The Thirty-Second AAAI Conference
on Artificial Intelligence (AAAI-18)

Semi-Supervised Bayesian Attribute Learning for Person Re-Identification

Wenhe Liu,[1] **Xiaojun Chang,**[2] **Ling Chen,**[1] **Yi Yang**[1]
[1]Centre for Artificial Intelligence, University of Technology Sydney, Sydney, Australia.
[2]Language Technologies Institute, Carnegie Mellon University, Pittsburgh, PA, USA.
{wenhe.liu, ling.chen, yi.yang}@uts.edu.au, cxj273@gmail.com

Abstract

Person re-identification (re-ID) tasks aim to identify the same person in multiple images captured from non-overlapping camera views. Most previous re-ID studies have attempted to solve this problem through either representation learning or metric learning, or by combining both techniques. Representation learning relies on the latent factors or attributes of the data. In most of these works, the dimensionality of the factors/attributes has to be manually determined for each new dataset. Thus, this approach is not robust. Metric learning optimizes a metric across the dataset to measure similarity according to distance. However, choosing the optimal method for computing these distances is data dependent, and learning the appropriate metric relies on a sufficient number of pair-wise labels. To overcome these limitations, we propose a novel algorithm for person re-ID, called semi-supervised Bayesian attribute learning. We introduce an Indian Buffet Process to identify the priors of the latent attributes. The dimensionality of attributes factors is then automatically determined by nonparametric Bayesian learning. Meanwhile, unlike traditional distance metric learning, we propose a *re-identification probability* distribution to describe how likely it is that a pair of images contains the same person. This technique relies solely on the latent attributes of both images. Moreover, pair-wise labels that are not known can be estimated from pair-wise labels that are known, making this a robust approach for semi-supervised learning. Extensive experiments demonstrate the superior performance of our algorithm over several state-of-the-art algorithms on small-scale datasets and comparable performance on large-scale re-ID datasets.

Introduction

In the field of computer vision, the study of person re-ID has attracted considerable attention in recent years (Zheng, Yang, and Hauptmann 2016). The main goal of re-ID is to use a *probe* set of images that capture a person from one camera view and re-identify that same person in a *gallery* set of images captured by other non-overlapping camera views. However, because the camera views do not overlap, there may be uncontrollable and/or unpredictable variations among the images in appearance, such as body pose, view angle, occlusion, illumination conditions, and so on. As a result, re-ID performance often degrades (Zheng, Yang, and Hauptmann 2016).

Existing approaches to re-ID have mainly focused on representation learning and/or metric learning to overcome these challenges. In representation learning, many frameworks learn a factor-based representation to enhance re-ID task performance (Kodirov, Xiang, and Gong 2015; Liu et al. 2014). Several recent works have turned to attribute learning methods for further improvement (Lin et al. 2017). In Re-ID, attributes are mid-level features shared by multiple instances, such as hair color or wearing/not wearing a dress. Overall, the general idea behind these methods is that there are only a certain number of feature subsets that contribute to image matching performance. In metric learning, algorithms learn a suitable metric in the given set of data, which is then used to measure similarity (Xiong et al. 2014; Hirzer et al. 2012).

Previous studies have demonstrated some exciting results, but there are still challenges associated with each approach. Determining the number of latent factors in factor-based representation models is a common problem. Typically, cross-validation forms the solution, where the model evaluates various numbers of latent factors that are manually predefined. However, this is a time-consuming task for large re-ID datasets, limiting the scalability of these methods. Another solution is to manually annotate the attributes to enhance learning performance (Lin et al. 2017; Su et al. 2016). However, on large-scale datasets, this method has high human labor costs.

Metric learning re-ID methods also have drawbacks. Because they rely on learning a metric suitable to the re-ID targets, the performance is sensitive to the given dataset. Additionally, choosing the optimal method for calculating similarity distances, e.g., using ℓ_1 norm or ℓ_2 norm, can be problematic (Wang, Nie, and Huang 2014). Moreover, metric learning methods rely on pair-wise label information, such that performance suffers when there are only a few labels (Yang, Jin, and Sukthankar 2012).

A few previous works have attempted to jointly apply both representation and metric learning to re-ID problems. Some use each technique independently to accomplish a specific goal. For instance, in (Liao et al. 2015; Zhang, Xiang, and Gong 2016), researchers use representation learning methods in pre-processing stage to generate useful features that can then be used in metric learning. Others combine both methods into a deep learning architecture. However, these methods still rely on pre-annotated attributes (Lin et al. 2017; Su et al. 2016) or labeled data (Ahmed, Jones, and Marks 2015).

To overcome these limitations with re-ID tasks, we propose a semi-supervised Bayesian attribute learning algorithm (SBAL). SBAL combines an Indian buffet process (IBP) (Ghahramani and Griffiths 2006) prior in an infinite latent factor model that enables adaptively learning attributes for re-ID (Broderick, Kulis, and Jordan 2013). Additionally, inspired by statistical relation learning, we also propose re-ID probability, which has been successfully used in knowledge graph learning on large-scale datasets, such as social networks (Nickel et al. 2016). Wrapped within a Bayesian framework, SBAL automatically determines the latent factors and simultaneously estimates a *re-identification probability*. The contributions of our work are as follows:

- We introduce IBP as the prior of latent factors for learning binary representations. A dictionary of attributes is adaptively determined using an efficient estimation method. Thus, our algorithm does not require the dimensionality of latent factors to be pre-defined, nor the attribute information to be pre-annotated for training, which are two major limitations of the existing frameworks.
- We propose a re-identification probability for predicting pair-wise relations in re-ID. The re-ID probability does not rely on distance computation and avoids the problem of determining the optimal method for computing distances inherent in traditional metric learning.
- We propose a Bayesian framework unifies representation learning and re-ID probability estimation and can simultaneously optimize both learning tasks.
- Our algorithm is also able to estimate unknown pair-wise labels using the probability distributions learned from known pair-wise labels, making our algorithm robust in semi-supervised learning scenarios.

