MLA 2023

Disambiguated Attention Embedding for Multi-Instance Partial-Label Learning

Wei Tang, Weijia Zhang, Min-Ling Zhang

NeurIPS 2023

November 4, 2023





MIPL¹ 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.

image multi-instance bag candidate label set

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

¹W. Tang, W. Zhang, M.-L. Zhang. <u>Multi-instance partial-label learning: Towards exploiting dual inexact supervision</u>. 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



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

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

$$\mathcal{L}_{a} = -\frac{1}{m} \sum_{i=1}^{m} \sum_{j=1}^{n_{i}} a_{i,j} \log a_{i,j}.$$
 (2)



(4)

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

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

Initialize the weights:

$$w_{i,c}^{(0)} = \left\{ egin{array}{cc} rac{1}{|\mathcal{S}_i|} & ext{if } Y_{i,c} \in \mathcal{S}_i, \ 0 & ext{otherwise,} \end{array}
ight.$$

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

Update the weights:

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

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



Table: Accuracy on the benchmark datasets.

Appendix 1 0.076,12008 0.8381140021 0.0744410016 0.035410011 DEMIRE 2 0.094440037 0.823410021 0.074440014 0.554440611 1 0.094490037 0.823410025 0.696440024 0.554440611 1 0.094490037 0.8474.00304 0.716440026 0.654440613 MIPLGP 2 0.8374.104034 0.716440026 0.66942.00105 0.66942.00105 MIPLGP 2 0.8374.104034 0.7014.10027 0.672.40015 0.66942.00102 PRODEN 2 0.4814.40036 0.7374.00026 0.2254.00102 0.6994.00102 1 0.6554.00514 0.7374.00026 0.2274.20015 0.1664.00174 2 0.4814.40036 0.7324.00126 0.3224.00126 0.2344.00126 1 0.6554.00151 0.7354.00126 0.3254.00126 0.2344.00126 2 0.2944.00126 0.7264.00137 0.2344.00126 0.2344.00126 1 0.6544.00126 0.7344.00149 0.2344.00146 0.2344.00116 3	Algorithm	æ	MNIST MIDI	EMNIST MIDI	Birdeong MIDI	SIVAL MIDI			
DEMIRI, 2 0.0441-0007 0.0821-0003 0.0711-0003 0.0835-10003 1 0.0994-0008 0.6571-0003 0.6994-0004 0.9393-00108 0.9393-00108 MIPLGP 2 0.0471-0004 0.9714-00074 0.6721-00158 0.6633-00208 0.6633-00208 0.6633-00208 0.6633-00208 0.6633-00208 0.6633-00208 0.6633-00208 0.6633-00208 0.6633-00208 0.6633-00208 0.6633-00208 0.6633-00208 0.6633-00208 0.6633-00208 0.6633-00208 0.6633-00208 0.633-00208 0.633-00208 0.2724-00109 0.1844-00149 0.2943-00028 0.3635-00178 0.3624-00178 0.3624-00178 0.3624-00178 0.3624-00178 0.3624-00178 0.3724-0028 0.3724-0028 0.2344-001078 0.3724-0028 0.2342-00178 0.3924-00318 0.7624-0028 0.3524-00178 0.3724-00278 0.2344-00178 0.3724-00278 0.2344-00178 0.3724-00278 0.2344-00178 0.3724-00278 0.3724-00278 0.3724-00278 0.3724-00278 0.3724-00278 0.3724-00278 0.3724-00278 0.3724-00278 0.3724-00278 0.3724-00278 0	Aigoriunn	L'	0.076±0.008	0.991±0.021	0.744±0.016	0.625±0.041			
