T-MDML: Triplet-based Multiple Distance Metric Learning for Multi-Instance Multi-Label Classification with Label Correlation

Dongyeon Kim[1], **Yejin Kan**[1], **Gangman Yi**[1, 2 *]

[1]Department of Computer Science & Artificial Intelligence, Dongguk University, Seoul, 04620, Korea
[2]Division of AI Software Convergence, Dongguk University, Seoul, 04620, Korea
dongyeon@dgu.ac.kr, kanyj@dgu.ac.kr, gangman@dongguk.edu

Abstract

The multi-instance multi-label (MIML) problem is a new supervised learning paradigm that has emerged to efficiently represent complex data. Therefore, various similarity-based algorithms have been proposed, but existing algorithms commonly measure similarity by considering only the structural relationships in the feature space without utilizing information from the label space. As these approaches do not adequately reflect the complex properties of MIML data, it is essential to improve the accuracy of MIML classification by utilizing information from both feature and label spaces. Thus, we propose a new algorithm, T-MDML: triplet-based multiple distance metric learning for MIML. T-MDML defines a distance metric by learning a global property shared by the entire label space and a label-specific property for each label. In addition, we simultaneously consider the structural characteristics of features and label space to extract label correlation and incorporate it into the optimization process. In experiments, we demonstrate the efficiency of our label correlation estimation method and verify its performance by applying it to MIMLkNN. We also demonstrate T-MDML's relative superiority over existing MIML algorithms, as well as its scalability when applied to similarity-based MIML methods.

Introduction

Multi-instance multi-label learning (MIML) is a new machine learning framework for efficiently representing semantically informative objects (Zhou and Zhang 2006; Zhou et al. 2012; Qiu et al. 2023). It is a two-way extension of the general single-instance single-label learning (SISL) framework. The extension to multiple instances enables a flexible representation of the input space, and the extension to multiple labels enables multiple representations of the output space. In multi-instance learning (MIL), input data are represented as a bag consisting of multiple instances. In each bag, the number of instances can be different; however, the number of attributes is always the same (Amores 2013). Multi-label learning (MLL) involves the problem of simultaneously assigning multiple labels to each instance (Zhang and Zhou 2007a, 2014). Specifically, MIML deals with the problem that multiple labels are simultaneously assigned to each bag, and only bag-level label descriptions are provided to represent complex objects (Briggs et al. 2012). Therefore, the MIML framework is efficient for various applications such as image classification (Zha et al. 2008), image annotation (Guo et al. 2018), text categorization (Zhang and Zhou 2008), and protein function prediction (Liu et al. 2022).

MIML problems can be solved using two general approaches. The degeneration method transforms the MIML problem into an MIL or MLL problem by degenerating it to a lower-dimensional problem (Zhou et al. 2012; Zafra and Gibaja 2023). The advantage of this approach is that existing MIL and MLL algorithms can be used to solve the MIML problem; however, important information may be removed during the degeneration process (Zafra and Gibaja 2023). The direct method solves this problem by directly measuring the similarity between bags or using a model-based algorithm (Li et al. 2021). Especially, various distance metrics are used to measure similarity between bags, and nearest neighbor (NN)-based MIML algorithms have been proposed (Zafra and Gibaja 2023). However, the proposed metrics are based on the Euclidean distance, and they cannot measure similarity reflecting label information. Specifically, in the complex structures of multiple labels, NN-based algorithms are dependent on distance, making it difficult to perform efficient classification. To solve this problem, (Jin, Wang, and Zhou 2009; Hu et al. 2019) introduced the distance metric learning (DML) to directly learn similarity based on labels.

DML is used to directly learn a metric that measures similarity by reflecting label information (Weinberger and Saul 2009; Guo et al. 2023; Zhao and Yang 2023). Specifically, it learns a transformation matrix from Mahalanobis distance to a metric space where data of the same class become more similar and different classes become more disjoint (Weinberger and Saul 2009). This approach has been extended to MLL problems and has helped improve the performance of similarity-based multi-label classification (Sun and Zhang 2021; Mao, Wang, and Zhang 2023). In addition, DML-based approaches have been proposed in the MIML framework. However, despite the complex structure of the data, existing approaches use a single metric for multi-label classification (Jin, Wang, and Zhou 2009; Hu et al. 2019). Because a single metric forces samples into a unified metric space, it does not fully include all label-specific characteristics and has limitations in MIML problems.

To address the limitations of single metric learning in the

*Corresponding Author

MIML setting, we propose the T-MDML: triplet-based multiple distance metric learning for MIML. T-MDML extracts triplet sets to learn metrics that transform data into a metric space for each label. Based on the extracted triplets, an objective function is designed to maximize the margin between high- and low-similarity bags. Furthermore, based on the smoothness assumption, we incorporate the extracted label correlation based on the feature and label spaces into the objective function. Finally, we apply the proposed method to the MIMLkNN to verify the performance improvement and extend it to NN-based MIML algorithms to determine its effectiveness. The main contributions are as follows:

- In the MIML environment, we propose T-MDML to accurately measure label-based similarity and to improve the performance of distance-based MIML classification.
- We define a method that extracts the correlation between label pairs by extending the structural information from the feature space to the label space. Further, we verify the improvement of the classification performance by applying it to the distance metric optimization process.
- We verify the effectiveness of the proposed method through extensive experiments on 10 datasets with different structures. We also apply T-MDML to other distance-based MIML algorithms to demonstrate its scalability.

Related Work

Multi-instance Multi-label Learning (MIML)

The MIML problem aims to predict bag-level labels for an unknown bag. It can generally be divided into two categories: degeneration methods and direct methods.

The degeneration method is used to solve a high-dimensional MIML problem by decomposing it into lower dimensions, such as multi-instance or multi-label. (Zhou et al. 2012) proposed two approaches to degenerate MIML problems into MIL and MLL, respectively. The authors proposed MIMLBoost, in which the original problem is decomposed into an MIL by using label decomposition, following which MIBoosting is applied (Xu and Frank 2004). Further, they proposed MIMLSVM by applying clustering to decompose the problem into an MLL and then applying MLSVM (Boutell et al. 2004). (Zafra and Gibaja 2023) defined an arithmetic, geometric, and min-max transformation metric for degenerating MIML into MLL and solved the problem through NN-based MLL algorithms such as MLkNN (Zhang and Zhou 2007b), IBLRML (Cheng and Hüllermeier 2009), and MLDGC (Reyes, Morell, and Ventura 2016).

The direct method complements the limitations of the degeneration method, which may cause information loss, and solves the problem by designing a model based on MIML data or directly measuring the similarity between bags. (Li et al. 2021) proposed a KISAR based on an alternating optimization solution using convex optimization, based on the fact that highly related labels share some patterns. (Zhang 2010) proposed an MIMLkNN, and defined the concept of citer, which considers itself as a neighbor as well as the nearest neighbor. Then, it uses both label information to predict test data. In addition, (Zafra and Gibaja 2023) defined three distance metrics for measuring similarity between bags and proposed NN-based MIML algorithms such as MIMLBRkNN, MIMLIBLR, and MIMLDGC.

