Cautious label-wise ranking with constraint satisfaction

18th International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems

CARRANZA-ALARCON Yonatan-Carlos
Ph.D. Candidate

Soundouss MESSOUDI
Ph.D. Candidate

DESTERCKE Sébastien
Ph.D Director

utc Université de Technologie Compiègne
heudiasyc
cnrs dépasser les frontières

15 Jun 2020 to 19 Jun 2020

Our approach in a nutshell

What ?

Cautious label-ranking by rank-wise decomposition

How ?

- Rank-wise decomposition
- For each decomposition, predict set of ranks using imprecise probabilities
- Use Constraint Satisfaction (CSP) to :
  - resolve inconsistencies
  - remove impossible assignments

Why ?

- weak information in structured settings more prone to be of use
- few rank-wise approaches (except score-based) for this problem

Overview

- Introduction to Label ranking problem
- Our proposal in details
- Evaluation
  - Settings and Datasets
  - Experimental results
- Conclusions

Label ranking problem

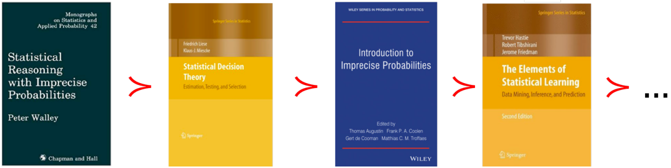
☞ Problem statement :

Let $\mathscr{K} = \{m_1, \ldots, m_K\}$ be a set of objects equipped with an order relation.

i.e. : ranked from "not relevant" → "relevant"

e.g. : Set of books ranked in order of preference (easy → hard)

$m_1 \qquad m_2 \qquad m_3 \qquad m_4$



A complete order relation on $\mathscr{K} \rightarrow \Lambda(\mathscr{K}) \subseteq \mathscr{K} \times \mathscr{K}$

$\Lambda(\mathscr{K})$: Set of complete rankings, $|\Lambda(\mathscr{K})| = !K$ (set of all permutations)

Label ranking problem

☞ **The goal of label ranking problem :**

Given a training data : $\mathscr{D} = \{\boldsymbol{x}_i, Y_i\}_{i=0}^{N} \subseteq \mathbb{R}^p \times \Lambda(\mathscr{K})$

Learning a complete ranking rule : $\varphi : \mathbb{R}^p \rightarrow \Lambda(\mathscr{K})$

☞ Example of training data :

X_1	X_3	Y
107.1	*Blue*	$m_1 \succ m_4 \succ m_2 \succ m_3$
−50	*Red*	$m_2 \succ m_3 \succ m_1 \succ m_4$
200	*Green*	$m_1 \succ m_3 \succ m_4 \succ m_2$
...	...	...

Overview

- Introduction to Label ranking problem
- Our proposal in details
- Evaluation
  - Settings and Datasets
  - Experimental results
- Conclusions

Cautious label-wise ranking with constraint satisfaction

Our proposal

Step ❶ Rank-wise decomposition

- for each label $m_i \in \mathscr{K}$, create a new data set $\mathbb{D}_{m_i}$

Step ❷ Cautious ordinal regression

- for each label $m_i \in \mathscr{K}$, predict a set of ranks using imprecise probabilities

Step ❸ Global inference with constraint satisfaction problem

- resolve inconsistencies
- remove impossibles assignments

Step ❶ : Rank-wise decomposition

$\mathbb{D}$

X_1	X_3	Y
107.1	*Blue*	$m_1 \succ m_3 \succ m_2 \succ m_4$
−50	*Red*	$m_2 \succ m_3 \succ m_1 \succ m_4$
200	*Green*	$m_1 \succ m_4 \succ m_3 \succ m_2$
…	…	…

$\mathbb{D}_1$

X_1	X_3	Y
107.1	*Blue*	1
−50	*Red*	3
200	*Green*	1
…	…	…

$\mathbb{D}_2$

X_1	X_2	Y
107.1	*Blue*	3
−50	*Red*	1
200	*Green*	4
…	…	…

$\mathbb{D}_3$

X_1	X_2	Y
107.1	*Blue*	2
−50	*Red*	2
200	*Green*	3
…	…	…

$\mathbb{D}_4$

X_1	X_2	Y
107.1	*Blue*	4
−50	*Red*	4
200	*Green*	2
…	…	…

FIGURE – Label-wise decomposition

☞ For each data set $\mathbb{D}_i$, solve an ordinal regression problem.

