Deep Kinship Verification

Mengyin Wang†, Zechao Li†, Xiangbo Shu†, Jingdong Wang†, Jinhui Tang†
†School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing, China
‡Microsoft Research, Beijing, China
wangmengyin11@gmail.com, zechao.li@njust.edu.cn, shuxb104@gmail.com
jingdw@microsoft.com, tangjh1981@acm.org

Abstract—To improve the performance of kinship verification, we propose a novel deep kinship verification (DKV) model by integrating excellent deep learning architecture into metric learning. Unlike most existing shallow models based on metric learning for kinship verification, we employ a deep learning model followed by a metric learning formulation to select nonlinear information of face images in the wild, we take advantage of deep learning architecture selecting nonlinear information of face images, keeping the most important information. Specifically, the stacked auto-encoder network [24] is employed, which is one of classical deep learning models. In order to reduce dimensions effectively, the stacked auto-encoder network learns the nonlinear abstract features by limiting the number of hidden layer cells. Simultaneously, we use the metric learning to make sure the margin between negative sample pairs as large as possible and the margin between positive sample pairs as small as possible. We conduct the experiments on KFW-I and KFW-II datasets and demonstrate the effectiveness of our method.

I. INTRODUCTION

How to judge the kinship relation of people? Besides DNA detection, people can also discovery the kinship relations between parents and their children, since they are look like each other. Kinship verification is an interesting and worthy research in computer vision areas, which can be applied to discover the human social relations, find the missing child/parent, annotate images, create a family tree, etc. In recent years, many experts have gained the great progress in kinship verification [5], [6], [7], [10], [11], [14], [15], [18], [20], [21], [23], [25], [26], [27], [28], [29], [30], [31]. Generally, there are many kinship relations between people, such as parent-child, brother-brother, sister-sister, sister-brother, etc. Guo et al. [11] study four typical parent-child relations and three sibling relations respectively. Shao et al. [21] divide kinship relations into two parts, i.e., sibling relation and parent-child relation. In this paper, we only consider four parent-child relations, including father-son (FS), father-daughter (FD), mother-son (MS) and mother-daughter (MD).

However, kinship verification is still a challenge topic, especially for the unconstrained scenario, because most face images are selected from the wild varied in pose, illumination, expression, and aging. In this paper, we propose a novel deep kinship verification (DKV) model by integrating excellent deep learning architecture into metric learning. Since the variations of face images in the wild, we take advantage of deep learning architecture selecting nonlinear information of face images and keeping the most important information.
the unlabeled dataset from LFW dataset, and then employ the SVM classifier to train the mid-level features. Finally, the labeled data from kinship dataset is represented by Mid-Level features, which are extracted the more discriminative information.

For the existing methods for kinship verification, we categorized them into four classes: (1) feature learning-based [5, 6], [7], [10], [15], [23], [5], [29], [30], [31]. (2) graph-based [11], [21], (3) transfer learning-based [20], [25], [27], (4) metric learning-based [14], [18], [28]. In particular, the state-of-the-art method is based on the metric model [18]. Therefore, we focus on the metric learning-based methods for kinship verification in this paper.

B. Metric Learning

Metric learning has been widely used in many computer vision fields, such as image annotation [8], face identification [9], face verification [16] and kinship verification [14], [18], [28]. Mahalanobis distance metric is the fundamental theory of metric learning. Most methods of metric learning aim to learn Mahalanobis distance. As far as the existing metric learning method for kinship verification, these methods aim to learn interclass samples with small margin and intraclass samples with large margin by creating the shallow models. However, they only consider linear information of face images and ignore the nonlinear information for the method of kinship verification. Since deep learning is widely used for extracting nonlinear information, we consider designing the deep metric learning model.

C. Deep Learning

Deep learning has become a popular topic and achieved satisfactory performance [2]. Many deep learning algorithms have been proposed and applied in various fields, such as image classification [30], natural language processing [4], face verification [13], etc. The classical models of deep learning are stacked auto-encoder [24], deep belief network [12], convolutional neural networks [17] and restricted boltzmann machine [19]. The purpose of deep learning is to learn hierarchical abstract nonlinear features by creating the multilayer networks. Therefore, we use the stacked auto-encoder network to extract nonlinear abstract features.

