

# 多标记数据的特征及标记降维方法综述\*

张平照, 张辉宜

(安徽工业大学 计算机科学与技术学院, 安徽 马鞍山 243000)

**摘要:**多标记学习降维方法在实际应用问题中用以处理特征, 标记或二者维度较高的数据集, 已成为研究热点; 针对目前多标记学习降维方法数量众多, 种类繁杂而导致缺乏科学分类的问题, 从多标记数据降维空间选择的角度, 提出将多标记学习降维方法按照特征空间降维, 标记空间降维和二者均降维的形式归纳为三类, 其中特征空间降维又分为特征降维和特征选择两类问题, 分别从独立于和依赖于彼此空间的角度对已有的 40 余篇文献中的典型多标记学习降维算法的研究现状进行了综述; 最后, 总结了多标记学习降维方法的研究现状和启示, 并提出了未来进一步的研究方向。

**关键词:**多标记学习; 特征; 标记; 降维

**中图分类号:**TP181

**文献标志码:**A

**文章编号:**1672-058X(2020)05-0023-07

## 0 引言

多标记学习(MLL)涉及监督学习问题, 其中每个实例可能同时与多个标记关联, 多标记学习与传统的单标记或者多类分类之间的主要区别在于, 多标记学习中的标记不是互斥的。多标记学习出现在许多实际应用中, 例如在网页分类中, 网页可以包含多个主题; 在基因和蛋白质功能预测中, 单个基因或蛋白质通常可以执行多项功能。

与单标记或者多类分类相比, 多标记学习更为通用, 因此解决起来更具有挑战性。其中一项挑战便是对具有大规模数据的多标记学习有效性和效率性的考量, 特别是当数据维度较大带来的“维度诅咒”<sup>[1]</sup>问题, 这意味着对于给定的样本大小, 存在最大数量的特征, 超过这些特征则算法的性能将会下降, 这是由于数据点变得稀疏且彼此之间距离较远。而如今, 数据维度正在呈爆炸式增长趋势, 随之产生了海量质量各异的原始数据, 最近这种规模

的演变还导致了极端多标记学习<sup>[2]</sup>的出现, 该情况考虑标记数量巨大(约 $10^6$ 或更多)的问题。以上这些问题日益增长的复杂性使降维方法得到了重大关注。降维方法旨在减少特征, 标记或两者的数量, 以提高分类器的缩放特性和预测性能。

降维有着悠久的历史, 并与不同的动机相关, 例如在数据可视化, 数据压缩, 数据去燥等方面。简而言之, 对原始数据使用降维处理可以获得一种综合的表现形式, 这种形式能挖掘出隐藏在整体中的关联和结构并指导算法的学习<sup>[3]</sup>。多标记学习中的降维作为处理大数据和嘈杂数据的一种有效手段, 在过去的十几年中一直是热门主题, 随之诞生了各种降维方法。然而却少有文章广泛探索现有方法并进行全局的比较。

分析了 40 余篇多标记降维方法的典型文献, 根据特征空间降维, 标记空间降维或者两者均降维对算法进行分类, 并考虑到两者之间的独立性和依赖性。总结在多标记学习中开发的降维方法, 旨在帮助读者更高效地选择更有效的策略。

收稿日期:2019-11-05; 修回日期:2019-12-10.

\* 基金项目:安徽省高校自然科学研究重点项目资助(KJ2017A063).

作者简介:张平照(1995—), 男, 安徽芜湖人, 硕士研究生, 从事机器学习研究.

