“十三五”科学技术专著丛书

基于多视角学习的图像语义分析技术

薛 哲 编著

Image Semantic Analysis Technology Based on Multi-View Learning

北京邮电大学出版社
www.buptpress.com

“十三五”科学技术专著丛书

基于多视角学习的图像语义分析技术

薛　哲　编著

北京邮电大学出版社
www. buptpress. com

内 容 简 介

随着互联网技术的快速发展和数字移动设备的广泛普及，大量的图像数据在互联网平台上出现并传播。如何准确地识别和理解这些真实场景下具有复杂视觉特性的图像充满挑战。本书围绕多视角学习中的融合与表示两个关键问题，从多视角图像聚类、多视角图像降维、多视角图像标注和基于深度模型的多视角学习等几个方面深入介绍目前主流的基于多视角学习的图像语义分析技术。针对多视角图像无监督聚类的任务，提出基于分组敏感多视角融合的图像聚类方法；针对多视角图像降维任务，提出基于双阶段子空间学习的多视角降维方法；针对多视角图像标注任务，提出图像多视角表示与标注的联合学习方法；针对基于深度模型的多视角学习任务，提出基于深度低秩子空间集成学习的多视角图像聚类方法。本书还通过在不同数据集上的对比实验验证了本书所提方法的有效性。

本书可作为高等院校计算机、多媒体等相关专业本科生和研究生的教材，也可作为相关领域的科研与工程技术人员的重要参考书。

图书在版编目(CIP)数据

基于多视角学习的图像语义分析技术 / 薛哲编著. -- 北京 : 北京邮电大学出版社，2019.8
ISBN 978-7-5635-5782-0

Ⅰ. ①基… Ⅱ. ①薛… Ⅲ. ①图象分析－语义分析－研究 Ⅳ. ①TP391.41

中国版本图书馆 CIP 数据核字（2019）第 162493 号

书　　名：基于多视角学习的图像语义分析技术
作　　者：薛　哲
责任编辑：满志文　穆菁菁
出版发行：北京邮电大学出版社
社　　址：北京市海淀区西土城路 10 号（邮编：100876）
发 行 部：电话：010-62282185　传真：010-62283578
E-mail：publish@bupt.edu.cn
经　　销：各地新华书店
印　　刷：北京玺诚印务有限公司
开　　本：787 mm×1 092 mm　1/16
印　　张：7.25
字　　数：178 千字
版　　次：2019 年 8 月第 1 版　2019 年 8 月第 1 次印刷

ISBN 978-7-5635-5782-0　　定　价：38.00 元

前　言

人工智能技术是当今最为活跃、发展最快的技术领域之一。2017 年 7 月，国务院发布了《新一代人工智能发展规划》，可见我国已经把发展人工智能上升至国家战略的高度。在国家战略支持和市场需求驱动的双重引力下，我国人工智能与实体经济融合速度不断加快，社会效益和经济效益明显。随着移动互联网的快速普及，每天网络上都会产生大量图像、视频、音频等多媒体数据。为了基于多媒体数据的内容提供精准的数据查询、搜索、推荐等服务，以互联网和人工智能技术为基础的图像语义分析技术应运而生，并受到了研究者广泛的关注和重视。本书介绍如何借助人工智能的技术手段，以实现高效、精准、智能地识别和理解互联网图像数据。

在人工智能的多种技术中，多视角学习是非常重要的一种技术手段。所谓多视角，是指数据可以通过多种特征或多种途径来描述，每种特征或途径就是描述数据的一个视角。多视角学习技术能够充分利用不同视角信息之间的互补性，并使用最优的方式整合多种信息，从而获得更好的学习性能。正是由于多视角学习技术与传统人工智能技术相比具有独特的优势，基于多视角学习的人工智能技术取得了快速的发展。为此，本书从多视角图像聚类、多视角图像降维、多视角图像标注和基于深度模型的多视角学习等几个方面详细介绍目前主流的基于多视角学习的图像语义分析技术。本书的组织结构如下：

第 1 章为绪论部分，主要介绍基于多视角学习的图像语义分析技术的研究背景、存在的问题和本书的主要内容。

第 2 章详细介绍多视角学习的研究现状、多视角学习的基本准则，并按照分类依次介绍了相关的研究工作。

第 3 章提出基于分组敏感多视角融合的图像聚类方法。为了克服基于全局融合策略的多视角聚类方法的不足，本章提出一种利用图像分组的局部学习策略来对多视角融合权重进行学习。通过迭代优化图像分组、融合权重以及聚类结果，该方法能够得到更准确的多视角融合结果，图像聚类的性能也可以进一步提升。

