弱监督场景下的支持向量机算法综述

丁世飞[1),2)] 孙玉婷[1)] 梁志贞[1),2)] 郭丽丽[1),2)] 张 健[1),2)] 徐 晓[1),2)]

[1)](中国矿业大学计算机科学与技术学院 江苏 徐州 221116)

[2)](矿山数字化教育部工程研究中心(中国矿业大学) 江苏 徐州 221116)

摘 要 支持向量机(Support Vector Machine,SVM)是一种建立在结构风险最小化原则上的统计学习方法,以其在非线性、小样本以及高维问题中的独特优势被广泛应用于图像识别、故障诊断以及文本分类等领域.但 SVM 是一种监督学习算法,它旨在利用大量的、唯一且明确的真值标记样本来训练学习器,在不完全监督、不确切监督以及多义监督等弱监督场景下难以取得较好的效果.本文首先阐述了弱监督场景的概念和 SVM 的相关理论,然后从弱监督场景角度出发,系统地梳理了目前 SVM 算法的研究现状和发展,包括基于半监督学习、多示例学习以及多标记学习的方法;其中基于半监督学习的方法根据数据假设可细分为基于聚类假设和基于流形假设的方法,基于多标记学习的方法根据解决方案可细分为基于示例水平空间、基于包水平空间以及基于嵌入空间的方法,基于多标记学习的方法根据处理思路可细分为基于问题转换和基于算法自适应的方法;随后,本文总结了部分代表性算法在公开数据集上的实验结果;最后,探讨并展望了未来可能的研究方向.

关键词 弱监督场景;支持向量机;半监督学习;多示例学习;多标记学习

中图法分类号 TP181 **DOI 号** 10.11897/SP.J.1016.2024.00987

Survey on Support Vector Machine Algorithms in Weakly Supervised Scenarios

DING Shi-Fei[1),2)] SUN Yu-Ting[1)] LIANG Zhi-Zhen[1),2)] GUO Li-Li[1),2)] ZHANG Jian[1),2)] XU Xiao[1),2)]

[1)](*School of Computer Science and Technology, China University of Mining and Technology, Xuzhou, Jiangsu* 221116)

[2)](*Mine Digitization Engineering Research Center of the Ministry of Education (China University of Mining and Technology), Xuzhou, Jiangsu* 221116)

Abstract Support Vector Machine (SVM) is a statistical learning method based on the principle of minimizing structural risk. It provides an intuitive geometric interpretation and rigorous mathematical derivation, showing the unique advantages in handling nonlinear, few shot, and high dimensional problems. SVM has garnered significant attention and widely applied in various fields such as image recognition, fault diagnosis, and text classification. SVM is a classical supervised machine learning algorithm designed to train the learner using samples with complete, unique, and unambiguous ground-truth labels to ensure the generalization ability. However, as real-world application tasks become increasingly complex, creating such a sample set is laborious and difficult. On the one hand, it requires a significant amount of time and cost for data collection, cleaning, and debugging. For specific domains, especially in the medical field, experts often need to combine domain knowledge to process and label the samples. On the other hand, learning tasks in the real world often undergo changes and evolution. For example, data annotation criteria, annota-

收稿日期:2023-04-20;在线发布日期:2024-01-26.本课题得到国家自然科学基金(62276265,61976216,62206297,62206296)资助.**丁世飞**(通信作者),博士,教授,博士生导师,中国计算机学会(CCF)杰出会员,主要研究领域为智能信息处理、模式识别、机器学习、数据挖掘、粒计算.E-mail:dingsf@cumt.edu.cn.**孙玉婷**(通信作者),博士研究生,主要研究方向为支持向量机、弱监督学习.E-mail:18270826309@163.com.**梁志贞**(通信作者),博士,副教授,主要研究方向为图像处理、机器学习、数据挖掘.E-mail:liang@cumt.edu.cn.**郭丽丽**,博士,讲师,主要研究方向为深度学习、多模态情感分析.**张 健**,博士,讲师,主要研究方向为深度学习、多标签学习.**徐 晓**,博士,讲师,主要研究方向为数据挖掘、密度峰值聚类.

tion granularity, or downstream use cases may frequently change, requiring the re-labeling of samples. Consequently, a large amount of samples in real-world applications lack complete and unambiguous labels for the high cost of sample labeling. Moreover, samples in most practical task scenarios may exhibit polysemous, that is, a sample can be associated with multiple labels at the same time. Therefore, standard SVM struggles to achieve satisfactory performance in weakly supervised scenarios such as incomplete supervision, inexact supervision, and polysemous supervision. Weakly supervised scenarios are contrasted with supervised scenarios. Unlike the latter, learning algorithms in weakly supervised scenarios are designed to train the learner using samples that may be limited, ambiguous, or only roughly labeled. From the perspective of weakly supervised scenarios, this survey systematically reviews the current research status and development of SVM algorithms. Firstly, the concept of weakly supervised scenarios and the basic mathematical principle of SVM are briefly introduced. Secondly, the existing SVM algorithms in weakly supervised scenarios are divided into three categories according to different learning paradigms, namely, the semi-supervised learning based methods, the multiple instance learning based methods, and the multi-label learning based methods. Specifically, the semi-supervised learning based methods can be further subdivided into clustering assumption based approaches and manifold assumption based approaches according to data assumptions. The multiple instance learning based methods can be further classified into instance level based approaches, bag level based approaches and embedded space based approaches according to problem solutions. The multi-label learning based methods can be further refined into problem transformation based approaches and algorithm adaptation based approaches according to processing ideas. This survey provides a detailed introduction to the representative methods within these categories, summarizes and analyzes their characteristics and shortcomings, offering a basis for selecting different SVM methods in various task scenarios. After that, the performance of some representative algorithms is evaluated and analyzed by carefully conducting experiments on publicly available datasets. Finally, potential research directions for the future development of SVM algorithms in weakly supervised scenarios are discussed, such as data imbalance, weakly supervised regression, mixed weakly supervised learning, large-scale deep-level tasks and learning problems for open enviroment.

Keywords weakly supervised scenarios; support vector machine (SVM); semi-supervised learning; multiple instance learning; multi-label learning

1 引 言

互联网技术的迅速发展使得现实中的数据不断增长,如何对这些大量甚至海量的数据进行高效的分析和管理,并挖掘出其中蕴含的有效信息和知识,已成为亟待解决的问题.机器学习[1]是人工智能领域中的重要组成部分,也是一个非常活跃的研究方向,其主要研究是从观测数据中寻找蕴含在数据中的规律,并根据这些规律对未来的数据或无法观测的数据进行预测,使其具备良好的泛化能力[2].经典的机器学习算法,特别是深度学习,通常对监督信息有以下要求:训练样本规模足够大,样本标记必须充分.这些特点决定了监督信息的质量,进而影响了模型的泛化能力.然而,在实际生产实践中,这些要求往往无法同时满足:一方面,样本的收集周期长、标注成本高,对于某些特定领域的样本集,尤其是医学领域,需要专家结合相关知识,耗费大量人力物力,依据统一准则对样本进行处理和标注,整个样本集的完善往往需要数月甚至数年的时间;另一方面,样本的稀有性也是一个问题,很多实际任务本身就是小样本学习问题,目标样本往往很难获取,例如野外珍稀物种的样本;此外样本的隐私性也限制了大规模获取和使用样本的可能性.可见,在实际场景中样本的收集和标注是非常困难的.2018 年南京大学周志华教授发表了题为"A brief introduction to weakly

supervised learning”的文章[3]，对机器学习任务给出了一个新的趋势和思路，希望机器学习算法能够在弱监督场景下进行工作，旨在研究通过较弱的监督信号来构建预测模型.

与深度学习不同，SVM 是一种基于统计学习理论[4-5]的小样本机器学习方法，与以往仅考虑如何将经验风险尽可能最小化的一类机器学习方法不同，SVM 在训练学习器时采用了结构风险最小化原理，并利用 VC 维对结构化风险进行度量[5]. 根据有限的数据样本信息，SVM 在模型复杂性与学习能力间寻找一个最佳折中，来实现最优的泛化性能，其目标优化问题可形式化为凸二次规划问题，可利用最优化理论保证其全局最优解. SVM 引入核函数的思想，将原始的样本空间直接映射到高维的特征空间上[6]，通过在高维空间中寻找线性超平面来实现原始空间的非线性决策，避免了维数灾难问题. SVM 以其严格的数学理论推导和直观的几何解释在图像分类[7-8]、文本分析[9]、故障诊断[10-11]等领域展开了广泛的应用.

然而 SVM 的标准形式只适用于有监督场景，即它需要通过大量的、具有单一且明确的真值标记样本来训练学习器，然而在很多实际场景中，由于数据样本标注过程耗时耗力，大量的数据样本的标签是缺乏的甚至是不明确的. 另外，在实际应用中样本对象不再具有唯一语义，即一个样本对象可以拥有多个标签. 在这种多义性的场景下，大规模的样本输出空间迫使学习器需要更多更丰富的监督信息[12]. 传统的 SVM 在不完全监督、不确切监督以及多义监督等弱监督场景下难以取得较好的效果. 而弱监督场景下的学习问题普遍存在于语义分割[13-15]、图像识别与理解[16-17]、目标检测[18-20]等各个领域，具有广泛而实际应用背景，因此如何使 SVM 方法在弱监督场景下更有效地学习建模，具有重大的理论与现实意义和应用前景. 目前，研究学者们针对弱监督场景下的支持向量机算法展开了研究，相应地出现了基于半监督学习、多示例学习、多标记学习的研究方法，缓解了传统 SVM 严重依赖于大量人工标注的不足，实现了更为高效的实际应用.

针对 SVM 的相关性研究成果，丁世飞等人[21]、Roy 等人[22]都进行了系统性的阐述和评价，但这些工作主要集中于有监督场景下 SVM 算法的总结概括，而对于弱监督场景下的 SVM 算法综述性研究仍非常少见. 为此，本文主要从弱监督场景下的半监督学习、多示例学习以及多标记学习出发，对基于这些学习问题下的支持向量机算法展开综述. 并对比分析了部分代表性算法，最后对未来的工作进行了展望.

2　相关知识

2.1　弱监督场景

机器学习算法通过学习大量的训练样本来构建预测模型，在很多任务中取得了巨大成功，大部分学习算法，例如深度学习，都严重依赖于具有单一且明确的真值标记的大规模训练样本集，然而在很多任务中，创建这样的训练样本集既昂贵又耗时，需要花费大量时间和成本进行数据的收集、清理和调试，尤其是涉及到领域专业知识的情况下. 除此之外，学习任务经常会发生变化和演变. 例如数据标注准则、标注的粒度或下游的用例都会经常变化，需要重新标记. 因此，由于数据标注代价高昂，很多任务很难获得如全部真值标签这样的强监督信息. 针对这一问题，研究者希望机器学习算法能在弱监督场景下工作，弱监督场景是相对于监督场景而言的，同监督场景不同，弱监督场景下的学习算法可以通过有限的，多义的或者粗糙标注的样本来进行训练. 根据标注数据及信息的提供情况，弱监督场景可以分为不完全监督、不确切监督甚至是多义监督等多种场景，其中不完全监督场景是指训练样本集中只有少量样本具有标签，而绝大部分样本是没有标签的. 例如在进行 Web 网页推荐中，用户需要标注出自己感兴趣的网页，但很少会有用户愿意花大量时间去提供标记，导致有标记的网页样本较少，但 Web 上却存在着无数的网页. 不确切监督是指训练样本只给出粗粒度的标签. 例如在图像分类时，希望图片中的每个对象都被标注，但现实中只有整张图片的标签，而具体图片中的每个对象则是没有标签的. 多义监督是指训练样本集中每一个样本同时具有多个标签. 例如在文本分类中，一篇新闻可能同时涉及到多个主题，如“政治”“经济”“体育”“科技”等. 针对这些具体的弱监督场景下的建模问题，学者们提出了不同的学习范式以克服相应的问题.

如图 1 所示. 面对不完全监督场景下的学习问题，研究学者们提出了半监督学习[24-26]，它可以通过大量的未标记样本以及少量的标记样本来训练学习器，以获取比仅通过少量标记样本训练的学习器更为出色的性能，在一定程度上弥补了标记样本不足的缺陷. 为了解决不确切监督场景下的学习问题，研究学者们提出了多示例学习[27-28]，在该学习框架中一个学习对象可被描述成一个示例包，且一个示例

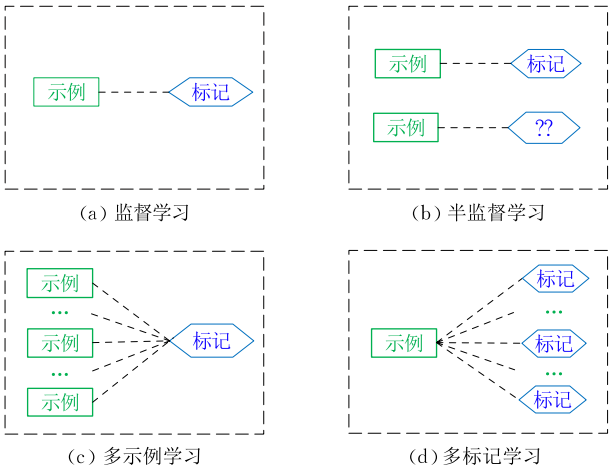

图 1 四种不同的学习框架[23]

包是由若干个没有标签的示例构成,当示例包中所有示例均为负,则该包被标记为负,否则被标记为正.多示例学习通过在包含粗粒度标记的弱监督信息下进行学习,以获得泛化性能更好的学习器.面对多义监督场景下的学习问题,多标记学习[29-30]被研究学者们广泛地研究,在多标记学习中一个学习对象隶属于多个而不是唯一的类别标签,学习目的是给未知的示例赋予多个合适的标记.