III. OUR METHODS

This section briefly introduces the main idea and formula derivation of our method. Firstly, we use stacked auto-encoder network to select nonlinear low-dimension features from original image features. Then, our deep kinship verification (DKV) model combines stacked auto-encoder network and metric learning to fine-tune the deep metric network. The training phase of our method is shown as Fig. 1.

A. Stacked Auto-encoder Network

As one of classical deep learning models, stacked auto-encoder network learns the nonlinear abstract features to represent original features by the proceeding of encoding and decoding. Fig. 1 shows the structure of stacked auto-encoder network. The blue layers stand for encoder process, while the orange layers mean decoder process.

Suppose \( \mathbf{X} \in \mathbb{R}^{n \times d} \), \( \mathbf{Y} \in \mathbb{R}^{n \times d'}, \mathbf{Z} \in \mathbb{R}^{n \times d} \), where \( \mathbf{X} \) denotes the training set with \( d \) dimensions (first blue layer), \( \mathbf{Y} \) is the hidden layer presentation with \( d' \) dimensions (middle blue layer) and \( \mathbf{Z} \) is the reconstruction with \( d \) dimensions (last orange layer); \( n \) is the number of samples. \( \mathbf{Y} \) and \( \mathbf{Z} \) is defined as follow

\[
\mathbf{Y} = s(\mathbf{XW^T}), \mathbf{W} \in \mathbb{R}^{d' \times d} \tag{1}
\]

\[
\mathbf{Z} = s(\mathbf{YW'^T}), \mathbf{W'} \in \mathbb{R}^{d' \times d} \tag{2}
\]

where \( s(\cdot) \) is the activation function, and \( \mathbf{W}, \mathbf{W'} \) are the map matrix of encoder and decoder. Empirically, we set \( \mathbf{W'} = \mathbf{W}^T \).

The \( \ell_{2,1} \)-norm can avoid overfit and reduce redundancy, defined as

\[
\| \mathbf{W} \|_{2,1} = \sum_{i=1}^{d'} \left( \sum_{j=1}^{d} w_{ij}^2 \right)^{\frac{1}{2}} \tag{3}
\]

where \( w_{ij} \) denotes the \( i \)th row and \( j \)th column element of map matrix. Therefore, we use the \( \ell_{2,1} \)-norm to restrict the map matrix.

Stacked auto-encoder network aims to let \( \mathbf{X} \) and \( \mathbf{Z} \) as close as possible. We formulate the optimization problem as follow

\[
\min_{\mathbf{W}} L(\mathbf{X}, \mathbf{Z}) = \frac{1}{2} \| \mathbf{X} - \mathbf{Z} \|_F^2 + \lambda \| \mathbf{W} \|_{2,1} \tag{4}
\]

where \( \| \mathbf{A} \|_F \) is the Frobenius norm of matrix \( \mathbf{A} \) and \( \lambda \) is the parameter of \( \ell_{2,1} \)-norm. We use the back propagation algorithm to solve the optimization problem of Eq. (4), namely, to get the optimal map matrix \( \mathbf{W} \) which realizes the minimum value of Eq. (4). The gradient of the loss function \( L \) on map matrix \( \mathbf{W} \) is formulated as Eq. (5), and the update way of map matrix \( \mathbf{W} \) is shown as Eq. (6).

\[
\frac{\partial L}{\partial \mathbf{W}} = (\mathbf{Y} + \mathbf{XW^T} \odot s'( \mathbf{XW^T} ))^T ((\mathbf{Z} - \mathbf{X}) \odot s'( \mathbf{YW} )) + \lambda \mathbf{W} \tag{5}
\]

\[
\mathbf{W} = \mathbf{W} - \eta \frac{\partial L}{\partial \mathbf{W}} \tag{6}
\]

where \( s'(\cdot) \) is the derivative of activation function, \( \mathbf{D} \) is a diagonal matrix \( D_{ii} = 1/\| \mathbf{W}_i \|_2 \), \( \mathbf{W}_i \) represents the \( i \)th row vector of map matrix \( \mathbf{W} \). The symbol \( \odot \) means point-wise multiplication. The parameter \( \eta \) in Eq. (5) denotes the learning rate of the back propagation algorithm.