Distance Metric Learning (DML)

The DML method directly learns the Mahalanobis distance so that the distance can be measured by considering label information. In SISL, various DML algorithms have been proposed based on the margin maximization framework (Weinberger and Saul 2009; Nguyen et al. 2019; Wu et al. 2020; Duan et al. 2020). However, in multi-label environments, because multiple labels are assigned simultaneously, only some labels of neighboring instances may be different, making existing approaches inapplicable. To address this problem and reflect the characteristics of multiple labels, various algorithms have been proposed. The LM-kNN proposed by (Liu and Tsang 2015) projects the input and output into the same embedding space to integrate feature and label correlation. Then, a large margin formulation for distance metric learning from embedding spaces is proposed. Commu (Sun and Zhang 2021) performs compositional metric learning by modeling the structural interaction between the instance and label spaces to extract the integrated meaning. LIMIC (Mao, Wang, and Zhang 2023) considers that it is difficult to capture all semantic meanings in complex multi-label structures with a single metric, and it learns label-specific metrics. Studies have also applied DML to similarity-based MIML algorithms. (Jin, Wang, and Zhou 2009; Hu et al. 2019) reflected the properties of MIML data by constructing a label bag to represent each label and learning a metric to increase the similarity between the label bags and the related bags.

However, despite the large number of similarity-based MIML algorithms, few studies have investigated DML to improve the performance. In addition, existing DML methods for MIML learn a single metric even in complex datasets. This makes it difficult to reflect the abundant properties of MIML data and may lead to inadequate performance. Therefore, we propose a novel approach to extract label correlations by considering the characteristics of MIML data and learn a distance metric specialized for each label.

Proposed Method

We propose the T-MDML for application to MIML data. It consists of two steps: label correlation estimation and triplet-based multiple distance metric learning. Figure 1 shows an overview of T-MDML. In the first step, the label correlation is extracted for incorporation into the optimization process. This differs from statistical-based methods because it uses information from both the feature and label spaces. Second, we propose a distance metric learning method that can be applied to MIML data. Unlike existing single-metric learning that does not reflect label-specific properties, we propose a triplet-based distance metric that can be transformed for each label. Finally, we perform a classification for each label based on the trained metrics.

Preliminaries

Notations Formally, let $\mathcal{X}$ and $\mathcal{Y} = \{\lambda_1, \lambda_2, \cdots, \lambda_q\}$ denote d-dimensional instance space and label space. We de-

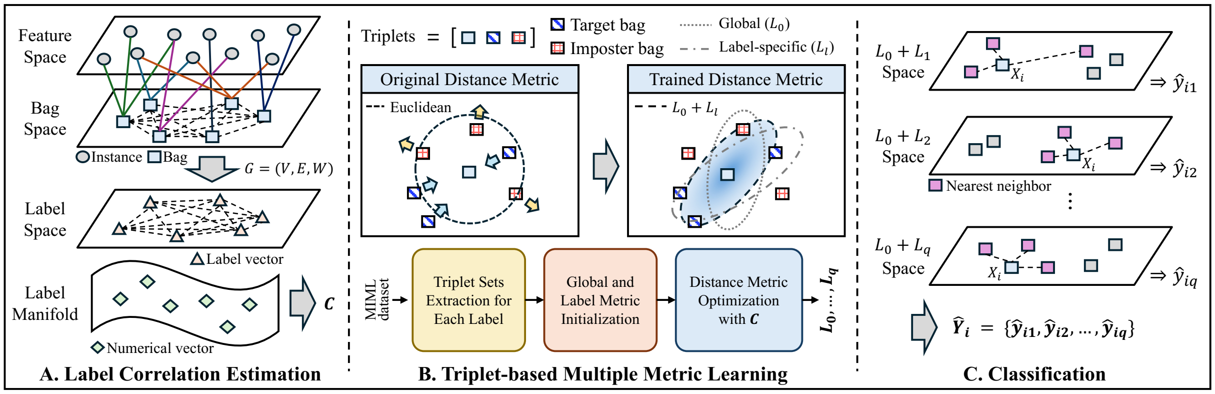


Figure 1: Overview of proposed T-MDML.

note the MIML dataset as $D = \{(X_i, Y_i), 1 \le i \le m\}$. $X_i = \{x_{i1}, x_{i2}, \cdots, x_{in_i}\} \subseteq \mathcal{X}$ represents a bag, consisting of n_i instances of $x_{ij} \in \mathbb{R}^d$, which is a vector of d dimensions. $Y_i = \{y_{i1}, y_{i2}, \cdots, y_{iq}\} \subseteq \mathcal{Y}$ denotes a set of labels, where $y_{ik} \in \{-1, +1\}$ $(k = 1, \cdots, q)$. q is the number of candidate labels, and $y_{ik} = +1$ means that bag X_i has the k-th label, otherwise, X_i does not have the k-th label. The aim of the MIML framework is to learn a function $f : 2^{\mathcal{X}} \to 2^{\mathcal{Y}}$.

Distance Metric Learning The Mahalanobis distance between two instances x_i and x_j is $(x_i - x_j)^T M (x_i - x_j) = Tr(MU_{ij})$, where $M \in \mathbb{R}^{d \times d}$ and $U_{ij} = (x_i - x_j)(x_i - x_j)^T$. Here, M is a positive semi-definite matrix, it can be decomposed into LL^T, where $L \in \mathbb{R}^{d \times d}$:

$$\begin{aligned} d_M^2(x_i, x_j) &= (x_i - x_j)^T M (x_i - x_j) \Leftrightarrow d_L^2(x_i, x_j) \\ &= (x_i - x_j)^T LL^T (x_i - x_j) = \left\| L^T (x_i - x_j) \right\|_2^2, \end{aligned} \tag{1}$$

This is equivalent to measuring the Euclidean distance in a new space mapped by the transformation matrix L. Furthermore, by decomposing the original PSD matrix M into LL^T, the optimization procedure becomes simpler as the PSD constraints are eliminated. Therefore, the proposed metric learning method learns the transformation matrix L.

Label Correlation Estimation

Many studies have investigated label correlation in MLL and MIML because it has an important impact on performance improvement (Huang and Zhou 2012; Yang, Tang, and Min 2022). For example, probability-based methods in the label space are commonly used because they are simple but efficient (Su, Yan, and Yu 2021; Mao, Wang, and Zhang 2023). In this approach, the correlation is estimated based on the conditional probability $P(\lambda_i | \lambda_j)$ for each label pair. Another approach learns the correlation directly based on the label matrix (Hu et al. 2019). However, all of these approaches only use information in the label space. Therefore, noise problems or inconsistent label correlation patterns can disrupt the learning process and degrade the performance. To solve these, the T-MDML computes the label correlation by considering both the feature and label spaces information.