第 4 章提出基于双阶段子空间学习的多视角降维方法。该方法是一种无监督的多视角数据降维方法，能够有效地利用多视角的互补信息以及结构信息，通过所提出的双阶段学习机制，准确地把多视角信息保留在低维表示中。在相关图像数据集上的实验结果表明，该方法得到的低维表示更具有判别力，验证了所提方法的有效性。

第 5 章提出图像多视角表示与标注的联合学习方法。该方法是一种针对多视角数据的图像标注方法，它能够利用多视角表示学习以及图像标注之间的相关性，同时进行图像表示以及标签预测学习。实验结果表明，该方法能够学习得到合适的图像表示结果并可以有效提升图像标注的性能。

第6章提出基于深度低秩子空间集成学习的多视角图像聚类方法。该方法是一种基于深度学习模型的多视角图像数据聚类方法，它能够提取图像数据多层次的聚类结构信息，并通过集成学习方法将多个视角、多个层次的聚类结构进行有效融合，获得鲁棒、准确的多视角融合结果。实验结果表明，该方法能够获得更具有判别力的图像低维嵌入表示，并进一步提升了图像聚类的性能。

第7章对全书的工作进行总结，并对未来本领域的研究方向进行展望。

我们希望读者通过阅读本书，能够对书中介绍的人工智能技术形成初步的认识，对基于多视角学习的图像语义分析技术存在的问题以及最新的进展有大致的了解，并掌握如何将人工智能技术应用于图像语义分析的基本原理。

本书的编写得到了国家自然科学基金（项目编号：61802028，61532006，61772083）和智能通信软件与多媒体北京市重点实验室主任基金（项目编号：ITSM20180102）的资助。以及在本书的编写过程中得到了北京邮电大学出版社姚顺编辑和有研科技集团邵旭老师的支持和帮助，特此鸣谢。

本书由北京邮电大学计算机学院教师薛哲编著完成。由于作者的水平有限，加之技术和相关学术领域的不断变化和更新，书中的不足之处在所难免，恳请各位专家和读者批评、指正。

作　者

目　录

第1章 绪 论

1.1 研究背景

在人类所获取的外界信息中，大约有60%的信息来自于视觉[1]。由于图像具有生动直观的特点，一幅图像通常可以给人们带来非常丰富的信息，因此图像数据是人们记录、传递信息的重要手段。近年来，随着互联网技术的快速发展和数字移动设备特别是手机的迅速普及，网络用户可以很方便地通过手机上传和分享图像、视频等数据。同时，多种多样的社交媒体平台如Facebook、QQ空间、微信以及微博等迅速发展，每天都有大量的多媒体数据在互联网上出现并传播。图像作为多媒体数据的重要组成部分，成为人们展示日常生活状态以及获取外界信息的主要形式。据统计，在著名的社交网站Facebook上，用户每天上传的图片数目约三亿张[2]；而在图像分享平台Flickr上，2016年平均每月上传的图片数目超过五千万张[3]。互联网的图像数据呈现爆炸式增长，这就需要研究者们提出有效的分析处理图像数据的方法，从而在大量的图像数据中及时准确地获取感兴趣的知识和信息。然而用户在拍摄照片时由于拍照角度、光照变化以及照相设备等情况各不相同，会对照片质量造成很多不利影响。例如，光照、模糊的问题会造成图像成像不清晰；遮挡会影响物体轮廓的提取并干扰图像类别的推断；拍照角度的不同会带来物体背景剧烈的变化。对研究者来说，准确识别和理解这些在真实场景下具有复杂视觉特性的图像是一项充满挑战的任务。

为了更加准确地分析并处理数据，通常从多种角度对数据进行观察和表示。例如，一个互联网的网页可以被其包含的文字、图像以及超链接等信息表示；电影、电视等数据则可以通过视频、音频和字幕等信息表示；一场体育比赛中，某个运动员的画面可以由多个角度的摄像机同时捕捉；人类对外部事物的感知是通过视觉、听觉、嗅觉、味觉、触觉等多种渠道完成的，通过对不同类型的感知，让我们从多种角度来认识事物。“横看成岭侧成峰，远近高低各不同”，在不同的角度上对同一物体进行观察，可能会得到不同的现象和结论。单一角度的观察只能得到在该角度下的一些特定信息，很难掌握原数据的全部特性和规律。这就需要我们能够充分利用多个视角的信息，从而更加完整准确地理解事物的本质。

