A Robust Distance with Correlated Metric Learning for Multi-Instance Multi-Label Data

Yashaswi Verma
CVIT, IIIT Hyderabad, India
yashaswi.verma@research.iiit.ac.in

C. V. Jawahar
CVIT, IIIT Hyderabad, India
jawahar@iiit.ac.in

ABSTRACT

In multi-instance data, every object is a bag that contains multiple elements or instances. Each bag may be assigned to one or more classes, such that it has at least one instance corresponding to every assigned class. However, since the annotations are at bag-level, there is no direct association between the instances within a bag and the assigned class labels, hence making the problem significantly challenging.

While existing methods have mostly focused on Bag-to-Bag or Class-to-Bag distances, in this paper, we address the multiple instance learning problem using a novel Bag-to-Class distance measure. This is based on two observations: (a) existence of outliers is natural in multi-instance data, and (b) there may exist multiple instances within a bag that belong to a particular class. In order to address these, in the proposed distance measure (a) we employ L_1-distance that brings robustness against outliers, and (b) rather than considering only the most similar instance-pair during distance computation as done by existing methods, we consider a subset of instances within a bag while determining its relevance to a given class. We parameterize the proposed distance measure using class-specific distance metrics, and propose a novel metric learning framework that explicitly captures inter-class correlations within the learned metrics. Experiments on two popular datasets demonstrate the effectiveness of the proposed distance measure and metric learning.

CCS Concepts

•**Computing methodologies** → *Computer vision tasks; Machine learning;*

Keywords

Multiple instance learning; Distance metric learning

1. INTRODUCTION

In the conventional image categorization task, each image is considered as a single instance, and is assigned to one (single-instance single-label problem) or more (single-instance multi-label problem) classes. However, this ignores the fact that usually only some region(s) in an image correspond(s) to a particular semantic concept. E.g., as shown in Figure 1, though the images are tagged with several classes, each of these concepts represent only few specific region(s).

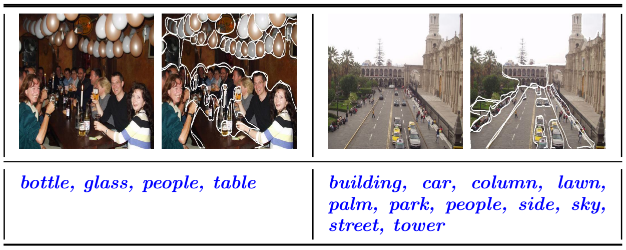
bottle, glass, people, table
building, car, column, lawn, palm, park, people, side, sky, street, tower

Figure 1: Samples from IAPR TC-12 dataset [6], which show that a whole image is not representative of all the assigned classes. Moreover, there could be multiple instances within an image that denote a particular semantic concept. E.g., there are multiple segments in the left and right pairs that belong to the classes "people" and "car" respectively.

Multiple Instance Learning (MIL) [2] is a machine learning paradigm that has lately achieved significant attention [17, 8, 10, 11, 1, 19, 14, 16, 15]. In MIL, each object/sample is assumed to be a bag consisting of a collection of instances, and is assigned to one or more classes such that there is at least one instance corresponding to every assigned class. MIL framework has been shown to be particularly useful for capturing inherent structure in data, such as indirect associations between an image's regions and its assigned classes [8, 14, 15], or reducing the labeling cost of videos [16]. However, along with these advantages, it also presents several challenges [8, 10, 14, 16, 15]. First, computing distance between two object-bags is not straightforward since they are *bags* of multiple instances rather than a single instance, and traditional distance measures such as Manhattan distance or Euclidean distance cannot be applied directly. Second, there exists a weak association between the instances of a bag and the classes assigned to it. Due to this, learning class-specific models remains non-trivial. And third, by the definition of MIL, an object bag is assigned to a class if at least one of its instances belongs to that class. In such a scenario, most of the other instances which do not represent that class act as outliers for the class under consideration.