Related Work

Various classical and state-of-the-art machine learning models have been proposed to solve problems with re-ID, such as (Ma and Li 2014; Xiao et al. 2016; Ahmed, Jones, and Marks 2015). Most existing re-ID methods can be classified into two categories: representation learning and metric learning. Representation learning methods aim to learn appropriate representations of captured images from different camera views to enhance re-ID.

One straightforward solution in representation learning is to learn the most representative features directly from images. Factor-based representation, such as dictionary learning (Kodirov, Xiang, and Gong 2015; Karanam, Li, and Radke 2015), has shown promising performance. Additionally, in recent years, some approaches have begun to incorporate deep learning (Xiao et al. 2016) with even better results. In some recently works (Lin et al. 2017; Su et al. 2015), it proposed manually annotating attributes for use in a deep learning framework. In (Su et al. 2016; Schumann and Stiefelhagen 2017), deep learning models were designed to be trained on separate datasets with attribute labels, then fine-tuned on target datasets without attribute labels. However, pre-training such algorithms is still limited by the number and type of attributes in the datasets that have been manually annotated.

Metric learning methods comprise distance metric learning (Hirzer et al. 2012; Liao et al. 2015; Liao and Li 2015) and learning-to-rank methods (Prosser et al. 2010; Paisitkriangkrai, Shen, and van den Hengel 2015). Metric learning aims to learn a metric that brings images of the same person closer together than images of different people. Learning-to-rank methods aim to rank the gallery of images given a probe dataset according to the likelihood that the same person is pictured. Both methods are highly reliant on pair-wise labeled data, and performance may degrade heavily when lacking of labeled pairs.

Previous studies have seldom considered combining both representation and metric learning to boost re-ID performance. Those algorithms combining both learning have typically separate the representation and metric learning stages. For instance, the joint learning methods in (Zhang, Xiang, and Gong 2016; Liao et al. 2015) use the Local Maximal Occurrence (LOMO) algorithm(Liao et al. 2015) for representation learning followed by a metric learning schema for re-ID. Neither simultaneously optimizes representation and metric learning. In (Su et al. 2016), Su et al. proposed a semi-supervised deep attribute learning (SSDAL) algorithm to enhance the metric learning method cross-view quadratic discriminant analysis (XQDA) proposed in (Liao et al. 2015). However, the representation and metric learning components are still separate sequential operations. A recent work in (Ahmed, Jones, and Marks 2015) uses a deep learning architecture to learn features and a corresponding similarity metric for re-ID. However, the dimensionality of features to be learned still needs to be pre-defined, as required by learning neural networks.

Some semi-supervised re-ID methods have also been proposed (Ma and Li 2014; Liu et al. 2014). Commonly, the training models in semi-supervised methods rely on both labeled and unlabeled data. Hence, they produce acceptable performance compared to supervised methods without an abundance of labeled data (Ma and Li 2014).

IBP is used as a nonparametric prior for infinite latent factor models (Ghahramani and Griffiths 2006; Broderick, Kulis, and Jordan 2013; Ghahramani and Griffiths 2006). In recently works, IBP was used as prior of the latent factors for matrix factorization (Xu, Zhu, and Zhang 2012) and link prediction (Zhu, Song, and Chen 2016).

The Proposed Methods

The following section formally describes each component of the proposed framework. This paper focuses on two-view re-ID problems, where data is recorded from two non-overlapping camera views. However, the schema is easy to extend to multi-view re-ID scenarios.

Let $\mathbf{X} = \{\mathbf{X}_1; \mathbf{X}_2\}$ be the training data, in which $\mathbf{X}_1 \in \mathbb{R}^{d\times n_1}$ and $\mathbf{X}_2 \in \mathbb{R}^{d\times n_2}$ are image sets of people from two non-overlapping cameras, camera 1 and 2, that containing

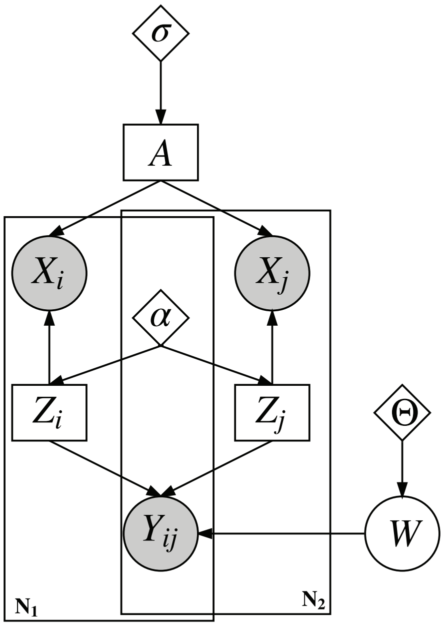

Figure 1: The plate notation of our model.

n_1 and n_2 images respectively. We also have a matrix of pair-wise labels $\mathbf{Y} \in \mathbb{R}^{n_1 \times n_2}$, where $\mathbf{y}_{ij} = 1$ if $\mathbf{x}_i$ and $\mathbf{x}_j$ are the same person; otherwise, $\mathbf{y}_{ij} = -1$. However, not all the pairs have labels, i.e., $\mathbf{Y}$ is not fully observed. Let $\mathbf{y}_{ij} = 0$ indicate the unknown pair-wise labels for observations $\mathbf{x}_i$ and $\mathbf{x}_j$. The set of pairs with known labels is denoted as $\mathcal{I} = \{(i,j)|y_{ij} \in \{-1,1\}\}$ and the set of pairs without labels is denoted as $\mathcal{U} = \{(i,j)|y_{ij} = 0\}$.