在图像语义分析中的一个基本问题就是如何描述和表示一幅图像的内容。为此，研究者们提出了多种视觉描述子，如SIFT[4]、HOG[5]、LBP[6]等特征。每种视觉描述子都可以看作在某个角度上对原始图像数据的一种描述。例如，当使用颜色直方图特征时，图像的颜色信息就能够很好地被提取出来；使用HOG特征时，图像的形状轮廓信息就能够有很好的描述；使用SIFT特征时，图像的局部特征点就能够被很好地表示和利用。我们把图像数据

的每种视觉特征当作描述它的一个视角，同时利用多种视觉特征就构成了多视角数据(Multi-View Data)。一些研究工作表明[7-11]，仅仅利用单一视角特征在表示图像时的描述能力有限，不能很好地表达出图像的内容和语义信息。如果利用多个视角特征，让各个视角的特征能够相互补充、相互促进，图像的高层语义信息就能够被更好地发掘出来。

不同的视角通常具有不同的物理意义，在各个视角的表示空间中，数据的统计特性和分布规律也不一样，因此不同视角之间是不可以直接进行比较的。例如，在使用 SIFT-BoW 特征来描述图像时，它的向量空间中的每一维表示的是该图像包含某个视觉单词的情况。而在使用 HOG 特征时，其向量空间的某个维度则表示的是图像的梯度直方图在某个方向上的分量大小。一些传统的机器学习方法，如支持向量机、逻辑斯蒂回归等大多是针对单视角数据的，在处理多视角数据时，需要把不同视角的特征拼接为一个新的特征才能进行处理。但是，这种简单的拼接操作使得数据维度变大，很容易造成模型的过拟合问题，而且其物理意义也不明确。为了克服以上的问题，多视角学习(Multi-View Learning)考虑到了各个视角不同的物理特性，能够在复杂、异质的多视角特征中找到有效的信息成分，并能够充分利用多个视角包含的丰富信息，使得各视角之间可以互相补充，进一步提升学习任务的性能。因此，多视角学习已经成为人工智能、机器学习等领域的研究热点，并受到研究者越来越多的重视。

1.2 存在的问题

近年来，随着研究者对多视角学习方法研究的不断深入，多视角学习领域取得了较大的进展和丰富的成果[12-16]，但仍然存在着一些关键问题亟待解决。首先，为了有效利用多个视角的信息，我们需要对多个视角的数据进行融合。然而图像数据的视觉特性通常是复杂多变的，如何根据图像自身的特点来准确地进行多视角信息融合具有重要意义。其次，由于多视角数据通常维度较高、物理意义和统计特性不同，直接使用原始的多视角特征存在着诸多问题。因此，我们需要对多视角数据学习出适合于后续任务的数据表示，从而更有效地对图像内容进行分析与理解。总之，在图像语义分析中，多视角学习面临的挑战主要体现在多视角数据“融合”与“表示”这两个问题上。下面我们就详细介绍这两个问题。

1. 多视角数据融合

融合多视角数据的难点在于如何对各个数据学习出合适的融合权重。在图像聚类的任务中，现有的多视角融合方法通常基于多核学习或谱嵌入学习的框架，对全部数据学习统一的融合权重。在进行多视角融合时，各个视角的重要性对所有数据是相同的。然而这种方式忽略了不同的数据具有不同的视觉特点和统计特性。例如，在区分“苹果”和“梨”这两个类别的图像时，使用颜色特征可以得到很好的判别效果，那么颜色视角的权重应该更高；而在区分“香蕉”和“梨”时，形状特征比颜色特征更具有判别力，此时形状视角应该获得比颜色视角更高的权重。若采用全局一致的融合权重，则不能根据图像数据自身的特性来选择合适的视角，也就无法对全部样本获得准确的多视角融合结果。此外，在多视角信息融合的过程中，多视角数据包含的一些不准确的描述和噪声会影响最终融合结果的准确性，因此多视角融合方法应该具备一定的鲁棒性，从而有效地克服噪声带来的影响。

