MLA 2023

Disambiguated Attention Embedding for Multi-Instance Partial-Label Learning

Wei Tang, Weijia Zhang, Min-Ling Zhang

NeurIPS 2023

November 4, 2023

東南大學 SOUTHEAST UNIVERSITY

THE UNIVERSITY OF NEWCASTLE AUSTRALIA

Multi-Instance Partial-Label Learning (MIPL)

MIPL[1] can be seen as a generalized framework of multi-instance learning and partial-label learning. In MIPL, a training sample is represented as a multi-instance bag associated with a bag-level candidate label set, which comprises a ground-truth label along with several false positive labels.

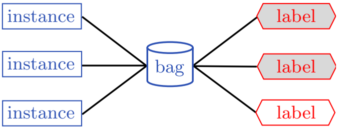

Figure: The framework of MIPL.

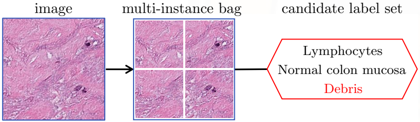

Figure: colorectal cancer classification (sourced from CRC-MIPL dataset).

[1] W. Tang, W. Zhang, M.-L. Zhang. Multi-instance partial-label learning: Towards exploiting dual inexact supervision. **SCIENCE CHINA Information Sciences (SCIS)**, in press.

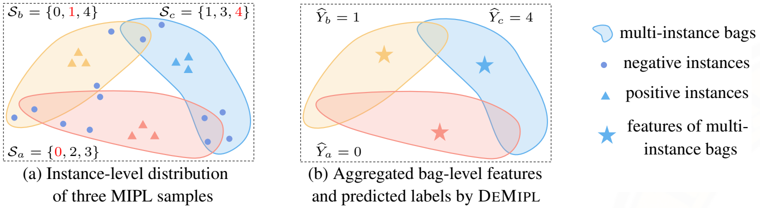

Figure: A brief illustration of DEMIPL.

The scheme based on the instance-space paradigm may be suboptimal as global bag-level information is ignored and the predicted labels of bags are sensitive to predictions of negative instances.

- **Algorithm**: we propose the first algorithm, named DEMIPL, based on the embedded-space paradigm for MIPL.
- **Dataset**: we introduce a real-world MIPL dataset CRC-MIPL for colorectal cancer classification.

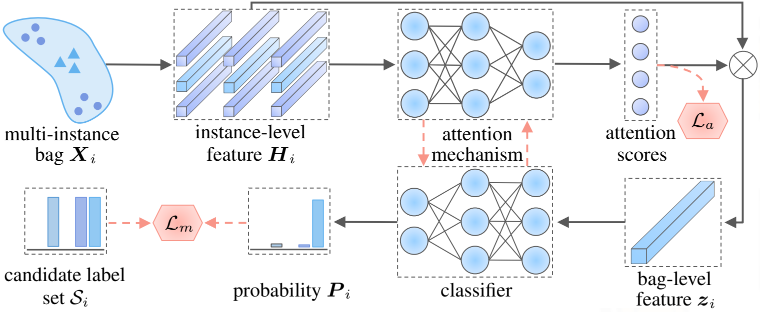

Figure: The framework of DEMIPL.

Two key procedures in DEMIPL:

- ▷ Feature aggregation: disambiguation attention mechanism
- ▷ Label disambiguation: momentum-based disambiguation strategy

Disambiguation Attention Mechanism

We propose a multi-class attention mechanism to calculate the attention score $a_{i,j}$ of $\boldsymbol{x}_{i,j}$ as follows:

$$a_{i,j} = \frac{1}{1 + \exp\left\{-\boldsymbol{W}^\top \left(\tanh\left(\boldsymbol{W}_v^\top \boldsymbol{h}_{i,j} + \boldsymbol{b}_v\right) \odot \operatorname{sigm}\left(\boldsymbol{W}_u^\top \boldsymbol{h}_{i,j} + \boldsymbol{b}_u\right)\right)\right\}}. \quad (1)$$

To ensure that the attention scores of positive instances should be higher than those of negative instances, the proposed attention loss is shown below:

$$\mathcal{L}_a = -\frac{1}{m} \sum_{i=1}^{m} \sum_{j=1}^{n_i} a_{i,j} \log a_{i,j}. \quad (2)$$

Momentum-based Disambiguation Strategy

We propose momentum-based disambiguation loss to accurately identify the ground-truth label from the candidate label set:

$$\mathcal{L}_m = \frac{1}{m}\sum_{i=1}^{m}\sum_{c=1}^{k} w_{i,c}^{(t)} \ell\left(f_c^{(t)}(\boldsymbol{z}_i^{(t)}), \mathcal{S}_i\right). \tag{3}$$

Initialize the weights:

$$w_{i,c}^{(0)} = \begin{cases} \frac{1}{|\mathcal{S}_i|} & \text{if } Y_{i,c} \in \mathcal{S}_i, \\ 0 & \text{otherwise,} \end{cases} \tag{4}$$

where $\frac{1}{|\mathcal{S}_i|}$ is the cardinality of the candidate label set $\mathcal{S}_i$.

Update the weights:

$$w_{i,c}^{(t)} = \begin{cases} \lambda^{(t)} w_{i,c}^{(t-1)} + (1-\lambda^{(t)})\frac{f_c^{(t)}(\boldsymbol{z}_i^{(t)})}{\sum_{j\in\mathcal{S}_i} f_j^{(t)}(\boldsymbol{z}_j^{(t)})} & \text{if } Y_{i,c} \in \mathcal{S}_i, \\ 0 & \text{otherwise,} \end{cases} \tag{5}$$

where t refers to the t-th epoch and $\lambda^{(t)} = \frac{T-t}{T}$.

