Information Sciences 689 (2025) 121473

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

Granular correlation-based label-specific feature augmentation for multi-label classification

Tianna Zhao [a,b,c], Yuanjian Zhang [d,*], Duoqian Miao [e,f]

[a] Institute of Artificial Intelligence on Education, Shanghai Normal University, 200234, China
[b] The Research Base of Online Education for Shanghai Middle and Primary Schools, Shanghai Normal University, 200234, China
[c] Shanghai Engineering Research Center of Intelligent Education and Big data, Shanghai Normal University, 200234, China
[d] School of Computer Engineering and Science, Shanghai University, 200444, China
[e] Department of Computer Science and Technology, Tongji University, Shanghai, 201804, China
[f] Key Laboratory of Embedded System and Service Computing, Ministry of Education, Tongji University, Shanghai, 201804, China

ARTICLE INFO

Keywords:
Multi-granularity
Uncertainty
Feature augmentation
Label correlation
Label-specific features
Multi-label classification

ABSTRACT

Multi-label classification is an extension of single-label classification with generations of multi-output for unseen instances. Label correlation is an essential component in constructing multi-label classifiers. How to optimize the representation of label correlation while preserving the semantics of label-specific remains an uncertain issue. Instead of estimating label correlation by a holistic feature representation, we present an augmented label correlation model by generating multi-granularity label-specific features. Firstly, we devise a mixture distance measure to characterize the closeness of an instance by weighing the Pearson correlation coefficient with cosine similarity. Secondly, we explore the local label-specific relative discrimination by leveraging from both the instance-level and class-level correlation distribution within k nearest neighborhood. Finally, we conduct an information fusion strategy to comprehensively integrate the positive and the negative tendencies at the neighborhood level. Instances with salient positive tendency and compact neighborhood structure receive larger values while receiving smaller values with salient negative tendency and sparse neighborhood structure. With the concatenation of original features and augmented features, we examine the classification performance of the proposed granule correlation-based feature augmentation (GOFA) on well-established second-order multi-label classification methods. Extensive comparisons on thirteen benchmarks demonstrate the statistical superiority of GOFA over state-of-the-art multi-label classifications.

1. Introduction

Multi-label classification [44,28,17] attempts to determine the label association of instance-label pairs by learning a mapping from features to labels. It is an extension of the single-label case in both the fashion of outputs and the semantics of associations. For the fashion of outputs, the cardinality of average associated labels is much smaller than the total count of labels. Meanwhile, the associated label count varies for different instances. For the semantics of associations, the associated labels are different in composition, and such differences imply a particular state of the scene. Such phenomena are omnipresent in real applications involving emotion analysis [13], video annotation [50], and medical diagnosis [34].

* Corresponding author.
E-mail address: zhangyuanjian@shu.edu.cn (Y. Zhang).

https://doi.org/10.1016/j.ins.2024.121473
Received 21 May 2024; Received in revised form 7 September 2024; Accepted 9 September 2024

Available online 13 September 2024

Label correlation is an essential component in regularizing effective feature representation. However, there are two issues with developing a robust multi-label classifier. Firstly, the composition of discriminative features is uncertain. Secondly, even if the desirable features are available, the relative importance of each label is uncertain as the composition of correlated labels is undetermined. For the first point, recent studies attempt to learn an optimal latent feature space. Zhang et al. [38] developed a bi-sparsity regularization by exploring underlying correlation in the label enrichment matrix. Zhang et al. [45] demonstrated explorations of relative implicit importance can boost the model's generalization. Qin et al. [24] proposed a multi-label feature selection method with adaptive graph learning on label enhancement. For the second issue, many studies [9,16,46,3] conduct the label-specific learning with second-order or high-order label correlation assumption. With label-specific assumptions, instances with similar associated labels have similar weights on relevant features. For example, the co-existence of blue and yellow color may imply a pairwise relationship between *sea* and *beach*, which yields a lower probability of other scenes such as *buildings* and *traffic*. However, multi-label annotations are not limited to specific representations on instances. Leveraging the fragmented information from different perspectives (e.g., similarity between instances, occlusion between objects) can potentially imply a robust correlation, which is conducive to correcting the bias from original features. Unfortunately, such a decision-making process is beyond the scope of the previous studies.

Granular computing (GrC) [8,23] is a structural cognitive methodology that simulates human processing of problems with uncertainty. With information granulation, the underlying problem structure is approximately represented. Many scholars [47,22,1,31] demonstrate the superiority of employing granular computing on multi-label classification. With granular computing, it is possible to determine the label association by integrating it with information from different granular layers. Particularly, the feature augmentation technique [11,12] is competent for the granular layer transformation. To simulate the adaptive utilization of multifaceted information, we devise a multi-granulation feature augmentation model to explore the local intrinsic relationship within instances. With the refinement of granularity, the integrations of essential factors like the components of neighborhoods, the instance-level similarity, the positive and negative class distribution, and the compactness of neighborhoods boost the discrimination of labels. These high-level features can be regarded as supplementary to original features. Our contributions are enumerated below:

1. We construct a granular structure to enrich the feature representation of multi-label by exploiting both instance-level and class-level correlation distribution. The augmented features are intuitive in semantics, revealing the strength of relative positive tendency locally.
2. The feature augmentation generation takes a double-quantitative viewpoint in measuring the positive and negative influential degree without introducing additional parameters. Accordingly, the contribution of each instance within k nearest neighborhood is automatically determined.
3. We present a novel model for multi-label classification called granule correlation-based feature augmentation (GOFA). The superiority of GOFA over state-of-the-art algorithms is demonstrated.

The remaining parts are organized as follows: Section 2 reviews the related work. Section 3 outlines the pipeline and explains model details. Section 4 presents experimental results on benchmarks. Finally, Section 5 concludes the paper.

2. Related work

2.1. Label correlation

The label correlations describe the statistical dependency information among the high-level concepts (i.e., labels), thus adjusting the feature weights via feature-label dependency. However, the scope and space for a trustworthy label correlation is previously unknown. To this challenge, many studies assume a particular order of label correlation. The solutions include first-order, second-order, or high-order. Typically, the first-order (e.g., ML-kNN [42] and LIFT [43]) learns each label without considering correlations among labels, while the second-order (e.g., MGT-LEML [49] and FIGR_LC [35]) and high-order (e.g., MLR [21] and FL_MLC [2]) exploits pairwise or high-order fashion of label correlations. In contrast, the construction of appropriate feature space is the extension of feature selection and feature extraction for original and latent space, respectively. In particular, the embedding space works for noisy (e.g., PML-NI [32] and nEM [4]) or missing labels cases (e.g., C2ML [29] and MML-TLNM [30]).

2.2. Feature augmentation

Feature augmentation exploits the implicit relationships of instances from the explicit original features and is an informative component in developing a classifier. Recently, many multi-granularity-based feature augmentations have been examined in the single-label case. Gong et al. [6] constructed a local attention-guided network to integrate the information from both the most salient area of a person and the body parts for person re-identification. The body parts features were extracted by a region-interest-map generator and augmented by a residual structure. Guo et al. [7] developed a multiple granularity semantic learning network for the image-based scene classification of remote sensing. The multi-scale region features were extracted by a multiple granularity semantic learning module and enhanced via a granular mapping operator. Zhang et al. [39] formulated a feature enhancement network for small object detection. A multi-granular deformable convolution network was devised to capture and fuse the offset feature representation in different granularity. Xue et al. [33] proposed a multi-granularity relational augmentation network to optimize knowledge graph embedding. A relational augmentation convolution module preserved the distinctive relational properties and enriched the feature informativeness afterward.

Table 1
Summary of notations.

Notations	Meanings
$\mathbf{X}$	Original training set with n instances and m features
$\mathbf{Y}$	Labelset of $\mathbf{X}$ on q labels
$Dis(\cdot,\cdot)$	Instance-based distance operator
μ	weight of cosine similarity and Pearson correlation coefficient in instance-based distance
k	size of nearest neighbor $\forall \mathbf{x}_b$
$\mathcal{I}(\mathbf{x}_b)$	Indexes of k nearest neighborhood for $\mathbf{x}_b$
$\mathcal{P}_j^{\mathbf{x}_b}$	Indicators of the nearest neighborhood with label association on the jth label for $\mathbf{x}_b$
$\mathcal{N}_j^{\mathbf{x}_b}$	Indicators of the nearest neighborhood without label association on the jth label for $\mathbf{x}_b$
$\mathbf{W}^{\mathbf{x}_b}$	Instance weights of k nearest neighborhood for $\mathbf{x}_b$
$rip_j^{\mathbf{x}_b}$	Relative influential ratio of positive instances within neighborhood of $\mathbf{x}_b$ based on distance
$rin_j^{\mathbf{x}_b}$	Relative influential ratio of negative instances within neighborhood of $\mathbf{x}_b$ based on distance
$\mathbf{rip}_j$	Relative influential ratio of positive instances based on distance
$\mathbf{rin}_j$	Relative influential ratio of negative instances based on distance
$WSP_j^{\mathbf{x}_b}$	Weighted score of positive tendency for instances in neighborhood of $\mathbf{x}_b$ on label l_j
$WSN_j^{\mathbf{x}_b}$	Weighted score of negative tendency for instances in neighborhood of $\mathbf{x}_b$ on label l_j
$\mathbf{WSP}_j$	Weighted score vectors for positive tendency of instances on label l_j
$\mathbf{WSN}_j$	Weighted score vectors for negative tendency of instances on label l_j
$\mathbf{AF}_j$	Augmented feature for label l_j
$\mathbf{X}_{aug}$	Multigranularity features
$\mathbf{V}$	Embedding matrix for generation of multi-granularity features
$\mathbf{U}$	Weight matrix for multi-granularity features
$\lvert S\rvert$	Instance count in benchmark
$dim(S)$	Feature dimensionality in benchmark
$L(S)$	Label count in benchmark
$LCard(S)$	Cardinality of average associated labels per instance

2.3. Granular computing

Granular computing provides an uncertainty-driven method for data processing by identifying and inferring the homogeneous/heterogeneous correlations among multifaceted granules. Relative values serve as prototypes of correlations and thus become an essential component in granulation. Zhao et al. [48] introduced a relation matrix to explore the relative information measure across granularity. A unified framework describes the coarser-finer relation among granules, and the granularity itself is interpreted as the sum of the relative information measure of its atomic granules. Liang et al. [15] developed a group of relative-value-based loss functions for decision-theoretic rough sets. The ambiguous relative ratios between loss functions are determined via an analytic hierarchy process for pairwise misclassification cost. Li et al. [14] explored the acceleration of attribute reduction by defining the granularity space. With the unified representation of granularity space, a granularity search strategy is developed by focusing on relative significance variation instead of searching the core attributes. Huang et al. [10] developed a tournament selection operator-guided particle swarm optimization to address the linguistic information granulation for multi-attribute group decision-making. The relative importance of experts is obtained by balancing the pairwise similarity and group-based consensus on each attribute. Zhan et al. [37] investigated the triangular fuzzy number-based tripartition on an incomplete multiscale information system. With regret-rejoicing values measured by utility difference, the thresholds of three-way decision are relatively defined.

3. Proposed model

3.1. Notations

Given a multi-label dataset $\mathcal{D} = \{(\mathbf{x}_i, Y_i \,|\, 1 \leqslant i \leqslant n)\}$, let $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n]^\top \in \mathbb{R}^{n\times m}$ denote the n instances with m-dimensional features and $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_n]^\top \in \{0,1\}^{n\times q}$ denote the q-dimensional labels on n instances, where $\mathbf{y}_i = [y_{i1}, y_{i2}, \ldots, y_{iq}]^\top$. $y_{ij} = 1$ if $l_j \in Y_i$ and $y_{ij} = 0$ otherwise. GOFA generates the label-specific features for each label from observable features. For readability, we enumerate major notations in Table 1.

3.2. Basic idea

We optimize the label-specific feature representations by learning discriminative but limited[1] features. The granularity in our procedures refers to the hierarchy of feature structure. Given that the accurate classification on each label boosts the overall performance, we develop augmented features with q-dimensionality. As depicted in Fig. 1, the feature augmentation procedures are sequentially refined in a label-by-label manner. Without losing generality, we explain how to deduce the jth augmented feature $\mathbf{AF}_j \in \mathbb{R}^{n\times 1}$ for label l_j. Concretely, starting with an appropriate instance similarity measure, the feature prototype of l_j is automatically determined

[1] The phrase limited means the count of augmented features is much smaller than that of original feature dimensionality.

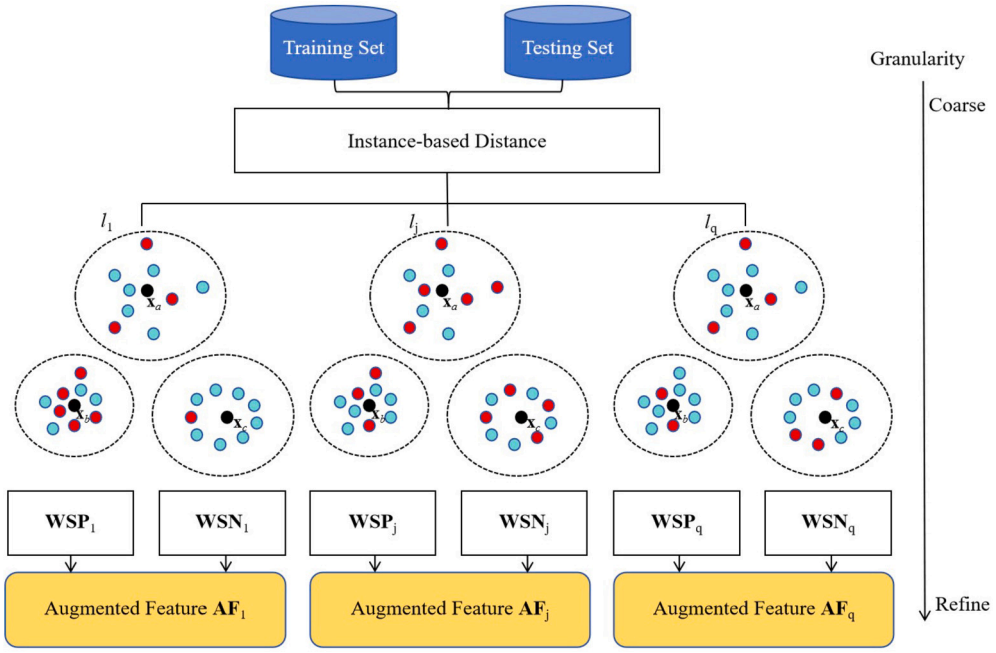

Fig. 1. Framework of GOFA: The j-th augmented feature $\mathbf{AF}_j$ considers both the neighborhood and label association difference on l_j, where $\mathbf{x}_a, \mathbf{x}_b, \mathbf{x}_c$ are three arbitrary instances from Training/Testing set. The red and blue dots represent the positive and negative label association, while the dashed circles represent the k-nearest neighborhood.

via k-nearest neighborhood w.r.t. an arbitrary instance (see $\mathbf{x}_b$ in Fig. 1). By combining the qualitative positive and negative instance distribution with the quantitative similar-aware weight of positive/negative instances, we refine the prototypes with the overall weighted scores for positive and negative, respectively (see $\mathbf{WSP}_j$ and $\mathbf{WSN}_j$ in Fig. 1). The two weighted scores are two prototypes for measuring relative class-dependent tendency, which is the implementation of granularity from different perspectives. Ultimately, the augmented features are a concatenation of fractions from all labels $l_j, j = 1, 2, \ldots, q$, where each component is the fusion of positive and negative tendencies. Accordingly, we have a fine-grained representation of features by concatenating the augmented features.

