**提示信息：**为保证本文中的链接能够使用，请您先将本文件按规定的命名方式命名，并与“[USTB硕士学位论文规范及论文制作指南.exe](file:///C:\Users\a\AppData\Roaming\Microsoft\Word\USTB硕士学位论文规范及论文制作指南.exe)”保存于同一目录。

不用此信息时，删除此框。

（鼠标移到此框四边，鼠标变为十字箭头，点击边框选中此框，然后按Del删除）

测度学习研究及其应用

吴翔

北京科技大学

**加密论文编号：\_\_\_\_\_\_\_\_\_\_\_\_\_**

**密　　　　级：\_\_\_\_\_\_\_\_\_\_\_\_\_**

论文题目：测度学习研究及其应用

S20130946

学　 　号：\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

吴翔

作　 　者：\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

电子科学与技术

专 业 名 称：\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

年月日

测度学习研究及其应用

The Research on Metric Learning and its Application

研究生姓名：吴翔

指导教师姓名：石志国

北京科技大学计算机与通信工程学院学院

北京100083，中国

Master Degree Candidate： Xiang Wu

Supervisor： Zhiguo Shi

School of Computer and Communication Engineering

University of Science and Technology Beijing

30 Xueyuan Road，Haidian District

Beijing 100083，P.R.CHINA

分类号：\_\_\_\_\_\_\_\_\_\_\_\_ 密　　级：\_\_\_\_\_\_\_\_\_\_\_\_\_\_

１０００８

ＵＤＣ：\_\_\_\_\_\_\_\_\_\_\_\_ 单位代码：\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**北京科技大学硕士学位论文**

测度学习研究及其应用

**论文题目：**

吴翔

**作者：\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

北京科技大学

石志国 副教授

**指 导 教 师： 单位：**

**指导小组成员： 单位：**

**单位：**

**论文提交日期：**年 月 日

**学位授予单位：北 京 科 技 大 学**

致 谢

摘 要

关键词： （用逗号“，”分隔）

Abstract

Key Words： （用英文逗号“,”分隔）

目 录

[致 谢 I](#_Toc435115433)

[摘 要 III](#_Toc435115434)

[Abstract V](#_Toc435115435)

[插图和附表清单 IX](#_Toc435115436)

[1 引言 1](#_Toc435115437)

[2 文献综述 2](#_Toc435115438)

[2.1 测度学习的发展背景 2](#_Toc435115439)

[2.2 测度学习 3](#_Toc435115440)

[2.2.1 马氏距离 4](#_Toc435115441)

[2.2.2 线性测度学习 5](#_Toc435115442)

[2.2.3 非线性测度学习 7](#_Toc435115443)

[2.2.4 结构化测度学习 7](#_Toc435115444)

[2.3 测度学习相关领域 8](#_Toc435115445)

[3 近似相似度测度学习 10](#_Toc435115446)

[3.1 相似度测度学习 10](#_Toc435115447)

[3.1.1 余弦距离 10](#_Toc435115448)

[3.1.2 余弦相似度测度学习理论 11](#_Toc435115449)

[3.1.3 缺陷 12](#_Toc435115450)

[3.2 近似余弦相似度测度学习 12](#_Toc435115451)

[3.2.1 理论 12](#_Toc435115452)

[3.2.2 算法复杂度分析 13](#_Toc435115453)

[3.3 实验 15](#_Toc435115454)

[3.3.1 UCI数据集分类 15](#_Toc435115455)

[3.3.2 LFW人脸验证数据集分类 17](#_Toc435115456)

[3.4 结论 19](#_Toc435115457)

[4 深度测度学习 20](#_Toc435115458)

[4.1 深度学习简介 20](#_Toc435115459)

[4.2 深度测度学习 20](#_Toc435115460)

[4.2.1 深度测度学习理论 21](#_Toc435115461)

[4.2.2 混合深度测度学习 23](#_Toc435115462)

[4.3 实验 25](#_Toc435115463)

[4.3.1 UCI数据集分类 25](#_Toc435115464)

[4.3.2 LFW人脸验证数据集分类 26](#_Toc435115465)

[4.4 结论 29](#_Toc435115466)

[5 测度学习在图像检索中的应用 30](#_Toc435115467)

[5.1 图像检索概述 30](#_Toc435115468)

[5.2 基于测度学习的图像检索算法框架 30](#_Toc435115469)

[5.2.1 费舍尔特征 31](#_Toc435115470)

[5.2.2 最大边际测度学习算法 32](#_Toc435115471)

[5.2.3 图像检索系统框架 34](#_Toc435115472)

[5.3 实验 35](#_Toc435115473)

[5.4 结论 37](#_Toc435115474)

[6 结论 38](#_Toc435115475)

[参考文献 39](#_Toc435115476)

[作者简历及在学研究成果 45](#_Toc435115477)

[独创性说明 47](#_Toc435115478)

[关于论文使用授权的说明 47](#_Toc435115479)

[学位论文数据集 49](#_Toc435115480)

1. 引言

二十一世纪以来，人类进入了一个全面的信息时代。随着科学技术的发展和进步，信息在人类生活中的地位和作用越发的突显，人类对信息质量的要求也越来越高。因此，如何高质量的传递、储存和处理信息成为了人类亟待解决的问题。

在机器学习、模式识别、数据挖掘等领域，选择合适的测量准则衡量数据特征之间的距离或者相似度通常对解决问题起重要的作用。例如，K邻近算法(K-Nearest Neighbor, kNN)[1] 通常选择欧几里得距离计算邻近元素；K均值算法(K-means)[2] 基于数据特征之间的距离度量进行聚类分析。此外在信息检索领域，搜索算法也是基于查询关键字与数据库中相关词条的相似度进行排序。然而在实际系统中，仅仅使用欧几里得距离或者余弦距离度量数据特征向量之间的相关度往往很难表示数据之间本征特征关系。因此选择合适的测量准则对特征向量进行合理有效的度量便成为了机器学习领域的一个重要课题。但是在解决特定问题时，我们通常很难人工设计出合理有效的度量准则，因此2002年卡内基梅隆大学的Eric Xing教授提出了测度学习(Metric Learning)[2] 。测度学习的目标是根据数据之间的关系学习一个有效的度量准则，以此解决实际系统中数据之间相关性计算、分析等问题。传统的测度学习方法大多基于半正定优化问题(Semi-Definite Programming, SDP)。随着大数据的发展，海量数据和高维特征成为机器学习在实际系统应用的主流，但是由于传统的SDP问题优化求解时需要应用的特征根分解算法(Eigenvalue Decomposition) 的算法复杂度为O(n3)，所以当特征向量维度过高时求解的时间成本和空间成本都会大幅增加。

本文针对传统测度学习SDP问题优化求解高维特征向量的劣势，利用随机梯度下降算法(Stochastic Gradient Descent, SGD)和深度学习(Deep Learning, DL)层次结构的思想, 对测度学习优化过程进行改进，并在人脸识别、图像检索等实际应用上进行测试。解决大数据高维特征的测度学习优化问题，可以有效的适应当前机器学习和模式识别领域发展的趋势，为理论算法在计算机视觉实际系统中的应用提供可能。

本论文主体由四部分组成，第二章主要综述了测度学习近年来的发展；第三章介绍了余弦相似度测度学习的优势和不足，并提出了近似余弦相似度测度学习理论；第四章主要借鉴深度学习的优势，提出了深度测度学习理论；第五章具体阐述了测度学习在图像索引中的应用。

1. 文献综述
   1. 测度学习的发展背景

图像信息作为人们获取信息的重要来源，有着作用距离远、传输速度快、存储信息量大等其他形式的信息不具备的优点。图像信息的重要性无可置否，但是在实际应用中，我们所关心的往往不是图像所搭载的全部信息，只是其中的一部分信息，这时我们就需要进行对图像进行处理分析，以此来实现我们最终目的。所谓图像处理，就是利用计算机对图像进行分析，达到所需要的结果。图像处理主要包括图像压缩，增强和复原，匹配、描述和识别这几个部分，常见的图像处理方法有图像编码、图像复原、图像分割和图像分析等。

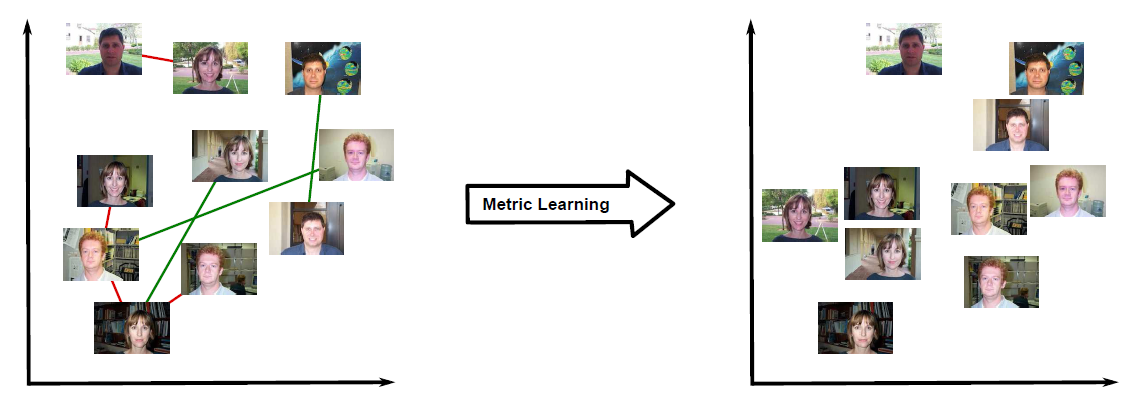


图 2-1 测度学习在人脸识别系统中的应用

图像检索和人脸识别问题是计算机视觉发展和研究的重要领域，而在处理图像检索以及人脸识别问题时，人们发现通过特征提取的方法提取图像特征往往在高维空间中呈现非线性或某种流形分布。如图 2-1左图所示，由于姿态、光照、背景拍摄环境的影响，在高维空间中，同一个人的人脸照片特征分布可能距离较远，而不同人的特征距离则相对较近，这就导致我们使用KNN或者K-means算法进行特征分类时很难获得良好的分类效果。在2002年NIPS会议上，Eric Xing提出在衡量数据之间距离时不能简单使用特征向量之间的欧式距离或余弦距离进行表征，因为在高维空间中同类特征并非呈现线性分布。基于这一特性，Eric Xing提出了测度学习的概念，即通过现有标定数据的统计特性，学习一个合适的评价数据样本间距离的测量准则，减小同类数据距离，增大不同类数据间距离，如图 2-1右图所示，经过测度学习投影，使得KNN或者K-means算法在处理高维复杂数据分类问题上更为有效。

* 1. 测度学习

最早的测度学习相关工作的研究可以追溯的上个世纪八十到九十年代[5] [6] ，然而卡内基梅隆大学的Eric Xing教授于2002年在NIPS会议上发表的论文[3] 则被认为是测度学习领域的先驱工作。

测度学习目的是寻找一个合适的投影矩阵，表征高维空间中不同类特征，增加类间距离、减小类内距离以此简化分类问题。因此，我们定义

(2-1)

其中集合表示同类数据构成集合，集合表示不同类数据构成集合。因此，我们可以简单的将测度学习目标函数定义为

(2-2)

其中表示测度学习投影矩阵，表示M的相关约束项，为损失函数(loss function)，为正则项(regularization)，为常数。不同的测度学习方法会选择不同的距离表示方法、损失函数、正则项。在经过测度学习后，通过投影矩阵M对特征进行映射，我们可以更好的使用k-means、kNN、SVM等分类器对数据进行进一步的分析。

我们可以根据学习方式、距离表示的形式、优化求解方法、是否降维等特性对各种测度学习算法进行分类。根据测度学习的学习方式可以将其分为监督学习(fully supervised)、弱监督学习(weakly supervised)以及半监督学习(semi-supervised)。