DEMIR. 2 0.9331002 0.03231002 0.0321002 0.0321002 0.0321002 0.0321002 0.0321002 0.0321002 0.0321002 0.03210010 0.0321002 0.03210010 0.03210000 0.02310000 0.02310000 0.02310000 0.02310000 0.02310000 0.02310000 0.02310000 0.02310000 0.02310000 0.02310000 0.02310000 0.02310000 0.023100000 0.023100000 0.023100000 <t< td=""><td rowspan="2">DEMIPL</td><td></td><td>0.970±0.008</td><td>0.881±0.021</td><td>0.744 ± 0.010 0.701 ± 0.024</td><td>0.055 ± 0.041 0.554±0.051</td></t<>	DEMIPL		0.970±0.008	0.881±0.021	0.744 ± 0.010 0.701 ± 0.024	0.055 ± 0.041 0.554 ±0.051			
1 0.4911006 0.00711002 0.49121006 0.2011006 MIPLGP 2 0.49121006 0.49121007 0.20121006 0.20121006 MIPLGP 2 0.49121007 0.49121007 0.20231006 0.20121007 MIPLGP 2 0.69121007 0.20231007 0.20231007 0.20321007 PRODEN 1 0.6052140028 0.697210072 0.20231007 0.20321007 1 0.605210028 0.697210072 0.202410014 0.211210014 0.211210014 2 0.48140.0036 0.5734.00267 0.2214.10013 0.1844.0014 0.2344.0017 0.3024.0037 0.3024.0037 0.3024.0037 0.2344.0017 0.3024.0037 0.2344.0016 0.2344.0017 0.3024.0037 0.2344.0017 0.3024.0037 0.2344.0017 0.3024.0037 0.2344.0017 0.2344.0017 0.2344.0017 0.2344.0017 0.2344.0017 0.2344.0017 0.2344.0017 0.2344.0017 0.2344.0017 0.2344.0017 0.2344.0017 0.2344.0017 0.2344.0017 0.2344.0017 0.2344.0017 0.2344.0017 0.2344.0017 <td>2</td> <td>0.943±0.027</td> <td>0.823±0.028</td> <td>0.701 ± 0.024 0.606 ± 0.024</td> <td>0.334 ± 0.031 0.502±0.018</td>		2	0.943±0.027	0.823±0.028	0.701 ± 0.024 0.606 ± 0.024	0.334 ± 0.031 0.502 ±0.018			
MIPLGP 0 0.347±0036 0.71±0.027 0.672±0.015 0.6613±0.026 MIPLGP 0.317±0.003 0.672±0.015 0.6613±0.026 0.6613±0.026 MIR 0.605±0.015 0.6613±0.026 0.6613±0.026 0.6613±0.026 PRODEN 1 0.695±0.015 0.562±0.015 0.562±0.015 0.562±0.015 PRODEN 1 0.681±0.016 0.573±0.016 0.272±0.010 0.184±0.014 3 0.328±0.026 0.355±0.017 0.325±0.017 0.325±0.017 0.325±0.017 0.325±0.017 0.325±0.017 0.255±0.016 0.275±0.0116 0.275±0.0116 0.255±0.016 0.255±0.016 0.255±0.016 0.255±0.016 0.255±0.016 0.255±0.016 0.255±0.016 0.235±0.016 0.235±0.016 0.235±0.016 0.235±0.016 0.235±0.016 0.235±0.016 0.235±0.015 0.236±0.018 0.35±0.015 0.236±0.018 0.35±0.015 0.236±0.018 0.35±0.015 0.236±0.018 0.35±0.014 0.35±0.015 0.236±0.018 0.345±0.014 0.315±0.0105 0.35±0.014 0.35±0.014 0.345±0.014 0.315±0.016		3	0.709±0.088	0.037±0.023	0.090±0.024	0.003±0.018			
$\begin{array}{rrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrr$	MIPLGP		0.949±0.016•	0.84/±0.030•	0.716±0.026•	0.009±0.0190			
1 0.0631:00.064 0.0690:00.025 0.0635:00.015 0.0639:00.0125 PRODEN 1 0.6655:00.224 0.6697:00.042 0.296:10.0144 0.219:10.0144 PRODEN 1 0.6655:10.025 0.6697:00.042 0.296:10.0144 0.219:10.0144 1 0.6655:10.025 0.673:10.024 0.2724:0.0104 0.219:10.0144 1 0.6654:10.025 0.2724:0.0104 0.184:4.00.0144 2 0.283:4.0025 0.234:5.0027 0.211:10.013 0.1664:0.0174 Rc 2 0.384:10.013 0.573:4.0014 0.3624:0.015 0.279:4.00114 1 0.455:10.0168 0.773:4.00142 0.3264:10.010 0.237:4.00174 2 0.394:10.0128 0.726:4.00174 0.2364:10.0108 0.2364:10.018 1 0.455:10.0168 0.779:4.0014 0.2364:10.018 0.3364:10.018 2 0.394:10.0128 0.737:4.0014 0.3354:10.018 0.3364:10.018 2 0.394:10.0128 0.374:10.014 0.3364:10.018 0.3364:10.018 3 0.345:10.0158 <t< td=""><td>2</td><td>0.81/±0.030•</td><td>0./91±0.02/•</td><td>0.6/2±0.015•</td><td>0.613±0.0260</td></t<>		2	0.81/±0.030•	0./91±0.02/•	0.6/2±0.015•	0.613±0.0260			
