



中国计算机学会学术著作丛书
——知识科学系列 11

机器学习及其应用 2015

高新波 张军平 主编

清华大学出版社



中国计算机学会学术著作丛书
——知识科学系列 **11**

机器学习及其应用2015

高新波 张军平 主编



清华大学出版社
北京

内 容 简 介

本书是对第十一届和十二届中国机器学习及其应用研讨会的一个总结,共邀请了会议中的10位专家就其研究领域撰文,以综述的形式探讨了机器学习不同分支及相关领域的研究进展。全书共分10章,内容分别涉及稀疏学习、众包数据中的隐类别分析、演化优化、深度学习、半监督支持向量机、差分隐私保护等技术,以及机器学习在图像质量评价、图像语义分割、多模态图像分析等方面的应用,此外,还介绍了新硬件寒武纪神经网络计算机的研究进展。

本书可供计算机、自动化及相关专业的研究人员、教师、研究生和工程技术人员参考。

版权所有,侵权必究。侵权举报电话:010-62782989 13701121993

图书在版编目(CIP)数据

机器学习及其应用 2015/高新波,张军平主编.--北京:清华大学出版社,2015

(中国计算机学会学术著作丛书.知识科学系列)

ISBN 978-7-302-40659-4

I. ①机… II. ①高… ②张… III. ①机器学习 IV. ①TP181

中国版本图书馆CIP数据核字(2015)第153201号

责任编辑:薛 慧

封面设计:傅瑞学

责任校对:刘玉霞

责任印制:杨 艳

出版发行:清华大学出版社

网 址: <http://www.tup.com.cn>, <http://www.wqbook.com>

地 址:北京清华大学学研大厦A座 邮 编:100084

社总机:010-62770175 邮 购:010-62786544

投稿与读者服务:010-62776969, c-service@tup.tsinghua.edu.cn

质 量 反 馈:010-62772015, zhiliang@tup.tsinghua.edu.cn

印 装 者:三河市金元印装有限公司

经 销:全国新华书店

开 本:185mm×230mm 印 张:14.25 插 页:1 字 数:275千字

版 次:2015年10月第1版 印 次:2015年10月第1次印刷

印 数:1~1500

定 价:49.00元

产品编号:064797-01

评 审 委 员 会

名誉主任委员：张效祥

主 任 委 员：唐泽圣

副主任委员：陆汝铃

委 员：（以姓氏笔画为序）

王 珊 吕 建 李晓明

林惠民 罗军舟 郑纬民

施伯乐 焦金生 谭铁牛

前言

随着大数据时代的来临,来自互联网、安全、金融、医疗、科学观测等众多领域的数据呈爆发式增长。在享受巨量数据提供的丰富信息的同时,我们也淹没在数据的海洋中,很难挖掘出急需的信息和最有用的知识。要解决这一矛盾,一个重要的策略是利用机器学习。

机器学习发源于人工智能,近三十年来已经逐渐发展成为一门相对完备且独立的学科,广受计算机科学、统计学、认知科学等相关领域的关注。在理论方面,针对数据采样分布与真实分布的差异,形成了概率近似逼近(PAC)的学习机制,并在此基础上发展了传统的统计学习理论。为避免数据预测中目标函数的病态问题,一系列正则化理论被提出,如侧重于可解释性的稀疏学习技术、侧重于保持数据非线性几何结构的流形正则化理论、期望保持最优分类性能的最大间隔正则技术等。不仅如此,应用驱动的机器学习也推动了众多新兴研究方向的产生,如考虑无标签数据的半监督学习,考虑不同数据分布的迁移学习、领域自适应学习,考虑数据“多”特性的多标签、多示例、多视角、多任务学习,考虑网络数据标记策略的众包学习等。与优化技术如随机梯度下降、半正定规划等的融合,也促进了对大规模数据的处理和对全局最优求解。值得一提的是,近年来深度(神经网络)学习通过逐阶递减的特征提取技术以及大数据的训练策略,在多个层面如图像、语音甚至文本分类性能上都超越了统计学习为主导的机器学习方法。这使得神经网络在经历了近二十多年的低谷后,再次把广大研究人员的眼球重新吸引回来。它不仅掀起了机器学习的新一轮热潮,也直接导致了工业界对机器学习的研究和发展前所未有的关注。

