

基于PCA和支持向量机的人脸识别

罗 强

长安大学, 陕西 西安

收稿日期: 2021年11月16日; 录用日期: 2021年12月30日; 发布日期: 2022年1月14日

摘 要

随着大数据和人工智能时代的来临, 人脸识别已经成为一个非常重要的研究领域。在这项工作中, 输入对象是一张张人脸的照片, 我们可以将每张照片视为一个多维向量, 该向量的分量代表图片某处的像素值, 那么我们要处理的将是维数高达数万乃至数十万维的向量, 这意味着非常巨大的运算量, 即使如今计算机性能十分强大, 然而我们要处理的是一般是规模比较大的人脸数据库, 如果我们对采集的人脸数据不加处理, 必将导致时间和资源的浪费。本文基于主成分分析(Principal Component Analysis, PCA)和支持向量机(Support Vector Machine, SVM)实现人脸识别。PCA算法能够将高维数据简化成低维问题, 简单、快速, 且主成分之间相互正交, 可消除原始数据成分间的影响, 基于PCA算法的人脸识别技术能够在一定程度上去除光照、姿态、遮挡产生的噪音。使用核函数的SVM方法对于非线性问题具备良好的分类效果。该算法结合PCA和SVM方法, 基于ORL人脸数据库先对训练数据进行PCA降维及特征提取, 然后将降维后的数据输入到使用高斯核函数的SVM中进行训练, 最后将测试数据输入训练好的模型中测试模型的准确率。实验证明该算法具有较高的识别率。

关键词

人脸识别, 主成分分析, 支持向量机, SVD奇异值分解

Face Recognition Based on PCA and Support Vector Machine

Qiang Luo

Chang'an University, Xi'an Shaanxi

Received: Nov. 16th, 2021; accepted: Dec. 30th, 2021; published: Jan. 14th, 2022

Abstract

With the advent of the era of big data and artificial intelligence, face recognition has become a very

important research field. In this work, the input object is a picture of a face. We can treat each picture as a multi-dimensional vector. The components of this vector represent the pixel value somewhere in the picture, so what we have to deal with will be the dimensionality, up to tens of thousands or even hundreds of thousands of dimensional vectors, which means a large number of calculations. Even though the computer performance is very powerful now, what we have to deal with is generally a relatively large-scale face database. If we analyze the face dataset untreated, it will inevitably lead to a waste of time and resources. This paper implements face recognition based on Principal Component Analysis (PCA) and Support Vector Machine (SVM). The PCA algorithm can simplify high-dimensional data into low-dimensional problems. It is simple, fast, and the principal components are orthogonal with each other, which eliminates the influence of the original data among variable components. The face recognition technology based on the PCA algorithm can remove noises caused by lighting, posture, and cover. The SVM method using the kernel function has a good classification effect for nonlinear problems. The algorithm combines PCA and SVM methods, based on the ORL face database, first performs PCA dimensionality reduction and feature extraction on the training data, and then inputs the reduction data into the SVM using the Gaussian kernel function for training, and finally inputs the test data into the trained model to test the accuracy of the model. Experiments indicate that the algorithm has a high recognition rate.

Keywords

Face Recognition, Principal Component Analysis, Support Vector Machine, Singular Value Decomposition

Copyright © 2022 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

PCA [1]是一种经典的多元统计方法,是由 Hotelling 于 1933 年首先提出的,多元统计分析处理的是多变量(多指标)问题,由于变量较多,增加了分析问题的复杂性。但在实际问题中,变量之间可能存在一定的相关性,因此,多变量中可能存在信息的重叠。人们自然希望通过克服相关性、重叠性,用较少的变量来代替原来较多的变量,而这种代替可以反映原来多个变量的大部分信息,这实际上是一种“降维”的思想。