The first step is to learn representations of the training data with a Bayesian generative model. Let $\mathbf{A} \in \mathbb{R}^{d \times k}$ be a dictionary of basic patterns (attributes) on k basis. Let $\mathbf{Z} \in \mathbb{R}^{k \times n}$ be a binary representation matrix of $\mathbf{X}$ where $z_{ik} \in \{0,1\}$ and $z_{ik} = 1$ indicates the presence of attribute $\mathbf{a}_k$ for the image otherwise $\mathbf{x}_i$ and $z_{ik} = 0$. Given a set of images $\mathbf{X}$ we therefore have $\mathbf{X} \approx \mathbf{AZ}$. After learning a dictionary of attributes $\mathbf{A}$, the binary representation of a new image $\mathbf{x}$ can be obtained by $\mathbf{z} = \arg\min_{\hat{z} \in \{0,1\}} \|\mathbf{x} - \mathbf{A}\hat{z}\|_2^2$. The prior distributions of $\mathbf{A}$ and $\mathbf{X}$ are usually assumed to be Gaussian (Broderick, Kulis, and Jordan 2013):

$$\mathrm{P}(\mathbf{A}|0, \sigma_{\mathbf{A}}^2) = \prod_{k=1}^{K} \prod_{d=1}^{D} \mathcal{N}(a_{dk}; 0, \sigma_{\mathbf{A}}^2), \tag{1}$$

and

$$\mathrm{P}(\mathbf{X}|\mathbf{Z}, \mathbf{A}, \sigma_{\mathbf{X}}^2) = \prod_{n=1}^{N} \mathcal{N}(\mathbf{x}_i; \mathbf{A}z_i, \sigma_{\mathbf{X}}^2 I). \tag{2}$$

The above formulations assume that the dimensionality K of the latent factor $\mathbf{Z}$ is known as a priori. However, this assumption is often unrealistic in practice, particularly with large-scale datasets, as the possible attributes in image data become more complex when the size of the dataset increases. Conventional methods (Kodirov, Xiang, and Gong 2015; Li, Shao, and Fu 2015) usually include a model selection stage, such as cross-validation, to select an appropriate value for K by retraining and evaluating the model. This is an expensive process when the training data is large and may even miss the optimal value of K if it is outside the range of the search.

We overcome this problem by introducing Indian Buffet Process (IBP) as the prior of $\mathbf{Z}$. IBP is a nonparametric prior and has been widely used in infinite latent factor models (Broderick, Kulis, and Jordan 2013; Ghahramani and Griffiths 2006). These models are based on the assumption that an infinite number of latent factors have a distribution using an IBP prior. Considering finite latent factor models first, our model assumes there is a binary feature vector $\mathbf{z}_i$ with K elements for each instance $\mathbf{x}_i$, i.e., $\mathbf{z}_i \in \{0,1\}^K$. Further, we assume $\mathbf{Z}$ has a prior distribution of:

$$\mathrm{P}(\mathbf{Z}|\alpha) = \prod_{k=1}^{K} \frac{\frac{\alpha}{K}\Gamma(m_k + \frac{\alpha}{K})\Gamma(N - m_k + 1)}{\Gamma(N + 1 + \frac{\alpha}{K})}. \tag{3}$$

The binary latent factor z_{ik} is drawn from a Bernoulli distribution, Bernoulli(π_k), and parameterized by π_k. Furthermore, we assume π_k is sampled from a Beta distribution Beta$(\alpha/K, 1)$ where α is the hyper-parameter and K is the number of basis (i.e. attributes). $m_k = \sum_{i=1}^{N} z_k$ denotes the total number of times the kth attribute in the N samples is found. Then, according to the infinite assumption, i.e. letting $K \to \infty$, we obtain the IBP prior of the binary representations (Broderick, Kulis, and Jordan 2013):

$$\begin{aligned}&\lim_{K \to \infty} \mathrm{P}(\mathbf{Z}|\alpha) \\ &\quad = \frac{\alpha^{K_+} \exp(-\alpha H_N)}{K_+!} \prod_{k=1}^{K_+} \frac{(N - m_k)!(m_k - 1)!}{N!},\end{aligned} \tag{4}$$

where $H_N = \sum_{i=1}^{N} i^{-1}$ is the Nth harmonic number and K_+ denotes the number of determined attributes corresponding to the dataset $\mathbf{X}$. Several methods for inferring the prior in (4) have been proposed in previous works, such as sampling methods and variational methods (Ghahramani and Griffiths 2006). However, they can be computationally expensive when the number of instances N becomes large. As a more efficient alternative, we propose learning the joint probability $\mathrm{P}(\mathbf{X}, \mathbf{A}, \mathbf{Z}) = \mathrm{P}(\mathbf{X})\mathrm{P}(\mathbf{A})\mathrm{P}(\mathbf{Z})\mathrm{P}(\mathbf{X}|\mathbf{Z}, \mathbf{A})$ with an asymptotic limitation as in (Broderick, Kulis, and Jordan 2013). The details of this approach are provided in the next section.

In the second step, with the binary representations of images, we formalize the re-ID task as a probabilistic relation learning schema. The *re-identification probability* that the image representations $\mathbf{z}_i$ and $\mathbf{z}_j$ include the same person is calculated by

$$\mathrm{P}(y_{ij} = 1|\mathbf{Z}, \mathbf{W}) = \eta(\mathbf{z}_i W \mathbf{z}_j^T), \tag{5}$$

where $\mathbf{z}_i \in \mathbf{Z}_1$ and $\mathbf{z}_j \in \mathbf{Z}_2$, $\eta(v) = \frac{1}{1+\exp(-v)}$ is the sigmoid function. We assume the real value matrix $\mathbf{W} \in$

$\mathbb{R}^{K\times K}$ is drawn from a Gaussian prior:

$$\mathrm{P}(\mathbf{W}|\Theta,\sigma^2_{\mathbf{W}}) = \prod_{(k,k')\in\mathcal{I}} \mathcal{N}(w_{kk'};,\theta_{kk'},\sigma^2_{\mathbf{W}}). \tag{6}$$

For simplicity, we let $\sigma^2_{\mathbf{W}} = 1$. When both $z_{ik} = 1$ and $z_{jk'} = 1$, the element $w_{kk'}$ of $\mathbf{W}$ indicates the joint weight of the kth attribute in $\mathbf{z}_i$ and the k'th attribute in $\mathbf{z}_j$. Once $\mathbf{W}$ is determined, the prediction rule for our binary classifier becomes $\hat{y}_{ij} = \mathrm{sign}(\mathbf{z}_i\mathbf{W}\mathbf{z}_j^T)$. The putative pair-wise labels y^* for unknown pair-wise labels can also be generated with this prediction rule for re-ID probability learning. Thus, the joint probability of the discriminative model becomes:

$$\mathrm{P}(\mathbf{Y}|\mathbf{Z},\mathbf{W}) = \prod_{(i,j)\in\mathcal{I}} \mathrm{P}(y_{ij}|\mathbf{Z},\mathbf{W}) \prod_{(i,j)\in\mathcal{U}} \mathrm{P}(y^*_{ij}|\mathbf{Z},\mathbf{W}). \tag{7}$$

