Progressive Label Propagation for Semi-Supervised Multi-Dimensional Classification

Teng Huang[1,3] , **Bin-Bin Jia**[2] , **Min-Ling Zhang**[1,3*]
[1]School of Computer Science and Engineering, Southeast University, Nanjing 210096, China
[2]College of Electrical and Information Engineering, Lanzhou University of Technology, Lanzhou 730050, China
[3]Key Lab. of Computer Network and Information Integration (Southeast University), MOE, China
{tengh, zhangml}@seu.edu.cn, jiabinbin@lut.edu.cn ,

Abstract

In multi-dimensional classification (MDC), each training example is associated with multiple class variables from different class spaces. However, it is rather costly to collect labeled MDC examples which have to be annotated from several dimensions (class spaces). To reduce the labeling cost, we attempt to deal with the MDC problem under the semi-supervised learning setting. Accordingly, a novel MDC approach named PLAP is proposed to solve the resulting semi-supervised MDC problem. Overall, PLAP works under the label propagation framework to utilize unlabeled data. To further consider dependencies among class spaces, PLAP deals with each class space in a progressive manner, where the previous propagation results will be used to initialize the current propagation procedure, and all processed class spaces and the current one will be regarded as an entirety. Experiments validate the effectiveness of the proposed approach.

1 Introduction

In supervised learning, one traditional task is to learn a model under the supervision of one single class variable, e.g., multi-class classification. However, the simplifying assumption does not fit well in many real-world applications because some objects with rich semantics should be classified along different dimensions. For instance, a piece of song can be classified from `emotion` dimension (with possible labels *happy*, *sad*, *cathartic*, etc.), from the `scenario` dimension (with possible labels *walk*, *wedding*, *nightclub*, etc.), and from the `language` dimension (with possible labels *English*, *Chinese*, *Spanish*, etc.). To build classification models for such kind of problems, the multi-dimensional classification (MDC) framework [Read *et al.*, 2014a; Cambuí *et al.*, 2021] is a more natural solution which associates multiple class variables to the object. Here, each class variable corresponds to one specific class space which is used to characterize the object's semantics from one dimension. In fact, the needs of learning from objects with multi-dimensional semantics widely exist in diverse application scenarios, such as text mining [Shatkay *et al.*, 2008], computer vision [Lian *et al.*, 2020], bioinformatics [Borchani *et al.*, 2013].

The MDC problem can be solved via learning a multi-class classifier for each dimension. However, this independent decomposition strategy ignores the dependencies among different class spaces, which might lead to performance degeneration. Thus, most existing approaches aim at modeling the class dependencies in either explicit ways [Bielza *et al.*, 2011; Read *et al.*, 2014b; Arias *et al.*, 2016] or implicit ways [Zhang *et al.*, 2022]. It is worth noting that a common assumption for these works is that there are enough labeled MDC samples available for model induction, while it is rather costly to annotate the semantics of objects from several dimensions. To reduce the labeling cost, this paper investigates the feasibility of learning from a few labeled MDC samples with the help of a large number of unlabeled data, which formalizes the framework *semi-supervised multi-dimensional classification* (SSMDC). As a mixture of MDC and semi-supervised learning, both class dependencies and unlabeled data should be carefully utilized when designing SSMDC approaches.

In this paper, we propose a novel approach named PLAP (i.e., *Progressive LAbel Propagation*) to solve the SSMDC problem. Specifically, to utilize the unlabeled data, PLAP propagates the labeling information from labeled samples to unlabeled data via constructing a weighted directed graph. To utilize the class dependencies, PLAP progressively deals with each class space where the current propagation procedure will be initialized by the previous propagation results. To evaluate PLAP's effectiveness, we construct SSMDC testbed with ten real-world MDC data sets and compare PLAP with its degenerated version which deals with each class space independently via label propagation as well as six state-of-the-art MDC approaches which only use labeled MDC samples to induce supervised models. Experimental results clearly show that PLAP achieves superior performance against both its degenerated version and existing MDC approaches. To the best of our knowledge, this paper makes the first attempt towards studying the MDC problem in data utilization level and is different from existing works which are in method design level.

The rest of this paper is organized as follows. Firstly, the related works on MDC are briefly discussed in Section 2. Then, the technical details of the proposed PLAP approach are presented in Section 3. After that, the experimental results of comparative studies are reported in Section 4. Finally, we

*Corresponding author

conclude this paper in Section 5.

2 Related Work

Multi-class classification (MCC) and multi-label classification (MLC) are the two most related frameworks to MDC. Specifically, MCC can be regarded as a special case of MDC where the number of dimensions is one (i.e., single dimensional classification), while MLC [Zhang and Zhou, 2014; Gibaja and Ventura, 2015; Liu *et al.*, 2022] can also be regarded as a special case of MDC where the number of class labels in each dimension is two. Conceptually, both MCC and MLC assume one homogeneous class space while MDC assumes multiple heterogeneous class spaces in output space. Compared with MCC, MLC no longer restricts single relevant label for each instance to model ambiguous objects in a better way, while MDC employs multiple class spaces to model the multi-dimensional semantics of objects more conveniently.

The MDC problem can be solved by decomposing it into multiple independent MCC problems, one per dimension, which is usually termed as binary relevance (BR). However, BR doesn't consider potential class dependencies which should be utilized for model induction. According to the different ways of modeling class dependencies, existing works can be roughly categorized into explicitly and implicitly modeling methods. The first category aims at explicitly modeling class dependencies with some structures. Representative approaches include the family of multi-dimensional Bayesian classifiers which assume directed acyclic graph over class spaces [Gil-Begue *et al.*, 2021], the family of classifier chains which learn a chain of classifiers where the labeling information for preceding classifiers on the chain will be used as features for subsequent classifiers [Zaragoza *et al.*, 2011; Jia and Zhang, 2022a], the family of two-level methods which consider low-order and high-order dependencies in the frst and second level respectively [Arias *et al.*, 2016; Jia and Zhang, 2020b; Jia and Zhang, 2021a].

The second category aims at implicitly modeling class dependencies via manipulating either feature or output space. For manipulating feature space, KRAM generates *k*NN-augmented features to enrich the original feature space which can facilitate subsequent model induction [Jia and Zhang, 2020a], and the following works LEFA [Wang *et al.*, 2020] and SFAM [Jia and Zhang, 2022b] respectively improve the generation and utilization of augmented features. For manipulating output space, MLKT [Ma and Chen, 2018] and SLEM [Jia and Zhang, 2021b] transform the original categorical output space into a binary and real-valued one and then induce the predictive model in their transformed label space.