在 20 世纪 90 年代 Vapnik 等提出了一种新的机器学习方法——支持向量机[2] (Support Vector Machine, SVM)。支持向量机(SVM)是一种基于统计学习理论的采用结构风险最小化原则的机器学习方法,它通过构建最优分类超平面,将未知样本的分类误差降低到最小,并表现出很高的泛化能力。相比于传统的机器学习方法,它能够很好地克服非线性、小样本、过拟合等问题。SVM 被广泛应用于机器学习中的分类和回归问题如模式识别[3]、信号处理[4]、函数拟合、预测分析[5]、故障分类等领域。SVM 还被应用于计量生物学[6],生物信息学[7]和化学信息学[8]等学科中。

随着人工智能技术的发展,人脸识别技术也得到了前所未有的关注。目前,人脸识别技术已经广泛应用于安防、金融、数码相机、门禁系统、身份识别、网络应用等场景中[9]。人脸识别技术包括基于几何特征的方法(Geometrical Features Based, GFB)、局部特征分析方法(Local Face Analysis, LFA)、基于动态链接模型的方法(Dynamic Link Architecture, DLA)、卷积神经网络算法(Convolutional Neural Networks, CNN)和主成分分析(Principal Component Analysis, PCA)算法。

本文基于 ORL 人脸数据集，将数据集分为训练集和测试集，首先使用 PCA 对训练数据降维，将降维后的数据进行归一化处理，输入使用高斯核函数(RBF)的 SVM 训练，在 PCA 降维的过程中，我们可以得到用于降维的若干个特征向量，我们将这些特征向量称为特征脸，使用这些特征脸将测试数据映射到低维空间，将得到的数据归一化处理，输入到训练好的模型进行预测，从而实现人脸识别功能。

2. 主成分分析

PCA 本质是找出一组投影空间，使得在投影空间的数据能最大限度地代表原始数据。PCA 法能够降低数据的维数，减少图像中冗余的信息和噪声并且使降维后的数据不能失真。其中，冗余信息指数据中与其他数据线性相关的数据，这些数据被其他数据线性表示；噪声代表矩阵较小的特征根对应的特征向量，常见的噪声包括光照、阴影等。而通常的方法是通过保留数据的低阶主成分，去除高阶主成分，从而能够保留数据中的核心部分去除冗余和噪声。

2.1. PCA 的算法原理

主成分分析的数学模型是，设 p 个变量构成的 p 维随机向量为 $X = (X_1, X_2, \dots, X_p)$ ，假定其存在二阶矩，设 X 的均值向量和协方差阵分别为 $\mu = E(X), \Sigma = D(X)$ 。对 X 作正交变换，令 $Y = T'X$ ，其中 T 为正交阵，要求 Y 的各分量是不相关的，并且 Y 的第一个分量的方差是最大的，第二个分量的方差次之，……，等等。为了保持信息不丢失， Y 的各分量方差和与 X 的各分量方差和相等。

为了防止大数吃小数的问题，我们首先将输入数据作去中心化处理，在本文中，假设数据集的每张人脸图可以表示为 $S_i = (Z_1, Z_2, \dots, Z_p)^T, i = 1, 2, \dots, m$ ，均值为 $\Psi = \frac{1}{m} \sum_{i=1}^m S_i$ ，则去中心化后的人脸为 $\Phi_i = S_i - \Psi, i = 1, 2, \dots, m$ ，图 1 展示了这一过程。将去中心化的人脸作为 PCA 的输入，即上文中的 X 是去中心化后的结果。