2.2 支持向量机

SVM 是针对二分类问题提出的一种监督学习方法,其学习目的是通过寻求一个最优超平面将两种不同类别的样本分开,并保证两种类别的样本间的间隔最大.给定训练数据集 $T=\{(\boldsymbol{x}_i,y_i),i=1,2,\cdots,l\}$,其中 $\boldsymbol{x}_i\in R^n$ 为样本的特征向量,$y_i\in\{-1,+1\}$ 为样本的类别标签,下面给出线性支持向量机和非线性支持向量机的数学模型.

如图 2 所示,针对线性分类问题,SVM 通过寻找一个间隔最大且分类错误最少的超平面 $f(x)=\boldsymbol{w}^{\mathrm{T}}x+b$,其目标优化问题可以表示如下:

$$\min_{\boldsymbol{w},b}\frac{1}{2}\|\boldsymbol{w}\|^2+C\sum_{i=1}^{l}\xi_i$$
$$\text{s.t.}\begin{cases}y_i(\boldsymbol{w}^{\mathrm{T}}\cdot\boldsymbol{x}_i+b)\geqslant 1-\xi_i,\ i=1,2,\cdots,l\\ \xi_i\geqslant 0\end{cases}\tag{1}$$

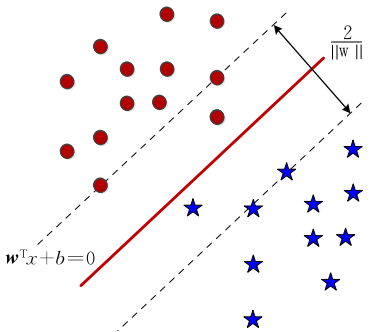

图 2 线性支持向量机的示例

其中,$\boldsymbol{w}$ 表示法向量,b 表示偏置,C 表示惩罚参数,ξ_i 表示松弛变量.

通过引入拉格朗日函数,式(1)的优化问题可转换成对偶优化问题:

$$\max_{\alpha}\sum_{i=1}^{l}\alpha_i-\frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}\alpha_i\alpha_j y_i y_j\boldsymbol{x}_i\cdot\boldsymbol{x}_j$$
$$\text{s.t.}\begin{cases}\sum_{i=1}^{l}\alpha_i y_i=0\\ 0\leqslant\alpha_i\leqslant C,\ i=1,2,\cdots,l\end{cases}\tag{2}$$

式(2)中 $\alpha_i>0$ 为拉格朗日乘子,求解上述问题后得到的决策函数 $f(x)$ 为

$$f(x)=\mathrm{sgn}\Big(\sum_{i=1}^{n}\alpha_i y_i\boldsymbol{x}_i\cdot x+b\Big)\tag{3}$$

针对非线性分类问题,SVM 通过引入核函数的概念,将原始的低维度样本空间映射高维度的特征空间,利用非线性变换来实现样本的线性可分,如图 3 所示.因此式(1)可以变为

$$\min_{\boldsymbol{w},b}\frac{1}{2}\|\boldsymbol{w}\|^2+C\sum_{i=1}^{l}\xi_i$$
$$\text{s.t.}\begin{cases}y_i(\boldsymbol{w}^{\mathrm{T}}K(\boldsymbol{x}_i,\boldsymbol{x}_j)+b)\geqslant 1-\xi_i,\ i=1,2,\cdots,l\\ \xi_i\geqslant 0\end{cases}\tag{4}$$

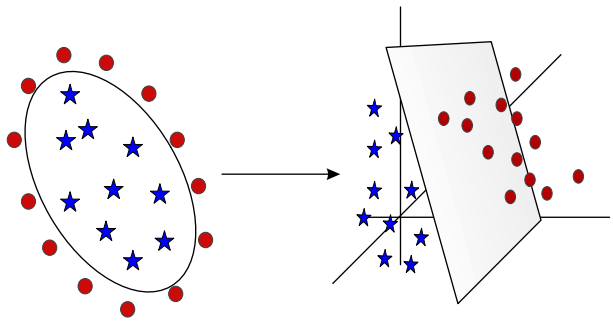

图 3 非线性支持向量机的示例

此时对偶优化问题可变为

$$\max_{\alpha}\sum_{i=1}^{l}\alpha_i-\frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}\alpha_i\alpha_j y_i y_j K(\boldsymbol{x}_i,\boldsymbol{x}_j)$$
$$\text{s.t.}\begin{cases}\sum_{i=1}^{l}\alpha_i y_i=0\\ 0\leqslant\alpha_i\leqslant C,\ i=1,2,\cdots,l\end{cases}\tag{5}$$

相应的决策函数 $f(x)$ 变为

$$f(x)=\mathrm{sgn}\Big(\sum_{i=1}^{n}\alpha_i y_i K(\boldsymbol{x}_i,\boldsymbol{x})+b\Big)\tag{6}$$

3 弱监督场景下的支持向量机算法

本文从弱监督场景角度出发,分别从基于半监督学习、多示例学习和多标记学习这 3 个方面来介绍 SVM 算法的研究进展,其整体架构如图 4 所示.

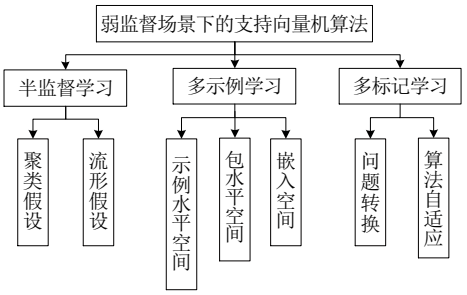

图 4　弱监督场景下的支持向量机算法分类

3.1　基于半监督学习的支持向量机算法

半监督学习是一种介于监督学习和无监督学习间的机器学习方法，它可以利用大量无标记样本来辅助少量标记样本来确定问题的整体分布，也可以利用少量标记样本来指导无监督学习；从而有效地克服了监督学习过程中学习器缺少足够标记样本的不足，同时也缓解了无监督学习过程中未使用标记样本导致学习器性能下降的问题. 半监督学习的主要思想是根据样本数据分布上的模型假设给未标记样本赋予标签，可形式化定义为给定样本集 $R=L\cup U$，其中 $L=\{\boldsymbol{x}_i,y_i\}_{i=1}^{l}$ 是有标记样本集，$U=\{\boldsymbol{x}_i\}_{i=l+1}^{l+u}$ 是未标记样本集，希望可以得到一个映射函数 $f:X\to Y$ 能够准确地预测样本 x 的标签 y，l 和 u 分别为有标记样本和未标记样本的数目. 对于如何构建未标记样本和目标标签之间的联系，半监督学习中存在着两个常用的基本假设：聚类假设[31]和流形假设[32-33].

3.1.1　聚类假设

聚类假设认为同属于一个类簇中的样本极有可能具有相同的标签. 在此假设下，未标记样本可以帮助探明样本分布的稀疏区域和密集区域，引导学习算法根据标记样本所获得的决策边界进行调整，使其尽可能地通过样本分布的稀疏区域[31]. 1999 年，Bennett 等人[34]提出了一种典型的基于聚类假设的半监督支持向量机算法 S3VM，为了有效地利用未标记样本，在 SVM 的基础上，S3VM 添加了两种对未标记样本的约束，这两个约束分别计算未标记样本属于正样本或负样本(即类别 1 或−1)的错分率，目标函数则是求解两种约束下的最小错分率. 具体来说，S3VM 可以定义为

$$\min_{w,b,\eta,\xi,z} C\Big[\sum_{i=1}^{l}\xi_i+\sum_{j=1}^{u}\min(\eta_j,z_j)\Big]+\|\boldsymbol{w}\|$$
$$\text{s.t.}\begin{cases} y_i[\boldsymbol{w}\cdot\boldsymbol{x}_i+b]+\xi_i\geqslant 1, & \xi_i\geqslant 0\\ \boldsymbol{w}\cdot\boldsymbol{x}_j+b+\eta_j\geqslant 1, & \eta_j\geqslant 0\\ -(\boldsymbol{w}\cdot\boldsymbol{x}_i-b)+z_j\geqslant 1, & z_j\geqslant 0 \end{cases} \tag{7}$$

其中 $C>0$ 是错分率的惩罚参数.

Bennett 利用整数规划来求解上述问题，其基本思想为每一个未标记样本点 $\boldsymbol{x}_j$ 添加一个 0 或 1 的决策变量 d_j. 因此，式(7)可转化为以下混合整数规划问题：

$$\min_{w,b,\eta,\xi,z,d} C\Big[\sum_{i=1}^{l}\xi_i+\sum_{j=1}^{u}\min(\eta_j,z_j)\Big]+\|\boldsymbol{w}\|$$
$$\text{s.t.}\begin{cases} y_i[\boldsymbol{w}\cdot\boldsymbol{x}_i-b]+\xi_i\geqslant 1, & \eta_i\geqslant 0\\ \boldsymbol{w}\cdot\boldsymbol{x}_j-b+\eta_j+M(1-d_j)\geqslant 1, & \eta_j\geqslant 0\\ -(\boldsymbol{w}\cdot\boldsymbol{x}_i-b)+z_j+Md_j\geqslant 1, & z_j\geqslant 0, d_j=\{0,1\} \end{cases} \tag{8}$$

其中 $M>0$ 为常量.

同年，Joachims[35]针对文本分类问题提出了直推式支持向量机算法 TSVM，其基本思路与 S3VM 所要解决的最优问题相似，因而这两种算法的概念可以相互替换，但 TSVM 是从文本分类的背景下提出的，更侧重于直推式的概念，即它仅考虑一个待定的测试数据集，并尽量减少这个测试集的错分率，而不考虑一般的情况. 有别于传统 SVM 能够被现有凸优化技术有效解决，S3VM 和 TSVM 的目标函数都具有非凸性，这使得它们难以直接求解最优解. 为了解决这个问题，学者们提出了许多优化方法，包括分支定界法、凹凸优化法、半定规划方法以及判定性模拟退火方法等，但这些方法的计算复杂度会随着样本数目的增加显著增加.

Zhao 等人[36]提出了割平面半监督支持向量机算法 CutS3VM，通过构造了一个嵌套序列，使得该序列中的每个优化问题都可以使用约束凹凸过程(Concave-Convex Procedure，CCCP)进行有效求解；此外，该算法通过理论分析证明了其计算时间与样本大小和稀疏度呈线性关系，在保证算法精度的同时有效降低了算法的计算复杂度. Li 等人[37]提出了一种快速的半监督支持向量机算法 MeanS3VM，与需要估计所有未标记样本标记的 S3VM 不同，它只需通过大间隔准则来估计类中心，并利用凸松弛算法和交替优化算法来解决非凸性问题，获得了两个版本的 MeanS3VM：MeanS3VM-mkl 和 MeanS3VM-iter，发现随着数据集规模的增大其计算效率比 TSVM 快 100 倍，当未标记数据的类中心已知时，算法的分类性能可近似于获得所有未标记样本标记的监督 SVM 算法. 但 MeanS3VM 通过随机选取无标记样本来确定标签均值来训练分类器严重影响了算法的稳定性，针对这个问题，Tian 等人[38]提出了一种改进的标签均值半监督支持向量机算法 CMeanS3VM，该算法将 MeanS3VM 算法中标签均值替换为聚类标签均值，并修改原算法中对

无标记样本惩罚项的约束条件,不仅明显提高了算法的正确率,还提高了算法的稳定性.