2002年,陆汝钤院士在复旦大学智能信息处理实验室发起组织了“智能信息处理系列研讨会”,并将“机器学习及其应用”列为当年支持的研讨会之一。2002年11月,研讨会成功举行,并确定了会议不征文、不收费、报告人由组织者邀请,以及“学术至上,其他从简”的办会宗旨。2004年11月,在复旦大学举行了第二届“机器学习及其应用”研讨会,两天半的会议一直有100余人旁听。2005年起,研讨会由南京大学软件新技术国家重点

实验室举办。2005年11月举办的第三届研讨会吸引了来自全国近10个省市的250余人旁听;2006年11月、2007年11月分别由南京航空航天大学信息科学与技术学院、南京师范大学数学与计算机学院协办第四届和第五届研讨会,两次均吸引了来自全国10余个省市的约300人旁听;2008年11月举行的第六届研讨会,适逢南京大学计算机学科建立50周年,吸引了来自全国10余个省市的380余人旁听;此后在2009年11月和2010年11月在南京大学分别举行了第七、八届研讨会,均有约400人旁听。2011年11月和2012年11月由清华大学自动化系、智能科学与系统国家重点实验室、清华大学信息科学与技术国家实验室(筹)举办第九届和第十届研讨会,两次会议均有500多人旁听。2013年11月由复旦大学计算机科学技术学院和上海市智能信息处理实验室举行了第十一届研讨会,2014年11月在西安电子科技大学举办了第十二届研讨会,这两次会议均有600多人旁听。可以说,“机器学习及其应用”研讨会已经成为机器学习及其相关领域研究人员的盛会。

本书是对第十一届和十二届中国机器学习及其应用研讨会的一个总结,共邀请了会议中的10位专家就其研究领域撰文,以综述的形式探讨了机器学习不同分支及相关领域的研究进展。全书共分10章,内容分别涉及稀疏学习、众包数据中的隐类别分析、演化优化、深度学习、半监督支持向量机、差分隐私保护等技术,以及机器学习在图像质量评价、图像语义分割、多模态图像分析等方面的应用,此外,还介绍了新硬件寒武纪神经网络计算机的研究进展。

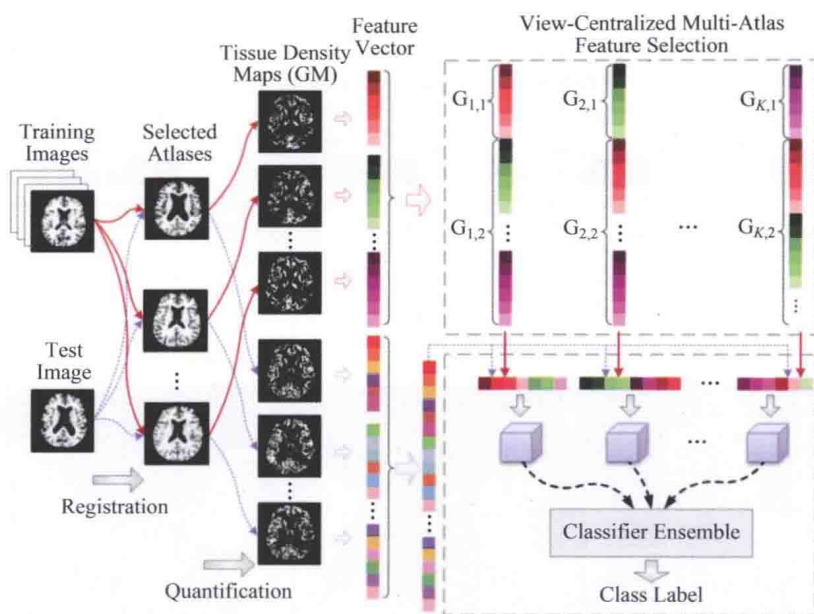
其中,龚平华博士与张长水教授在第1章研究了稀疏学习在鲁棒多任务特征学习和多阶段多任务特征学习中的理论和算法。田天博士与朱军教授在第2章综述了众包标注问题和标注整合的两种基本模型后,提出了众包学习中的隐类别结构模型。针对多数演化算法常依赖于启发式算法的不足,俞扬教授在第3章分析了演化优化的理论基础。通过借鉴深度学习的多层结构框架,陈渤教授在第4章发展了层次化贝叶斯分析以及在线变分贝叶斯推断方法。李宇峰博士与周志华教授在第5章从“多”“快”“好”“省”四个方面详细讨论和分析了半监督支持向量机近十年来的研究新进展。考虑到多数机器学习算法是建立在包含用户敏感信息的数据集上,王立威教授与郑凯博士在第6章分析现有隐私保护模型的优势和不足后,并基于差分隐私策略提出了针对光滑查询的隐私保护机制。作为视觉大数据中最重要的载体之一,图像的质量评价是视觉信息质量评价的基本问题。高新波教授与何立火博士在第7章引入了基于特征表示、回归分析和贝叶斯推理等机器学习方法对图像质量进行了客观评价,提出了一系列的无参考型图像质量评价的测度。除此以外,考虑到图像的高层语义提取一直是计算机视觉中的“难”问题,薛向阳教授在第