Based on the smoothness assumption (Yang, Tang, and Min 2022) that if the values in the feature space are similar, the assigned label values are also similar, we construct an NN graph $G = (V, E, W)$ to exploit the inherent information in the bag space. Let V be the vertex set consisting of each bag X_i and E be the edge set consisting of the relation e_{ij}, which represents the relation between bags X_i and X_j. W is a weight matrix, where each element w_{ij} is the weight of e_{ij}. It is computed based on $\mathcal{K}_i$, as follows:

$$w_{ij} = \begin{cases} 1/D_M^2(X_i, X_j), & if \quad X_j \in \mathcal{K}_i \\ 0, & Otherwise. \end{cases} \tag{2}$$

where, $\mathcal{K}_i$ denotes the nearest neighbor set of bag X_i. To measure the distance between two bags, the average Hausdorff distance (Zhang and Zhang 2006; Zafra and Gibaja 2023) is used, which is defined as follows:

$$\begin{aligned} D_M^2(X_i, X_j) = &\frac{\sum_{x_{ik} \in X_i} min_{x_{jl} \in X_j} d_M^2(x_{ik}, x_{jl})}{n_i + n_j} \\ &+ \frac{\sum_{x_{jl} \in X_j} min_{x_{ik} \in X_i} d_M^2(x_{jl}, x_{ik})}{n_i + n_j}, \end{aligned} \tag{3}$$

where $d_M(\cdot)$ is the Euclidean distance; therefore $M = I$. The NN graph implies the structural properties of the bag space, and we can extend this information to the label space. Therefore, based on the local relationship in the bag space, we compute a new numerical label vector $S_i = \{s_{i1}, \cdots, s_{iq}\} \in \mathbb{R}^q$ for each bag from the label space:

$$s_{ij} = \sum_{X_k \in \mathcal{K}_i} \frac{exp(w_{ik}) \times y_{kj}}{\sum_{X_k \in \mathcal{K}_i} exp(w_{ik})} + y_{ij}, \tag{4}$$

Next, we apply the softmax function to S_i to transform it to a probability value and form the label manifold $\mathcal{S}$:

$$s_{ij} = \frac{exp(s_{ij})}{\sum_{k=1}^{q} exp(s_{ik})}, \tag{5}$$

where $\forall S_i \in \mathcal{S}$. Because the label manifold reflects the local structure of the bag space, it represents a smooth change between neighboring elements. Finally, we derive the label correlation matrix $C \in R^{q \times q}$ based on the matrix $S = [S_i, \cdots, S_m]^T \in R^{m \times q}$, which consists of each numerical vector S_i in the label manifold:

$$c_{ij} = \frac{P_i \cdot P_j}{\|P_i\| \, \|P_j\|}, where \, 1 \le i, j \le q, \tag{6}$$

where P_i is the i-th column vector of S. c_{ij} is an element of C that denotes the correlation value between λ_i and λ_j.

Dataset	Labels	Bags	Inst.	Dim.	Card.	Domain
scene	5	2000	18000	15	1.24	Image
MSRC_v2	23	591	1758	48	2.51	Image
news	10	612	24406	50	1.45	Text
birds	13	548	10232	38	2.14	Audio
reuters	7	2000	7119	243	1.15	Text
H. marismortui	234	304	950	216	3.25	Protein
P. furiosus	321	425	1317	216	4.48	Protein
A. vinelandii	340	407	1251	216	4.00	Protein
C. elegans	940	2512	8509	216	6.07	Protein
D. melanogaster	1035	2605	9146	216	6.02	Protein

Table 1: Details of 10 datasets used for evaluation.

Triplet-based Multiple Distance Metric Learning

Unlike SISL and MLL, we cannot perform DML based on Eq. (1) for MIML because the data is structured as a bag. Therefore, we use Eq. (7) to measure the distance between bags, and we learn the transformation matrix L:

$$Dis_L^2(X_i, X_j) = \min_{x_{ik} \in X_i, x_{jl} \in X_j} d_L^2(x_{ik}, x_{jl}) \quad (7)$$

MIML data have a complex label structure that is difficult to transform into metric space through single metric learning. Therefore, we propose a triplet-based optimization method to learn individual transformation matrices $L_1, \cdots, L_q$ that reflect the properties of each label. Furthermore, because independent learning for each label may miss the global properties, we simultaneously learn a global metric matrix L_0 that incorporates the properties of all labels. Finally, we combine L_0, L_l to derive the transformation matrix L_l of λ_l.

To learn the transformation matrix L_l for the l-th label, we form a triplet set $\mathcal{T}_l$. A triplet consists of a target bag X_j with the same assignment for λ_l as bag X_i and an imposter bag X_k with a different assignment, and it is defined as follows:

$$\mathcal{T}_l = \{(X_i, X_j, X_k) | X_j \in T_i, X_k \in I_i\} \quad (8)$$

where T_i represents the set of nearest target neighbors of X_i, and I_i represents the set of nearest imposter neighbors. It also satisfies $|T_i| = k_1, |I_i| = k_2, |\mathcal{T}_l| = m \times k_1 \times k_2, y_{il} = y_{jl} \neq y_{kl}$, and $1 \leq i \leq m$. Although triplets can be constructed for all possible combinations of each training bag, in practice, optimizing for all triplet combinations significantly increases the computational complexity. Therefore, we construct a nearest neighbor set of k_1, k_2 (Weinberger, Blitzer, and Saul 2005; Wang, Kalousis, and Woznica 2012).

To learn a transformation L_l such that the distance from bag X_i to target X_j is closer than that to imposter X_k, we define a margin ξ_{ijk}^l, that is the difference in distances:

$$\xi_{ijk}^l = Dis_{L_0+L_l}^2(X_i, X_k) - Dis_{L_0+L_l}^2(X_i, X_j), \quad (9)$$

We learn the transformation matrix $L_1, \cdots, L_q$ to maximize the margin over all elements of the constructed triplet set for each label. This allows us to learn the individual properties of each label space. Also, we learn a global transformation matrix L_0 that reflects the common properties of all labels and update L_l based on L_0 through the optimization process:

$$\min_{L_0, \cdots, L_q} \sum_{l=1}^{q} \frac{1}{|\mathcal{T}_l|} \sum_{(X_i, X_j, X_k) \in \mathcal{T}_l} \mathcal{L}(-\xi_{ijk}^l) + \mu \sum_{l=0}^{q} \|L_l\|_F^2 \quad (10)$$

Here, $\mathcal{L}$ is the loss function; it is defined as $\frac{e^{2x}}{e^{2x}+1}$. The first term represents the sum of the loss for the margins of all elements in the triplet set. As the distance to the target becomes smaller than the distance to the imposter, the loss converges to zero. In the opposite case, the loss will increase. We designed the loss function such that the loss becomes larger as it approaches zero, even if the margin is positive, so that the target and imposter are far enough apart. We also adjusted the variable to $2x$ to make the loss more sensitive. The optimization process is performed for each of the q labels. The second is a regularization term and μ is a weight parameter.

If the transformation matrix for each label is learned independently, it cannot reflect the label correlation. Therefore, T-MDML adjusts the similarity between the transformation matrices based on the calculated label correlation. The objective function with label correlation is as follows:

$$\begin{aligned} \min_{L_0, \cdots, L_q} & \sum_{l=1}^{q} \frac{1}{|\mathcal{T}_l|} \sum_{(X_i, X_j, X_k) \in \mathcal{T}_l} \mathcal{L}(-\xi_{ijk}^l) \\ & + \mu_1 \sum_{l=0}^{q} \|L_l\|_F^2 + \mu_2 \sum_{i=1}^{q} \sum_{j=1}^{q} c_{ij} \|L_i - L_j\|_F^2, \end{aligned} \quad (11)$$

The added third term incorporates the calculated label correlation matrix into the optimization process. The higher the correlation between two labels λ_i, λ_j, the more the transformation matrices L_i, L_j are adjusted to be similar; conversely, the lower the correlation, the more different L_i and L_j are. Finally, μ_1, μ_2 are non-negative weight values that balance the regularization term and correlation term.