2. 多视角数据表示

多视角数据的不同视角之间具有不同的物理意义和统计特性，并且构成多视角数据的特征通常是成千上万维的。在使用多视角数据时，若简单拼接多视角特征矩阵，则会面临物理意义不明确以及维度过高等问题。为了有效地使用高维、异质的多视角数据，对其学习出合适的数据表示是非常必要的。由于不同视角的物理意义不同，在多视角表示学习时需要考虑到不同视角之间的不可比较性，才能让多视角信息更准确地保留在所学的表示空间中。另外，现有的多视角表示学习方法通常忽视了表示学习与后续任务之间相互关联、耦合的特点。例如，在图像标注任务中，传统方法在进行标签预测时没有考虑到多视角表示学习与标签预测之间的相关性，这两个任务的学习过程是相互独立进行的。因此所学的多视角表示不能很好地用于后续图像标注任务中，所得到的标注性能也会受到限制。

多视角数据的融合与表示这两个问题是紧密关联的：一方面，为了得到有效的多视角数据表示，我们必须通过多视角融合把各个视角的信息保留在所学的数据表示中；另一方面，多视角表示是多视角融合结果的一种外在表达形式，经过融合后的多视角信息需要进行有效的表示才能用于后续的学习任务。总之，多视角学习的方法需要解决融合与表示这两个挑战性的问题，才能更有效地利用多视角数据，进一步提升多视角学习任务的性能。针对以上两个挑战性的问题，我们将介绍本书的主要内容。

1.3 本书主要内容

本书在图像语义分析的应用背景下，围绕多视角学习中的融合与表示两个关键问题，分别对多视角图像聚类、多视角图像降维、多视角图像标注和基于深度模型的多视角学习四个任务开展介绍和研究，提出了相应的学习模型和优化方法。在多视角图像聚类的任务中，本书提出了一种局部敏感融合的方法来克服全局权重估计方法所造成的融合不准确等问题，获得了更准确的融合结果并进一步提升了图像聚类性能。在多视角图像降维的任务中，我们利用双阶段矩阵分解的方法来逐级地对多视角数据进行成分提取与融合，从而获得了有效的多视角低维表示。在多视角图像标注任务中，我们利用子空间学习的方法来同时进行多视角表示学习和图像标注学习，充分利用了两个任务之间的相关性，在得到多视角表示结果的同时也提升了图像标注的性能。在基于深度模型的多视角学习任务中，本书提出了深度低秩子空间集成学习模型，通过在深度矩阵分解的隐空间中学习低秩子空间，然后利用集成学习方法融合多个子空间，最终实现多视角图像聚类。从解决问题的角度来看，多视角图像聚类的方法主要是针对多视角融合问题提出的，而多视角图像降维和多视角图像标注的方法主要是针对多视角表示问题提出的，同时也涉及一些多视角融合的技术。总之，通过对以上四个任务开展介绍与研究，本书的相关内容能够帮助我们有效地分析和处理多视角数据，从而实现对多视角图像数据的一致性表达和语义分析。本书的主要内容总结如下。

1. 基于分组敏感多视角融合的图像聚类方法研究

图像聚类就是对无类别信息的图像样本划分图像类别，从而让计算机能够自动地按照视觉内容对图像进行分门别类，更好地组织管理图像数据。聚类时为了得到图像之间的相似度，需要对图像的多视角信息进行融合。然而传统图像聚类方法在多视角融合时通常对

整个样本空间采用统一的融合权重，无法得到准确的融合结果。为此，本书提出了一种基于分组敏感多视角融合的图像聚类方法。本方法首先将图像数据分成若干图像组，使每个组内的图像具有相似的视觉特性，可以使用同一组融合权重进行多视角融合。然后提出了两种多视角融合权重的学习准则，对每个图像组的融合权重进行学习。为了得到合适的融合权重以及图像聚类结果，本书提出了一种交替迭代的优化策略。首先给定初始图像分组和融合权重，可以得到图像聚类结果。然后通过图像聚类后的类别信息，进一步优化更新图像分组和融合权重。通过以上两个步骤的交替迭代进行不断优化，就能够得到最终的多视角融合结果以及图像聚类结果。相比于传统的全局融合方法，本书提出的局部融合方法能够更灵活地为不同特性的图像数据估计融合权重，从而获得更准确的多视角融合结果。本书所引入的图像“组”是一种介于图像类别与单个图像之间的子类结构，它能够有效地应对真实图像数据的分布特性复杂和类内差异性较大等问题。此外，相比于为每个图像样本都进行权重学习的局部融合方式，分组融合策略能够有效地降低算法的复杂度，并获得更准确的多视角融合结果。