8 章从特征融合、深度网络、弱监督策略等方面分析了图像语义分割问题。在脑图像分析中,多种采集设备可形成多模态的图像。为了有效融合多模态异质脑图像数据,张道强教授与程波博士等人在第 9 章从多任务学习、多模态流形正则迁移学习、多视图分类的角度分析了机器学习在多模态脑图像分析的应用。最后,陈天石与陈云霄教授从硬件的角度探讨了机器学习固化的可能性,并对中国科学院计算技术研究所研发的寒武纪系列处理器进行了简要介绍。

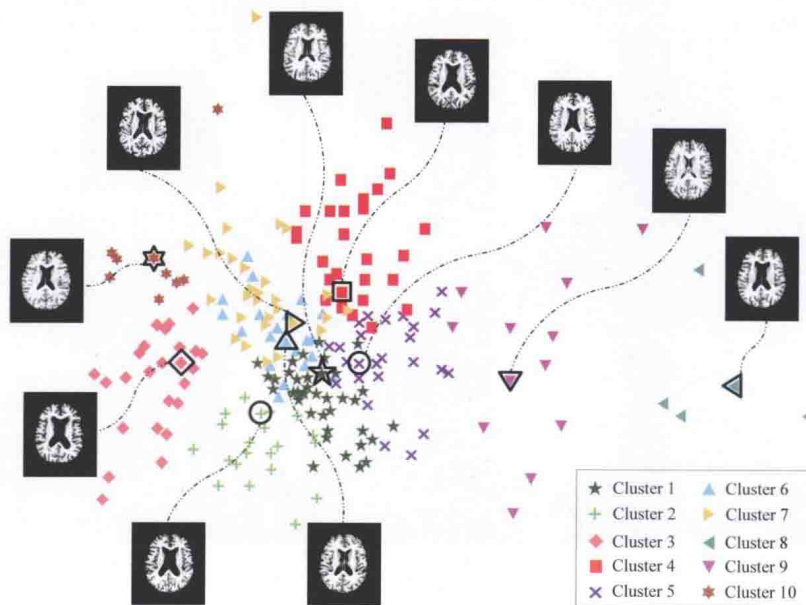
本书概括了国内机器学习及其应用的最新研究进展,可供计算机、自动化、信息处理及相关专业的研究人员、教师、研究生和工程技术人员参考,也可作为人工智能、机器学习课程的辅助内容,希望对有志于从事机器学习研究的人员有所帮助。