3.3. Granular label-specific feature augmentation

3.3.1. k nearest neighborhood generation

The instance similarity measures how many degrees an instance contributes to the classifier. We introduce k-nearest neighborhood to characterize the local similarity between pairwise instances. For each instance $\mathbf{x}_b$, let $\mathcal{I}(\mathbf{x}_b) = \{\mathbf{x}_r \,|\, Dis(\mathbf{x}_b, \mathbf{x}_r) \in Dis_k(\mathbf{x}_b)\}$ denote the set of indexes identified in k-nearest neighborhood of $\mathbf{x}_b \in \mathbf{X}$, where $Dis_k(\mathbf{x}_b)$ denotes the set of the k-smallest elements measured by the instance similarity operator $Dis(\cdot,\cdot)$. $Dis(\cdot,\cdot)$ is a mixture of Pearson correlation coefficient and cosine similarity.[2]

$$Dis(\mathbf{x}_b, \mathbf{x}_r) = \mu\left(\frac{1 - \frac{Cov(\mathbf{x}_b, \mathbf{x}_r)}{\sigma_{\mathbf{x}_b}\sigma_{\mathbf{x}_r}}}{2}\right) + (1-\mu)\left(\frac{1 - \frac{\mathbf{x}_b\mathbf{x}_r}{\|\mathbf{x}_b\|\|\mathbf{x}_r\|}}{2}\right) \tag{1}$$

Where $\mu \in [0,1]$, $\frac{Cov(\mathbf{x}_b,\mathbf{x}_r)}{\sigma_{\mathbf{x}_b}\sigma_{\mathbf{x}_r}}$ and $\frac{\mathbf{x}_b\mathbf{x}_r}{\|\mathbf{x}_b\|\|\mathbf{x}_r\|}$ represent Pearson correlation coefficient and cosine similarity, respectively, μ is a trade-off factor that balances the contributions of the Pearson similarity coefficient and cosine similarity in the sense of distance. Since both $\frac{Cov(\mathbf{x}_b,\mathbf{x}_r)}{\sigma_{\mathbf{x}_b}\sigma_{\mathbf{x}_r}} \in [-1,1]$ and $\frac{\mathbf{x}_b\mathbf{x}_r}{\|\mathbf{x}_b\|\|\mathbf{x}_r\|} \in [-1,1]$ hold, it is straightforward to deduce that $Dis(\cdot,\cdot) \in [0,1]$. The smaller the value of $Dis(\cdot,\cdot)$, the more similar the $\mathbf{x}_r$ for the given $\mathbf{x}_b$.

Example. Given a multi-label survey dataset on credit card issuance [1] in Table 2 with the instance count $n = 9$ representing nine customers, feature dimensionality $m = 6$ representing the relative values of spending power, risky preference, deposit, luxury consumption, card activity, and repayment cycle, and label count $q = 3$ representing three financial products. The customer $\mathbf{x}_i$ will purchase the jth product if $y_{ij} = 1$, otherwise $y_{ij} = 0$.

[2] We consider the weighted form instead of applying a particular distance to accommodate data characteristics. Here, we focus on the relative contribution of linear correlation strength and direction consistency.

Table 2
A multi-label user survey for product recommendation.

$\mathbf{X}$	c_1	c_2	c_3	c_4	c_5	c_6	l_1	l_2	l_3
x_1	0.8	0.1	0.1	0.5	0.2	0.3	1	0	1
x_2	0.3	0.5	0.2	0.8	0.1	0.1	1	0	0
x_3	0.2	0.2	0.6	0.7	0.3	0.2	0	0	1
x_4	0.6	0.3	0.1	0.2	0.5	0.3	1	0	1
x_5	0.3	0.4	0.3	0.3	0.6	0.1	0	1	0
x_6	0.2	0.3	0.5	0.3	0.5	0.2	1	0	0
x_7	0.3	0.3	0.4	0.2	0.6	0.2	0	1	1
x_8	0.3	0.4	0.3	0.1	0.4	0.5	0	1	0
x_9	0.3	0.2	0.5	0.4	0.4	0.2	1	0	0

Let $\mu = 0.5$, $k = 5$ then the distance operator $Dis(\cdot,\cdot)$ is computed by Equ (1) as:

$$Dis(\cdot,\cdot) = \begin{bmatrix} 0.0000 & 0.2548 & 0.3499 & 0.1425 & 0.3963 & 0.5016 & 0.4294 & 0.4368 & 0.3399 \\ 0.2548 & 0.0000 & 0.1607 & 0.4175 & 0.3266 & 0.3882 & 0.4742 & 0.5502 & 0.3040 \\ 0.3499 & 0.1607 & 0.0000 & 0.5147 & 0.3064 & 0.1688 & 0.3217 & 0.5306 & 0.0672 \\ 0.1425 & 0.4175 & 0.5147 & 0.0000 & 0.1927 & 0.3749 & 0.2034 & 0.2155 & 0.3815 \\ 0.3963 & 0.3266 & 0.3064 & 0.1927 & 0.0000 & 0.0984 & 0.0462 & 0.3033 & 0.1883 \\ 0.5016 & 0.3882 & 0.1688 & 0.3749 & 0.0984 & 0.0000 & 0.0539 & 0.3008 & 0.0653 \\ 0.4294 & 0.4742 & 0.3217 & 0.2034 & 0.0462 & 0.0539 & 0.0000 & 0.2192 & 0.1475 \\ 0.4368 & 0.5502 & 0.5306 & 0.2155 & 0.3033 & 0.3008 & 0.2192 & 0.0000 & 0.4343 \\ 0.3399 & 0.3040 & 0.0672 & 0.3815 & 0.1883 & 0.0653 & 0.1475 & 0.4343 & 0.0000 \end{bmatrix} \quad \square$$

3.3.2. Label-specific tendency estimation

With well-defined k-nearest neighborhood structures, we can estimate the positive/negative tendency on an arbitrary label from both qualitative and quantitative perspectives. The positive/negative tendency estimation follows the following assumptions:

(a) (Qualitative assumption) The k-nearest neighborhood of $\mathbf{x}_b$ exhibits stronger positive tendency on l_j if there are more instances with label l_j, and vice versa.
(b) (Quantitative assumption) The k-nearest neighborhood of $\mathbf{x}_b$ exhibits stronger positive tendency on l_j if the instances that are more similar to $\mathbf{x}_b$ are with label l_j, and vice versa.

We realize the qualitative assumption by calculating the instance count with positive/negative class on l_j. Suppose the distance between $\mathbf{x}_b$ and an arbitrary $\mathbf{x}_r$ ranks in ascending order according to $Dis(\cdot,\cdot)$, where $r = 1, 2, \ldots, n$. The indicators $\mathcal{P}_j^{\mathbf{x}_r}$ and $\mathcal{N}_j^{\mathbf{x}_r}$ are defined as:

$$\mathcal{P}_j^{\mathbf{x}_b} = \left([\![y_{rj} = 1]\!] \right)_{k\times 1} \tag{2}$$

where $\mathbf{x}_r \in \mathcal{I}(\mathbf{x}_b)$, $[\![\pi]\!]$ returns 1 if it holds and 0 otherwise. Similarly, we have:

$$\mathcal{N}_j^{\mathbf{x}_b} = \left([\![y_{rj} = 0]\!] \right)_{k\times 1} \tag{3}$$

where $\mathbf{x}_r \in \mathcal{I}(\mathbf{x}_b)$, $[\![\pi]\!]$ returns 1 if it holds and 0 otherwise.

For the quantitative assumption, we consider both instance similarity weight and relative influential ratio. The instance similarity weight measures how much percentage an instance $\mathbf{x}_b$ will be represented by an arbitrary instance $\mathbf{x}_r$. Formally, we devise a similar-aware weight vector (i.e., $\mathbf{W}^{\mathbf{x}_b}$) in computing varying $\mathcal{I}(\mathbf{x}_b)$.

$$\mathbf{W}^{\mathbf{x}_b} = \left(\sqrt{1 - Dis(\mathbf{x}_b, \mathbf{x}_r)} \right)_{k\times 1} \tag{4}$$

where $\mathbf{x}_r \in \mathcal{I}(\mathbf{x}_b)$. It is straightforward to deduce the component $w_r^{\mathbf{x}_b} \in [0,1]$, with the larger weight on $\mathbf{x}_r$ than $\mathbf{x}_s$ ($w_r^{\mathbf{x}_b} > w_s^{\mathbf{x}_b}$) if $\mathbf{x}_r$ is more similar to $\mathbf{x}_b$ than $\mathbf{x}_s$ and vice versa.[3] In such settings, the instance weight is defined in a data-driven fashion for the neighborhood of $\mathbf{x}_b$ (i.e., $\mathcal{I}(\mathbf{x}_b)$) and $\mathbf{x}_c$ (i.e., $\mathcal{I}(\mathbf{x}_c)$). Typically, $w_r^{\mathbf{x}_b}$ reaches maximum value of 1 if the original features of $\mathbf{x}_b$ and $\mathbf{x}_r$ are indistinguishable.

Example *(Continued)*. For the survey dataset mentioned in Table 2, we obtain the similar-aware weight vector $\mathbf{W}$ by firstly finding all weights as:

[3] For generalization, the instance itself is excluded in the corresponding k nearest neighborhood.

$$\left(\sqrt{1-Dis\left(\cdot,\cdot\right)}\right)_{9\times9}=\begin{bmatrix}1.0000 & 0.8632 & 0.8063 & 0.9260 & 0.7770 & 0.7060 & 0.7554 & 0.7505 & 0.8124\\ 0.8632 & 1.0000 & 0.9161 & 0.7632 & 0.8206 & 0.7822 & 0.7251 & 0.6707 & 0.8342\\ 0.8063 & 0.9161 & 1.0000 & 0.6966 & 0.8329 & 0.9117 & 0.8236 & 0.6851 & 0.9658\\ 0.9260 & 0.7632 & 0.6966 & 1.0000 & 0.8985 & 0.7906 & 0.8925 & 0.8857 & 0.7864\\ 0.7770 & 0.8206 & 0.8329 & 0.8985 & 1.0000 & 0.9495 & 0.9766 & 0.8347 & 0.9010\\ 0.7060 & 0.7822 & 0.9117 & 0.7906 & 0.9495 & 1.0000 & 0.9727 & 0.8362 & 0.9668\\ 0.7554 & 0.7251 & 0.8236 & 0.8925 & 0.9766 & 0.9727 & 1.0000 & 0.8836 & 0.9233\\ 0.7505 & 0.6707 & 0.6851 & 0.8857 & 0.8347 & 0.8362 & 0.8836 & 1.0000 & 0.7521\\ 0.8124 & 0.8342 & 0.9658 & 0.7864 & 0.9010 & 0.9668 & 0.9233 & 0.7521 & 1.0000\end{bmatrix}$$

By comparing the values of weights, we have the components of $\mathbf{W}$ by following Equ (4) as:

$$\mathbf{W}_{9\times5}=\begin{bmatrix}0.9260 & 0.8632 & 0.8124 & 0.8063 & 0.7770\\ 0.9161 & 0.8632 & 0.8342 & 0.8206 & 0.7822\\ 0.9658 & 0.9161 & 0.9117 & 0.8329 & 0.8236\\ 0.9260 & 0.8985 & 0.8925 & 0.8857 & 0.7906\\ 0.9766 & 0.9495 & 0.9010 & 0.8985 & 0.8347\\ 0.9727 & 0.9668 & 0.9495 & 0.9117 & 0.8362\\ 0.9766 & 0.9727 & 0.9233 & 0.8925 & 0.8836\\ 0.8857 & 0.8836 & 0.8362 & 0.8347 & 0.7521\\ 0.9668 & 0.9658 & 0.9233 & 0.9010 & 0.8342\end{bmatrix}$$

where the corresponding neighborhood $\mathcal{I}\left(\mathbf{X}\right)$ is:

$$\mathcal{I}\left(\mathbf{X}\right)_{9\times5}=\begin{bmatrix}\mathcal{I}\left(\mathbf{x}_1\right)\\ \mathcal{I}\left(\mathbf{x}_2\right)\\ \mathcal{I}\left(\mathbf{x}_3\right)\\ \mathcal{I}\left(\mathbf{x}_4\right)\\ \mathcal{I}\left(\mathbf{x}_5\right)\\ \mathcal{I}\left(\mathbf{x}_6\right)\\ \mathcal{I}\left(\mathbf{x}_7\right)\\ \mathcal{I}\left(\mathbf{x}_8\right)\\ \mathcal{I}\left(\mathbf{x}_9\right)\end{bmatrix}=\begin{bmatrix}\mathbf{x}_4 & \mathbf{x}_2 & \mathbf{x}_9 & \mathbf{x}_3 & \mathbf{x}_5\\ \mathbf{x}_3 & \mathbf{x}_1 & \mathbf{x}_9 & \mathbf{x}_5 & \mathbf{x}_6\\ \mathbf{x}_9 & \mathbf{x}_2 & \mathbf{x}_6 & \mathbf{x}_5 & \mathbf{x}_7\\ \mathbf{x}_1 & \mathbf{x}_5 & \mathbf{x}_7 & \mathbf{x}_8 & \mathbf{x}_6\\ \mathbf{x}_7 & \mathbf{x}_6 & \mathbf{x}_9 & \mathbf{x}_4 & \mathbf{x}_8\\ \mathbf{x}_7 & \mathbf{x}_9 & \mathbf{x}_5 & \mathbf{x}_3 & \mathbf{x}_8\\ \mathbf{x}_5 & \mathbf{x}_6 & \mathbf{x}_9 & \mathbf{x}_4 & \mathbf{x}_8\\ \mathbf{x}_4 & \mathbf{x}_7 & \mathbf{x}_6 & \mathbf{x}_5 & \mathbf{x}_9\\ \mathbf{x}_6 & \mathbf{x}_3 & \mathbf{x}_7 & \mathbf{x}_5 & \mathbf{x}_2\end{bmatrix}$$