针对不同的错误分类可能会导致不同的代价,Li 等人[39]还提出了一种 MeanS3VM 的代价敏感版本 CS4VM,将代价敏感学习与半监督学习相结合,通过优化有标记和未标记样本上的总体代价,有效地减少了样本总体的错误代价.针对 S3VM 处理数据规模的可扩展性进行研究,Li 等人[40]还提出了一种基于标记生成的 WELLSVM 算法,它是一种能处理大规模数据且对具有全局理论保证的半监督支持向量机方法,通过不断产生最违反的标记赋值来最大化间隔,利用多核学习获得这些标记赋值的最优线性组合,使得 WELLSVM 可形式化为 S3VM 的凸松弛,保证了求解结果的全局性;另外,WELLSVM 的求解涉及一系列两类 SVM 子问题,从而可以很容易利用目前最好的可扩展 SVM 软件包来处理大规模问题.Geng 等人[41]提出了一种适用于 S3VM 的三重随机梯度算法 TSGS3VM,它在每次迭代中对标记样本和未标记样本以及随机特征进行采样,以计算三重随机梯度,并使用近似梯度来更新求解,通过对 TSGS3VM 进行了新的理论分析,保证了在弱假设下,对于一般的非凸学习问题,TSGS3VM 能够以次线性收敛速度收敛到一个平稳点.Wang 等人[42]基于 Bézier 函数对非光滑项的逼近性质,提出了一类新的 Bézier 光滑半监督支持向量机算法 BS4VM,由于这种近似,可以使用快速拟牛顿法求解 BS4VM 来减少计算时间;另外针对不同规模的数据集,通过选择最佳平滑函数,增强了 S3VM 在非线性情况下的泛化能力和鲁棒性.

当标记样本有限时,我们期望通过探索未标记样本来提高模型的性能.然而很多文献的实验结果表明,利用未标记样本可能会导致性能下降,在某些情况下甚至比只使用标记样本获得的性能还要差.Li 等人[43]发现对于给定的少量标记样本和大量未标记样本,可能会出现多个大间隔低密度划分,如果选择一个错误的低密度划分则会造成算法性能的大幅降低;针对这种情况,作者提出了一种安全的半监督支持向量机算法 S4VM,对于多个间隔较大的边界,对未标记样本的类别进行优化,使得算法在最坏情况下性能得到最大限度的提升.Wang 等人[44]发现当来自不同类别的样本严重重叠时,聚类假设不能很好地反映真实数据分布;基于聚类调整的聚类假设,作者利用无监督聚类方法,通过聚类调整的聚类假设计了一种新的安全半监督分类算法 ACA-S3VM,该算法考虑了学习过程中单个未标记样本到分布边界的距离,将聚类边界中的样本分类为类边界;因此,在样本严重重叠的情况下,类边界被引导通过簇边界而不是低密度区域,进一步提高学习安全性.安全半监督方法 S4VM 和 ACA-SVM 旨在对无标记样本的利用,但忽略了对标记样本隐含的监督信息的利用.为此,Huang 等人[45]提出了一种调整聚类假设联合成对约束半监督分类方法 ACA-JPA-SVM,它不仅继承了 ACA-SVM 算法的安全性,还通过成对约束将有标记样本的类标签转化为成对约束信息,弥补了算法对监督信息的利用不足,该方法不仅继承和缓解了不同样本严重重叠时分类边界造成的不安全学习情况,还在一定程度上提升了算法性能.

此外还有学者将基于聚类假设的半监督信息用于核函数的构造中,与 SVM 相结合,分别提出了基于谱聚类核[46]以及基于高斯混合模型核[47]的半监督 SVM 方法.

3.1.2 流形假设

流形假设认为在相同局部邻域内的样本性质相似,那么样本标记也应该相似[32].根据这一假设,未标记样本令样本空间更为密集,更有助于局部区域特征的描述,从而使决策函数能够更好地与样本进行拟合.流形正则化[38]是建立在流形假设上的一种半监督学习算法框架,它假设样本分布在外围空间的子流形上,可利用大量的无标记样本估计出样本数据的内在流形结构,通过将流形结构信息嵌入到分类器中,既保持了分类器光滑性又保持了样本的流形结构.流形正则化框架可以表示为

$$\begin{aligned} f^{*}&=\arg\min\frac{1}{l}\sum_{i=1}^{l}V(\boldsymbol{x}_i,y_i,f)+\gamma_1\|f\|_K^2+\gamma_2\|f\|_I^2\\ &=\arg\min\frac{1}{l}\sum_{i=1}^{l}V(\boldsymbol{x}_i,y_i,f)+\gamma_1\|f\|_K^2+\frac{\gamma_2}{(u+l)^2}f^{\mathrm{T}}\boldsymbol{L}f \end{aligned} \tag{9}$$

其中,$V(\boldsymbol{x}_i,y_i,f)$表示损失函数,用于度量期望输出与实际输出之间的损失值.正则化参数 γ_1 和 γ_2 分别控制函数 f 在再生核希尔伯特空间中和数据流形结构上的复杂性,$\|f\|_K^2$和$\|f\|_I^2$分别为再生核希尔伯特空间的惩罚项和流形正则化项.$\boldsymbol{L}$ 为拉普拉斯矩阵,且 $f=[f(\boldsymbol{x}_1),f(\boldsymbol{x}_2),\cdots,f(\boldsymbol{x}_{l+u})]$.

2006 年,Belkin 等人[32]在上述流形正则化框架中令损失函数为铰链损失函数,提出了一种典型的基于流形正则化的半监督学习方法——拉普拉斯支持向量机(Laplacian Support Vector Machine,LapSVM),LapSVM 的学习问题可以表示为以下优

化问题：

$$f^*=\arg\min_{f\in H_K}\frac{1}{l}\sum_{i=1}^{l}(1-y_if(\boldsymbol{x}_i))+\gamma_1\|f\|_K^2+\frac{\gamma_2}{(u+l)^2}f^{\mathrm{T}}\boldsymbol{L}f \tag{10}$$

该问题的解可以表示成如下形式：

$$f^*(x)=\sum_{i=1}^{l+u}\alpha_i^*K(x,\boldsymbol{x}_i) \tag{11}$$

因此，LapSVM 的学习问题可以等价为

$$\min_{\alpha\in R^{l+u},\xi\in R^l}\frac{1}{l}\sum_{i=1}^{l}\xi_i+\gamma_1\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{K}\boldsymbol{\alpha}+\frac{\gamma_2}{(l+u)^2}\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{K}\boldsymbol{L}\boldsymbol{K}\boldsymbol{\alpha}$$

$$\text{s.t.}\begin{cases}y_i\left(\sum_{j=1}^{l+u}\alpha_jK(\boldsymbol{x}_i,\boldsymbol{x}_j)+b\right)\geq 1-\xi_i\\ \xi_i\geq 0,\ i=1,2,\cdots,l\end{cases} \tag{12}$$

LapSVM 将拉普拉斯正则项与 SVM 优化目标函数相结合，把样本数据的流形结构信息嵌入到传统的 SVM 算法中，在继承了传统 SVM 优点的同时，又在一定程度上解决了训练不足的问题. 由于拉普拉斯矩阵的运算开销较大，未标记样本的增加使得矩阵的维数增大，显著地增加了时间以及内存的消耗，因此 LapSVM 在大规模数据集中难以得到广泛的使用. 受原始空间求解 SVM 方法的启发，Melacci 等人[48]提出了一种原始空间下的半监督学习算法 LapSVM-PCG，它通过预优共轭梯度方法来训练原始 LapSVM，并采用早期停止策略来加速训练，使得该算法能够快速计算近似解，从而大大减少了训练时间. Qi 等人[49]提出了一种快速的半监督学习算法 FLapSVM，它不需要处理额外的矩阵，也不需要处理与变量切换相关的计算负担，通过传统的 SVM 过程进行推导，将核函数直接应用到优化模型中，利用连续超松弛(Successive Over-Relaxation，SOR)技术对目标函数求解，使其线性收敛，能更有效处理大规模数据问题. Yang 等人[50]从问题本身的规模出发，提出了一种有效的安全筛选方法来加速 LapSVM，该方法通过结合 KKT 条件和变分不等式得到可行集，可以在求解优化问题之前剔除了大量的训练样本，既有效的提升了算法的计算效率，同时也保证了最优解.

LapSVM 采用铰链损失函数作为损失函数，该函数不是二次可微的，可能会对噪声标记的样本进行过多的惩罚. 为降低 LapSVM 对噪声数据的敏感性，Khan 等人[51]提出了 Huber-LapSVM 算法，该算法在 LapSVM 中引入了 Huber 铰链损失函数，Huber 铰链损失函数是铰链损失函数的可微近似，在衡量误分类时会使训练更加鲁棒；另外，作者采用了 Melacci 等人[48]提出的预优共轭梯度方法提高该算法的效率，不仅减少了训练时间，还保证了与原始问题相近的精确度. Dong 等人[52]使用 Laplace 核函数来度量两个变量之间基于相关熵的相似性，将导出的相关熵损失函数(称为 LK 损失)，应用于 LapSVM，以构建鲁棒半监督分类器 LK-LapSVM，通过分析 LK 损失的鲁棒性、对称性、有界性、Fisher 一致性以及渐近逼近性，引入了 LK 损失的非对称版本，来获得更好的鲁棒性；此外，还利用 CCCP 技术迭代处理该损失函数的非凸性，算法不仅具有更好的鲁棒性，还具有更好的泛化性能. Pei 等人[53]提出了一种新的具有非对称平方损失的半监督支持向量机算法 asy-LapSVM，它不仅能够充分利用丰富的未标记样本，并且对噪声样本更为鲁棒；为了加快训练过程，作者还提出了一种简单有效的函数迭代方法来代替传统的二次规划法来解决所涉及的优化问题，算法取得了显著的性能.

Liu 等人[54]发现 LapSVM 在构建图的过程中忽略了样本的类别分布信息，提出了一种基于局部行为相似性的半监督学习算法 LBS-LapSVM，该算法根据人类行为认知思想的特点，通过结合标记信息的行为相似性权重来重新构建图，更好地提取出样本内在的几何结构信息，同时利用局部分布参数来避免热核参数选择问题，有效地改善了算法的性能. Sun 等人[55]发现 LapSVM 中图的构造仅考虑两两样本之间的序对关系，而忽略样本之间多元复杂关系，利用超图来代替图设计了一种超图正则化的半监督支持向量机算法 HGSVM，由于构造超图需要更高的计算复杂度，作者还提出了一种边界样本的方法来筛选信息量更丰富的未标记样本，从而减少样本的训练规模，所提出的 fast-HGSVM 算法不仅在提高分类准确率的同时，也保证不错的计算效率.

此外，受 PSVM[56-57]、TWSVM[58-59]以及 LST-SVM[60-61]算法的启发，学者们对这些改进的 SVM 算法与流形正则化相结合，提出了 LapTSVM[62]、Lap-MNPSVM[63]、SSSR-LTPSVM[64]、FLap-TW-SVM[65]、HSR-LSTSVM[66]以及 Lap-LpTSVM[67]的方法，降低了算法的复杂度并提高其性能.

表 1 对基于半监督学习的支持向量机算法进行总结和对比.

表 1 基于半监督学习的支持向量机算法的对比

划分类型	算法名称	特 点	不 足
聚类假设	S3VM[34]	在 SVM 基础上加入两个用于计算未标记样本错分率的约束	非凸性;计算复杂度高
	TSVM[35]	基于直推式学习思想,将无标签样本中隐含的分布信息引入到 SVM 训练中	算法执行之前需要确定无标签样本中的正样本数目;计算复杂度高
	CutS3VM[36]	构造一个嵌套序列,使该序列中的每个优化问题都可以使用约束 CCCP 有效地求解	非凸性;算法鲁棒性低;计算效率低
	MeanS3VM[37]	只需要估计标签均值训练,提高了算法的效率	算法准确率低且稳定性差
	CMeanS3VM[38]	采用聚类标签均值替换标签均值,提供了更多的类别信息,更适合多分类任务	类簇数目难以确定
	CS4VM[39]	代价敏感学习与半监督学习相结合,有效地减少了样本总体的错误代价	未考虑样本不平衡问题,噪声问题;忽略了样本间的结构信息
	WELLSVM[40]	通过不断产生最违反的标记赋值来最大化间隔,结合多核学习,保证了解的全局性,更适合大规模任务	侧重于二分类问题,未考虑多分类问题
	TSGS3VM[41]	利用三重随机梯度算法来解决非凸学习问题,并且能在弱假设条件下实现快速收敛	算法涉及的参数太多
	BS4VM[42]	利用一类新的 Bézier 函数来逼近 Hinge 损失函数,使目标函数光滑可微,并采用快速的拟牛顿法求解来减少计算时间	如何选择最优的 Bézier 函数
	S4VM[43]	构造多个候选低密度划分,选择最优划分在最坏情况下实现性能最大化提升	当无标签样本数量大,算法效率低;未考虑样本严重重叠的情况
	ACA-S3VM[44]	将未标记样本到类边界的距离引入到模型训练中,缓解了不同样本严重重叠在分类边界时可能造成不安全学习的情况	未充分利用标记样本中的隐含的监督信息
	ACA-JPC-S3VM[45]	结合成对约束信息,有效利用了标记样本隐含的监督信息,弥补了不安全造成的性能损失	如何调整成对约束正则化项参数;计算效率低
流形假设	LapSVM[32]	通过流形正则化项引入了无标签样本的内在几何结构信息,较好地解决了标记样本有限情况下训练不充分的问题	对噪声敏感;图的构建简单丢失了相关信息
	LapSVM-PCG[48]	在原始空间下,使用预共轭梯度和早期停止策略实现了快速求解	核矩阵的存储和运算随数据量增加而变大
	FLapSVM[49]	不需要计算逆矩阵,通过逐次超松弛技术使算法线性收敛,更适合大规模问题	未考虑多分类情况
	SSR- LapSVM[50]	通过变分不等式提出了一种有效的筛选方法,减少了样本的数量,还有效地解决了多参数问题	只能筛选有标记样本
	Huber-LapSVM[51]	利用连续可微的 Huber 损失使算法具有更强的鲁棒性,通过预优共轭梯度方法减少训练时间	未考虑多分类情况
	LK-LapSVM[52]	利用 LK 损失及其不对称形式增强算法鲁棒性,通过 CCCP 来迭代求解实现了线性收敛	算法的计算复杂度高
	asy-LapSVM[53]	利用非对称平方损失降低对噪声的敏感性,通过简单有效的函数迭代法快速求解优化问题	未考虑多分类情况
	LBS-LapSVM[54]	利用标记信息的行为相似性权值构造邻接图,引入局部分布参数解决了热核参数的选择	计算复杂度高,不适合大规模数据集
	HGSVM[55]	利用超图进一步挖掘样本之间的高阶几何关系,构造样本的多元流形结构,提高了算法的分类性能	超图的构建存在极高的计算复杂度
	fast- HGSVM[55]	利用边界样本来筛选信息量更丰富的未标记样本,缩减了样本的规模	如何选取更具启发性的标记样本
	其他[62-67]	将 PSVM,TWSVM,LSTSVM 与流形正则化结合,推广到半监督形式,提高算法的泛化性能	未考虑多分类情况