MM '16, October 15-19, 2016, Amsterdam, Netherlands
© 2016 ACM. ISBN 978-1-4503-3603-1/16/10...$15.00
DOI: http://dx.doi.org/10.1145/2964284.2967259

Most of the existing MIL methods [17, 8, 10, 19, 14, 16, 15] have focused on developing sophisticated distance measures, and/or learning dataset-specific or class-specific distance metrics (also called metric learning) for MIL. While metric learning for single-instance data (single-label [5, 4, 18] or multi-label [7, 13]) is a well-studied topic, there have been few attempts that perform metric learning for multi-instance data. The first metric learning formulation for multi-instance data was proposed in [10]. Since it is based on computing bag-to-bag (or B2B) distance, there exists no direct association between instances within a bag and their class labels. Hence, to overcome this, it learns a single distance metric by simultaneously learning associations between instances and class labels, which results into a complex optimization problem. After this, few other methods have focused on learning distance metric for multi-instance data by using either B2B distance [8] or class-to-bag (or C2B) distance [14, 16, 15]. Since [14, 16, 15] are based on C2B distance, they were able to learn class-specific distance metrics unlike [10]. They represent every class by a *super-bag*, which consists of all the instances from the object-bags that belong to that class. Given a new object bag and a particular class, its distance from every instance of the corresponding super-bag is computed using the class-specific distance metric. Finally, the classes are assigned based on decreasing order of their distance from that object bag.

In this paper, we present a novel bag-to-class (or B2C) distance measure for computing distance of an object-bag from a super-bag. The major differences between the proposed distance measure and those in the earlier works are:
(1) Instead of using L_2-distance [17, 8, 10, 14, 16, 15], we propose to use L_1-distance while computing distance between two instances. This is because L_1-distance is known to be more robust against outliers than L_2-distance, which are frequent in multi-instance data.
(2) Rather than using C2B distance [14, 15, 16] or B2B distance [10, 8], our distance measure is based on computing B2C distance. This is because while C2B distance may be affected by the presence of large number of outliers within a super-bag, B2B distance complicates learning distance metrics since it requires computing distance between bags belonging to different classes which involves distance metrics from different classes (as also mentioned in [15]).
(3) While existing methods such as [8, 10, 14, 16, 15] consider only the single most similar pair of instances for computing C2B or B2B distance, we compute the distance of an object-bag from a given super-bag based on a subset of top few instances from both the sets that are most similar to each other, with similarity being computed from object-bag towards super-bag. This is based on the observation that there could be multiple instances within a bag that denote a particular semantic concept (Figure 1). Each of these in turn may match with a potentially different subset of instances within the super-bag under consideration. We believe that this further adds robustness to our distance measure since instead of relying only on the most similar instance which may be an outlier, we consider a set of top few most similar instances while computing the B2C distance.

Due to these considerations specifically adopted to introduce robustness while computing distance in MIL setting, we call our proposed distance as Robust B2C (or RB2C) distance. Other important contributions of this work are: (1) Similar to [14, 16, 15], we integrate class-specific distance metrics into RB2C-distance. (2) We formulate a novel metric learning framework that explicitly captures inter-class correlations, which are particularly observed in multi-label datasets. To our knowledge, this is the first metric learning formulation of its kind in this domain.

To validate our approach, we extensively experiment on two popular multi-label datasets: Corel-5K [3] and IAPR TC-12 [6]. Experiments demonstrate that the baseline RB2C-distance itself outperforms most of the existing techniques, and achieves further improvements after incorporating learned distance metrics.

2. PROPOSED DISTANCE MEASURE

In this section, first we briefly discuss the multi-instance setting [14, 15, 16]. Then, we describe the proposed RB2C-distance for multi-instance data.