Thus, the neighborhood-based label indicator $\mathcal{P}$ is computed by Equ (2) as:

$$\mathcal{P}_{9\times3}=\begin{bmatrix}\mathcal{P}_1^{\mathbf{x}_1} & \mathcal{P}_2^{\mathbf{x}_1} & \mathcal{P}_3^{\mathbf{x}_1}\\ \mathcal{P}_1^{\mathbf{x}_2} & \mathcal{P}_2^{\mathbf{x}_2} & \mathcal{P}_3^{\mathbf{x}_2}\\ \mathcal{P}_1^{\mathbf{x}_3} & \mathcal{P}_2^{\mathbf{x}_3} & \mathcal{P}_3^{\mathbf{x}_3}\\ \mathcal{P}_1^{\mathbf{x}_4} & \mathcal{P}_2^{\mathbf{x}_4} & \mathcal{P}_3^{\mathbf{x}_4}\\ \mathcal{P}_1^{\mathbf{x}_5} & \mathcal{P}_2^{\mathbf{x}_5} & \mathcal{P}_3^{\mathbf{x}_5}\\ \mathcal{P}_1^{\mathbf{x}_6} & \mathcal{P}_2^{\mathbf{x}_6} & \mathcal{P}_3^{\mathbf{x}_6}\\ \mathcal{P}_1^{\mathbf{x}_7} & \mathcal{P}_2^{\mathbf{x}_7} & \mathcal{P}_3^{\mathbf{x}_7}\\ \mathcal{P}_1^{\mathbf{x}_8} & \mathcal{P}_2^{\mathbf{x}_8} & \mathcal{P}_3^{\mathbf{x}_8}\\ \mathcal{P}_1^{\mathbf{x}_9} & \mathcal{P}_2^{\mathbf{x}_9} & \mathcal{P}_3^{\mathbf{x}_9}\end{bmatrix}$$

where

$$\begin{bmatrix}\mathcal{P}_1^{\mathbf{x}_1}\\ \mathcal{P}_1^{\mathbf{x}_2}\\ \mathcal{P}_1^{\mathbf{x}_3}\\ \mathcal{P}_1^{\mathbf{x}_4}\\ \mathcal{P}_1^{\mathbf{x}_5}\\ \mathcal{P}_1^{\mathbf{x}_6}\\ \mathcal{P}_1^{\mathbf{x}_7}\\ \mathcal{P}_1^{\mathbf{x}_8}\\ \mathcal{P}_1^{\mathbf{x}_9}\end{bmatrix}=\begin{bmatrix}[\![y_{41}=1]\!] & [\![y_{21}=1]\!] & [\![y_{91}=1]\!] & [\![y_{31}=1]\!] & [\![y_{51}=1]\!]\\ [\![y_{31}=1]\!] & [\![y_{11}=1]\!] & [\![y_{91}=1]\!] & [\![y_{51}=1]\!] & [\![y_{61}=1]\!]\\ [\![y_{91}=1]\!] & [\![y_{21}=1]\!] & [\![y_{61}=1]\!] & [\![y_{51}=1]\!] & [\![y_{71}=1]\!]\\ [\![y_{11}=1]\!] & [\![y_{51}=1]\!] & [\![y_{71}=1]\!] & [\![y_{81}=1]\!] & [\![y_{61}=1]\!]\\ [\![y_{71}=1]\!] & [\![y_{61}=1]\!] & [\![y_{91}=1]\!] & [\![y_{41}=1]\!] & [\![y_{81}=1]\!]\\ [\![y_{71}=1]\!] & [\![y_{91}=1]\!] & [\![y_{51}=1]\!] & [\![y_{31}=1]\!] & [\![y_{81}=1]\!]\\ [\![y_{51}=1]\!] & [\![y_{61}=1]\!] & [\![y_{91}=1]\!] & [\![y_{41}=1]\!] & [\![y_{81}=1]\!]\\ [\![y_{41}=1]\!] & [\![y_{71}=1]\!] & [\![y_{61}=1]\!] & [\![y_{51}=1]\!] & [\![y_{91}=1]\!]\\ [\![y_{61}=1]\!] & [\![y_{31}=1]\!] & [\![y_{71}=1]\!] & [\![y_{51}=1]\!] & [\![y_{21}=1]\!]\end{bmatrix}=\begin{bmatrix}1 & 1 & 1 & 0 & 0\\ 0 & 1 & 1 & 0 & 1\\ 1 & 1 & 1 & 0 & 0\\ 1 & 0 & 0 & 0 & 1\\ 0 & 1 & 1 & 1 & 0\\ 0 & 1 & 0 & 0 & 0\\ 0 & 1 & 1 & 1 & 0\\ 1 & 0 & 1 & 0 & 1\\ 1 & 0 & 0 & 0 & 1\end{bmatrix}$$

$$\begin{bmatrix} \mathcal{P}_2^{\mathbf{x}_1} \\ \mathcal{P}_2^{\mathbf{x}_2} \\ \mathcal{P}_2^{\mathbf{x}_3} \\ \mathcal{P}_2^{\mathbf{x}_4} \\ \mathcal{P}_2^{\mathbf{x}_5} \\ \mathcal{P}_2^{\mathbf{x}_6} \\ \mathcal{P}_2^{\mathbf{x}_7} \\ \mathcal{P}_2^{\mathbf{x}_8} \\ \mathcal{P}_2^{\mathbf{x}_9} \end{bmatrix} = \begin{bmatrix} [\![y_{42}=1]\!] & [\![y_{22}=1]\!] & [\![y_{92}=1]\!] & [\![y_{32}=1]\!] & [\![y_{52}=1]\!] \\ [\![y_{32}=1]\!] & [\![y_{12}=1]\!] & [\![y_{92}=1]\!] & [\![y_{52}=1]\!] & [\![y_{62}=1]\!] \\ [\![y_{92}=1]\!] & [\![y_{22}=1]\!] & [\![y_{62}=1]\!] & [\![y_{52}=1]\!] & [\![y_{72}=1]\!] \\ [\![y_{12}=1]\!] & [\![y_{52}=1]\!] & [\![y_{72}=1]\!] & [\![y_{82}=1]\!] & [\![y_{62}=1]\!] \\ [\![y_{72}=1]\!] & [\![y_{62}=1]\!] & [\![y_{92}=1]\!] & [\![y_{42}=1]\!] & [\![y_{82}=1]\!] \\ [\![y_{72}=1]\!] & [\![y_{92}=1]\!] & [\![y_{52}=1]\!] & [\![y_{32}=1]\!] & [\![y_{82}=1]\!] \\ [\![y_{52}=1]\!] & [\![y_{62}=1]\!] & [\![y_{92}=1]\!] & [\![y_{42}=1]\!] & [\![y_{82}=1]\!] \\ [\![y_{42}=1]\!] & [\![y_{72}=1]\!] & [\![y_{62}=1]\!] & [\![y_{52}=1]\!] & [\![y_{92}=1]\!] \\ [\![y_{62}=1]\!] & [\![y_{32}=1]\!] & [\![y_{72}=1]\!] & [\![y_{52}=1]\!] & [\![y_{22}=1]\!] \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 \end{bmatrix}$$

$$\begin{bmatrix} \mathcal{P}_3^{\mathbf{x}_1} \\ \mathcal{P}_3^{\mathbf{x}_2} \\ \mathcal{P}_3^{\mathbf{x}_3} \\ \mathcal{P}_3^{\mathbf{x}_4} \\ \mathcal{P}_3^{\mathbf{x}_5} \\ \mathcal{P}_3^{\mathbf{x}_6} \\ \mathcal{P}_3^{\mathbf{x}_7} \\ \mathcal{P}_3^{\mathbf{x}_8} \\ \mathcal{P}_3^{\mathbf{x}_9} \end{bmatrix} = \begin{bmatrix} [\![y_{43}=1]\!] & [\![y_{23}=1]\!] & [\![y_{93}=1]\!] & [\![y_{33}=1]\!] & [\![y_{53}=1]\!] \\ [\![y_{33}=1]\!] & [\![y_{13}=1]\!] & [\![y_{93}=1]\!] & [\![y_{53}=1]\!] & [\![y_{63}=1]\!] \\ [\![y_{93}=1]\!] & [\![y_{23}=1]\!] & [\![y_{63}=1]\!] & [\![y_{53}=1]\!] & [\![y_{73}=1]\!] \\ [\![y_{13}=1]\!] & [\![y_{53}=1]\!] & [\![y_{73}=1]\!] & [\![y_{83}=1]\!] & [\![y_{63}=1]\!] \\ [\![y_{73}=1]\!] & [\![y_{63}=1]\!] & [\![y_{93}=1]\!] & [\![y_{43}=1]\!] & [\![y_{83}=1]\!] \\ [\![y_{73}=1]\!] & [\![y_{93}=1]\!] & [\![y_{53}=1]\!] & [\![y_{33}=1]\!] & [\![y_{83}=1]\!] \\ [\![y_{53}=1]\!] & [\![y_{63}=1]\!] & [\![y_{93}=1]\!] & [\![y_{43}=1]\!] & [\![y_{83}=1]\!] \\ [\![y_{43}=1]\!] & [\![y_{73}=1]\!] & [\![y_{63}=1]\!] & [\![y_{53}=1]\!] & [\![y_{93}=1]\!] \\ [\![y_{63}=1]\!] & [\![y_{33}=1]\!] & [\![y_{73}=1]\!] & [\![y_{53}=1]\!] & [\![y_{23}=1]\!] \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \end{bmatrix}$$

Analogously, the neighborhood-based label indicator $\mathcal{P}$ is computed by Equ (3) as:

$$\mathcal{N}_{9\times 3} = \begin{bmatrix} \mathcal{N}_1^{\mathbf{x}_1} & \mathcal{N}_2^{\mathbf{x}_1} & \mathcal{N}_3^{\mathbf{x}_1} \\ \mathcal{N}_1^{\mathbf{x}_2} & \mathcal{N}_2^{\mathbf{x}_2} & \mathcal{N}_3^{\mathbf{x}_2} \\ \mathcal{N}_1^{\mathbf{x}_3} & \mathcal{N}_2^{\mathbf{x}_3} & \mathcal{N}_3^{\mathbf{x}_3} \\ \mathcal{N}_1^{\mathbf{x}_4} & \mathcal{N}_2^{\mathbf{x}_4} & \mathcal{N}_3^{\mathbf{x}_4} \\ \mathcal{N}_1^{\mathbf{x}_5} & \mathcal{N}_2^{\mathbf{x}_5} & \mathcal{N}_3^{\mathbf{x}_5} \\ \mathcal{N}_1^{\mathbf{x}_6} & \mathcal{N}_2^{\mathbf{x}_6} & \mathcal{N}_3^{\mathbf{x}_6} \\ \mathcal{N}_1^{\mathbf{x}_7} & \mathcal{N}_2^{\mathbf{x}_7} & \mathcal{N}_3^{\mathbf{x}_7} \\ \mathcal{N}_1^{\mathbf{x}_8} & \mathcal{N}_2^{\mathbf{x}_8} & \mathcal{N}_3^{\mathbf{x}_8} \\ \mathcal{N}_1^{\mathbf{x}_9} & \mathcal{N}_2^{\mathbf{x}_9} & \mathcal{N}_3^{\mathbf{x}_9} \end{bmatrix}$$

where

$$\begin{bmatrix} \mathcal{N}_1^{\mathbf{x}_1} \\ \mathcal{N}_1^{\mathbf{x}_2} \\ \mathcal{N}_1^{\mathbf{x}_3} \\ \mathcal{N}_1^{\mathbf{x}_4} \\ \mathcal{N}_1^{\mathbf{x}_5} \\ \mathcal{N}_1^{\mathbf{x}_6} \\ \mathcal{N}_1^{\mathbf{x}_7} \\ \mathcal{N}_1^{\mathbf{x}_8} \\ \mathcal{N}_1^{\mathbf{x}_9} \end{bmatrix} = \begin{bmatrix} [\![y_{41}=0]\!] & [\![y_{21}=0]\!] & [\![y_{91}=0]\!] & [\![y_{31}=0]\!] & [\![y_{51}=0]\!] \\ [\![y_{31}=0]\!] & [\![y_{11}=0]\!] & [\![y_{91}=0]\!] & [\![y_{51}=0]\!] & [\![y_{61}=0]\!] \\ [\![y_{91}=0]\!] & [\![y_{21}=0]\!] & [\![y_{61}=0]\!] & [\![y_{51}=0]\!] & [\![y_{71}=0]\!] \\ [\![y_{11}=0]\!] & [\![y_{51}=0]\!] & [\![y_{71}=0]\!] & [\![y_{81}=0]\!] & [\![y_{61}=0]\!] \\ [\![y_{71}=0]\!] & [\![y_{61}=0]\!] & [\![y_{91}=0]\!] & [\![y_{41}=0]\!] & [\![y_{81}=0]\!] \\ [\![y_{71}=0]\!] & [\![y_{91}=0]\!] & [\![y_{51}=0]\!] & [\![y_{31}=0]\!] & [\![y_{81}=0]\!] \\ [\![y_{51}=0]\!] & [\![y_{61}=0]\!] & [\![y_{91}=0]\!] & [\![y_{41}=0]\!] & [\![y_{81}=0]\!] \\ [\![y_{41}=0]\!] & [\![y_{71}=0]\!] & [\![y_{61}=0]\!] & [\![y_{51}=0]\!] & [\![y_{91}=0]\!] \\ [\![y_{61}=0]\!] & [\![y_{31}=0]\!] & [\![y_{71}=0]\!] & [\![y_{51}=0]\!] & [\![y_{21}=0]\!] \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 \end{bmatrix}$$

$$\begin{bmatrix} \mathcal{N}_2^{\mathbf{x}_1} \\ \mathcal{N}_2^{\mathbf{x}_2} \\ \mathcal{N}_2^{\mathbf{x}_3} \\ \mathcal{N}_2^{\mathbf{x}_4} \\ \mathcal{N}_2^{\mathbf{x}_5} \\ \mathcal{N}_2^{\mathbf{x}_6} \\ \mathcal{N}_2^{\mathbf{x}_7} \\ \mathcal{N}_2^{\mathbf{x}_8} \\ \mathcal{N}_2^{\mathbf{x}_9} \end{bmatrix} = \begin{bmatrix} [\![y_{42}=0]\!] & [\![y_{22}=0]\!] & [\![y_{92}=0]\!] & [\![y_{32}=0]\!] & [\![y_{52}=0]\!] \\ [\![y_{32}=0]\!] & [\![y_{12}=0]\!] & [\![y_{92}=0]\!] & [\![y_{52}=0]\!] & [\![y_{62}=0]\!] \\ [\![y_{92}=0]\!] & [\![y_{22}=0]\!] & [\![y_{62}=0]\!] & [\![y_{52}=0]\!] & [\![y_{72}=0]\!] \\ [\![y_{12}=0]\!] & [\![y_{52}=0]\!] & [\![y_{72}=0]\!] & [\![y_{82}=0]\!] & [\![y_{62}=0]\!] \\ [\![y_{72}=0]\!] & [\![y_{62}=0]\!] & [\![y_{92}=0]\!] & [\![y_{42}=0]\!] & [\![y_{82}=0]\!] \\ [\![y_{72}=0]\!] & [\![y_{92}=0]\!] & [\![y_{52}=0]\!] & [\![y_{32}=0]\!] & [\![y_{82}=0]\!] \\ [\![y_{52}=0]\!] & [\![y_{62}=0]\!] & [\![y_{92}=0]\!] & [\![y_{42}=0]\!] & [\![y_{82}=0]\!] \\ [\![y_{42}=0]\!] & [\![y_{72}=0]\!] & [\![y_{62}=0]\!] & [\![y_{52}=0]\!] & [\![y_{92}=0]\!] \\ [\![y_{62}=0]\!] & [\![y_{32}=0]\!] & [\![y_{72}=0]\!] & [\![y_{52}=0]\!] & [\![y_{22}=0]\!] \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 & 1 \end{bmatrix}$$