2. 基于双阶段子空间学习的多视角降维方法研究

多视角的数据通常由成千上万维的特征构成，然而模型在使用高维特征时计算复杂，并且很容易面临“维度灾难”的问题，因此有必要对高维的多视角数据提出有效的降维学习方法。此外，由于不同视角的数据具有不同的物理意义，无法直接进行比较，可以把它们先表示到一个能够相互比较的空间后再进行降维学习。基于以上考虑，本书提出一种基于双阶段子空间学习的多视角数据降维方法。每个阶段的子空间学习都是通过非负矩阵分解的方法实现的。在第一阶段的学习中，我们将各个视角的数据进行矩阵分解，然后得到各自的系数矩阵和基矩阵。在非负的表示空间中，各视角的系数矩阵就具有了统一的物理意义，从而是可以相互比较的。同时，我们对各视角的独立成分和共享成分加入相应的约束条件，从而让不同视角的信息能够相互补充。在第二阶段的学习中，我们把第一阶段的学习结果与多视角结构信息通过联合矩阵分解的方法学习得到最终的低维表示。为了克服多视角数据噪声的影响，我们使用 $\ell_{2,1}$-范数作为损失函数，增强了方法的鲁棒性。本书所提出的双阶段学习机制是一种灵活的学习策略，通过对不同阶段设计相应的目标函数，本方法能够有效地应对多视角数据物理含义不同以及数据噪声等问题，从而更准确地将多视角信息保留在低维表示中。

3. 图像多视角表示与标注的联合学习方法研究

图像标注就是对没有标签信息的图像分配与之内容相关的标签，从而让人们能够更准确的检索、组织和管理图像数据。然而由于语义鸿沟的存在，基于原始的多视角特征很难准确地预测图像的高层语义信息。我们希望能够得到一个更有利于标注任务的图像表示，而不是基于原始的多视角特征直接进行标签预测。为此，本书提出了联合学习图像的多视角表示与标签预测的方法。多视角表示学习是利用子空间学习的方法完成的，通过在多个视角的空间中学出一个共享的子空间，多视角数据就拥有了一致的表示形式。为了让所学的子空间充分地保留多视角信息，本书提出利用 softmax 激活函数来保留各个视角的结构信息。我们还把图像的语义信息嵌入到所学的子空间中，从而让所学的图像表示更具有判别性。此外，我们的方法考虑到了多视角表示学习与标签预测这两个任务之间的相关性，把它们各自的目标函数放到了一个统一的优化框架中。这样，给定标签预测器，我们就可以进一步提升子空间的判别性，使之更好地预测图像的标签信息。同时，给定一个较好的图像表

示，也会进一步提升标签预测器的预测性能。这样两个任务就可以相互促进，从而得到更优的标签预测结果。

4. 基于深度低秩子空间集成学习的多视角图像聚类方法研究

真实世界的图像数据由于光照、角度、遮挡和多属性等因素的影响，通常它们的分布特性复杂，并且类别结构呈现出多层次的特点。传统的浅层学习模型非线性学习能力较弱，因此它们无法有效获取图像数据本质的类别结构。鉴于此，本书提出一种基于深度低秩子空间集成学习的多视角图像聚类方法。首先，使用深度矩阵分解模型对多视角图像数据学习多层次的隐空间，每个隐空间有效包含了图像数据某个属性的特征表达。然后，利用低秩子空间学习模型，对每个层次的隐空间进一步提取出低秩的数据相似性矩阵。随后，利用集成学习的方法将多个层次、多个视角下的低秩子空间进行集成与融合。提出组稀疏正则范数对融合系数进行惩罚，可有效提升准确视角的重要性，抑制不准确视角对最终结果的干扰。对融合后的结果使用子空间嵌入方法进一步获得数据的低维表达。最后，使用谱聚类方法对所学的低维表达进行数据聚类，获得多视角数据的聚类结果。本书所提出的方法能够有效利用深度模型不同层次的类别结构信息，并通过集成学习的方式可以有效将不同视角、不同层次的聚类结构信息进行有效融合，从而获得更加鲁棒、准确的多视角融合结果。