We train each auto-encoder network respectively and then stack them together by throwing the part of decoder (last orange layer). The hidden layer presentation of first auto-encoder network acts as the input of second auto-encoder network and so on.
B. Mahalanobis Distance Metric

The stacked auto-encoder network can learn the nonlinear abstract features by limiting the number of hidden layer cells. However, the abstract features do not make sure the margin of negative samples pairs (i.e. parent and child pairs without kinship relation) as large as possible and the margin of positive samples pairs (i.e. parent and child pairs with kinship relation) as small as possible. The purpose of metric learning is to seek mapping space by learning samples, which can satisfy the above constraints. Therefore, we propose a novel deep kinship verification (DKV) model by integrating the stacked auto-encoder network into metric learning to improve the performance of kinship verification.

Mahalanobis distance metric is the fundamental theory of metric learning. The purpose of most metric learning is to learn optimal matrix $A \in \mathbb{R}^{d \times d}$ from the training set. For the kinship verification, denote $P \in \mathbb{R}^{n \times d}$ and $C \in \mathbb{R}^{n \times d}$ as the parent and child set, respectively. The Mahalanobis distance between $p_i$ and $c_i$ is described as follow

$$ M(p_i, c_i) = \sqrt{(p_i - c_i)A(p_i - c_i)^T} $$

where $p_i$ is the $i$th sample of parent set $P$ and $c_i$ is the $i$th sample of child set $C$. Since matrix $A$ is symmetric and positive semi-definite, we can decompose $A = GG^T$, where $d' \leq d$, $G \in \mathbb{R}^{d \times d'}$. Therefore, we rewrite the Mahalanobis distance of Eq. (7) as

$$ M(p_i, c_i) = \sqrt{(p_i - c_i)GG^T(p_i - c_i)^T} $$

$$ = \sqrt{(p_iG - c_iG)(p_iG - c_iG)^T} $$

$$ = \|p_iG - c_iG\|_2 $$

(8)

Mahalanobis distance metric projects the parent and child samples onto the low-dimensional subspace by the linear transformation. However, the distribution of face images is not linear but nonlinear. Since the Mahalanobis distance metric lack the nonlinear information, we utilize the deep learning model which can achieve nonlinear abstract features to address this limitation.

C. Deep Metric Learning

Our method DKV model fuses the advantage of deep learning and metric learning together. Specifically, we employ the stacked auto-encoder network to find appropriate project space and utilize metric learning to ensure the margin of negative kinship sample pairs as large as possible and the margin of positive kinship sample pairs as small as possible.

As similar as section III-A, there are $L$ auto-encoder networks and $W^{(l)} \in \mathbb{R}^{d(l) \times d(l-1)}$ is the $l$th map matrix. Given the input parent and child features, we construct the stacked auto-encoder network with $L + 1$ layers. The parent and child samples of first hidden layer can be presented as

$$ a_p^{(1)} = s \left( PW^{(1)} \right) \in \mathbb{R}^{n \times d(1)} $$

$$ a_c^{(1)} = s \left( CW^{(1)} \right) \in \mathbb{R}^{n \times d(1)} $$

where $W^{(1)} \in \mathbb{R}^{d(1) \times d}$, $d(l)$ denotes the size of dimension of the $l$th hidden layer cells and $n$ is the number of input samples. Likewise, the presentation of $l$th hidden layer is shown as

$$ a_p^{(l)} = s \left( a_p^{(l-1)} W^{(l)} \right) \in \mathbb{R}^{n \times d(l)} $$

$$ a_c^{(l)} = s \left( a_c^{(l-1)} W^{(l)} \right) \in \mathbb{R}^{n \times d(l)} $$

(11)