All the above approaches work by assuming enough labeled MDC samples available for model induction. However, it is rather costly to collect labeled MDC samples which have to be annotated from different dimensions. Semi-supervised learning [van Engelen and Hoos, 2020] aims at learning supervised models with a few labeled samples as well as a large number of unlabeled data and then reducing the need of labeled samples. To utilize unlabeled data, label propagation [Zhou *et al.*, 2003] is one of commonly-used techniques which aims at making similar instances have similar labeling information. In the next section, we will present the technical details of the proposed PLAP approach which learns from a few labeled MDC samples as well as a large number of unlabeled data via progressive label propagation.

3 The PLAP Approach

Let $\mathcal{X} = \mathbb{R}^d$ be the d-dimensional input (feature) space and $\mathcal{Y} = C_1 \times C_2 \times \cdots \times C_q$ be the output space which corresponds to the Cartesian product of q class spaces. Here, each class space C_j consists of K_j class labels, i.e., $C_j = \{c_1^j, c_2^j, \ldots, c_{K_j}^j\}$ $(1 \le j \le q)$. Given a set of labeled MDC samples $\mathcal{D}_l = \{(\boldsymbol{x}_i, \boldsymbol{y}_i) \mid 1 \le i \le L\}$, for each labeled MDC sample $(\boldsymbol{x}_i, \boldsymbol{y}_i)$, $\boldsymbol{y}_i = [y_{i1}, y_{i2}, \ldots, y_{iq}]^\top \in \mathcal{Y}$ is the ground-truth class vector of $\boldsymbol{x}_i \in \mathcal{X}$ where $y_{ij} \in C_j$ corresponds to $\boldsymbol{x}_i$'s class label in the j-th dimension. Furthermore, given a set of unlabeled samples $\mathcal{D}_u = \{\boldsymbol{x}_j \mid L{+}1 \le j \le L{+}U\}$, the task of SSMDC is to induce a predictive model $f : \mathcal{X} \mapsto \mathcal{Y}$ from $\mathcal{D} = \mathcal{D}_l \cup \mathcal{D}_u$. Generally, it is assumed that $L \ll U$ and we further denote $N = L + U$ for convenience.

To propagate the labeling information from labeled samples $\mathcal{D}_l$ to unlabeled data $\mathcal{D}_u$, PLAP constructs a weighted directed graph $G = (V, E)$. Here, the vertex set V corresponds to the instances in $\mathcal{D}_l \cup \mathcal{D}_u$, i.e., $V = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_L, \ldots, \boldsymbol{x}_N\}$, while the edge set E corresponds to the affinity matrix $\mathbf{W} \in \mathbb{R}^{N \times N}$ defined based on the Gaussian function:

$$W_{ij} = \begin{cases} e^{-\frac{\|\boldsymbol{x}_i - \boldsymbol{x}_j\|^2}{2\sigma^2}}, & \text{if } i \neq j \\ 0, & \text{otherwise} \end{cases} \tag{1}$$

where W_{ij} is the (i,j)-th item of $\mathbf{W}$ and σ is the bandwidth parameter of Gaussian function. Based on $\mathbf{W}$, PLAP further constructs the propagation matrix $\mathbf{S} = \mathbf{D}^{-\frac{1}{2}}\mathbf{W}\mathbf{D}^{-\frac{1}{2}}$ where $\mathbf{D} = diag(d_1, d_2, \ldots, d_N)$ is a diagonal matrix with $d_i = \sum_{j=1}^{N} W_{ij}$. With the propagation matrix $\mathbf{S}$, PLAP conducts label propagation procedure via the following iterative equation [Zhou *et al.*, 2003]:

$$\mathbf{F}(t) = \alpha\mathbf{S}\mathbf{F}(t-1) + (1-\alpha)\mathbf{Y} \tag{2}$$

Here, $\mathbf{Y}$ is the initial label matrix, $\mathbf{F}(t)$ is the propagated label matrix in the t-th round ($t \in \{1, 2, \ldots\}$), and $\alpha \in (0, 1)$ is a hyper-parameter to be specified for balancing the importance of $\mathbf{S}\mathbf{F}(t)$ and $\mathbf{Y}$. Besides, $\mathbf{F}(0)$ is usually initialized as $\mathbf{Y}$ for the first iteration. In this paper, PLAP progressively deals with each class space, the definitions of $\mathbf{Y}$ and the corresponding $\mathbf{F}$ are different for different class spaces.

To consider the first class space C_1, the current $\mathbf{Y} \in \{0,1\}^{N \times K_1}$ is set as follows:

$$Y_{ia} = \begin{cases} 1, & \text{if } (1 \le i \le L) \wedge (y_{i1} = c_a^1) \\ 0, & \text{otherwise} \end{cases} \tag{3}$$

where Y_{ia} is the (i,a)-th item of $\mathbf{Y}$. In other words, Y_{ia} is set to 1 only for labeled sample $\boldsymbol{x}_i$ (i.e., $1 \le i \le L$) if its class label w.r.t. C_1 is c_a^1 (i.e., $y_{i1} = c_a^1$) and 0 otherwise. With $\mathbf{Y}$, we can obtain the final $\mathbf{F} \in \mathbb{R}^{N \times K_1}$ via iterating Eq.(2) until convergence. Let F_{ia} denote the (i,a)-th item of $\mathbf{F}$, the class label $\hat{y}_{i1}$ of unlabeled sample $\boldsymbol{x}_i$ $(L+1 \le i \le N)$ w.r.t. C_1 can be obtained via the following rule:

$$\hat{y}_{i1} = c_{\hat{a}}^1, \;\; \text{where } \hat{a} = \arg\max_{1 \le a \le K_1} F_{ia} \tag{4}$$

Then, to consider the second class space C_2 and model the dependencies among C_1 and C_2, we deal with the first two class spaces as an entirety. Moreover, the previous predictions w.r.t. C_1 for unlabeled samples will also be used to determine the current $\mathbf{Y}$ in Eq.(2). Specifically, let $\phi_2(a_1, a_2)$ be some injective function from the Cartesian product $\{1, 2, \ldots, K_1\} \times \{1, 2, \ldots, K_2\}$ to natural numbers $\{1, 2, \ldots, K_1 \times K_2\}$ and we denote $K_{12} = K_1 \times K_2$ for convenience. To determine the current $\mathbf{Y} \in \{0, 1\}^{N \times K_{12}}$, consider the two cases as follows:

1. For labeled sample $\boldsymbol{x}_i$ $(1 \leq i \leq L)$:

$$Y_{i\phi_2(a_1,a_2)} = \begin{cases} 1, & \text{if } \mathrm{CL}_i^2 = \text{true} \\ 0, & \text{otherwise} \end{cases} \tag{5}$$

   where $\mathrm{CL}_i^2 \triangleq (y_{i1} = c_{a_1}^1) \wedge (y_{i2} = c_{a_2}^2)$. In other words, the corresponding item $Y_{i\phi_2(a_1,a_2)}$ for labeled sample $\boldsymbol{x}_i$ is directly determined by its class labels y_{i1} and y_{i2} w.r.t. C_1 and C_2.