# 1 多标记学习

单标记分类和多类分类是目前较为流行的分类方式,对于前者而言,它的每个实例只会关联一个标记,例如每个时刻只对应一个温度;而后者表示分类任务可以有多个类别,但一个实例只属于其中一

个类别,例如一张纸币可以是十元或者百元的,但不可以同时是二者。上述情况统称为单标记学习(SLL)。但是在很多实际应用中这些任务可以用多个标记来描述,如图 1(a)所示,一部电影既可以划分为“音乐片”,又可以用“剧情片”来形容;如图 1(b)所示,一个博客网页可以拥有多个标记便于快速定位文章位置。上述情况被称作为多标记学习(MLL)。

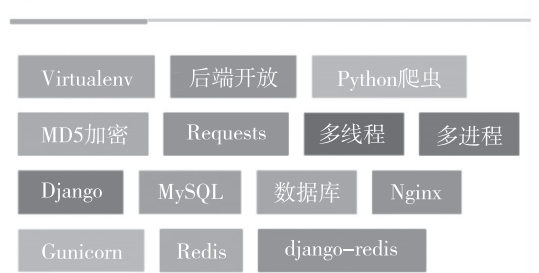
海上钢琴师 La leggenda del pianista sull'



导演: 朱塞佩·托纳多雷  
 编剧: 亚利桑德罗·巴里克/朱塞佩·托纳多雷  
 主演: 蒂姆·罗斯/普路特·泰勒·文斯/比尔·努恩/克兰伦斯·威廉姆斯三世/梅兰尼·蒂埃里/更多  
 类型: 剧情/音乐  
 纸片国家/地区: 意大利  
 语言: 英语/法语/意大利  
 上映日期: 2019-11-15 (中国大陆)/1998-10-28 (意大利)  
 片长: 165 min/120 min (法国版)/169 min (加长版)/125 min (中国大陆)

(a) 一部电影

标签云



(b) 一个网页

图 1 多标记学习两个例子

Fig. 1 Two examples of multi label learning

多标记学习和单标记学习之间的重要区别在于,多个标记通常是相互关联的,这意味着它们彼此间并非相互独立。而在单标记学习中,类别之间是互斥的。例如在图像注解中,包含“海洋”的图像比包含“汽车”的图像更有可能包含“船只”这一元素。在此实例中,“海洋”跟“船只”是高度相关的标记,而跟“汽车”的相关度较低。这类知识能够帮助分类器获取更好的性能,基于标记相关性的顺序可以从 3 种策略进行考虑。一阶策略不考虑其他标记的共存,而是通过单独考虑每个标记来执行学习任务;二阶策略考虑标记间的成对关系,因而标记间的相关性可以在某种程度上被二阶策略利用;高级策略则需要考虑更多标记间的相关性。不论采用何种策略,这些方法均包含在“问题转化”和“算法适应”这两类机器学习算法之中。

## 1.1 问题转化

问题转化方法首先将多标记学习问题转化为一组单标记问题,然后使用现有的单标记学习方法对问题进行求解。关键思想是消除原始标记空间中的重叠标记。与算法适应相比,问题转化更加灵活,因为在问题转化后可以应用任何现成的单标记分类器。将介绍 3 种不同的转化方法,包括复制转化(Co),二元相关(BR)和标记幂集(LP)。

(1) 复制转化通过对数据集执行转换将多标记问题转换为多类问题。具体来说,对于原始多标记问题,用一组新实例替换原始实例,通过这种复制转化,原始多标记学习问题被转换为多类问题。

(2) 二元相关是文献中用于多标记学习最著名和最常用的转化方法。具体来说,二元相关为所有标记独立学习与标记相等数量的二元分类器,对于每个标记,它将与之相关联的所有实例视为正训练样本,其余实例视为负训练样本。基于新的训练数据集,为标记学习分类器。二元相关作为一种直接的转化方法,可以与许多最新的二元分类算法相结合,例如支持向量机和人工神经网络。而二元相关的缺点也很明显,由于所有标记均被独立处理,因此它在转换中没有考虑标记相关性,无法有效处理较大数据集。

(3) 标记幂集将多标记训练数据中的每个不同的标记组合视作新标记,多标记问题因而被转化为一组单标记分类问题。具体来说,给定一个数据实例,标记幂集将会输出最具可能的类别,即原始问题中的一组标记。与复制转化与二元相关相比,标记幂集考虑了不同标记间的相关性和依赖性。而缺点是在进行预测时不会考虑未出现在训练数据中的标记组合。