Therefore, the Mahalanobis distance of parent and child samples can be computed by the Euclidean distance between the $l$th hidden layers output of parent and child samples. Eq. (8) can be also rewritten as

$$ M(p_i, c_i) = \|a_p^{(L)} - a_c^{(L)}\|_2 $$

where $a_p^{(l)}$ and $a_c^{(l)}$ denote the $l$th hidden layer output of $p_i$ and $c_i$, respectively. The major reason of misclassification is possibly that the margin between parent (child) samples and neighbourhood of corresponding positive child (parent) may be larger than the margin of positive sample pairs as shown in Fig. 2. To obtain more discriminative information of kinship
verification, we prompt the margin between parent (child) samples and k-nearest neighbours of corresponding positive child (parent) samples as large as possible on the training set. For our deep kinship verification model, we fine-tune the stacked auto-encoder network to purify the more discriminative information when learning the high-level features, where the corresponding objective function is formally formulated as follow

\[
\min \mathbf{W} = \frac{1}{2} \sum_{i=1}^{n} \sum_{c_i \in T} \mathbf{M}^2 (p_i, c_i) - \frac{\alpha}{2} \sum_{i=1}^{n} \sum_{c_{it} \in T} \mathbf{M}^2 (p_i, c_{it}) - \frac{\beta}{2} \sum_{i=1}^{n} \sum_{c_{it} \in T} \mathbf{M}^2 (p_{ir}, c_i) + \lambda \sum_{l=1}^{L} \| \mathbf{W}^{(l)} \|_{2,1}
\]

(14)

where \( c_{it} \) is the \( t \)th nearest neighbour of \( c_i \), \( p_{ir} \) is the \( r \)th nearest neighbour of \( p_i \) and parameters \( \alpha, \beta, \lambda \) are the weighting coefficients to control the corresponding term. To solve the above optimization problem, we use the back propagation algorithm. The gradient of the objective function \( J \) w.r.t. \( \mathbf{W}^{(l)} \) is derived as follow

\[
\frac{\partial J}{\partial \mathbf{W}^{(l)}} = \frac{1}{n} \sum_{i=1}^{n} \left( \delta^{(l)}_{p_i, c_i} \mathbf{T} \mathbf{a}^{(l-1)}_{p_i} - \delta^{(l)}_{c_i, p_i} \mathbf{T} \mathbf{a}^{(l-1)}_{c_i} \right) - \alpha \sum_{i=1}^{n} \sum_{c_{it} \in T} \left( \delta^{(l)}_{p_i, c_{it}} \mathbf{T} \mathbf{a}^{(l-1)}_{p_i} - \delta^{(l)}_{c_{it}, p_i} \mathbf{T} \mathbf{a}^{(l-1)}_{c_{it}} \right) - \beta \sum_{i=1}^{n} \sum_{c_{it} \in T} \left( \delta^{(l)}_{p_{ir}, c_i} \mathbf{T} \mathbf{a}^{(l-1)}_{p_{ir}} - \delta^{(l)}_{c_i, p_{ir}} \mathbf{T} \mathbf{a}^{(l-1)}_{c_i} \right) + \lambda \mathbf{D} \mathbf{W}^{(l)}
\]

(15)

where \( D_{ii} = 1/\| \mathbf{W}^{(l)} \|_2, i = 1, 2, \ldots, L, k \) is the number of nearest neighbours,

\[
\delta^{(l)}_{s_1, s_2} = \left( \delta^{(l+1)}_{s_1, s_2} \mathbf{W}^{(l+1)} \right) \odot s' \left( \mathbf{z}^{(L)}_{s_1} \right)
\]

(16)

\[
\delta^{(L)}_{s_1, s_2} = \left( \mathbf{a}^{(L)}_{s_1} - \mathbf{a}^{(L)}_{s_2} \right) \odot s' \left( \mathbf{z}^{(L)}_{s_1} \right)
\]

(17)

\[
\mathbf{z}^{(l)}_{s_1} = \mathbf{a}^{(l-1)}_{s_1} \mathbf{W}^{(l)} \mathbf{T}
\]

(18)

and \( s_1, s_2 \) can be replaced by \( p_i, c_i, p_{ir}, c_{it} \). We suppose \( \mathbf{a}^{(0)}_{c} = \mathbf{C} \) and \( \mathbf{a}^{(0)}_{p} = \mathbf{P} \). The update way of map matrix \( \mathbf{W}^{(l)} \) is similar as Eq. (6).