In terms of representation learning, all the samples are combined and used for training, whether or not they have a known pair-wise label. Given this learning schema handles both labeled and unlabeled pairs, our algorithm can be considered for semi-supervised re-ID tasks. Overall, our model is formulated as

$$\begin{aligned}&\mathrm{P}(\mathbf{X},\mathbf{Y},\mathbf{Z},\mathbf{A},\mathbf{W})\\ &= \mathrm{P}(\mathbf{X})\mathrm{P}(\mathbf{A})\mathrm{P}(\mathbf{Z})\mathrm{P}(\mathbf{W})\mathrm{P}(\mathbf{Y}|\mathbf{Z},\mathbf{W})\mathrm{P}(\mathbf{X}|\mathbf{Z},\mathbf{A}).\end{aligned} \tag{8}$$

The related parameters have been omitted from for simplicity. In (8), both the representation learning and the reidentification learning models shared the same prior of latent factors $\mathrm{P}(\mathbf{Z})$. A plate notation of our model is illustrated in Figure 1.

Optimization

This section outlines the algorithms for efficiently learning the proposed Bayesian model in (8). The generative model for attribute learning is considered first. Using the priors from the last section, we have the joint distribution:

$$\begin{aligned}&\mathrm{P}(\mathbf{X},\mathbf{A},\mathbf{Z}) = \mathrm{P}(\mathbf{X})\mathrm{P}(\mathbf{A})\mathrm{P}(\mathbf{Z})\mathrm{P}(\mathbf{X}|\mathbf{Z},\mathbf{A})\\ &= \frac{1}{(2\pi\sigma^2_{\mathbf{X}})^{ND/2}}\exp\{-\frac{1}{2\sigma^2_{\mathbf{X}}}\mathrm{Tr}((\mathbf{X}-\mathbf{Z}\mathbf{A})^T(\mathbf{X}-\mathbf{Z}\mathbf{A}))\}\\ &\cdot\frac{\alpha^{K_+}\exp(-\alpha H_N)}{K_+!}\prod_{k=1}^{K_+}\frac{(N-m_k)!(m_k-1)!}{N!}\\ &\cdot\frac{1}{(2\pi\sigma^2_{\mathbf{A}})^{(K+D)/2}}\exp\{-\frac{1}{2\sigma^2_{\mathbf{A}}}\mathrm{Tr}(\mathbf{A}^T\mathbf{A})\}.\end{aligned} \tag{9}$$

Following (Broderick, Kulis, and Jordan 2013), we let $\sigma_{\mathbf{X}} \to 0$ and $\alpha = \exp(-\lambda^2/2\sigma^2\mathbf{X})$. Then

$$-\log\mathrm{P}(\mathbf{X},\mathbf{A},\mathbf{Z}) \sim \|\mathbf{X}-\mathbf{A}\mathbf{Z}\|^2_F + \lambda^2 K_+, \tag{10}$$

where λ can be treated as a penalty parameter as K_+ increases. It is easy to verify that $\mathbf{A}$ has a closed formed solution when $\mathbf{Z}$ is fixed. Then, according to Bayesian theory, the posterior distribution of the uncertain remainder in (8) is

$$\mathrm{P}(\mathbf{Y},\mathbf{W}|\mathbf{Z}) = \mathrm{P}(\mathbf{Y}|\mathbf{Z},\mathbf{W})\mathrm{P}(\mathbf{W}). \tag{11}$$

According to the definition of re-identification possibility in (5) we have

$$-\log\mathrm{P}(\mathbf{Y},\mathbf{W}|\mathbf{Z}) \propto \sum_{(i,j)\in\mathcal{I}\cup\mathcal{U}}^{|\mathcal{I}|+|\mathcal{U}|} \mathrm{sign}(\mathbf{Z}_1\mathbf{W}\mathbf{Z}_2). \tag{12}$$

The remaining subproblem is to infer the probability $\mathrm{P}(\mathbf{W})$ when the other parameters are fixed. A straight forward method is to estimate a single value of $\mathbf{W}$ using $\mathrm{P}(\mathbf{W}) \propto \beta\|\mathbf{W}\|^2_F$ where β represents a leverage parameter as in previous works (Nickel et al. 2016; Feng et al. 2014). However, our framework exploits the maximum entropy discrimination (MED) method (Jaakkola, Meila, and Jebara 2000) to learn the distribution of $\mathrm{P}(\mathbf{W})$. According to the MED theory, we can learn $\mathrm{P}(\mathbf{W})$ by estimating the expectation of $\mathbf{W}$ and solving the optimization problem

Algorithm 1 Semi-supervised Bayesian Attribute Learning

```
Initialize $K_+ = 1$,$\mathbf{A} = [\sum_i \mathbf{x}_i/N]$.
while objective value in (18) deceasing do
    for $n = 1, \cdots, N$ do
        for $n = 1, \cdots, K_+$ do
            Determine $z_{ij} \in \{0,1\}$ to minimize the objective value in (18) greedily;
        end for
    end for
    $\mathbf{A} \leftarrow \mathbf{X}\mathbf{Z}^T(\mathbf{Z}\mathbf{Z}^T)^{-1}$.
    Sample a new basis $\mathbf{a}_{K_+}$ with probability
    $\mathrm{P}(\mathbf{a}_{K_+} = \mathbf{x}_i - \mathbf{A}\mathbf{z}_i) \propto \|\mathbf{x}_i - \mathbf{A}\mathbf{z}_i\|^2_2$.
    update $\mathbf{A} \leftarrow [\mathbf{A}, \mathbf{a}_{K_+}]$;
    update $K_+ \leftarrow K_+ + 1$.
    update $\Theta$ which is the expectation of $\mathbf{W}$ as in (17)
    update $\mathbf{y}^*$.
end while
```

$$\min_{\mathrm{P}(\mathbf{W})\in\mathcal{P}} \mathrm{KL}(\mathrm{P}(\mathbf{W})||\mathrm{P}_0(\mathbf{W})) + C\mathcal{E}_\ell(\mathbb{E}(W)), \tag{13}$$

where $C > 0$ is a regularization parameter that leverages the influence of the prior and the max-margin hinge loss. $\mathcal{P}$ denotes the space of distributions of $\mathrm{P}(\mathbf{W})$. KL(p||q) denotes the KullbackLeibler divergence, which is used to evaluate the distribution divergence between the distributions p and q. $\mathbb{E}(\mathbf{W})$ is the expectation of $\mathbf{W}$, and $\mathcal{E}(\cdot)$ is a loss function.