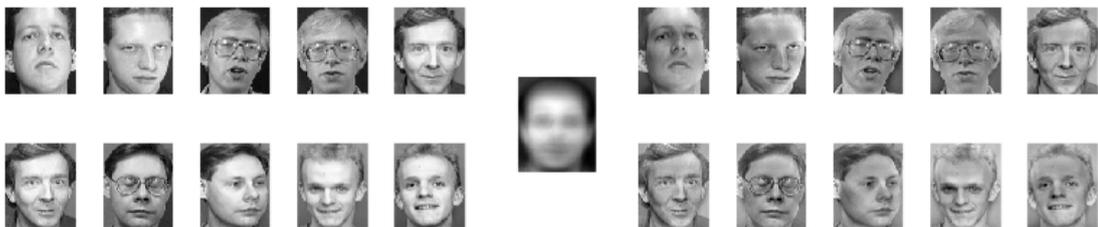


Figure 1. Face decentralization
图 1. 人脸去中心化

考虑如下的线性变换：

$$\begin{cases} Y_1 = t_{11}X_1 + t_{12}X_2 + \dots + t_{1p}X_p = T_1'X \\ Y_2 = t_{21}X_1 + t_{22}X_2 + \dots + t_{2p}X_p = T_2'X \\ \vdots \\ Y_p = t_{p1}X_1 + t_{p2}X_2 + \dots + t_{pp}X_p = T_p'X \end{cases}$$

用矩阵表示为 $Y = T'X$ ，其中 $Y = (Y_1, Y_2, \dots, Y_p)'$ ， $T = (T_1, T_2, \dots, T_p)$ 。

我们希望寻找一组新的变量 $Y_1, \dots, Y_m (m \leq p)$ ，这组新的变量要求充分地反映原变量 X_1, \dots, X_p 的信息，而且相互独立。这里我们应该注意到，对于 Y_1, \dots, Y_m 有

$$D(Y_i) = D(T_i'X) = T_i'D(X)T_i = T_i'\Sigma T_i, \quad i=1,2,\dots,m$$

$$\text{Cov}(Y_i, Y_k) = \text{Cov}(T_i'X, T_k'X) = T_i'\text{Cov}(X, X)T_k = T_i'\Sigma T_k, \quad i, k=1,2,\dots,m$$

这样, 我们所要解决的问题就转化为, 在新的变量 Y_1, \dots, Y_m 相互独立的条件下, 求 T_i 使得 $D(Y_i) = T_i'\Sigma T_i$, $i=1,2,\dots,m$, 达到最大。

首先应该注意到, 使得 $D(Y_i)$ 达到最大的线性组合, 显然用常数乘以 T_i 后, $D(Y_i)$ 也随之增大, 为了消除这种不确定性, 不妨假设 T_i 满足 $T_i'T_i = 1$ 或者 $|T_i| = 1$ 。那么, 问题可以更加明确。

第一主成分为, 满足 $T_1'T_1 = 1$, 使得 $D(Y_1) = T_1'\Sigma T_1$ 达到最大的 $Y_1 = T_1'X$ 。

第二主成分为, 满足 $T_2'T_2 = 1$, 且 $\text{Cov}(Y_2, Y_1) = \text{Cov}(T_2'X, T_1'X) = 0$, 使得 $D(Y_2) = T_2'\Sigma T_2$ 达到最大的 $Y_2 = T_2'X$ 。

一般情形, 第 k 主成分为, 满足 $T_k'T_k = 1$, 且 $\text{Cov}(Y_k, Y_i) = \text{Cov}(T_k'X, T_i'X) = 0 (i < k)$, 使得 $D(Y_k) = T_k'\Sigma T_k$ 达到最大的 $Y_k = T_k'X$ 。

第一主成分的目标函数为

$$\varphi_1(T_1, \lambda) = T_1'\Sigma T_1 - \lambda(T_1'T_1 - 1)$$

对 T_1 求导, 可得 $(\Sigma - \lambda I)T_1 = 0$, $T_1'\Sigma T_1 = \lambda$ 。

由于 X 的协方差阵 Σ 为非负定的, 其特征根均大于零, 不妨设 $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$, 那么, Y_1 的最大方差值为 λ_1 , 其相应的单位化特征向量为 T_1 。

