank you !

Making Set-valued Predictions in Evidential Classification: A Comparison of Different Approaches

Liyao Ma, Thierry Denœux

Two families of set-valued decision strategies

Partial preorders among precise assignments

Patterns are assigned to one and only one of the n classes: $\mathcal{F} = \{f_1, \cdots, f_n\}$

$$\underline{\mathbb{E}}_m(f_i) = \sum_{B \subseteq \Omega} m(B) \min_{\omega_j \in B} u_{ij}$$

$$\overline{\mathbb{E}}_m(f_i) = \sum_{B \subseteq \Omega} m(B) \max_{\omega_j \in B} u_{ij}$$

decision criterion	preference relation
interval dominance	$f_i \succcurlyeq_{ID} f_j \iff \underline{\mathbb{E}}_m(f_i) \geq \overline{\mathbb{E}}_m(f_j)$
maximality	$f_i \succcurlyeq_{max} f_j \iff \underline{\mathbb{E}}_m(f_i - f_j) \geq 0$
weak dominance	$f_i \succcurlyeq_{WD} f_j \iff \left(\underline{\mathbb{E}}_m(f_i) \geq \underline{\mathbb{E}}_m(f_j)\right) \wedge \left(\overline{\mathbb{E}}_m(f_i) \geq \overline{\mathbb{E}}_m(f_j)\right)$

Complete preorders among partial assignments

Patterns are assigned partially to a non-empty subset of Ω: $\mathcal{F} = \{f_A, A \in 2^\Omega \setminus \{\emptyset\}\}$

- generalized maximin $f_{A_i} \succcurlyeq_* f_{A_j} \iff \underline{\mathbb{E}}_m(f_{A_i}) \geq \underline{\mathbb{E}}_m(f_{A_j})$
- generalized maximax $f_{A_i} \succcurlyeq^* f_{A_j} \iff \overline{\mathbb{E}}_m(f_{A_i}) \geq \overline{\mathbb{E}}_m(f_{A_j})$
- generalized Hurwicz $f_{A_i} \succcurlyeq_\alpha f_{A_j} \iff \mathbb{E}_{m,\alpha}(f_{A_i}) \geq \mathbb{E}_{m,\alpha}(f_{A_j})$
- pignistic criterion $f_{A_i} \succcurlyeq_p f_{A_j} \iff \mathbb{E}_p(f_{A_i}) \geq \mathbb{E}_p(f_{A_j})$
- generalized OWA $f_{A_i} \succcurlyeq_\beta f_{A_j} \iff \mathbb{E}^{owa}_{m,\beta}(f_{A_i}) \geq \mathbb{E}^{owa}_{m,\beta}(f_{A_j})$
- generalized minimax regret $f_{A_i} \succcurlyeq_r f_{A_j} \iff \overline{R}(f_{A_i}) \leq \overline{R}(f_{A_j})$
- maximum expected utility $f_{A_i} \succcurlyeq_m f_{A_j} \iff EU(f_{A_i}) \geq EU(f_{A_j})$

Extending utility matrix via an OWA operator

The extended utility matrix $\hat{\mathbf{U}}_{(2^n-1)\times n}$ is crucial to both decision-making and performance evaluation. The utility of assigning one instance to set A should intuitively be a function of those utilities of each precise assignments within A:

$$\hat{u}_{A,j} = F\left(\{u_{ij} \mid \omega_i \in A\}\right) = \sum_{k=1}^{|A|} w_k u^A_{(k)j}.$$

Given the DM's *tolerance degree of imprecision*

$$TOL(\boldsymbol{w}) = \sum_{k=1}^{|A|} \frac{|A| - k}{|A| - 1} w_k = \gamma,$$

the weights corresponding to the OWA operator are obtained by maximizing the entropy

$$ENT(\boldsymbol{w}) = -\sum_{k=1}^{|A|} w_k \log w_k,$$

subject to $TOL(\boldsymbol{w}) = \gamma$ and $\sum_{k=1}^{|A|} w_k = 1$.

Example: the utility matrix extended by an OWA operator with $\gamma = 0.8$

acts	states of nature ω_1	ω_2	ω_3
$f_{\{\omega_1\}}$	1.0000	0.2000	0.1000
$f_{\{\omega_2\}}$	0.2000	1.0000	0.2000
$f_{\{\omega_3\}}$	0.1000	0.2000	1.0000
$f_{\{\omega_1,\omega_2\}}$	0.8400	0.8400	0.1800
$f_{\{\omega_1,\omega_3\}}$	0.8200	0.2000	0.8200
$f_{\{\omega_2,\omega_3\}}$	0.1800	0.8400	0.8400
$f_{\{\omega_1,\omega_2,\omega_3\}}$	0.7373	0.7455	0.7373

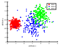
Evaluation of set-valued predictions

The classification performance is evaluated by the *averaged utility* in the test set T:

$$Acc(T) = \frac{1}{|T|} \sum_{i=1}^{|T|} \hat{u}_{\mathcal{F}^*_i, i^*}.$$

Experimental data

UCI Balance scale dataset and simulated Gaussian datasets



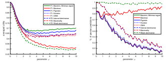
Experiments

Belief functions concerning the states of nature were generated through the DS theory-based neural network classifier.

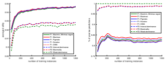
Classification performances with varying γ (UCI Balance scale dataset)

		DC1	DC2	DC3	DC4	DC5	DC6	DC7	DC8	DC9
averaged utility	γ=0.5	0.9186	**0.9188**	0.9186	0.9186	0.9186	0.9186	0.9187	0.9187	0.9187
	γ=0.6	0.9179	0.9184	0.9176	0.9179	0.9184	0.9176	**0.9188**	**0.9188**	0.9187
	γ=0.7	0.9059	0.9064	0.9052	0.9059	0.9056	0.9054	**0.9190**	**0.9190**	0.9187
	γ=0.8	0.9043	0.9032	0.9028	0.9043	0.9030	0.9024	**0.9191**	**0.9191**	0.9188
	γ=0.9	**0.9339**	0.9319	0.9325	**0.9339**	0.9331	0.9319	0.9192	0.9192	0.9188
	γ=1.0	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	0.9194	0.9194	0.9188
% of precision	γ=0.5	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	97.44%	97.44%	99.97%
	γ=0.6	88.96%	89.47%	88.96%	88.96%	89.18%	89.06%	97.44%	97.44%	99.97%
	γ=0.7	80.10%	80.77%	80.06%	80.10%	80.22%	80.26%	97.44%	97.44%	99.97%
	γ=0.8	69.70%	70.14%	69.63%	69.70%	69.82%	69.63%	97.44%	97.44%	99.97%
	γ=0.9	57.02%	57.76%	57.12%	57.02%	57.38%	57.12%	97.44%	97.44%	99.97%
	γ=1.0	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	97.44%	97.44%	99.97%

Performances with noised test sets (Gaussian dataset)



Performances with increasing training set size (Gaussian)



Conclusions

The set-valued predictions induced by a partial preorder turn into precise ones when information becomes more precise. In contrast, the criteria based on a complete preorder can provide set-valued predictions even when uncertainty is quantified by probabilities. Set-valued predictions perform better than precise ones in the case of complex data sets: therefore, the most cautious rules should be preferred in highly uncertain environments.