Optimization Strategy

We denote the Eq. (11) as $\mathcal{O}$. To optimize $\mathcal{O}$, we use the gradient-based optimization, which is commonly used in DML (De Vazelhes et al. 2020; Mao, Wang, and Zhang 2023). The global matrix L_0, which is the basis for learning each label metric, is set to the identity matrix I to initialize to the Euclidean distance as in existing methods. Because $L_1, \cdots, L_q$, which learns local properties for each label, is optimized based on L_0, it is set to the zero matrix. These are updated based on L_0 as the optimization process progresses.

Specifically, we optimize the $L_0, \cdots, L_q$ by using the gradient descent (Mao, Wang, and Zhang 2023), where the derivatives of the $\mathcal{O}$ with respect to L_0, L_l are as follows:

$$\begin{aligned} \frac{\partial \mathcal{O}}{\partial L_0} = & \sum_{l=1}^{q} \frac{-2}{|\mathcal{T}_l|} \sum_{(X_i, X_j, X_k) \in \mathcal{T}_l} \theta_{ijk}^l (U_{ij} - U_{ik})(L_0 + L_l) \\ & + 2\mu_1 L_0 \end{aligned} \quad (12)$$

$$\begin{aligned} \frac{\partial \mathcal{O}}{\partial L_l} = & \frac{-2}{|\mathcal{T}_l|} \sum_{(X_i, X_j, X_k) \in \mathcal{T}_l} \theta_{ijk}^l (U_{ij} - U_{ik})(L_0 + L_l) \\ & + 2\mu_1 L_l + 4\mu_2 \sum_{i=1}^{q} c_{li}(L_l - L_i) \end{aligned} \quad (13)$$

where U_{ij} is derived from Eq. (7) based on the instance pair measured by the minimum distance between the two bags. θ_{ijk}^l is the derivative of the $\mathcal{L}$ and is computed as follows:

$$\theta_{ijk}^l = \frac{2e^{-2\xi_{ijk}^l}}{(e^{-2\xi_{ijk}^l} + 1)^2} \quad (14)$$

	***Hamming Loss* ↓**		***One Error* ↓**		***Coverage* ↓**		***Average Precision* ↑**	
	Label Correlation	Non-Label Correlation	Label Correlation	Non-Label Correlation	Label Correlation	Non-Label Correlation	Label Correlation	Non-Label Correlation
scene	0.1823 $_{0.0096}$	0.1823 $_{0.0083}$	0.3490 $_{0.0296}$	0.3490 $_{0.0277}$	**1.0485** $_{0.0327}$	1.0495 $_{0.0419}$	**0.7704** $_{0.0176}$	0.7703 $_{0.0179}$
MSRC_v2	**0.1510** $_{0.0099}$	0.1511 $_{0.0124}$	0.4705 $_{0.0389}$	0.4705 $_{0.0615}$	**7.8133** $_{0.5213}$	7.9776 $_{0.3831}$	0.5653 $_{0.0254}$	0.5653 $_{0.0309}$
news	0.1802 $_{0.0347}$	0.1802 $_{0.0375}$	0.3807 $_{0.0598}$	0.3807 $_{0.0609}$	2.2893 $_{0.2087}$	2.2893 $_{0.2460}$	0.7121 $_{0.0371}$	0.7121 $_{0.0403}$
birds	0.0910 $_{0.0100}$	0.0910 $_{0.0126}$	**0.1626** $_{0.0437}$	0.1644 $_{0.0459}$	**2.6767** $_{0.3304}$	2.6803 $_{0.3053}$	**0.8455** $_{0.0221}$	0.8454 $_{0.0258}$
reuters	**0.0579** $_{0.0060}$	0.0581 $_{0.0041}$	0.1290 $_{0.0118}$	0.1290 $_{0.0154}$	0.5685 $_{0.0521}$	0.5685 $_{0.0461}$	0.9098 $_{0.0072}$	0.9098 $_{0.0075}$
H. marismortui	**0.0921** $_{0.0173}$	0.1000 $_{0.0218}$	**0.6219** $_{0.0432}$	0.6263 $_{0.0478}$	**23.5107** $_{4.9117}$	23.8970 $_{4.9172}$	**0.4146** $_{0.0377}$	0.4087 $_{0.0457}$
P. furiosus	**0.0816** $_{0.0179}$	0.0881 $_{0.0170}$	**0.6250** $_{0.0414}$	0.6383 $_{0.0541}$	**41.2518** $_{4.0562}$	42.5882 $_{4.6997}$	**0.3932** $_{0.0356}$	0.3817 $_{0.0231}$
A. vinelandii	**0.1208** $_{0.0243}$	0.1219 $_{0.0293}$	**0.6209** $_{0.0273}$	0.6240 $_{0.0183}$	**40.5078** $_{2.2633}$	41.0950 $_{3.0755}$	**0.3622** $_{0.0247}$	0.3597 $_{0.0225}$
C. elegans	**0.0374** $_{0.0059}$	0.0386 $_{0.0052}$	**0.5507** $_{0.0247}$	0.5555 $_{0.0241}$	96.1684 $_{3.0567}$	96.1684 $_{3.0567}$	**0.4470** $_{0.0139}$	0.4457 $_{0.0075}$
D. melanogaster	**0.0370** $_{0.0077}$	0.0378 $_{0.0078}$	**0.5441** $_{0.0176}$	0.5463 $_{0.0148}$	**113.0491** $_{5.4233}$	113.4860 $_{3.9413}$	**0.4416** $_{0.0137}$	0.4402 $_{0.0165}$

Table 2: Compared performance ($mean_{std}$) with and without proposed label correlation estimation method.

Consequently, the overall process of T-MDML is as follows. First, we extract structural features from the bag space and extend them to the label space to compute the label correlation. Next, we extract the triplet sets $\mathcal{T}_l$ for optimizing the distance metric for each label. After initializing $L_0, \cdots, L_q$, we optimize iteratively using Eqs. (11), (12) and (13).

The trained distance metric can be applied to distance-based algorithms. The label prediction for unknown data is performed independently for each label based on the learned transformation matrix $L_0, \cdots, L_q$. For example, the l-th label is determined based on the similarity between the bags measured by $Dis^2_{L_0+L_l}(\cdot)$ in Eq. (7). The step C of Figure 1 shows an overview of classification procedure.

Experiments

Experimental Setting

Datasets For evaluation, we use 10 datasets. Table 1 describes the dataset. Because MIML datasets have a complex structure, we can consider various properties. Inst, Dim and Card represent the total number of instances, size of the dimension, and cardinality, where cardinality is the average number of labels assigned per bag.

Evaluation Metrics To evaluate the performance of our experiments, we used five evaluation metrics that are commonly used in multi-label learning: *Hamming Loss, Ranking Loss, One Error, Coverage,* and *Average Precision*. More detailed descriptions about datasets and evaluation metrics can be found in (Zafra and Gibaja 2023).

Implementation Details Theoretically, T-MDML is applicable to similarity-based MIML methods. In our experiments, we applied the learned metric to MIMLkNN that uses both the NN and citer for prediction. Therefore, it is dependent on the distance metric, and the impact of the T-MDML can be distinctively verified. In addition, all classifiers were evaluated by applying five-fold cross-validation. The hyperparameters of the algorithm are set to the best values from the results in the Parameter Sensitivity section.