### 1.2 算法适应

算法适应通过将现有方法用于单标记学习来解决多标记学习问题。算法适应的一个关键研究问题是如何通过考虑标记空间中的不同标记间的相关性来提高预测性能。现将介绍适用于现有单标记分类方法的多标记学习方法,包括排名支持向量机(RankSVM),多标记  $k$  最近邻居(MLkNN)和增强文本(BoosTexter)。

(1) 排名支持向量机基于支持向量机(SVM)方法,其思想是使用新颖的排名损失函数来捕获多标记学习问题的特征,然后相应地解决由此产生的优化问题。与 SVM 相比,RankSVM 尝试控制模型复杂度,同时最大程度地减小误差。

(2) 多标记  $k$  最近邻居是通过扩展  $k$  最近邻居(kNN)来处理多标记数据的,其基本思想是首先在框架中识别未见实例的  $k$  个最近的邻居,然后基于这些实例的标记集获得的统计信息,应用最大后验来预测未见实例的标记集。

(3) 增强文本将适应增强(AdaBoost)算法拓展到了多标记学习范畴,其基本思想是在训练实例和标记上维持一组权重用以处理多个标记,在增强过程中,难以分类的训练实例及其标记将获得更高权重,而易于分类的实例及其标记将获得更低的权重,从而提高方法的整体性能。

## 2 多标记数据降维方法

类似于许多其他机器学习任务,多标记学习也在遭受“维度诅咒”,为了解决这类问题,很多针对多标记数据的降维方法应运而生,将对它们一一介绍。

假设一个数据集中有  $n$  个实例,每个实例由  $d$  个特征和  $q$  个标记描述。分别用  $X$  和  $Y$  表示  $n \times d$  的特征矩阵和  $n \times q$  的标记矩阵,在没有歧义的情况下, $X$  和  $Y$  也指特征空间和标记空间,多标记学习分类的目的是为任意一个特征向量  $x \in R^d$  预测出正确的标记向量  $y \in R^q$ 。在训练阶段,给定一个特征矩阵  $X$ ,调整分类器以使其预测适合的标记矩阵  $Y$ 。

基于降维的大多数多标记分类方法遵循两个步骤:缩减  $X$  或  $Y$  或同时缩减两者;根据缩减后的空间使用分类器预测标记。降维通常被作为预测之前

的独立数据预处理过程,但也有有的研究者对降维和分类之间的耦合进行探索。无论采用何种策略,降维对分类器性能的影响都能够通过所预测标记的质量来进行评估(如 Hamming Loss, F1)。

分别用  $X'$  和  $Y'$  表示  $X$  和  $Y$  的缩减空间,其中  $X' = n \times q, Y' = n \times l, q$  和  $l$  分别是缩减空间  $X'$  和  $Y'$  的维度。如图 2 所示,有 3 种不同的方法可以解决多标记分类的降维问题:将特征空间  $X$  缩减为  $X'$ ,并根据缩减的特征矩阵  $X'$  预测标记矩阵  $Y$ ;将标记空间  $Y$  缩减为  $Y'$ ,并根据特征矩阵  $X$  预测缩减的标记矩阵  $Y'$ ;将特征空间和标记空间均缩减为  $X'$  和  $Y'$ ,并根据缩减的特征矩阵  $X'$  预测缩减的标记矩阵  $Y'$ 。