图 1.1 所示的是本书各章的组织结构。

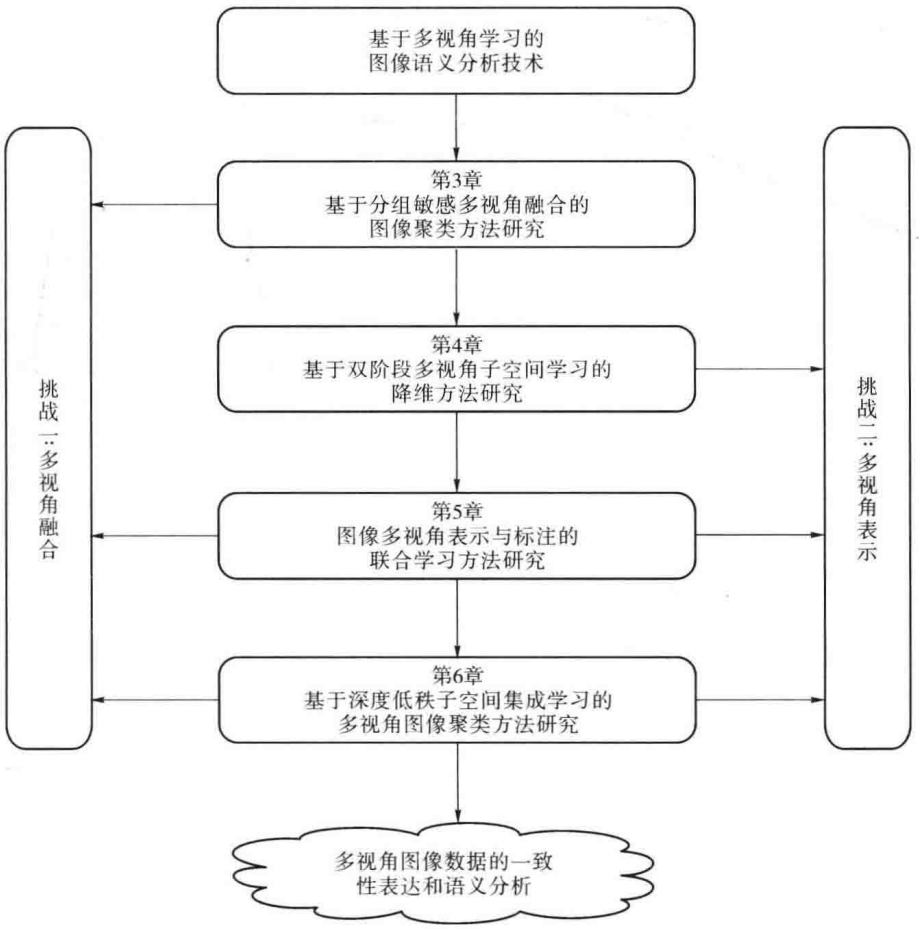

图 1.1　本书各章的组织结构

1.4 符号说明

最后在这里对本书的符号进行说明。本书采用黑体大写字母表示矩阵。对于任意矩阵 $\boldsymbol{A}$，$\boldsymbol{A}_{ij}$ 表示其第(i,j)位置上的元素，$\boldsymbol{A}_{i\cdot}$ 表示第 i 行，$\boldsymbol{A}_{\cdot j}$ 表示第 j 列，它的迹表示为 $\mathrm{Tr}[\boldsymbol{A}]$。$\boldsymbol{A}\odot\boldsymbol{B}$ 和 $\boldsymbol{A}\oslash\boldsymbol{B}$ 表示两个矩阵各个对应元素的乘法和除法。$(\boldsymbol{A},\boldsymbol{B})$表示水平拼接两个矩阵 $\boldsymbol{A}$ 和 $\boldsymbol{B}$，$(\boldsymbol{A};\boldsymbol{B})$表示垂直拼接它们。$\boldsymbol{1}_M\in\boldsymbol{R}^{M\times 1}$ 表示一个元素都是 1 的向量。对于矩阵 $\boldsymbol{A}\in\boldsymbol{R}^{N\times M}$，其 Frobenius 范数表示为 $\|\boldsymbol{A}\|_{\mathrm{F}}$，$\ell_{2,1}$-范数定义为

$$\|\boldsymbol{A}\|_{2,1}=\sum_{i=1}^{N}\sqrt{\sum_{j=1}^{M}\boldsymbol{A}_{ij}^{2}} \tag{1-1}$$