高新波 张军平

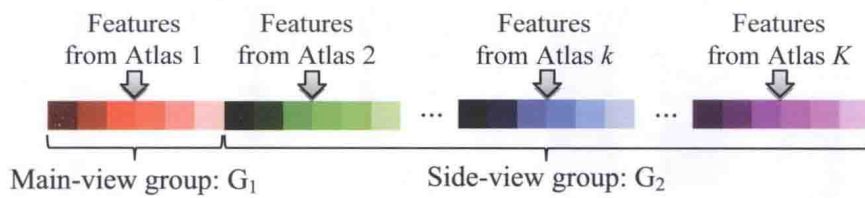
2015 年 7 月



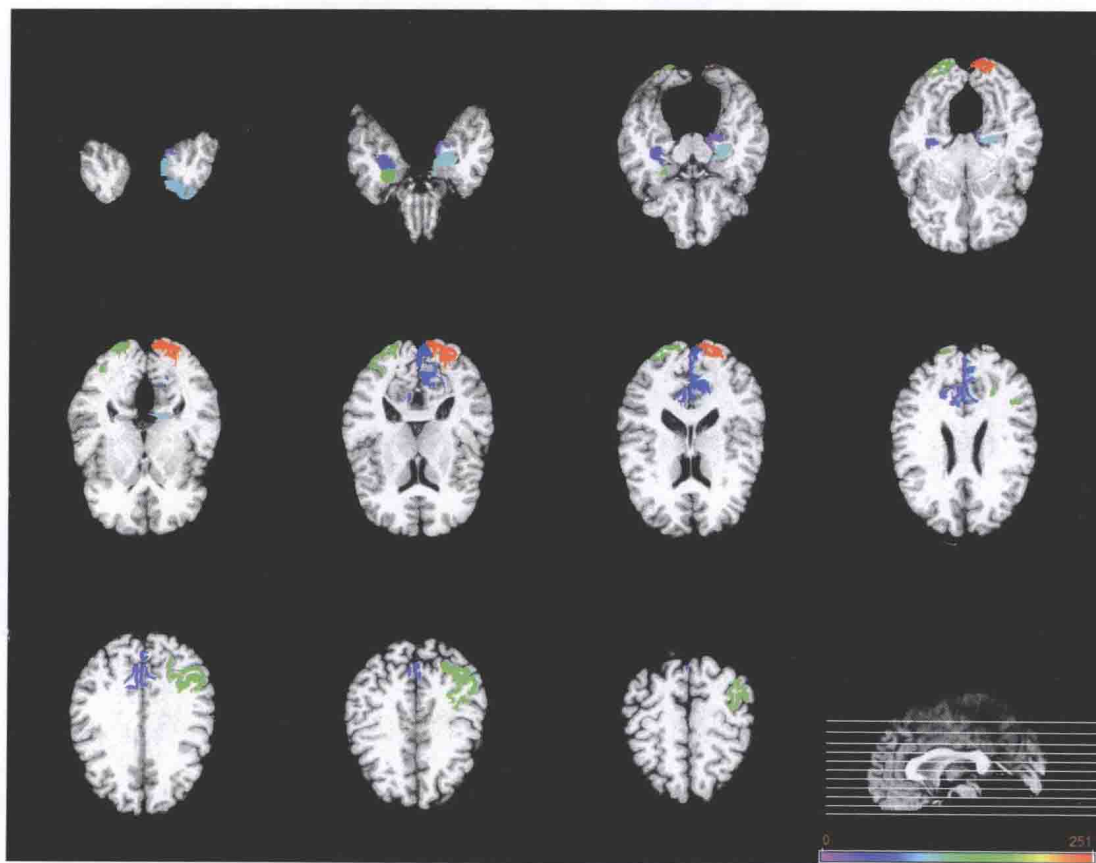
彩插 1



彩插 2



彩插 3



彩插 4

目 录

稀疏学习在多任务学习中的应用	龚平华 张长水	1
1 引言		1
2 鲁棒多任务特征学习		2
3 多阶段多任务特征学习		8
4 结语		15
参考文献		15
众包数据标注中的隐类别分析	田 天 朱 军	19
1 引言		19
2 众包标注问题		20
3 标注整合的几种基本模型		23
3.1 多数投票模型		23
3.2 混淆矩阵模型		24
4 众包标注中的隐类别结构		26
5 隐类别估计		28
6 实验表现		30
7 结语		32
参考文献		33
演化优化的理论研究进展	俞 扬	35
1 引言		35
2 演化优化算法		36
3 演化优化的理论发展		38
4 运行时间分析方法		39
5 逼近性能分析		41