针对一般情形, 第 k 主成分应该是在 $T_k'T_k = 1$ 且 $T_k'T_i = 0$ 或 $T_i'T_k = 0 (i < k)$ 的条件下, 使得 $D(Y_k) = T_k'\Sigma T_k$ 达到最大的 $Y_k = T_k'X$ 。这样我们构造目标函数为

$$\varphi_k(T_k, \lambda, \rho_i) = T_k'\Sigma T_k - \lambda(T_k'T_k - 1) - 2\sum_{i=1}^{k-1} \rho_i(T_i'T_k)$$

对目标函数 $\varphi_k(T_k, \lambda, \rho_i)$ 求导有:

$$\frac{\partial \varphi_k}{\partial T_k} = 2\Sigma T_k - 2\lambda T_k - 2\sum_{i=1}^{k-1} \rho_i T_i = 0$$

$$\text{用 } T_i' \text{ 左乘上式有 } T_i'\Sigma T_k - \lambda T_i'T_k - T_i'\left(\sum_{i=1}^{k-1} \rho_i T_i\right) = 0$$

即有 $\rho_i T_i'T_k = 0$, 那么, $\rho_i = 0 (i=1,2,\dots,k-1)$ 。从而

$$(\Sigma - \lambda I)T_k = 0, \quad T_k'\Sigma T_k = \lambda$$

因此 Y_k 的最大方差值为第 k 大特征根 λ_k , 其相应的单位化的特征向量为 T_k 。

综上所述, 设 $X = (X_1, \dots, X_p)'$ 的协方差阵为 Σ , 其特征根为 $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$, 相应的单位化的特征向量为 T_1, T_2, \dots, T_p 。那么, 由此所确定的主成分为 $Y_1 = T_1'X$, $Y_2 = T_2'X$, \dots , $Y_m = T_m'X$, 其方差分别为 Σ 的特征根。

主成分分析把 p 个原始变量 X_1, X_2, \dots, X_p 的总方差 $\text{tr}(\Sigma)$ 分解成了 p 个相互独立的变量 Y_1, Y_2, \dots, Y_p 的方差之和 $\sum_{k=1}^p \lambda_k$ 。主成分分析的目的是减少变量的个数, 所以一般不会使用所有 p 个主成分的, 忽略一些

带有较小方差的主成分将不会给总方差带来太大的影响。这里我们称 $\varphi_k = \lambda_k / \sum_{k=1}^p \lambda_k$ 为第 k 个主成分 Y_k 的贡献率。第一主成分的贡献率最大, 这表明 $Y_1 = T_1'X$ 综合原始变量 X_1, X_2, \dots, X_p 的能力最强, 而 Y_2, Y_3, \dots, Y_p 的综合能力依次递减。若只取 $m (< p)$ 个主成分, 则称

$$\psi_m = \frac{\sum_{k=1}^m \lambda_k}{\sum_{k=1}^p \lambda_k}$$

为主成分 Y_1, \dots, Y_m 的累计贡献率, 累计贡献率表明 Y_1, \dots, Y_m 综合 X_1, X_2, \dots, X_p 的能力。通常取 m , 使得累计贡献率达到一个较高的百分数(如 85% 以上)。

2.2. 奇异值分解

在 2.1 中, 我们通过计算样本(去中心化后)的协方差矩阵 Σ 的特征值和特征向量来求解主成分, 然而, 在人脸识别的研究中, $X = (X_1, X_2, \dots, X_p)$ 的维数 p 很大, 导致协方差矩阵 Σ 的维数也很大, 那么在求特征值和特征向量时会消耗大量的时间。为此, 我们通过奇异值分解的方法来简化这一过程。

假设有 m 张人脸图片, 我们将每张照片拉直成一个列向量, 并且去中心化后得到 $\Phi_1, \Phi_2, \dots, \Phi_m$, $\Phi_i, i=1, \dots, m$ 是 p 维的($P \gg m$)。现在, 我们要求解

$$\Sigma = \frac{1}{m} \Phi \cdot \Phi^T, \text{ 其中 } \Phi = (\Phi_1, \Phi_2, \dots, \Phi_m)$$