$$\begin{bmatrix} \mathcal{N}_3^{\mathbf{x}_1} \\ \mathcal{N}_3^{\mathbf{x}_2} \\ \mathcal{N}_3^{\mathbf{x}_3} \\ \mathcal{N}_3^{\mathbf{x}_4} \\ \mathcal{N}_3^{\mathbf{x}_5} \\ \mathcal{N}_3^{\mathbf{x}_6} \\ \mathcal{N}_3^{\mathbf{x}_7} \\ \mathcal{N}_3^{\mathbf{x}_8} \\ \mathcal{N}_3^{\mathbf{x}_9} \end{bmatrix} = \begin{bmatrix} [\![y_{43}=0]\!] & [\![y_{23}=0]\!] & [\![y_{93}=0]\!] & [\![y_{33}=0]\!] & [\![y_{53}=0]\!] \\ [\![y_{33}=0]\!] & [\![y_{13}=0]\!] & [\![y_{93}=0]\!] & [\![y_{53}=0]\!] & [\![y_{63}=0]\!] \\ [\![y_{93}=0]\!] & [\![y_{23}=0]\!] & [\![y_{63}=0]\!] & [\![y_{53}=0]\!] & [\![y_{73}=0]\!] \\ [\![y_{13}=0]\!] & [\![y_{53}=0]\!] & [\![y_{73}=0]\!] & [\![y_{83}=0]\!] & [\![y_{63}=0]\!] \\ [\![y_{73}=0]\!] & [\![y_{63}=0]\!] & [\![y_{93}=0]\!] & [\![y_{43}=0]\!] & [\![y_{83}=0]\!] \\ [\![y_{73}=0]\!] & [\![y_{93}=0]\!] & [\![y_{53}=0]\!] & [\![y_{33}=0]\!] & [\![y_{81}=0]\!] \\ [\![y_{53}=0]\!] & [\![y_{61}=0]\!] & [\![y_{93}=0]\!] & [\![y_{43}=0]\!] & [\![y_{83}=0]\!] \\ [\![y_{43}=0]\!] & [\![y_{73}=0]\!] & [\![y_{63}=0]\!] & [\![y_{53}=0]\!] & [\![y_{93}=0]\!] \\ [\![y_{63}=0]\!] & [\![y_{33}=0]\!] & [\![y_{73}=0]\!] & [\![y_{53}=0]\!] & [\![y_{23}=0]\!] \end{bmatrix} = \begin{bmatrix} 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 & 1 \end{bmatrix} \quad \square$$

In contrast, the relative influential ratio measures the degree of proximity of the sets composed of positive and negative class instances within the neighborhood to their respective optimal distributions. By referring to optimal class distribution, we mean the positive and negative classes distributed in the nearest neighbors have the same instance count. Formally, we define the relative influential ratio for a positive class within $\mathbf{x}_b$ (i.e., $rip_j^{\mathbf{x}_b}$) as:

$$rip_j^{\mathbf{x}_b} = \begin{cases} \frac{\mathbf{w}^{\mathbf{x}_b\top}\mathcal{P}_j^{\mathbf{x}_b}}{\sum_{r=1}^{P} w_r^{\mathbf{x}_b}} & \exists y_{rj}=1 \\ 0 & otherwise \end{cases} \tag{5}$$

where $P=\sum sgn(\mathcal{P}_j^{\mathbf{x}_b})$, denotes the count of instances with label l_j in the neighborhood of $\mathbf{x}_b$. The semantics of relativeness in Equ (5) refers to the proportion of instances with positive class of l_j in terms first P weights if there are $P\,(P>0)$ instances with positive class of l_j in the neighborhood of $\mathbf{x}_b$, and 0 if $P=0$. Consequently, the relative influential ratio for the positive class is defined as:

$$\mathbf{rip}_j = \left(rip_j^{\mathbf{x}_b}\right)_{n\times 1} \tag{6}$$

Analogously, we define the relative influential ratio for the negative class within $\mathbf{x}_b$ (i.e., $rin_j^{\mathbf{x}_b}$) as:

$$rin_j^{\mathbf{x}_b} = \begin{cases} \frac{\mathbf{w}^{\mathbf{x}_b\top}\mathcal{N}_j^{\mathbf{x}_b}}{\sum_{r=1}^{N} w_r^{\mathbf{x}_b}} & \exists y_{rj}=0 \\ 0 & otherwise \end{cases} \tag{7}$$

where $N=\sum sgn(\mathcal{N}_j^{\mathbf{x}_b})$, denotes the count of instances without label l_j in the neighborhood of $\mathbf{x}_b$. The semantics of relativeness in Equ (7) refers to the proportion of instances with positive class of l_j in terms first N weights if there are $N\,(N>0)$ instances with positive class of l_j in the neighborhood of $\mathbf{x}_b$, and 0 if $N=0$. Consequently, the relative influential ratio for the negative class is defined as:

$$\mathbf{rin}_j = \left(rin_j^{\mathbf{x}_b}\right)_{n\times 1} \tag{8}$$

It is straightforward to deduce that both $rip_j^{\mathbf{x}_b}\in[0,1]$ and $rin_j^{\mathbf{x}_b}\in[0,1]$ hold. $rip_j^{\mathbf{x}_b}\in[0,1]$ reaches the maximal value of 1 if the instance with label l_j are the nearest neighbors of $\mathbf{x}_b$, and reaches the minimum value of 0 if no positive association on l_j occurs. Similar value variations also apply to the $rin_j^{\mathbf{x}_b}\in[0,1]$.

Example *(Continued)*. For the survey dataset mentioned in Table 2, we have the relative influential ratio for the positive class by Equ (5) as:

$$\mathbf{rip} = \begin{bmatrix} \mathbf{rip}_1 & \mathbf{rip}_2 & \mathbf{rip}_3 \end{bmatrix} = \begin{bmatrix} rip_1^{\mathbf{x}_1} & rip_2^{\mathbf{x}_1} & rip_3^{\mathbf{x}_1} \\ rip_1^{\mathbf{x}_2} & rip_2^{\mathbf{x}_2} & rip_3^{\mathbf{x}_2} \\ rip_1^{\mathbf{x}_3} & rip_2^{\mathbf{x}_3} & rip_3^{\mathbf{x}_3} \\ rip_1^{\mathbf{x}_4} & rip_2^{\mathbf{x}_4} & rip_3^{\mathbf{x}_4} \\ rip_1^{\mathbf{x}_5} & rip_2^{\mathbf{x}_5} & rip_3^{\mathbf{x}_5} \\ rip_1^{\mathbf{x}_6} & rip_2^{\mathbf{x}_6} & rip_3^{\mathbf{x}_6} \\ rip_1^{\mathbf{x}_7} & rip_2^{\mathbf{x}_7} & rip_3^{\mathbf{x}_7} \\ rip_1^{\mathbf{x}_8} & rip_2^{\mathbf{x}_8} & rip_3^{\mathbf{x}_8} \\ rip_1^{\mathbf{x}_9} & rip_2^{\mathbf{x}_9} & rip_3^{\mathbf{x}_9} \end{bmatrix}$$

$$= \begin{bmatrix} \frac{0.9260+0.8632+0.8124}{0.9260+0.8632+0.8124} & \frac{0.7770}{0.9260} & \frac{0.9260+0.8063}{0.9260+0.8632} \\ \frac{0.8632+0.8342+0.7822}{0.9161+0.8632+0.8342} & \frac{0.8206}{0.9161} & \frac{0.9161+0.8632}{0.9161+0.8632} \\ \frac{0.9658+0.9161+0.9117}{0.9658+0.9161+0.9117} & \frac{0.8329+0.8236}{0.9658+0.9161} & \frac{0.8236}{0.9658} \\ \frac{0.9260+0.7906}{0.9260+0.8985} & \frac{0.8985+0.8925+0.8857}{0.9260+0.8985+0.8925} & \frac{0.9260+0.8925}{0.9260+0.8985} \\ \frac{0.9495+0.9010+0.8985}{0.9766+0.9495+0.9010} & \frac{0.9766+0.8347}{0.9766+0.9495} & \frac{0.9766+0.8985}{0.9766+0.9495} \\ \frac{0.9668}{0.9727} & \frac{0.9727+0.9495+0.8362}{0.9727+0.9668+0.9495} & \frac{0.9727+0.9117}{0.9727+0.9668} \\ \frac{0.9727+0.9233+0.8925}{0.9766+0.9727+0.9233} & \frac{0.9766+0.8836}{0.9766+0.9727} & \frac{0.8925}{0.9766} \\ \frac{0.8857+0.8362+0.7521}{0.8857+0.8836+0.8362} & \frac{0.8836+0.8347}{0.8857+0.8836} & \frac{0.8857+0.8836}{0.8857+0.8836} \\ \frac{0.9668+0.8342}{0.9668+0.9658} & \frac{0.9233+0.9010}{0.9668+0.9658} & \frac{0.9658+0.9233}{0.9668+0.9658} \end{bmatrix} = \begin{bmatrix} 1.0000 & 0.8391 & 0.9682 \\ 0.9488 & 0.8958 & 1.0000 \\ 1.0000 & 0.8802 & 0.8528 \\ 0.9409 & 0.9852 & 0.9967 \\ 0.9724 & 0.9404 & 0.9735 \\ 0.9939 & 0.9548 & 0.9716 \\ 0.9707 & 0.9543 & 0.9139 \\ 0.9495 & 0.9712 & 1.0000 \\ 0.9319 & 0.9440 & 0.9775 \end{bmatrix}$$

Analogously, the relative influential ratio for the negative class by Equ (7) as:

$$\mathbf{rin} = \begin{bmatrix} \mathbf{rin}_1 & \mathbf{rin}_2 & \mathbf{rin}_3 \end{bmatrix} = \begin{bmatrix} rin_1^{\mathbf{x}_1} & rin_2^{\mathbf{x}_1} & rin_3^{\mathbf{x}_1} \\ rin_1^{\mathbf{x}_2} & rin_2^{\mathbf{x}_2} & rin_3^{\mathbf{x}_2} \\ rin_1^{\mathbf{x}_3} & rin_2^{\mathbf{x}_3} & rin_3^{\mathbf{x}_3} \\ rin_1^{\mathbf{x}_4} & rin_2^{\mathbf{x}_4} & rin_3^{\mathbf{x}_4} \\ rin_1^{\mathbf{x}_5} & rin_2^{\mathbf{x}_5} & rin_3^{\mathbf{x}_5} \\ rin_1^{\mathbf{x}_6} & rin_2^{\mathbf{x}_6} & rin_3^{\mathbf{x}_6} \\ rin_1^{\mathbf{x}_7} & rin_2^{\mathbf{x}_7} & rin_3^{\mathbf{x}_7} \\ rin_1^{\mathbf{x}_8} & rin_2^{\mathbf{x}_8} & rin_3^{\mathbf{x}_8} \\ rin_1^{\mathbf{x}_9} & rin_2^{\mathbf{x}_9} & rin_3^{\mathbf{x}_9} \end{bmatrix}$$

$$= \begin{bmatrix} \frac{0.8063+0.7770}{0.9260+0.8632} & \frac{0.9260+0.8632+0.8124+0.8063}{0.9260+0.8632+0.8124+0.8063} & \frac{0.8632+0.8124+0.7770}{0.9260+0.8632+0.8124} \\ \frac{0.9161+0.8206}{0.9161+0.8632} & \frac{0.9161+0.8632+0.8342+0.7822}{0.9161+0.8632+0.8342+0.8206} & \frac{0.8342+0.8206+0.7822}{0.9161+0.8632+0.8342} \\ \frac{0.8329+0.8236}{0.9658+0.9161} & \frac{0.9658+0.9161+0.9117}{0.9658+0.9161+0.9117} & \frac{0.9658+0.9161+0.9117+0.8329}{0.9658+0.9161+0.9117+0.8329} \\ \frac{0.8985+0.8925+0.8857}{0.9260+0.8985+0.8925} & \frac{0.9260+0.7906}{0.9260+0.8985} & \frac{0.8985+0.8857+0.7906}{0.9260+0.8985+0.8925} \\ \frac{0.9766+0.8347}{0.9766+0.9495} & \frac{0.9495+0.9010+0.8985}{0.9766+0.9495+0.9010} & \frac{0.9495+0.9010+0.8347}{0.9766+0.9495+0.9010} \\ \frac{0.9727+0.9495+0.9117+0.8362}{0.9727+0.9668+0.9495+0.9117} & \frac{0.9668+0.9117}{0.9727+0.9668} & \frac{0.9668+0.9495+0.8362}{0.9727+0.9668+0.9495} \\ \frac{0.9766+0.8836}{0.9766+0.9727} & \frac{0.9727+0.9233+0.8925}{0.9766+0.9727+0.9233} & \frac{0.9766+0.9727+0.9233+0.8836}{0.9766+0.9727+0.9233+0.8925} \\ \frac{0.8836+0.8347}{0.8857+0.8836} & \frac{0.8857+0.8362+0.7521}{0.8857+0.8836+0.8362} & \frac{0.8836+0.8347+0.7521}{0.8857+0.8836+0.8362} \\ \frac{0.9658+0.9233+0.9010}{0.9668+0.9658+0.9233} & \frac{0.9668+0.9658+0.8342}{0.9668+0.9658+0.9233} & \frac{0.9668+0.9010+0.8342}{0.9668+0.9658+0.9233} \end{bmatrix} = \begin{bmatrix} 0.8849 & 1.0000 & 0.9427 \\ 0.9761 & 0.9888 & 0.9325 \\ 0.8802 & 1.0000 & 1.0000 \\ 0.9852 & 0.9409 & 0.9477 \\ 0.9404 & 0.9724 & 0.9498 \\ 0.9656 & 0.9685 & 0.9528 \\ 0.9543 & 0.9707 & 0.9976 \\ 0.9712 & 0.9495 & 0.9481 \\ 0.9770 & 0.9688 & 0.9461 \end{bmatrix} \square$$