1. 监督学习

定义训练数据集，其中特征向量、标签。监督学习是在明确的指导信息下进行训练的。在训练过程中，我们根据训练数据集的标签指导信息以及的约束条件求解投影矩阵M。

1. 弱监督学习

对于弱监督学习，其训练数据集没有明确的标签指导信息，我们通过数据集只能获得边缘信息，例如搜索引擎中用户的点击信息、文章引用信息等。

1. 半监督学习

对于半监督学习，其训练数据集分为两部分，一部分是包含有标签的数据项，另外一部分是未标定数据集。在互联网大数据时代，大部分数据都是未标定数据，因而在各种工业系统应用测度学习时通常采用半监督学习方法进行训练，该种训练方法在充分利用数据的同时，也可以防止标签信息货边缘信息不足而导致的模型过拟合现象。

根据距离表示形式可以分为线性距离(linear metrics)、非线性距离(nonlinear metrics)和局部距离(local metrics)。

1. 线性距离

线性距离(例如欧氏距离)通常表现形式简单，但由于其损失函数为凸函数，因此在进行优化求解时可以直接通过梯度下降算法求解全局最优解。

1. 非线性距离

非线性距离(例如距离)，其目标函数非凸，因而其求解投影矩阵为局部最优解，求解难度较大并且容易过拟合。

测度学习的优化方法根据目标函数的不同分为凸优化方法和非凸优化方法。对于目标函数为凸函数的问题，采用凸优化方法，求解的投影矩阵为全局最优解(global optimum)；对于目标函数非凸的问题，最终求解的投影矩阵为局部最优解(local optimum)。

此外，不同的测度学习算法根据是否降维也可以进一步分类。一些测度学习算法通过将数据特征投影到新的低维特征空间获得更紧致的特征已达到更好的分类效果和更快的计算速度。

* + 1. 马氏距离

所谓马氏距离(Malahanobis Distance)，其定义为

(2-3)

其中为特征向量，，表示维空间上对称半正定(symmetric positive semi-definite)锥所围成的空间，如图 2-2所示。当为单位矩阵时，马氏距离即为欧几里得距离。

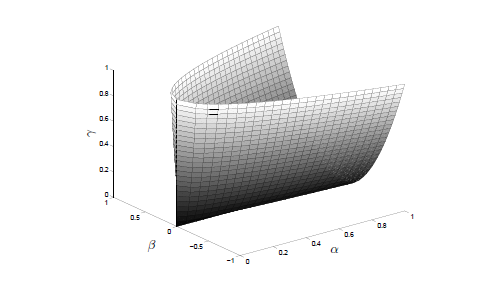


图 2-2 对称半正定矩阵所表征的空间

为半正定矩阵可以确保马氏距离有如下性质

1. 非负性：
2. 同一性：
3. 对称性：
4. 三角不等关系：

此外，因为M为对称半正定矩阵，因为我们可以将M表示为，则马氏距离可表示为

(2-4)

如果M为低秩矩阵，即，则矩阵M相当于将原特征向量映射到一个低维空间，这将保证我们获得更紧致的特征向量从而减小距离计算时的时间和空间复杂度、较小存储特征向量所用的空间以及获取更好的分类效果。

在测度学习求解马氏距离中，要保证投影矩阵M始终满足对称半正定约束，最有效的办法是使用投影梯度法(Projection Gradient Method)，该种方法通过将负特征向量置零，保证每次将计算的梯度均可投影到一个对称半正定锥所围成的求解空间内。但是由于投影梯度法的时间复杂度为，因此当特征向量维度*d*过大时，投影梯度法的计算复杂度增加。此外，由于优化低秩矩阵约束为NP问题，因此如何学习一个低秩投影矩阵代替满秩矩阵达到更好的分类效果，也是测度学习中亟需解决的问题。

* + 1. 线性测度学习

线性测度学习通常基于马氏距离表达式，求解对称半正定投影矩阵M。

最早提出测度距离，并将马氏距离应用在测度距离求解的方法是Eric Xing[2] ，他提出的测度学习算法基于无正则项的凸优化目标函数

(2-5)

Eric Xing通过在梯度下降算法中使用投影梯度法保证M的对称半正定特性，求解投影矩阵M。

Weinberger等人提出的最大边缘最邻近算法(Large Margin Nearest Neighbors, LMNN)[7] [8] [9] 是线性测度学习算法中应用相当广泛的算法之一。LMNN算法其优化目标函数为

(2-6)

其中。LMNN与SVM类似[10] ，使用hinge loss作为损失函数，因而在求解LMNN时，可以采用传统的梯度下降法(gradient descent method)或者基于在线学习(online learning)的随机梯度下降算法(stochastic gradient descent, SGD)。

信息论测度学习算法(Information-Theoretic Metric Learning, ITML)[11] 是另一基于马氏距离测度学习的经典算法。该算法引入对数散度正则项(LogDet divergence regularization)

(2-7)

其中d为输入特征向量的维度，为对称半正定矩阵。在实际应用中，通常将先验对称半正定矩阵为单位矩阵或者训练数据的协方差矩阵。单位矩阵约束可以保证学到的马氏距离更接近欧氏距离，协方差矩阵约束则可以保证投影矩阵在训练数据集上有最大类间距离。对数散度正则项是根据信息论中最小化两个参数分别为和的多维高斯分布的KL散度(KL divergence)获得的，其意义在于传统的马氏距离测度学习中引入了概率和信息论的概念，使所学的投影矩阵更符合某种统计学特性。ITML优化的目标函数为

(2-8)

其中为截断阈值。ITML算法的缺陷在于先验对称半正定矩阵需要手工给定，并且其正则项推倒假设基于高斯分布，因而并非适用于各种类型的数据。

随着稀疏表示(Sparse Representation)的发展，在各种机器学习问题中引入稀疏性可以增加机器学习系统的鲁棒性，在2009年ICML会议上，Qi等人提出了稀疏距离测度学习(Sparse Distance Metric Learning, SDML)[12] 。SDML的目标函数可以表示为

(2-9)

其中为除对角线元素外的L-1范数，通过该范数约束保证求解投影矩阵的稀疏性。SDML通过引入辅助变量将目标函数转换为半定优化问题(Semi-Definite Programming, SDP)使用区域坐标下降法(block coordinate descent)[14] 进行求解。

* + 1. 非线性测度学习

非线性测度学习较线性测度学习而言，虽然由于其目标函数通常为非凸函数，求解投影矩阵通常为优化问题的局部最优解而非全局最优解，在实际问题中容易对训练数据产生过拟合现象，但因为其非线性特性因而具有比线性测度学习更强的表达能力，因此也被广泛应用。

非线性测度学习通常分为两类问题，一类问题成为核函数化(kernelization)，该种方法与核函数学习方法类似，其目的是通过核函数映射的方法，在非线性特征空间中学习线性测度表示[15] [16] 。尽管通过核函数非线性映射，但是由于线性测度学习表达式均为矩阵内积运算(即线性投影)，因此会造成导数难以求解等问题。此外由于核函数映射需要矩阵映射原数据特征向量(其中n为训练数据集总量)，因此当训练数据集过大时，该优化问题将变得很难求解。另一类非线性测度学习目标是学习一个非线性距离。Chopra提出非线性测度学习方法[17] ，该方法通过学习非线性映射最小化L1距离。其中非线性映射函数并没有具体的函数形式或假设，参数W通过卷积神经网络(convolution neural network, CNN)利用后向传播算法(back-propagation, BP)与随机梯度下降算法(stochastic gradient descent, SGD)最小化损失函数学习。

* + 1. 结构化测度学习

结构化测度学习基于结构化支持向量机(Structural SVM)的求解框架，最早由McFee在2010年ICML会议上提出，是用于解决信息检索中的排序问题而提出的测度学习排序算法(Metric Learning to Rank, MLR)[18] 。其目标函数为

(2-10)

其中

(2-11)

分别为重排序标签以及原标签，为排序函数。该优化问题通过割平面(cut plane method)进行优化求解。

由于测度学习容易受噪声数据影响，因而McFee在2013年ICML上基于MLR算法又提出了鲁棒性结构化测度学习(Robust Structural Metric Learning to Rank, R-MLR)[19] ，其目标函数为

(2-12)

其中，即该范数表示投影矩阵W每行L2范数求和。由于结构化支持向量机割平面优化求解问题的方法的算法复杂度为，因此McFee将原优化问题分解为三个子优化问题并引入拉格朗日乘子进行求解。

* 1. 测度学习相关领域

测度学习作为机器学习领域的一个重要分支，在计算机视觉、信息检索等领域扮演着重要角色。对于计算机视觉问题，一个合适的测度投影矩可以有效的表示和处理从图像和视频中通过词袋模型(Bags-of-Visual-Words, BoVW)[13] 或者卷积神经网络(Convolution Neural Network, CNN)[20] 等特征提取算法提取出的特征向量。近年来，测度学习在大规模数据计算机视觉系统中被广泛应用，其应用方面涉及图像分类(image classification)[20] 、目标检测(object recognition)[21] [23] 、人脸识别(face recognition)[24] 、目标跟踪(visual tracking)[25] 以及图像检索与标注(image retrieval and annotation)[26] 等领域。对于信息检索问题，大多数信息检索系统都是通过计算特征向量之间的相似性进行检索结果的查找和排序的，因此，选择合理的相似性计算准则对信息检索系统的准确率和召回率均有很大的提升。

测度学习作为机器学习领域内的一个重要分支，与其他机器学习问题如核方法(kernel learning)[27] [28] 、降维问题(dimensionality reduction)[29] 有密切的联系和区别。

核方法通常都是无参的(nonparamteric)。核方法是通过数据本身的统计特性，在没有对核函数附加任何约束条件下，利用产生式模型(generative model)的方法生成核矩阵(kernel matrix)，实现数据映射。因此核方法通常受限于数据分布类型，对于未知或者与训练数据分布不相符的测试数据集，核方法所学的投影矩阵很难达到较好的分类或聚类效果。不同于核方法，多核学习方法(multiple kernel learning)是一种有参模型，该方法通过学习预先定义的核函数中的超参数投影表征特征。但是由于核函数中超参数是根据先验知识设定，因此该方法在应用上比核方法更受限制。而对于测度学习问题，通常都是有参问题(parametric problem)的研究(即通常我们需要学习一个参数矩阵表征特征)，但是其参数投影矩阵的求解通常只要求投影矩阵半正定，因此其在实际系统中比起核方法和多核方法具有更好的适应性和灵活性。

对于特征降维问题，通常分为监督降维(supervised dimensionality reduction)与无监督降维(unsupervised dimensionality reduction)。对于监督学习降维问题，其目标通常是寻找一个低维表征空间，可以最大化不同标签数据之间的空间距离分布。对于无监督学习降维问题，又称为流形学习(manifold learning)，其通常认为数据在低维空间中呈某种流形分布，通过线性或非线性映射可以在保持原有数据统计特征(方差、距离等)不变并投影特征是数据具有不再具有复杂的流形分布特性。相较各种特征降维算法，一些测度学习通过构造新的低维特征空间映射原数据特征向量，减小类内距离、增大类间距离，其目的与降维算法异曲同工。

1. 近似相似度测度学习
   1. 相似度测度学习
      1. 余弦距离

欧几里得距离是我们经常衡量数据之间相关性的距离准则之一，除此之外，余弦相似度(Cosine Similarity)也是我们经常用来衡量数据之间关系的重要指标。

首先根据定义余弦距离为

(3-1)

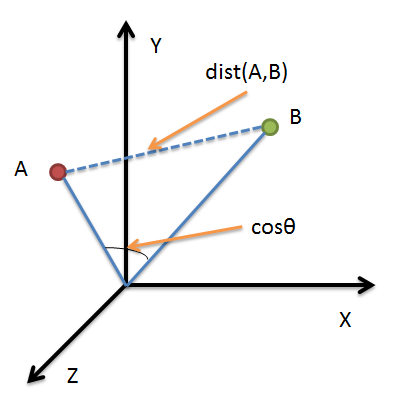


图 3-1 欧氏距离与余弦距离示意

如图 3-1，欧氏距离衡量空间中两点之间的距离，余弦距离衡量空间中两个向量之间方向上的差异。由于余弦距离更加注重维度之间的差异性，因此其可以有效的规避个体相同认知中不同程度的差异表现，而欧氏距离更加注重数值上的差异，衡量的是两个向量之间的相同程度。根据公式(3-2)我们可以发现，当对特征向量进行归一化处理后，欧氏距离等价于余弦距离。此时我们可以认为在一个归一化超球面上，欧氏距离表示的是球面上两点的直线距离，余弦距离表示的是球面上两点的球面距离，本质上相同。