2. For unlabeled sample $\boldsymbol{x}_i$ $(L+1 \leq i \leq N)$, the class label w.r.t. C_1 has already been predicted in the previous label propagation step while the class label w.r.t. C_2 is to be determined. Let $\mathcal{N}(\boldsymbol{x}_i)$ be $\boldsymbol{x}_i$'s k nearest neighbors identified in labeled samples $\mathcal{D}_l$, and $n_{ia_2}^2$ be the number of samples with class label $c_{a_2}^2$ w.r.t. C_2 in $\mathcal{N}(\boldsymbol{x}_i)$, it is easy to know that $\sum_{a_2=1}^{K_2} n_{ia_2}^2 = k$. Suppose $\boldsymbol{x}_i$ is predicted as $c_{\hat{a}_1}^1$ w.r.t. C_1, we set $\mathbf{Y}$ as:

$$Y_{i\phi_2(a_1,a_2)} = \begin{cases} \frac{n_{ia_2}^2}{k}, & \text{if } \mathrm{CU}_i^2 = \text{true} \\ 0, & \text{otherwise} \end{cases} \tag{6}$$

   where $\mathrm{CU}_i^2 \triangleq (a_1 = \hat{a}_1)$ and $1 \leq a_2 \leq K_2$.

To facilitate understanding, here is an example:

Example 1. *Suppose that the output space of $\mathcal{D}$ is $\mathcal{Y} = C_1 \times C_2$ where $C_1 = \{c_1^1, c_2^1\}$, $C_2 = \{c_1^2, c_2^2, c_3^2\}$. Given the MDC data set $\mathcal{D} = \{(\boldsymbol{x}_1, \boldsymbol{y}_1), (\boldsymbol{x}_2, \boldsymbol{y}_2), (\boldsymbol{x}_3, \boldsymbol{y}_3)\} \cup \{\boldsymbol{x}_4, \boldsymbol{x}_5\}$, where $\boldsymbol{y}_1 = [c_1^1, c_2^2]^\top$, $\boldsymbol{y}_2 = [c_2^1, c_3^2]^\top$ and $\boldsymbol{y}_3 = [c_1^1, c_1^2]^\top$. Let $k = 2$ and suppose $\mathcal{N}(\boldsymbol{x}_4) = \{\boldsymbol{x}_2, \boldsymbol{x}_3\}$, $\mathcal{N}(\boldsymbol{x}_5) = \{\boldsymbol{x}_1, \boldsymbol{x}_3\}$, then $n_{41}^2 = 1$, $n_{42}^2 = 0$, $n_{43}^2 = 1$ because there is respectively one sample with c_1^2 (i.e., $\boldsymbol{x}_3$) and c_3^2 (i.e., $\boldsymbol{x}_2$) w.r.t. C_2 in $\mathcal{N}(\boldsymbol{x}_4)$, and $n_{51}^2 = 1$, $n_{52}^2 = 1$, $n_{53}^2 = 0$ because there is respectively one sample with c_1^2 (i.e., $\boldsymbol{x}_3$) and c_2^2 (i.e., $\boldsymbol{x}_1$) w.r.t. C_2 in $\mathcal{N}(\boldsymbol{x}_5)$. Moreover, suppose the predicted class labels w.r.t. C_1 in the previous label propagation step for unlabeled samples $\boldsymbol{x}_4$, $\boldsymbol{x}_5$ are $\hat{y}_{41} = c_1^1$, $\hat{y}_{51} = c_2^1$, given the injective function $\phi_2(a_1, a_2) = 3 \times (a_1 - 1) + a_2$, $\mathbf{Y}$ is set as:*

$$\mathbf{Y} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 \end{bmatrix}$$

Here, in the 1st row, $Y_{12} = 1$ because $\boldsymbol{y}_1 = [c_1^1, c_2^2]^\top$ and $\phi_2(1, 2) = 3 \times (1 - 1) + 2 = 2$. In the 2nd row, $Y_{26} = 1$ because $\boldsymbol{y}_2 = [c_2^1, c_3^2]^\top$ and $\phi_2(2, 3) = 3 \times (2 - 1) + 3 = 6$. In the 3rd row, $Y_{31} = 1$ because $\boldsymbol{y}_3 = [c_1^1, c_1^2]^\top$ and $\phi_2(1, 1) = 1 \times (1 - 1) + 1 = 1$; In the 4th row, given $\hat{y}_{41} = c_1^1$ (i.e., $a_1 = 1$), $Y_{41} = \frac{n_{41}^2}{k} = \frac{1}{2}$ because $\phi_2(1, 1) = 1$, $Y_{42} = \frac{n_{42}^2}{k} = 0$ because $\phi_2(1, 2) = 2$, $Y_{43} = \frac{n_{43}^2}{k} = \frac{1}{2}$ because $\phi_2(1, 3) = 3$; In the 5th row, given $\hat{y}_{51} = c_2^1$ (i.e., $a_1 = 2$), $Y_{51} = \frac{n_{51}^2}{k} = \frac{1}{2}$ because $\phi_2(2, 1) = 4$, $Y_{52} = \frac{n_{52}^2}{k} = \frac{1}{2}$ because $\phi_2(2, 2) = 5$, $Y_{53} = \frac{n_{53}^2}{k} = 0$ because $\phi_2(2, 3) = 6$.

With the $\mathbf{Y}$ determined by Eq.(5) and Eq.(6), we can also obtain the corresponding $\mathbf{F} \in \mathbb{R}^{N \times K_{12}}$ via iterating Eq.(2) until convergence. For unlabeled sample $\boldsymbol{x}_i$ $(L+1 \leq i \leq N)$, its class labels w.r.t. C_1 and C_2 can be determined via the following rule:

$$\hat{y}_{i1} = c_{\hat{a}_1}^1, \hat{y}_{i2} = c_{\hat{a}_2}^2, \quad \begin{array}{l} \text{where } [\hat{a}_1, \hat{a}_2] = \phi_2^{-1}(\hat{a}), \\ \hat{a} = \arg\max_{1 \leq a \leq K_{12}} F_{ia} \end{array} \tag{7}$$

where $\phi_2^{-1}(\cdot)$ is the corresponding inverse function of $\phi_2(\cdot, \cdot)$.