The overall positive tendency and negative tendency can be estimated from the triple tuple $(\mathcal{P}_j^{\mathbf{x}_b}, \mathbf{W}^{\mathbf{x}_b}, rip_j^{\mathbf{x}_b})$ and $(\mathcal{N}_j^{\mathbf{x}_b}, \mathbf{W}^{\mathbf{x}_b}, rin_j^{\mathbf{x}_b})$, respectively. For $\mathbf{x}_b$ on l_j, the weighted score of positive class is:

$$WSP_j^{\mathbf{x}_b} = {rip_j^{\mathbf{x}_b}}^{\top} \mathcal{P}_j^{\mathbf{x}_b} \tag{9}$$

Consequently, the collection of the weighted score of positive class on l_j is defined as:

$$\mathbf{WSP}_j = \left(WSP_j^{\mathbf{x}_b} \right)_{n \times 1} \tag{10}$$

Analogously, the weighted score of negative class for $\mathbf{x}_b$ on l_j is:

$$WSN_j^{\mathbf{x}_b} = {rin_j^{\mathbf{x}_b}}^{\top} \mathcal{N}_j^{\mathbf{x}_b} \tag{11}$$

Consequently, the collection of negative score of positive class on l_j is defined as:

$$\mathbf{WSN}_j = \left(WSN_j^{\mathbf{x}_b} \right)_{n \times 1} \tag{12}$$

Example *(Continued)*. For the survey dataset mentioned in Table 2, we have the weighted score of positive class by Equ (10) as:

$$\mathbf{WSP} = \begin{bmatrix} \mathbf{WSP}_1 & \mathbf{WSP}_2 & \mathbf{WSP}_3 \end{bmatrix}$$

$$= \begin{bmatrix} 1.0000\times 3 & 0.8391\times 1 & 0.9682\times 2 \\ 0.9488\times 3 & 0.8958\times 1 & 1.0000\times 2 \\ 1.0000\times 3 & 0.8802\times 2 & 0.8528\times 1 \\ 0.9409\times 2 & 0.9852\times 3 & 0.9967\times 2 \\ 0.9724\times 3 & 0.9404\times 2 & 0.9735\times 2 \\ 0.9939\times 1 & 0.9548\times 3 & 0.9716\times 2 \\ 0.9707\times 3 & 0.9543\times 2 & 0.9139\times 1 \\ 0.9495\times 3 & 0.9712\times 2 & 1.0000\times 2 \\ 0.9319\times 2 & 0.9440\times 2 & 0.9775\times 2 \end{bmatrix} = \begin{bmatrix} 3.0000 & 0.8391 & 1.9364 \\ 2.8464 & 0.8958 & 2.0000 \\ 3.0000 & 1.7604 & 0.8528 \\ 1.8818 & 2.9556 & 1.9934 \\ 2.9172 & 1.9086 & 0.9139 \\ 0.9939 & 2.8644 & 1.9432 \\ 2.9121 & 1.9086 & 0.9139 \\ 2.8485 & 1.9424 & 2.0000 \\ 1.8638 & 1.8880 & 1.9550 \end{bmatrix}$$

Analogously, we have the weighted score of negative class by Equ (12) as:

$$\mathbf{WSN} = \begin{bmatrix} \mathbf{WSN}_1 & \mathbf{WSN}_2 & \mathbf{WSN}_3 \end{bmatrix}$$

$$= \begin{bmatrix} 0.8849\times 2 & 1.0000\times 4 & 0.9427\times 3 \\ 0.9761\times 2 & 0.9888\times 4 & 0.9325\times 3 \\ 0.8802\times 2 & 1.0000\times 3 & 1.0000\times 4 \\ 0.9852\times 3 & 0.9409\times 2 & 0.9477\times 3 \\ 0.9404\times 2 & 0.9724\times 3 & 0.9498\times 3 \\ 0.9656\times 4 & 0.9685\times 2 & 0.9528\times 3 \\ 0.9543\times 2 & 0.9707\times 3 & 0.9976\times 4 \\ 0.9712\times 2 & 0.9495\times 3 & 0.9481\times 3 \\ 0.9770\times 3 & 0.9688\times 3 & 0.9461\times 3 \end{bmatrix} = \begin{bmatrix} 1.7698 & 4.0000 & 2.8281 \\ 1.9522 & 3.9552 & 2.7975 \\ 1.7604 & 3.0000 & 4.0000 \\ 2.9556 & 1.8818 & 2.8431 \\ 1.8808 & 2.9172 & 2.8494 \\ 3.8624 & 1.9370 & 2.8584 \\ 1.9086 & 2.9121 & 3.9904 \\ 1.9424 & 2.8485 & 2.8443 \\ 2.9310 & 2.9064 & 2.8383 \end{bmatrix} \quad \square$$

3.3.3. Label-specific augmented features

Having completed the positive/negative tendency measure (i.e., $\mathbf{WSP}_j$ and $\mathbf{WSN}_j$), we can estimate the local relative discrimination with available features. As evaluation metrics highlight the accuracy of positive classes, we devise a measure to encourage a larger value for the instances with salient positive tendency and compactness of the neighborhood and constrain a smaller value otherwise. For this sake, we define the following augmented feature for l_j:

$$\mathbf{AF}_j = \frac{\mathbf{WSP}_j - \mathbf{WSN}_j}{Me(\mathbf{W}^{\mathbf{x}_b})} \tag{13}$$

where $Me(\cdot)$ denotes the median distance of the instances within the neighborhood. The term $\mathbf{WSP}_j - \mathbf{WSN}_j$ describes the saliency of the positive tendency to the negative tendency on label l_j. The larger the difference is, the stronger the positive tendency becomes. In contrast, the term $Me(\mathbf{W}^{\mathbf{x}_b})$[4]describes the compactness of k nearest neighborhood. The smaller the median value is, the more compact a k nearest neighborhood becomes. In this way, the larger the component in $\mathbf{AF}_j$ is, the more reliable positive tendency of a k nearest neighborhood becomes, and vice versa.

Example *(Continued)*. For the survey dataset mentioned in Table 2, we have the augmented features by following Equ (13) as:

$$\mathbf{AF} = \begin{bmatrix} \mathbf{AF}_1 & \mathbf{AF}_2 & \mathbf{AF}_3 \end{bmatrix}$$

$$= \begin{bmatrix} \frac{3.0000-1.7698}{0.8124} & \frac{0.8391-4.0000}{0.8124} & \frac{1.9364-2.8281}{0.8124} \\ \frac{2.8464-1.9522}{0.8342} & \frac{0.8958-3.9552}{0.8342} & \frac{2.0000-2.7975}{0.8342} \\ \frac{3.0000-1.7604}{0.9117} & \frac{1.7604-3.0000}{0.9117} & \frac{0.8528-4.0000}{0.9117} \\ \frac{1.8818-2.9556}{0.8925} & \frac{2.9556-1.8818}{0.8925} & \frac{1.9934-2.8431}{0.8925} \\ \frac{2.9172-1.8808}{0.9010} & \frac{1.9086-2.9172}{0.9010} & \frac{0.9139-2.8494}{0.9010} \\ \frac{0.9939-3.8694}{0.9495} & \frac{2.8644-1.9370}{0.9495} & \frac{1.9432-2.8584}{0.9495} \\ \frac{2.9121-1.9086}{0.9233} & \frac{1.9086-2.9121}{0.9233} & \frac{0.9139-3.9904}{0.9233} \\ \frac{2.8485-1.9424}{0.8362} & \frac{1.9424-2.8485}{0.8362} & \frac{2.0000-2.8443}{0.8362} \\ \frac{1.8638-2.9310}{0.9233} & \frac{1.8880-2.9064}{0.9233} & \frac{1.9550-2.8383}{0.9233} \end{bmatrix} = \begin{bmatrix} 1.5143 & -3.8908 & -1.0976 \\ 1.0719 & -3.6675 & -0.9560 \\ 1.3597 & -1.3597 & -3.4520 \\ -1.2031 & 1.2031 & -0.9520 \\ 1.1503 & -1.1194 & -2.1482 \\ -3.0284 & 0.9767 & -0.9639 \\ 1.0869 & -1.0869 & -3.3321 \\ 1.0836 & -1.0836 & -1.0097 \\ -1.1559 & -1.1030 & -0.9567 \end{bmatrix} \quad \square$$

After generating augmented features for all labels, we deduce the multi-granularity feature $\mathbf{X}_{aug}$ as:

$$\mathbf{X}_{aug} = \mathbf{X} \cup \left(z\left(\mathbf{AF}_j\right)\right)_{n\times q} \tag{14}$$

[4] We use the median instead of mean value of k nearest neighborhood to guarantee the stable estimation when the outlier is available.

where $j = 1, 2, \ldots, q$ and $z(\cdot)$ represents the z-score function. By z-score scaling, the bias incurred by absolute values of augmented features are reduced.

Example *(Continued)*. For the user survey mentioned in Table 2, the representations after performing feature augmentation are shown in Table 3.

Table 3
A multi-label user survey for product recommendation.

$\mathbf{X}$	c_1	c_2	c_3	c_4	c_5	c_6	$z(\mathbf{AF}_1)$	$z(\mathbf{AF}_2)$	$z(\mathbf{AF}_3)$	l_1	l_2	l_3
x_1	0.8	0.1	0.1	0.5	0.2	0.3	0.8149	-1.5356	0.5239	1	0	1
x_2	0.3	0.5	0.2	0.8	0.1	0.1	0.5387	-1.4064	0.6577	1	0	0
x_3	0.2	0.2	0.6	0.7	0.3	0.2	0.7184	-0.0711	-1.7008	0	0	1
x_4	0.6	0.3	0.1	0.2	0.5	0.3	-0.8813	1.4117	0.6615	1	0	1
x_5	0.3	0.4	0.3	0.3	0.6	0.1	0.5877	0.0679	-0.4688	0	1	0
x_6	0.2	0.3	0.5	0.3	0.5	0.2	-2.0207	1.2807	0.6502	1	0	0
x_7	0.3	0.3	0.4	0.2	0.6	0.2	0.5481	0.0867	-1.5875	0	1	1
x_8	0.3	0.4	0.3	0.1	0.4	0.5	0.5461	0.0886	0.6069	0	1	0
x_9	0.3	0.2	0.5	0.4	0.4	0.2	-0.8519	0.0774	0.6570	1	0	0

□

3.4. Corresponding algorithm

Algorithm 1 summarizes the main procedures of the proposed GOFA model. Step 2 corresponds to instance-based neighborhood construction, which takes the complexity of $O(n^2m)$. Step 4–Step 9 corresponds to the neighborhood-aware label-dependent feature augmentation procedure, which takes the complexity of $O(nkq)$. Step 11 is the concatenation of augmented features, which takes the complexity of $O(q)$. Since $m \gg q$ and $m > n$ hold in most cases, the overall complexity of GOFA is $O(n^2m)$.

Algorithm 1: Granule cOrrelation-based Feature Augmentation (GOFA).

Input: Training set $\mathbf{X}_1$, ground-truth labels $\mathbf{Y}_1$, Testing set $\mathbf{X}_2$ Nearest neighborhood count k, balance factor μ
Output: Augmented features $\mathbf{X}_{aug}$
1 **for** $b = 1$ *to* n **do**
2 Find the k nearest neighborhood w.r.t. $\mathbf{x}_b$ from $\mathbf{X}_1$ based on Equ (1).
3 Compute the instance weight $\mathbf{W}^{\mathbf{x}_b}$ w.r.t. $\mathbf{x}_b$ based on Equ (4).
4 **for** $j = 1$ *to* q **do**
5 Find the positive and negative class (i.e., $\mathcal{P}_j^{\mathbf{x}_b}$, $\mathcal{N}_j^{\mathbf{x}_b}$) based on Equ (2) and Equ (3).
6 Compute the relative influential ratio of positive and negative class (i.e., $\mathbf{rip}_j$ and $\mathbf{rin}_j$) based on Equ (6) and Equ (8).
7 Compute weighted score of positive and negative class (i.e., $\mathbf{WSP}_j$ and $\mathbf{WSN}_j$) based on Equ (10) and Equ (12).
8 Generate augmented feature $\mathbf{AF}_j$ based on Equ (13).
9 **end**
10 **end**
11 Output augmented features $\mathbf{X}_{aug}$ based on Equ (14).

4. Experiments

4.1. Datasets

Table 4 enumerates characteristics of thirteen datasets, including the instance count ($|S|$), feature dimensionality ($dim(S)$), label count ($L(S)$), the cardinality of average associated labels per instance ($LCard(S)$), and domain (Domain). Among them, datasets *birds*, *cal500*, *emotions*, and *scene* come from Mulan,[5] datasets *art*, *business*, *computers*, *education*, *entertainment*, *health*, *recreation*, *science*, and *social* come from Lamda.[6]

4.2. Experimental settings

To evaluate the effectiveness of GOFA, we adopt five widely used metrics [25], including *Hamming Loss*, *One Error*, *Ranking Loss*, *Average Precision* and *Macro-averaging AUC*. Except for the last two metrics that achieve better performance if the values are larger, the first three metrics achieve better performance if the values are smaller.

[5] http://mulan.sourceforge.net/datasets-mlc.html.
[6] https://www.lamda.nju.edu.cn/code_MDDM.ashx.

Table 4
Dataset characteristics.

Data set	$\|S\|$	$dim(S)$	$L(S)$	$LCard(S)$	Domain
art	5000	462	26	1.64	text
birds	645	260	19	1.014	audio
business	5000	438	30	1.59	text
cal500	502	68	174	26.044	music
computers	5000	681	33	1.51	text
education	5000	550	33	1.46	text
emotions	593	72	6	1.869	music
entertainment	5000	640	21	1.42	text
health	5000	612	32	1.66	text
recreation	5000	606	22	1.42	text
scene	2407	294	6	1.074	images
science	5000	743	40	1.45	text
social	5000	1047	39	1.28	text

We compare GOFA against eight state-of-the-art multi-label algorithms for performance evaluations. The configurations for these algorithms take the recommended values via five-fold cross-validation.

- WRAP [36][7]: A label-specific multi-label classification which takes a wrapped approach w.r.t. each label[8] [parameter configurations: grid search for $\lambda_1, \lambda_2 \in \{0, 1, \ldots, 10\}$. $\lambda_3 \in \{0.1, 1\}$ and $\alpha = 0.9$].
- HOMI [27][9]: A high-order label correlation learning with self-representation and local geometric structure [parameter configurations: grid search for $\beta, \lambda, \gamma \in \{10^{-5}, 10^{-4}, \ldots, 1\}$ and $s \in \{5, 10\}$].
- SLOFS [26][10]: A shared latent sublabel structure for adaptive feature selection [parameter configurations: grid search for $\alpha, \lambda_1, \lambda_2, \beta, \delta \in \{0.01, 0.1, 0.3, 0.5, 0.7, 0.9, 1\}$] The selected features are examined by the WRAP classifier.
- MDFS [40][11]: An embedded feature selection with manifold regularization [parameter configurations: $\alpha = 1$, grid search for $\beta, \gamma \in \{10^{-3}, 10^{-2}, \ldots, 10^{3}\}$]. The selected features are examined by the WRAP classifier.
- TIFS [20][12]: A latent topic-based instance and feature selection with global and local label correlation [parameter configurations: $\lambda = 0.5, k = 10, s = 50$, grid search for $\tau, \delta \in \{10^{-4}, 10^{-3}, \ldots, 10^{-1}\}$].
- RLFSCL [18][13]: A low-rank feature and label representation learning [parameter configurations: $\rho = 1.1$, grid search for $\lambda_1, \lambda_2 \in \{10^{-3}, 10^{-2}, \ldots, 10^{3}\}$, $\mu \in \{1, 10^{1}, \ldots, 10^{6}\}$].
- GLFS [41][14]: A group-preserving label-specific feature selection learning [parameter configurations: grid search for $\alpha, \lambda \in \{0, 0.2, \ldots, 1\}$, $\beta, \gamma \in \{10^{-3}, 10^{-2}, \ldots, 10^{3}\}$, $K = 5, M = 16$]. The selected features are examined by the WRAP classifier.
- MC-GM [19][15]: A group-specific feature selection with label-specific group selection [parameter configurations: grid search for $\lambda, \beta \in \{10^{-4}, 10^{-2}, \ldots, 1\}$, $\delta, \alpha \in \{10^{-2}, 10^{-1}, \ldots, 10^{2}\}$, $s = 50$].