Results and Analysis

Label Correlation Analysis First, we analyzed the impact of our label correlation estimation method on performance. In Table 2, the label correlation and non-label correlation represent the results of optimizing with and without applying the label correlation. Each value represents the mean of the cross-validation. The value in the lower right of each result indicates the standard deviation. Further, ↑ and ↓ mean that the higher or lower the value, the better is the performance. Results in bold indicate better performance than the comparisons. We can confirm that considering the label correlation results in at least the same performance as that in the opposite case. We also found that the performance improves on all datasets except news. Notably, for datasets with a large number of labels, such as protein data, the performance was improved on most metrics. T-MDML efficiently extracted correlations in the complex label space based on the structural information in the bag space; therefore, data with a large number of labels showed a higher performance improvement. We also demonstrated the effectiveness of an objective function designed to ensure that pairs of related labels perform similar transformations, compared to a method that learns the metric independently for each label.

Comparative Analysis with MIMLkNN To evaluate the performance improvement of MIMLkNN with T-MDML, we performed a comparative analysis with the MIMLkNN. Table 3 shows that the performance is improved compared to that of the MIMLkNN for almost all datasets for all metrics. In particular, the algorithm with T-MDML outperform all datasets in terms of the *Hamming Loss* and *Average Precision*. These results indicate the importance of using the label information to measure the similarity in distance-based algorithms. When T-MDML is applied, the performance improvement is more significant for datasets with a large number of labels. Compared to existing methods that do not extract information from the label space, the T-MDML learns the properties of each label and considers the label correlation, thereby enabling more accurate multi-label prediction.

Comparative Analysis with Other MIML Algorithms To evaluate the relative performance, we performed a comparative analysis with the following MIML algorithms:

- **MIMLkNN** (Zhang 2010): A kNN-based algorithm that considers the labels of its neighbors and the citers.
- **MIMLBRkNN** (Zafra and Gibaja 2023): A BRkNN-

	Hamming Loss ↓		*One Error* ↓		*Coverage* ↓		*Average Precision* ↑	
	T-MDML +MIMLkNN	MIMLkNN	T-MDML +MIMLkNN	MIMLkNN	T-MDML +MIMLkNN	MIMLkNN	T-MDML +MIMLkNN	MIMLkNN
scene	**0.1823** $_{0.0096}$	0.1827 $_{0.0096}$	**0.3490** $_{0.0296}$	0.3495 $_{0.0296}$	**1.0495** $_{0.0334}$	1.0535 $_{0.0322}$	**0.7703** $_{0.0176}$	0.7699 $_{0.0174}$
MSRC_v2	**0.1510** $_{0.0099}$	0.1647 $_{0.0069}$	**0.4705** $_{0.0389}$	0.4823 $_{0.0363}$	**7.8369** $_{0.5249}$	7.8422 $_{0.2615}$	**0.5653** $_{0.0254}$	0.5517 $_{0.0139}$
news	**0.1802** $_{0.0347}$	0.1922 $_{0.0343}$	**0.3807** $_{0.0598}$	0.3971 $_{0.0598}$	2.2828 $_{0.2065}$	**2.2579** $_{0.2051}$	**0.7124** $_{0.0370}$	0.7040 $_{0.0368}$
birds	**0.0910** $_{0.0100}$	0.0942 $_{0.0104}$	0.1626 $_{0.0437}$	**0.1607** $_{0.0465}$	**2.6858** $_{0.3296}$	2.7681 $_{0.3195}$	**0.8455** $_{0.0221}$	0.8444 $_{0.0226}$
reuters	**0.0579** $_{0.0060}$	0.0627 $_{0.0071}$	**0.1290** $_{0.0118}$	0.1315 $_{0.0113}$	0.5700 $_{0.0537}$	**0.5685** $_{0.0567}$	**0.9098** $_{0.0072}$	0.9088 $_{0.0070}$
H. marismortui	**0.0921** $_{0.0173}$	0.1099 $_{0.0153}$	**0.6219** $_{0.0432}$	0.6340 $_{0.0493}$	**23.5107** $_{4.9117}$	25.3334 $_{4.9499}$	**0.4146** $_{0.0377}$	0.4095 $_{0.0337}$
P. furiosus	**0.0816** $_{0.0179}$	0.0915 $_{0.0179}$	**0.6250** $_{0.0414}$	0.6517 $_{0.0400}$	**41.4400** $_{4.0465}$	44.4024 $_{3.6139}$	**0.3932** $_{0.0356}$	0.3674 $_{0.0345}$
A. vinelandii	**0.1208** $_{0.0243}$	0.1322 $_{0.0239}$	**0.6240** $_{0.0340}$	0.6604 $_{0.0386}$	**40.5395** $_{2.2963}$	42.4599 $_{4.4778}$	**0.3622** $_{0.0247}$	0.3355 $_{0.0287}$
C. elegans	**0.0374** $_{0.0059}$	0.0394 $_{0.0058}$	**0.5507** $_{0.0247}$	0.5701 $_{0.0246}$	**97.2525** $_{3.1033}$	103.9441 $_{3.1088}$	**0.4470** $_{0.0139}$	0.4409 $_{0.0139}$
D. melanogaster	**0.0378** $_{0.0086}$	0.0426 $_{0.0110}$	**0.5433** $_{0.0190}$	0.5646 $_{0.0174}$	**113.2046** $_{5.4493}$	124.9163 $_{3.3663}$	**0.4422** $_{0.0152}$	0.4314 $_{0.0122}$

Table 3: Performance ($mean_{std}$) comparison with MIMLkNN.

based algorithm that applies kNN for each label.

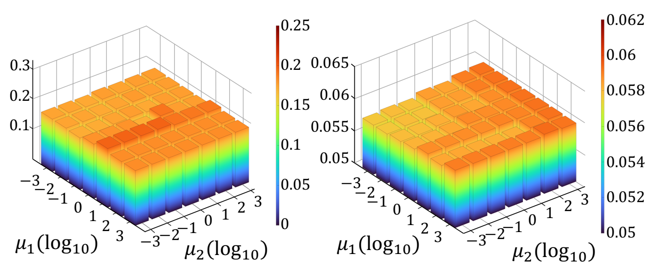
- **MIMLIBLR** (Zafra and Gibaja 2023): An IBLRML-based method that combines instance-based learning and logistic regression.
- **MIMLDGC** (Zafra and Gibaja 2023): An MLDGC-based method that considers neighbor labels and NGC calculated with neighbor density and weights.
- **MIMLBoost** (Zhou et al. 2012): An algorithm that decomposes the label space and applies the MIBoosting.
- **MIMLSVM** (Zhou et al. 2012): An algorithm that applies MLSVM after performing clustering degeneration.
- **KISAR** (Li et al. 2021): An algorithm that exploits the patterns between instances and their relevant labels.

The parameters of the comparison algorithms are set to the values in the released software. In the proposed method, the number of elements of the nearest neighbor set k_1, k_2 is set to 5. Table 4 shows the experimental results for the three metrics. AVG RANK indicates the average of each algorithm's ranking for each dataset, and RANK indicates the rank of the AVG RANK. MIMLkNN with T-MDML shows the best performance in terms of the *Ranking Loss, Coverage, Average Precision*. MIMLSVM shows relatively good performance on complex datasets with a large number of labels; however, it shows lower performance on relatively simple datasets. In other words, the classification results are not consistent across all datasets. T-MDML considers the label correlation and learns a metric for each label; this can improve the prediction accuracy of the overall labels. Therefore, it has stable and top-ranked performance on these metrics regardless of the complexity of the data. These results indicate that the application of T-MDML can stably improve the classification performance of comprehensive labels.

