Matrix Completion for Resolving Label Ambiguity

Ching-Hui Chen, Vishal M. Patel and Rama Chellappa
Department of Electrical and Computer Engineering and the Center for Automation Research, UMIACS, University of Maryland, College Park, USA

Learning a visual classifier requires a large amount of labeled images and videos. However, labeling images is expensive and time-consuming due to the significant amount of human efforts involved. As a result, brief descriptions such as tags, captions and screenplays accompanying the images and videos become important for training classifiers. Although such information is publicly available, it is not as explicitly labeled as human annotation. For instance, names in the caption of a news photo provide possible candidates for faces appearing in the image [1]. The names in the screenplays are only weakly associated with faces in the shots [4]. The problem in which instead of a single label per instance, one is given a candidate set of labels, of which only one is correct is known as ambiguously labeled learning [2, 6].

The ambiguously labeled data is denoted as $\mathcal{L} = \{(\mathbf{x}_j, L_j),\ j = 1, 2, \ldots, N\}$, where N is the number of instances. There are c classes, and the class labels are denoted as $\mathcal{Y} = \{1, 2, \ldots, c\}$. Note that $\mathbf{x}_j$ is the feature vector of the j^{th} instance, and its ambiguous labeling set $L_j \subseteq \mathcal{Y}$ consists of the candidate labels associated with the j^{th} instance. The true label of the j^{th} instance is $l_j \in L_j$. In other words, one of the labels in L_j is the true label of $\mathbf{x}_j$. The objective is to resolve the ambiguity in $\mathcal{L}$ such that each predicted label $\hat{l}_j$ of $\mathbf{x}_j$ matches its true label l_j.

We interpret the ambiguous labeling set L_j with soft labeling vector $\mathbf{p}_j$, where $p_{i,j}$ indicates the probability that instance j belongs to class i. This allows us to quantitatively assign the likelihood of each class the instance belongs to if such information is provided. Without any prior knowledge, we assume equal probability for each candidate label. Let $\mathbf{P} \in \mathbb{R}^{c \times N}$ denotes the ambiguous labeling matrix with $\mathbf{p}_j$ in its j^{th} column. With this, one can model the ambiguous labeling as $\mathbf{P} = \mathbf{P}^0 + \mathbf{E}_P$, where $\mathbf{P}^0$ and $\mathbf{E}_P$ denote the true labeling matrix and the labeling noise, respectively. The j^{th} column vector of $\mathbf{P}^0$ is $\mathbf{p}^0_j = \mathbf{e}_{l_j}$, where $\mathbf{e}_{l_j}$ is the canonical vector corresponding to the 1-of-K coding of its true label l_j. Similarly, assuming that the feature vectors are corrupted by some noise or occlusion, the feature matrix $\mathbf{X}$ with $\mathbf{x}_j$ in its j^{th} column can be modeled as $\mathbf{X} = \mathbf{X}^0 + \mathbf{E}_X$, where $\mathbf{X} \in \mathbb{R}^{m \times N}$ consists of N feature vectors of dimension m, $\mathbf{X}^0$ represents the feature matrix in the absence of noise and $\mathbf{E}_X$ accounts for the noise.

Figure 1 shows the geometric interpretation of our proposed method, Matrix Completion for Ambiguity Resolving (MCar). When each element in the ambiguous labeling set is trivially treated as the true label, the convex hulls of each class are erroneously expanded. MCar reassigns the ambiguous labels such that each over-expanded convex hull shrinks to its actual contour, and the convex hulls becomes potentially separable.

In the paper, we show that the heterogeneous feature matrix, which is the concatenation of the labeling matrix $\mathbf{P}$ and feature matrix $\mathbf{X}$, is ideally low-rank in the absence of noise (Figure 2), which allows us to convert the aforementioned label reassignment problem as a matrix completion problem [5]. The proposed MCar takes the heterogeneous feature matrix as input, and returns the predicted labeling matrix $\mathbf{Y}$ by solving the following optimization problem

$$
\begin{aligned}
&\min_{\mathbf{Y}, \mathbf{E}_X} \operatorname{rank}(\mathbf{H}) + \lambda \|\mathbf{E}_X\|_0 + \gamma \|\mathbf{Y}\|_0 \\
&\text{s.t. } \mathbf{H} = \begin{bmatrix} \mathbf{Y} \\ \mathbf{Z} \end{bmatrix} = \begin{bmatrix} \mathbf{P} \\ \mathbf{X} \end{bmatrix} - \begin{bmatrix} \mathbf{E}_P \\ \mathbf{E}_X \end{bmatrix}, \\
&\quad \mathbf{1}_c^T \mathbf{Y} = \mathbf{1}_N^T,\ \mathbf{Y} \in \mathbb{R}_+^{c \times N}, \\
&\quad y_{i,j} = 0 \text{ if } p_{i,j} = 0,
\end{aligned} \tag{1}
$$