Step ❷ : Cautious ordinal regression

$\mathbb{D}_1$

X_1	X_3	Y
107.1	*Blue*	1
−50	*Red*	3
…	…	…

$\mathbb{D}_2$

X_1	X_2	Y
107.1	*Blue*	3
−50	*Red*	1
…	…	…

$\mathbb{D}_3$

X_1	X_2	Y
107.1	*Blue*	2
−50	*Red*	2
…	…	…

$\mathbb{D}_4$

X_1	X_2	Y
107.1	*Blue*	4
−50	*Red*	4
…	…	…

Learning with a set of probabilities $\mathscr{P}$

Maximality criterion

$$m \succ m' \iff \inf_{P \in \mathscr{P}_{Y|x}} \sum_{y \in \mathcal{K}} P(y|x)\left(\ell_{m'}(y) - \ell_{m}(y)\right) \geq 0$$

Loss functions of choice $m \in \mathcal{K}$

$$\ell_m : \mathcal{K} \to \mathbb{R}$$

$\mathbb{D}_1$

X_1	X_3	$\hat{Y}$
107.1	*Blue*	$\{1,2\}$
−50	*Red*	$\{3\}$
…	…	…

$\mathbb{D}_2$

X_1	X_2	$\hat{Y}$
107.1	*Blue*	$\{1,2,3\}$
−50	*Red*	$\{1\}$
…	…	…

$\mathbb{D}_3$

X_1	X_2	$\hat{Y}$
107.1	*Blue*	$\{2\}$
−50	*Red*	$\{1,2\}$
…	…	…

$\mathbb{D}_4$

X_1	X_2	$\hat{Y}$
107.1	*Blue*	$\{1,2,3,4\}$
−50	*Red*	$\{3,4\}$
…	…	…

Step ❷ : Cautious ordinal regression

$\mathbb{D}_1$

X_1	X_3	Y
107.1	*Blue*	1
−50	*Red*	3
…	…	…

$\mathbb{D}_2$

X_1	X_2	Y
107.1	*Blue*	3
−50	*Red*	1
…	…	…

$\mathbb{D}_3$

X_1	X_2	Y
107.1	*Blue*	2
−50	*Red*	2
…	…	…

$\mathbb{D}_4$

X_1	X_2	Y
107.1	*Blue*	4
−50	*Red*	4
…	…	…

Learning with a set of probabilities $\mathscr{P}$

Maximality criterion

$$m \succ m' \iff \inf_{P \in \mathscr{P}_{Y|x}} \sum_{y \in \mathcal{K}} P(y|x)(\ell_{m'}(y) - \ell_m(y)) \geq 0$$

Loss functions of choice $m \in \mathcal{K}$

$$\ell_m : \mathcal{K} \to \mathbb{R}$$

But, what loss ?

On the median in imprecise ordinal problems [2]

$\ell = L_1$ norm between ranks, loss of predicting rank j if k is true

$$\ell_j(k) = |j - k|$$

$\mathscr{P}$ described by lower/upper cumulative distributions $\underline{F}, \overline{F}$

Why L_1 ?

☞ prediction is guaranteed to be an "interval" of ranks,
☞ it corresponds to the set of possible medians (very easy to get).

$\mathbb{D}_1$

X_1	X_3	$\hat{Y}$
107.1	*Blue*	{1,2}
−50	*Red*	{3}
…	…	…

$\mathbb{D}_2$

X_1	X_2	$\hat{Y}$
107.1	*Blue*	{1,2,3}
−50	*Red*	{1}
…	…	…

$\mathbb{D}_3$

X_1	X_2	$\hat{Y}$
107.1	*Blue*	{2}
−50	*Red*	{1,2}
…	…	…

$\mathbb{D}_4$

X_1	X_2	$\hat{Y}$
107.1	*Blue*	{1,2,3,4}
−50	*Red*	{3,4}
…	…	…

Step ❷ : Cautious ordinal regression

Input datasets

$\mathbb{D}_1$

X_1	X_3	Y
107.1	*Blue*	1
−50	*Red*	3
...	...	...

$\mathbb{D}_2$

X_1	X_2	Y
107.1	*Blue*	3
−50	*Red*	1
...	...	...