(3-2)

根据余弦距离的表达式我们可以发现，而欧氏距离，因此余弦距离在衡量数据相关性做进一步归一化处理更为直观和方便。

* + 1. 余弦相似度测度学习理论

Qamar & Gaussier提出广义余弦相似度学习方法(Generalized Cosine Similarity Learning Algorithm, gCosLA) [30] ，其余弦相似度表达式为

(3-3)

其中。当M矩阵为单位阵时该距离为余弦距离。Qamar使用Hinge Loss作为损失函数，基于约束投影矩阵M具有对称半正定性质，作者提出了一种基于批数据的在线学习方法，利用特征根分解求解投影矩阵M。

在2010年亚洲计算机视觉会议上，Nguyen & Bai同样提出了余弦相似度测度学习算法(Cosine Similarity Metric Learning, CSML)[31] 。与Qamar不同，Nguyen的优化目标函数基于相似度表达式为

(3-4)

根据公式(2-1)，Nguyen定义了测度学习的损失函数为

(3-5)

其中集合表示同类数据构成集合，集合表示不同类数据构成集合，损失函数公式(3-5)通过最大化正负样本之间的距离求解投影矩阵。参数用来处理正负样本不平衡问题，我们可以简单的将赋值为

(3-6)

也可以利用交叉验证(cross validation)的方法对进行计算求解。为正则项(regularizer)，用来约束求解投影矩阵参数，抑制过拟合问题。

Nguyen使用共轭梯度法(Conjugate Gradient Descent Method, CGD)对公式(3-5)进行优化，根据公式(3-4)和(3-5)可以计算出余弦距离的梯度为

(3-7)

其中

(3-8)

* + 1. 缺陷

根据公式(3-5)、(3-7)、(3-8)，我们可以得出余弦距离梯度的复杂度为O()，其中为训练样本的数量，为特征向量的维度，为经过测度学习投影后特征向量的维度。因此我们可以得到余弦测度学习算法的复杂度为O()，其中为优化目标函数迭代次数，*b*为交叉验证过程的次数，*g*为共轭梯度法迭代的次数。

因此我们可以发现，当训练数据集的特征向量维度过高时，相似度测度学习的时间复杂度和空间复杂度会迅速增加，导致我们很难计算出测度学习所需要的投影矩阵。因此，本文提出了近似余弦相似度测度学习方法(Qusai Cosine Similarity Metric Learning, QCSML)[32] ，以此解决相似度测度学习在处理高维特征向量所遇到的问题。

* 1. 近似余弦相似度测度学习

本章节主要介绍近似余弦相似度测度学习方法，引入hinge-loss作为目标函数，使用随机梯度下降进行目标函数优化。

* + 1. 理论

首先我们定义训练集，余弦相似度的定义如公式(3-4)所示，根据不等式，我们可以对余弦相似度作如下变换:

(3-9)

因此余弦相似度可以表示为

(3-10)

为了简化余弦距离表达式，我们可以对投影矩阵W增加先验约束，即

(3-11)

其中表示矩阵的F正则项(Frobenius norm)，Tr()表示矩阵的迹(trace)。同时，我们可以将训练集特征向量进行归一化处理，即令，这样我们可以将余弦距离化简为

(3-12)

定义为一对训练样本的标签信息，则可以定义

(3-13)

同时，定义余弦距离的参数阈值*b*，以此区分是否属于同一类，则可以得到约束

(3-14)

根据公式(3-12)和公式(3-14)，可以将该问题的约束条件转化为

(3-15)

根据支持向量机(Support Vector Machine, SVM)[33] 的思想，利用hinge loss最大化两类问题之间的间隔，构建目标函数可以得到

(3-16)

对于有约束优化问题，使用拉格朗日乘子法(Lagrange Multipleier Method)转化为无约束优化进行求解，首先对训练集特征向量进行归一化处理，则原目标函数可以转化为

(3-17)

其中为拉格朗日乘子并且满足约束条件。

* + 1. 算法复杂度分析

根据公式(3-17)所示，近似余弦测度学习的目标函数为凸函数，因此我们可以利用梯度下降方法去优化求解该目标函数。

为了适应高维特征向量和大数据，我们采用随机梯度下降方法(Stochastic Gradient Descent, SGD)[34] 代替Nguyen使用的共轭梯度法。根据目标函数，我们可以得到投影矩阵的优化梯度更新公式为

(3-18)

其中为学习率，表示目标函数，则目标函数的梯度为

(3-19)

同时阈值*b*的更新公式为

(3-20)

其中为调整阈值*b*的偏置系数。

表 3-1 近似余弦相似度优化算法流程伪代码

|  |
| --- |
| 算法：近似余弦相似度测度学习优化算法 |
| **输入**：训练数据  **输出**：参数  **Begin**  参数初始化;      ;  ;  **for** to *n* **do**  **switch****do**  **case** 1:  score = ;  **if** score **then**  ;  ;  **break;**  **case** -1:  score = ;  **if** score **then**  ;  ;  **break;**  **return** ; |

近似余弦相似度测度学习算法优化的具体流程如所示，其中参数的设置对测度学习参数优化的结果起重要的作用。学习率控制梯度下降的速度，由于通常随机梯度下降算法使用batch进行迭代优化，因此学习率设置与其他梯度下降算法例如L-BFGS、共轭梯度法等应该较小，以此保证优化迭代的损失函数不会因为局部采样不具有显著性而产生较大波动。此外，也是优化目标函数中非常重要的参数之一。

尽管公式(3-17)所示的优化目标函数为凸函数，但是一个较好的投影矩阵初始值可以很快的是随机梯度收敛到全局最优解。本文中使用白化主成分分析(Principal Components Analysis Whitening, PCA-Whitening)和K均值(K-means)的方法初始化投影矩阵*M*。

根据可知目标函数梯度计算的复杂度为O()，其中为输入特征向量的维度，为经过投影矩阵计算输出的维度。因此近似余弦相似度测度学习的算法复杂度为O()，较余弦相似度测度学习算法的时间复杂度O()大大地提高。

* 1. 实验
     1. UCI数据集分类

本章节分别选择Iris[[1]](#footnote-1)、Ionosphere[[2]](#footnote-2)和wine[[3]](#footnote-3)三个UCI常用数据集对论文算法进行测试评估。上述三个数据集的详细信息和处理方法如下:

1. Iris数据集：该数据集有3类共计150个样本组成(每类50个样本)，每个样本的特征向量维度为4维，实验中设置训练集为120个样本(每类数据40个样本)，测试集为30个样本。
2. Ionosphere数据集：该数据集有2类共计351个样本组成，其特征向量维度为34维，实验中随机抽取200个样本为训练集，其余151个样本为测试集。
3. Wine数据集：该数据集共有3类数据总计178个样本，其特征向量维度为13维。我们随机抽取150个样本作为训练集，其余28个样本作为测试集。

实验从分类准确率和训练速度两个方面进行评估，并与以下方法进行对比：

1. 欧氏距离(Euclidean)：欧氏距离表达式
2. 协方差距离(InvCov)：协方差距离为特殊的马氏距离，其投影矩阵为训练集的协方差矩阵，该投影矩阵等即为对训练数据集做主成分分析求得投影矩阵。
3. 最大近邻测度学习(LMNN)[8] ：最大近邻测度学习(Large Margin Nearest Neighbor Metric Learning Method, LMNN)由Weinberger提出，利用hinge loss优化欧氏距离最大间距求解投影矩阵。
4. 信息论测度学习(ITML)：信息论测度学习(Information-Theoretuc Metric Learning, ITML)引入LogDet正则项并使用相对熵损失函数对测度学习进行优化。
5. 稀疏距离测度学习(SDML)：稀疏距离测度学习(Sparse Distance Metric Learning)使用L1正则项作为稀疏约束优化距离测度学习目标函数。

表 3-2 不同测度学习算法在UCI数据集上的分类错误率(%)比较

|  |  |  |  |
| --- | --- | --- | --- |
|  | IRIS | IONOSPHERE | WINE |
| Euclidean | 4.00 | 14.86 | 4.5 |
| InvCov | 8.67 | 17.71 | 43.82 |
| LMNN [8] | 3.34 | 14.29 | 2.25 |
| ITML[11] | 3.00 | 17.14 | 3.94 |
| SDML(IDENTITY MATRIX) [12] | 2.00 | 13.71 | 0.5618 |
| SDML(Inverse Covariance) [12] | 2.00 | 12 | **0** |
| QCSML(random) | 4.34 | 13.49 | 5.81 |
| QCSML(K-means) | **1.04** | 7.99 | **0** |
| QCSML(PCAW) | 2.22 | **6.93** | **0** |

图 3-2 不同测度学习算法训练时间比较

不同测度学习在UCI三个数据集上分类结果如所示，可以看出QCSML算法计算出的投影矩阵在分类精度上要明显优于其他测度学习算法。同时我们针对不同的*W*投影矩阵初始化方法进行了对比发现，对训练集采用白化主成分分析或者K均值方法进行*W*初始化，有助于随机梯度下降方法在有限迭代次数内收敛到更优解。

如比较了QCSML与其他测度学习算法在训练时间上的差异，我们可以看到QCSML在UCI三个数据集上的训练时间远远快于LMNN和ITML，与SDML算法的计算时间成本基本在一个数量级上。此外，由于QCSML采用随机梯度下降算法，因此在针对高维向量和大数据训练集时会有更明显的训练速度差异。

* + 1. LFW人脸验证数据集分类

本章节采用QCSML算法对人脸识别领域的公共测试数据库LFW[[4]](#footnote-4)进行测试评估。

LFW数据集[36] 是人脸识别领域知名的公共测试数据集，数据集包含5749个人总计13233张人脸图片。LFW将数据分成10个部分，其中每个部分包含600对人脸测试样本(300对正样本，300对负样本)，其总共包含两种测试方法——受限配置测试方法和非受限配置测试方法。对于受限测试方法，每次选择9个部分5400对样本作为训练数据集，余下的600对样本作为测试数据及，依照此规则分别测试10个部分，计算平均准确率和方差；对于非受限测试方法，可以选择除测试600对样本之外的任何数据作为训练，然后进行测试并计算准确率和方差。

对于人脸验证问题，如何寻找一个鲁棒的特征表征人脸图片同样是影响验证准确率的重要因素。本文采用费舍尔向量(Fisher Vector, FV)作为人脸图像的特征向量。费舍尔向量在诸多计算机视觉领域如图像分类[38] 、图像索引[39] 、人脸识别[40] 上均有优异的表现。本文使用ROC曲线评价人脸验证的准确率。

对于受限测试方法，我们比较B/G sample方法[41] 、LDML方法、DML-eig方法、LBP-CSML方法、SML方法[44] 以及Fisher Vector Face方法与QCSML方法的异同，不同人脸验证方法的准确率如表 3-3所示，QCSML在128维特征向量时，准确率为87.47%，相较FV特征直接做PCA降维到128维在准确率上提升了9%左右，相较CSML提升了将近2%左右。同时，QCSML算法也在LFW受限测试方法上超越了其他人脸识别算法。

表 3-3 QCSML方法与其他人脸验证方法在LFW受限测试方法上准确率的比较

|  |  |  |
| --- | --- | --- |
| 方法 | 维度 | 准确率(%) |
| Combined B/G[41] | - | 86.83±0.34 |
| LDML[42] | - | 79.27±0.60 |
| DML-eig[43] | - | 85.65±0.56 |
| LBP+CSML[31] | 200 | 85.57±0.52 |
| Sub-SML[44] | 300 | 86.73±0.53 |
| FV+PCA-Whitening | 128 | 78.60±N/A |
| Fisher Vector Face | 128 | **87.47±1.49** |
| FV+QCSML | 256 | 87.10±1.25 |
| FV+QCSML | 128 | **87.47±1.99** |
| FV+QCSML | 64 | 85.20±1.39 |
| FV+QCSML | 32 | 84.53±1.74 |