Now, we turn to the error function. As a binary model, the training error of our model would be $\mathcal{E}_{tr} = \sum_{(i,j)\in\mathcal{I}\cup\mathcal{U}} \delta(y_{ij} \neq \hat{y}_{ij})$ where $\delta(\cdot)$ is an indicator function that equals 1 if the predicate holds, and 0 otherwise. However, the non-convexity of this error function makes it difficult to deal with, so instead we have used the well-studied convex hinge loss in our model as a surrogate loss $\mathcal{E}_\ell(\mathbb{E}(W)) = \sum_{(i,j)\in\mathcal{I}\cup\mathcal{U}} h_\ell(y_{ij}f(\mathbf{x}_i,\mathbf{x}_j))$, where $f(\mathbf{z}_i,\mathbf{z}_j) = \mathbf{z}_i\mathbb{E}(\mathbf{W})\mathbf{z}_j^T$ denotes the latent discriminant function (Xu, Zhu, and Zhang 2013). After eliminating irrelevant terms, the subproblem can be written as

$$\begin{aligned}&\min_{\mathrm{P}(\mathbf{W})\in\mathcal{P}} \mathrm{KL}(\mathrm{P}(\mathbf{W})||\mathrm{P}_0(\mathbf{W})) + C\sum_{(i,j)\in\mathcal{I}} \xi_{ij}\\ &\forall(i,j)\in\mathcal{I}\cup\mathcal{U},\ \text{s.t.}\ y_{ij}(\mathrm{Tr}(\mathbb{E}(\mathbf{W})\mathbf{Z}^*_{ij}) \geq \ell - \xi_{ij},\end{aligned} \tag{14}$$

Category	Dataset	VIPeR	PRID
metric learning for re-ID	PRML(Hirzer et al. 2012)	27.0	4.8
	LMF(Zhao, Ouyang, and Wang 2014)	29.1	12.5
	KISSME(Koestinger et al. 2012)	25.4	10.2
	kLFDA(Xiong et al. 2014)	40.7	19.7
	KCCA(Lisanti, Masi, and Del Bimbo 2014)	37.2	14.5
	MLAPG(Liao and Li 2015)	40.7	16.6
Representation learning for re-ID	DLLR(Kodirov, Xiang, and Gong 2015)	38.9	25.2
	SSDAL(Su et al. 2016)	37.9	20.1
Joint learning for re-ID	LORAE(Su et al. 2015)	42.3	18.0
	LOMO+KISSME(Zhang, Xiang, and Gong 2016)	34.81	-
	LOMO+kLFDA(Zhang, Xiang, and Gong 2016)	38.58	22.40
	LOMO+XQDA(Liao et al. 2015)	40.0	26.70
	LOMO+NullSpace(Zhang, Xiang, and Gong 2016)	42.28	29.80
	SSDAL+XQDA(Su et al. 2016)	43.5	22.6
	ImprovedDeep(Ahmed, Jones, and Marks 2015)	34.81	-
	SBAL(Ours)	**45.2**	**32.4**

Table 1: Supervised re-ID result of Rank One Matching Accuracy(%) on two benchmarks. Best result of each Re-ID algorithm is marked as bold numbers.

where $\mathbf{Z}^*_{ij} = \mathbf{z}_j^T\mathbf{z}_i$ and $\{\xi_{ij}\}_{(i,j)\in\mathcal{I}\cup\mathcal{U}}$ are slack variables. According to Lagrangian duality theory, the optimal problem can be calculated by

$$\mathrm{P}(\mathbf{W}) \propto \mathrm{P}_0(\mathbf{W}) \exp\{\sum_{(i,j\in\mathcal{I}\cup\mathcal{U})} \omega_{ij} y_{ij} \mathrm{Tr}(\mathbf{W}\mathbf{Z}^*_{ij})\}, \quad (15)$$

where $\{\omega_{ij}\}_{(i,j)\in\mathcal{I}\cup\mathcal{U}}$. Let Θ be the expectation of $\mathbf{W}$, and the dual problem becomes

$$\begin{aligned} &\max_{\omega} \ell \sum_{(i,j)\in\mathcal{I}} \omega_{ij} - \frac{1}{2}(\|\Theta\|_2^2) \\ &\text{s.t.}\forall(i.j) \in \mathcal{I}\cup\mathcal{U},\ 0 \le \omega_{ij} \le C. \end{aligned} \quad (16)$$

This optimization problem can be solved by solving the equivalent primal problem

$$\begin{aligned} &\min_{\Theta} \frac{1}{2}(\|\Theta\|_2^2) + C \sum_{(i,j)\in\mathcal{I}} \xi_{ij} \\ &\forall(i,j) \in \mathcal{I}\cup\mathcal{U},\ \text{s.t. } y_{ij}(\mathrm{Tr}(\Theta\mathbf{Z}^*_{ij})) \ge \ell - \xi_{ij}. \end{aligned} \quad (17)$$

Eq. (17) can be efficiently solved as a standard binary SVM problem with a vectorized matrix $\mathbf{Z}$ and Θ (Pirsiavash, Ramanan, and Fowlkes 2009) using public SVM solvers.[1] Once the optimal expectation of $\mathbf{W}$, i.e., Θ^*, has been derived and the distribution of $\mathbf{W}$ has been certified, $\mathbf{Z}$ can be updated by greedily minimizing the following joint objective loss function:

$$\|\mathbf{X} - \mathbf{A}\mathbf{Z}\|_F^2 + \lambda^2 K_+ + \mathcal{E}_\ell(\Theta^*) + \frac{1}{2}\|\Theta\|_2^2, \quad (18)$$

where $\mathcal{E}_\ell(\mathbb{E}(W)) = \sum_{(i,j)\in\mathcal{I}\cup\mathcal{U}} h_\ell(y_{ij}(\mathbf{z}_i\Theta^*\mathbf{x}_j))$. As $K_+ \to \infty$, the algorithm alternately updates $\mathbf{A}$, Θ and $\mathbf{Z}$, along with the putative pair-wise labels y^*. The overall algorithm is provided as Algorithm. 1.