2.1 Preliminaries

Let $\mathcal{D} = \{(X_1, \mathbf{y}_1), \ldots, (X_{|\mathcal{D}|}, \mathbf{y}_{|\mathcal{D}|})\}$ be a dataset consisting of $|\mathcal{D}|$ input-output pairs and L classes. Each $X_i = [\mathbf{x}_{i1}, \ldots, \mathbf{x}_{in_i}]$ is a bag of n_i instances, where $\mathbf{x}_{ij}$ is the j^{th} instance of X_i and is represented by an N-dimensional feature vector ($\mathbf{x}_{ij} \in \mathbb{R}^N\ \forall j = \{1, \ldots, n_i\}$). Each output vector $\mathbf{y}_i \in \{0,1\}^L$ is a binary vector. Under the MIL setting, for a given bag X_i, if $\exists j \in \{1, \ldots, n_i\}$ such that the instance $\mathbf{x}_{ij}$ belongs to the l^{th} class ($1 \leq l \leq L$), then the whole bag X_i belongs to the l^{th} class and $\mathbf{y}_i(l) = 1$; otherwise $\mathbf{y}_i(l) = 0$. Also, if $\sum_{l=1}^{L} \mathbf{y}_i(l) = 1$, $\forall i \in \{1, \ldots, |\mathcal{D}|\}$, then each bag X_i belongs to exactly one class and the dataset $\mathcal{D}$ is a single-label dataset. While, if $\sum_{l=1}^{L} \mathbf{y}_i(l) \geq 1\ \forall i \in \{1, \ldots, |\mathcal{D}|\}$, then each bag X_i may belong to one or more classes, and the dataset $\mathcal{D}$ is a multi-label dataset. Thus, single-label data is a special case of multi-label data.

2.2 Robust B2C Distance for MIL

Let each class l be represented by a super-bag U_l that consists of all instances from all the training bags that belong to that class [14, 15, 16]. Precisely,

$$U_l = \{\mathbf{x}_{ij} \mid \mathbf{y}_i(l) = 1\} \tag{1}$$

Let $m_l = |U_l| = \sum_{i \mid \mathbf{y}_i(l)=1} n_i$ be the size of the super-bag U_l. In case of single-label data where each bag belongs to exactly one class, all the super-bags are non-overlapping; i.e., (i) $\sum_{j=1}^{|\mathcal{D}|} n_j = \sum_{l=1}^{L} m_l$, and (ii) $|U_g \cap U_h| = 0$, $\forall\ g \neq h$. Whereas, in case of multi-label data where each bag may be labeled with one or more classes, there may be an overlap among different super-bags; i.e., (i) $\sum_{j=1}^{|\mathcal{D}|} n_j \leq \sum_{l=1}^{L} m_l$, and (ii) $|U_g \cap U_h| \geq 0$, $\forall\ g \neq h$.

Based on this, now we present the RB2C-distance for MIL. Given an object bag $A = [\mathbf{a}_1, \ldots, \mathbf{a}_{n_A}]$ and a class l, let $\mathbf{d}_{ilj}^{A} \in \mathbb{R}^N$ denote element-wise L_1-distance between an instance $\mathbf{a}_i$ of A and an instance $\mathbf{x}_{lj} \in U_l$. Precisely,

$$\mathbf{d}_{ilj}^{A}(k) = |\mathbf{a}_i(k) - \mathbf{x}_{lj}(k)|, \ \ \forall\ k \in \{1, \ldots, N\} \tag{2}$$

Using this, distance between the instances $\mathbf{a}_i$ and $\mathbf{x}_{lj}$ (instance-to-instance distance or D_{i2i}) is defined as

$$D_{i2i}(\mathbf{a}_i, \mathbf{x}_{lj}) = \sum_{k=1}^{N} |\mathbf{a}_i(k) - \mathbf{x}_{lj}(k)| = \mathbf{e}^T \mathbf{d}_{ilj}^{A}\ , \tag{3}$$

where $\mathbf{e} = [1, \ldots, 1]^T$ is a vector with all entries equal to 1. Let $\mathcal{N}_{il}^{K_1} \subseteq U_l$ be the set of the top K_1 nearest-neighbours of

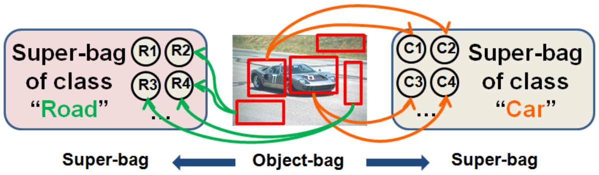

Figure 2: The proposed RB2C-distance is based on the assumption that a bag may have multiple instances (illustrated by red boxes) that represent some particular class. Additionally, each of these instances may match with a potentially different (may or may not be disjoint) subset of instances within the super-bag of the concerned class. E.g., in this figure, there are two instances in the object-bag that match with the super-bag of "car". Moreover, each of these matched instances match with two distinct sets of instances $\{C_1, C_2\}$ and $\{C_3, C_4\}$. Note that in both the datasets used in our experiments, an instance is a segment of an image. Here we use box just for illustrative purpose.