The configurations of GOFA are as follows: The balance factor μ is searched in $[0, 1]$ at a step of 0.1. We stipulate $k = 10$ for neighborhood construction if the instance count reaches 5000 and define $k = 5$ otherwise. We assess the functionality of augmented features by employing the concatenation with reduced features on the WRAP classifier. In this way, we simulate the cognitive processing of "data + relation" for multi-label classification. To be compatible with the feature dimensionality from WRAP, the embedding dimensionality d follows the setting of WRAP and is determined as $d = \lceil \alpha \min(m, q) \rceil$, where $\alpha = 0.9$. We take a grid search manner for trade-off $\lambda_1, \lambda_2 \in \{1, 2, \ldots, 10\}$. The trade-off λ_3 is fixed as $\lambda_3 = 1$.

For fair comparisons, all experiments are executed via five-fold cross-validation and are implemented on a desktop with Matlab R2017b.

4.3. Results

Tables 5–9 show the detailed classification performance ranking information over thirteen benchmarks and eight comparing methods. From the metric view, GOFA ranks first at 100% cases. From the dataset view, GOFA ranks first at 70.8% $\left(\frac{46}{65}\right)$, ranks second at 26.2% $\left(\frac{17}{65}\right)$.

[7] Code available at: https://palm.seu.edu.cn/zhangml/files/WRAP.rar.
[8] For fair comparisons, we use the linear version instead of kernel version.
[9] Code available at: https://github.com/Chongjie-Si/HOMI.
[10] Code available at: https://github.com/zhongjingyu1/SLOFS.
[11] Code available at: https://github.com/jiazhang-ml/MDFS.
[12] Code available at: https://github.com/JianghongMA/TIFS.
[13] Code available at: https://github.com/JingChuanTang/RLFSCL.
[14] Code available at: https://github.com/jiazhang-ml/GLFS.
[15] Code available at: https://github.com/JianghongMA/MC-GM.

Table 5
Comparisons (mean±std) on metric *Hamming Loss* ↓: The smaller the better.

Dataset	*Hamming Loss* ↓								
	GOFA	WRAP	HOMI	SLOFS	MDFS	TIFS	RLFSCL	GLFS	MC-GM
art	**0.053±0.002**	0.054±0.001	0.060±0.001	0.063±0.001	0.063±0.001	0.124±0.020	0.057±0.001	0.063±0.001	0.086±0.002
birds	0.049±0.008	0.050±0.005	0.056±0.008	0.053±0.004	0.054±0.004	0.057±0.004	**0.046±0.004**	0.053±0.006	0.075±0.004
business	**0.025±0.002**	**0.025±0.001**	0.026±0.001	0.029±0.001	0.028±0.001	0.044±0.006	0.026±0.000	0.028±0.002	0.032±0.002
cal500	**0.136±0.001**	0.137±0.003	0.137±0.002	0.137±0.003	0.137±0.001	0.211±0.003	0.137±0.003	0.137±0.005	0.138±0.001
computers	**0.033±0.001**	0.034±0.001	0.035±0.001	0.040±0.001	0.038±0.001	0.074±0.008	0.035±0.001	0.039±0.001	0.063±0.005
education	**0.037±0.001**	0.038±0.001	0.043±0.000	0.044±0.001	0.044±0.001	0.072±0.007	0.039±0.001	0.044±0.000	0.060±0.001
emotions	**0.206±0.023**	0.208±0.012	0.208±0.011	0.238±0.018	0.224±0.009	0.269±0.021	0.223±0.009	0.272±0.014	0.228±0.012
entertainment	**0.051±0.001**	0.052±0.001	0.063±0.002	0.068±0.001	0.067±0.001	0.112±0.008	0.055±0.001	0.067±0.001	0.087±0.004
health	0.035±0.001	0.035±0.001	0.036±0.001	0.050±0.001	0.036±0.001	0.044±0.001	**0.032±0.001**	0.047±0.001	0.048±0.002
recreation	**0.054±0.001**	**0.054±0.001**	0.062±0.001	0.065±0.001	0.065±0.001	0.124±0.014	0.056±0.001	0.065±0.001	0.098±0.007
scene	0.111±0.002	0.117±0.005	0.115±0.007	0.167±0.006	0.128±0.011	0.133±0.007	**0.091±0.002**	0.139±0.004	0.166±0.008
science	**0.031±0.006**	0.032±0.001	0.035±0.000	0.036±0.001	0.036±0.000	0.061±0.004	0.034±0.001	0.036±0.000	0.057±0.001
social	**0.020±0.001**	0.021±0.001	0.025±0.001	0.028±0.001	0.024±0.001	0.042±0.004	0.022±0.001	0.026±0.001	0.033±0.003
Avg rank	1.3462(1)	2.4231(2)	4.1154(4)	6.6154(7)	5.3077(5)	8.3846(9)	2.8077(3)	6.0769(6)	7.9231(8)

Table 6
Comparisons (mean±std) on metric *Ranking Loss* ↓: The smaller the better.

Dataset	*Ranking Loss* ↓								
	GOFA	WRAP	HOMI	SLOFS	MDFS	TIFS	RLFSCL	GLFS	MC-GM
art	0.132±0.005	0.133±0.006	**0.124±0.006**	0.176±0.004	0.172±0.002	0.189±0.018	0.135±0.006	0.174±0.003	0.156±0.006
birds	**0.089±0.023**	0.098±0.017	0.118±0.011	0.146±0.022	0.123±0.017	0.161±0.024	0.114±0.015	0.158±0.011	0.115±0.015
business	**0.039±0.004**	0.040±0.002	0.040±0.003	0.048±0.004	0.048±0.002	0.050±0.009	0.061±0.006	0.047±0.004	0.054±0.003
cal500	**0.175±0.006**	0.177±0.005	0.181±0.004	0.178±0.007	0.178±0.004	0.190±0.006	0.186±0.002	0.181±0.003	0.179±0.004
computers	**0.090±0.007**	0.094±0.008	0.102±0.008	0.096±0.005	0.095±0.001	0.136±0.013	0.097±0.004	0.094±0.004	0.127±0.010
education	0.105±0.006	0.107±0.007	**0.093±0.003**	0.108±0.004	0.106±0.003	0.128±0.011	0.097±0.006	0.108±0.003	0.127±0.010
emotions	**0.154±0.026**	0.161±0.014	0.180±0.016	0.215±0.027	0.197±0.012	0.217±0.013	0.185±0.017	0.251±0.017	0.198±0.025
entertainment	0.111±0.003	0.114±0.007	**0.100±0.005**	0.155±0.004	0.137±0.003	0.152±0.005	0.115±0.002	0.149±0.002	0.128±0.050
health	0.063±0.004	0.063±0.005	**0.062±0.003**	0.082±0.003	0.070±0.006	0.056±0.003	0.075±0.006	0.073±0.006	0.079±0.006
recreation	0.145±0.004	0.146±0.007	**0.136±0.009**	0.213±0.005	0.197±0.003	0.212±0.013	0.149±0.006	0.205±0.005	0.162±0.008
scene	**0.081±0.008**	0.090±0.009	0.109±0.007	0.179±0.023	0.095±0.018	0.107±0.064	**0.081±0.005**	0.101±0.007	0.100±0.012
science	0.129±0.002	0.131±0.007	**0.119±0.006**	0.153±0.002	0.149±0.005	0.152±0.009	0.133±0.006	0.149±0.002	0.138±0.006
social	0.079±0.006	0.082±0.003	**0.065±0.003**	0.078±0.004	0.078±0.002	0.091±0.004	0.084±0.007	0.078±0.005	0.095±0.008
Avg rank	2.0000(1)	3.1154(3)	3.0769(2)	6.9615(8)	4.9615(5)	7.6923(9)	4.8846(4)	6.0769(6)	6.2308(7)

Table 7
Comparisons (mean±std) on metric *One Error* ↓: The smaller the better.

Dataset	*One Error* ↓								
	GOFA	WRAP	HOMI	SLOFS	MDFS	TIFS	RLFSCL	GLFS	MC-GM
art	**0.460±0.019**	0.462±0.018	0.492±0.006	0.749±0.009	0.730±0.012	0.689±0.066	0.524±0.020	0.738±0.010	0.471±0.014
birds	0.665±0.009	0.673±0.032	0.732±0.032	0.816±0.049	0.771±0.024	**0.389±0.054**	0.671±0.020	0.842±0.036	0.721±0.045
business	**0.111±0.008**	0.113±0.008	0.114±0.016	0.135±0.011	0.135±0.004	0.293±0.054	0.115±0.007	0.135±0.008	0.127±0.018
cal500	0.116±0.019	0.118±0.020	0.116±0.019	0.116±0.021	**0.115±0.035**	0.386±0.043	0.122±0.033	**0.115±0.023**	0.123±0.027
computers	**0.348±0.013**	0.350±0.009	0.371±0.014	0.476±0.020	0.476±0.012	0.616±0.041	0.383±0.010	0.476±0.011	0.368±0.014
education	**0.469±0.008**	0.471±0.006	0.507±0.015	0.685±0.007	0.685±0.008	0.689±0.049	0.519±0.015	0.685±0.017	0.477±0.009
emotions	**0.276±0.005**	0.278±0.009	0.287±0.033	0.374±0.020	0.364±0.011	0.388±0.034	0.320±0.014	0.412±0.045	0.341±0.047
entertainment	**0.398±0.012**	0.403±0.015	0.436±0.022	0.715±0.009	0.650±0.009	0.608±0.031	0.459±0.010	0.700±0.025	0.404±0.007
health	0.263±0.017	0.270±0.013	0.283±0.011	0.493±0.016	0.289±0.013	0.364±0.020	**0.254±0.008**	0.448±0.021	0.324±0.024
recreation	**0.449±0.008**	0.459±0.021	0.488±0.022	0.805±0.012	0.750±0.024	0.691±0.027	0.521±0.017	0.799±0.011	0.469±0.018
scene	0.249±0.024	0.261±0.021	0.282±0.020	0.465±0.042	0.281±0.042	0.289±0.020	**0.246±0.007**	0.303±0.016	0.288±0.036
science	**0.496±0.009**	0.501±0.029	0.524±0.016	0.728±0.015	0.682±0.027	0.706±0.037	0.565±0.013	0.699±0.011	0.502±0.015
social	**0.287±0.006**	0.289±0.016	0.341±0.007	0.446±0.014	0.404±0.004	0.554±0.047	0.325±0.015	0.431±0.017	0.294±0.016
Avg rank	1.4615(1)	2.6154(2)	4.1538(3)	7.8462(9)	5.9615(6)	7.2308(7)	4.1538(4)	7.2692(8)	4.3077(5)

We employ Friedman test [5] to examine whether the statistical difference on relative performance hold for all metrics. Let N, T and r_i^j denote the comparing approaches count, the dataset count and the rank of the j-th algorithm on the i-th dataset, respectively. Given the average rank (i.e., $R_i^j = \frac{1}{T}\sum_{i=1}^{T} r_i^j$) information induced from Table 5 to Table 9, Friedman statistic F_F follows the F-distribution under the null hypothesis that all algorithms are statistically indistinguishable on classification performance, with $N-1$ numerator degrees of freedom and $(N-1)(T-1)$ denominator degrees of freedom (see (15)).

Table 8
Comparisons (mean±std) on metric *Average Precision* ↑: The bigger the better.

Dataset	*Average Precision* ↑								
	GOFA	WRAP	HOMI	SLOFS	MDFS	TIFS	RLFSCL	GLFS	MC-GM
art	**0.585±0.008**	0.583±0.009	0.573±0.013	0.419±0.005	0.429±0.006	0.441±0.047	0.555±0.013	0.424±0.004	0.572±0.011
birds	**0.664±0.048**	0.635±0.039	0.567±0.028	0.473±0.061	0.548±0.042	0.638±0.040	0.660±0.030	0.438±0.024	0.585±0.029
business	**0.855±0.009**	**0.855±0.004**	0.854±0.008	0.826±0.010	0.827±0.004	0.777±0.033	**0.855±0.007**	0.829±0.008	0.835±0.010
cal500	0.521±0.011	0.521±0.009	0.505±0.009	0.515±0.010	0.516±0.008	0.463±0.012	0.516±0.006	0.505±0.005	**0.525±0.027**
computers	**0.695±0.014**	0.690±0.008	0.683±0.011	0.587±0.017	0.590±0.007	0.510±0.033	0.671±0.006	0.588±0.011	0.660±0.010
education	**0.623±0.007**	**0.623±0.006**	0.602±0.009	0.479±0.004	0.485±0.002	0.457±0.038	0.607±0.009	0.481±0.012	0.612±0.011
emotions	**0.829±0.003**	0.824±0.007	0.813±0.012	0.771±0.011	0.785±0.008	0.735±0.016	0.802±0.144	0.735±0.019	0.789±0.019
entertainment	**0.680±0.015**	0.676±0.012	0.659±0.015	0.489±0.005	0.535±0.007	0.538±0.020	0.646±0.008	0.503±0.014	0.667±0.009
health	0.753±0.012	0.750±0.010	0.746±0.008	0.610±0.007	0.737±0.014	0.716±0.013	**0.772±0.009**	0.644±0.011	0.698±0.020
recreation	**0.606±0.007**	0.602±0.016	0.589±0.016	0.383±0.007	0.422±0.017	0.450±0.022	0.563±0.010	0.394±0.007	0.591±0.012
scene	0.855±0.011	0.845±0.012	0.826±0.011	0.717±0.026	0.836±0.026	0.823±0.010	**0.858±0.006**	0.824±0.009	0.830±0.019
science	**0.566±0.006**	0.560±0.018	0.547±0.009	0.388±0.007	0.425±0.013	0.422±0.028	0.528±0.012	0.406±0.003	0.553±0.011
social	**0.719±0.011**	0.717±0.009	0.702±0.009	0.616±0.014	0.640±0.005	0.585±0.028	0.709±0.008	0.625±0.014	0.697±0.015
Avg rank	1.3846(1)	2.3077(2)	4.4231(5)	8.2308(9)	6.0385(6)	7.3462(7)	3.5000(3)	7.6154(8)	4.1538(4)

Table 9
Comparisons (mean±std) on metric *Macro-averaging AUC* ↑: The bigger the better.

