Multi-label Ranking from Positive and Unlabeled Data

Kanehira A, Harada T. CVPR, 2016

Outline

- Introduction
- Multi-label PU ranking
- Experiment

Introduction

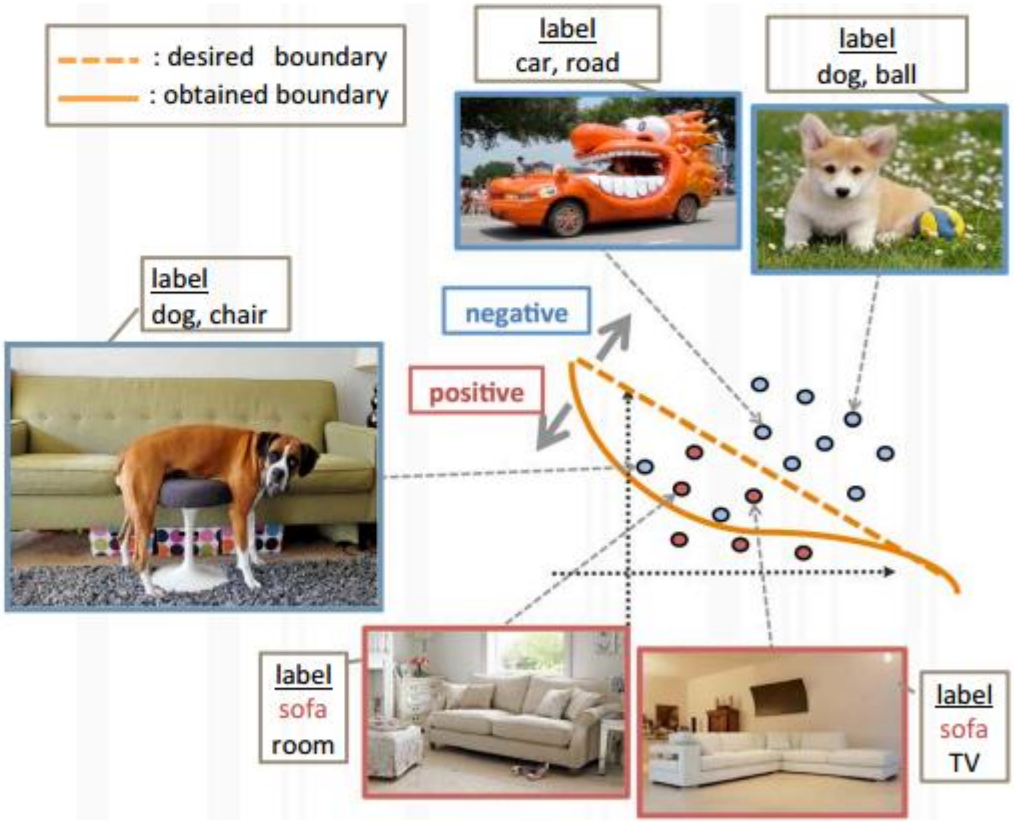


Figure 1. Multi-label dataset tends to have partially labeled samples. Absent labels are regarded as negative, and it affects classification performance.

Problem Setting

- 1. assigned labels are definitely positive,
- 2. absent labels are Not necessarily negative,
- 3. samples are allowed to take more than one label

PU Classification + **Multi-label classification** = **Multi-label PU classification**

Formulation

$$\min L_{\text{true}} = \mathrm{E}_{\mathrm{xy}}\left[R(\mathrm{f}(\mathrm{x}),\mathrm{y})\right]$$

$$R(\mathrm{f}(\mathrm{x}),\mathrm{y}) = p\left(f_i < f_j \middle| y_i = 1, y_j = 0\right)$$ mis-rank rate

$$\mathrm{x} \in \mathrm{R}^d : \text{sample}$$

$$\cancel{\mathrm{y} \in \{0,1\}^m : \text{true label}}$$ unknown

$$s \in \{0,1\}^m : \text{observed label}$$ known

where d is feature dimension and m is the number of classes

minimizing ranking loss, with only *observed s*

Relationship between *y* and *s*

$s \in \{0,1\}^m$: observed label

$\mathrm{y} \in \{0,1\}^m$: true label

$$
\begin{tikzcd}
\boxed{s_i = 1(\text{assigned})} \arrow[d] & \boxed{s_i = 0(\text{assigned})} \arrow[dl] \arrow[d] \\
\boxed{y_i = 1(\text{positive})} & \boxed{y_i = 0(\textit{negative})}
\end{tikzcd}
$$

Analysis of multi-label PU ranking (a)

$$\min L_{\mathrm{true}} = \mathrm{E}_{\mathrm{xy}}[R(\mathrm{f}(\mathrm{x}),\mathrm{y})]$$

$$R(\mathrm{f}(\mathrm{x}),\mathrm{y}) = p\left(f_i < f_j \middle| y_i = 1, y_j = 0\right)$$ (mis-rank rate)

We can't estimate mis-rank rate
because it depends on true (unknown) label

$$R_X(\mathrm{f}(\mathrm{x}),\mathrm{s}) = p\left(f_i < f_j \middle| s_i = 1, s_j = 0\right)$$ (pseudo mis-rank rate)