将根据要降维的目标空间详细介绍多标记数据的降维方法,并按照它们缩减的空间( $X, Y, X'$  和  $Y'$ )对它们进行分组(图 2)。

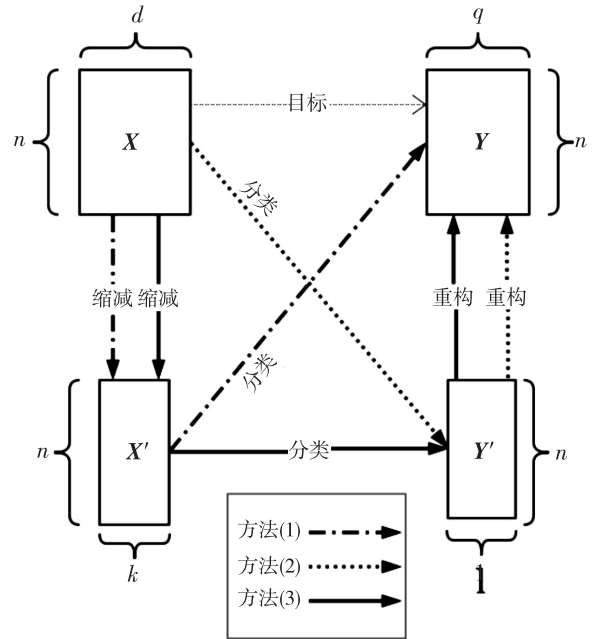


图 2 多标记分类降维策略示意图

Fig. 2 Schematic diagram of dimensionality reduction strategy for multi label classification

### 2.1 特征空间降维

特征空间降维方法旨在将原始较大的特征空间  $X$  转换为缩减空间  $X'$ ,目的是提取数据的基本信息。多标记数据的特征空间降维方法可以分为特征降维和特征选择。

#### 2.1.1 特征降维

特征降维方法可以细分为独立或依赖于标记所携带的信息。

(1) 独立于标记的很多方法最初并不是为了多标记学习而开发的,但是却经常应用到其中,根据所考虑的缩减信息,可以将这些方法分为 3 个系列:一是最大程度保留特征协方差。例如,主成分分析(PCA)<sup>[4]</sup>是通过特征协方差矩阵的特征分解得到最大方差的正交方向(即主成分);约束非负矩阵分解(CNMF)<sup>[5]</sup>通过对  $X$  上的非负矩阵进行分解使第一个因子被作为缩减的特征空间  $X'$ ;随机主成分分析(RPCA)<sup>[6]</sup>通过随机化代数技术来逼近  $X$  的主成分。二是最小化由  $X$  和  $X'$  之间的距离引起的重构误差。例如,自动编码器(AE)<sup>[7]</sup>通过非线性缩减(投影、激活)和解码来有效地进行缩减和重构原始特征空间。三是最大限度地保持用  $X$  和  $X'$  描述的项之间的距离。例如,局部保留投影(LPP)是通过实例邻接图的频谱分析以计算最大限度地保留它的简化特征空间;正交保留局部投影(OLPP)<sup>[8]</sup>是对 LPP 在缩减特征空间上具有正交性约束的扩展。

(2) 依赖于标记的方法旨在通过标记信息来指导缩减过程,这有助于加强提取的缩减特征空间  $X'$  与标记空间  $Y$  之间的联系,主要包括 3 种主要策略:

通过标准准则(协方差、Hilbert-Schmidt 独立性)实现  $X-Y$  关系的最大化。例如,依赖最大化多标记降维(MDDM)。通过特征空间的线性投影产生与标记空间具有最小 Hilbert-Schmidt 独立性的缩减空间;直接多标记线性判别分析(DMLDA)<sup>[9]</sup>重新定义多标记线性判别分析(MLDA)<sup>[10]</sup>的类间方差矩阵,以克服 MLDA 对缩减空间维度的限制。

保持原始标记空间  $Y$  中描述的实例与缩减特征空间  $X'$  中描述的实例之间的等距关系。例如,超图谱学习(HSL)<sup>[11]</sup>通过超图谱分解将实例与多个公共标记联系起来以获得一个缩减的特征空间,使它们之间具有局部性;监督正交保留局部投影(SOLPP)<sup>[12]</sup>对特征或标记邻接图进行频谱分解,以获得一个缩减的特征空间,其中邻居具有共同的特征和标记。