Furthermore, we can generalize above propagation procedure to the j-th class space $(j \geq 2)$. Specifically, we deal with the first three class spaces as an entirety and define injective function $\phi_j(a_1, \ldots, a_j)$ which maps the Cartesian product $\{1, 2, \ldots, K_1\} \times \ldots \times \{1, 2, \ldots, K_j\}$ to natural numbers $\{1, 2, \ldots, K_1 \times \ldots \times K_j\}$. To determine the corresponding $\mathbf{Y}$, Eq.(5) will be generalized as follows:

$$Y_{i\phi_j(a_1,\ldots,a_j)} = \begin{cases} 1, & \text{if } \mathrm{CL}_i^j = \text{true} \\ 0, & \text{otherwise} \end{cases} \tag{8}$$

where $\mathrm{CL}_i^j \triangleq (y_{i1} = c_{a_1}^1) \wedge \ldots \wedge (y_{ij} = c_{a_j}^j)$. Eq.(6) will be generalized as follows:

$$Y_{i\phi_j(a_1,\ldots,a_j)} = \begin{cases} \frac{n_{ia_j}^j}{k}, & \text{if } \mathrm{CU}_i^j = \text{true} \\ 0, & \text{otherwise} \end{cases} \tag{9}$$

where $\mathrm{CU}_i^j \triangleq (a_1 = \hat{a}_1) \wedge \ldots \wedge (a_{j-1} = \hat{a}_{j-1})$. $n_{ia_j}^j$ denotes the number of samples with class label $c_{a_j}^j$ w.r.t. C_j in $\mathcal{N}(\boldsymbol{x}_i)$ and $\boldsymbol{x}_i$ is respectively predicted as $c_{\hat{a}_1}^1$, ..., $c_{\hat{a}_{j-1}}^{j-1}$ w.r.t. C_1, ..., C_{j-1} in the previous label propagation step. For unlabeled sample $\boldsymbol{x}_i$ $(L+1 \leq i \leq N)$, its class labels w.r.t. C_1, ..., C_j can be determined with the help of the inverse function of $\phi_j(a_1, \ldots, a_j)$ like Eq.(7).

Table 1: Basic information for data sets. Here, n, x and b in last column represent numeric, nominal and binary type features.

Data set	#Exam.	#Dim.	#Feat.
Edm	154	2	$16n$
Song	785	3	$98n$
WQpla.	1060	7	$16n$
WQani.	1060	7	$16n$
WQ	1060	14	$16n$
BeLaE	1930	5	$1n, 44x$
Thyroid	9172	7	$7n, 20b, 2x$
Pain	9734	10	$136n$
Disfa	13095	12	$136n$
Adult	18419	4	$5n, 5x$

Table 2: Experimental results (mean±std.) of each MDC approach with $L = 40$ labeled samples. In addition, •/○ indicates whether PLAP is significantly superior/inferior to other compared approaches on each data set with pairwise t-test at 0.05 significance level.

(a) Hamming Score

Data Set	PLAP	PLAP$_{\mathrm{d}}$	BR	CC	SLEM	KRAM$_{\mathrm{d}}$	KRAM$_{\mathrm{c}}$	MDKNN
Edm	.724±.033	.709±.031	.684±.022•	.657±.033•	.718±.039	.661±.014•	.654±.022•	.647±.017•
Song	.740±.010	.692±.012•	.676±.041•	.671±.037•	.658±.065•	.711±.018•	.707±.018•	.716±.028•
WQpla.	.623±.029	.627±.015	.483±.054•	.475±.038•	.589±.029•	.517±.019•	.518±.024•	.546±.022•
WQani.	.615±.009	.611±.012	.448±.050•	.433±.043•	.546±.015•	.476±.025•	.473±.025•	.506±.020•
WQ	.620±.013	.618±.013	.473±.028•	.459±.017•	.567±.018•	.474±.016•	.475±.018•	.512±.013•
BeLaE	.333±.011	.332±.011	.342±.012	.343±.011	.288±.015•	.313±.010•	.313±.008•	.290±.007•
Thyroid	.960±.001	.956±.004•	.305±.013•	.677±.008•	.933±.042	.940±.015•	.944±.013•	.951±.009•
Pain	.947±.001	.894±.009•	.914±.012•	.908±.016•	.931±.010•	.887±.017•	.885±.016•	.926±.015•
Disfa	.872±.000	.859±.001•	.862±.007•	.813±.015•	.857±.021	.779±.020•	.784±.024•	.840±.009•
Adult	.593±.014	.549±.018•	.548±.020•	.551±.019•	.564±.014•	.552±.018•	.552±.018•	.520±.016•

(b) Exact Match

Data Set	PLAP	PLAP$_{\mathrm{d}}$	BR	CC	SLEM	KRAM$_{\mathrm{d}}$	KRAM$_{\mathrm{c}}$	MDKNN
Edm	.504±.057	.517±.044	.462±.054	.478±.047	.525±.067	.432±.039•	.423±.045•	.457±.032
Song	.390±.019	.322±.032•	.279±.058•	.271±.055•	.329±.056•	.349±.029•	.342±.031•	.351±.044•
WQpla.	.057±.035	.063±.029	.014±.009•	.007±.003•	.061±.015	.021±.010•	.023±.012•	.030±.014•
WQani.	.041±.014	.041±.014	.009±.003•	.008±.004•	.021±.007•	.008±.003•	.009±.004•	.011±.005•
WQ	.003±.002	.006±.002○	.002±.000	.002±.001	.006±.002○	.000±.001•	.000±.001•	.000±.000•
BeLaE	.009±.004	.009±.002	.008±.002	.009±.002	.005±.001•	.005±.002•	.005±.002•	.003±.001•
Thyroid	.736±.004	.713±.015•	.077±.014•	.082±.013•	.607±.204	.642±.077•	.663±.069•	.688±.049•
Pain	.739±.009	.567±.039•	.544±.080•	.513±.100•	.666±.063•	.493±.069•	.489±.062•	.647±.071•
Disfa	.366±.000	.347±.003•	.280±.038•	.079±.032•	.303±.074•	.084±.031•	.091±.051•	.261±.054•
Adult	.070±.010	.104±.018○	.086±.015○	.089±.014○	.114±.016○	.078±.017	.077±.017	.065±.022

(c) Sub-Exact Match

Data Set	PLAP	PLAP$_{\mathrm{d}}$	BR	CC	SLEM	KRAM$_{\mathrm{d}}$	KRAM$_{\mathrm{c}}$	MDKNN
Edm	.945±.029	.901±.022•	.906±.030•	.837±.027•	.911±.019•	.889±.035•	.886±.041•	.838±.032•
Song	.835±.014	.783±.012•	.769±.056•	.761±.053•	.710±.099•	.798±.025•	.793±.025•	.812±.038
WQpla.	.216±.069	.239±.033	.074±.037•	.051±.023•	.191±.038	.104±.026•	.111±.032•	.133±.029•
WQani.	.189±.021	.186±.029	.060±.014•	.053±.012•	.112±.017•	.059±.019•	.058±.021•	.072±.019•
WQ	.018±.013	.023±.013	.004±.002•	.003±.001•	.015±.006	.002±.002•	.002±.001•	.003±.002•
BeLaE	.059±.008	.059±.007	.059±.010	.063±.009	.033±.007•	.042±.008•	.043±.008•	.031±.004•
Thyroid	.982±.000	.976±.010	.148±.018•	.297±.049•	.931±.084	.942±.030•	.948±.027•	.969±.014•
Pain	.845±.002	.693±.021•	.763±.037•	.734±.062•	.806±.022•	.686±.053•	.679±.052•	.800±.036•
Disfa	.564±.000	.534±.004•	.529±.029•	.311±.064•	.512±.067•	.268±.065•	.284±.081•	.467±.034•
Adult	.491±.033	.389±.034•	.394±.042•	.401±.042•	.422±.026•	.409±.037•	.410±.039•	.349±.035•