Nikelin Nikelin PRODEN 1 0.005140024 0.009720024 0.2082400144 0.2194100145 2 0.0181410026 0.0272401019 0.1844400104 0.1844400144 3 0.038140026 0.027240107 0.2194100145 0.1844400144 2 0.038140026 0.035540017 0.2114100115 0.1844400144 2 0.0384400316 0.7353400426 0.335240017 0.2294400176 3 0.324240315 0.6494200286 0.3354200176 0.2374400204 1 0.4634400486 0.7264400276 0.2384400010 0.2374400204 1 0.4634200486 0.7264400276 0.3344400104 0.2374400204 2 0.2095400276 0.7344400276 0.3344400104 0.03154400176 1 0.4574400176 0.3774400176 0.3374400176 0.3316400176 2 0.4596400376 0.3774400376 0.3374400146 0.3164400128 1 0.5198400276 0.3374400146 0.3164400128 0.3164400128 1 0.5198400276		3	0.621±0.064•	0.670±0.052	0.625±0.015•	0.569±0.0520			
PRODEN 1 0.0605-0102: 0.0971-01042: 0.2952-0014: 0.2192-0014: RCDEN 0.4841:0.036: 0.5732-0.026: 0.2724:0.019: 0.1844:0.0174: 3 0.2832-0.028: 0.3452-0.027: 0.2114:0.0114: 0.1646:0.0174: RC 2 0.5984:0.038: 0.6732-0.027: 0.3214:0.0124: 0.2794:0.0114: 3 0.3224:0.033: 0.6491:0.028: 0.3284:0.0104: 0.2374:0.0124: 1 0.4324:0.048: 0.7264:0.0134: 0.2374:0.0264: 0.2374:0.0104: LWS 1 0.4324:0.014: 0.354:0.0104: 0.2374:0.0104: 0.2374:0.0104: LWS 1 0.4324:0.014: 0.5794:0.0114: 0.2374:0.0104: 0.354:0.0114: 2 0.3554:0.015: 0.5794:0.0114: 0.2374:0.0104: 0.3154:0.0119: 2 0.3554:0.0124: 0.4714:0.0157: 0.2964:0.015: 0.354:0.0118: 2 0.354:0.0144: 0.3774:0.0144: 0.3154:0.0104: 0.3154:0.0118: 3 0.384:0.0144: 0.3774:0.0144: 0.3154:0.0118: 0.3154:0.0118: <td colspan="9">Mean</td>	Mean								
PRODEN 2 0.4481:e00.66 0.57:e00.026 0.27:e200.018 0.18:e400.014 Q 233:e0028 0.435:e0027 0.211:e0013 0.166:e0017 RC 2 0.598:e10031 0.733:e0142 0.322:e00116 0.279:e00111 RC 2 0.598:e10031 0.649:e0028 0.325:e0101 0.258:e01071 3 0.392:e10035 0.649:e10208 0.355:e0110 0.258:e01071 4 0.432:e100436 0.726:e100108 0.228:e100108 0.228:e100108 1 0.463:e10048 0.726:e100108 0.226:e100108 0.249:e10148 2 0.209:e10128 0.726:e100128 0.237:e100148 0.238:e100178 2 0.209:e10028 0.721:e100278 0.338:e100198 0.338:e100198 1 0.671:e100279 0.733:e100108 0.335:e100158 0.355:e100158 2 0.398:e100478 0.378:e100108 0.355:e100158 0.355:e100158 1 0.598:e100478 0.378:e100148 0.316:e100198 0.356:e100158 0.356:e100158 1 0.598:e100248	PRODEN	1	0.605±0.023•	0.69/±0.042•	0.296±0.014•	0.219±0.014•			
$\begin{array}{c c c c c c c c c c c c c c c c c c c $		2	0.481±0.036•	0.573±0.026•	0.272±0.019•	$0.184 \pm 0.014 \bullet$			
$\begin{array}{rrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrr$		3	0.283±0.028•	0.345±0.027•	0.211±0.013•	0.166±0.017•			
$\begin{array}{rrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrr$	RC	1	0.658±0.031•	0.753±0.042•	$0.362 \pm 0.015 \bullet$	$0.279 \pm 0.011 \bullet$			