$\mathbf{a}_i$ from the super-bag U_l determined using Eq. 3. Then, the distance of instance $\mathbf{a}_i$ from super-bag U_l (instance-to-class distance or D_{i2c}) is defined as:

$$D_{i2c}(\mathbf{a}_i, U_l) = \frac{1}{K_1} \sum_{\mathbf{x}_{lj}\in\mathcal{N}_{il}^{K_1}} D_{i2i}(\mathbf{a}_i, \mathbf{x}_{lj}) \ . \qquad (4)$$

Now, let $\mathcal{N}_l^{K_2}$ be the set that contains the top K_2 instances of A with least distance from super-bag U_l computed using Eq. 4. Then, distance of A from U_l (bag-to-class distance or D_{b2c}) is defined as:

$$D_{b2c}(A, U_l) = \frac{1}{K_2} \sum_{\mathbf{a}_i\in\mathcal{N}_l^{K_2}} D_{i2c}(\mathbf{a}_i, U_l) \ . \qquad (5)$$

Substituting Eq. 3 and Eq. 4 into the above equation gives

$$D_{b2c}(A, U_l) = \frac{1}{K_1K_2}\ \mathbf{e}^T \left(\sum_{\mathbf{a}_i\in\mathcal{N}_l^{K_2}} \sum_{\mathbf{x}_{lj}\in\mathcal{N}_{il}^{K_1}} \mathbf{d}_{ilj}^A \right). \qquad (6)$$

The distance function defined by the above equation is our proposed RB2C-distance. Intuitively, it computes distance of a given object from a class based on distance between the few instances from both the bags that are most similar to each other, with similarity being computed from object bag towards super-bag. Figure 2 illustrates the gist of RB2C-distance for multi-instance data.

3. METRIC LEARNING FOR RB2C

The distance function defined in Eq. 6 is based on simple L_1-distance that considers every dimension of a feature vector as equally important. However, it is well-known that computing distance in a learned projected space better captures data properties as compared to unprojected space, and hence also improves quantitative performance. Here we discuss how to learn linear distance metrics for the proposed RB2C-distance. Also, since there exists large diversity among samples from different classes, we learn L different class-specific metrics $\mathbf{w}_l \in \mathbb{R}^N$, $\forall\ l \in \{1, \dots, L\}$. For this, we re-define the distance in Eq. 6 as:

$$D^{\mathbf{w}}_{b2c}(A, U_l) = \frac{1}{K_1K_2}\ \mathbf{w}_l^T \left(\sum_{\mathbf{a}_i\in\mathcal{N}_l^{K_2}} \sum_{\mathbf{x}_{lj}\in\mathcal{N}_{il}^{K_1}} \mathbf{d}_{ilj}^A \right). \qquad (7)$$

We shall refer this as metric based RB2C-distance (or M-RB2C). To learn the metrics $\mathbf{w}_l\ \forall l \in \{1, \dots, L\}$, we follow the approach of metric learning using pair-wise comparisons [18, 4, 5, 13]. For a given bag A, let $L_A^+ \subseteq \{1, \dots, L\}$ be the set of classes to which it belongs; and $L_A^- = \{1, \dots, L\} \setminus L_A^+$ be the set of classes to which it does not belong. We are interested in learning $\mathbf{w}_l$'s such that the distance of A from $U_l\ \forall\ l \in L_A^+$ is less than its distance from $U_k\ \forall\ k \in L_A^-$ by a margin. This results in the following constraints:

$$\forall l \in L_A^+, \forall\ k \in L_A^- : \quad D^{\mathbf{w}}_{b2c}(A, U_k) - D^{\mathbf{w}}_{b2c}(A, U_l) \geq 1 - \xi_{kl}^A \quad (8)$$