4 Experiments

4.1 Experimental Setting

Data Sets

In this paper, we use ten real-world MDC data sets for experimental studies where Table 1 summarizes their basic characteristics, including the number of examples (#Exam.), the number of class spaces (#Dim.), and the number of features (#Feat.). For each data set, we randomly sample 40, 50 and 60 examples to form the labeled data set $\mathcal{D}_l$ while the remaining examples are used to form the unlabeled data set $\mathcal{D}_u$. The sampling procedure is repeated ten times, and the mean metric values as well as standard deviations are recorded.

Evaluation Metrics

In this paper, we consider three commonly used metrics for performance evaluation, namely *hamming score* (a.k.a. global accuracy [Bielza *et al.*, 2011]), exact match (a.k.a. example accuracy [Read *et al.*, 2014a]) and sub-exact match [Jia and Zhang, 2020a]. Given the test set $\mathcal{S} = \{(\boldsymbol{x}_i, \boldsymbol{y}_i) \mid 1 \leq i \leq p\}$ and the MDC model f to be evaluated, the definitions of these three evaluation metrics are given as follows:

1. Hamming Score:

$$\mathrm{HS}_{\mathcal{S}}(f) = \frac{1}{p}\sum_{i=1}^{p}\frac{1}{q}\cdot r^{(i)}$$

Table 3: Experimental results (mean±std.) of each MDC approach with $L = 50$ labeled samples. In addition, •/○ indicates whether PLAP is significantly superior/inferior to other compared approaches on each data set with pairwise t-test at 0.05 significance level.

(a) Hamming Score

Data Set	PLAP	PLAP$_{\mathrm{d}}$	BR	CC	SLEM	KRAM$_{\mathrm{d}}$	KRAM$_{\mathrm{c}}$	MDKNN
Edm	.749±.030	.730±.028	.690±.024•	.663±.031•	.652±.068•	.673±.035•	.670±.035•	.642±.031•
Song	.739±.012	.699±.008•	.662±.047•	.658±.043•	.682±.057•	.729±.013	.727±.012	.736±.014
WQpla.	.637±.010	.637±.013	.458±.061•	.445±.047•	.596±.019•	.522±.026•	.518±.025•	.556±.035•
WQani.	.618±.009	.615±.015	.421±.057•	.414±.048•	.563±.028•	.478±.020•	.477±.018•	.527±.020•
WQ	.624±.013	.625±.014	.444±.031•	.436±.022•	.574±.018•	.487±.013•	.492±.013•	.534±.013•
BeLaE	.341±.010	.335±.010•	.343±.012	.345±.010	.293±.015•	.317±.007•	.317±.008•	.296±.007•
Thyroid	.960±.001	.954±.004•	.283±.026•	.661±.019•	.946±.013•	.933±.018•	.933±.018•	.940±.014•
Pain	.948±.001	.906±.005•	.903±.015•	.894±.017•	.916±.034•	.905±.008•	.906±.009•	.932±.008•
Disfa	.874±.001	.863±.003•	.850±.013•	.799±.021•	.868±.006•	.804±.016•	.808±.016•	.849±.007•
Adult	.603±.010	.558±.013•	.541±.029•	.546±.028•	.578±.012•	.564±.014•	.563±.012•	.523±.018•

(b) Exact Match

Data Set	PLAP	PLAP$_{\mathrm{d}}$	BR	CC	SLEM	KRAM$_{\mathrm{d}}$	KRAM$_{\mathrm{c}}$	MDKNN
Edm	.544±.044	.539±.039	.460±.047•	.481±.042•	.416±.097•	.438±.047•	.436±.052•	.452±.050•
Song	.387±.023	.334±.023•	.264±.066•	.258±.063•	.337±.056•	.371±.022	.371±.021	.387±.024
WQpla.	.065±.028	.081±.025	.009±.007•	.007±.006•	.062±.019	.027±.016•	.024±.017•	.032±.019•
WQani.	.045±.013	.046±.015	.010±.003•	.010±.004•	.027±.009•	.010±.005•	.010±.005•	.015±.008•
WQ	.004±.003	.008±.002○	.001±.000•	.002±.000•	.007±.002○	.001±.001•	.001±.001•	.000±.001•
BeLaE	.010±.004	.010±.002	.008±.002	.008±.002	.005±.002•	.006±.002•	.005±.002•	.004±.001•
Thyroid	.736±.004	.707±.016•	.069±.017•	.072±.018•	.675±.056•	.623±.070•	.623±.065•	.627±.077•
Pain	.745±.003	.614±.023•	.501±.073•	.449±.089•	.613±.168•	.556±.040•	.564±.042•	.673±.040•
Disfa	.367±.001	.354±.003•	.216±.071•	.070±.038•	.348±.016•	.156±.040•	.163±.045•	.292±.027•
Adult	.080±.008	.110±.013○	.082±.023	.086±.023	.132±.016○	.093±.014○	.091±.014○	.070±.019

(c) Sub-Exact Match

Data Set	PLAP	PLAP$_{\mathrm{d}}$	BR	CC	SLEM	KRAM$_{\mathrm{d}}$	KRAM$_{\mathrm{c}}$	MDKNN
Edm	.954±.026	.920±.021•	.921±.015•	.845±.025•	.888±.059•	.909±.039•	.905±.040•	.833±.024•
Song	.836±.017	.787±.012•	.749±.067•	.741±.067•	.754±.088•	.824±.016	.820±.014	.830±.017
WQpla.	.249±.027	.259±.028	.057±.034•	.042±.025•	.193±.037•	.113±.034•	.107±.033•	.140±.053•
WQani.	.196±.021	.194±.031	.058±.014•	.055±.013•	.133±.036•	.059±.020•	.061±.020•	.090±.023•
WQ	.026±.014	.034±.013○	.003±.001•	.002±.001•	.022±.007	.004±.004•	.005±.004•	.004±.003•
BeLaE	.064±.009	.062±.006	.060±.009	.064±.009	.035±.008•	.043±.005•	.046±.006•	.036±.005•
Thyroid	.982±.000	.975±.010	.139±.017•	.231±.077•	.958±.023•	.920±.057•	.921±.061•	.951±.022•
Pain	.846±.002	.733±.013•	.730±.045•	.692±.067•	.770±.084•	.738±.025•	.740±.024•	.811±.022•
Disfa	.569±.003	.545±.006•	.487±.047•	.252±.087•	.556±.013•	.359±.049•	.360±.052•	.495±.021•
Adult	.509±.025	.405±.025•	.382±.055•	.391±.050•	.446±.023•	.431±.029•	.429±.026•	.351±.042•