6 算法参数分析.....	43
7 结语.....	45
参考文献	47
基于贝叶斯卷积网络的深度学习算法	陈 渤 53
1 引言.....	53
2 多层稀疏因子分析.....	57
2.1 单层模型	57
2.2 抽取和最大池化	58
2.3 模型特征和可视化	59
3 层次化贝叶斯分析.....	59
3.1 层级结构	59
3.2 计算	60
3.3 贝叶斯输出的应用	60
3.4 与之前模型的相关性	61
4 推理中发掘卷积.....	61
4.1 Gibbs 采样	61
4.2 VB 推理	61
4.3 在线 VB	62
5 实验结果.....	63
5.1 参数设定	63
5.2 合成数据以及 MNIST 数据	64
5.3 Caltech 101 数据分析	67
5.4 每层的激活情况	71
5.5 稀疏性	71
5.6 对于 Caltech 101 的分类	73
5.7 在线 VB 和梵·高油画分析	76
6 结语.....	77
参考文献	79
半监督支持向量机学习方法的研究	李宇峰 周志华 83
1 引言.....	83
2 半监督支持向量机简介.....	84

3 半监督支持向量机学习方法	86
3.1 多：用于多训练示例的大规模半监督支持向量机	87
3.2 快：用于提升学习效率的快速半监督支持向量机	90
3.3 好：用于提供性能保障的安全半监督支持向量机	94
3.4 省：用于代价抑制的代价敏感半监督支持向量机	98
4 结语	102
参考文献	103
差分隐私保护的机器学习	王立威 郑 凯 109
1 引言	109
2 相关定义及性质	111
3 常用机制	113
4 针对光滑查询的隐私保护机制	117
5 实验结果	121
6 结语	122
参考文献	123
学习无参考型图像质量评价方法研究	高新波 何立火 127
1 引言	127
2 基于特征表示的图像质量评价方法	128
2.1 基于特征降维的无参考型图像质量评价方法	129
2.2 基于图像块学习的无参考型图像质量评价方法	129
2.3 基于稀疏表示的无参考型图像质量评价方法	129
3 基于回归分析的图像质量评价方法	132
3.1 基于支撑矢量回归的无参考型图像质量评价方法	132
3.2 基于神经网络的无参考型图像质量评价方法	133
3.3 基于多核学习的无参考型图像质量评价方法	133
4 基于贝叶斯推理的图像质量评价方法	137
4.1 简单概率模型图像质量评价方法	138
4.2 基于主题概率模型的图像质量评价方法	138
4.3 基于深度学习的图像质量评价方法	139
5 实验结果	141
6 结语	145
参考文献	146

图像语义分割	薛向阳	151
1 引言		151
2 无监督图像区域分割		154
3 全监督语义分割方法		155
3.1 基于多尺度分割的语义分割方法		156
3.2 基于多特征融合的语义分割方法		157
3.3 基于深度网络的语义分割方法		159
4 弱监督语义分割方法		161
4.1 带 Bounding Box 训练图像数据		161
4.2 有精确图像层标签的训练图像数据		161
4.3 带噪声标签的训练图像数据		165
5 面向语义图像分割的常用数据集		167
6 不同监督条件下 state of the art 方法对比		168
7 结语		169
参考文献		170
机器学习在多模态脑图像分析中的应用	张道强 程波 接标 刘明霞	175
1 引言		175
2 流形正则化多任务特征学习		176
3 多模态流形正则化迁移学习		179
4 视图中心化的多图谱分类		181
5 实验结果		185
5.1 流形正则化多任务特征学习		185
5.2 多模态流形正则化迁移学习		189
5.3 视图中心化的多图谱分类		193
6 结语		196
参考文献		197
寒武纪神经网络计算机	陈天石 陈云霁	201
1 人工神经网络		201
2 曾经的失败		202
2.1 算法: SVM 的崛起		203
2.2 应用: 认知任务被忽略		203
2.3 工艺: 通用处理器享受摩尔定律红利		203

3	神经网络计算机的涅槃	204
3.1	算法：深度学习的有效训练算法	204
3.2	应用：认知任务的普遍化	205
3.3	工艺：暗硅时代的到来	205
3.4	第二代神经网络的兴起	206
4	主要挑战	206
5	寒武纪神经网络(机器学习)处理器	207
5.1	DianNao	207
5.2	DaDianNao	208
5.3	PuDianNao	210
6	未来工作	212
	参考文献	212