$\mathbb{D}_3$

X_1	X_2	Y
107.1	*Blue*	2
−50	*Red*	2
...	...	...

$\mathbb{D}_4$

X_1	X_2	Y
107.1	*Blue*	4
−50	*Red*	4
...	...	...

Cautious ordinal classifier [2]

$$\mathscr{R}_{m_i} : \mathbb{R}^p \rightarrow \widehat{Y} \subseteq \mathscr{K}$$

Predictions

$\mathbb{D}_1$

X_1	X_3	$\widehat{Y}$
107.1	*Blue*	{1,2}
−50	*Red*	{3}
...	...	...

$\mathbb{D}_2$

X_1	X_2	$\widehat{Y}$
107.1	*Blue*	{1,2,3}
−50	*Red*	{1}
...	...	...

$\mathbb{D}_3$

X_1	X_2	$\widehat{Y}$
107.1	*Blue*	{2}
−50	*Red*	{1,2}
...	...	...

$\mathbb{D}_4$

X_1	X_2	$\widehat{Y}$
107.1	*Blue*	{1,2,3,4}
−50	*Red*	{3,4}
...	...	...

Step ❸ : Global inference with constraint satisfaction problem (CSP)

☞ Removal of impossible solutions

Consider the rank-wise $\mathscr{R}_i(\boldsymbol{x})$ cautious prediction of the first observation

$$\widehat{\mathscr{R}}_{m_1} = \{1,2\},\ \widehat{\mathscr{R}}_{m_2} = \{1,2,3\},\ \widehat{\mathscr{R}}_{m_3} = \{2\},\ \widehat{\mathscr{R}}_{m_4} = \{1,2,3,4\}$$

As $\mathscr{R}_{m_3}(\boldsymbol{x})$ has to take single value $\{2\}$,then the others should not retain it :

$$\widehat{\mathscr{R}}^*_{m_1} = \{1\},\ \widehat{\mathscr{R}}^*_{m_2} = \{1,3\},\ \widehat{\mathscr{R}}^*_{m_3} = \{2\},\ \widehat{\mathscr{R}}^*_{m_4} = \{1,3,4\},$$

....until removing all inconsistencies, we get :

$$\widehat{\mathscr{R}}'_{m_1} = \{1\},\ \widehat{\mathscr{R}}'_{m_2} = \{3\},\ \widehat{\mathscr{R}}'_{m_3} = \{2\},\ \widehat{\mathscr{R}}'_{m_4} = \{4\}$$

✌ It is well-known as the *all different constraint* which forces every label-wise prediction to assume a value different from the value of every other.

Overview

- Introduction to Label ranking problem
- Our proposal in details
- Evaluation
  - Settings and Datasets
  - Experimental results
- Conclusions

Datasets and experimental setting

Material and method

- ☞ 14 data sets issued from UCI repository [1].
- ☞ 10×10-fold cross-validation procedure.
- ☞ Binary decomposition + Naive imprecise classifier (NCC)
- ☞ Adaptive method to obtain the “optimal” value *s* of hype-param.
- ☞ Comparing with other precise approaches (LRT, RPC, SVM-LR)

Measuring results quality

Completeness (CP)

$$CP(\widehat{R}) = \frac{k^2 - \sum_{i=1}^{k} |\widehat{R}_i|}{k^2 - k}$$

- ☞ Max if one ranking possible
- ☞ Min if all rankings possible

Correctness (CR)

$$CR(\widehat{R}) = 1 - \frac{\sum_{i=1}^{k} \min_{\widehat{r}_i \in \widehat{R}_i} |\widehat{r}_i - r_i|}{0.5k^2}$$

☞ Equivalent to Spearman footrule if one ranking predicted

Completeness/Correctness trade-off

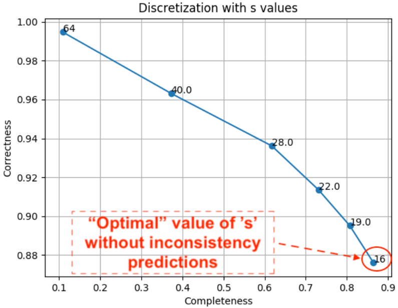


FIGURE – Evolution of the hyper-parm. s on glass

#	Data set	Type	#Inst	#Attributes	#Labels
a	authorship	classification	841	70	4
b	bodyfat	regression	252	7	7
c	calhousing	regression	20640	4	4
d	cpu-small	regression	8192	6	5
e	fried	regression	40768	9	5
f	glass	classification	214	9	6
g	housing	regression	506	6	6
h	iris	classification	150	4	3
i	pendigits	classification	10992	16	10
j	segment	classification	2310	18	7
k	stock	regression	950	5	5
l	vehicle	classification	846	18	4
m	vowel	classification	528	10	11
n	wine	classification	178	13	3