通过学习一个子空间  $X'$  来重构  $X$  和  $Y$ ,以最大化特征与标记空间的联系。例如多标记潜在语义索引(MLSI)<sup>[13]</sup>通过线性特征空间投影来优化原始特征空间的重构以及投影特征空间与标记空间的相关性;多标记共享子空间(ML<sub>IS</sub>)假设一个子空间被多个标记共享,并计算一个线性变换来发现这个

子空间,在特征空间与标记空间的重构中产生嵌入的结果。

### 2.1.2 特征选择

特征选择旨在减少特征数量、达到降低维度的效果,关于特征选择的多标记学习算法可以分为过滤式,包裹式和嵌入式这 3 类。

(1) 过滤式方法的思想是通过特征进行评分以保留得分高的特征,再去学习分类器,优点是计算复杂度低,适合用于高维数据。例如,基于互信息的粒状多标记特征选择<sup>[14]</sup>通过利用最大相关性和最小冗余度这两个标准来筛选出具有判别性的特征子集。

(2) 包裹式方法的思想是将特定的学习算法去选择具有较高预测性能的特征,通常可以获得较好的效果,但计算量较大时可能会无法使用。例如,基于可扩展标准的大标记集多标记特征选择(SCLS)<sup>[15]</sup>使用一种可伸缩的相关性评估方式去评估特征子集;多标记模因特征选择(MFSA)<sup>[16]</sup>利用局部优化方法来加强遗传算法的搜索性能来选择高性能的特征子集。

(3) 嵌入式方法的思想是将特征选择过程嵌入到模型训练过程中,利用了前面两类方法的优势。例如,多标记通知特征选择(MIFS)<sup>[17]</sup>通过将多标记分解为低维的缩小空间以捕获标记之间的相关性,然后使用缩小的空间去引导特征选择;缺失标记的多标记特征选择(MLMLFS)<sup>[18]</sup>考虑了这一点,它使用线性回归模型恢复了缺失的标记,然后将  $L_{2,p}$  范数作用在特征选择矩阵上,以选择出最具判别性的特征同时去除噪声特征;基于特征流形学习和稀疏正则化的多标记特征选择(MDFS)<sup>[19]</sup>通过将特征信息投影到低维空间中以利用局部标记相关性,然后使用低维空间来设计基于流形正则化的标记信息,以考虑标记间的全局相关性,最后通过交替优化,学习稀疏表示以用于多标记特征选择。

## 2.2 标记空间降维

由于标记之间往往是相互关联的,因此通过考虑这些相关性来提高分类的质量和可伸缩性是必要的,这可以通过学习标记降维空间来实现,标记空间降维旨在将原始较大的标记空间  $Y$  转换为缩减空间  $Y'$ 。与特征空间降维方法类似,标记空间降维方法细分为独立或依赖于特征所携带的信息。

(1) 独立于特征的方法可以分为 3 类:一是最

大程度保留标记协方差。例如,主标记空间转换(PLST)是基于PCA的,对标记空间中最大方差的正交方向(即主成分)进行计算;多标记列子集选择问题(ML-CSSP)从标记协方差矩阵的谱中推导标记权重并用加权概率进行标记抽样以产生缩减空间 $Y'$ 。二是最小化由 $Y$ 和 $Y'$ 之间的距离引起的重构误差。例如,基于布尔矩阵分解进行多标记分类(MLC-BMaD)通过布尔矩阵分解构造二元的缩减标记空间,使其能够最优地重构原始标记空间;多输出预测地标选择方法(MOPLMS)通过对一个强正则化标记空间编码或解码问题的求解,并选择解中非零标记作为缩减的标记空间。三是最大限度地保持用 $Y$ 和 $Y'$ 描述的项之间的距离。例如,基于多维缩放实现代价敏感的标记嵌入(CLEMS)<sup>[20]</sup>通过标记空间的多维缩放,以根据选择的实例对代价来嵌入实例,该代价反映了标记间的相似性。