3.2 基于多示例学习的支持向量机算法

多示例学习是 Dietterich 等人[27]在研究药物分子活性预测问题上提出的概念,在该学习框架中,样本集由多个带标签的示例包组成,而每个包中又包含了若干个无标签的示例,若包中至少含有一个正示例则标记为正,否则为负. 给定一个训练包的集合 $D=\{(B_1,Y_1),(B_2,Y_2),\cdots,(B_m,Y_m)\}$,$B_I$表示一个示例包,$Y_I$表示示例包 B_I的标签. $B_I=\{\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_n\}$,其中 $\boldsymbol{x}_i$表示示例包的一个示例. 设 $x\in R^d$ 为一个 d 维的示例空间,$Y=\{-1,1\}$是标记空间,多示例学习的目标是通过对 D 的学习,得到一个映射函数 $f:2^x\rightarrow Y$ 来预测未知包的标签.

多示例学习中样本的层次化表示结构使其更能反映实际问题的内在逻辑结构,因此在医学诊断[68-70]、文本分类[71]以及目标检测[72-74]等多个领域有着广泛的应用. 目前,学者们已经提出了大量的多示例学习算法,主要可分为示例水平空间、包水平空间以及嵌入空间的方法.

3.2.1 示例水平空间

这类方法采用训练集中的所有示例来学习一个

示例水平空间的分类器，并通过包中示例的标记来推断整个包的标记，其基本思想如图 5 所示.

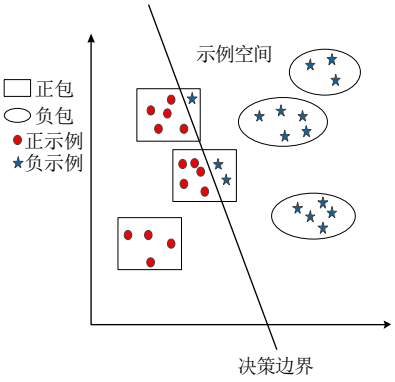

图 5　基于示例水平空间方法的示意图[75]

mi-SVM[76]是 Andrews 等人对 SVM 在基于示例水平空间的多示例学习上的推广，它假定训练集中所有正包中的示例全部为正，通过这些“正”示例样本和负包中的所有示例样本来训练 SVM 分类器，利用训练后的 SVM 重新标记正包中的示例；若一个正包中的所有示例均被判定为负，则标记该包中具有最大决策函数值的示例为正；然后再通过这些新标记的正示例和负包中的所有示例重新训练 SVM，反复训练和标记，直到所有训练示例的标签不再发生变化为止. mi-SVM 可表示如下：

$$\min_{\{y_i\}}\min_{\{\boldsymbol{w},b,\xi\}}\frac{1}{2}\|\boldsymbol{w}\|^2+C\sum_i\xi_i$$

$$\text{s.t. } \forall i: y_i(\langle\boldsymbol{w},\boldsymbol{x}_i\rangle+b)\geq 1-\xi_i,\ \xi_i\geq 0,\ y_i\in\{-1,1\}$$

$$\sum_{i\in I}\frac{y_i+1}{2}\geq 1,\ \forall I\ \text{ s.t. } Y_I=1,\ \text{and } y_i=-1,\ \forall I\ Y_I=-1 \tag{13}$$

其中，I 指包的序号，i 指包中示例的序号，Y_I 表示包 I 的标签，y_i 表示示例 i 的标签.

由于 mi-SVM 算法待求解的目标函数是一个混合整数非线性规划问题，所以不能直接求其代数解，Astorino 等人[77]引入了拉格朗日松弛技术，提出了一种用于二元多示例分类的支持向量机方法 mi-SVM*，该算法利用拉格朗日松弛技术的特性，将一个复杂的混合整数非线性问题转换为一组相对简单的拉格朗日子问题，同时采用块坐标下降的方法近似求解有限个拉格朗日子问题，通过拉格朗日子问题的最优解为原始问题提供了最优解，在分类性能和计算效率之间取得了满意的结果. mi-SVM 在每次迭代中需要求解一个 SVM 二次规划子问题，若子问题的数量太多，则总体计算工作量会显著增加. 为此，Avolio 等人[78]提出了一个新的示例空间算法 mi-SPSVM，它是 SVM 和 PSVM 模型之间的折中，旨在生成一个放置于两个平行超平面中间的超平面：其中第一个超平面是聚集正包示例的近端正平面，而第二个超平面是构成负包示例的支持超平面，使得算法不仅继承了 SVM 的高精度还满足了 PSVM 高效率. Yuan 等人[79]将基于 CCCP 算法的安全示例筛选规则引入到多示例支持向量机中，通过识别非活动示例可以显著减轻存储负担. 该规则分为两个步骤构建：首先设计动态筛选规则来筛选凸优化子问题中的非活动示例，其次针对示例标签可能发生变化的问题，设计了在 CCCP 迭代之间传播非活动示例的规则. 为了进一步提高求解效率，在内部求解器中引入了智能双坐标下降法 SDCDM，该方法跳过了许多导致当前迭代没有变化的更新，并证明了 SDCDM 的安全性. Li 等人[80]提出了 I-KI-SVM，在每一次迭代中，该方法基于示例水平生成一个最违规的关键示例来最大化间隔，并通过多核学习将其进行组合，由于 I-KI-SVM 的求解过程设计到一系列标准的 SVM 子问题，从而可以方便地利用可扩展 SVM 软件包来处理大规模问题.

包中示例的模糊性问题是多示例学习中的关键难点之一，即仅知道正包中至少有一个示例是正示例，但并不清楚具体是哪些示例是正的. 针对这个问题，Xiao 等人[81]提出了一种基于相似性的多示例学习算法 SMILE，通过考虑模糊示例与正、负类示例的相似性来处理模糊示例，具体地说，从正包中选取示例的子集作为正候选示例，分别计算每个示例与正、负类示例的相似性权重；将这些模糊示例及其相似性权重输入到 SVM 的训练过程中，另外利用一个启发式算法更新正候选示例及相似度权重，来细化分类边界，该方法不仅具有更高的分类精度，还对标记噪声具有更强的鲁棒性.

3.2.2　包水平空间

这类方法将训练集的每个训练包看作一个独立的整体，通过获得样本的包级别特征来学习相应的包级分类器，其基本思想如图 6 所示.

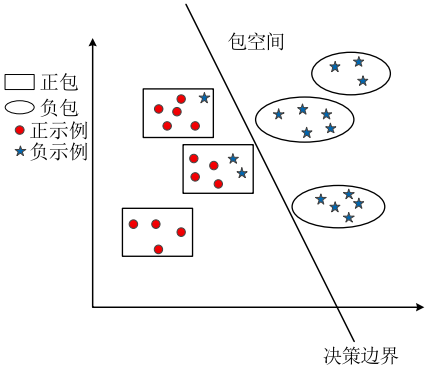

图 6　基于包水平空间方法的示意图[75]

与 mi-SVM 类似，MI-SVM[76] 同样是 Andrew 等人在 SVM 基础上设计的算法，但它是一种基于包空间水平的多示例方法. MI-SVM 首先将包中所有示例的特征相加取均值，得到相应的单特征向量，然后利用这些得到单特征向量训练 SVM 分类器，对正包中的所有示例进行预测，将正包中具有最大决策函数值的示例更新为该包的单特征向量，重新训练 SVM 分类器；经过反复训练和标记，直到所有示例包的单特征向量均不再发生变化. 因此，MI-SVM 的目标函数如下：

$$\min_{\{\boldsymbol{w},b,\xi\}} \frac{1}{2}\|\boldsymbol{w}\|^2 + C\sum_I \xi_i$$

$$\text{s.t. } \forall I: Y_I \max_{\{i\in I\}}(\langle \boldsymbol{w}, \boldsymbol{x}_i\rangle + b) \geqslant 1-\xi_I,\ \xi_I \geqslant 0 \quad (14)$$

由于 MI-SVM 的目标函数是一个混合整数优化问题，Andrews 等人[76] 采用启发式方法来求解由此产生的非线性优化问题，但并没有给出收敛性证明. 而 Cheung 等人[82] 则是采用约束 CCCP 来求解 MI-SVM 的目标函数，提出了 MI-SVM*，它通过定义包的输出与其相关示例之间的损失函数，使包和示例直接参与优化过程，为多示例学习提供了一个更完整的正则化框架，采用 CCCP 过程来解决非线性优化问题保证了收敛到局部最优. 在 MI-SVM 算法的基础上，Melki 等人[83] 提出了一种多示例代表分类算法 MIRSVM，它利用基于包级信息的包代表选择器来训练 SVM，通过识别高度影响分类的示例作为包代表，寻找最佳的分类超平面；该方法通过允许正包和负包最多拥有一个代表，以克服可能出现的类不平衡问题，此外，算法还具有良好的收敛性和可扩展性，适合于大规模的学习任务.

相比于上述基于示例水平空间的 I-KI-SVM，Li 等人[80] 还提出了 B-KI-SVM，它基于包水平来选择一个关键示例，即利用每个负包中示例的平均值作为关键示例，从而避免优化过程中产生大量约束的问题. Cheplygina 等人[84] 提出一种通用的多示例差异化算法 MIND，该方法利用训练集中每个包与其他包的差异向量来表示每个包，并将这些差异视为特征表示，因此多示例学习问题可被转化为监督学习问题，Cheplygina 等人还证明了不同的相异性定义对示例的信息性有不同的隐含假设，其中基于包之间最小示例距离平均值的差异性表现出了良好的性能，通过在 SVM 分类器上进行学习，实现更好的分类性能和更高的计算效率.

3.2.3 嵌入空间

与基于包水平空间的方法不同，基于嵌入空间的方法通过一种隐形的方式来抽取全局包级别信息，即利用某种映射函数将示例包从原始特征空间映射到新的特征空间中，然后在新特征空间中重新训练学习器，其基本思想如图 7 所示.

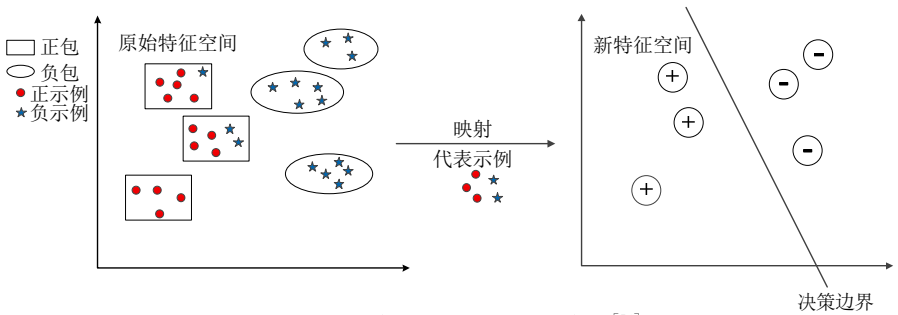

图 7 基于嵌入空间方法的示意图[75]

2004 年，Chen 等人[85] 提出了 DD-SVM 算法，它利用多样性密度(Diverse Density，DD)算法选择训练集中具有局部多样性密度最大的示例作为原型，并根据这些示例原型定义一个非线性映射，将每个示例包映射到一个新的特征空间中，即包特征空间. 包到包特征空间的映射公式如下：

$$\phi(B_i) = \begin{bmatrix} \min\limits_{j=1,\cdots,N_I} \|\boldsymbol{x}_{ij} - x_1^*\|_{w_1^*} \\ \min\limits_{j=1,\cdots,N_I} \|\boldsymbol{x}_{ij} - x_2^*\|_{w_2^*} \\ \cdots \\ \min\limits_{j=1,\cdots,N_I} \|\boldsymbol{x}_{ij} - x_n^*\|_{w_n^*} \end{bmatrix} \quad (15)$$

通过上述映射公式，多示例学习问题可被转化为单示例学习问题，因此，多示例学习问题可表示为下面的二次优化问题：

$$\alpha^* = \arg\max \sum_{i-1}^{l}\alpha_i - \frac{1}{2}\sum_{i,j=1}^{l} y_i y_j \alpha_i \alpha_j K(\phi(B_i), \phi(B_j))$$

$$\text{s.t. } \begin{cases} \sum\limits_{i=1}^{l} y_i \alpha_i = 0 \\ C \geqslant \alpha_i \geqslant 0,\ i = 1, \cdots, l \end{cases} \quad (16)$$

通过求解上述优化问题，将得到的 SVM 分类器对新包的标签进行预测.