表 3-4 QCSML方法与其他人脸验证方法在LFW非受限测试方法上准确率的比较

|  |  |
| --- | --- |
| 方法 | 准确率(%) |
| LDML-MKNN[42] | 87.50±0.40 |
| PLDA combined[44] | 90.07±0.51 |
| Joint Bayesian combined[46] | 90.90±1.48 |
| Sub-SML combined[44] | 90.75±0.64 |
| Fisher Vector Faces[40] | **93.03±1.05** |
| FV+QCSML | **92.33±1.12** |

对于非受限测试方法，其实验结果如和所示。QCSML在LFW非受限测试方法下获得了92.33%的准确率，其结果及其接近Fisher Vector Face 93.03%的准确率。根据所示，QCSML方法在LFW人脸验证的测试结果也超过了LDML-MKNN[42]、PLDA[44] 、Joint Bayesian[46]以及Sub-SML[44]方法。

此外如 (b)中ROC曲线所示，尽管QCSML算法在准确率的实验结果略低于Fisher Vector Face 93.03%的结果，但是我们可以发现在误检率较低时，QCSML算法ROC曲线明显高于Fisher Vector Face的ROC曲线，在误检率(false positive rate)为0.01时，QCSML算法的准确率为82.03%，而Fisher Vector Face的准确率仅为73.41%。对于人脸验证系统，系统本身性能往往关注的不是等错率而是低误检率时，算法准确率的高低。因此。QCSML算法相较Fisher Vector Face在实际系统中有更好的性能表现。

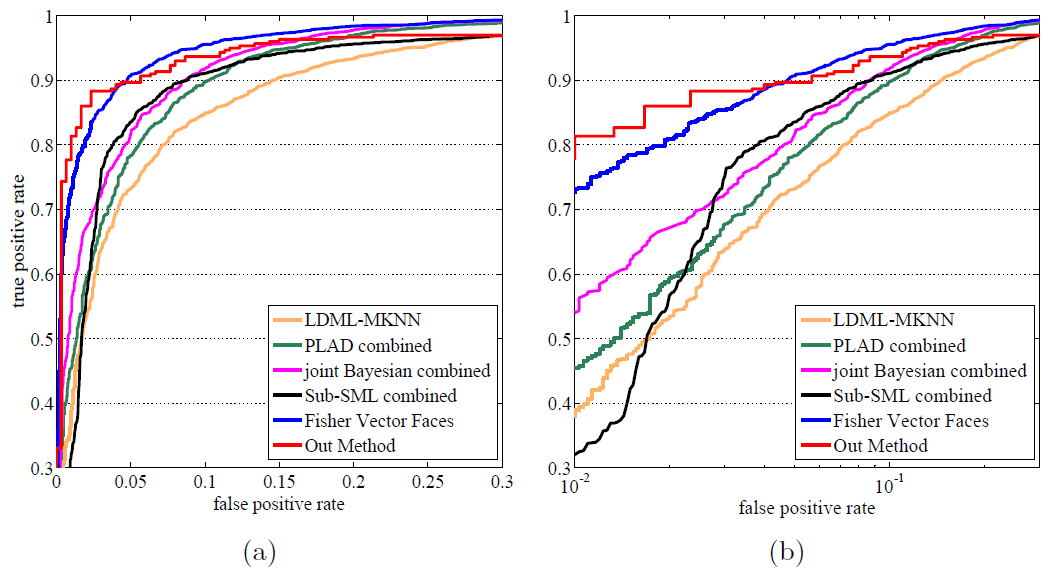


图 3-3 QCSML方法与其他人脸识别算法在LFW非受限测试下的ROC曲线。(a)为线性坐标系；(b)为对数坐标系

* 1. 结论

本章节主要介绍了余弦相似度测度学习(CSML)，并针对CSML的不足，我们提出了近似余弦相似度测度学习(QCSML)，利用向量不等式简化余弦距离，构造有约束的测度学习目标函数，然后根据拉格朗日乘子法将约束优化转化为无约束优化并使用随机梯度下降方法对目标函数进行优化求解。章节3.2的理论分析和章节的实验结果表明QCSML在速度和准确率上较CSML算法和其他测度学习算法有明显的提升。

1. 深度测度学习
   1. 深度学习简介

对于传统的机器学习问题，通常应用统计学或者线性代数的理论设计一个合理有效的分类器，解决实践应用中遇到的各种问题。这些算法如支持向量机(Support Vector Machine, SVM)、逻辑回归(Logistic Regression, LR)[48] 、随机森林(Random Forest)等方法均是利用数据的空间分布特性选择合适的超平面进行分类，因此当数据在特征空间上不具有可分性时，分类器分类性能下降。为了更好的表征度量方式，发现数据在特征空间上具备的各种性质，进一步提出了核方法(kernel method)。核方法固然可以通过将低维度的特征映射到无限维提高数据的可分性，然而由于计算时间复杂度和空间复杂的限制，一直很难在工业界推广，在大数据时代简单的机器学习模型会比复杂的机器学习模型更加有效。然而传统的机器学习方法，通常均是通过人工经验抽样提取特征，强调模型的分类预测性能，因而人工特征的好坏就成为机器学习系统的瓶颈。

2006年，加拿大多伦多大学Geoffrey E. Hinton教授在Science上发表了其最新的研究成果——深度置信网络(Deep Belief Network, DBN)[50] [51] ，提出了深度学习(Deep Learning, DL)概念，将机器学习研究推向了一个新的高潮。深度置信网络利用多层神经网络应用监督学习的方法学习通过一个多隐层模型利用机器学习和海量数据学习更有用的特征，以此提升分类或预测精度。区别于传统机器学习算法，深度学习突出了特征学习的重要性，通过逐层次的特征映射，将数据原空间的特征映射到一个新的特征空间中，使得分类和预测更加容易。与人工特征相比，深度学习可以利用数据学习更符合需求的专用特征，克服了人工特征不可扩展的缺陷。

* 1. 深度测度学习

所谓深度测度学习，即利用深度学习层次结构的优势，通过多次空间投影和非线性映射，利用后向传播算法(backpropogation, BP)求解测度矩阵，寻找更鲁棒的特征空间，以此提高分类精度。

本章节主要介绍深度测度学习理论、优势以及其存在的缺点，并提出了混和深度学习方法，抑制深度测度学习可能出现的诸多问题。

* + 1. 深度测度学习理论

给定训练数据集，同时定义一个对称半正定矩*M*，则马氏距离为

(4-1)

根据*M*矩阵对称半正定的性质，我们可以将其分解为，则公式(4-1)可以表示为

(4-2)

根据公式(4-2)，我们可以将参数矩阵*W*看做是将特征向量从欧式空间投影到一个分布在希尔伯特空间(Hilbert Space)的高维流行(manifold)上，同时也可以看做是人工神经网络中输入层(input layer)向隐层(hidden layer)的映射。

因此根据神经网络理论，我们可以定义

(4-3)

其中为输入层，为激活函数(activation function)。在传统的神经网络中我们定义sigmoid函数或者tanh函数作为激活函数，然而由于sigmoid函数或者tanh函数对于多层神经网络存在梯度消失现象，因此导致在很长一段时间多层神经网络并不能取得良好的分类效果。在2012年的imagenet竞赛上，多伦多大学的Hinton教授使用ReLU激活函数代替sigmoid和tanh，取得了长足的进步。ReLU激活函数如下

(4-4)

同时根据马氏距离计算公式(4-1)，可以得到

(4-5)

其中，将数据特征通过逐层非线性变化映射到分布的高维特征空间中。通用逻辑回归损失函数(Generalized Logistic loss)[52] 为

(4-6)

其中表达式通过定义是否属于同一类，为判断是否属于同一类的阈值，函数。

如果采用公式(4-6)作为目标函数，可以发现与其他最小均方误差问题相同，公式(4-5)中使用L2距离可能对异常值(outliers)比较敏感，为了克服这一缺点，我们引入L1距离

(4-7)

此外，选择一个合适的初值对于目标函数求解优化到一个较优解会产生很大的影响。因此，为了克服公式(4-6)，我们使用softmax损失函数

(4-8)

根据公式(4-8)，我们可以定义深度测度学习优化目标函数为

(4-9)

其中定义

(4-10)

表 4-1 深度测度学习算法流程

|  |
| --- |
| 算法：深度测度学习训练算法 |
| **输入**：训练数据，深度测度学习总层数*n*，学习率，迭代次数*T*，正则项参数  **输出**：参数  **Begin**  使用公式(4-11)初始化参数;  **for** to *T* **do**  随机选择迭代batch ;  ;  根据公式(4-9)以及后向传播算法(backpropagation, BP)[54] 计算梯度;  更新参数;  **return** ; |

我们使用随机梯度下降(Stochastic Gradient Descent, SGD)优化目标函数公式(4-9)。因为公式(4-9)中进过多层非线性映射，所以其优化目标函数非凸，因此一个好的初始值对优化求解问题十分关键，我们设置偏置参数*b*的初始值为0，而对于*W*采用Bengio提出的初始化方法[55]

(4-11)

其中为区间上的均匀分布，为深度测度学习*m*层输入节点的个数，深度测度学习优化求解方法如表 4-1所示。

* + 1. 混合深度测度学习

由于深度测度学习是通过样本对的相异性优化目标函数，求解参数矩阵。因此，如何克服训练数据集之间正负样本对的不平衡问题，成为影响深度测度学习性能的一个重要因素。

表 4-2 混合深度测度学习训练算法

|  |
| --- |
| 算法：混合深度测度学习训练算法 |
| **输入**：训练数据，深度测度学习总层数*n*，学习率，迭代次数*T*，正则项参数，损失权值参数  **输出**：参数  **Begin**  使用公式(4-11)初始化参数;  **for** to *n* **do**  随机选择迭代batch ;  ;  根据前向传播公式(4-9)和(4-12)计算和;  根据公式(4-14)计算混合损失函数的梯度;  根据后向传播算法(backpropagation, BP)计算深度测度学习逐层梯度;  更新参数;  **return** ; |

采样算法或者价值函数敏感(cost-sensitive)方法是机器学习中常用的解决不平衡数据学习问题的方法[56] 。采样方法通常试图通过对少数样本进行上采样(over-sampling)或者对多数样本进行降采样(under-sampling)平衡数据之间的分布，而价值函数敏感方法则通过修改损失函数技巧平衡不同样本之间对目标优化带来的影响。

对于测度学习算法，对负样本进行降采样或者对正负样本使用不同的损失函数均可以有效解决数据分布不平衡的问题。本文则提出了混合深度测度学习(Hybrid Deep Metric Leanring, HDML)的概念，通过结合分类损失函数(classification loss)和验证损失函数(verification loss)解决数据不平衡问题。

我们定义训练集为作为分类问题的输入数据，同时构造。根据多层神经网络的损失函数，我们采用相互熵损失(cross-entropy loss)函数作为分类的损失函数

(4-12)

分类信号可以保证经过多层映射后的特征在空间上具有较好的类间可分性，可以通过cross-entropy损失函数的放大有效的增加不同类之间的类间距离(inter-class distance)。而验证信号则可以有效的减小映射到同一凸锥上的样本特征的类内距离(intra-class distance)。混合深度测度学习模型表示如图 4-1所示，并且我们定义混合深度测度学习目标函数为

(4-13)