Expansion to Other MIML Algorithms Theoretically, T-MDML can be applied to any distance-based MIML algorithm. Therefore, we applied T-MDML to other distance-based algorithms and analyzed the results. Specifically, we analyzed the results for *Coverage, Average Precision* to verify the accuracy of the overall predicted label, as shown in Table 5. The results in bold represent improved performance over each of the original algorithms in Table 4. Similar to the results for MIMLkNN, T-MDML improves the overall accuracy of the predicted labels by learning individual label properties. Notably, the results show a significant performance improvement on datasets with a large number of labels. These results indicate that learning metrics that consider label-specific properties with correlations and applying them to distance-based algorithms can effectively improve the classification performance with MIML datasets.

Further Analysis

Parameter Sensitivity In Eq. (11), μ_1, μ_2 are hyperparameters representing the weights of the normalization and correlation term. Figure 2 shows how the *Hamming Loss* of T-MDML changes depending on the settings of the μ_1, μ_2 for scene and reuters. We set the values of $\mu_1, \mu_2 \in \{10^{-3}, 10^{-2}, 10^{-1}, 10^0, 10^1, 10^2, 10^3\}$. As a result, the performance of T-MDML varies relatively stably, even though the values of μ_1, μ_2 are set in a wide range. From this result, we can verify the stability of T-MDML.



(a) scene (b) reuters

Figure 2: Performance of T-MDML with varying parameter configurations in terms of *Hamming Loss*.

Conclusion

In this study, we propose the T-MDML to improve the performance of similarity-based MIML. We extract the label correlation by exploiting the structural information of the bag and label spaces. Also, we learn the distance metric of each label and the global metric by maximizing the triplet-based margin. The experimental results show that for most datasets, the algorithm with the T-MDML outperforms the existing methods and demonstrate its scalability.

	T-MDML +MIMLkNN	MIMLkNN	MIMLBRkNN	MIMLIBLR	MIMLDGC	KISAR	MIMLSVM	MIMLBoost
				***Ranking Loss* ↓**				
scene	0.1838 $_{0.0146}$	0.1960 $_{0.0151}$	0.2307 $_{0.0136}$	0.2062 $_{0.0127}$	0.2224 $_{0.0132}$	0.1677 $_{0.0082}$	0.4685 $_{0.0228}$	0.9957 $_{0.0016}$
MSRC_v2	0.1750 $_{0.0123}$	0.1772 $_{0.0148}$	0.1529 $_{0.0268}$	0.1028 $_{0.0158}$	0.1497 $_{0.0267}$	0.5023 $_{0.0194}$	0.1010 $_{0.0152}$	0.4019 $_{0.0229}$
news	0.1787 $_{0.0362}$	0.1774 $_{0.0315}$	0.2541 $_{0.0417}$	0.1840 $_{0.0213}$	0.3883 $_{0.0317}$	0.2535 $_{0.0247}$	0.2785 $_{0.0367}$	0.3406 $_{0.0134}$
birds	0.0886 $_{0.0174}$	0.0913 $_{0.0180}$	0.0670 $_{0.0108}$	0.0291 $_{0.0073}$	0.0432 $_{0.0114}$	0.1106 $_{0.0375}$	0.3249 $_{0.0194}$	0.8604 $_{0.0243}$
reuters	0.0547 $_{0.0036}$	0.0556 $_{0.0034}$	0.1465 $_{0.0092}$	0.1256 $_{0.0122}$	0.1456 $_{0.0087}$	0.0231 $_{0.0055}$	0.2402 $_{0.0159}$	0.8716 $_{0.0129}$
H. marismortui	0.2619 $_{0.0317}$	0.2641 $_{0.0563}$	0.3976 $_{0.0267}$	0.3718 $_{0.0324}$	0.3964 $_{0.0269}$	0.8672 $_{0.0189}$	0.2654 $_{0.0321}$	0.8928 $_{0.0106}$
P. furiosus	0.2703 $_{0.0354}$	0.2991 $_{0.0357}$	0.3781 $_{0.0178}$	0.3902 $_{0.0315}$	0.3745 $_{0.0192}$	0.8680 $_{0.0183}$	0.2124 $_{0.0162}$	0.7436 $_{0.0440}$
A. vinelandii	0.3247 $_{0.0293}$	0.3453 $_{0.0276}$	0.4425 $_{0.0340}$	0.4217 $_{0.0404}$	0.4417 $_{0.0336}$	0.9088 $_{0.0106}$	0.2295 $_{0.0229}$	0.7660 $_{0.0561}$
C. elegans	0.1623 $_{0.0074}$	0.1719 $_{0.0073}$	0.2861 $_{0.0126}$	0.2741 $_{0.0098}$	0.2855 $_{0.0127}$	0.7321 $_{0.0145}$	0.1127 $_{0.0074}$	-
D. melanogaster	0.1619 $_{0.0066}$	0.1804 $_{0.0066}$	0.3061 $_{0.0161}$	0.3065 $_{0.0118}$	0.3057 $_{0.0162}$	0.7714 $_{0.0128}$	0.1144 $_{0.0035}$	-
AVG RANK	2.40	3.20	5.20	3.80	4.60	5.70	3.50	7.50
RANK	**1**	2	6	4	5	7	3	8
				***Coverage* ↓**				
scene	1.001 $_{0.041}$	1.053 $_{0.039}$	1.183 $_{0.022}$	1.097 $_{0.034}$	1.151 $_{0.021}$	0.946 $_{0.029}$	2.096 $_{0.022}$	2.112 $_{0.043}$
MSRC_v2	7.519 $_{0.383}$	7.842 $_{0.411}$	7.168 $_{0.937}$	5.229 $_{0.586}$	7.058 $_{0.951}$	12.633 $_{0.656}$	5.140 $_{0.549}$	11.023 $_{0.452}$
news	2.282 $_{0.248}$	2.257 $_{0.220}$	2.971 $_{0.307}$	2.294 $_{0.167}$	4.237 $_{0.205}$	2.894 $_{0.225}$	3.235 $_{0.402}$	3.702 $_{0.177}$
birds	2.670 $_{0.308}$	2.673 $_{0.317}$	2.199 $_{0.320}$	1.752 $_{0.211}$	2.181 $_{0.330}$	3.04 $_{0.584}$	5.938 $_{0.468}$	6.565 $_{0.386}$
reuters	0.509 $_{0.044}$	0.514 $_{0.043}$	1.065 $_{0.064}$	0.930 $_{0.062}$	1.059 $_{0.057}$	0.301 $_{0.024}$	1.584 $_{0.081}$	3.957 $_{0.106}$
H. marismortui	24.946 $_{4.911}$	25.126 $_{6.192}$	130.010 $_{9.919}$	128.000 $_{10.006}$	129.797 $_{10.010}$	180.531 $_{2.293}$	25.511 $_{3.932}$	45.717 $_{4.830}$
P. furiosus	41.440 $_{4.582}$	44.402 $_{4.452}$	191.635 $_{3.639}$	198.637 $_{10.228}$	190.687 $_{3.874}$	260.174 $_{6.154}$	34.696 $_{2.815}$	57.472 $_{5.290}$
A. vinelandii	40.539 $_{3.075}$	42.459 $_{2.801}$	204.389 $_{13.256}$	211.170 $_{12.302}$	204.139 $_{13.260}$	284.090 $_{4.315}$	31.586 $_{4.372}$	61.894 $_{4.299}$
C. elegans	97.242 $_{4.241}$	103.944 $_{5.305}$	375.124 $_{10.173}$	387.629 $_{9.091}$	374.420 $_{10.409}$	681.710 $_{10.858}$	68.421 $_{1.063}$	-
D. melanogaster	113.204 $_{3.966}$	124.916 $_{3.121}$	456.629 $_{25.026}$	483.726 $_{19.124}$	456.049 $_{25.051}$	799.918 $_{11.973}$	80.465 $_{3.484}$	-
AVG RANK	2.40	3.20	5.30	4.50	4.70	5.80	3.50	6.25
RANK	**1**	2	6	4	5	7	3	8
				***Average Precision* ↑**				
scene	0.7800 $_{0.0179}$	0.7699 $_{0.0173}$	0.7396 $_{0.0186}$	0.7559 $_{0.0179}$	0.7526 $_{0.0161}$	0.7995 $_{0.0111}$	0.5190 $_{0.0108}$	0.5146 $_{0.0150}$
MSRC_v2	0.5749 $_{0.0309}$	0.5686 $_{0.0315}$	0.6742 $_{0.0454}$	0.7381 $_{0.0190}$	0.6822 $_{0.0434}$	0.4515 $_{0.0213}$	0.7025 $_{0.0254}$	0.3629 $_{0.0142}$
news	0.7124 $_{0.0404}$	0.7040 $_{0.0303}$	0.6250 $_{0.0472}$	0.6745 $_{0.0442}$	0.4864 $_{0.0215}$	0.6525 $_{0.0364}$	0.5915 $_{0.0309}$	0.4396 $_{0.0293}$
birds	0.8455 $_{0.0258}$	0.8444 $_{0.0244}$	0.9204 $_{0.0137}$	0.9351 $_{0.0194}$	0.9223 $_{0.0169}$	0.8223 $_{0.0566}$	0.4408 $_{0.0169}$	0.3466 $_{0.0161}$
reuters	0.9155 $_{0.0075}$	0.9088 $_{0.0066}$	0.8077 $_{0.0062}$	0.8254 $_{0.0170}$	0.8108 $_{0.0044}$	0.9528 $_{0.0090}$	0.6548 $_{0.0224}$	0.3679 $_{0.0121}$
H. marismortui	0.4061 $_{0.0432}$	0.4003 $_{0.0259}$	0.2009 $_{0.0235}$	0.1709 $_{0.0233}$	0.2368 $_{0.0339}$	0.0542 $_{0.0119}$	0.4614 $_{0.0322}$	0.0652 $_{0.0060}$
P. furiosus	0.3932 $_{0.0391}$	0.3674 $_{0.0414}$	0.2038 $_{0.0218}$	0.1555 $_{0.0282}$	0.2611 $_{0.0237}$	0.0580 $_{0.0167}$	0.4143 $_{0.0204}$	0.1348 $_{0.0088}$
A. vinelandii	0.3622 $_{0.0225}$	0.3355 $_{0.0232}$	0.1829 $_{0.0311}$	0.1318 $_{0.0299}$	0.2086 $_{0.0255}$	0.0223 $_{0.0047}$	0.4141 $_{0.0424}$	0.1080 $_{0.0164}$
C. elegans	0.4470 $_{0.0078}$	0.4409 $_{0.0058}$	0.4703 $_{0.0122}$	0.3387 $_{0.0101}$	0.5214 $_{0.0175}$	0.1559 $_{0.0150}$	0.4904 $_{0.0155}$	-
D. melanogaster	0.4422 $_{0.0157}$	0.4314 $_{0.0099}$	0.4394 $_{0.0202}$	0.2892 $_{0.0178}$	0.4835 $_{0.0216}$	0.1297 $_{0.0081}$	0.4577 $_{0.0133}$	-
AVG RANK	2.70	3.80	4.60	4.30	3.60	5.70	3.60	7.63
RANK	**1**	4	6	5	2	7	2	8