Experiments

Datasets and Settings

The following set of experiments compares the performance of various classical and state-of-the-art algorithms on two small-scale datasets and one large-scale dataset that are widely referred to in re-ID studies.

Datasets The **VIPeR** dataset (Gray, Brennan, and Tao 2007) collects 1,264 images of 632 people from two non-overlapping camera views. There are two images of each person, each captured by a different camera. Variations in viewpoint and illumination conditions are frequent in VIPeR. We randomly select 316 people as the testing set for the experiment; the ramaining people were used as the training set. The **PRID** dataset (Hirzer et al. 2011) contains images of individuals from two distinct cameras. Camera B has captured 749 persons and Camera A records 385 persons. In the dataset, 200 peoples are captured by both cameras. We selected images of 100 people taken by both cameras as the testing sets for the experiment and used the remaining images for the training sat. The **DukeMTMC-reID** dataset (Zheng, Zheng, and Yang 2017) is a subset of the DukeMTMC dataset. It collects 1,404 re-ID targets and 408 distractors. The dataset comprises 17,661 gallery images and 2,228 probe images captured by eight cameras , with 1404 individuals appearing in more than two cameras. We split the dataset equally using 702 people for the training set and 702 people for the testing set.

Evaluation Metrics and Preprocessing We used a cumulative matching characteristic (CMC) curve and mean average precision (mAP) as performance evaluation metrics. Both are widely used in the evaluation of re-ID models (Zheng, Yang, and Hauptmann 2016). In mAP evaluation, average precision is calculated for each probe, and the mAP

[1] For large-scale datasets, the numbers of their pair-wise labels are huge, we use a Stochastic Gradient SVM package *SvmSgd* : http://leon.bottou.org/projects/sgd.

Category	Dataset	VIPeR	PRID
metric learning for re-ID	RankSVM(Prosser et al. 2010)	20.7	-
	KISSME(Koestinger et al. 2012)	18.5	5.1
	kLFDA(Xiong et al. 2014)	27.5	14.1
	KCCA(Lisanti, Masi, and Del Bimbo 2014)	24.6	5.3
	MFA(Xiong et al. 2014)	25.3	-
Representation learning for re-ID	SSCDL(Liu et al. 2014)	25.6	-
	DLLR(Kodirov, Xiang, and Gong 2015)	32.5	22.1
Joint learning for re-ID	SBAL(Ours)	**33.6**	**24.4**

Table 2: Semi-supervised re-ID results in terms of rank-1 matching accuracy(%) for VIPeR and PRID datasets. The best result from each re-ID algorithm is shown in bold.

Category	Dataset	mAP(%)	CMC R1 (%)
(1)	Attributes+KISSME(Schumann and Stiefelhagen 2017)	12.83	21.97
	APR(Lin et al. 2017)	51.88	70.69
	ACRN(Schumann and Stiefelhagen 2017)	**51.96**	**72.58***
(2)	BoW+KISSME(Zheng et al. 2015)	12.17	25.13
	Basel.(Zheng, Yang, and Hauptmann 2016)	44.99	65.22
	LOMO+XQDA(Liao et al. 2015)	17.04	30.75
	SBAL(Ours)	**52.42***	**71.03**

Table 3: Attribute learning results on DukeMTMC-reID dataset. (1) Learning with predefined attributes (2) Learning with no pre-defined attributes. The best result for each category is shown in bold. The overall best results are marked with an asterisk (*)

is then calculated across all probe images. CMC calculates the probability that an image in the first rank k gallery set matches the probe image. Unlike previous works, such as (Kodirov, Xiang, and Gong 2015; Xiong et al. 2014) that rank gallery images according to their similarity with the probe image, our model ranks the gallery images according to their re-ID probability $\mathrm{P}(y = 1|\mathbf{Z}_{prob}, \mathbf{W}, \mathbf{Z}_{gallery})$. A higher probability implies the probe and the gallery image are more likely to be the same person.

In the experiments that test two-view re-ID models, we randomly selected a set of images captured by one of the cameras to form the probe set. The images captured by the other camera view(s) were used as gallery images. Following the pre-processing procedure outlined in (Lin et al. 2017), all images were first rescaled to 224×224 pixels. Then, we extracted 2048 dimensional feature vectors from the images using a pre-trained ResNet-50 deep neural network (He et al. 2016). We conducted experiments over ten splits and report the average results.

Experimental Study

In this section, we compared our algorithm in supervised/semi-supervised learning and attribute learning scenario with several other algorithms.

Supervised Person Re-ID In the experiments, we first compare our algorithms with several supervised re-ID models on the VIPeR and PRID datasets. As shown in Table 1, we compared SBAL with various metric learning re-ID, metric learning re-ID algorithms, and joint learning re-ID methods. Some representative learning methods, such as LOMO (Liao et al. 2015), were included as feature generation methods in joint learning algorithms. Overall, we observed that most of the metric and representation learning re-ID methods reported lower performance than the joint learning methods. Direct joint representation learning (e.g., LOMO) to metric learning re-ID methods, i.e., KISSME (Koestinger et al. 2012) and kLFDA (Xiong et al. 2014), enhanced the performance of metric learning re-ID methods by 10% at most. In terms of attribute learning methods, deep attribute driven re-ID (SSDAL) (Su et al. 2016) and our algorithm delivered higher performance than the others. Moreover, our method consistently reported the best performance of all the algorithms on both the VIPeR and PRID datasets and surpassed SSDAL by at most 3%.