Let $\mathbf{W} = [\mathbf{w}_1, \dots, \mathbf{w}_L] \in \mathbb{R}^{N\times L}$ be the concatenation of all class-specific metrics. Based on the above constraints, we solve the following convex optimization problem:

$$OP1: \min_{\mathbf{W}, \xi_{kl}^A \geq 0} \ \frac{1}{2} Trace(\mathbf{W}^T\mathbf{W}) + C \sum_{A, l\in L_A^+, k\in L_A^-} \xi_{kl}^A$$

$$s.t.\quad \forall A, l \in L_A^+, k \in L_A^- : D^{\mathbf{w}}_{b2c}(A, U_k) - D^{\mathbf{w}}_{b2c}(A, U_l) \geq 1 - \xi_{kl}^A$$
$$\forall\ 1 \leq l \leq L,\ 1 \leq i \leq N : \quad \mathbf{w}_l(i) \geq 0$$

where ξ_{kl}^A are slack variables, and $C > 0$ is a constant that controls the trade-off between the two terms. The second set of constraints impose non-negativity on the elements of $\mathbf{w}_l$. This is necessary since it is used to define a distance metric, which should be a positive semi-definite operator. We optimize $OP1$ in the primal form itself using a batch gradient-descent and projection method similar to [18]. In the beginning, we initialize $\mathbf{w}_l = \mathbf{e}$, $\forall\ l \in \{1, \dots, L\}$.

4. CORRELATED METRIC LEARNING (CML)

In multi-label datasets where each bag may belong multiple classes, presence of one class gives hint about the presence of its correlated classes. E.g., presence of "car" may imply presence of "road". In $OP1$, a linear metric $\mathbf{w}_l$ is learned corresponding to every class l. However, it ignores to capture inter-class correlations into the learned metrics. In this section, we extend $OP1$ by explicitly incorporating inter-class correlations within the metric learning framework.

Consider the distance computation of a bag A from super-bag U_l as given in Eq. 7. There, let $\tilde{\mathbf{d}}_l^A \in \mathbb{R}^N$ denote the normalized sum of all distance vectors; i.e.,

$$\tilde{\mathbf{d}}_l^A = \frac{1}{K_1K_2} \sum_{\mathbf{a}_i\in\mathcal{N}_l^{K_2}} \sum_{\mathbf{x}_{lj}\in\mathcal{N}_{il}^{K_1}} \mathbf{d}_{ilj}^A \qquad (9)$$

Now, let $\tilde{\mathbf{D}}^A = [\tilde{\mathbf{d}}_1^A, \dots, \tilde{\mathbf{d}}_L^A] \in \mathbb{R}^{N\times L}$. Also, $\forall l \in \{1, \dots, L\}$, let $\mathbf{e}_l \in \mathbb{R}^L$ be a binary vector with $\mathbf{e}_l(l) = 1$ and rest all entries being 0. Then, it is easy to verify that the constraints given by Eq. 8 are equivalent to

$$\forall l \in L_A^+, \forall\ k \in L_A^- : (\mathbf{W}\mathbf{I}\mathbf{e}_k)^T\tilde{\mathbf{D}}^A\mathbf{e}_k - (\mathbf{W}\mathbf{I}\mathbf{e}_l)^T\tilde{\mathbf{D}}^A\mathbf{e}_l \geq 1 - \xi_{kl}^A$$

where $\mathbf{I}$ denotes the identity matrix. Let $\mathbf{P} \in \mathbb{R}^{L\times L}$ be a matrix such that $\mathbf{P}(k, l)$ denotes the correlation between the k^{th} and l^{th} classes. Using $\mathbf{P}$, we introduce inter-class correlations into the above constraints as follows:

$$\forall l \in L_A^+, \forall\, k \in L_A^- : (\mathbf{WPe}_k)^T \tilde{\mathbf{D}}^A \mathbf{e}_k - (\mathbf{WPe}_l)^T \tilde{\mathbf{D}}^A \mathbf{e}_l \geq 1 - \xi_{kl}^A$$