第2章 研究现状

2.1 本章导读

随着互联网和信息技术的快速发展，多媒体数据呈现爆炸式的增长，图像、视频和音频等数据在网络上大量的出现并传播，对人们的生活、工作、娱乐等方面产生了巨大的影响。随着机器学习、模式识别等相关学科的发展，数据从以往仅依靠单一视角的表达方式发展为基于多个视角的表达方式。通过利用多个视角的信息，我们能够获得对数据更加全面准确的描述，从而更好地进行数据分类、聚类等任务。由于各个视角数据的物理意义和统计特性不同，传统处理单一视角数据的方法无法很好地适应多视角数据的特性，无法让多个视角的信息相互补充，不能发挥出多视角数据的优势。为了有效地分析和处理多视角数据，研究者开展了一系列多视角学习的研究工作。在本章中我们将详细介绍多视角学习的相关方法。首先简要介绍多视角学习的两个基本准则，这两个基本准则是多视角学习方法中常用到的两个准则。基于这两个准则，研究者提出了多种多样的多视角学习模型。然后我们将现有的多视角研究工作按照方法划分为协同学习、多核学习和子空间学习三类，并依次介绍各个类别中多视角研究工作的详细情况。

2.2 基本准则

为了更好地使用多视角数据，多视角学习方法需要能够有效地利用多个视角的信息。由于不同视角具有不同的特性和物理意义，如果不能合适地处理来自不同视角的信息，模型将无法发挥多视角数据的优势，所得到的学习性能也会受到限制。针对多视角数据的特点，现有的多视角学习工作主要基于一致性准则和互补性准则[17]两个学习准则。下面我们就依次介绍。

2.2.1 一致性准则

一致性准则(Consensus Principle)的学习目标是最大化不同视角的一致性来得到统一的学习结果。Dasgupta 等人[18]指出了在两个视角下所训练的学习器的一致性和它们的错误率之间的关系。在一些温和的假设下，有如下的不等式成立：

$$P(f^1 \neq f^2) \geqslant \max\{P_{\text{err}}(f^1), P_{\text{err}}(f^2)\} \tag{2-1}$$

式中，f^1 和 f^2 分别是在两个视角上训练的学习器，$P_{\text{err}}(\cdot)$ 表示产生误差的概率。从以上不等式我们可以看出，两个假设 f^1 和 f^2 的不一致性就是它们各自误差的上界。因此，我们通过减小两个假设之间的不一致性，就可以使它们各自的错误率降低。此外，还有许多的研究工作使用了一致性准则对多视角数据进行分析和处理。例如，在协同训练(Co-Training)的方法中[19,20]，通过在两个不同的视角上分别训练分类器，然后把它们在另一个视角上进行预测，所得到的置信度较高的样本作为标记样本。通过最小化在标记样本上的预测误差以及最大化在无标记样本上的预测一致性，协同训练最终会得到更准确的分类器。此外，典型相关分析(Canonical Correlation Analysis，CCA)也利用了一致性准则，它的目标是对两个不同的视角找出一个共享的子空间来使两个视角的相关性最大。在学得的共享空间中，若来自两个视角的数据相关性越高，则说明它们的空间表示越趋于一致。在一些多视角聚类的工作中[21]，通过使用聚类集成的方法在多个视角中学出一个统一的相似度矩阵，该矩阵可以最小化不同视角间的差异性，那么最终的聚类结果就可以通过对一个矩阵进行聚类来完成。此外，一些半监督学习方法[22]首先在各个视角上训练出 SVM 分类器，然后结合 CCA 的思想，让它们的预测结果趋于一致。假设多视角数据集 X 由两个视角构成，每个样本可以写为 (x_i^1, x_i^2, y_i)，其中 y_i 是样本的标签，则其约束形式可以写为

$$\| f^1(x_i^1) - f^2(x_i^2) \| \leqslant \eta_i + \varepsilon \tag{2-2}$$

式中，$f^1(\cdot)$ 和 $f^2(\cdot)$ 分别是两个视角下 SVM 的决策面函数；η_i 约束了两个视角之间的一致性，ε 是 SVM 分类器的松弛变量。还有一些多视角嵌入学习的方法[9,23]首先对多视角数据的局部结构进行建模，然后在所学的嵌入空间中让局部结构进行保留，通过对不同视角的局部结构信息进行一致性约束，使它们在嵌入空间中能够对齐，从而得到多视角数据在嵌入空间中的表示。