其中和为分类信号和验证信号的权值参数，为正则项，用于抑制过拟合。

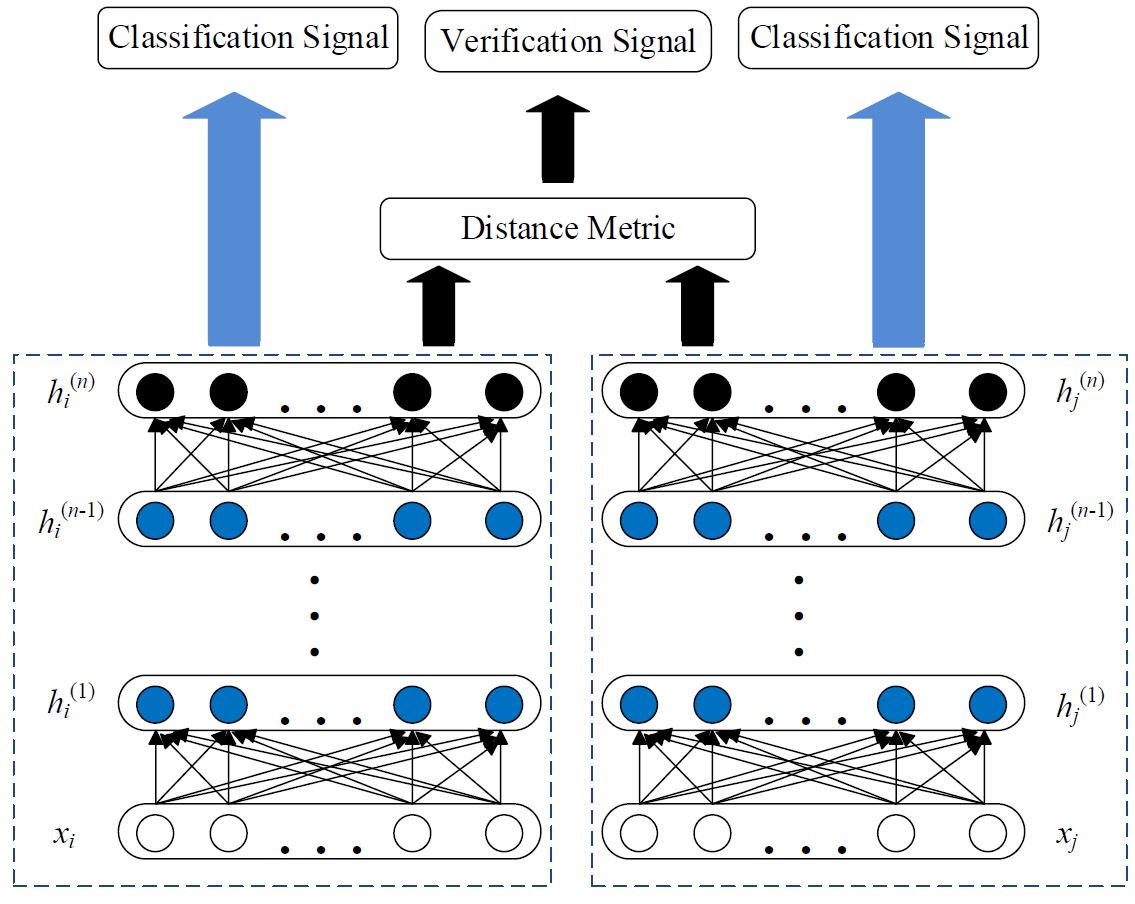


图 4-1 混合深度测度学习模型

根据公式(4-13)，可以计算混合深度测度学习损失函数层的梯度公式为

(4-14)

分类信号可以看做对于深度测度学习的一项数据依赖约束项，我们采用随机梯度下降方法优化目标函数，优化算法如所示。

* 1. 实验
     1. UCI数据集分类

本章节分别选择Diabetes[[5]](#footnote-5)、Ionosphere[[6]](#footnote-6)和Sonar[[7]](#footnote-7)三个UCI常用数据集对论文算法进行测试评估。上述三个数据集的详细信息和处理方法如下:

1. Diabetes数据集：该数据集有2类共计768个样本，其特征向量长度为8维，实验中抽取600个样本作为训练数据集，其余168个样本作为测试数据集。
2. Ionosphere数据集：该数据集有2类共计351个样本组成，其特征向量维度为34维，实验中随机抽取200个样本为训练集，其余151个样本为测试集。
3. Sonar数据集：该数据集共208个数据，特征向量长度为60维，实验中选择150个数据作为训练集，其余58个数据作为测试集。

针对UCI数据集，我们设置HDML算法训练参数为。此外，我们设置学习率，dropout参数设置为0.5，以此来抑制可能出现的过拟合现象。

表 4-3 不同测度学习算法在UCI数据集上的分类精度(%)比较

|  |  |  |  |
| --- | --- | --- | --- |
|  | Diabetes | Ionosphere | Sonar |
| Euclidean | 55.83% | 63.38% | 54.76% |
| InvCov | 58.12% | 82.29% | 60.58% |
| LMMN[8] | 60.44% | 85.71% | 85.58% |
| ITML[11] | 60.21% | 82.86% | 76.44% |
| DML | 65.58% | 91.55% | 64.29% |
| HDML | **74.68%** | **92.96%** | **85.71%** |

我们使用caffe[57] 进行模型训练，然后使用kmeans算法对学到的特征进行聚类测试，DML以及HDML算法与其他测度学习算法聚类准确率对比如表 4-3所示。我们可以发现DML和HDML较其他测度学习算法可以明显的学到更优的特征空间。此外，我们还发现HDML引入分类信号可以有效的提高特征鲁棒性，寻找到更可分的高维特征空间。

* + 1. LFW人脸验证数据集分类

人脸验证实验我们使用LFW作为测试数据集，WDRef[[8]](#footnote-8)数据集作为训练数据，WDRef数据集包含2995个人总计71846张人脸图片，并且该数据集提供了LBP特征[58] 以及LE特征[59] 。

(a)

(b)

图 4-2 DML和HDML比较。(a) WDRef验证集准确率。蓝线表示HDML在验证集比对信号(verification)的准确率；绿线表示HDML在验证集分类(classification)信号准确率; 橙线表示DML在验证集比对信号(verifiction)的准确率.(b) LFW准确率。

我们将WDRef数据集按照4:1的比例分割为训练数据集和验证数据集。其中训练集包含56641个人总计1378668对样本对，测试集包含15205个人总计82082对样本对。对于HDML训练模型，我们设置验证损失函数、分类损失函数和正则项系数分别为，同时我们设置学习率，dropout为0.5。

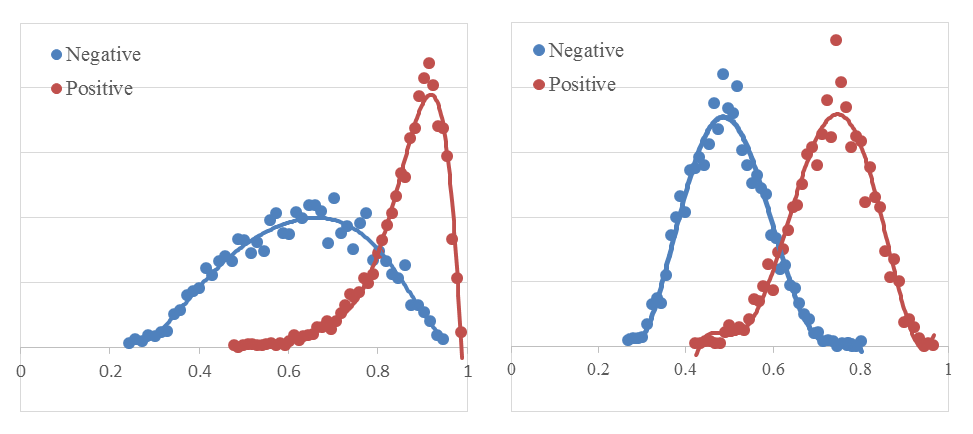


图 4-3 LFW数据集6000对测试样本分数分布：左图为DML算法结果；右图为HDML算法结果

根据图 4-2(a)(b)所示，对比蓝线和橙线很明显可以发现，引入分类信号的HDML算法在WDRef验证集和LFW测试集上均有较好的表现，由此可以证明分类信号对于提高深度测度学习泛化能力有很大的影响。此外，根据图 4-3所示，我们可以发现HDML算法在LFW数据集6000对测试样本中负样本分数分布比DML算法的分布更靠近0，同样可以证明分类信号对于增大类间距离有着显著的作用。

此外由于测度学习比对问题是一个典型的不平衡问题，因此如何对负样本对进行采样对训练结果通常会产生很大的影响。针对数据不平衡问题，我们在构成训练集时选择WDRef数据集可生成全部正样本对作为训练集的正样本数据，同时设置比例为1:1, 1:5, 1:10分别生成负样本对构成三个不同的训练集。如图 4-4所示，我们可以发现HDML算法对于不平衡训练集问题有很好的适应能力，实验结果表明即使采样时损失部分负样本信息，HDML算法同样可以求解到一个较优解，这是因为HDML算法中分类信号可以帮助增加类间距离，而在测度学习算法中负样本的主要作用同样是增加类间距离，因此分类信号在一定程度上可以代替采样损失的负样本。

(a)

(b)

(c)

图 4-4 正负样本不均匀训练集时HDML算法性能。(a)WDRef训练集准确率趋势图； (b) WDRef验证集准确率趋势图；(c)LFW准确率趋势图。

同时根据表4-4，HDML算法在LFW数据集上利用LBP特征获得了88.50%的准确率，而LE特征取得了90.53%的准确率，综合两种特征则将准确率提高到了91.57%，将Joint Bayesian的准确率提高了0.6%，也超过了其他算法。同时，对比LE特征利用DML和HDML算法的准确率我们也可以发现加入分类信号对深度测度学习学习更鲁棒的特征有显著的效果。

表 4-4 不同测度学习算法在LFW数据集上的精度(%)比较

|  |  |
| --- | --- |
| 方法 | 准确率(%) |
| LDML-MKNN, funneled | 87.50±0.40 |
| PLDA combined, funneled & aligned | 90.07±-.51 |
| Joint Bayesian combined | 90.90±1.48 |
| Sub-SML, funneled & aligned | 90.75±0.64 |
| LE+DML | 84.83±1.28 |
| LBP+HDML | 88.50±0.51 |
| LE+HDML | 90.53±0.53 |
| LE+LBP+HDML | 91.57±0.52 |

* 1. 结论

本章节主要结合深度学习思想提出了深度测度学习算法，并结合分类信号和验证信号提出了混合深度测度学习算法，利用分类信号的特性，抑制了验证问题中正负样本分布不均匀的影响，并且通过实验可以发现验证信号有助于减小类内距离而分类信号有助于增大类间距离，通过验证信号和分类信号共同优化可以通过数据学习一个鲁棒性较强的特征空间表征数据分布。

1. 测度学习在图像检索中的应用
   1. 图像检索概述

随着大数据和机器学习的发展，图像检索[60] [61] [62] 和图像标注[63] [64] 问题越来越得到众多科学家的关注。基于词袋模型(bag-of-visual-words)[65] , 越来越多的研究者将图像像素通过SIFT[66] 、GIST[67] 等特征提取算法提取低级特征，并通过Kmeans等聚类算法抽象高阶语义信息，实现图像检索系统。然而，词袋模型由于其计算复杂度和拓展性问题很难适用于大规模数据集。

为了克服词袋模型缺点，Jaakkola提出了费舍尔特征(fisher vector, FV)[68] 并Perronnin[69] 将其应用在了图像分类上。费舍尔特征在imagenet竞赛上被证明，相较于传统的词袋模型，其具有更优的泛化能力和表达能力。费舍尔特征可以看做是词袋模型的扩展和延伸，其不仅限于挖掘图像本身的像素信息并对其进行字典编码学习，同时还抽象了额外的一阶梯度信息进行像素分布的描述，增加了特征的鲁棒性。

尽管费舍尔特征在图像领域取得了广泛的应用，但是由于其特征向量维度过高，耗费存储空间和索引计算时间，因此很难适用于实际的图像检索系统。因此一个有效的降维方法就成为了亟待解决的问题。Perronnin等人基于二值化的方法[71] [72] 提出了可压缩的费舍尔特征[70] ，并将其应用到图像检索上。

本章节提出一种新的图像检索算法结构框架，利用denseSIFT算法提取低阶像素特征，使用费舍尔向量对低阶局部特征进行编码，然后通过最大边际测度学习方法一方面提高特征向量的表达性，另一方面有效的降低特征向量的维度。该图像检索系统结合了生成式模型和判决式模型的优点，可以有效的提高图像检索算法的性能。

* 1. 基于测度学习的图像检索算法框架

如图 5-1所示，本章节主要阐述基于测度学习的图像索引算法的框架。首先，根据输入图像提取Dense SIFT特征，然后利用高斯混合模型算法对局部Dense SIFT特征进行编码并构成费舍尔特征，然后利用测度学习对费舍尔特征进行降维和监督学习，并生成最后的图像特征表示。