Now, let $\mathbf{V} = \mathbf{WP} \in \mathbb{R}^{N \times L}$, and $\mathbf{R} = \mathbf{P}^T\mathbf{P} \succ 0$. Then, the above constraints and the identity $Trace(\mathbf{G}_1\mathbf{G}_2\mathbf{G}_3) = Trace(\mathbf{G}_3\mathbf{G}_1\mathbf{G}_2)$ result into the following optimization:

$$OP2 : \min_{\mathbf{V}, \xi_{kl}^A \geq 0} \frac{1}{2} Trace(\mathbf{V}\mathbf{R}^{-1}\mathbf{V}^T) + C \sum_{A, l \in L_A^+, k \in L_A^-} \xi_{kl}^A$$

$$s.t.\ \forall A, l \in L_A^+, k \in L_A^- : (\mathbf{Ve}_k)^T \tilde{\mathbf{D}}^A \mathbf{e}_k - (\mathbf{Ve}_l)^T \tilde{\mathbf{D}}^A \mathbf{e}_l \geq 1 - \xi_{kl}^A$$
$$\forall\, 1 \leq l \leq L,\ 1 \leq i \leq N : \quad \mathbf{v}_l(i) \geq 0$$

where $\mathbf{v}_l$ denotes the l^{th} column of $\mathbf{V}$. Similar to $OP1$, we use a batch gradient-descent and projection method to solve $OP2$. Thus, $\mathbf{v}_l$ is the new learned distance metric that explicitly encodes correlations of the l^{th} class with all other classes. Moreover, since each of these metrics are learned jointly in a max-margin discriminative manner, it also ensures optimal predictive performance. One can observe that if the prior matrix $\mathbf{P}$ is taken to be $\mathbf{I}$ (i.e., there are no prior correlations among the classes), then $OP2$ becomes the same as $OP1$, which makes $OP1$ a special case of $OP2$. Also, since there is no particular restriction on the matrix $\mathbf{R}$ except that it must be a positive-definite matrix, it can be either sparse or dense depending on the given application.

Using these new class-correlated metrics, we update the distance function of Eq. 7 by replacing $\mathbf{w}_l$ with $\mathbf{v}_l$ as below:

$$D_{b2c}^{\mathbf{v}}(A, U_l) = \mathbf{v}_l^T \left(\sum_{\mathbf{a}_i \in \mathcal{N}_l^{K_2}} \sum_{\mathbf{x}_{lj} \in \mathcal{N}_{il}^{K_1}} \mathbf{d}_{ilj}^A \right). \qquad (10)$$

We shall refer the above distance as correlated metric based RB2C (or CM-RB2C). We believe ours is the first work that incorporates inter-class correlations into metric learning framework for multi-instance multi-label data.

The proposed CML formulation is motivated by [9] that incorporates inter-class correlations into SVM classifiers. Our framework differs from theirs in two important ways: (1) Their approach is meant to learn binary one-vs.-rest SVM classifiers, while our aim is to learn class-specific distance-metrics in a nearest-neighbour scenario. (2) Since they learn one-vs.-rest SVM classifiers, their optimization problem boils-down into disjoint optimization problems for individual classes, which can be optimized easily. However, in our case, the pair-wise constraints defined over pairs of positive and negative classes for each bag (Eq. 10) result in a more complex optimization problem, where we need to learn metrics for all the classes simultaneously in a joint manner.

Defining P: Since $\mathbf{R}$ is a positive-definite matrix, the capacity of the prior correlation matrix $\mathbf{P}$ is quite large (it can also have negative entries to model negative correlations). We define the correlation between k^{th} and l^{th} class as [12]:

$$\mathbf{P}(k, l) = \frac{f_{kl}}{f_k + f_l - f_{kl}}\ , \qquad (11)$$

where f_k and f_l denote the frequencies of the k^{th} and l^{th} classes respectively, and f_{kl} denotes their co-occurrence frequency. Higher the value of $P(k, l)$, more is the correlation between these two classes, and vice-versa.

Label Prediction: Given an object A, we compute its distance from every class using Eq. 6, 7, and 10 for different forms of parameterization. Then we rank all the classes based on the decreasing order of their distance from A, with smaller distance implying higher relevance and vice-versa.

5. EXPERIMENTS

Now we evaluate and compare different variants of the proposed distance measures (Eq. 6, Eq. 7 and Eq. 10).