2.2.2 互补性准则

互补性准则(Complementary Principle)是指多视角数据的各个视角信息之间存在着互补性，某个视角可能含有一些其他视角没有的信息。因此，如果我们同时使用多个视角的信息就可以更加全面准确地理解数据之间的关系。研究工作[24]表明，协同学习的方法之所以成功，是因为当不同视角下的分类器之间具有一定的差异时，协同学习的方法就能够更好地利用各视角间的差异来相互补充，从而进一步提高分类器的分类性能。例如，当一个视角下训练的分类器 f_1 将某些样本进行分类后，对另一个视角的分类器 f_2 产生了帮助，那么 f_1 包含了某些 f_2 所没有的信息，这样两个视角间的信息就产生了互补。随着协同训练的不断迭代，不同的视角之间就能够相互提升学习性能。

在多核学习(Multiple Kernel Learning，MKL)[25-27]中，不同视角的数据分别构建了不同的核矩阵，这些核矩阵描述了不同视角上数据之间的相似性。多核学习的目标就是希望找到一个最优的核矩阵，该核矩阵能够很好地将各个核矩阵融合起来从而达到更优的分类或聚类结果。多核学习就是利用了多视角数据的互补特性，让不同的核矩阵能够互补形成新的核矩阵，该核矩阵能够得到比使用任意一个核矩阵更优的学习性能。一些多视角降维学习的方法[9,28]同样利用了互补性准则。一种简单的使用多视角特征的方法就是把各种特

征拼接起来作为新的数据表示，然后使用单视角学习的方法进行后续的处理。但是这种方法忽略了不同的视角，其物理意义也不同，并且拼接后的特征维度较高，容易造成模型的过拟合等问题。为了解决这个问题，研究者提出了多视角谱嵌入的方法[9]，该方法利用图嵌入学习将多视角信息保留在低维表示中。Han 等人[28]使用组稀疏约束将多视角的互补信息保留在低维表示中。还有一些多视角隐空间学习的方法[29]，在所学的共享子空间中整合了各个视角的信息，从而有效利用多视角信息的互补性，进一步提升低维表示的判别力。为了在不同视角的空间中学习出更优的度量空间，Yu 等人[30]提出了一种多视角度量学习方法，通过利用数据的标签信息和多视角的互补性，该方法能够更有效地学习新的度量空间，进一步提升学习性能。

2.3 方法分类

2.3.1 协同训练

协同训练的思想可上溯于 Yarowsky 提出的一个消除语义模糊问题的无监督算法[31]，该算法使用了两个分类器，分别是在不同的角度上对单词词性进行提取的分类器。这两个分类器通过迭代训练，其性能优于有监督的机器学习方法的性能。协同训练一般在两个视角上分别训练分类器，并使得两个分类器的预测结果在未标记样本上保持相似，其方法流程如图 2.1 所示。Blum 和 Mitchell 第一次提出了协同训练的算法[19]，该算法在两个视角上面分别训练各自的分类器，然后使用训练后的分类器去标记另一个视角上置信度高的未标记样本。在实验中，一个视角使用的是网页的单词特征，另一个视角则使用网页的超链接特征。实验结果表明，通过协同训练后，算法取得了更优的分类性能，甚至高于有监督学习的分类性能。其具体训练过程见算法 1。

算法 1 协同训练算法

输入：带标记的训练样本集合 L，未标记的样本集合 U；
1：在集合 U 中随机选取子集 U′；
2：for 进行 k 次迭代 do
3：使用训练集 L 的第一个视角信息 X_1 训练分类器 h_1；
4：使用训练集 L 的第一个视角信息 X_2 训练分类器 h_2；
5：使用 h_1 标记 U′ 中置信度最高的 p 个正例样本和 n 个负例样本；
6：使用 h_2 标记 U′ 中置信度最高的 p 个正例样本和 n 个负例样本；
7：将被标记的样本加入集合 L 中；
8：从集合 U 中随机挑选 2n + 2p 个样本来更新 U′；
9：end for
输出：分类器 h_1、h_2