图 5-1 基于测度学习的图像检索框架

* + 1. 费舍尔特征

首先我们定义局部特征集合，因此我们可以定义局部特征集合*X*的梯度特征为

(5-1)

其中表示*X*的概率密度函数。根据公式(5-1)我们可以发现梯度特征描述了参数的贡献程度，并且费舍尔特征向量的维度取决于参数的取值而不是训练集的大小*T*。因此我们可以定义核函数

(5-2)

其中核函数具有对称正定性质，因此我们可以将核函数表示为，并且可以表示归一化特征向量为

(5-3)

同时，对数似然函数可以表示为

(5-4)

其中对于，其概率分布可以表示为一个高斯混合模型(Gaussian Mixture Model, GMM)

(5-5)

因此其梯度特征我们可以表示为

(5-6)

(5-7)

(5-8)

其中, *N*为高斯混合模型中高斯分布的数量，为每个高斯分布的均值和方差，并且参数定义为

(5-9)

因此根据公式(5-6)，我们可以发现高斯混合模型的梯度特征维度仅与高斯混合模型中的高斯分布数量*N*决定，与训练样本数量无关。

核函数可以表示为通过表示为对角矩阵，其中

(5-10)

(5-11)

因此我们可以根据公式(5-3)、(5-7)、(5-8)利用核函数将高斯混合模型梯度归一化为

(5-12)

(5-13)

因此我们可以将费舍尔特征向量表示为

(5-14)

* + 1. 最大边际测度学习算法

对于图像检索问题，费舍尔特征向量是一个很好特征表示方法，但是由于其维度过高，不利于图像特征的存储以及大规模图像检索中的计算，因此需要采用降维的方法对费舍尔特征向量做处理，直接的降维手段是主成分分析方法 (Principal component analysis, PCA)，然而对于无监督学习的PCA方法，很难根据训练集获得本征特征，因此我们提出了一种线性映射方法用于将费舍尔特征向量压缩到一个低维空间中。

首先定义投影矩阵, 因此可以定义特征。使用该投影矩阵，我们可以将高维的费舍尔特征向量压缩到低维空间中，我们定义在一个图像子空间中的欧氏距离为

(5-15)

同时我们定义另一种距离表述

(5-16)

其中表示欧氏距离度量，表示相似度度量。

我们可以定义阈值判断两张图像和是否是同一类图像，因此根据距离表达公式和阈值可以定义约束

(5-17)

表 5-1 最大边际测度学习算法优化过程

|  |
| --- |
| 算法：最大边际测度学习优化算法 |
| **输入**：训练数据  **输出**：参数  **Begin**  参数初始化;      ;  ;  **for** to *n* **do**  **switch****do**  **case** 1:  score = ;  **if** score **then**  ();  ;  **break;**  **case** -1:  score = ;  **if** score **then**  ;  ;  **break;**  **return** ; |

其中表示图像不是一类，表示图像是一类。此时，我们可以根据公式(5-17)约束，构造关于参数的优化目标函数

(5-18)

根据公式(5-18)我们可以发现目标函数非凸，因此一个较优的初始值对获得具有泛化能力的参数至关重要。本文采用白化主成分分析(PCA-whitening)初始化*W*投影矩阵，并且使用随机梯度下降算法优化目标函数。

根据公式(5-15)可以计算出投影矩阵的更新公式为

(5-19)

其中，为更新参数的学习率。同时我们可以计算出公式(5-16)的参数*W*更新公式也为公式(5-19)，而相似度距离中参数*V*的更新公式为

(5-20)

其中。根据公式(5-19)和(5-20)我们可以发现，当训练的图像对满足约束条件公式(5-17)时，参数*W*与*V*并不产生变化。最大边际测度学习算法的详细流程如表 5-1所示。

* + 1. 图像检索系统框架

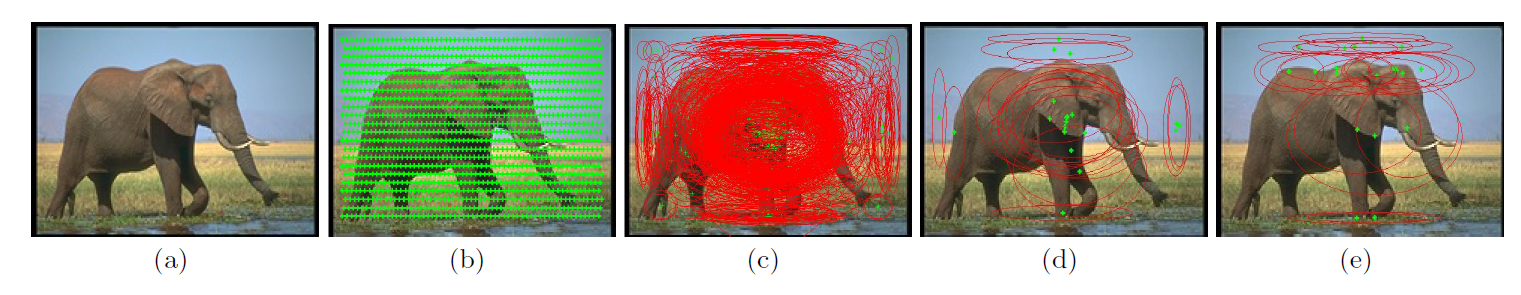


图 5-2 图像检索系统处理流程表示. (a) 输入图像；(b) Dense SIFT 特征；(c) GMM聚类模型；(d) 根据最大边际测度学习算法获得能量值最高的20个的高斯分布；(e)根据最大边际测度学习算法获取能量值最低的20个高斯分布。

本文在的图像块上提取Dense SIFT特征作为图像的局部描述子，遍历5个尺度空间，尺度变化的因数为，同时使用root-SITF[73] 表征每个图像块的特征。我们在的图像上提取局部特征描述子，因此对于每张图像最终可以获得33000个128维的特征向量，如图 5-2(b)所示。然后，根据PCA-SIFT[74] ，我们采用PCA算法将Dense SIFT特征从128维降到64维。尽管Dense SIFT是一个具有空间不变性的鲁棒特征，但是由于其在提取特征时只考虑局部梯度，并没有包含像素点全局的位置信息，因此我们引入位置信息和Dense SIFT特征共同构成图像局部特征描述子，其中表示64维的SITF特征向量，表示局部图像块的中心坐标，经过这种处理，我们获得33000个66维的特征向量。之后我们采用高斯混合模型对图像局部描述子进行编码。本文设置高斯混合模型中高斯分布的个数，因此我们可知最终经过GMM编码获得的费舍尔特征向量的维度为维()。最后，我们使用最大边际测度学习方法对高维费舍尔特征向量进行压缩并获得图像的特征描述子。

* 1. 实验

本文使用Corel-5K数据集评价基于最大边际测度学习算法的图像检索框架。Corel-5K数据集包含5000张图片，260个描述单词作为训练集，测试集包含499张图片。

我们使用平均准确率(mean average precision, mAP)作为衡量算法的指标。mAP指标指的是检索系统在全部相关文档上的性能指标，其检索出的相关文档和词条排名越靠前，mAP就越高。

本文提出的图像检索框架的处理流程如图 5-2所示，首先对图像提取Dense SIFT特征，然后利用GMM学习的参数对局部Dense SIFT特征编码并生成图像描述子费舍尔特征向量，最后根据*W*投影矩阵对特征进行降维。如图 5-2(c)所示，每个高斯分布可以描述图像的联合分布统计特性，而经过测度学习投影矩阵计算后，能量高的高斯分布表示的是图像局部区域中的重要信息(如图 5-2(d))所示。

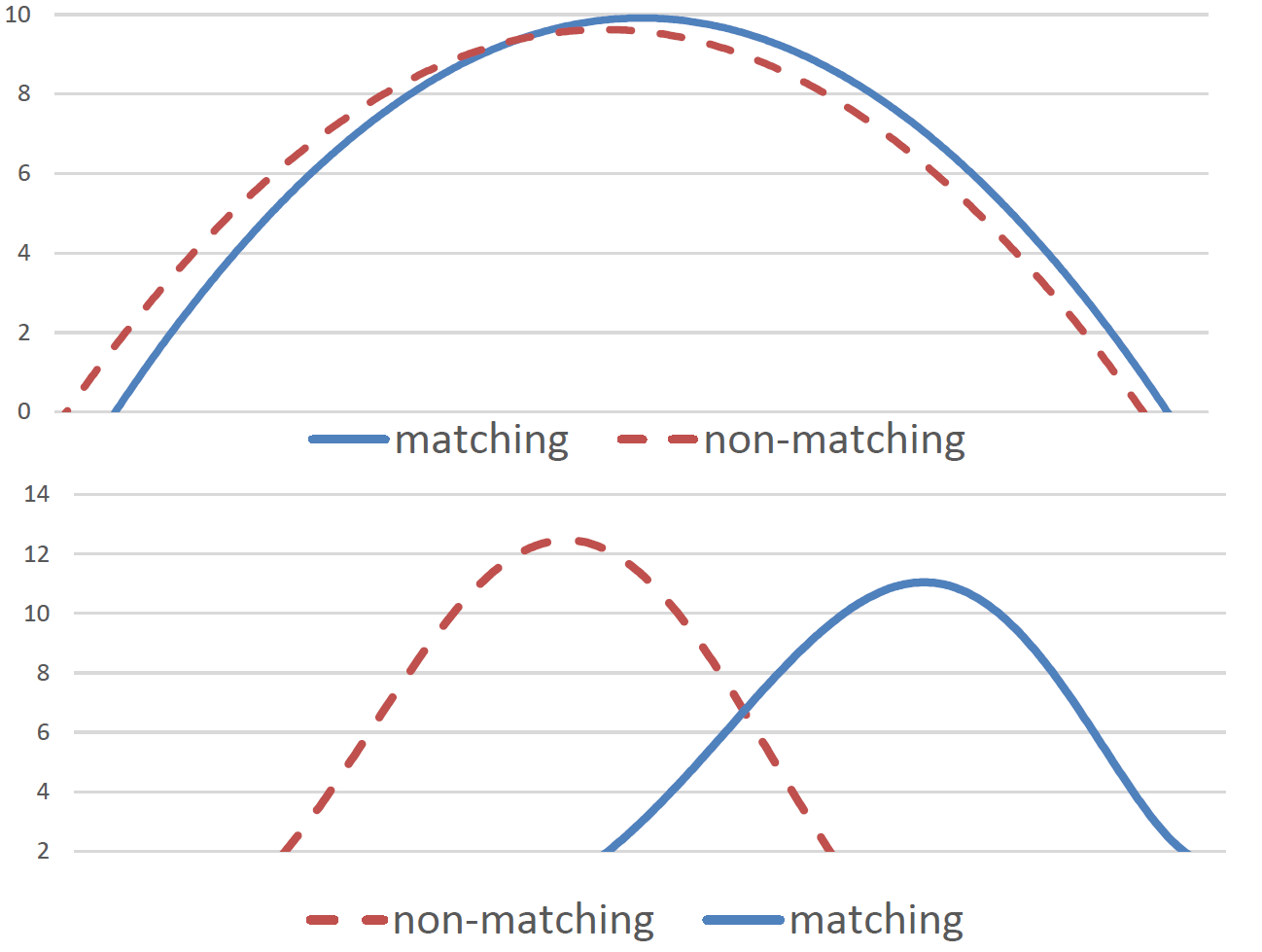


图 5-3 Corel-5K数据集相同类与不同类分数的分布曲线。上图表示费舍尔特征向量计算分数得到的结果；下图表示经过W投影矩阵映射后获得的特征向量计算的结果。

我们采用特征向量之间的欧式距离作为衡量图像对之间是否相似标准，欧氏距离越小代表两张图像越相似，欧氏距离越大表示两张图像越不相似。根据这样的准则，我们可以获取相似图像分数和不相似图像分数的分布曲线，如所示。上图是费舍尔特征向量作为图像特征描述子时计算的分数分布，下图是经过最大边界测度学习映射后的图像分布，可以很明显的发现，经过测度学习，我们寻找到了一个有效的投影矩阵将相似图像投影到同一个高维流行上，从而增加了图像检索系统的准确性，同时测度学习还具有降维的作用，可以减少存储和计算的代价。