Semi-supervised Person Re-ID In comparing our algorithms with other semi-supervised re-ID models on the VIPeR and PRID datasets, we set two-thirds of the training data as unlabeled. As in previous works (Liu et al. 2014; Kodirov, Xiang, and Gong 2015), we also introduced supervised metric learning methods RankSVM, KISSME, kLFDA, KCCA and MFA as baselines. In the experiments, they are training with only the labeled data. we introduced the semi-supervised version of DLLR (Kodirov, Xiang, and Gong 2015) as another baseline. As shown in Table 2, all semi-supervised methods demonstrated lower performance than supervised learning in Table 1. More specifically, supervised learning methods such as kLFDA, KCCA and KISSME We also observed that the representation learning re-ID methods showed better performance than the metric learning methods. The reason for this could be that metric learning methods rely on pair-wise labels. Overall, our method consistently reported the best performance on both the VIPeR and PRID datasets, which implies that our algorithm is robust even with few labeled pairs.

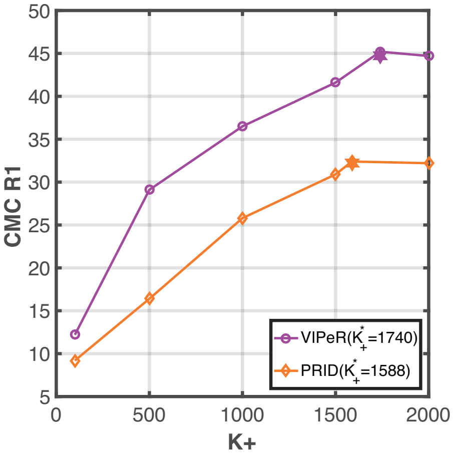

Figure 2: Influence of K_+ w.r.t. CMC Rank One accuracy. The automatically leaned attribute numbers are $K_+^* = 1740$ for VIPeR dataset and $K_+^* = 1588$ for PRID dataset(marked with asterisk symbol (*)).

Attributes Learning in Re-ID We further compared our algorithms with several state-of-the-art attribute learning re-ID methods on the large-scale dataset DukeMTMC-reID. We divide the comparison algorithms into two category, learning methods with pre-defined attributes and those without. The learning methods with pre-defined attributes included three algorithms. APR (Lin et al. 2017) utilizes manually annotated attributes from DukeMTMC-reID to enhance deep learning re-ID. ACRN (Schumann and Stiefelhagen 2017) trains an attribute classifier using separate re-ID data from PETA (Deng et al. 2014), which is then used in the training stage to learn the attributes for DukeMTMC-reID and subsequently learn the re-ID model. We also use attributes generated by ACRN as pre-defined attributes and combined them with KISSME as a baseline method, denoted as Attributes+KISSME. The learning methods without pre-defined attributes assume that no attribute information has been provided in the training stage. Following the settings in (Zheng et al. 2015), we used BoW features and KISSME (BoW+KISSME) and LOMO features and XQDA (LOMO+XQDA) as the baseline methods for joint learning. We also included a recently presented method Basel.(Zheng, Yang, and Hauptmann 2016) as a baseline.

The mAP and rank one accuracy for CMC performance is listed in Table 3. our method delivered the best performance in the comparison between attribute learning methods without pre-defined attributes. Comparing the learning methods with pre-defined attributes, our method performed 2% worse than the state-of-the-art method, ACRN, in terms of rank-1 accuracy. It implies our algorithm is very comparable as our algorithm did not require any pre-defined attributes.

The Influence of Latent Factor Dimensionality To gauge the influence of automatically learned attributes, we used the settings specified for supervised learning on the VIPeR and PRID datasets and forced our algorithm to run after researching the optimal K_+ and stopped at $K_+ = 2000$. As Figure 2 shows, performance generally increased as K_+ increased. However, at an optimal $K_+^* = 1740$ for the VIPeR dataset and an optimal $K_+^* = 1588$ for PRID dataset, performance slightly degraded on both datasets. This implies that our algorithm is able to detect representative attributes with optimal numbers and can provide reliable re-ID performance.

Conclusion

This paper proposed a novel semi-supervised Bayesian attribute learning framework, called SBAL, for person re-ID. Through this framework, representation learning and re-ID probability estimation are simultaneously optimized. The algorithm relies on semi-supervised learning to handle both labeled and unlabeled pairs of re-ID data. It is based on factor-based attribute learning and can, therefore, adaptively learn binary latent factors that do not have pre-defined dimensionality. Through extensive experiments on two small datasets, we show that our algorithm outperforms various state-of-the-art methods. Further, the results reveal comparable performance on large-scale datasets without the pre-defined attribute information required by existing methods. For future works, we suggest extending our algorithm for non-linear applications by using deep generative models.

Acknowledgments

This work was partially supported by the Data to Decisions Cooperative Research Centre (www.d2dcrc.com.au) and partially supported by the National Science Foundation under Grant No. IIS-1638429.

References

Ahmed, E.; Jones, M.; and Marks, T. K. 2015. An improved deep learning architecture for person re-identification. In *CVPR*, 3908–3916.

Broderick, T.; Kulis, B.; and Jordan, M. 2013. Mad-bayes: Map-based asymptotic derivations from bayes. In *International Conference on Machine Learning*, 226–234.

Deng, Y.; Luo, P.; Loy, C. C.; and Tang, X. 2014. Pedestrian attribute recognition at far distance. In *Proceedings of the 22nd ACM international conference on Multimedia*, 789–792. ACM.

Feng, J.; Jegelka, S.; Yan, S.; and Darrell, T. 2014. Learning scalable discriminative dictionary with sample relatedness. In *CVPR*, 1645–1652.

Ghahramani, Z., and Griffiths, T. L. 2006. Infinite latent feature models and the indian buffet process. In *Advances in neural information processing systems*, 475–482.

Gray, D.; Brennan, S.; and Tao, H. 2007. Evaluating appearance models for recognition, reacquisition, and tracking. In *Proc. IEEE International Workshop on Performance Evaluation for Tracking and Surveillance (PETS)*, volume 3.