Dataset	*Macro-averaging AUC* ↑								
	GOFA	WRAP	HOMI	SLOFS	MDFS	TIFS	RLFSCL	GLFS	MC-GM
art	**0.959±0.001**	0.958±0.001	0.957±0.001	0.933±0.002	0.936±0.001	0.604±0.022	0.956±0.001	0.932±0.002	0.957±0.002
birds	**0.959±0.007**	0.953±0.011	0.949±0.008	0.948±0.014	0.957±0.003	0.697±0.030	**0.959±0.004**	0.957±0.003	0.948±0.011
business	**0.960±0.001**	0.959±0.001	0.959±0.001	0.933±0.007	0.925±0.002	0.562±0.013	0.957±0.001	0.917±0.008	0.956±0.001
cal500	**0.852±0.005**	0.848±0.006	0.831±0.001	0.847±0.004	0.845±0.005	0.521±0.009	0.846±0.004	0.826±0.008	0.850±0.005
computers	**0.970±0.001**	0.969±0.001	0.968±0.001	0.945±0.002	0.949±0.002	0.594±0.013	0.968±0.000	0.940±0.005	0.968±0.002
education	**0.971±0.001**	0.970±0.001	0.967±0.003	0.950±0.004	0.952±0.002	0.594±0.025	0.968±0.003	0.946±0.006	0.967±0.002
emotions	0.796±0.012	**0.797±0.007**	0.794±0.007	0.787±0.013	0.791±0.004	0.732±0.016	0.795±0.005	0.762±0.009	0.792±0.011
entertainment	**0.957±0.002**	**0.957±0.001**	0.956±0.001	0.926±0.002	0.942±0.001	0.637±0.014	0.956±0.001	0.930±0.010	0.956±0.002
health	**0.967±0.001**	0.965±0.002	0.965±0.001	0.954±0.007	0.965±0.002	0.591±0.006	0.966±0.001	0.961±0.000	0.965±0.001
recreation	**0.959±0.007**	0.958±0.002	0.958±0.002	0.931±0.002	0.940±0.002	0.623±0.013	0.956±0.001	0.933±0.001	0.957±0.001
scene	0.899±0.001	0.898±0.005	0.898±0.001	0.892±0.004	0.899±0.002	0.821±0.011	**0.900±0.001**	0.898±0.002	0.899±0.001
science	**0.976±0.000**	**0.976±0.001**	0.975±0.001	0.959±0.002	0.966±0.001	0.615±0.016	**0.976±0.001**	0.958±0.002	0.974±0.002
social	**0.979±0.001**	0.977±0.003	0.977±0.001	0.959±0.007	0.968±0.002	0.627±0.011	0.976±0.003	0.956±0.003	0.978±0.000
Avg rank	1.3846(1)	2.8846(2)	4.3077(5)	7.0385(7)	5.5385(6)	9.0000(9)	3.4231(3)	7.2692(8)	4.1538(4)

Table 10
Summary of the Friedman Statistics F_F ($N = 9$, $T = 13$) and the critical value at significance level $\alpha = 0.05$ in terms of each evaluation metric.

Evaluation metric	F_F	Critical value
Hamming Loss	56.2063	
Ranking Loss	12.0284	
One Error	22.7310	2.0363
Average Precision	45.5047	
Macro − averaging AUC	42.6385	

Table 10 summarizes the results for all considered five metrics, where $N = 9, T = 13$. Given critical value $\alpha = 0.05$, the null hypothesis of statistically indistinguishable performance among all considered algorithms is clearly rejected for all evaluations.

$$F_F = \frac{(T-1)\chi_F^2}{T(N-1)-\chi_F^2} \tag{15}$$

Where $\chi_F^2 = \frac{12T}{N(N+1)}\left[\sum_{j=1}^{N} R_j^2 - \frac{N(N+1)^2}{4}\right]$.

To further analysis whether GOFA is significantly superior over the remaining algorithms on different metrics, we employ Holm's procedure [5] as the post-hoc test by regarding GOFA as the control algorithm (denoted as $\mathcal{A}_1$). For the remaining $N-1$ comparing algorithms (denoted as $\mathcal{A}_j$, where $2 \leqslant j \leqslant N$), the one obtaining the $j-1$-th largest average rank over all datasets is denoted as $\mathcal{A}_j$. Then we have the test statistic for comparing $\mathcal{A}_1$ (i.e., GOFA) and $\mathcal{A}_j$ in (16).

Table 11
Comparisons of GOFA (control algorithm) against the remaining algorithms on *Hamming Loss*.

j	Algorithm	z_j	p_j	$\alpha/(k-j+1)$
2	**TIFS**	-6.552451	5.6600e-11	0.00625
3	**MC-GM**	-6.122782	9.1955e-10	0.00714
4	**SLOFS**	-4.905387	9.3243e-7	0.00833
5	**GLFS**	-4.404106	1.0622e-5	0.01000
6	**MDFS**	-3.687992	2.2603e-4	0.01250
7	**HOMI**	-2.578014	0.009937	0.01667
8	RLFSCL	-1.360618	0.173634	0.02500
9	WRAP	-1.002561	0.316073	0.05000

Table 12
Comparisons of GOFA (control algorithm) against the remaining algorithms on *Ranking Loss*.

j	Algorithm	z_j	p_j	$\alpha/(k-j+1)$
2	**TIFS**	-5.29925	1.1628e-7	0.00625
3	**SLOFS**	-4.618941	3.857e-6	0.00714
4	**MC-GM**	-3.938632	8.1948e-5	0.00833
5	**GLFS**	-3.795409	1.474e-4	0.01000
6	**MDFS**	-2.757042	0.005833	0.01250
7	**RLFSCL**	-2.685431	0.007244	0.01667
8	WRAP	-1.038367	0.299099	0.02500
9	HOMI	-1.002561	0.316073	0.05000

Table 13
Comparisons of GOFA (control algorithm) against the remaining algorithms on *One Error*.

j	Algorithm	z_j	p_j	$\alpha/(k-j+1)$
2	**SLOFS**	-5.943753	2.7857e-9	0.00625
3	**GLFS**	-5.406667	6.4208e-8	0.00714
4	**TIFS**	-5.370862	7.8361e-8	0.00833
5	**MDFS**	-4.189272	2.7985e-5	0.01000
6	**MC-GM**	-2.649625	0.008058	0.01250
7	**HOMI**	-2.506402	0.012197	0.01667
8	**RLFSCL**	-2.506402	0.012197	0.02500
9	WRAP	-1.074172	0.282745	0.05000

Table 14
Comparisons of GOFA (control algorithm) against the remaining algorithms on *Average Precision*.

j	Algorithm	z_j	p_j	$\alpha/(k-j+1)$
2	**SLOFS**	-6.373422	1.8486e-10	0.00625
3	**GLFS**	-5.80053	6.6106e-9	0.00714
4	**TIFS**	-5.54989	2.8585e-8	0.00833
5	**MDFS**	-4.332495	1.4743e-5	0.01000
6	**HOMI**	-2.828654	0.004674	0.01250
7	**MC-GM**	-2.578014	0.009937	0.01667
8	RLFSCL	-1.969316	0.048917	0.02500
9	WRAP	-0.859338	0.390154	0.05000

$$z_j = \left(R_1 - R_j\right) \Big/ \sqrt{\frac{N(N+1)}{6T}} \quad (2 \leqslant j \leqslant N) \tag{16}$$

We stipulate p_j as the p-value of z_j under the normal distribution. Given the significance level $\alpha = 0.05$, the Holm's procedure works by sequentially examining whether $p_j < \alpha/(N-j+1)$ holds in ascending order of j. Typically, the Holm's procedure continues until the first j (denoted as j^*) which suggests $p_j < \alpha/(N-j+1)$ does not hold.[16] Consequently, GOFA is deemed to be significantly superior over $\mathcal{A}_j$ with $j \in \{2, \ldots, j^* - 1\}$.

[16] j^* is set to be $N+1$ if $p_j < \alpha/(N-j+1)$ holds $\forall j$.

Table 15
Comparisons of GOFA (control algorithm) against the remaining algorithms on *Macro-averaging AUC*.

j	Algorithm	z_j	p_j	$\alpha/(k-j+1)$
2	**TIFS**	-7.089537	1.3456e-12	0.00625
3	**GLFS**	-5.478279	4.2948e-8	0.00714
4	**SLOFS**	-5.263444	1.4138e-7	0.00833
5	**MDFS**	-3.86702	1.1017e-4	0.01000
6	**HOMI**	-2.721237	0.006504	0.01250
7	**MC-GM**	-2.578014	0.009937	0.01667
8	RLFSCL	-1.897704	0.057735	0.02500
9	WRAP	-1.396424	0.162587	0.05000

Tables 11–15 enumerate the results of Holm's procedure. We can infer that GOFA is the most statistically superior at metric *One Error*. Additionally, it achieves statistical superior performance over TIFS, MC-GM, SLOFS, GLFS, MDFS on all five metrics, HOMI on all metrics except for *Ranking Loss*, RLFSCL on *Ranking Loss* and *One Error*.

4.4. Sensitivity analysis

To explore the performance fluctuations w.r.t. parameter settings, we conduct sensitivity analysis for GOFA on datasets *cal500*, *computers*, and *scene* over the parameters μ and k in Figs. 2–6. The parameter μ adjusts the relative importance between the Pearson correlation coefficient and cosine similarity and is searched from 0.1 to 0.9 at a step of 0.1. Meanwhile, the parameter k controls the size of the nearest neighborhood, and is searched from 3 to 12 at a step of 1.

The results clearly illustrate that the best performance is both dataset-dependent and metric-dependent. Typically, the results of *cal500* become best if μ is medium while k is smaller; the results of *computers* become best if both μ and k are larger; the results of *scene* become best if μ is larger while k is smaller. It implies that Pearson correlation coefficient contributes more than cosine distance in measuring instance similarity, and a reasonable setting of k can consolidate the robustness of augmented features.

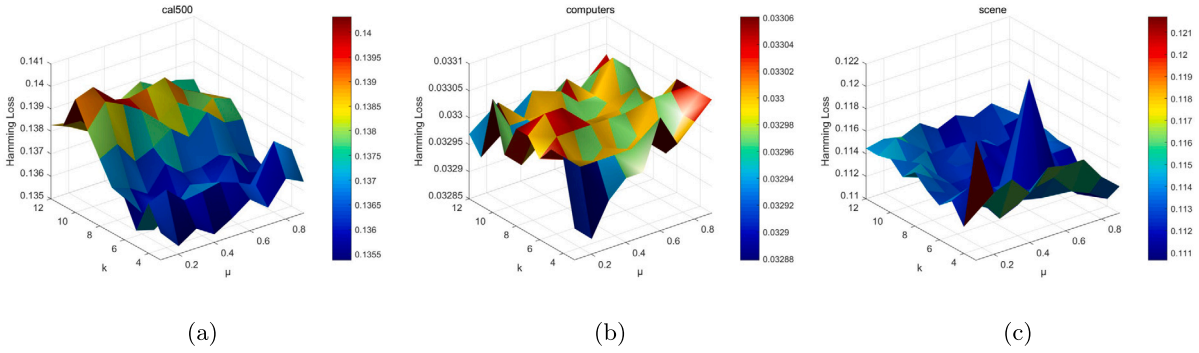

Fig. 2. *Hamming Loss* with varying parameters μ and k on (a) *cal500*, (b) *computers*, and (c) *scene*.

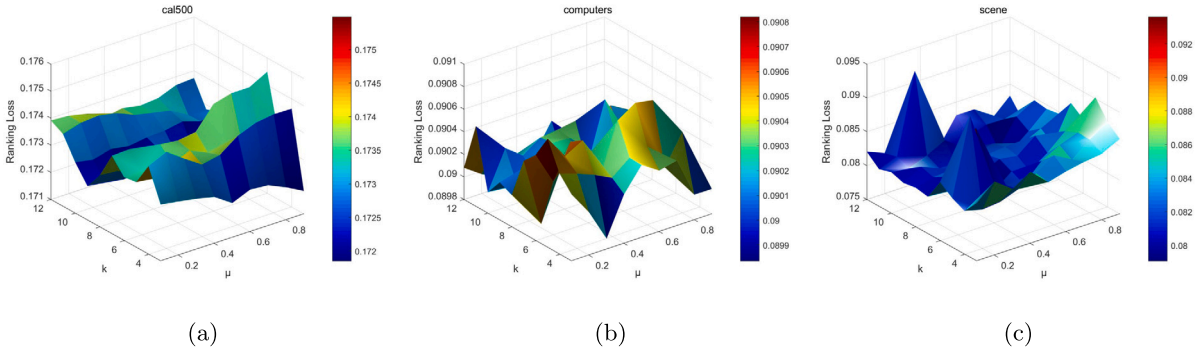

Fig. 3. *Ranking Loss* with varying parameters μ and k on (a) *cal500*, (b) *computers*, and (c) *scene*.

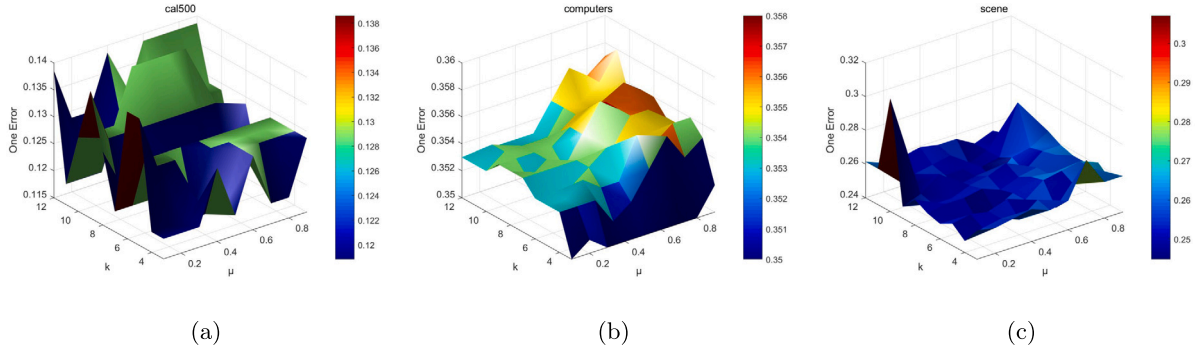

Fig. 4. *One Error* with varying parameters μ and k on (a) *cal500*, (b) *computers*, and (c) *scene*.