(2) 依赖于特征的方法旨在通过特征信息来指导缩减过程,减少标记的同时加强与特征的联系。例如,条件主标记空间转换(CPLST)将PLST和规范相关分析(CCA)结合使用并通过特征信息引导标记空间的缩减;特征感知隐式标记空间编码(FaIE)隐式地减少标记以最大化它们重构原始标记的能力和它们与特征空间的相关性;基于随机技术的响应嵌入(Rembrandt)<sup>[21]</sup>通过随机线性代数技术对特征-标记协方差矩阵进行特征值分解以缩减标记空间,提高了其与特征空间的联系;基于依赖最大化的标记空间降维(DMLR)<sup>[22]</sup>使用PLST中用过的最小二乘损失函数来测量编码损失,并基于Hilbert-Schmidt独立性标准来测量依赖性损失以得到缩减的标记空间;基于特征感知的非线性标记空间变换的多标记分类(COMB)<sup>[23]</sup>将线性的CCA和非线性的内核规范相关分析(KCCA)集成在一起以特征感知的方式获取标记缩减空间;极端多标记分类的稀疏局部嵌入(SLEEC)<sup>[24]</sup>通过保留最近标记间的成对距离来捕获标记相关性以学习标记的低维嵌入,并在嵌入空间中使用kNN分类器进行预测;基于动态超图融合的多标记分类鲁棒嵌入框架(REFDHF)<sup>[25]</sup>能够获得强大抗噪和预测能力的潜在标记空间,同时使用动态超图融合技术通过特征空间中实例的高阶相关性校正噪声。

### 2.3 特征与标记空间降维

当特征空间和标记空间都被缩减时,每个空间

的缩减依赖于另一个空间,研究了两种主要策略。

(1) 在标记空间和特征空间中寻找彼此的线性相关最大化。例如,最小二乘规范相关分析(LS-CCA)将CCA表述为等效最小二乘问题,并将特征和标记投影到低维空间中以获得二者之间的最大相关性;有监督的双重空间缩减(2SDSR)在 $X$ 上应用现有的依赖于标记的特征空间降维方法以及在 $Y$ 上应用已有的依赖于特征的标记空间降维方法;嘈杂的多标记半监督降维(NMLSDR)<sup>[26]</sup>使用标记传播方法并允许数据在传播过程中更改标记以处理嘈杂的标记,然后通过最大化特征和标记的依赖性来学习数据的低维表示。

(2) 最小化 $X'$ 和 $Y'$ 之间的距离函数。例如,凸共嵌入(ILA)利用标记空间与特征空间的投影以优化缩减的标记和特征向量之间的相似性;目标定向归纳矩阵补全(GIMC)<sup>[27]</sup>通过多标记矩阵补全技术缩减了实例的特征与标记空间;典型相关自动编码器(C2AE)<sup>[28]</sup>通过将标记空间自动编码器与CCA相结合以缩减特征和标记,并解码所预测的缩减标记;基于交叉视图搜索的多标记预测(CoE)<sup>[29]</sup>学习一个特征和标记公共的潜在子空间并使输入与输出很好地对齐,以提升kNN的搜索性能并用于预测;一种有效的基于聚类的极端多标记随机森林(CRAFTML)<sup>[30]</sup>通过随机森林策略将特征和标记投影到低维空间中以随机缩减特征和标记空间。

## 3 结束语

在多标记学习日益受到关注并满足如今处理高维数据需求的背景下撰写的,为了解决这类问题的复杂性,分析了在最近十几年中已经发布了的多个多标记降维的典型方法。尽管描述这些方法的文献很丰富,但将它们之间联系起来难度较大,如何为眼前的问题选择合适的方法成了亟待考虑的需求。试图将这些方法联系起来,从特征空间降维,标记空间降维或者两者均降维这几个角度作为切入点和方法进行了分类和概述,希望能够为以后的研究者在算法选择上提供参考。

对多标记数据降维方法的分类遵循特征空间降维,标记空间降维或者二者均降维这一标准,特征空间降维方法目前较为流行,但随着人们对极端多标记学习的研究,标记空间降维方法也迅速发展