图 5-4 Corel-5K数据集中不同类别图像的mAP结果。其中每个类别图像的mAP值由左到右一次为费舍尔特征结果、费舍尔特征PCA降维后结果、公式(5-15)作为目标函数距离度量结果、公式(5-16)作为目标函数距离度量的结果。

表 5-2 本文图像检索框架在Corel-5K数据集结果以及与其他算法的比较

|  |  |  |
| --- | --- | --- |
| method | Dim. | mAP |
| TagProp[64] |  | 0.4180 |
| MTH[75] | 82 | 0.4998 |
| MSD[76] | 72 | 0.5592 |
| CDH[77] | 108 | 0.5723 |
| Fisher Vector | 67584 | 0.5567 |
| PCA-whitening | 128 | 0.5579 |
| Metric | 64 | 0.5603 |
| Metric | 128 | 0.5666 |
| Metric-Simlarity | 2x128 | **0.6828** |

本文提出的图像检索框架的算法性能如图 5-4、所示。根据图 5-4我们可以发现对于Corel-5K数据集，最大边界测度学习对图像检索的性能有很大的提升，而根据表 5-3我们可以发现本文图像检索最终的mAP为68.28%，超过了CHD[77]方法11.05%，如图 5-5为图像检索系统效果演示图，我们可以发现最大边际测度学习算法对图像检索性能做出了巨大的提升。



图 5-5 最大边际测度学习算法图像检索结果示例；左一列为输入图像，右边五张图为排名前五的图像检索结果。其中红框表示检索类别正确，绿框表示检索不正确。

* 1. 结论

本章节提出了基于最大边际测度学习算法算法的图像检索系统，将高维费舍尔特征向量压缩到低维空间当中，并且在Corel-5K数据集上获得了68.28%的mAP，较为有效的提高了图像检索系统的性能。当然由于本文所使用的优化目标函数为非凸函数，因此目标函数参数的初始值对实验结果将产生很大的影响，如何有效的初始化优化参数，也是影响图像检索系统最终表现的重要方面之一。

1. 结论

本文主要阐述了测度学习在人脸识别、图像检索等领域的应用。

第三章节提出的近似相似度测度学习算法利用向量不等式简化余弦距离，构造有约束的测度学习目标函数，然后根据拉格朗日乘子法将约束优化转化为无约束优化并使用随机梯度下降方法对目标函数进行优化求解。对改善余弦相似度测度学习算法消耗内存、无法处理大规模数据等缺点有很大的作用。

第四章节提出的混合深度测度学习算法借鉴深度学习分布求精逐层学习的思想，结合分类信号和验证信号提高了测度学习特征的泛化能力，并且有效的解决了测度学习面临的正负样本不均匀问题。

第五章根据费舍尔特征向量和测度学习理论提出了一个有效的图像检索框架，利用最大边界测度学习对费舍尔特征向量进行压缩降维，不仅提高了图像检索精度，同时还降低了特征向量维度，解决了费舍尔特征在图像检索系统中特征存储代价和计算代价。

测度学习算法是机器学习领域中一种分析高维特征之间的相关性，寻找一个合适的特征空间更有效的表征高维特征的学习算法框架。测度学习通过目标函数优化，目的是尽可能减小目标问题中同一类目标之间的相对距离，并且增加不同类目标之间的相对距离，从而获得良好的分类效果。本文从人脸识别和图像检索两个计算机热门领域出发，证明了测度学习在机器学习和计算视觉领域的重要作用和价值。

参考文献

1. Cover T M, Hart P E. Nearest neighbor pattern classification[J]. Information Theory IEEE Transactions on, 1967, 13(1):21-27.
2. Lloyd S P. Least Squares Quantization in PCM's[J]. IEEE Transactions on Information Theory, 1982, 28(2):129-136.
3. Xing E P, Ng A Y, Jordan M I, et al. Distance Metric Learning, with Application to Clustering with Side-information[J]. Advances in Neural Information Processing Systems, 2002: 505-512.
4. Short R D, Fukunaga K. The optimal distance measure for nearest neighbor classification[C]. IEEE Transactions on Information Theory1981:622 - 627.
5. Hastie, T., & Tibshirani, R. Discriminant adaptive nearest neighbor classification[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1996, 18(6): 607-616.
6. Baxter, J., & Bartlett, P. The canonical distortion measure in feature space and 1-NN classification[C]. Advances in neural information processing systems, 1998: 245-251.
7. Weinberger K Q, Blitzer J, Saul L K. Distance metric learning for large margin nearest neighbor classification[C]. Advances in neural information processing systems, 2006:207-244.
8. Kilian Q Weinberger, John Blitzer, and Lawrence K Saul. Distance metric learning for large margin nearest neighbor classification[J]. The Journal of Machine Learning Research, 2009, 10:207-244.
9. Kilian QWeinberger and Lawrence K Saul. Fast solvers and efficient implementations for distance metric learning[C]. Proceedings of the 25th international conference on Machine learning, ACM, 2008: 1160-1167.
10. Huyen Do, Alexandros Kalousis, JunWang, and AdamWoznica. A metric learning perspective of SVM: on the relation of LMNN and SVM[C]. In Proceedings of the 15th International Conference on Artificial Intelligence and Statistics (AISTATS), 2012: 308–317.
11. Jason V Davis, Brian Kulis, Prateek Jain, Suvrit Sra, and Inderjit S Dhillon. Information-theoretic metric learning[C]. Proceedings of the 24th international conference on Machine learning, ACM, 2007: 209-216.
12. Guo-Jun Qi, Jinhui Tang, Zheng-Jun Zha, Tat-Seng Chua, and Hong-Jiang Zhang, An efficient sparse metric learning in high-dimensional space via l1-penalized log-determinant regularization[C]. Proceedings of the 26th Annual International Conference on Machine Learning, ACM, 2009: 841-848.
13. Fei-Fei Li and Pietro Perona. A Bayesian Hierarchical Model for Learning Natural Scene Categories[C]. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2005:524–531
14. Jerome Friedman, Trevor Hastie, and Robert Tibshirani. Sparse inverse covariance estimation with the graphical lasso[J]. Biostatistics 9, 2008, no. 3: 432-441.
15. Matthew Schultz and Thorsten Joachims. Learning a Distance Metric from Relative Comparisons[C]. In Advances in Neural Information Processing Systems (NIPS) 16, 2003.
16. Steven C. Hoi, Wei Liu, Michael R. Lyu, and Wei-Ying Ma. Learning Distance Metrics with Contextual Constraints for Image Retrieval[C]. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2006: 2072–2078.
17. Sumit Chopra, Raia Hadsell, and Yann LeCun. Learning a Similarity Metric Discriminatively with Application to Face Verification[C]. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2005: 539–546.
18. Brian McFee and Gert R Lanckriet, Metric learning to rank[C]. Proceedings of the 27th International Conference on Machine Learning (ICML-10), 2010: 775-782.
19. Daryl Lim, Gert Lanckriet, and Brian McFee, Robust structural metric learning[C]. Proceedings of The 30th International Conference on Machine Learning, 2013: 615-623.
20. Bengio Y, Lecun Y. Convolutional Networks for Images, Speech, and Time-Series[J]. The Handbook of Brain Theory & Neural, 1997.
21. Mensink T, Verbeek J, Perronnin F, et al. Metric Learning for Large Scale Image Classification: Generalizing to New Classes at Near-Zero Cost[M]. Computer Vision – ECCV 2012 Springer Berlin Heidelberg, 2012:488-501.
22. Andrea Frome, Yoram Singer, Fei Sha, and Jitendra Malik. Learning Globally-Consistent Local Distance Functions for Shape-Based Image Retrieval and Classification[C]. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2007: 1–8.
23. Nair V, Sellamanickam S, Mahajan D, et al. Learning hierarchical similarity metrics[C]. 2013 IEEE Conference on Computer Vision and Pattern RecognitionIEEE, 2012:2280-2287.
24. Guillaumin M, Verbeek J, Schmid C. Is that you? Metric learning approaches for face identification[J]. Proceedings, 2009, 30(2):498 - 505.
25. Xi Li, Chunhua Shen, Qinfeng Shi, Anthony Dick, and Anton van den Hengel. Nonsparse Linear Representations for Visual Tracking with Online Reservoir Metric Learning[C]. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2012: 1760–1767
26. Matthieu Guillaumin, Thomas Mensink, Jakob J. Verbeek, and Cordelia Schmid. TagProp: Discriminative metric learning in nearest neighbor models for image auto-annotation[C]. In Proceddings of the IEEE International Conference on Computer Vision (ICCV), 2009:309–316..
27. M. Ehsan Abbasnejad, Dhanesh Ramachandram, and Mandava Rajeswari. A survey of the state of the art in learning the kernels[J]. Knowledge and Information Systems (KAIS), 2012, 31 (2):193–221.
28. Mehmet G¨onen and Ethem Alpaydin. Multiple Kernel Learning Algorithms[J]. Journal of Machine Learning Research (JMLR), 2011, 12:2211–2268
29. Maaten L J P V D, Postma E O, Herik H J V D. Dimensionality reduction: A comparative review[J]. Journal of Machine Learning Research, 2007, 10.
30. Ali M. Qamar and Eric Gaussier. Online and Batch Learning of Generalized Cosine Similarities[C]. In Proceedings of the IEEE International Conference on Data Mining (ICDM), 2009: 926–931.
31. Hieu V. Nguyen and Li Bai. Cosine Similarity Metric Learning for Face Verification[C]. In Proceedings of the 10th Asian Conference on Computer Vision (ACCV), 2010: 709–720.
32. Wu, X., Shi, Z. G., & Liu, L. Quasi Cosine Similarity Metric Learning[C]. In Computer Vision-ACCV 2014 Workshops. Springer International Publishing, 2014: 194-205
33. Burges C J C. A tutorial on support vector machines for pattern recognition[C]. Data Mining and Knowledge Discovery 1998.
34. Bousquet O, Bottou L. The tradeoffs of large scale learning[J]. Advances in Neural Information Processing Systems, 2008:161--168.
35. Zhu C. L-BFGS-B - Fortran Subroutines for Large-Scale Bound Constrained Optimization[J]. Acm Trans.math.software, 1999, 23(4):550-560.
36. Huang G B, Mattar M, Berg T, et al. Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments[J]. Workshop on Faces in 'Real-Life' Images: Detection, Alignment, and Recognition, 2007.
37. Perronnin F, Dance C. Fisher Kernels on Visual Vocabularies for Image Categorization[C]. 2007 IEEE Conference on Computer Vision and Pattern RecognitionIEEE Computer Society, 2007:1-8.
38. Perronnin F, Sánchez J, Mensink T. Improving the Fisher Kernel for Large-Scale Image Classification[J]. Lecture Notes in Computer Science, 2010, 6314:143-156.
39. Perronnin F, Liu Y, Sanchez J, et al. Large-scale image retrieval with compressed Fisher vectors[C]// 2013 IEEE Conference on Computer Vision and Pattern RecognitionIEEE, 2010:3384-3391.
40. Simonyan, K., Parkhi, O.M., Vedaldi, A., Zisserman, A.: Fisher vector faces in the wild[C]. In: Proc. BMVC. Volume 1. 2013.
41. Wolf L, Hassner T, Taigman Y. Similarity Scores Based on Background Samples[M]. Computer Vision – ACCV 2009 Springer Berlin Heidelberg, 2010:88-97.
42. Guillaumin M, Verbeek J, Schmid C. Is that you? Metric learning approaches for face identification[J]. Proceedings, 2009, 30(2):498 - 505.
43. Ying, Y., & Li, P. Distance metric learning with eigenvalue optimization[J]. The Journal of Machine Learning Research, 2012, 13(1): 1-26.
44. Cao Q, Ying Y, Li P. Similarity Metric Learning for Face Recognition[C] 2013 IEEE International Conference on Computer Vision (ICCV) IEEE Computer Society, 2013:2408-2415.
45. Li P, Fu Y, Mohammed U, et al. Probabilistic Models for Inference about Identity[J]. Pattern Analysis & Machine Intelligence IEEE Transactions on, 2011, 34(1):144-157.
46. Chen D, Cao X, Wang L, et al. Bayesian Face Revisited: A Joint Formulation[M]. Computer Vision – ECCV 2012Springer Berlin Heidelberg, 2012:566-579.
47. 1 C C. Support-Vector Networks[J]. Machine Learning, 1995, volume 20(3):273-297(25).
48. Freedman D A. Statistical Models: Theory and Practice[J]. Cambridge University Press Cambridge, 2005, 48(2):xiv+442.
49. Cutler D R, Edwards T C, Beard K H, et al. Random Forests for classification in ecology. Ecology[J]. Ecology, 2007, 88(11):2783-2792.
50. Hinton G E, Salakhutdinov R R. Reducing the dimensionality of data with neural networks[J]. Science, 2006, 313(5786):504-507.
51. Hinton G E, Osindero S. A fast learning algorithm for deep belief nets[C]// Neural Computation 2006:2006.
52. Krizhevsky A, Sutskever I, Hinton G E. ImageNet Classification with Deep Convolutional Neural Networks[J]. Advances in Neural Information Processing Systems, 2012:2012.
53. Mignon A, Jurie F. PCCA: A new approach for distance learning from sparse pairwise constraints[C]. 2012 IEEE Conference on Computer Vision and Pattern RecognitionIEEE Computer Society, 2012:2666-2672.
54. Rumelhart D E, Hinton G E, Williams R J, et al. Learning internal representation by back-propagation of errors[J]. Nature, 1986, 323(1116):533-536.
55. Glorot X, Bengio Y. Understanding the difficulty of training deep feedforward neural networks[C], International Conference on Artificial Intelligence & Statistics Society for Artificial Intelligence & Statistics2010:249-256.
56. He H, Garcia E /. Learning from Imbalanced Data[J]. IEEE Transactions on Knowledge & Data Engineering, 2009, 21(9):1263-1284.
57. Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., & Darrell, T. Caffe: Convolutional architecture for fast feature embedding[C]. In Proceedings of the ACM International Conference on Multimedia, ACM, 2014:675-678.
58. Ojala T, Pietikäinen M, Mäenpää T. Multiresolution Gray-Scale and Rotation Invariant Texture Classification with Local Binary Patterns[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2002, 24(7):971-987.
59. Cao Z, Yin Q, Tang X, et al. Face recognition with learning-based descriptor[C]. Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference onIEEE, 2010:2707-2714.
60. Ladel C H, Daugelat S, Kaufmann S H E. Immune response to Mycobacterium bovis bacille Calmette Guérin infection in major histocompatibility complex class I- and II-deficient knock-out mice: contribution of CD4 and CD8 T cells to acquired resistance.[J]. European Journal of Immunology, 1995, 25(2):377–384.
61. Lin J, Duan L Y, Huang T, et al. Robust fisher codes for large scale image retrieval[C]. Acoustics, Speech, and Signal Processing, 1988. ICASSP-88., 1988 International Conference on. 2013:1513-1517.
62. Perronnin F, Liu Y, Sanchez J, et al. Large-scale image retrieval with compressed Fisher vectors[C]// 2013 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2010:3384-3391.
63. Lu Y, Zhang W, Zhang K, et al. Semantic context learning with large-scale weakly-labeled image set[C]. Proceedings of the 21st ACM international conference on Information and knowledge management. ACM, 2012:1859-1863.
64. Guillaumin M, Mensink T, Verbeek J, et al. TagProp: Discriminative metric learning in nearest neighbor models for image auto-annotation[C]// Computer Vision, 2009 IEEE 12th International Conference on. IEEE, 2009:309-316.
65. Sivic J, Zisserman A. Video Google: A Text Retrieval Approach to Object Matching in Videos[C]. null. IEEE Computer Society, 2003:1470.
66. Lowe D G. Distinctive Image Features from Scale-Invariant Keypoints[C]. International Journal of Computer Vision. 2004:91-110.
67. Oliva A, Torralba A. Modeling the Shape of the Scene: A Holistic Representation of the Spatial Envelope[J]. International Journal of Computer Vision, 2001, 42(3):145-175.
68. Jaakkola T S, Haussler D. Exploiting generative models in discriminative classiers[J]. Advances in Neural Information Processing Systems, 1999.
69. Perronnin F, Dance C. Fisher Kernels on Visual Vocabularies for Image Categorization[C]. 2007 IEEE Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, 2007:1-8.
70. Perronnin F, Liu Y, Sanchez J, et al. Large-scale image retrieval with compressed Fisher vectors[C]. 2013 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2010:3384-3391.
71. Weiss Y, Torralba A, Fergus R. Spectral hashing[J]. Proc Nips, 2008, 282(3):1753-1760.
72. Charikar M S. Similarity estimation techniques from rounding algorithms[J]. Proc of Stoc, 2002:380-388.
73. Arandjelović R, Zisserman A. Three things everyone should know to improve object retrieval[C]. Proceedings / CVPR, IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2012:2911-2918.
74. Ke Y, Sukthankar R. PCA-SIFT: a more distinctive representation for local image descriptors[C]. Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on. IEEE, 2004:II-506-II-513 Vol.2.
75. Liu G H, Zhang L, Hou Y K, et al. Image retrieval based on multi-texton histogram[J]. Pattern Recognition, 2010, 43(7):2380–2389.
76. Liu G H, Li Z Y, Zhang L, et al. Image retrieval based on micro-structure descriptor[J]. Pattern Recognition, 2011, 44(9):2123–2133.
77. Liu, G H, Jing-Yu Yang. Content-based image retrieval using color difference histogram[J]. Pattern Recognition 2013, 46(1): 188-198.