2. Exact Match:

$$\mathrm{EM}_{\mathcal{S}}(f) = \frac{1}{p}\sum_{i=1}^{p}[\![r^{(i)} = q]\!]$$

3. Sub-Exact Match:

$$\mathrm{SEM}_{\mathcal{S}}(f) = \frac{1}{p}\sum_{i=1}^{p}[\![r^{(i)} \geq q-1]\!]$$

Here, $r^{(i)} = \sum_{j=1}^{q}[\![y_{ij} = \hat{y}_{ij}]\!]$ denotes the number of class spaces which are predicted correctly, y_{ij} and $\hat{y}_{ij}$ denote the ground-truth and predicted label w.r.t. the j-th class space for the i-th test sample, $[\![\pi]\!]$ returns 1 if π holds and 0 otherwise.

Compared Approaches

As the first attempt towards solving the SSMDC problem, there are not existing SSMDC works to be used as comparing approaches. Thus, we mainly compare the proposed PLAP approach with existing MDC approaches. In this paper, a total of six well-established MDC approaches are used, including BR, CC, SLEM, KRAM$_{\mathrm{d}}$, KRAM$_{\mathrm{c}}$, MDKNN. Specifically, BR solves the MDC problem via learning a multi-class classifier for each dimension. CC [Jia and Zhang, 2022a] solves the MDC problem via learning a chain of multi-class classifiers, one per dimension. SLEM [Jia and Zhang, 2021b] transforms the MDC problem into a multi-output regression problem via spare label encoding. Both KRAM$_{\mathrm{d}}$ and KRAM$_{\mathrm{c}}$ [Jia and Zhang, 2020a] consider class dependencies via kNN feature

Table 4: Experimental results (mean±std.) of each MDC approach with $L = 60$ labeled samples. In addition, •/○ indicates whether PLAP is significantly superior/inferior to other compared approaches on each data set with pairwise t-test at 0.05 significance level.

(a) Hamming Score

Data Set	PLAP	PLAP_d	BR	CC	SLEM	KRAM_d	KRAM_c	MDKNN
Edm	.766±.028	.736±.028•	.685±.024•	.660±.041•	.714±.058•	.679±.035•	.675±.037•	.657±.025•
Song	.738±.011	.696±.013•	.633±.041•	.632±.037•	.680±.067•	.733±.016	.733±.015	.735±.011
WQpla.	.643±.008	.640±.007	.447±.060•	.434±.044•	.589±.026•	.538±.028•	.535±.029•	.582±.017•
WQani.	.624±.005	.622±.004	.405±.047•	.404±.042•	.555±.027•	.492±.015•	.489±.015•	.540±.018•
WQ	.629±.012	.630±.011	.415±.045•	.417±.029•	.581±.012•	.495±.016•	.494±.016•	.547±.010•
BeLaE	.342±.010	.334±.009•	.347±.010	.349±.010	.293±.015•	.324±.006•	.322±.007•	.301±.007•
Thyroid	.960±.001	.955±.003•	.279±.024•	.658±.019•	.953±.003•	.941±.012•	.945±.008•	.948±.013•
Pain	.948±.000	.912±.003•	.891±.022•	.879±.029•	.932±.021	.915±.007•	.913±.007•	.937±.006•
Disfa	.876±.001	.867±.004•	.834±.014•	.776±.027•	.866±.012•	.816±.016•	.820±.017•	.855±.006•
Adult	.606±.013	.564±.011•	.530±.026•	.536±.026•	.577±.011•	.565±.010•	.565±.013•	.545±.014•

(b) Exact Match

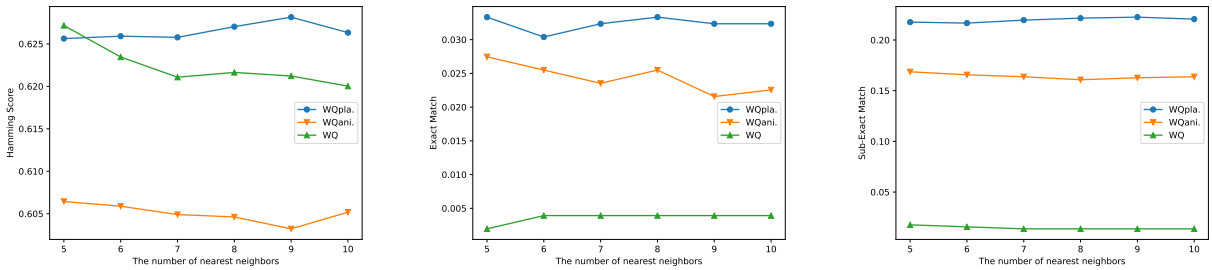
Data Set	PLAP	PLAP_d	BR	CC	SLEM	KRAM_d	KRAM_c	MDKNN
Edm	.586±.036	.544±.038•	.460±.048•	.477±.059•	.521±.077•	.447±.058•	.448±.056•	.469±.043•
Song	.380±.021	.334±.026•	.232±.064•	.226±.060•	.343±.069	.381±.029	.378±.028	.381±.023
WQpla.	.080±.024	.083±.021	.009±.008•	.004±.004•	.065±.018	.032±.017•	.031±.018•	.046±.016•
WQani.	.052±.009	.052±.008	.007±.004•	.007±.004•	.023±.011•	.009±.003•	.009±.003•	.016±.005•
WQ	.005±.002	.009±.002○	.001±.001•	.001±.001•	.007±.003○	.001±.001•	.000±.001•	.001±.001•
BeLaE	.009±.004	.010±.002	.008±.003	.009±.003	.004±.001•	.006±.002	.006±.001•	.004±.002•
Thyroid	.735±.008	.708±.013•	.069±.017•	.072±.018•	.698±.013•	.644±.064•	.665±.049•	.673±.069•
Pain	.745±.003	.628±.019•	.459±.092•	.396±.118•	.669±.101	.593±.037•	.591±.035•	.699±.032•
Disfa	.368±.003	.358±.004•	.154±.048•	.045±.035•	.344±.030•	.188±.046•	.198±.046•	.313±.019•
Adult	.085±.007	.114±.012○	.072±.018•	.075±.020	.130±.012○	.094±.014○	.095±.014○	.085±.011