Table 4: Classification performance ($mean_{std}$) of each compared MIML algorithms. The '-' indicates that it did not end within one week.

	***Coverage* ↓**			***Average Precision* ↑**		
	T-MDML +MIMLBRkNN	T-MDML +MIMLIBLR	T-MDML +MIMLDGC	T-MDML +MIMLBRkNN	T-MDML +MIMLIBLR	T-MDML +MIMLDGC
scene	**1.1815** $_{0.0225}$	**1.0935** $_{0.0326}$	**1.1490** $_{0.0187}$	0.7394 $_{0.0185}$	**0.7562** $_{0.0177}$	0.7526 $_{0.0152}$
MSRC_v2	**6.9499** $_{0.7871}$	**5.0013** $_{0.7633}$	**6.8025** $_{0.6822}$	**0.6880** $_{0.0386}$	0.7323 $_{0.0365}$	**0.6971** $_{0.0375}$
news	**2.9567** $_{0.3032}$	**2.2732** $_{0.1421}$	**4.1800** $_{0.2053}$	0.6240 $_{0.0459}$	0.6740 $_{0.0432}$	**0.4968** $_{0.0278}$
birds	2.2988 $_{0.3695}$	**1.6758** $_{0.2602}$	2.2952 $_{0.3690}$	0.9095 $_{0.0192}$	0.9343 $_{0.0189}$	0.9107 $_{0.0207}$
reuters	**1.0625** $_{0.0637}$	**0.9060** $_{0.0590}$	1.0595 $_{0.0567}$	**0.8099** $_{0.0041}$	**0.8285** $_{0.0160}$	0.8106 $_{0.0045}$
H. marismortui	**31.8976** $_{5.3266}$	**20.5156** $_{5.3139}$	**31.7203** $_{5.2811}$	**0.3201** $_{0.0411}$	**0.3802** $_{0.0558}$	**0.3529** $_{0.0436}$
P. furiosus	**50.6776** $_{3.3448}$	**37.8635** $_{2.6394}$	**49.4941** $_{3.2614}$	**0.2994** $_{0.0188}$	**0.3023** $_{0.0220}$	**0.3757** $_{0.0224}$
A. vinelandii	**50.0296** $_{3.5960}$	**33.6849** $_{1.3570}$	**49.7836** $_{3.6585}$	**0.3106** $_{0.0304}$	**0.3015** $_{0.0277}$	**0.3400** $_{0.0322}$
C. elegans	**149.4584** $_{7.0089}$	**92.2812** $_{4.2609}$	**149.0085** $_{7.1971}$	**0.5187** $_{0.0134}$	**0.4690** $_{0.0118}$	**0.5617** $_{0.0168}$
D. melanogaster	**182.3551** $_{10.2690}$	**116.6376** $_{4.8035}$	**181.7582** $_{10.2814}$	**0.4695** $_{0.0179}$	**0.4240** $_{0.0130}$	**0.5180** $_{0.0178}$

Table 5: Classification performance ($mean_{std}$) of extending to other distance-based MIML algorithms.

Acknowledgments

This work was supported by Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (RS-2024-00400368, Development of Image Integrity Verification Technology for Forgery Prevention and Original Proof).

References

Amores, J. 2013. Multiple instance classification: Review, taxonomy and comparative study. *Artificial Intelligence*, 201: 81–105.

Boutell, M. R.; Luo, J.; Shen, X.; and Brown, C. M. 2004. Learning multi-label scene classification. *Pattern Recognition*, 37(9): 1757–1771.