由于 DD-SVM 中的 DD 算法训练速度比较慢，

使得整个算法的速度受到很大的影响.针对这个问题,Chen等人[86]提出了MILES算法,该算法没有示例标签与包标签相关联的假设,即没有强制约束负包中所有的示例均为负标签,通过定义一个示例包和示例之间的相似性度量,将包映射到基于示例的特征空间中,然后使用1范数SVM进行训练,剔除了大量冗余或不相关的特征,从而得到一个稀疏解,不仅提高了算法训练速度,还在一定程度上降低了标签噪声的敏感性.多示例学习问题的复杂性在很大程度上取决于训练数据集中示例的数量,如何设计有效的示例选择技术以加快训练过程是非常重要的.Li等人[87]根据示例水平和包水平的特征表示方案提出了两种基于消歧的多示例学习算法MILD_I和MILD_B,利用条件概率模型在各正包中预测一个代表性正示例,在每个负包中选择一个代表性负示例,并将代表性示例基于不同的特征表示方案映射到特征空间上,最后利用SVM设计分类器进行求解,算法具有良好的效率、鲁棒性以及准确性.Fu等人[88]提出了一种基于自适应示例选择的多示例学习算法MILIS,该方法利用核密度估算函数来实现初始示例选择,通过迭代的方式将示例选择和SVM分类器学习的步骤结合从而保证收敛,实现了更高的训练速度,更适用于大规模多示例学习问题.Huang等人[89]合并了多种包表示方法,提出了BDR-SVM,他利用提出的包不相似正则化框架对隐式嵌入空间中数据的局部几何结构建模,并将其纳入SVM分类器中,另外由于BDR框架依赖于隐式和显式包表示,为有效地保留多示例样本中的判别信息,还提出了一种基于因子分析的Fisher评分的方法,提出的BDR方法可以通过同时使用多个包表示来实现算法性能的即时改进.在不同包表示的帮助下,BDR方法可以适应各种实际任务.

表2对基于多示例学习的支持向量机算法进行总结和对比.

表2　基于多示例学习的支持向量机算法比较

划分类型	算法名称	特　点	不　足
示例水平空间	mi-SVM[76]	将SVM引入多示例学习问题,基于包中示例最大化分类边界	非凸性,计算复杂度高,混合整数非线性问题
	mi-SVM*[77]	利用拉格朗日松弛技术和块坐标下降法来获得最优解	仅适合二分类多示例问题
	mi-SPSVM[78]	结合SVM和PSVM,生成一个位于两个平行超平面中间的分离超平面,不仅具有高精度还满足高效率	未将算法扩展到核变换,仅支持线性分离
	SSR-SDCDM[79]	引入了安全筛选规则显著减轻存储负担,为内部求解器引入了SDCDM方法进一步提高求解效率	内部求解器依赖于CCCP,计算负担仍然很大
	I-KI-SVM[80]	基于示例水平生成关键示例,并联合多核学习对一个不满足要求的关键示例进行处理求出最优解	算法稳定性差
	SMILE[81]	考虑模糊示例与类的相似性,分类器具有更好的判别能力和抗噪性	计算复杂度高,不适合大规模数据集
包水平空间	MI-SVM[76]	将SVM引入多示例学习问题,基于包最大化分类边界	过早收敛,易陷入局部最优
	MI-SVM*[82]	定义包输出与相关示例的损失函数,采用CCCP保证了收敛性	依赖于CCCP算法,计算负担增加
	MIRSVM[83]	通过识别高度影响分类的示例来寻找最佳超平面,仅允许正包和负包最多有一个代表性,消除了可能的类不平衡问题	未考虑多分类情况
	B-KI-SVM[80]	基于包水平生成关键示例,并联合多核学习对一个不满足要求的关键示例进行处理求出最优解	算法稳定性差
	MIND[84]	利用训练集中包之间的相异性来表示每个包,将多示例问题被转换为监督学习问题,扩展性好	如何确定最佳的相异性表示
嵌入空间	DD-SVM[85]	利用DD算法将多示例问题转为监督问题	计算复杂度极高
	MILES[86]	不强加示例标签与包标签相关联的假设,采用1范数SVM剔除冗余和不相关特征,提高计算效率,还提高噪声鲁棒性	特征向量不稀疏,所需存储空间大
	MILD[87]	通过一种新的消歧方法选择正包中真正示例,采用示例级和包级两种特征表示方案,效率高,鲁棒性好,且扩展性强	示例选择中忽略了负示例包中负示例原型
	MILIS[88]	采用基于负示例分布的示例选择方法降低问题规模,通过交替优化框架将示例选择和分类器相结合,保证了收敛性	算法依赖的迭代优化机制耗时长,执行效率低,难以应用大规模问题
	BDR-SVM[89]	利用因子分析的Fisher评分有效保留多示例样本中的判别信息,通过一种包不相似正则化框架,提高算法性能	计算复杂度高

3.3　基于多标记学习的支持向量机算法

不同于传统的单标记学习,多标记学习主要用于解决单个样本同时属于多个标签的问题[29-30].多标记学习可定义如下:假设$X=R^d$表示d维特征空间,$Y=\{y_1,y_2,\cdots,y_q\}$表示标记集合,多标记样本集$D=\{(\boldsymbol{x}_1,Y_1),(\boldsymbol{x}_2,Y_2),\cdots,(\boldsymbol{x}_m,Y_m)\}$,其中$\boldsymbol{x}_i\in X$表示第$i$个样本,$Y_i\in Y$表示$\boldsymbol{x}_i$的相关标记集合.多标记学习的目标是通过训练样本集$D$,学习到一个映射$f:X\to 2^Y$来预测未知样本的标签集合.近年来,研究学者们提出了许多的多标记学习算法,其理论成

果在情感识别[90-92]、图像标注[93]、文本分类[94-96]、生物信息[97-98]等领域也得到了广泛的应用. 目前,多标记学习算法可大致分为两类:问题转换法以及算法自适应法.

3.3.1 问题转换

问题转换法旨在将多标记分类任务转换为其他成熟的学习问题,如单标记分类和标记排序. Boutell 等人[99]提出了第一个用于多标签分类的 SVM 算法 BR-SVM,它借助 Binary Relevance(BR)策略,将具有 k 个标签的样本集划分为 k 个二分类样本集,其中每个二分类样本集是由包含指定标签的正样本和不包含该指定标签的负样本组成,并且每个二分类样本集的样本数目与多标签样本集的样本数目相同,然后为每个二分类样本集分别训练一个独立的二分类 SVM 分类器. 受 DC-SVM[100]算法的启发,Liu 等人[101]提出了 MLDC-SVM 算法,它利用基于分而治之策略的 DC-SVM 替代 SVM 达到了全局的快速收敛,并通过一对多分解技术将 DC-SVM 拓展到多分类问题上,实现了快速多标签分类;另外算法还采用 DEC[100]方法改进 DC-SVM,改善了标签数据不平衡的问题. Classifier Chains(CC)[102]是 BR 的扩展,它将 BR 策略产生的分类器连成一条链,下一个分类器的特征空间用前一个分类器的预测标签进行扩展,通过在分类器之间传递信息,有效地利用了标签的相关性. Ensemble of Classifier Chains (ECC)[103]通过集成多条随机标记序列的 CC,减轻 CC 中分类器的随机顺序给性能造成的影响. Calibrated Label Ranking(CLR)[104]是一种将多标记任务转换为标签排序任务的策略,通过引入一个人工校准标签将每个示例中的相关标签和不相关标签进行分离. Label Powerset(LP)[105]是一种多标记任务转化为多类分类任务的策略,但 LP 不能泛化新的标签集合,且计算复杂度高,Tsoumakas 等人[105]通过结合 LP 策略和集成学习,提出了 RAKEL 策略,它通过在标签集上大小为 k 的小随机子集上训练 LP 分类器来考虑标签之间的关联,避免了 LP 的计算复杂度问题. Yapp 等人[106]综合分析了上述五种转换策略:BR、CC、ECC、CLR 和 RAKEL 在四种基分类器的分类性能,这四种基分类器包括 k 最近邻、决策树、朴素贝叶斯、支持向量机,实验结果表明支持向量机在这五种问题转换策略上能取得最好的分类性能.

3.3.2 算法自适应

算法自适应法旨在改进传统的监督学习算法,使其直接应用于多标记样本的学习. Elisseeff 等人[107]利用相关标签和不相关标签之间的成对关系充分描述了单个样本的标签相关性,通过扩展多类支持向量机并将排序损失作为其经验损失,提出了一种多标记分类的支持向量机算法 RankSVM. 该算法的基本思想是通过最小化 L2 范数正则化项来最大化样本中相关标签与不相关标签之间的差值,通过最小化排序损失使任何相关标签的排序都高于任何不相关标签. RankSVM 需要解决的优化问题为

$$\min_{\boldsymbol{w},\boldsymbol{\xi}} \frac{1}{2}\sum_{k=1}^{q}\|\boldsymbol{w}_k\|^2 + C\sum_{i=1}^{m}\frac{1}{|Y_i||\overline{Y}_i|}\sum_{(y_j,y_j)\in Y_i\times\overline{Y}_i}\xi_{ijk}$$
$$\text{s.t.}\begin{cases}\langle \boldsymbol{w}_j,\boldsymbol{x}_i\rangle + b_j - \langle \boldsymbol{w}_k,\boldsymbol{x}_i\rangle - b_k \geqslant 1-\xi_{ijk},\\ \xi_{ijk}\geqslant 0,\ 1\leqslant i\leqslant m\end{cases} \tag{17}$$

其中 Y_i 为样本的标签集,$\overline{Y}_i$ 为其补集,q 表示样本标签数目.

此外,RankSVM 还需要一个阈值学习步骤. 设阈值函数 $t(x)=\langle \boldsymbol{w}^*,\boldsymbol{f}^*(x)\rangle + b^*$,其中 $\boldsymbol{w}^*\in R^q$ 是权值向量,$\boldsymbol{f}^*(x)=(f(x,y_1),f(x,y_2),\cdots,f(x,y_q))^{\mathrm{T}}\in R^q$ 是属性向量,$b^*\in R$ 是偏置. RankSVM 可通过线性最小二乘法来求解对应参数:

$$\min_{\{\boldsymbol{w}^*,b^*\}}\sum_{i=1}^{m}(\langle \boldsymbol{w}^*,\boldsymbol{f}^*(\boldsymbol{x}_i)\rangle + b^* - s(\boldsymbol{x}_i))^2 \tag{18}$$

其中 $s(\boldsymbol{x}_i)$ 的取值范围为一个实数区间,

$$s(\boldsymbol{x}_i)=\arg\min_{a\in R}(|\{y_i\,|\,y_i\in Y_i, f(\boldsymbol{x}_i,y_i)\leqslant a\}| + |\{y_k\,|\,y_k\in\overline{Y}_i, f(\boldsymbol{x}_i,y_i)\geqslant a\}|) \tag{19}$$

因此,可得多标记分类器:

$$h(x)=\{y_j\,|\,\langle \boldsymbol{w}_j,x\rangle + b_j > \langle \boldsymbol{w}^*,\boldsymbol{f}^*(x)\rangle + b^*, 1\leqslant j\leqslant q\} \tag{20}$$