IV. EXPERIMENTS

A. Data Sets

In our experiments, we use the data sets KFW-I and KFW-II released by Lu et al. [18], the largest datasets on kinship relations. The data sets KFW-I and KFW-II are collected from internet, including many public figures and their parent or child. There are four kinship relations in data sets: father-son (FS), father-daughter (FD), mother-son (MS) and mother-daughter (MD). In the KFW-I dataset, the pairs of kinship relation images are collected from different pictures and the number of four kinship-relationship pairs: father-son, father-daughter, mother-son and mother-daughter are 134, 156, 127 and 116 respectively. In the KFW-II dataset, the pairs of kinship relation images come from same picture and each kinship relation contains 250 pairs.

B. Feature and Classifier

Local binary patterns (LBP) [1] operator describes the local texture feature, which is extensively applied to face recognition, face verification and so on. The pixel size of images in the KFW-I and KFW-II dataset are all 64×64. We segment images into 4×4 non-overlapped blocks and extract classical 256 dimensions LBP feature for each block. Then, we obtain 4096 dimensions LBP feature.

Support Vector Machine (SVM) [22] has been widely used in pattern recognition, classification and regression, which is originally designed for binary classification problem and shows a great effect on it. Whether parent and child sample pairs have kinship relations or not is a binary classification problem. Hence, we choose SVM as the classifier in our model.

C. Experiment settings

In the experiments, we set the parameters \( \ell_{2,1} \)-norm coefficient \( \lambda \) and learning rate \( \eta \) as \( 10^{-2} \) and \( 10^{-3} \) respectively. For the k-nearest neighbours, we set \( k \) as 5 and parameters \( \alpha, \beta \) equal the inverse of \( k \). Our deep kinship verification model train three layers network, i.e., \( L=2 \). Activation function of network transfers the input of network layers to another space by the nonlinear transformation. In our experiments, we choose \( \tanh \) as the activation function. The \( \tanh \) function is defined as

\[
\tanh(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}}
\]

(19)

and the derivative of \( \tanh \) is

\[
\tanh'(z) = 1 - \tanh^2(z)
\]

(20)

For a fair comparison, selecting the positive sample pairs and negative sample pairs on the two datasets follows the strategy introduced in the paper [18]. We also adopt fivefold cross validation in the experiments. The datasets are divide into five folds. Four folds positive samples pairs act as the training set of our DKV model. The four folds positive sample pairs combined with the same number negative pairs are the training...
In this paper, we propose a novel deep kinship verification (DKV) model. In this model, we integrate the excellent deep learning architecture into metric learning, which can further improve the performance of kinship verification. Overall, the experimental results show that the proposed method achieves the competitive performance on both of KFW-I and KFW-II datasets generally. Our further work will be considered pushing the proposed deep kinship verification into the social-aware multimedia applications.

V. CONCLUSION

In this paper, we propose a novel deep kinship verification (DKV) model. In this model, we integrate the excellent deep learning architecture into metric learning, which can further improve the performance of kinship verification. Overall, the experimental results show that the proposed method achieves the competitive performance on both of KFW-I and KFW-II datasets generally. Our further work will be considered pushing the proposed deep kinship verification into the social-aware multimedia applications.

ACKNOWLEDGMENT

This work was partially supported by the 863 Program (Project No. 2014AA015101), the National Natural Science Foundation of China (Grant No. 6152203 and 61402228) and the Open Projects Program of National Laboratory of Pattern Recognition.

REFERENCES


<table>
<thead>
<tr>
<th>Datasets</th>
<th>Methods</th>
<th>FS</th>
<th>FD</th>
<th>MS</th>
<th>MD</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>KFW-I</td>
<td>NRML</td>
<td>64.7</td>
<td>65.2</td>
<td>59.4</td>
<td>65.4</td>
<td>63.7</td>
</tr>
<tr>
<td></td>
<td>DKV</td>
<td>71.3</td>
<td>62.7</td>
<td>66.4</td>
<td>66.6</td>
<td>66.9</td>
</tr>
<tr>
<td>KFW-II</td>
<td>NRML</td>
<td>69.0</td>
<td>69.5</td>
<td>69.8</td>
<td>69.0</td>
<td>69.5</td>
</tr>
<tr>
<td></td>
<td>DKV</td>
<td>73.4</td>
<td>68.2</td>
<td>71.0</td>
<td>72.8</td>
<td>71.3</td>
</tr>
</tbody>
</table>

TABLE I

ACCURACY (%) ON TWO DATASETS