Experimental Results

Table: Accuracy on the benchmark datasets.

Algorithm	r	MNIST-MIPL	FMNIST-MIPL	Birdsong-MIPL	SIVAL-MIPL
DeMipl	1	0.976±0.008	0.881±0.021	0.744±0.016	0.635±0.041
	2	0.943±0.027	0.823±0.028	0.701±0.024	0.554±0.051
	3	0.709±0.088	0.657±0.025	0.696±0.024	0.503±0.018
MiplGp	1	0.949±0.016•	0.847±0.030•	0.716±0.026•	0.669±0.019◦
	2	0.817±0.030•	0.791±0.027•	0.672±0.015•	0.613±0.026◦
	3	0.621±0.064•	0.670±0.052	0.625±0.015•	0.569±0.032◦
Mean					
Proden	1	0.605±0.023•	0.697±0.042•	0.296±0.014•	0.219±0.014•
	2	0.481±0.036•	0.573±0.026•	0.272±0.019•	0.184±0.014•
	3	0.283±0.028•	0.345±0.027•	0.211±0.013•	0.166±0.017•
Rc	1	0.658±0.031•	0.753±0.042•	0.362±0.015•	0.279±0.011•
	2	0.598±0.033•	0.649±0.028•	0.335±0.011•	0.258±0.017•
	3	0.392±0.033•	0.401±0.063•	0.298±0.009•	0.237±0.020•
Lws	1	0.463±0.048•	0.726±0.031•	0.265±0.010•	0.240±0.014•
	2	0.209±0.028•	0.720±0.025•	0.254±0.010•	0.223±0.008•
	3	0.205±0.013•	0.579±0.041•	0.237±0.005•	0.194±0.026•
PL-AGGD	1	0.671±0.027•	0.743±0.026•	0.353±0.019•	0.355±0.015•
	2	0.595±0.036•	0.677±0.028•	0.314±0.018•	0.315±0.019•
	3	0.380±0.032•	0.474±0.057•	0.296±0.015•	0.286±0.018•
MaxMin					
Proden	1	0.508±0.024•	0.424±0.045•	0.387±0.014•	0.316±0.019•
	2	0.400±0.037•	0.377±0.040•	0.357±0.012•	0.287±0.024•
	3	0.345±0.048•	0.309±0.058•	0.336±0.012•	0.250±0.018•
Rc	1	0.519±0.028•	0.731±0.027•	0.390±0.014•	0.306±0.023•
	2	0.469±0.035•	0.666±0.027•	0.371±0.013•	0.288±0.021•
	3	0.380±0.048•	0.524±0.034•	0.363±0.010•	0.267±0.020•
Lws	1	0.242±0.042•	0.435±0.049•	0.225±0.038•	0.289±0.017•
	2	0.239±0.048•	0.406±0.040•	0.207±0.034•	0.271±0.014•
	3	0.218±0.017•	0.318±0.064•	0.216±0.029•	0.244±0.023•
PL-AGGD	1	0.527±0.035•	0.391±0.040•	0.383±0.014•	0.397±0.028•
	2	0.439±0.020•	0.371±0.037•	0.372±0.020•	0.360±0.029•
	3	0.321±0.043•	0.327±0.028•	0.344±0.011•	0.328±0.023•

Table: Accuracy on the CRC-MIPL.

Algorithm	CRC-MIPL-Row	CRC-MIPL-SBN	CRC-MIPL-KMeansSeg	CRC-MIPL-SIFT
DeMipl	0.408±0.010	0.486±0.014	0.521±0.012	0.532±0.013
MiplGp	0.432±0.005◦	0.335±0.006•	0.329±0.012•	–
Mean				
Proden	0.365±0.009•	0.392±0.008•	0.233±0.018•	0.334±0.029•
Rc	0.214±0.011•	0.242±0.012•	0.226±0.009•	0.209±0.007•
Lws	0.291±0.010•	0.310±0.006•	0.237±0.008•	0.270±0.007•
PL-AGGD	0.412±0.008	0.480±0.005•	0.358±0.008•	0.363±0.012•
MaxMin				
Proden	0.401±0.007	0.447±0.011•	0.265±0.027•	0.291±0.011•
Rc	0.227±0.012•	0.338±0.010•	0.208±0.007•	0.246±0.008•
Lws	0.299±0.008•	0.382±0.009•	0.247±0.005•	0.230±0.007•
PL-AGGD	0.460±0.008◦	0.524±0.008◦	0.434±0.009•	0.285±0.009•

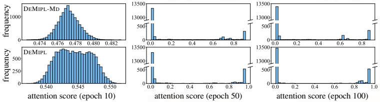

Figure: The frequency distribution.

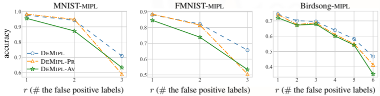

Figure: Accuracy of DeMipl and variants.

- ▷ DeMipl outperforms the compared algorithms in 96.3% of cases on benchmark datasets and in 88.6% of cases on the CRC-MIPL dataset.
- ▷ Both the attention loss and momentum-based disambiguation strategy are conducive to improving accuracy.

Thank you for listening!

More resources are available at `http://palm.seu.edu.cn/zhangml/` and `https://github.com/tangw-seu/DEMIPL`.

Codes & Datasets

GitHub