5.1 Datasets and Experimental Details

We use two popular multi-instance multi-label datasets Corel-5K [3] and IAPR TC-12 [6]. In both these datasets, each image (bag) consists of multiple segments (instances), and each segment is represented by a feature vector. We use the same train/test partitions for both the datasets as in [7, 13]. Corel-5K dataset consists of 4500 training images, 500 testing images, and a vocabulary of 260 classes. IAPR TC-12 dataset consists of 17665 training images, 1962 testing images, and a vocabulary of 291 classes. In all our experiments, we consider the top 20 most frequent classes from each dataset since others have very few occurrences.

We compare with following benchmark methods: (a) Citation kNN [17], (b) MIMLSVM [19], (c) MildML [8], (d) S-C2B and its variants (C2B and M-C2B) [15], (e) TagProp [7], and (f) 2PKNN [13]. For MIMLSVM [19], MildML [8], TagProp [7] and 2PKNN [13], we use publicly available codes. For Citation-kNN [17] and S-C2B (and its variants) [15], we have implemented these methods by following the details from the respective papers. For RB2C and its variants, we keep $K_1 = 5$ and $K_2 = 8$ for Corel-5K dataset, and $K_1 = 10$ and $K_2 = 4$ for IAPR TC-12 dataset.

Note that both TagProp and 2PKNN consider whole image as a single instance and used a set of 15 features [7]. This gives a 37152-dimensional feature vector per image, which is more than 100 times the number of features used by all other methods listed above. Since this would result in an unfair comparison, we project each of these 15 features using Principle Component Analysis (PCA) and keep only the top 5% dimensions from every feature. This gives a 1858-dimensional feature vector per image.

For quantitative evaluations, we use five popular metrics: Hamming loss (HL), One-error (OE), Coverage (Co), Ranking loss (RL) and Average precision (AP). More details on these measures can be found in [15, 16, 10]. Among these, while HL and OE measure multi-label classification performance, Co, RL and AP measure multi-label ranking performance, Also, for HL, OE, Co and RL, smaller score means better performance; and for AP, higher score means better performance. Since Hamming loss is based on binary assignment of classes, we assign the top μ classes to each test-bag while computing it. Here, μ is set to be the (rounded) average number of classes per bag in training data ($\mu = 2$ for Corel-5K dataset, and $\mu = 3$ for IAPR TC-12 dataset).

5.2 Results and Discussion

Table 1 compares the performance of different methods. We can observe that: (1) Citation-kNN, which although is quite old, achieves very competitive results and performs either better than or comparable to several recent methods. This demonstrates the effectiveness of this classical MIL method, and also reflects the need of revisiting such methods for developing better methods. (2) The simple RB2C-

Table 1: Performance using different methods (↓: lower is better; ↑: higher is better).

	Corel-5K					IAPR TC-12				
Method	**HL ↓**	**OE ↓**	**Co ↓**	**RL ↓**	**AP ↑**	**HL ↓**	**OE ↓**	**Co ↓**	**RL ↓**	**AP ↑**
Citation-kNN [17]	0.088	0.514	5.288	0.193	0.558	0.121	0.502	7.682	0.249	0.521
MIMLSVM [19]	0.100	0.590	5.849	0.195	0.530	0.133	0.593	8.929	0.281	0.456
MildML [8]	0.090	0.511	5.851	0.210	0.548	**0.120**	0.520	7.968	0.262	0.504
S-C2B [16]	0.519	0.669	7.094	0.240	0.438	0.487	0.608	9.706	0.304	0.420
C2B [16]	0.479	0.671	7.016	0.237	0.438	0.452	0.605	9.611	0.300	0.422
M-C2B [16]	0.465	0.664	6.988	0.229	0.445	0.439	0.597	9.497	0.293	0.431
TagProp [7]	0.122	0.603	5.217	0.174	0.520	0.168	0.546	7.406	0.223	0.516
2PKNN [13]	0.120	0.628	5.524	0.188	0.516	0.172	0.591	7.748	0.240	0.498
RB2C (this work)	0.107	0.468	4.207	0.131	0.607	0.148	0.480	6.912	0.199	0.554
M-RB2C (this work)	0.095	0.462	4.124	0.126	0.613	0.142	0.469	6.727	0.190	0.562
CM-RB2C (this work)	**0.087**	**0.457**	**4.086**	**0.119**	**0.620**	0.139	**0.460**	**6.659**	**0.184**	**0.568**

distance itself outperforms (sometimes significantly) most of the other methods, and the performance of M-RB2C is consistently better than RB2C. It further improves by using distance metrics learned with inter-class correlations (CM-RB2C), thus validating the utility of the proposed CML formulation. (3) CM-RB2C provides the best performance for both the tasks, except on IAPR TC-12 where its HL is slightly inferior to competing methods.