稀疏学习在多任务学习中的应用

龚平华 张长水

清华大学 自动化系, 北京 100084

1 引言

近年来,具有稀疏结构的机器学习问题成为机器学习领域一个很活跃的研究课题,尤其是在互联网数据呈现爆炸式增长的今天,它已经成为人们从海量数据中提取有用信息的重要工具。简单来讲,我们把具有稀疏结构的机器学习问题称为稀疏学习(sparse learning)。在稀疏学习中,我们往往会得到稀疏的解,即解向量或者矩阵中很多元素为零,因而稀疏学习自动具有特征选择的直观解释,即某个元素为零表示相应的特征没有选上,相反,则表示相应的特征被选中。稀疏学习具有特征选择的功能,使得它在机器学习领域有着广泛的应用,特别是在多任务学习问题中,稀疏学习得到了特别的关注。多任务学习^[1]通过把相关的任务放到一起学习,并发掘任务之间的共享信息,从而达到提高推广性能的目的,它是近年来稀疏学习领域的一个十分重要的应用问题。随着稀疏学习研究的不断深入,多任务学习的各种应用成果层出不穷,在目标识别^[1],语音分类^[2]、手写字符识别^[3]、生物医药信息挖掘^[4,5]等领域都有很成功的应用。在多任务学习中,一个关键的问题是如何实现任务之间信息的共享。在现有的一些多任务学习算法中,任务之间共享的信息有很多种,包括神经网络中的隐层单元^[1,6]、贝叶斯模型中的先验^[7-10]、高斯过程的参数^[11]、特征映射矩阵^[12]、分类权重向量^[13]、相似度量矩阵^[2,14]、低秩子空间^[15,16]、一组相关的特征^[17-24]等。多任务特征学习,旨在任务之间学习出相关的特征,最近引起了人们广泛的关注,特别是在各种稀疏学习模型在很多问题中取得成功的应用后,人们对稀疏模型在多任务特征学习的应用研究更是热情高涨,但现有的一些多任务特征学习模型要求任务之间要么同时共享某一个特征,要么同时不共享某一个特

征^[19-21,24],这个要求在实际问题中有些苛刻,因为实际问题中很可能只有部分任务共享某些特征,而且还可能存在异常的任务。为此,本文介绍两种多任务特征学习的稀疏模型,它们分别从从不从的角度来放松“所有任务共享某些特征”的要求。具体说,第2节将介绍一个鲁棒的多任务特征学习算法,该算法能够检测异常任务。第3节将介绍一个非凸的多任务特征学习模型,该模型能够使得一些特征被某些任务共享而不是被所有的任务共享。第4节将对全文做一个总结。

2 鲁棒多任务特征学习

我们首先介绍一个鲁棒的多任务特征学习模型^[25],该模型在学习到共享特征的同时,还能够检测到异常任务。具体来说,我们将权重矩阵分解成两个矩阵之和。对第一个矩阵施加一个组稀疏惩罚,以达到学习共享特征的目的。同时,对第二个矩阵的转置也施加一个组稀疏惩罚,以达到检测异常任务的目的。我们利用加速梯度下降法来有效求解相应的优化问题,使得提出的算法能够适用于大规模的问题。此外,我们对提出的模型进行了理论上的分析。具体来说,我们给出了参数估计误差和预测误差的界。另外,在假设真实的权重幅度大于噪声这个基础上,理论分析指出,在一定条件下,我们的模型能够获得真实共享的特征和异常任务。

下面介绍鲁棒多任务特征学习模型。假定有 m 个任务,分别对应训练数据 $\{(X_1, y_1), \dots, (X_m, y_m)\}$, 其中 $X_i \in \mathbb{R}^{n_i \times d}$ 是第 i 个任务的数据矩阵(每一行是一个样本); $y_i \in \mathbb{R}^{n_i}$ 是第 i 个任务的预测目标(在回归问题中, y_i 是连续的回归值,在分类问题中, y_i 是离散的分类标签); d 是样本的维数; n_i 是第 i 个任务的样本个数。我们归一化数据,使得 X_i 每一列的长度为 1, 即

$$\sum_{j=1}^{n_i} (x_{jk}^{(i)})^2 = 1, \quad \forall k \in \mathbb{N}_d \quad (1)$$