RankSVM 不仅存在着极高的计算复杂度,而且没有定义相关标签的内部零点. 针对这个问题,Xu 等人[108]提出 RankSVMz 算法,它通过添加一个零标签将相关标签和不相关标签分隔开,并定义了一种新的排序损失函数,简化了原有的 RankSVM 形式;另外,该算法采用 Frank-Wolfe 方法来求解优化问题,并将整个线性规划问题通过一对多分解技术划分为一系列子问题,计算成本显著降低. 为进一步加快 RankSVMz 算法的训练过程,Xu 等人[109]进而提出了 RankSVMzR 算法,它利用随机块坐标下降法[110]替代 RankSVMz 中的 Frank-Wolfe 方法,根据不相交的等式约束,通过一对多分解技术将大规模二次规划问题分解成一系列小规模二次规划子问题,利用二分类 SVM 中广泛使用的序列最小化优化算法来有效地解决每个子问题,实现了比

RankSVMz 更低的时间复杂度和更好的分类性能. 由于 RankSVM 不能删除高维问题的冗余特征，且对噪声点敏感，Wang 等人[111]提出了一种具有 Pinball 损失的稀疏弹性网络多标记 RankSVM 算法 pin-ENR 来提高泛化性；该算法利用 Pinball 损失[112]替代铰链损失，降低其对噪声点的敏感性，同时它采用稀疏弹性网络正则化代替 L2 范数正则化实现变量选择；此外，为了加速 pin-ENR，作者基于原目标函数和对偶目标函数的强凸性，构造了一个安全的特征和标签对消除规则来加速算法的训练过程，通过在训练前删除大量非活动特征和标签对，在不牺牲分类准确性的同时提高了算法的计算效率. Wu 等人[113]认为 RanSVM 阈值学习过程中阈值化的堆叠方式易造成错误累积，从而降低最终分类性能，提出了一种新的多标签分类算法 RBRL，它采用 BR 策略将阈值化步骤融入 RankSVM 的学习过程，同时利用参数矩阵的低秩约束进一步挖掘标签相关性；另外，通过加速近端梯度法来有效求解目标函数，不仅克服了 BR 存在类不平衡问题和忽略标签相关性的缺点，还避免了 RankSVM 中还需阈值学习步骤的过程. 尽管如此，解决大规模问题仍然是 RBRL 算法面临的挑战，Zhong 等人[114]采用了一种用于 RBRL 的子空间筛选规则来加速求解过程，提出了 SSR-RBRL，这是一种基于矩阵的稀疏表示及其最优性条件的迭代方法. 具体地说，在每次迭代中，通过矩阵分解和最优性条件删除了最优解中大多数系数为零的子空间，这使得需要解决的问题的规模能够显著减小；为了进一步提高求解速度，在不同阶段采用了近似奇异值分解和加速近端梯度来加快算法迭代过程.

此外，针对一些改进的 SVM 算法，研究学者们通过结合多标记学习，先后提出了 RankCVM[115]、RankLSVM[116]、MLTSVM[117]等多标记算法. Xu 等人[115]通过结合 RankSVM 和核向量机(Core Vector Machine，CVM)，设计并实现一种新的 SVM 型多标签分类算法 RankCVM，其优化形式与 CVM 相同，具有一个特殊的单纯形约束；当采用 Frank-Wolfe 方法时，RankCVM 在任何迭代中都存在解析解和步长，以及一些有用的代理解、梯度向量和目标函数值的递推公式，大大加快了 RankCVM 的训练过程，降低了计算量. Xu 等人[116]还提出了一种新型高效的 SVM 型多标签算法，通过在相关标签和不相关标签之间应用成对约束，并定义一个近似的排名损失，将二分类拉格朗日支持向量机推广到多标签形式 RankLSVM，得到一个仅有非负约束条件的严格凸二次规划问题；另外，该算法利用基于带有启发式收缩策略的随机块坐标下降法来加速训练过程，避免了一些不会导致排名损失的训练样本的更新，减少了支持向量的数量. 借助 TWSVM 的思想，Chen 等人[117]提出了第一个应用于多标签学习的非平行超平面 SVM 分类器 MLTSVM，它通过确定多个非平行超平面来捕获数据样本中嵌入的多标签信息，同时采用一种高效的 SOR 算法来解决 MLTSVM 中的二次规划问题；此外，根据样本到不同超平面的距离来设计决策函数，可以方便地为新样本指定标签集，不仅可以克服测试过程中的模糊情况，而且可以降低误分类的概率. 由于数据样本的结构信息可能包含用于训练分类器的有用的先验领域知识，Azad-Manjiri 等人[118]提出了一种用于多标签学习的结构最小二乘双支持向量机算法 ML-SLSTSVM，它扩展了 MLTSVM 方法，利用求解二次规划问题为每个标签寻找一个近端超平面，通过考虑训练样本的结构信息并使用最小二乘思想提高了算法的泛化性能和运行速度.

表 3 对基于多标记学习的支持向量机算法进行总结和对比.

表 3　基于多标记学习的支持向量机算法比较

划分类型	算法名称	特　点	不　足
问题转换法	BR-SVM[99]	利用 BR 策略将具有 k 个标签的问题转化为 k 个独立的二值分类问题，算法简单易实现	忽略标签之间的相互关系，无法应对标记空间类不平衡问题
	MLDC-SVM[101]	结合 BR 策略和 DEC 方法，不仅提高了算法的计算效率，还克服了标签样本不平衡的问题	未考虑样本标签间的相关性信息
	CC-SVM[106]	BR 策略的改进，通过标记链式排列，利用前序标记的预测值增强后序标记的预测性能	分类器排列的顺序对算法影响较大
	ECC-SVM[106]	CC 策略的改进，采取多个随机产生的不同标记序列集成，减轻分类器排列顺序问题的影响	学习空间中存在的大量冗余，计算复杂度高，无法适用大规模数据集
	CLR-SVM[106]	引入一个人工校准标签分离每个样本中的相关标签和不相关标签，将多标签分类转化为标签排序问题	分类器的数量与标签的数量成二次方
	RAKEL-SVM[106]	结合 LP 策略和集成学习，既考虑标签间的相关性同时降低了 LP 算法的复杂度	随机性选择子集大小可能会破坏算法的稳定性

（续 表）

划分类型	算法名称	特 点	不 足
算法自适应	RankSVM[107]	考虑了两两样本间的标签相关性，利用最大边际策略处理多标记样本	计算复杂度极高，对噪声敏感
	RankSVMz[108]	定义新的排序损失，简化了 RankSVM 形式，增加了零标签，利用 Frank-Wolfe 迭代求解降低了计算量	算法依赖 Frank-Wolfe 迭代法，计算负担仍然很大
	RankSVMzR[109]	提出 RBCDM 训练算法求解问题，产生的支持向量更少，计算复杂度降低	如何选择最优参数
	Pin-ENR[111]	将弹性网正则化代替 L2 范数正则化为算法提供良好的解释性，采用 FLER 加速策略和 Pinball 损失函数，不仅减少了问题的规模，还降低对噪声的敏感性	FLER 加速策略仅依靠于一个权衡参数
	RBRL[113]	结合 BR 和 RankSVM 实现了一步训练，避免了误差累计，利用低秩约束进一步挖掘标签相关性，采用两种不同的加速近端梯度方法来有效地求解算法	算法参数多，计算复杂度高，难以处理大规模数据集
	SSR-RBRL[114]	通过矩阵分解和最优性条件去除具有零系数的参数矩阵子空间，缩小问题的规模，其次引入近似奇异值分解和加速近端梯度来进一步加快迭代过程	算法所依赖的参数较多
	RankCVM[115]	结合 CVM 和 RankSVM，具有更少的支持向量，计算复杂度更低	算法依赖 Frank-Wolfe 迭代法，计算负担仍然很大
	RankLSVM[116]	推广 LSVM 构造其多标签形式，利用基于启发式收缩策略的随机块坐标下降方法来求解算法，具有线性收敛速度	如何选择最优参数
	MLTSVM[117]	推广 TWSVM 构造其多标签形式，通过确定多个非平行超平面来捕获嵌入在样本的多标签信息，提出一种有效的逐步松弛算法加快训练过程	算法所依赖的参数较多
	ML-SLSTSVM[118]	考虑了训练样本的结构信息，采用最小二乘思想来提高算法泛化能力运行速度	算法所依赖的参数较多

4 算法分析与比较

为了分析弱监督场景下支持向量机算法的分类性能，本节对第 3 节综述的一些代表性算法进行评估和比较，实验的硬件测试环境为 Intel(R) Core(TM) i5-8300H CPU@2.30GHz，4GB RAM.

4.1 半监督学习

针对在半监督学习下支持向量机算法，实验选取了部分具有代表性的方法在 5 个 UCI 数据集进行测试，数据集的详细信息见表 4.

表 4 UCI 数据集

数据集	样本数	特征数	已标记数目
Australian	690	14	69
CMC	1473	9	147
German	1000	24	100
Ionosphere	351	34	35
WDBC	569	30	57

为保证实验对比的可靠性，每个 UCI 数据集以 7∶3 的比例划分为训练集和测试集，并在训练集上采用了十倍交叉验证的网格搜索法对参数进行选择，所有算法均采用 RBF 核函数，且其核参数的选取范围为 $\{2^{-6},2^{-5},\cdots,2^{5},2^{6}\}$. MeanS3VM-iter，MeanS3VM-mkl，CS4VM 和 WELLSVM 都属于基于聚类假设的半监督 SVM 方法，该类算法的主要参数是用于平衡模型复杂度、有标记样本与未标记样本重要程度的折中参数，分别固定为 1 和 0.1，CS4VM 中关于正类样本的错分代价的参数固定为 1. LapSVM、LapTSVM、LapTSVM、LapPPSVM、HGSVM 和 fast-HGSVM 都属于基于流形假设的半监督 SVM 方法，该类算法的主要参数是用于控制分类器在再生核希尔伯特空间中和数据流形结构上复杂度的两个正则化参数，其参数的选取范围均为 $\{2^{-6},2^{-5},\cdots,2^{5},2^{6}\}$，fast-HGSVM 中的边界参数固定为 0.8. 本节采用半监督学习中常用的评价指标准确率 acc 来评估不同算法的性能.

$$acc=\frac{TP+TN}{TP+TN+FP+FN} \tag{21}$$

其中，TP 表示正样本被预测为正类的数目，TN 表示负样本被预测为负类的数目，FP 表示负样本被预测为正类的数目，FN 表示正样本被预测为负类的数目.

从表 5 的实验结果可以明显看出，不同的半监督方法在 5 个 UCI 数据集上的分类性能均有所差异，其中 fast-HGSVM 算法性能表现最佳，其次是 HGSVM 算法. 这两种基于流形假设的方法通过超图来表征样本的流形结构，充分挖掘了样本间的高阶复杂关系，另外 fast-HGSVM 还利用边界样本的方法来预选最有可能成为支持向量的未标记样本更进一步提高了分类的准确率. 在基于聚类假设的方

法中，CS4VM 算法相较于其它三种算法表现更优，该算法利用代价敏感学习来优化有标记和未标记样本的总体代价，在一定程度上提高了分类准确率. 图 8 展示了部分半监督 SVM 方法在 UCI 数据集上的运行时间，基于聚类假设的四个算法表现了差不多的运行时间，而在基于流形假设的算法中 LapSVM 消耗了最少的运行时间，反之 HGSVM 却消耗了最多的运行时间. HGSVM 考虑了多个样本之间的多元关系，而 LapSVM 仅考虑了两两样本之间的序对关系，其计算成本必然增加. fast-HGSVM 利用边界样本筛选未标记，大大缩减了样本空间，显著地降低了计算成本，虽然没有取得最少的算法运行时间，但取得了最优的模型性能，更适合一些要求实时响应且精度高的学习任务.

表 5　半监督方法在 UCI 数据集上的实验结果

数据集	MeanS3VM -iter	MeanS3VM -mkl	CS4VM	WELLSVM	Lap SVM	Lap TSVM	Lap PPSVM	HG SVM	fast-HGSVM
Australian	0.837	0.802	0.838	0.852	0.652	0.666	0.829	0.868	0.876
CMC	0.554	0.574	0.917	0.913	0.587	0.614	0.776	0.931	0.933
German	0.652	0.616	0.661	0.650	0.598	0.606	0.707	0.713	0.716
Ionosphere	0.788	0.883	0.805	0.760	0.695	0.730	0.789	0.876	0.884
WDBC	0.914	0.864	0.941	0.925	0.841	0.864	0.915	0.943	0.950

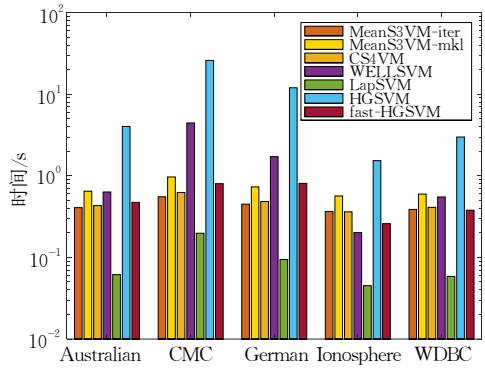

图 8　半监督方法在 UCI 数据集上的运行时间

4.2　多示例学习

为了分析在多示例学习下支持向量机算法的分类性能，本节选取了多示例学习基准数据集 Musk 对部分具有代表性的方法进行评价. Musk 是 Dietterich 等人提出的一种药物活性预测数据集，其详细信息见表 6，该数据集分为两个部分：Musk1 和 Musk2，其中 Musk2 的包总数比 Musk1 多 10 个，但示例数是后者的 13.8 倍多，因此 Musk2 的学习难度更大. 实验的设置参照文献[86]，本节采用准确率 *acc* 来评估不同算法的性能.