作者简历及在学研究成果

1. 作者入学前简历

|  |  |  |
| --- | --- | --- |
| 起止年月 | 学习或工作单位 | 备注 |
| 年月至年月 | 在攻读 |  |
| 年月至年月 | 在从事的工作 |  |
|  |  |  |
|  |  |  |

**提示信息：**硕士入学前简历，根据自己情况填写，备注中填写职务等。

不用此信息时，删除此框。

（鼠标移到此框四边，鼠标变为十字箭头，点击边框选中此框，然后按Del删除）

1. 在学期间从事的科研工作

（应注明课题名称、参加身份、通过时间、通过方式、评定单位等）。不用此信息时，请删除本段。

1. 在学期间所获的科研奖励

（应注明奖励名称、授奖单位、授奖时间等，请填写科研方面奖励，请勿填写其他奖励信息，如不得填写三好研究生等奖励信息。）不用此信息时，请删除本段。

1. 在学期间发表的论文

（应按照参考文献的格式来填写，包括编号。并在后面依次标明以下事项，各项之间用“．”分隔：1）标明“已发表”或“已录用”；2）是否“SCI/EI/STP/CSSCI刊源”；3）是否被“SCI/EI/STP/CSSCI检索”；4）检索号。第2、3项请标明具体检索名称）。不用此信息时，请删除本段。

盲审论文，请隐去所有可能影响盲审结果的信息，诸如作者姓名、导师姓名、作者学号等，具体隐去方法请详见“学位论文工具栏”中的“常用工具”中的“盲审论文制作方法”，或查看本文档最后的论文编制指南中的“13盲审论文制作方法”。另外在此处，研究成果中论文作者的发表文章列表中应隐去所有作者的名字，只标明论文作者是第几作者，具体如“[第二作者]．论文名称．……”。不用此信息时，请删除本段。

独创性说明

本人郑重声明：所呈交的论文是我个人在导师指导下进行的研究工作及取得研究成果。尽我所知，除了文中特别加以标注和致谢的地方外，论文中不包含其他人已经发表或撰写的研究成果，也不包含为获得北京科技大学或其他教育机构的学位或证书所使用过的材料。与我一同工作的同志对本研究所做的任何贡献均已在论文中做了明确的说明并表示了谢意。

本页的签名均应当是亲笔签名，不用输入，不用此信息时，删除此框。

签名：\_\_\_\_\_\_\_\_\_\_\_ 日期：\_\_\_\_\_\_\_\_\_\_\_\_

关于论文使用授权的说明

本人完全了解北京科技大学有关保留、使用学位论文的规定，即：学校有权保留送交论文的复印件，允许论文被查阅和借阅；学校可以公布论文的全部或部分内容，可以采用影印、缩印或其他复制手段保存论文。

**（保密的论文在解密后应遵循此规定）**

签名：\_\_\_\_\_\_\_\_\_\_\_ 导师签名：\_\_\_\_\_\_\_\_\_\_\_ 日期：\_\_\_\_\_\_\_\_\_\_\_\_

学位论文数据集

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **关键词\*** | **密级\*** | **中图分类号\*** | **UDC** | **论文资助** |
|  |  |  |  |  |
| **学位授予单位名称\*** | | **学位授予单位代码\*** | **学位类别\*** | **学位级别\*** |
| 北京科技大学  **提示：**  送交正式论文时，本页所有带星号的项目必须填写完整并确保正确。否则，可能重新制作并提交正式论文。  不需此提示时，请删除此提示框。  **学位类别：**请填写哲学、经济学、法学、教育学、文学、理学、工学、管理学等，详细请见[USTB硕士学位论文规范及论文制作指南.exe](file:///C:\Users\a\AppData\Roaming\Microsoft\Word\USTB硕士学位论文规范及论文制作指南.exe)中的我校学科专业目录。  **学位级别：**请填写博士、硕士。  不用此提示时，请删除此信息提示框。  **提示：**  在答辩后给学校上交最终正式论文时，请删除本提示框并填写本页。  在答辩及答辩前，请保持本页及本框不变；建议将本页其他提示框删除。  填写此页时，请删除此信息提示框。 | | 10008 |  |  |
| **论文题名\*** | | **并列题名** | | **论文语种\*** |
|  | |  | |  |
| **作者姓名\*** |  | | **学号\*** |  |
| **培养单位名称\*** | | **培养单位代码\*** | **培养单位地址** | **邮编** |
| 北京科技大学 | | 10008 | 北京市海淀区学院路30号 | 100083 |
| **学科专业\*** | | **研究方向\*** | **学制\*** | **学位授予年\*** |
|  | |  |  |  |
| **论文提交日期\*** |  | | | |
| **导师姓名\*** |  | | **职称\*** |  |
| **评阅人** | **答辩委员会主席\*** | | **答辩委员会成员** | |
|  |  | |  | |
| **电子版论文提交格式** 文本（ ） 图像（ ） 视频（ ） 音频（ ） 多媒体（ ） 其他（ ） **推荐格式：**application/msword；application/pdf | | | | |
| **电子版论文出版（发布）者** | | **电子版论文出版（发布）地** | | **权限声明** |
|  | |  | |  |
| **论文总页数\*** |  | | | |
| 共33项，其中带\*为必填数据，为22项。 | | | | |

**提示信息：**带\*为必填数据，必填数据共22项。所填数据的格式为“宋体、五号”

不用此信息时，删除此框。

（鼠标移到此框四边，鼠标变为十字箭头，点击边框选中此框，然后按Del删除）

1. <https://archive.ics.uci.edu/ml/datasets/Iris> [↑](#footnote-ref-1)
2. <https://archive.ics.uci.edu/ml/datasets/Ionosphere> [↑](#footnote-ref-2)
3. <https://archive.ics.uci.edu/ml/datasets/Wine> [↑](#footnote-ref-3)
4. <http://vis-www.cs.umass.edu/lfw/> [↑](#footnote-ref-4)
5. <https://archive.ics.uci.edu/ml/datasets/Diabetes> [↑](#footnote-ref-5)
6. <https://archive.ics.uci.edu/ml/datasets/Ionosphere> [↑](#footnote-ref-6)
7. <https://archive.ics.uci.edu/ml/datasets/Connectionist+Bench+%28Sonar%2C+Mines+vs.+Rocks%29> [↑](#footnote-ref-7)
8. <http://home.ustc.edu.cn/chendong/JointBayesian/index.html> [↑](#footnote-ref-8)