Briggs, F.; Lakshminarayanan, B.; Neal, L.; Fern, X. Z.; Raich, R.; Hadley, S. J. K.; Hadley, A. S.; and Betts, M. G. 2012. Acoustic classification of multiple simultaneous bird species: A multi-instance multi-label approach. *The Journal of the Acoustical Society of America*, 131(6): 4640–4650.

Cheng, W.; and Hüllermeier, E. 2009. Combining instance-based learning and logistic regression for multilabel classification. *Machine Learning*, 76: 211–225.

De Vazelhes, W.; Carey, C.; Tang, Y.; Vauquier, N.; and Bellet, A. 2020. metric-learn: Metric learning algorithms in python. *Journal of Machine Learning Research*, 21(138): 1–6.

Duan, Y.; Lu, J.; Zheng, W.; and Zhou, J. 2020. Deep Adversarial Metric Learning. *IEEE Transactions on Image Processing*, 29: 2037–2051.

Guo, H.-F.; Han, L.; Su, S.; and Sun, Z.-B. 2018. Deep multi-instance multi-label learning for image annotation. *International Journal of Pattern Recognition and Artificial Intelligence*, 32(03): 1859005.

Guo, X.; Li, L.; Dang, C.; Liang, J.; and Wei, W. 2023. Multiple metric learning via local metric fusion. *Information Sciences*, 621: 341–353.

Hu, H.; Cui, Z.; Wu, J.; and Wang, K. 2019. Metric Learning-Based Multi-Instance Multi-Label Classification With Label Correlation. *IEEE Access*, 7: 109899–109909.

Huang, S.-J.; and Zhou, Z.-H. 2012. Multi-label learning by exploiting label correlations locally. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 26, 949–955.

Jin, R.; Wang, S.; and Zhou, Z.-H. 2009. Learning a distance metric from multi-instance multi-label data. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 896–902.

Li, Y.-F.; Hu, J.-H.; Jiang, Y.; and Zhou, Z.-H. 2021. Towards Discovering What Patterns Trigger What Labels. *Proceedings of the AAAI Conference on Artificial Intelligence*, 26(1): 1012–1018.

Liu, J.; Tang, X.; Cui, S.; and Guan, X. 2022. Predicting the function of rice proteins through Multi-instance Multi-label Learning based on multiple features fusion. *Briefings in Bioinformatics*, 23(3): bbac095.

Liu, W.; and Tsang, I. 2015. Large margin metric learning for multi-label prediction. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 29.

Mao, J.-X.; Wang, W.; and Zhang, M.-L. 2023. Label specific multi-semantics metric learning for multi-label classification: global consideration helps. In *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence*, IJCAI '23. ISBN 978-1-956792-03-4.

Nguyen, B.; Ferri, F. J.; Morell, C.; and De Baets, B. 2019. An efficient method for clustered multi-metric learning. *Information Sciences*, 471: 149–163.

Qiu, S.; Wang, M.; Yang, Y.; Yu, G.; Wang, J.; Yan, Z.; Domeniconi, C.; and Guo, M. 2023. Meta Multi-Instance Multi-Label learning by heterogeneous network fusion. *Information Fusion*, 94: 272–283.

Reyes, O.; Morell, C.; and Ventura, S. 2016. Effective lazy learning algorithm based on a data gravitation model for multi-label learning. *Information Sciences*, 340-341: 159–174.

Su, C.; Yan, Z.; and Yu, G. 2021. Cost-effective multi-instance multilabel active learning. *International Journal of Intelligent Systems*, 36(12): 7177–7203.

Sun, Y.-P.; and Zhang, M.-L. 2021. Compositional metric learning for multi-label classification. *Frontiers of Computer Science*, 15: 1–12.

Wang, J.; Kalousis, A.; and Woznica, A. 2012. Parametric Local Metric Learning for Nearest Neighbor Classification. In *Neural Information Processing Systems*.

Weinberger, K. Q.; Blitzer, J.; and Saul, L. 2005. Distance Metric Learning for Large Margin Nearest Neighbor Classification. In Weiss, Y.; Schölkopf, B.; and Platt, J., eds., *Advances in Neural Information Processing Systems*, volume 18. MIT Press.

Weinberger, K. Q.; and Saul, L. K. 2009. Distance metric learning for large margin nearest neighbor classification. *Journal of machine learning research*, 10(2).

Wu, H.; Zhou, Q.; Nie, R.; and Cao, J. 2020. Effective metric learning with co-occurrence embedding for collaborative recommendations. *Neural Networks*, 124: 308–318.

Xu, X.; and Frank, E. 2004. Logistic regression and boosting for labeled bags of instances. In *Pacific-Asia conference on knowledge discovery and data mining*, 272–281. Springer.

Xu, Y.; Min, H.; Song, H.; and Wu, Q. 2016. Multi-instance multi-label distance metric learning for genome-wide protein function prediction. *Computational biology and chemistry*, 63: 30–40.

Yang, M.; Tang, W.-T.; and Min, F. 2022. Multi-instance Multi-label Learning Based on Parallel Attention and Local Label Manifold Correlation. In *2022 IEEE 9th International Conference on Data Science and Advanced Analytics (DSAA)*, 1–10.

Zafra, A.; and Gibaja, E. 2023. Nearest neighbor-based approaches for multi-instance multi-label classification. *Expert Systems with Applications*, 232: 120876.

Zha, Z.-J.; Hua, X.-S.; Mei, T.; Wang, J.; Qi, G.-J.; and Wang, Z. 2008. Joint multi-label multi-instance learning for image classification. In *2008 ieee conference on computer vision and pattern recognition*, 1–8. IEEE.

Zhang, M.-L. 2010. A k-Nearest Neighbor Based Multi-Instance Multi-Label Learning Algorithm. In *2010 22nd IEEE International Conference on Tools with Artificial Intelligence*, volume 2, 207–212.

Zhang, M.-L.; and Zhou, Z.-H. 2007a. ML-KNN: A lazy learning approach to multi-label leaming. *Pattern recognition*, 40(7): 2038–2048.

Zhang, M.-L.; and Zhou, Z.-H. 2007b. ML-KNN: A lazy learning approach to multi-label learning. *Pattern Recognition*, 40(7): 2038–2048.

Zhang, M.-L.; and Zhou, Z.-H. 2008. M3MIML: A Maximum Margin Method for Multi-instance Multi-label Learning. In *2008 Eighth IEEE International Conference on Data Mining*, 688–697.

Zhang, M.-L.; and Zhou, Z.-H. 2014. A Review on Multi-Label Learning Algorithms. *IEEE transactions on knowledge and data engineering*, 26(8): 1819–1837.

Zhang, Z.-L.; and Zhang, M.-l. 2006. Multi-Instance Multi-Label Learning with Application to Scene Classification. In Schölkopf, B.; Platt, J.; and Hoffman, T., eds., *Advances in Neural Information Processing Systems*, volume 19. MIT Press.

Zhao, Y.; and Yang, L. 2023. Distance metric learning based on the class center and nearest neighbor relationship. *Neural Networks*, 164: 631–644.

Zhou, Z.-H.; and Zhang, M.-L. 2006. Multi-Instance Multi-Label Learning with Application to Scene Classification. 1609 – 1616. Cited by: 50.

Zhou, Z.-H.; Zhang, M.-L.; Huang, S.-J.; and Li, Y.-F. 2012. Multi-instance multi-label learning. *Artificial Intelligence*, 176(1): 2291–2320.