(c) Sub-Exact Match

Data Set	PLAP	PLAP_d	BR	CC	SLEM	KRAM_d	KRAM_c	MDKNN
Edm	.947±.027	.929±.024	.911±.018•	.844±.029•	.907±.048•	.911±.025•	.902±.035•	.846±.028•
Song	.838±.016	.780±.016•	.706±.062•	.704±.058•	.749±.091•	.828±.018	.830±.018	.834±.014
WQpla.	.263±.024	.263±.022	.047±.030•	.031±.019•	.192±.040•	.131±.035•	.125±.037•	.173±.030•
WQani.	.208±.013	.209±.010	.050±.012•	.047±.009•	.121±.031•	.068±.014•	.069±.012•	.104±.015•
WQ	.031±.013	.038±.011○	.003±.001•	.002±.001•	.024±.007	.005±.004•	.004±.003•	.008±.004•
BeLaE	.066±.010	.061±.005	.063±.010	.065±.010	.034±.008•	.050±.005•	.048±.004•	.038±.004•
Thyroid	.982±.001	.976±.010	.131±.018•	.213±.080•	.972±.009•	.948±.021•	.953±.015•	.964±.021•
Pain	.846±.001	.753±.010•	.689±.077•	.639±.115•	.812±.055	.767±.017•	.761±.018•	.822±.018•
Disfa	.572±.004	.554±.011•	.426±.062•	.190±.088•	.546±.042	.390±.048•	.402±.049•	.516±.025•
Adult	.515±.029	.417±.021•	.364±.051•	.373±.050•	.442±.025•	.431±.024•	.430±.027•	.398±.025•

Table 5: Win/tie/loss counts of pairwise t-test (at 0.05 significance level) between PLAP and each comparing approach.

#labeled samples (L)	Evaluation Metric	PLAP against							In Total
		PLAP_d	BR	CC	SLEM	KRAM_d	KRAM_c	MDKNN	
$L = 40$	Hamming Score	5/5/0	9/1/0	9/1/0	7/3/0	10/0/0	10/0/0	10/0/0	60/10/0
	Exact Match	4/4/2	6/3/1	6/3/1	5/2/3	9/1/0	9/1/0	8/2/0	47/16/7
	Sub-Exact Match	5/5/0	9/1/0	9/1/0	7/3/0	10/0/0	10/0/0	9/1/0	59/11/0
$L = 50$	Hamming Score	6/4/0	9/1/0	9/1/0	10/0/0	9/1/0	9/1/0	9/1/0	61/9/0
	Exact Match	4/4/2	8/2/0	8/2/0	7/1/2	8/1/1	8/1/1	8/2/0	51/13/6
	Sub-Exact Match	5/4/1	9/1/0	9/1/0	9/1/0	9/1/0	9/1/0	9/1/0	59/10/1
$L = 60$	Hamming Score	7/3/0	9/1/0	9/1/0	9/1/0	9/1/0	9/1/0	9/1/0	61/9/0
	Exact Match	5/3/2	9/1/0	8/2/0	5/3/2	7/2/1	8/1/1	8/2/0	50/14/6
	Sub-Exact Match	4/5/1	9/1/0	9/1/0	7/3/0	9/1/0	9/1/0	9/1/0	56/13/1
In Total		45/37/8	77/12/1	76/13/1	66/17/7	80/8/2	81/7/2	79/11/0	504/105/21



(a) Hamming Score (b) Exact Match (c) Sub-Exact Match

Figure 1: Performance of PLAP changes as k varies from 5 to 10

augmentation, where the former only generates discrete kNN statistics while the latter further generates continuous kNN statistics. MDKNN [Jia and Zhang, 2021a] solves the MDC problem via instance-based learning techniques where class dependencies are considered in a two-level strategy. Moreover, we further compare the proposed PLAP approach with its own degenerated version (denoted as PLAP_d) which independently deals with each class space the same as PLAP deals with the first class space.

Note that the six MDC approaches train their models only over labeled MDC samples $\mathcal{D}_l$ while PLAP and PLAP_d train their models over both labeled MDC samples $\mathcal{D}_l$ and unlabeled data $\mathcal{D}_u$. For the six MDC approaches, the suggested parameters in their respective literature are used. For PLAP and PLAP_d, the bandwidth parameter in Eq.(1) is fixed as 50, the trade-off parameter α in Eq.(2) is fixed as 0.99, and the number of nearest neighbors k in Eq.(6) is fixed as 7.

4.2 Experimental Results

The detailed experimental results are reported in Tables 2-4. Moreover, pairwise t-test [Demšar, 2006] at 0.05 significance level is conducted to show whether PLAP achieves significantly superior/inferior performance against other comparing approaches on each data set. Accordingly, the resulting win/tie/loss counts are summarized in Table 5.

According to the reported experimental results, we can make the following observations:

- With 40, 50 and 60 labeled samples, PLAP respectively achieves superior or at least comparable performance against the seven comparing approaches in 203, 203 and 204 cases across all the 210 configurations (10 data sets $\times$ 3 metrics $\times$ 7 comparing approaches).
- PLAP_d independently deals with each dimension via label propagation. It is shown that PLAP achieves superior or at least comparable performance against PLAP_d in 91.1% cases which clearly validates the benefits of modeling class dependencies in PLAP.
- BR, CC, SLEM, KRAM_d, KRAM_c and MDKNN train their models only over labeled MDC samples $\mathcal{D}_l$. It is worth noting that PLAP achieves superior or at least comparable performance against these six approaches in 98.9%, 98.9%, 92.2%, 92.2% and 100.0% cases, respectively. These experimental results clearly validate the benefits of utilizing unlabeled data in PLAP.
- For *WQpla.*, *WQani.*, *WQ* and *BeLaE*, it is shown that PLAP usually achieves comparable performance against PLAP_d. Possible reason is that the class dependencies in these MDC tasks are very weak. For example, there are 989 distinct class combinations appearing within the 1060 examples for *WQ* which also supports our conjecture to some extent.

4.3 Parameter Sensitivity Analysis

In the progressive label propagation procedure, for the current class space (except for the first one), PLAP utilizes the propagation results for previous class spaces to initialize the value of $\mathbf{Y}$ for unlabeled samples. Specifically, PLAP estimates the probability of several possible items in $\mathbf{Y}$ which are filtered by the predicted labels in previous label propagation via k nearest neighbors.

Figure 1 shows how the performance of PLAP fluctuates with different values of k. It is shown that PLAP achieves relatively stable performance when the value of k increases from 5 to 10. In this paper, the value of k is moderately set to 7 which can be used as the default parameter setting.

5 Conclusion

The major contributions of our work are two-fold: 1) Different from existing MDC works which aim at designing novel MDC approaches by assuming that enough labeled samples are available, we investigated the SSMDC problem which aims at utilizing unlabeled data to help induce MDC models with a few labeled samples and then labeling costs can be reduced. 2) We proposed a SSMDC approach named PLAP which can utilize both class dependencies and unlabeled data via progressive label propagation. Experiments clearly validate the effectiveness of the proposed PLAP approach.

PLAP works in the transductive learning schema where the test samples are exactly the unlabeled samples (i.e., closed-world assumption). In the future, it is interesting to further design SSMDC approaches working under open-world assumption [Zhou, 2022; Parmar *et al.*, 2023] where the test samples are completely unavailable to model training.