表 7 展示了 9 个代表性方法的分类准确率，9 种多示例方法在 Musk1 和 Musk2 数据集上均能取得较高的分类准确率，其中 SMILE 算法获得了最佳的分类准确率：0.913 和 0.916. 该算法不同于其他通过训练分类器来消除正包中的大量模糊示例的方法，而是通过考虑模糊示例与正、负类示例的相似度来显示地处理这些示例，从而提高了分类的准确率. 对于 KI-SVM 算法，在 Musk1 数据集上 B-KI-SVM 算法的分类准确率比 I-KI-SVM 高 0.040，而在 Musk2 数据集上，B-KI-SVM 算法的分类准确率比 I-KI-SVM 低 0.024. 对于 MILD 算法，在 Musk1 数据集上 MILD_I 算法的分类准确率比 MILD_B 高 0.009，而在 Musk2 数据集上 MILD_I 算法的分类准确率比 MILD_B 高 0.031. 图 9 展示了多示例 SVM 方法在 Musk 数据集上的运行时间，无论是 Musk1 还是 Musk2，MILIS 所占用运行时间最少，对于 Musk2 数据集，MILIS 相对于 MILES 的运行时间加速因示例数量较多而更加明显，影响 MILES 运行效率的主要障碍是训练 1 范数 SVM，其在处理大量示例时会非常耗时，因此导致这两种方法在训练速度上存在很大差异. 另外 MILIS(L1)的训练速度也证实了这一点，其中 1 范数 SVM 分类器在训练中需要更多的时间，比标准的 MILIS 要多. mi-SVM 和 MI-SVM 虽然比 MILES 快，但仍比 MILIS 慢. 因为 mi-SVM 和 MI-SVM 都是在示例级别处理 SVM 分类问题，在每次迭代中，mi SVM 使用所有输入示例作为训练集在示例特征空间中学习 SVM. MI-SVM 通过使用所有负示例和每个正袋中的单个正示例来训练类似的 SVM，当数据集每个包包含大量示例时，它们的计算成本大大增加.

表 6　Musk 数据集

数据集	包　数			特征数	示例数	包中示例数		
	正包数	负包数	总数			最少	最多	平均
Musk1	47	45	92	166	476	2	40	5.17
Musk2	39	63	102	166	6598	1	1044	64.69

表 7 多示例方法在 Musk 数据集上的实验结果

数据集	mi-SVM	I-KI-SVM	SMILE	MI-SVM	B-KI-SVM	DD-SVM	MILD_I	MILD_B	MILES
Musk1	0.874	0.840	0.913	0.779	0.880	0.858	0.911	0.902	0.863
Musk2	0.836	0.844	0.916	0.843	0.820	0.913	0.896	0.865	0.876

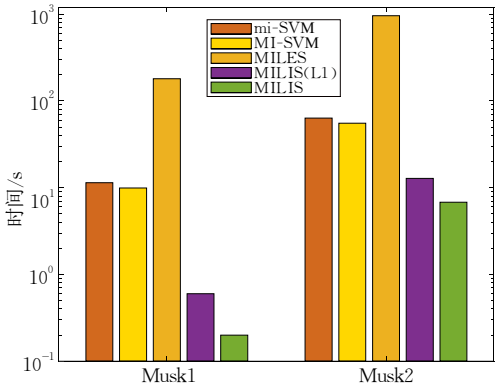

图 9 多示例方法在 *Musk* 数据集上的运行时间

4.3 多标记学习

由于样本可能同时具有多个标记的特性，传统单标记学习中的评价指标并不适用于多标记学习算法. 为了分析在多标记学习下支持向量机算法的性能，本节通过多标记领域中 5 种常用的评价指标来综合评价第 3.3 节中的部分代表性算法，评价指标包括 *Hamming loss*、*Ranking loss*、*Coverage*、*Subset accuracy*、*Average precision*，分别对应于 *hloss*、*rloss*、*cov*、*subsetacc*、*avgprec*；其中前 3 种评价指标取值越小，则算法的性能越好，而后 2 种评价指标取值越大，则算法的性能越好. 假设多标记测试样本集 $T=\{(\boldsymbol{x}_i,Y_i)\mid i=1,2,\cdots,m\}$，其中 $Y_i\in\{-1,1\}^l$，$h(\cdot)$ 为多标记分类器，$f(\cdot,\cdot)$ 为预测函数，$rank_f(\cdot,\cdot)$ 为排序函数.

(1) *Hamming loss*：用于评估样本中单个标记误匹配的情况.

$$hloss(h)=\frac{1}{m}\sum_{i=1}^{m}\frac{1}{|Y|}\left\|h(\boldsymbol{x}_i)\Delta Y_i\right\| \tag{21}$$

其中 Δ 表示两个集合的对称差.

(2) *Ranking loss*：用于评估不相关标记的排序高于相关标记的排序的情况.

$$\mathrm{rloss}(f)=\frac{1}{n}\sum_{i=1}^{n}\frac{1}{|Y_i|\,|\overline{Y}_i|}\left|\{(y_1,y_2)\mid f(\boldsymbol{x}_i,y_1)\le f(\boldsymbol{x}_i,y_2),(y_1,y_2)\in Y_i\times\overline{Y}_i\}\right| \tag{22}$$

(3) *Coverage*：用于评估覆盖样本所有相关标记所需的搜索步数.

$$cov(f)=\frac{1}{n}\sum_{i=1}^{n}\max_{y\in Y_i}rank_f(\boldsymbol{x}_i,y)-1 \tag{23}$$

(4) *Subset accuracy*：用于评估预测的标记集合与真实的标记集合一致的情况.

$$subsetacc(h)=\frac{1}{n}\sum_{i=1}^{n}\left\|h(\boldsymbol{x}_i=Y_i)\right\| \tag{24}$$

(5) *Average precision*：

$$avgprec(h)=\frac{1}{n}\sum_{i=1}^{n}\frac{1}{|Y_i|}\sum_{y\in Y_i}\frac{|y'|rank_f(\boldsymbol{x}_i,y')\le rank_f(\boldsymbol{x}_i,y),y'\in Y_i|}{rank_f(\boldsymbol{x}_i,y)} \tag{25}$$

实验数据集选取了来自音乐、图像、生物以及文本领域的六个多标记数据集：Emotions①、Scene①、Yeast①、Enron①、Arts①、Education①，其具体信息如表 8 所示. 为保证实验对比的可靠性，每个数据集以 3∶2 的比例划分为训练集和测试集，所有算法均采用 RBF 核函数，核参数为 $1/m$（其中 m 为数据集的特征数），另外采用了五倍交叉验证的网格搜索法对参数进行选择. BR-SVM、CLR-SVM、RAKEL-SVM 都属于基于问题转换的多标记 SVM 方法，该类算法的主要参数是用于折中错分样本与模型复杂度的惩罚参数，其选取范围为 $\{2^{-4},2^{-3},\cdots,2^3,2^4\}$，RAKEL-SVM 中标签子集数目 3，分类器数目为 $2l$，其中 l 为标签总数. RBRL 中三个折中参数的选取范围均为 $\{2^{-4},2^{-3},\cdots,2^3,2^4\}$.

表 8 多标记数据集

数据集	示例数	特征数	标记类别	标记基数	标记密度	领域
Emotions	593	72	6	1.869	0.311	音乐
Scene	2407	294	6	1.074	0.179	图像
Yeast	2417	103	14	4.237	0.303	生物
Enron	1702	1001	53	3.378	0.064	文本
Arts	5000	462	26	1.636	0.063	文本
Education	5000	550	33	1.461	0.044	文本

表 9 展示了 6 个代表性方法的分类效果. 对于评价指标 *Subset accuracy*，在 Emotions、Yeast 和 Enron 数据集上，基于问题转换的方法 RAKEL 更优于其他算法，而在 Scene、Arts 和 Education 数据集上，基于算法适应的方法 RBRL 取得最优的效果. 对于评价指标 *Hamming loss*，RBRL 在 Scene、Yeast 和 Enron 数据集上取得了最优性能，在 Emotions 数据集上仅次于 RAKEL. 对于评价指标 *Ranking loss*、*Coverage* 和 *Average precision*，RBRL 可在这六个多标记数据集上表现更优. 由此可见，RBRL 将

① http://mulan.sourceforge.net/datasets-mlc.html

表 9　多标记方法在多标记数据集上的实验结果

数据集	评价指标	BR-SVM	CLR-SVM	RAKEL	RankSVM	RankSVMz	RBRL
Emotions	*hloss* ↓	0.183	0.182	0.177	0.189	0.201	0.181
	rloss ↓	0.246	0.149	0.192	0.155	0.149	0.138
	coverage ↓	0.386	0.283	0.338	0.294	0.291	0.277
	subsetacc ↑	0.313	0.318	0.356	0.291	0.292	0.334
	avgprec ↑	0.760	0.813	0.801	0.808	0.819	0.828
Scene	*hloss* ↓	0.077	0.078	0.075	0.092	0.113	0.073
	rloss ↓	0.128	0.061	0.087	0.065	0.072	0.058
	coverage ↓	0.119	0.064	0.089	0.068	0.075	0.062
	subsetacc ↑	0.655	0.650	0.696	0.563	0.500	0.735
	avgprec ↑	0.834	0.887	0.875	0.882	0.874	0.895
Yeast	*hloss* ↓	0.188	0.188	0.195	0.203	0.207	0.187
	rloss ↓	0.308	0.158	0.244	0.170	0.172	0.157
	coverage ↓	0.627	0.436	0.543	0.446	0.458	0.436
	subsetacc ↑	0.190	0.194	0.248	0.156	0.179	0.192
	avgprec ↑	0.680	0.773	0.727	0.755	0.765	0.777
Enron	*hloss* ↓	0.052	0.050	0.052	0.051	0.061	0.046
	rloss ↓	0.298	0.071	0.208	0.081	0.098	0.072
	coverage ↓	0.580	0.210	0.472	0.235	0.267	0.214
	subsetacc ↑	0.128	0.131	0.154	0.119	0.060	0.137
	avgprec ↑	0.482	0.705	0.592	0.672	0.630	0.709
Arts	*hloss* ↓	0.054	0.055	0.057	0.061	0.106	0.059
	rloss ↓	0.344	0.112	0.261	0.109	0.116	0.106
	coverage ↓	0.432	0.170	0.352	0.166	0.181	0.164
	subsetacc ↑	0.241	0.237	0.319	0.274	0.098	0.348
	avgprec ↑	0.445	0.618	0.557	0.617	0.616	0.635
Education	*hloss* ↓	0.038	0.039	0.040	0.046	0.079	0.042
	rloss ↓	0.458	0.081	0.358	0.072	0.081	0.070
	coverage ↓	0.518	0.110	0.424	0.100	0.148	0.099
	subsetacc ↑	0.250	0.227	0.315	0.249	0.096	0.349
	avgprec ↑	0.376	0.605	0.503	0.613	0.578	0.643

RankSVM 和 BR 与鲁棒低秩学习相结合，不仅解决了 RankSVM 易造成错误累积的缺点，还充分挖掘低维标签空间假设下的高阶标签相关性，从而提高了分类性能.

图 10 展示了多标记 SVM 方法在多标记数据集上的运行时间，基于问题转换的方法要比基于算法适应的方法消耗更少的运行时间，因为基于算法适应的方法是通过改进传统的单标记算法来适应多标记样本的学习，面临的标记空间更复杂、更高维，

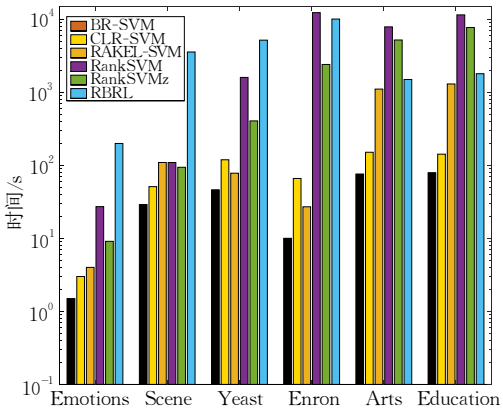

图 10　多标记方法在多标记数据集上的运行时间

计算复杂度会随之大幅度增加. 在基于问题转换的方法中，BR-SVM 消耗了最少的运行时间，相应的算法性能也最差. 相比之下，CLR-SVM 和 RAKEL-SVM 占用了更多的运行时间，但多标记学习任务上表现好，这意味带来额外性能同时消耗了更多的计算资源. 在基于算法适应的方法中，对比 RankSVM，RankSVMz 需要更少的运行时间，这得益于算法通过添加零标签简化了原始 RankSVM 的原始形式，并且在每个迭代步骤中采用 OVR 策略将整个线性规划问题划分一系列子问题，减少了计算成本；在前三个数据集中，RBRL 的运行时间比 RankSVM 慢，而在后三个文本数据集中，RBRL 的运行时间比 RankSVM 快，文本数据集具有更高维的、数量更多的样本，RBRL 可通过加速近端梯度方法使其快速收敛，更适合于大规模数据问题.