TABLE – Experimental data sets

Example : Inconsistency prediction :

$$\widehat{\mathscr{R}}'_{m_1} = \{1\},\ \widehat{\mathscr{R}}'_{m_2} = \{3\},\ \widehat{\mathscr{R}}'_{m_3} = \emptyset,\ \widehat{\mathscr{R}}'_{m_4} = \{4\}$$

Global solution is empty $\mathscr{R} = \emptyset$.

Experimental results

	LR-CSP-6	LRT	RPC	SVM-LR
a	93.90 ± 0.69 (1)	91.53 ± 0.31 (3)	93.21 ± 0.23 (2)	64.42 ± 0.36 (4)
b	54.12 ± 3.73 (1)	41.70 ± 1.48 (4)	50.43 ± 0.39 (3)	51.10 ± 0.49 (2)
c	61.05 ± 0.80 (1)	58.37 ± 0.28 (2)	51.85 ± 0.02 (3)	38.45 ± 0.02 (4)
d	68.72 ± 1.42 (1)	60.76 ± 0.30 (3)	61.93 ± 0.04 (2)	46.71 ± 0.87 (4)
e	99.20 ± 0.07 (2)	91.26 ± 0.06 (3)	99.92 ± 0.01 (1)	84.18 ± 2.67 (4)
f	91.95 ± 2.90 (1)	91.59 ± 0.47 (2)	90.83 ± 0.24 (3)	85.68 ± 0.33 (4)
g	79.21 ± 3.37 (2)	85.09 ± 0.46 (1)	74.86 ± 0.16 (3)	70.16 ± 0.46 (4)
h	99.36 ± 1.28 (1)	97.16 ± 0.55 (2)	92.75 ± 0.58 (3)	87.39 ± 0.37 (4)
i	91.31 ± 0.14 (3)	95.14 ± 0.05 (1)	94.12 ± 0.01 (2)	58.75 ± 2.71 (4)
j	91.20 ± 0.85 (3)	96.11 ± 0.10 (1)	94.52 ± 0.03 (2)	66.25 ± 3.05 (4)
k	88.63 ± 1.53 (2)	91.64 ± 0.27 (1)	82.23 ± 0.08 (3)	75.20 ± 0.17 (4)
l	85.29 ± 1.91 (3)	88.03 ± 0.44 (2)	89.24 ± 0.14 (1)	81.93 ± 1.00 (4)
m	88.23 ± 1.00 (1)	84.40 ± 0.62 (2)	72.88 ± 0.06 (3)	65.41 ± 1.21 (4)
n	98.20 ± 1.19 (1)	91.80 ± 0.87 (4)	94.58 ± 0.61 (2)	94.56 ± 0.50 (3)
avg.	85.03 ± 1.49(1.64)	83.18 ± 0.45(2.21)	81.67 ± 0.19(2.36)	69.30 ± 1.02(3.79)

TABLE – AVERAGE CORRECTNESS (%) COMPARED TO LR-CSP-6(%)

Overview

- Introduction to Label ranking problem
- Our proposal in details
- Evaluation
  - Settings and Datasets
  - Experimental results
- Conclusions

Conclusions and Perspectives

- ✓ Our proposal is also competitive w.r.t. other precise ones.
- ✓ We can use any imprecise classifier producing a set of probabilities.
- ✗ As dealing with inconsistent predictions (∼10% test dataset).
- ✗ A way to adapt the hyper-param. s_i by imprecise classifier $\mathscr{R}_i$.
- ✗ Using other imprecise classifier (eg. continuous classifier)

References

A. FRANK et A. ASUNCION. *UCI Machine Learning Repository*. 2010. URL : http://archive.ics.uci.edu/ml.

Sébastien DESTERCKE. "On the median in imprecise ordinal problems". In : *Annals of Operations Research* 256.2 (2017), p. 375-392.


Thank
You!