的特征值和特征向量, 等价于求 $\Phi \cdot \Phi^T$ 的特征值和特征向量。

根据高等代数中奇异值分解的理论, $\Phi \cdot \Phi^T$ 的正特征值和 $\Phi^T \cdot \Phi$ 的正特征值相同。设 $\Phi^T \cdot \Phi$ 的正特征值为 $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_k > 0$, $\lambda_i, i=1, \dots, k$ 对应的特征向量为 $V_i, i=1, \dots, k$, 那么 $\Phi \cdot \Phi^T$ 的特征值 $\lambda_i, i=1, \dots, k$ 对应的特征向量为

$$U_i = \frac{1}{\sqrt{\lambda_i}} \cdot \Phi \cdot V_i, i=1, \dots, k$$

这样, 我们通过求解一个阶数较小的矩阵 $\Phi^T \cdot \Phi$ 的特征值和特征向量, 得到了原本阶数很大的矩阵 $\Phi \cdot \Phi^T$ 的特征值和特征向量, 这一求解方法其实是在奇异值分解过程中产生的推论。

我们将上述问题中得到的 $U_i, i=1, \dots, k$ 称为特征脸(如图 2 所示), 在本文中, 我们取前 20 个主成分, 即 $k=20$, 此时, 每张人脸看成一个 20 维的向量, 与未处理前每张人脸数万维的规模相比, 大大降低了数据的复杂度, 并且节省存储空间, 同时, 极大的降低了支持向量机的运算量。

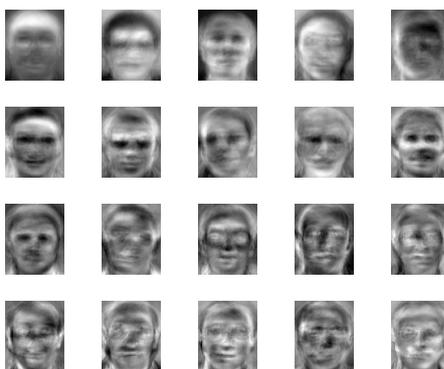


Figure 2. Grayscale image of eigenface
图 2. 特征脸灰度图像

3. 支持向量机

支持向量机(SVM)是在统计学习理论和结构风险最小原理基础上建立的, 根据有限的样本信息在模型的复杂性(即对特定训练样本的学习精度)和学习能力(即无错误地识别任意样本的能力)之间寻求平衡,

来获得最好的推广能力。SVM 是应用最广泛的分类器。

SVM 本身只能解决二分类问题，要将 SVM 应用到人脸识别实现多分类的功能，原始的 SVM 已经不能满足我们的要求，于是我们引入一对一(OVO)的 SVM。

所谓一对一(OVO)的 SVM，就是在任意两类样本之间设计一个 SVM，因此 N 个类别的样本就需要设计 $N(N-1)$ 个 SVM。当对一个未知样本进行分类时，得票最多的类别即为该未知样本的类别，如图 3 所示。