1 0.392-0033; 0.401-0005; 0.298-0009; 0.237+0020; 1 0.433+0048; 0.726+0.018; 0.265+0.010; 0.249+0.014; 2 0.209+0.028; 0.726+0.018; 0.255+0.010; 0.249+0.014; 3 0.205+0.013; 0.579±0.014; 0.237±0.005; 0.345±0.018; PL-AGGD 0.409±0.014; 0.731±0.002; 0.331±0.016; 0.355±0.015; 0.308±0.032; 0.731±0.026; 0.331±0.016; 0.315±0.019; 3 0.308±0.032; 0.474±0.037; 0.296±0.015; 0.286±0.018; MaxMin MaxMin 0.316±0.010; 0.2316±0.019; 0.316±0.019; PRODEN 1 0.5198±0.024; 0.371±0.004; 0.316±0.019; 0.316±0.019; RC 2 0.409±0.037; 0.371±0.010; 0.316±0.019; 0.316±0.019; 3 0.345±0.014; 0.391±0.014; 0.316±0.019; 0.316±0.019; 0.316±0.019; 1 0.519±0.025; 0.771±0.012; 0.391±0.014; 0.361±0.010; 0.281±0.014; 0.316±0.019; 2		2	0.598±0.033•	$0.649 \pm 0.028 \bullet$	$0.335 \pm 0.011 \bullet$	$0.258 \pm 0.017 \bullet$			
$ \begin{array}{c c c c c c c c c c c c c c c c c c c $		3	0.392±0.033•	0.401±0.063•	$0.298 \pm 0.009 \bullet$	$0.237 \pm 0.020 \bullet$			
$ \begin{array}{rrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrr$	Lws	1	0.463±0.048•	0.726±0.031•	0.265±0.010•	0.240±0.014•			
$\begin{array}{c c c c c c c c c c c c c c c c c c c $		2	0.209±0.028•	$0.720 \pm 0.025 \bullet$	$0.254 \pm 0.010 \bullet$	0.223±0.008•			
$\begin{tabular}{ c c c c c c c c c c c c c c c c c c c$		3	0.205±0.013•	$0.579 \pm 0.041 \bullet$	$0.237 \pm 0.005 \bullet$	0.194±0.026•			
$ \begin{array}{rrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrr$	PL-AGGD	1	0.671±0.027•	0.743±0.026•	0.353±0.019•	0.355±0.015•			
$\begin{tabular}{ c c c c c c c c c c c c c c c c c c c$		2	0.595±0.036•	0.677±0.028•	0.314±0.018•	0.315±0.019•			
$\begin{tabular}{ c c c c c c c c c c c c c c c c c c c$		3	0.380±0.032•	0.474±0.057•	0.296±0.015•	0.286±0.018•			
$\begin{tabular}{ c c c c c c c c c c c c c c c c c c c$	MaxMin								
$\begin{array}{rrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrr$	PRODEN	1	0.508±0.024•	0.424±0.045•	0.387±0.014•	0.316±0.019•			
$\begin{array}{cccccccccccccccccccccccccccccccccccc$		2	0.400±0.037•	0.377±0.040•	0.357±0.012•	0.287±0.024•			
$\begin{array}{c ccccccccccccccccccccccccccccccccccc$		3	0.345±0.048•	0.309±0.058•	0.336±0.012•	0.250±0.018•			
$\begin{array}{rrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrr$	Rc	1	0.519±0.028•	0.731±0.027•	0.390±0.014•	0.306±0.023•			
$\begin{array}{cccccccccccccccccccccccccccccccccccc$		2	0.469±0.035•	0.666±0.027•	0.371±0.013•	0.288±0.021•			
$\begin{array}{c c c c c c c c c c c c c c c c c c c $		3	0.380±0.048•	0.524±0.034•	0.363±0.010•	0.267±0.020•			
$ \begin{array}{cccccccccccccccccccccccccccccccccccc$	Lws	1	0.242±0.042•	0.435±0.049•	0.225±0.038•	0.289±0.017•			
$\begin{array}{ c c c c c c c c c c c c c c c c c c c$		2	0.239±0.048•	0.406±0.040•	0.207±0.034•	0.271±0.014•			
$\begin{array}{c ccccccccccccccccccccccccccccccccccc$		3	0.218±0.017•	0.318±0.064•	0.216±0.029•	0.244±0.023•			
PL-AGGD 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•	PL-AGGD	1	0.527±0.035•	0.391±0.040•	0.383±0.014•	0.397±0.028•			
3 0.321±0.043• 0.327±0.028• 0.344±0.011• 0.328±0.023•		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±0.0050	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	$0.480 \pm 0.005 \bullet$	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	$0.460 {\pm} 0.008 {\circ}$	$0.524 {\pm} 0.008 \circ$	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.

MLA 2023

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