He, K.; Zhang, X.; Ren, S.; and Sun, J. 2016. Deep residual learning for image recognition. In *CVPR*, 770–778.

Hirzer, M.; Beleznai, C.; Roth, P. M.; and Bischof, H. 2011. Person re-identification by descriptive and discriminative classification. In *Scandinavian conference on Image analysis*, 91–102. Springer.

Hirzer, M.; Roth, P. M.; Köstinger, M.; and Bischof, H. 2012. Relaxed pairwise learned metric for person re-identification. In *ECCV*, 780–793. Springer.

Jaakkola, T.; Meila, M.; and Jebara, T. 2000. Maximum entropy discrimination. In *Advances in neural information processing systems*, 470–476.

Karanam, S.; Li, Y.; and Radke, R. J. 2015. Person re-identification with discriminatively trained viewpoint invariant dictionaries. In *CVPR*, 4516–4524.

Kodirov, E.; Xiang, T.; and Gong, S. 2015. Dictionary learning with iterative laplacian regularisation for unsupervised person re-identification. In *British Machine Vision Conference (BMVC)*, volume 3, 8.

Koestinger, M.; Hirzer, M.; Wohlhart, P.; Roth, P. M.; and Bischof, H. 2012. Large scale metric learning from equivalence constraints. In *CVPR*, 2288–2295. IEEE.

Li, S.; Shao, M.; and Fu, Y. 2015. Cross-view projective dictionary learning for person re-identification. In *IJCAI*, 2155–2161.

Liao, S., and Li, S. Z. 2015. Efficient psd constrained asymmetric metric learning for person re-identification. In *CVPR*, 3685–3693.

Liao, S.; Hu, Y.; Zhu, X.; and Li, S. Z. 2015. Person re-identification by local maximal occurrence representation and metric learning. In *CVPR*, 2197–2206.

Lin, Y.; Zheng, L.; Zheng, Z.; Wu, Y.; and Yang, Y. 2017. Improving person re-identification by attribute and identity learning. *arXiv preprint arXiv:1703.07220*.

Lisanti, G.; Masi, I.; and Del Bimbo, A. 2014. Matching people across camera views using kernel canonical correlation analysis. In *Proceedings of the International Conference on Distributed Smart Cameras*, 10. ACM.

Liu, X.; Song, M.; Tao, D.; Zhou, X.; Chen, C.; and Bu, J. 2014. Semi-supervised coupled dictionary learning for person re-identification. In *CVPR*, 3550–3557.

Ma, A. J., and Li, P. 2014. Semi-supervised ranking for re-identification with few labeled image pairs. In *Asian Conference on Computer Vision*, 598–613. Springer.

Nickel, M.; Murphy, K.; Tresp, V.; and Gabrilovich, E. 2016. A review of relational machine learning for knowledge graphs. *Proceedings of the IEEE* 104(1):11–33.

Paisitkriangkrai, S.; Shen, C.; and van den Hengel, A. 2015. Learning to rank in person re-identification with metric ensembles. In *CVPR*, 1846–1855.

Pirsiavash, H.; Ramanan, D.; and Fowlkes, C. C. 2009. Bilinear classifiers for visual recognition. In *Advances in neural information processing systems*, 1482–1490.

Prosser, B. J.; Zheng, W.-S.; Gong, S.; Xiang, T.; and Mary, Q. 2010. Person re-identification by support vector ranking. In *British Machine Vision Conference (BMVC)*, volume 2, 6.

Schumann, A., and Stiefelhagen, R. 2017. Person re-identification by deep learning attribute-complementary information. In *Computer Vision and Pattern Recognition Workshops (CVPRW)*, 1435–1443. IEEE.

Su, C.; Yang, F.; Zhang, S.; Tian, Q.; Davis, L. S.; and Gao, W. 2015. Multi-task learning with low rank attribute embedding for person re-identification. In *ICCV*, 3739–3747.

Su, C.; Zhang, S.; Xing, J.; Gao, W.; and Tian, Q. 2016. Deep attributes driven multi-camera person re-identification. In *ECCV*, 475–491. Springer.

Wang, H.; Nie, F.; and Huang, H. 2014. Robust distance metric learning via simultaneous l1-norm minimization and maximization. In *International Conference on Machine Learning*, 1836–1844.

Xiao, T.; Li, H.; Ouyang, W.; and Wang, X. 2016. Learning deep feature representations with domain guided dropout for person re-identification. In *CVPR*, 1249–1258.

Xiong, F.; Gou, M.; Camps, O.; and Sznaier, M. 2014. Person re-identification using kernel-based metric learning methods. In *ECCV*, 1–16. Springer.

Xu, M.; Zhu, J.; and Zhang, B. 2012. Nonparametric max-margin matrix factorization for collaborative prediction. In *Advances in Neural Information Processing Systems*, 64–72.

Xu, M.; Zhu, J.; and Zhang, B. 2013. Fast max-margin matrix factorization with data augmentation. In *International Conference on Machine Learning*, 978–986.

Yang, L.; Jin, R.; and Sukthankar, R. 2012. Bayesian active distance metric learning. *arXiv preprint arXiv:1206.5283*.

Zhang, L.; Xiang, T.; and Gong, S. 2016. Learning a discriminative null space for person re-identification. In *CVPR*, 1239–1248.

Zhao, R.; Ouyang, W.; and Wang, X. 2014. Learning mid-level filters for person re-identification. In *CVPR*, 144–151.

Zheng, L.; Shen, L.; Tian, L.; Wang, S.; Wang, J.; and Tian, Q. 2015. Scalable person re-identification: A benchmark. In *CVPR*, 1116–1124.

Zheng, L.; Yang, Y.; and Hauptmann, A. G. 2016. Person re-identification: Past, present and future. *arXiv preprint arXiv:1610.02984*.

Zheng, Z.; Zheng, L.; and Yang, Y. 2017. Unlabeled samples generated by gan improve the person re-identification baseline in vitro. In *CVPR*.

Zhu, J.; Song, J.; and Chen, B. 2016. Max-margin nonparametric latent feature models for link prediction. *arXiv preprint arXiv:1602.07428*.