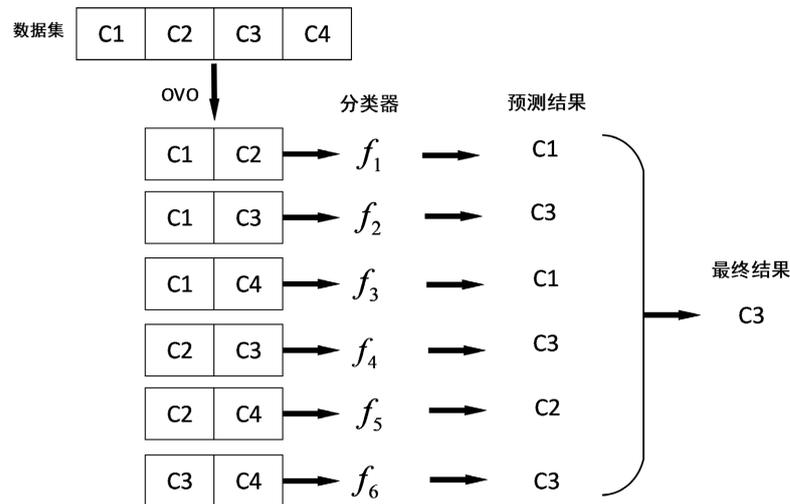


Figure 3. Schematic diagram of OVO
图 3. OVO 示意图

4. 实验

本文选择 ORL 人脸数据库作为实验数据，该数据库是英国剑桥的 Olivetti 研究实验室 1992 年 4 月至 1994 年 4 月期间创建，包含 40 个人的人脸照片，每人 10 张，共 400 张，像素均为 112×92 。我们将这 400 张照片分为训练集和测试集，取每个人的前八张照片，共 320 张作为训练集，取每人的后两张，共 80 张照片作为测试集。

本文的实验是在 Windows 10 环境下，使用 MATLAB 2020a 完成的。因为 MATLAB 自带的 svm 实现函数是 svmtrain 和 svmclassify 函数，且仅支持二分类问题，所以在训练 OVO 模型时，用到了 libsvm 工具包，libsvm 是台湾大学林智仁(Lin Chih-Jen)教授等开发设计的 MATLAB 工具包。实验前需要下载 libsvm 工具包安装到 MATLAB 中，并使用 Visual C++ 进行编译，完成了这两项工作之后就可以在 MATLAB 中调用 libsvm 中的 OVO 分类函数。

4.1. PCA 降维

将数据集的每张照片拉伸为一个 112×92 维的列向量。首先求训练集中 320 个列向量的均值，然后对每个训练样本去中心化后得到 $\Phi = (\Phi_1, \Phi_2, \dots, \Phi_{320})$ ，计算去中心化后样本的协方差矩阵 $\Sigma = \frac{1}{m} \Phi \cdot \Phi^T$ ，通过奇异值分解(SVD)求解 $\Phi \cdot \Phi^T$ 的特征值和特征向量，在本实验中，我们取前 20 个主成分，因此，经过 PCA 后，训练集中的每个人脸向量的维数从 112×92 维降到了 20 维，再将降维后的数据进行归一化处理，表 1 给出第一个人的前五张照片的一部分数据。

Table 1. Results of data dimensionality reduction

表 1. 数据降维的结果

s1	0.656968	0.695698	0.967977	0.478459	0.602719	...
s2	0.877324	0.670609	0.470037	0.264878	1	...
s3	0.833119	0.621005	0.777907	0.518892	0.664302	...
s4	0.899552	0.717222	0.534585	0	0.487597	...
s5	0.963199	0.680347	0.600285	0.10652	0.523953	...

PCA 降维后，我们还可以得到 $\Phi \cdot \Phi^T$ 的二十个最大特征值对应的特征向量，我们将之称为特征脸，如图 3 所示。最后，对测试集中数据去中心化，应用特征脸将去中心化的测试集映射到低维空间。

至此，我们应用 PCA 算法完成了人脸识别前非常重要的准备工作。

4.2. OVO 人脸识别

通常人脸数据都是非线性可分的，选取使用高斯核函数(RBF)的 SVM 对不同的人脸分类。高斯核函数的表达式为：

$$K(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right)$$

称参数 σ 为核参数， σ 值的选取会影响 SVM 的分类效果，此外，影响分类结果的还有(10)中的惩罚参数 C 。本文的参数组合 (C, σ) 时依靠经验选取的，我们取 $(C, \sigma) = (100, 0.08)$ 。