References

[Arias *et al.*, 2016] Jacinto Arias, José A. Gámez, Thomas D. Nielsen, and José Miguel Puerta. A scalable pairwise class interaction framework for multidimensional classification. *International Journal of Approximate Reasoning*, 68:194–210, 2016.

[Bielza *et al.*, 2011] Concha Bielza, Guangdi Li, and Pedro Larrañaga. Multi-dimensional classification with Bayesian networks. *International Journal of Approximate Reasoning*, 52(6):705–727, 2011.

[Borchani *et al.*, 2013] Hanen Borchani, Concha Bielza, Carlos Toro, and Pedro Larrañaga. Predicting human immunodeficiency virus inhibitors using multi-dimensional Bayesian network classifiers. *Artificial Intelligence in Medicine*, 57(3):219–229, 2013.

[Cambuí *et al.*, 2021] Brendon Gouveia Cambuí, Rafael Gomes Mantovani, and Ricardo Cerri. Exploring autoencoders for feature extraction in multi-target classification. In *Proceedings of the International Joint Conference on Neural Networks*, pages 1–8, Shenzhen, China, 2021.

[Demšar, 2006] Janez Demšar. Statistical comparisons of classifiers over multiple data sets. *Journal of Machine Learning Research*, 7:1–30, 2006.

[Gibaja and Ventura, 2015] Eva Gibaja and Sebastián Ventura. A tutorial on multilabel learning. *ACM Computing Surveys*, 47(3):Article 52, 2015.

[Gil-Begue *et al.*, 2021] Santiago Gil-Begue, Concha Bielza, and Pedro Larrañaga. Multi-dimensional Bayesian network classifiers: A survey. *Artificial Intelligence Review*, 54(1):519–559, 2021.

[Jia and Zhang, 2020a] Bin-Bin Jia and Min-Ling Zhang. Multi-dimensional classification via *k*NN feature augmentation. *Pattern Recognition*, 106:107423, 2020.

[Jia and Zhang, 2020b] Bin-Bin Jia and Min-Ling Zhang. Multi-dimensional classification via stacked dependency exploitation. *Science China Information Science*, 63(12):222102, 2020.

[Jia and Zhang, 2021a] Bin-Bin Jia and Min-Ling Zhang. MDKNN: An instance-based approach for multi-dimensional classification. In *Proceedings of the 25th International Conference on Pattern Recognition*, pages 126–133, Virtual Event / Milan, Italy, 2021.

[Jia and Zhang, 2021b] Bin-Bin Jia and Min-Ling Zhang. Multi-dimensional classification via sparse label encoding. In *Proceedings of the 38th International Conference on Machine Learning*, pages 4917–4926, Virtual Event, 2021.

[Jia and Zhang, 2022a] Bin-Bin Jia and Min-Ling Zhang. Decomposition-based classifier chains for multi-dimensional classification. *IEEE Transactions on Artificial Intelligence*, 3(2):176–191, 2022.

[Jia and Zhang, 2022b] Bin-Bin Jia and Min-Ling Zhang. Multi-dimensional classification via selective feature augmentation. *Machine Intelligence Research*, 19(1):38–51, 2022.

[Lian *et al.*, 2020] Zheng Lian, Ya Li, Jianhua Tao, Jian Huang, and Mingyue Niu. Expression analysis based on face regions in real-world conditions. *International Journal of Automation and Computing*, 17(1):96–107, 2020.

[Liu *et al.*, 2022] Weiwei Liu, Haobo Wang, Xiaobo Shen, and Ivor W. Tsang. The emerging trends of multi-label learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(11):7955–7974, 2022.

[Ma and Chen, 2018] Zhongchen Ma and Songcan Chen. Multi-dimensional classification via a metric approach. *Neurocomputing*, 275:1121–1131, 2018.

[Parmar *et al.*, 2023] Jitendra Parmar, Satyendra Singh Chouhan, Vaskar Raychoudhury, and Santosh Singh Rathore. Open-world machine learning: Applications, challenges, and opportunities. *ACM Computing Surveys*, 55(10):Article 205, 2023.

[Read *et al.*, 2014a] Jesse Read, Concha Bielza, and Pedro Larrañaga. Multi-dimensional classification with super-classes. *IEEE Transactions on Knowledge and Data Engineering*, 26(7):1720–1733, 2014.

[Read *et al.*, 2014b] Jesse Read, Luca Martino, and David Luengo. Efficient monte carlo methods for multi-dimensional learning with classifier chains. *Pattern Recognition*, 47(3):1535–1546, 2014.

[Shatkay *et al.*, 2008] Hagit Shatkay, Fengxia Pan, Andrey Rzhetsky, and W. John Wilbur. Multi-dimensional classification of biomedical text: Toward automated, practical provision of high-utility text to diverse users. *Bioinformatics*, 24(18):2086–2093, 2008.

[van Engelen and Hoos, 2020] Jesper E. van Engelen and Holger H. Hoos. A survey on semi-supervised learning. *Machine Learning*, 109(2):373–440, 2020.

[Wang *et al.*, 2020] Haobo Wang, Chen Chen, Weiwei Liu, Ke Chen, Tianlei Hu, and Gang Chen. Incorporating label embedding and feature augmentation for multi-dimensional classification. In *Proceedings of the 34th AAAI Conference on Artificial Intelligence*, pages 6178–6185, New York, NY, USA, 2020.

[Zaragoza *et al.*, 2011] Julio H. Zaragoza, Luis Enrique Sucar, Eduardo F. Morales, Concha Bielza, and Pedro Larrañaga. Bayesian chain classifiers for multidimensional classification. In *Proceedings of the 22nd International Joint Conference on Artificial Intelligence*, pages 2192–2197, Barcelona, Catalonia, Spain, 2011.

[Zhang and Zhou, 2014] Min-Ling Zhang and Zhi-Hua Zhou. A review on multi-label learning algorithms. *IEEE Transactions on Knowledge and Data Engineering*, 26(8):1819–1837, 2014.

[Zhang *et al.*, 2022] Wenbo Zhang, Yunhao Gou, Yuepeng Jiang, and Yu Zhang. Adversarial VAE with normalizing flows for multi-dimensional classification. In *Proceedings of the 5th Chinese Conference on Pattern Recognition and Computer Vision*, pages 205–219, Shenzhen, China, 2022.

[Zhou *et al.*, 2003] Dengyong Zhou, Olivier Bousquet, Thomas Navin Lal, Jason Weston, and Bernhard

Schölkopf. Learning with local and global consistency. In *Advances in Neural Information Processing Systems 16*, pages 321–328. MIT Press, 2003.

[Zhou, 2022] Zhi-Hua Zhou. Open-environment machine learning. *National Science Review*, 9(8):nwac123, 2022.