6. CONCLUSION

We have introduced a novel RB2C-distance for multi-instance multi-label data, and integrated class-specific distance metrics into it. Additionally, we have presented a novel metric learning framework that explicitly takes into account inter-class correlations in multi-label datasets, and also provide its principled interpretation. Our method demonstrates marginal improvements in performance on multi-label ranking and classification tasks compared to several benchmark techniques, thus confirming its effectiveness. Though in this work we have focused only on multi-instance multi-label data, the proposed CML framework can be adopted for single-instance multi-label data as well.

Acknowledgement: Yashaswi Verma is partially supported by Microsoft Research India PhD fellowship 2013.

7. REFERENCES

[1] Y. Chen and J. Z. Wang. Image categorization by learning and reasoning with regions. *The Journal of Machine Learning Research*, 5:913–939, 2004.

[2] T. Dietterich, R. Lathrop, and T. Lozano-Pérez. Solving the multiple-instance problem with axis-parallel rectangles. *Artificial Intelligence*, 89(1–2):31–71, 1997.

[3] P. Duygulu, K. Barnard, J. D. Freitas, and D. Forsyth. Object recognition as machine translation: Learning a lexicon for a fixed image vocabulary. In *ECCV*, 2002.

[4] A. Frome, Y. Singer, and J. Malik. Image retrieval and classification using local distance functions. In *NIPS*, 2007.

[5] A. Frome, Y. Singer, F. Sha, and J. Malik. Learning globally-consistent local distance functions for shape-based image retrieval and classification. In *ICCV*, 2007.

[6] M. Grubinger. *Analysis and Evaluation of Visual Information Systems Performance.* PhD thesis, Victoria University, Melbourne, Australia, 2007.

[7] M. Guillaumin, T. Mensink, J. Verbeek, and C. Schmid. Tagprop: Discriminative metric learning in nearest neighbor models for image auto-annotation. In *ICCV*, 2009.

[8] M. Guillaumin, J. Verbeek, and C. Schmid. Multiple instance metric learning from automatically labeled bags of faces. In *ECCV*, 2010.

[9] B. Hariharan, L. Z.-Manor, S. V. N. Vishwanathan, and M. Varma. Large scale max-margin multi-label classification with priors. In *ICML*, 2010.

[10] R. Jin, S. Wang, and Z. H. Zhou. Learning a distance metric from multi-instance multi-label data. In *CVPR*, 2009.

[11] O. Maron and A. L. Ratan. Multiple-instance learning for natural scene classification. In *ICML*, 1998.

[12] B. Sigurbjörnsson and R. van Zwol. Flickr tag recommendation based on collective knowledge. In *WWW*, 2008.

[13] Y. Verma and C. V. Jawahar. Image annotation using metric learning in semantic neighbourhoods. In *ECCV*, 2012.

[14] H. Wang, H. Huang, F. Kamangar, F. Nie, and C. Ding. Maximum margin multi-instance learning. In *NIPS*, 2011.

[15] H. Wang, F. Nie, and H. Huang. Learning instance specific distance for multi-instance classification. In *AAAI*, 2011.

[16] H. Wang, F. Nie, and H. Huang. Robust and discriminative distance for multi-instance learning. In *CVPR*, 2012.

[17] J. Wang and J. D. Zucker. Solving the multiple-instance problem: A lazy learning approach. In *ICML*, 2000.

[18] K. Q. Weinberger and L. K. Saul. Distance metric learning for large margin nearest neighbor classification. *The Journal of Machine Learning Research*, 10:207–244, 2009.

[19] Z. H. Zhou and M. L. Zhang. Multi-instance multi-label learning with application to scene classification. In *NIPS*, 2007.