将 $(C, \sigma) = (100, 0.08)$ 作为 OVO 模型的参数，降维后的训练数据、测试数据以及训练集标签、测试集标签作为 OVO 模型的输入，进行人脸识别。40 个人的标签分别为 1~40，用测试集标签和模型的预测结果比较，计算出 OVO 模型的准确率，预测结果如表 2 所示。

Table 2. Comparison of predicted labels and true labels

表 2. 预测标签和真实标签对比

预测	1	1	2	2	3	3	4	4	5	5
真实	1	1	2	2	3	3	4	4	5	5
预测	6	6	7	7	8	8	9	9	10	8
真实	6	6	7	7	8	8	9	9	10	10
预测	11	11	12	12	13	13	14	14	15	15
真实	11	11	12	12	13	13	14	14	15	15
预测	16	16	17	17	18	18	11	19	20	20
真实	16	16	17	17	18	18	19	19	20	20
预测	21	21	22	22	23	23	24	24	25	25
真实	21	21	22	22	23	23	24	24	25	25
预测	26	26	27	27	37	28	29	29	30	30
真实	26	26	27	27	28	28	29	29	30	30
预测	31	31	32	32	33	33	34	34	35	35

Continued

真实	31	31	32	32	33	33	34	34	35	35
预测	36	36	37	37	38	38	39	39	40	40
真实	36	36	37	37	38	38	39	39	40	40

模型的预测准确率为 96.25%，从表 2 可以看出第 10 人的第 10 张，第 19 人的第 9 张和第 28 人的第 9 张预测错误，在图 4 中标出这几人的照片。