起来。不论是特征空间还是标记空间的降维一开始都是独立进行的,但如果通过考虑彼空间所携带的信息来减少此空间,保留特征与标记之间的连接,那么对于分类任务便更加有效。

未来的工作将围绕将降维与分类耦合在一起展开,目前大多方法都是降维和分类独立进行,降维被视为与分类阶段无关的数据预处理过程,这会导致在某些情况下算法缺乏灵活性。因此如何在这两者之间建立耦合成为了亟待解决的问题。

### 参考文献(References):

- [1] HERRERA F, CHARTE F, RIVERA A J, et al. Dimensionality Reduction [M]. New York: Springer Cham, 2016
- [2] JAIN H, PRABHU Y, VARMA M. Extreme Multi-Label Loss Functions for Recommendation, Tagging, Ranking & Other Missing Label Applications [C]//Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM, 2016
- [3] BENGIO Y, COURVILLE A, VINCENT P. Representation Learning: A Review and New Perspectives [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35(8): 1798—1828
- [4] ABDI H, WILLIAMS L. Principal Component Analysis [J]. Wiley Interdisciplinary Reviews: Computational Statistics, 2010, 2(4): 433—459
- [5] LIU Y, JIN R, YANG L. Semi-Supervised Multi-Label Learning by Constrained Non-Negative Matrix Factorization [C]//Palo Alto, CA: AAAI, 2006
- [6] WARMUTH M, KUZMIN D. Randomized PCA Algorithms with Regret Bounds That Are Logarithmic in the Dimension [C]//Advances in Neural Information Processing Systems, Cambridge, MA: The MIT Press, 2007
- [7] HINTON G, SALAKHUTDINOV R. Reducing the Dimensionality of Data with Neural Networks [J]. Science, 2006, 313(6): 504—507
- [8] KOKIOPOULOU E, SAAD Y. Orthogonal Neighborhood Preserving Projections: A Projection-Based Dimensionality Reduction Technique [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2007, 29(12): 2143—2156
- [9] OIKONOMOU M, TEFAS A. Direct Multi-Label Linear Discriminant Analysis [C]//International Conference on Engineering Applications of Neural Networks Berlin: Heidelberg, 2013
- [10] WANG H, DING C, HUANG H. Multi-Label Linear Discriminant Analysis [C]//European Conference on Computer Vision. Berlin: Heidelberg, 2010
- [11] SUN L, JI S, YE J. Hypergraph Spectral Learning for Multi-Label Classification [C]//Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York, NY: ACM, 2008
- [12] MU T, ANANIADOU S. Proximity-Based Graph Embeddings for Multi-Label Classification [C]//Vienna Austria: KDIR, 2010
- [13] YU K, YU S, TRESP V. Multi-Label Informed Latent Semantic Indexing [C]//Proceedings of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. New York: ACM, 2005
- [14] LI F, MIAO D, PEDRYCZ W. Granular Multi-Label Feature Selection Based on Mutual Information [J]. Pattern Recognition, 2017, 67(2): 410—423
- [15] LEE J, KIM D. SCLS: Multi-Label Feature Selection Based on Scalable Criterion for Large Label Set [J]. Pattern Recognition, 2017, 66(5): 342—352
- [16] LEE J, KIM D. Memetic Feature Selection Algorithm for Multi-Label Classification [J]. Information Sciences, 2015, 293(8): 80—96
- [17] JIAN L, LI J, SHU K, et al. Multi-label Informed Feature Selection [C]//IJCAI. Burlington, MA: Morgan Kaufmann, 2016
- [18] ZHU P, XU Q, HU Q, et al. Multi-Label Feature Selection with Missing Labels [J]. Pattern Recognition, 2018, 74(12): 488—502
- [19] CAI Z, ZHU W. Multi-label Feature Selection via Feature Manifold Learning and Sparsity Regularization [J]. International Journal of Machine Learning and Cybernetics, 2018, 9(8): 1321—1334
- [20] HUANG K, LIN H. Cost-Sensitive Label Embedding for Multi-Label Classification [J]. Machine Learning, 2017, 106(9): 1725—1746
- [21] MINEIRO P, KARAMPATZIAKIS N. Fast Label Embeddings via Randomized Linear Algebra [C]//Joint European Conference on Machine Learning and Knowledge Discovery in Databases. Berlin, Heidelberg: Springer, 2015
- [22] ZHANG J, FANG M, WANG H, et al. Dependence Maximization Based Label Space Dimension Reduction for Multi-Label Classification [J]. Engineering Applications