where $\lambda \in \mathbb{R}_+$ and $\gamma \in \mathbb{R}_+$ control the sparsity of data noise and predicted labeling matrix, respectively. Consequently, the predicted label of instance j can be obtained as

$$
\hat{l}_j = \arg\max_{i \in \mathcal{Y}} y_{i,j}. \tag{2}
$$

This is an extended abstract. The full paper is available at the Computer Vision Foundation webpage.

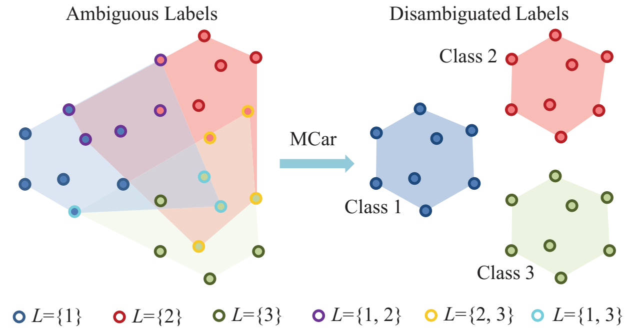

Figure 1: MCar reassigns the labels for those ambiguously labeled instances such that instances of the same subjects cohesively form potentially-separable convex hulls.

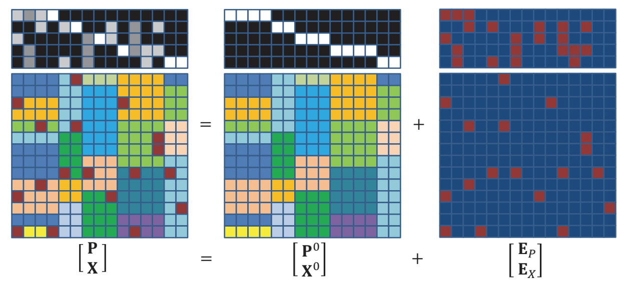

Figure 2: Ideal decomposition of heterogeneous feature matrix using MCar. The underlying low-rank structure and the ambiguous labeling are recovered simultaneously.

The proposed method inherits the benefit of low-rank recovery and possesses the capability to resolve the label ambiguity via low-rank approximation of the heterogeneous matrix. As a result, our method is more robust compared to some of the existing discriminative ambiguous learning methods [3, 7], sparsity/dictionary-based method [2], and low-rank representation-based method [8]. Moreover, we generalize MCar to include the labeling constraints between the instances for practical applications. Compared to the state of the arts, our proposed framework achieves 2.9% improvement on the labeling accuracy of the Lost dataset and performs comparably on the Labeled Yahoo! News dataset.

[1] T. L. Berg, A. C. Berg, J. Edwards, M. Maire, R. White, Y.-W. Teh, E. Learned-Miller, and D. A. Forsyth. Names and faces in the news. In *CVPR*, 2004.

[2] Y.-C. Chen, V. M. Patel, J. K. Pillai, R. Chellappa, and P. J. Phillips. Dictionary learning from ambiguously labeled data. In *CVPR*, 2013.

[3] T. Cour, B. Sapp, C. Jordan, and B. Taskar. Learning from ambiguously labeled images. In *CVPR*, 2009.

[4] M. Everingham, J. Sivic, and A. Zisserman. Hello! My name is... Buffy - Automatic naming of characters in TV video. In *BMVC*, 2006.

[5] A. B. Goldberg, X. Zhu, B. Recht, J.-M. Xu, and R. D. Nowak. Transduction with matrix completion: Three birds with one stone. In *NIPS*, 2010.

[6] E. Hüllermeier and J. Beringer. Learning from ambiguously labeled examples. In *Intell. Data Anal.*, 2006.

[7] J. Luo and F. Orabona. Learning from candidate labeling sets. In *NIPS*, 2010.

[8] Z. Zeng, S. Xiao, K. Jia, T.-H. Chan, S. Gao, D. Xu, and Y. Ma. Learning by associating ambiguously labeled images. In *CVPR*, 2013.