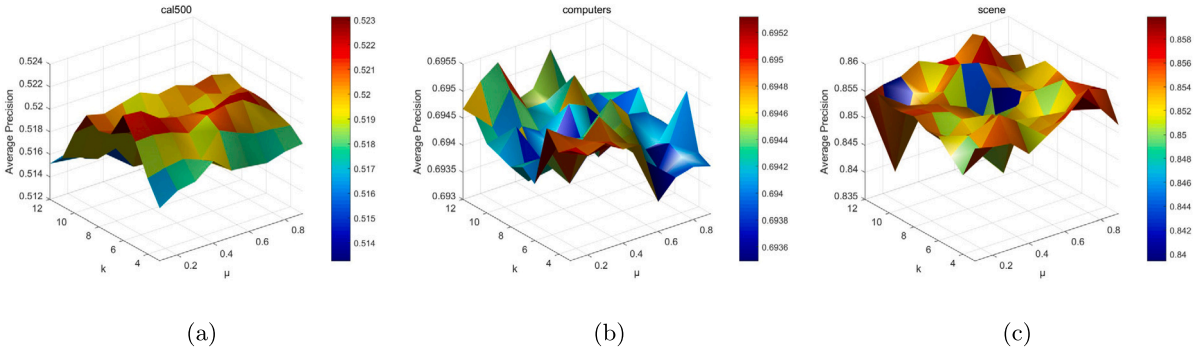

Fig. 5. *Average Precision* with varying parameters μ and k on (a) *cal500*, (b) *computers*, and (c) *scene*.

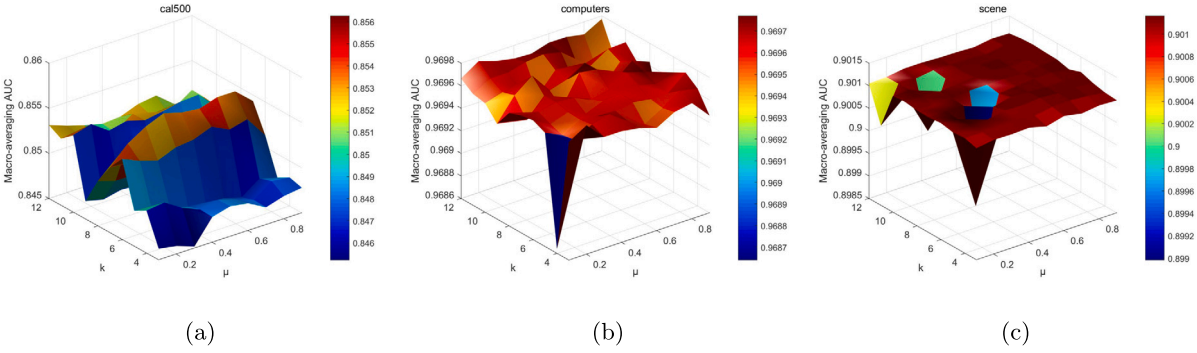

Fig. 6. *Macro-averaging AUC* with varying parameters μ and k on (a) *cal500*, (b) *computers*, and (c) *scene*.

5. Conclusions

We present a novel label-specific feature augmentation by constructing correlation-based granulation for multi-label classification. Unlike conventional multi-label classification solutions that emphasize discriminative feature learning merely from the raw data representation, we resort to the underlying instance-level and class-level correlation locally and refine the granularity of original features by feature concatenation. In this way, the discrimination differences between positive and negative classes across labels become explicit. We propose that these features bridge the information gap between low-level features and high-level concepts. The feature concatenation allows the classifier to automatically identify discriminative components. This process simulates the ability to handle uncertainty by leveraging various levels of granularity simultaneously. Extensive experiments validate the usefulness of augmented features. Specifically, we demonstrate that it is conducive to boosting the WRAP method if we incorporate the augmented features.

Despite the impressive classification performance, some open issues remain. In GOFA, we assume that all features contribute to the instance-based neighborhood construction. However, some features may contribute quite limited for a particular label, thus incurring the deformation of the nearest neighborhood structure. An alternative solution is to explore the neighborhood based on label-specific features. Furthermore, it is worth mentioning that GOFA depends on the quality of both ground-truth labels and the explicit features.

However, they may be degenerated in cases like noisy features and missing labels in practical applications. The biased tendency is two-fold. Firstly, the noisy features incur untrustworthy similarity among instances, which offers erroneous components and weights in the instance-based neighborhood. Secondly, the missing labels incur the indistinguishable positive and negative distribution among labels, which offers biased label correlation in classifier construction. The estimation of latent representation learning may be more desirable. We will consider the variation of GOFA to accommodate such cases in the future.

CRediT authorship contribution statement

Tianna Zhao: Writing – original draft, Software, Conceptualization, Investigation, Visualization, Formal analysis. **Yuanjian Zhang:** Writing – review & editing, Methodology, Conceptualization, Validation. **Duoqian Miao:** Supervision, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

Authors would like to thank the anonymous reviewers for their constructive comments and valuable suggestions. This work is supported by the National Key Research and Development Program of China (Grant No. 2022YFB3104700), National Natural Science Foundation of China (Grant No. 62376198).

References

[1] X.Y. Che, D.G. Chen, J. Deng, et al., Exploiting local label correlation from sample perspective for multi-label classification via three-way decision theory, Appl. Soft Comput. 149 (2023) 110950.
[2] X.Y. Che, D.G. Chen, J.S. Mi, Feature distribution-based label correlation in multi-label classification, Int. J. Mach. Learn. Cybern. 12 (2021) 1705–1719.
[3] X.Y. Che, D.G. Chen, J.S. Mi, Learning instance-level label correlation distribution for multilabel classification with fuzzy rough sets, IEEE Trans. Fuzzy Syst. 31 (8) (2023) 2871–2884.
[4] J.F. Chen, R.C. Zhang, J. Xu, et al., A neural expectation-maximization framework for noisy multi-label text classification, IEEE Trans. Knowl. Data Eng. 35 (11) (2023) 10992–11003.
[5] J. Demsar, Statistical comparisons of classifiers over multiple data sets, J. Mach. Learn. Res. 7 (2006) 1–30.
[6] X. Gong, Z. Yao, X. Li, et al., LAG-Net: multi-granularity network for person re-identification via local attention system, IEEE Trans. Multimed. 24 (2022) 217–229.
[7] W.L. Guo, S.Y. Li, J. Yang, Remote sensing image scene classification by multiple granularity semantic learning, IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 15 (2022) 2546–2562.
[8] J. Hobbs, Granularity, in: Proceedings of the 9th International Joint Conference on Artificial Intelligence, 1985, pp. 432–435.
[9] J. Huang, G.R. Li, Q.M. Huang, et al., Learning label-specific features and class-dependent labels for multi-label classification, IEEE Trans. Knowl. Data Eng. 28 (12) (2016) 3309–3323.
[10] T. Huang, X.A. Tang, S.Y. Zhao, et al., Linguistic information-based granular computing based on a tournament selection operator-guided PSO for supporting multi-attribute group decision-making with distributed linguistic preference relations, Inf. Sci. 610 (2022) 488–507.
[11] B.B. Jia, M.L. Zhang, Multi-dimensional classification via kNN feature augmentation, Pattern Recognit. 106 (2020) 107423.
[12] B.B. Jia, M.L. Zhang, Multi-dimensional classification via selective feature augmentation, Mach. Intell. Res. 19 (1) (2022) 38–51.
[13] X. Kang, X.F. Shi, Y.N. Wu, et al., Active learning with complementary sampling for instructing class-biased multi-label text emotion classification, IEEE Trans. Affect. Comput. 14 (1) (2023) 523–536.
[14] B.Z. Li, Z.H. Wei, D.Q. Miao, et al., Improved general attribute reduction algorithms, Inf. Sci. 536 (2020) 298–316.
[15] D.C. Liang, W. Pedrycz, D. Liu, Determining three-way decisions with decision-theoretic rough sets using a relative value approach, IEEE Trans. Syst. Man Cybern. Syst. 47 (8) (2017) 1785–1799.
[16] Y.J. Lin, Q.H. Hu, J.H. Liu, et al., MULFE: multi-label learning via label-specific feature space ensemble, ACM Trans. Knowl. Discov. Data 16 (1) (2022) 5.
[17] W.W. Liu, H.B. Wang, X.B. Shen, et al., The emerging trends of multi-label learning, IEEE Trans. Pattern Anal. Mach. Intell. 44 (11) (2022) 7955–7974.
[18] Z.F. Liu, C.J. Tang, S.E. Abhadiomhen, et al., Robust label and feature space co-learning for multi-label classification, IEEE Trans. Knowl. Data Eng. 35 (11) (2023) 11846–11859.
[19] J.H. Ma, B.C.Y. Chiu, T.W.S. Chow, Multilabel classification with group-based mapping: a framework with local feature selection and local label correlation, IEEE Trans. Cybern. 52 (6) (2022) 4596–4610.
[20] J.H. Ma, T.W.S. Chow, Topic-based instance and feature selection in multilabel classification, IEEE Trans. Neural Netw. Learn. Syst. 33 (1) (2022) 315–329.
[21] S. Nazmi, X.Y. Yan, A. Homaifar, et al., Evolving multi-label classification rules by exploiting high-order label correlations, Neurocomputing 417 (2020) 176–186.
[22] W.B. Qian, W.Y. Ruan, Y.H. Li, et al., Granular ball-based label enhancement for dimensionality reduction in multi-label data, Appl. Intell. 53 (20) (2023) 24008–24033.
[23] J.D. Qin, L. Martinez, W. Pedrycz, et al., An overview of granular computing in decision-making: extensions, applications, and challenges, Inf. Fusion 98 (2023) 101833.
[24] Z. Qin, H.M. Chen, Y. Mi, et al., Multi-label feature selection with adaptive graph learning and label information enhancement, Knowl.-Based Syst. 285 (2024) 111363.
[25] R. Schapire, Y. Singer, A boosting-based system for text categorization, Mach. Learn. 39 (2/3) (2000) 135–168.
[26] R.H. Shang, J.Y. Zhong, W.T. Zhang, et al., Multilabel feature selection via shared latent sublabel structure and simultaneous orthogonal basis clustering, IEEE Trans. Neural Netw. Learn. Syst. (2024), https://doi.org/10.1109/TNNLS.2024.3382911.

[27] C.J. Si, Y.H. Jia, R. Wang, et al., Multi-label classification with high-rank and high-order label correlations, IEEE Trans. Knowl. Data Eng. 36 (8) (2023) 4076–4088.
[28] W. Siblini, P. Kuntz, F. Meyer, A review on dimensionality reduction for multi-label classification, IEEE Trans. Knowl. Data Eng. 33 (3) (2021) 839–857.
[29] A.H. Tan, X.W. Ji, J.Y. Liang, et al., Weak multi-label learning with missing labels via instance granular discrimination, Inf. Sci. 594 (2022) 200–216.
[30] C.Z. Wang, Y. Wang, T.Q. Deng, Missing multi-label learning based on the fusion of two-level nonlinear mappings, Inf. Fusion 103 (2024) 102105.
[31] J.M. Wu, E.C.C. Tsang, W.H. Xu, et al., Correlation concept-cognitive learning model for multi-label classification, Knowl.-Based Syst. 290 (2024) 111566.
[32] M.K. Xie, S.J. Huang, Partial multi-label learning with noisy label identification, IEEE Trans. Pattern Anal. Mach. Intell. 44 (7) (2022) 3676–3687.
[33] Z.C. Xue, Z.L. Zhang, H. Liu, Learning knowledge graph embedding with multi-granularity relational augmentation network, Expert Syst. Appl. 233 (2023) 120953.
[34] L. Yi, L. Zhang, X.Y. Xu, et al., Multi-label softmax networks for pulmonary nodule classification using unbalanced and dependent categories, IEEE Trans. Med. Imaging 42 (1) (2023) 317–328.
[35] Y. Yu, M.Y. Lv, J. Qian, et al., Fuzzy information gain ratio-based multi-label feature selection with label correlation, Int. J. Mach. Learn. Cybern. 15 (2024) 2737–2747.
[36] Z.B. Yu, M.L. Zhang, Multi-label classification with label-specific feature generation: a wrapped approach, IEEE Trans. Pattern Anal. Mach. Intell. 44 (9) (2022) 5199–5210.
[37] J.M. Zhan, J. Deng, Z.S. Xu, et al., A three-way decision methodology with regret theory via triangular fuzzy numbers in incomplete multiscale decision information systems, IEEE Trans. Fuzzy Syst. 31 (8) (2023) 2773–2787.
[38] C.Q. Zhang, Z.W. Yu, H.Z. Fu, et al., Hybrid noise-oriented multilabel learning, IEEE Trans. Cybern. 50 (6) (2020) 2837–2850.
[39] H.Y. Zhang, M. Li, D.Q. Miao, et al., Construction of a feature enhancement network for small object detection, Pattern Recognit. 143 (2023) 109801.
[40] J. Zhang, Z.M. Luo, C.D. Li, et al., Manifold regularized discriminative feature selection for multi-label learning, Pattern Recognit. 95 (2019) 136–150.
[41] J. Zhang, H.R. Wu, M. Jiang, et al., Group-preserving label-specific feature selection for multi-label learning, Expert Syst. Appl. 213 (2023) 118861.
[42] M.L. Zhang, Z.H. Zhou, ML-KNN: a lazy learning approach to multi-label learning, Pattern Recognit. 40 (7) (2007) 2038–2048.
[43] M.L. Zhang, L. Wu, LIFT: multi-label learning with label-specific features, IEEE Trans. Pattern Anal. Mach. Intell. 37 (1) (2015) 107–120.
[44] M.L. Zhang, Z.H. Zhou, A review on multi-label learning algorithms, IEEE Trans. Knowl. Data Eng. 26 (8) (2014) 1819–1837.
[45] M.L. Zhang, Q.W. Zhang, J.P. Fang, et al., Leveraging implicit relative labeling-importance information for effective multi-label learning, IEEE Trans. Knowl. Data Eng. 33 (5) (2021) 2057–2070.
[46] X.Y. Zhang, F. Min, G.J. Song, et al., LSTC: when label-specific features meet third-order label correlations, Inf. Sci. 632 (2023) 617–636.
[47] Y.J. Zhang, T.N. Zhao, D.Q. Miao, et al., Granular multilabel batch active learning with pairwise label correlation, IEEE Trans. Syst. Man Cybern. Syst. 52 (5) (2022) 3079–3091.
[48] L.Q. Zhao, Y.Y. Yao, L. Zhang, Measurement of general granules, Inf. Sci. 415–416 (2017) 128–141.
[49] T.N. Zhao, Y.J. Zhang, D.Q. Miao, et al., Multi-granular labels with three-way decisions for multi-label classification, Int. J. Mach. Learn. Cybern. 14 (2023) 3737–3752.
[50] H. Zhou, C.Y. Zhang, Y. Luo, et al., Thinking inside uncertainty: interest moment perception for diverse temporal grounding, IEEE Trans. Circuits Syst. Video Technol. 32 (10) (2022) 7190–7203.