- of Artificial Intelligence,2015,45(10):453—463
- [23] LI X, GUO Y. Multi-Label Classification with Feature-Aware Non-Linear Label Space Transformation [C]//Twenty-Fourth International Joint Conference on Artificial Intelligence. Burlington, MA: Morgan-kaufmann, 2015
- [24] BHATIA K, JAIN H, KAR P, et al. Sparse Local Embeddings for Extreme Multi-Label Classification [C]//Advances in Neural Information Processing Systems. Cambridge, MA: The MIT Press, 2015
- [25] WANG K. Robust Embedding Framework with Dynamic Hypergraph Fusion for Multi-Label Classification [C]//2019 IEEE International Conference on Multimedia and Expo (ICME). Piscataway, NJ: IEEE, 2019
- [26] MIKALSEN K, SOGUERO-RUIZ C, BIANCHI F, et al. Noisy Multi-Label Semi-Supervised Dimensionality Reduction [J]. Pattern Recognition, 2019, 90(7): 257—270
- [27] SI S, CHIANG K, HSIEH C, et al. Goal-Directed Inductive Matrix Completion [C]//Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM, 2016
- [28] YEH C, WU W, KO W, et al. Learning Deep Latent Space for Multi-Label Classification [C]//Thirty-First AAAI Conference on Artificial Intelligence. Palo Alto, CA: AAAI, 2017
- [29] SHEN X, LIU W, TSANG I, et al. Multilabel Prediction via Cross-View Search [J]. IEEE Transactions on Neural Networks and Learning Systems, 2017, 29(9): 4324—4338
- [30] SIBLINI W, KUNTZ P, MEYER F. An Efficient Clustering-Based Random Forest for Extreme Multi-Label Learning [C]//ICML. New York: ACM, 2018

## A Review of Features and Labels Dimensionality Reduction Methods of Multi-Label Data

ZHANG Ping-zhao, ZHANG Hui-yi

(School of Computer Science and Technology, Anhui University of Technology, Anhui Maanshan 243000, China)

**Abstract:** Multi-label learning dimensionality reduction method has been a research hotspot in the practical application problem to deal with data sets with higher features dimensions, labels dimensions or both dimensions. In view of the large number of multi-label learning dimensionality reduction methods and the lack of scientific classification, from the perspective of the dimension reduction space selection of multi-label data, a multi-label learning dimension reduction method is proposed to be classified into three types according to feature space dimension reduction, label space dimension reduction and both. The feature space dimension reduction is divided into two categories: feature dimension reduction and feature selection. They are based on the independent and dependent space of each other. The research status of typical multi-label learning dimensionality reduction algorithms is summarized. Finally, the research status and inspiration of multi-label learning dimensionality reduction methods are reviewed, and further research directions are proposed for the future.

**Key words:** multi-label learning; feature; label; dimensionality reduction

责任编辑:田 静

引用本文/Cite this paper:

张平照,张辉宜. 多标记数据的特征及标记降维方法综述[J]. 重庆工商大学学报(自然科学版), 2020, 37(5): 23—29

ZHANG P Z, ZHANG H Y. A Review of Features and Labels Dimensionality Reduction Methods of Multi-Label Data [J]. Journal of Chongqing Technology and Business University (Natural Science Edition), 2020, 37(5): 23—29