其中, $x_{jk}^{(i)}$ 表示 X_i 的第 (j, k) 个元素; $\mathbb{N}_d$ 表示集合 $\{1, \dots, d\}$ 。我们为每个任务学习一个线性函数

$$y_i \approx f_i(X_i) = X_i w_i, \quad i \in \mathbb{N}_m \quad (2)$$

其中 $W = [w_1, \dots, w_m] \in \mathbb{R}^{d \times m}$ 是权重矩阵。我们将 W 分解成两个矩阵之和, 即 $W = P + Q$ (请参考图 1 所示的示意图), 同时分别对 P 和 Q 施加不同的惩罚, 以挖掘任务之间的关系。

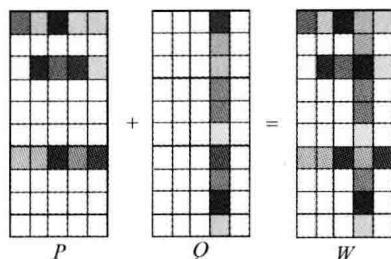


图1 鲁棒多任务特征学习(rMTFL)权重矩阵分解示意图(该图来源于文献[25])

其中有色方块表示相应的特征被选中。这里共有5个任务,其中第4个任务是异常任务

具体来说,我们将鲁棒多任务特征学习(robust multi-task feature learning, rMTFL)建模成如下的优化问题:

$$\min_{W, P, Q} \sum_{i=1}^m \frac{1}{m_i} \|X_i \mathbf{w}_i - \mathbf{y}_i\|^2 + \lambda_1 \|P\|_{1,2} + \lambda_2 \|Q^T\|_{1,2}, \quad \text{s. t. } W = P + Q \quad (3)$$

其中第一个对 P 的正则项用来获取任务之间的共享特征;第二个对 Q 的正则项用来发现异常的任务; λ_1 和 λ_2 是非负的正则项参数; P 的 $\ell_{1,2}$ 范数定义为: $\|P\|_{1,2} = \sum_{i=1}^d \|\mathbf{p}^i\|$, $\mathbf{p}^i$ 是 P 矩阵的第 i 行。具体来说,第一个正则项对 P 施加一个 $\ell_{1,2}$ 范数的惩罚,这使得问题(3)的最优解 P^* 的每一行要么都是零,要么都不是零^[18]。这样,所有的任务要么同时选上某一组特征,要么同时不选某一组特征,然而,限制所有的任务同时共享一组共同的特征在实际的应用中可能太苛刻,因为异常的任务往往会存在。为了解决这一问题,我们引入了一个对 Q 的正则项,用来发现异常的任务。类似地,问题(3)的最优解 Q^* 的每一列要么都是零,要么都不是零,不是零的列就对应异常的任务。直观上来讲,如果 Q^* 的第 i 列的每个分量都是非零的,那么 W^* 的第 i 列的每个分量都是非零的,这样第 i 个任务与其他任务就不再共享相同的特征,从而被检测为异常的任务,同时,对于其他的任务(对应于 Q^* 的非零列),则共享一组共同的特征(请参考图1)。

接着我们将详细介绍如何有效求解问题(3)的鲁棒多任务特征学习优化问题。记

$$l(P, Q) = \sum_{i=1}^m \frac{1}{m_i} \|X_i(\mathbf{p}_i + \mathbf{q}_i) - \mathbf{y}_i\|^2 \quad (4)$$

$$r(P, Q) = \lambda_1 \|P\|_{1,2} + \lambda_2 \|Q^T\|_{1,2}$$

其中, $l(P, Q)$ 是经验损失函数; $r(P, Q)$ 是正则项; $\mathbf{p}_i$ 表示矩阵 P 的第 i 列。我们注意到问题(3)的目标函数是一个连续可微的函数 $l(P, Q)$ 与一个不可微的函数 $r(P, Q)$ 的和,