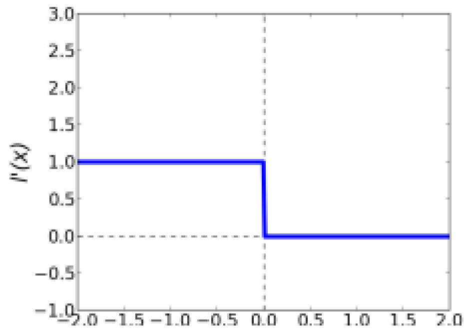
$$\begin{aligned} L_{\mathrm{PU}} &= \mathbb{E}_{\mathrm{xs}}[c_{ij} R_X(f(\mathrm{x}), \mathrm{s})] \\ &= L_{\mathrm{true}} - \mathrm{const} \end{aligned} \quad \left(\text{where} \quad c_{ij} = \frac{p(y_i = 1)}{p(s_i = 1, s_j = 0)}\right)$$

a) Loss function should be weighted properly

Analysis of multi-label PU ranking (b)

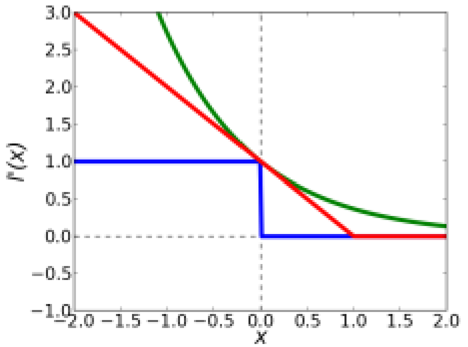
<u>optimization of loss function</u>

$$R(\mathrm{f}(\mathrm{x}), \mathrm{y}) = p(f_i < f_j | y_i = 1, y_j = 0)$$
$$= \mathbb{E}_{x|y_i=1, y_j=0}[l_{0\text{-}1}(f_i - f_j)]$$



Due to computationally complexity,
surrogate loss (e.g. hinge) is usually used.

$$= \mathbb{E}_{x|y_i=1, y_j=0}[l'_{\mathrm{sur}}(f_i - f_j)]$$

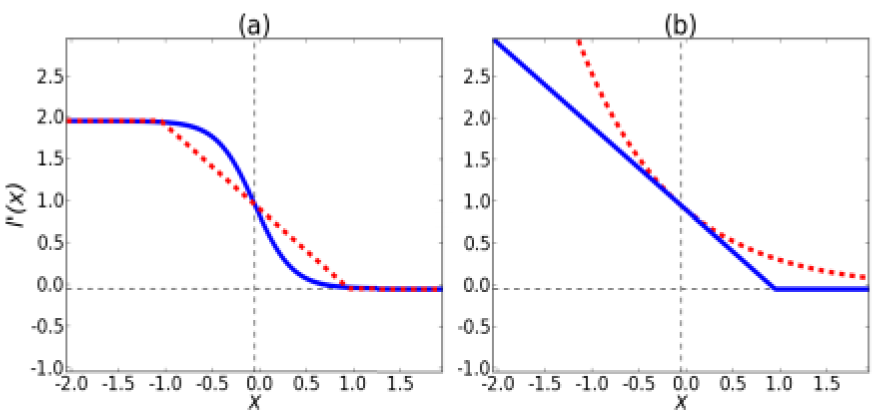


Analysis of multi-label PU ranking(b)

Using surrogate loss,

$$L'_{\mathrm{PU}} = L'_{\mathrm{true}} + \underbrace{p(y_i = 1, y_j = 1)\mathbb{E}_{x|y_i=1,y_j=1}[l'(f_i - f_j) + l'(f_j - f_i)]}_{\text{Surrogate loss generate bias}}$$

Bias can be cancelled
for symmetric surrogate loss.



b) Symmetric surrogate loss function should be used.

Experiment

❑ Setting:

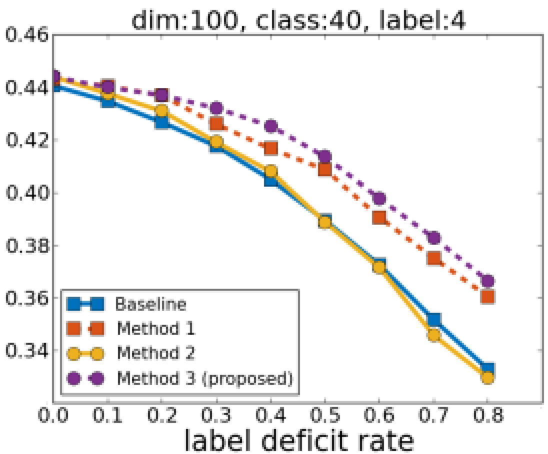
- synthetic dataset, image annotation dataset (MSCOCO, NUS-WIDE)
- compared 4 methods, which corresponding to each condition
- trained with data with 0-80% label deficit
- evaluated on Mean Average Precision

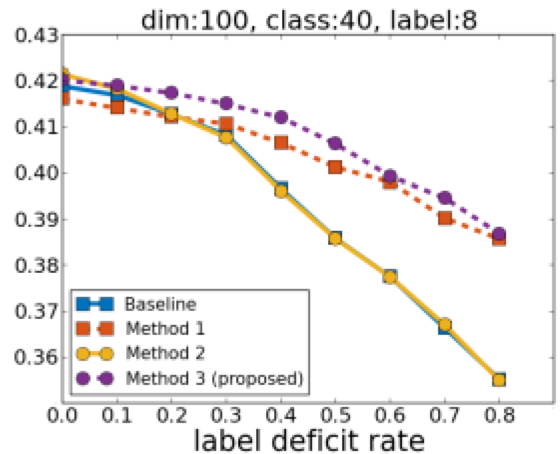
	Not symmetric (hinge loss)	Symmetric (ramp loss)
Not weighted	Baseline	Method ②
weighted	Method ①	Method ③(proposed)

Experiment

❑ Result:

synthetic





MSCOCO