Supplementary Material for "Similarity-based Multi-Dimensional Multi-Label Classification"

To facilitate understanding, Table 1 summarizes the notations used in The SIDLE Approach section.

Table 1: Summary of the notations used in our paper.

Notation	Descriptions
d	number of features in input space
q	number of label spaces (dimensions) in output space
K_j	number of class labels in the j-th label space ($1 \leq j \leq q$)
m	number of MDML training samples
$\mathcal{X}$	the d-dimensional input (feature) space, i.e., $\mathcal{X} = \mathbb{R}^d$
$\boldsymbol{l}_i$	the i-th label vector where $\boldsymbol{l}_i = [\boldsymbol{l}_i^1; \boldsymbol{l}_i^2; \cdots; \boldsymbol{l}_i^1] \in \{0,1\}^K$
$\boldsymbol{x}_i$	the i-th feature vector where $\boldsymbol{x}_i = [x_{i1}, x_{i2}, \ldots, x_{id}]^\top \in \mathcal{X}$
$\mathcal{Y}$	the output space where $\mathcal{Y} = \bigcup_{j=1}^{q} \mathcal{Y}^j$
$\mathcal{Y}^j$	the j-th label space where $\mathcal{Y}^j = \{y_1^j, y_2^j, \ldots, y_{K_j}^j\}$ ($1 \leq j \leq q$)
$\mathcal{D}$	the set of MDML samples where $\mathcal{D} = \{(\boldsymbol{x}_i, \boldsymbol{l}_i) \mid 1 \leq i \leq m\}$
$\boldsymbol{x}_*$	the unseen instance feature vector
$\hat{\boldsymbol{l}}_*$	the predicted label vector for unseen instance $\boldsymbol{x}_*$
f	the mapping function: $\mathcal{X} \mapsto \mathcal{Y}$ from $\mathcal{D}$
k	the parameter of the number of nearest neighbors considered
$\mathcal{N}_k^I(\boldsymbol{x}_*)$	the indices set for k nearest neighbors of unseen instance $\boldsymbol{x}_*$ where $\mathcal{N}_k^I(\boldsymbol{x}_*) = \{i_r \mid 1 \leq r \leq k\}$
$\mathcal{N}_k^L(\boldsymbol{x}_*)$	The indices set for k nearest neighbors of the labels of unseen instance $\boldsymbol{x}_*$ where $\mathcal{N}_k^L(\boldsymbol{x}_*) = \{i_r \mid 1 \leq r \leq k\}$
$\boldsymbol{\Theta}^j$	the predictive model parameter for the j-th label space where $\boldsymbol{\Theta}^j = [\boldsymbol{\theta}_1^j, \ldots, \boldsymbol{\theta}_{K_j}^j] \in \mathbb{R}^{d \times K_j}$
$\boldsymbol{p}_i$	the real-valued hypothesis label vector of instance $\boldsymbol{x}_i$
$\boldsymbol{p}_*$	the real-valued hypothesis label vector of unseen instance $\boldsymbol{x}_*$
$\delta_{ja}^{\boldsymbol{x}_*}$	the number of samples that are associated with the label y_a^j in $\boldsymbol{x}_*$'s k nearest neighbors
$\epsilon_{ja}^{\boldsymbol{x}_*}$	the corresponding bias for $\delta_{ja}^{\boldsymbol{x}_*}$
$\zeta_{ja}^{\boldsymbol{x}_*}$	the combination of $\delta_{ja}^{\boldsymbol{x}_*}$ and its bias $\epsilon_{ja}^{\boldsymbol{x}_*}$, where $\zeta_{ja}^{\boldsymbol{x}_*} = \delta_{ja}^{\boldsymbol{x}_*} + \epsilon_{ja}^{\boldsymbol{x}_*}$
$\boldsymbol{z}_{\boldsymbol{x}_*}$	the length-K similarity information vector for unseen instance $\boldsymbol{x}_*$
$\boldsymbol{z}_{\boldsymbol{x}_*}^I$	the instance-based similarity vector
$\boldsymbol{z}_{\boldsymbol{x}_*}^L$	the label-based similarity vector
$\boldsymbol{\xi}_i$	the augmented feature vector for instance $\boldsymbol{x}_i$ where $\boldsymbol{\xi}_i = \boldsymbol{z}_{\boldsymbol{x}_i}^I + \boldsymbol{z}_{\boldsymbol{x}_i}^L + \boldsymbol{p}_i$
$\widetilde{\mathcal{D}}$	the new version of MDML samples with feature augmented
$\widetilde{\boldsymbol{x}}_i$	the new instance feature vector with augmented features where $\widetilde{\boldsymbol{x}}_i = [\boldsymbol{x}_i; \boldsymbol{\xi}_i]$