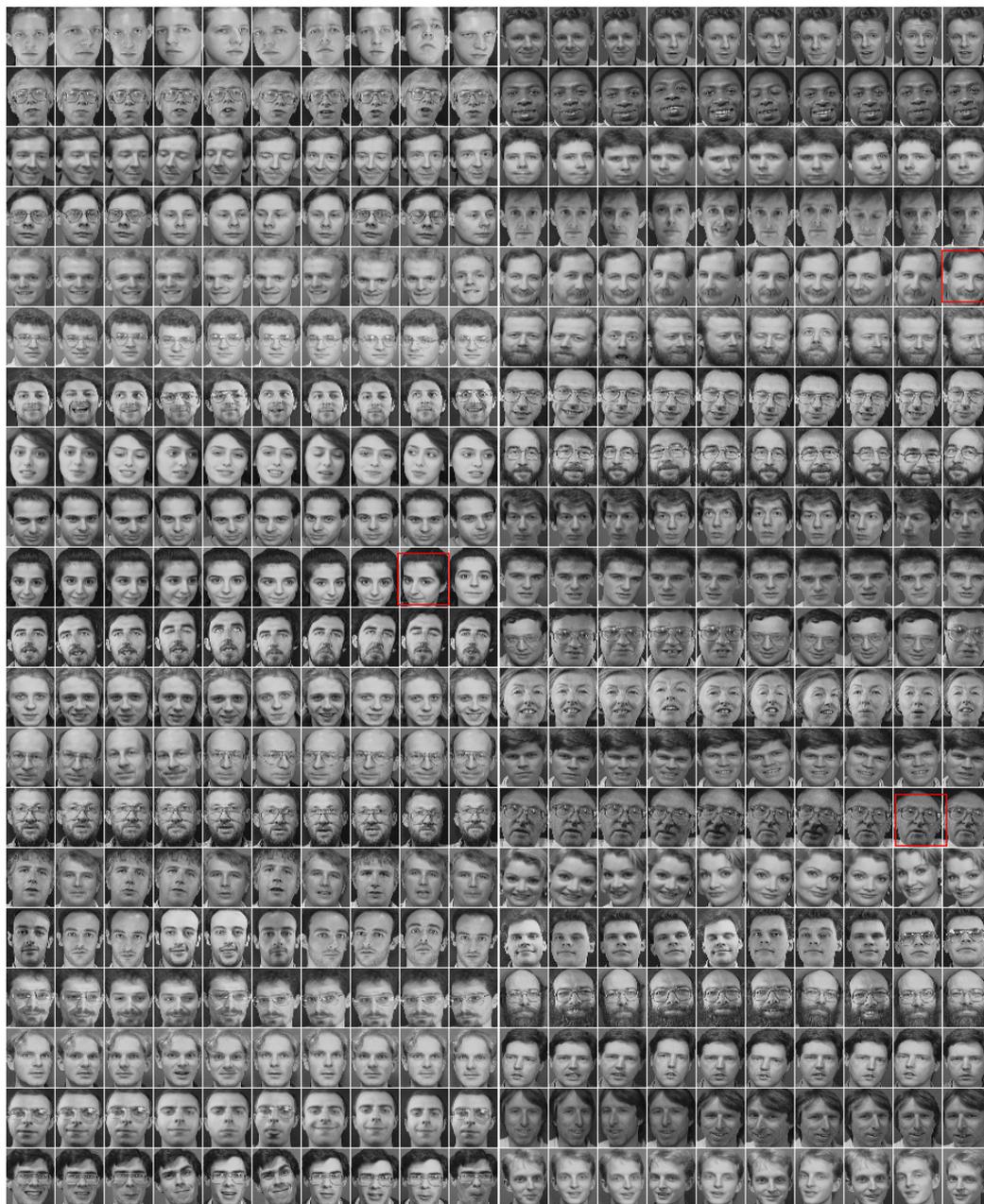


Figure 4. Mispredictions

图 4. 错误预测

5. 结论

本文阐述了 PCA 降维原理、SVD 奇异值分解、线性可分和非线性可分数据 SVM 的数学原理以及 SVM 在多分类问题中的使用方法。基于 ORL 人脸数据库对 400 张人脸图片进行数据降维，降维后的人脸向量作为多分类 SVM 模型 OVO 的样本，将样本分成训练集和测试集，并用训练集训练 OVO 模型，该 OVO 模型在测试集的精度为 96.25%，具有较高的识别率。

参考文献

- [1] Wold, S., Esbensen, K. and Geladi, P. (1987) Principal Component Analysis. *Chemometrics and Intelligent Laboratory Systems*, **2**, 37-52. [https://doi.org/10.1016/0169-7439\(87\)80084-9](https://doi.org/10.1016/0169-7439(87)80084-9)
- [2] Cortes, C. and Vapnik, V. (1995) Support-Vector Networks. *Machine Learning*, **20**, 273-297. <https://doi.org/10.1007/BF00994018>
- [3] 宋李俊, 赵虎. 基于融合特征与支持向量机的控制图模式识别[J]. 计算机应用研究, 2014, 31(3): 937-941.
- [4] 汪晓东. 基于支持向量机的气敏传感器阵列信号处理方法[J]. 仪器仪表学报, 2005, 26(8): 871-875.
- [5] 彭令, 牛瑞卿, 赵艳南, 邓清禄. 基于核主成分分析和粒子群优化支持向量机的滑坡位移预测[J]. 武汉大学学报(信息科学版), 2013, 38(2):148-152+161.
- [6] Vatsa, M., Singh, R. and Noore, A. (2005) Improving Biometric Recognition Accuracy and Robustness Using DWT and SVM Watermarking. *IEICE Electronics Express*, **2**, 362-367.
- [7] Byvatov, E. and Schneider, G. (2003) Support Vector Machine Applications in Bioinformatics. *Applied Bioinformatics*, **2**, 67-77.
- [8] Doucet, J.P., Barbault, F., Xia, H., *et al.* (2007) Nonlinear SVM Approaches to QSPR/QSAR Studies and Drug Design. *Current Computer-Aided Drug Design*, **3**, 263-289. <https://doi.org/10.2174/157340907782799372>
- [9] 白茹. 基于视频流媒体的人脸识别课堂监控系统[J]. 电子设计工程, 2021, 29(23): 166-170.