5　未来的研究方向

5.1　数据不平衡问题

数据不平衡在监督场景下的分类领域已是一个

经典的问题,但在弱监督场景中,数据不平衡现象更为严重.在半监督学习中,数据不平衡的现象不仅会造成分类边界偏向于多数类的一方,在极端情况下可能会造成某一个类别的样本不存在以致使学习器无法学习到该类样本的信息,从而降低半监督分类的准确性.在多示例学习中,数据不平衡现象不仅可以发生在示例空间层面,还可以发生包空间层面,这使得学习到的决策边界会更偏向于多数类.在多标记学习中,一个样本隶属于多个标签,不同标签的样本数量往往相差很多,这导致数据不平衡问题在多标记学习中更为普遍.目前,面向弱监督场景下的不平衡数据分类方法研究比较少,而面向监督场景下的不平衡数据分类方法更为常见,主要集中在数据层面和算法层面,其中数据层面是通过重采样来平衡数据集,使得各类别样本的数量相等,例如欠采样、过采样以及混合采样等,而算法层面则是通过调整算法本身,来提高起学习少数类的能力,例如代价敏感、集成学习等.因此,可借鉴监督场景下不平衡数据的处理方法,利用代价敏感、样本采样等技术中来缓解数据不平衡问题所带来的影响,使得弱监督下的支持向量机算法具有更好的泛化性能.

5.2 弱监督回归问题

尽管针对弱监督场景已经提出了各种类型的改进支持向量机算法,并已在文本分类、人脸识别等领域得到了成功应用,但是现有研究主要集中于解决支持向量机的分类问题,对回归问题的研究和利用还非常缺乏.回归问题比分类问题复杂,其样本标记是数值型输出,而半监督学习中聚类假设强调的是同一个类簇中的样本可能属于相同的类别标记,因此大多数基于聚类假设的半监督算法在支持向量机回归问题中不成立,与聚类假设不同,流形假设更强调相同局部邻域内的样本具有相似的输出,且回归问题的输出通常具有平滑性,因此流形假设更适用于回归问题,可尝试采用一些基于流形假设的半监督学习方法,例如图正则化算法,将其推广到半监督支持向量机回归中.在多示例学习中,Ray 等人[119]证明多示例回归问题的求解过程是个 NP 问题,并指出多示例回归方法要比分类方法更为有效,例如在研究药物活性预测问题上,数值型输出更能表现药物分子绑定的强弱,可为药物设计提供更大的帮助.因此,关于弱监督场景下的支持向量机回归问题的研究也是今后一个重要的研究方向.

5.3 混合弱监督学习问题

目前针对弱监督场景下的问题研究主要集中于一些单一的学习框架中.对于其他一些更为复杂的问题研究则相对比较少,这些问题大多数是以混合形式出现,例如半监督多标记、半监督多示例、多示例多标记等.然而这些混合弱监督学习问题都更具有广泛而实际的应用背景,例如在图像分类中,一幅图像不仅可以拥有多个标记,它还可以通过某种技术将图像划分成多个区域,从而被表示成由多个区域组成的一个集合;因此,如何设计面向混合弱监督场景下的支持向量机算法,具有重要的现实意义.

5.4 大规模深层次问题

目前弱监督场景下的支持向量机算法主要集中在一些小型的分类任务中,对于一些更高更深层次的任务如语义分割、自动控制等却极少涉及.虽然弱监督学习问题普遍存在于语义分割、图像识别与理解、目标检测等各个领域,但关于大规模弱监督场景下的学习问题研究尚处于起步阶段,大多数弱监督学习方法如半监督学习、多示例学习等扩展性并不好,随着数据集的增大,模型性能的衰退性十分明显.因此面对弱监督场景下的大规模深层次问题,如何合理有效地融合相关任务场景和支持向量机算法是一个具有挑战性的问题.

5.5 开放环境下的学习问题

目前大多数机器学习算法都是针对某一静态数据分布的样本集进行建模,而实际情况收集的样本集常常是动态变化的,可能会不断地发现新类别样本,这导致样本的数目、维度、标签的数目、标记缺失的概率分布等都会发生变化,因此真实应用场景往往具有“开放”属性.在线学习是一种模型训练方法,其特点是能够连续接收新样本并实时更新模型.这种方法适用于面对动态变化的样本集,使模型适应新的类别样本,并不断改进性能.因而,在面对开放环境下,如何结合弱监督场景和在线学习,使得 SVM 在学习和适应新的类别时,仍然能保持原有的性能,并提升算法对未知类别样本的泛化能力.

6 总 结

本文从弱监督场景入手,主要总结了基于半监督学习,多示例学习和多标记学习的支持向量机算法,并对近年的相关方法进行了分类和描述.针对基于半监督学习的支持向量机算法,根据其数据假设可划分为 2 类,分别为聚类假设和流形假设;针对基于多示例学习的支持向量机算法,根据其解决方案可划分为 3 类,分别为示例水平空间、包水平空间以

及嵌入空间的方法；针对基于多标记学习的支持向量机算法，根据其处理思路可划分为 2 类，分别为问题转换和算法自适应；通过在公共数据集进行实验，对比分析了部分代表性算法的分类性能；最后针对数据不平衡问题、弱监督回归问题、混合弱监督学习问题、大规模深层次问题以及开放环境下的学习问题，分析了弱监督场景下的支持向量机算法的未来研究方向.

参 考 文 献

[1] Zhou Zhi-Hua. Machine Learning. Beijing: Tsinghua University Press, 2016(in Chinese)
(周志华. 机器学习. 北京：清华大学出版社，2016)

[2] Wang Kuai-Ni. Research on Robust Learning Models and Algorithms of Support Vector Machine[Ph. D. dissertation]. China Agricultural University, Beijing, 2015(in Chinese)
(王快妮. 支持向量机鲁棒性模型与算法研究[博士学位论文]. 中国农业大学，北京，2015)

[3] Zhou Z. A brief introduction to weakly supervised learning. National Science Review, 2018, 5(1): 44-53

[4] Vapnik V. Statistical Learning Theory. New York, USA: Wiley, 1998

[5] Vapnik V. An overview of statistical learning theory. IEEE Transactions on Neural Network, 1999, 10(5): 988-999

[6] Yan X, Zhu H. A novel robust support vector machine classifier with feature mapping. Knowledge-Based Systems, 2022, 257: 109928

[7] Vadhnani S, Singh N. Brain tumor segmentation and classification in MRI using SVM and its variants: A survey. Multimedia Tools and Applications, 2022, 81(22): 31631-31656

[8] Sun G, Rong X, Zhang A, et al. Multi-scale Mahalanobis kernel-based support vector machine for classification of high-resolution remote sensing images. Cognitive Computation, 2021, 3: 787-794

[9] Bamgboye P, Adebiyi M, Adebiyi A, et al. Text classification on customer review dataset using support vector machine// Proceedings of the 6th World Conference on Smart Trends in Systems Security and Sustainability. London, UK, 2023, 579: 407-415

[10] Jeong K, Choi S. Takagi-Sugeno fuzzy observer-based magnetorheological damper fault diagnosis using a support vector machine. IEEE Transactions on Control Systems Technology, 2022,30(4): 1723-1735

[11] Zhang J, Zhang Q, Qin X, et al. A two-stage fault diagnosis methodology for rotating machinery combining optimized support vector data description and optimized support vector machine. Measurement, 2022, 200: 111651

[12] Chen Xia. Research on Weakly-Supervised Classification Methods Based on Samples and Labels Modeling[M. S. dissertation]. Southwest University, Chongqing, 2019(in Chinese)
(陈霞. 基于样本和标记建模的弱监督分类方法研究[硕士学位论文]. 西南大学，重庆，2019)

[13] Guo Y, Liang X, Wu B, et al. Dual-aware domain mining and cross-aware supervision for weakly-supervised semantic segmentation. ACM Transactions on Knowledge Discovery from Data, 2023, 17(7): 101

[14] Zhang B, Xiao J, Wei Y, et al. End-to-end weakly supervised semantic segmentation with reliable region mining. Pattern Recognition, 2022, 128: 108663

[15] Yi S, Ma H, Wang X, et al. Weakly-supervised semantic segmentation with superpixel guided local and global consistency. Pattern Recognition, 2022, 124: 108504

[16] Wang X, Liu J, Wang W, et al. Weakly supervised hyperspectral image classification with few samples based on intradomain sample expansion. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2023, 16: 5769-5781

[17] Yan T, Shi J, Li H, et al. Discriminative information restoration and extraction for weakly supervised low-resolution fine-grained image recognition. Pattern Recognition, 2022, 127: 108629

[18] Sui L, Zhang C, Wu J, et al. Salvage of supervision in weakly supervised object detection and segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45(8): 10394-10408

[19] Xia R, Li G, Huang Z, et al. CBASH: Combined backbone and advanced selection heads with object semantic proposals for weakly supervised object detection. IEEE Transactions on Circuits and Systems for Video Technology, 2022, 32(10): 6502-6514

[20] Zhang C, Li Y, Wu J. Weakly supervised foreground learning for weakly supervised localization and detection. Pattern Recognition, 2023, 137: 109279

[21] Ding Shi-Fei, Qi Bing-Juan, Tan Hong-Yan. An overview on theory and algorithm of support vector machines. Journal of University of Electronic Science and Technology of China, 2011, 40(1): 2-10(in Chinese)
(丁世飞，齐丙娟，谭红艳. 支持向量机理论与算法研究综述. 电子科技大学学报，2011，40(1)：2-10)

[22] Roy A, Chakraborty S. Support vector machine in structural reliability analysis: A review. Reliability Engineering and System Safety, 2023, 233: 109126

[23] Zhang Min-Ling, Wu Xuan. Disambiguation-free partial label learning. SCIENTIA SINICA Informationis, 2019, 49(9): 1083-1096(in Chinese)
(张敏灵，吴璇. 非消歧偏标记学习. 中国科学：信息科学，2019，49(9)：1083-1096)

[24] Liu Jian-Wei, Liu Yuan, Luo Xiong-Lin. Semi-supervised learning methods. Chinese Journal of Computers, 2015, 38(8): 1592-1617(in Chinese)
(刘建伟，刘媛，罗雄麟. 半监督学习方法. 计算机学报，2015，38(8)：1592-1617)

[25] Chapelle O, Scholkopf B, Zien A. Semi-Supervised Learning. Cambridge, USA: MIT Press, 2006

[26] Mey A, Loog M. Improved generalization in semi-supervised learning: A survey of theoretical results. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45(4): 4747-4767

[27] Dietterich T, Lathrop R, Lozano-Pérez T. Solving the multiple instance problem with axis-parallel rectangles. Artificial Intelligence, 1997, 89(1): 31-71

[28] Carbonneau M, Cheplygina V, Granger E, et al. Multiple instance learning: A survey of problem characteristics and applications. Pattern Recognition, 2018, 77: 329-353

[29] Zhang M, Zhou Z. A review on multi-label learning algorithms. IEEE Transactions on Knowledge and Data Engineering, 2014, 26(8): 1819-1837

[30] Liu W, Wang H, Shen X, et al. The emerging trends of multi-label learning. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 44(11): 7955-7974

[31] Chapelle O, Weston J, Schölkopf B. Cluster kernels for semi-supervised learning//Proceedings of the Advances in Neural Information Processing Systems. Cambridge, USA, 2002: 601-608

[32] Belkin M, Niyogi P, Sindhwani V. Manifold regularization: A geometric framework for learning from labeled and unlabeled examples. Journal of Machine Learning Research, 2006, 7: 2399-2434

[33] Dornaika F, Bi A, Zhang C. A unified deep semi-supervised graph learning scheme based on nodes re-weighting and manifold regularization. Neural Networks, 2023, 158: 188-196

[34] Bennett K, Demiriz A. Semi-supervised support vector machines //Proceedings of the Advances in Neural Information Processing Systems. Cambridge, USA, 1999, 11: 368-374

[35] Joachims T. Transductive inference for text classification using support vector machines//Proceedings of the 16th International Conference on Machine Learning. San Francisco, USA, 1999: 200-209

[36] Zhao B, Wang F, Zhang C. CutS3VM: A fast semi-supervised SVM algorithm//Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York, USA, 2008: 830-838

[37] Li Y, Kwok J, Zhou Z. Semi-supervised learning using label mean//Proceedings of the 26th International Conference on Machine Learning. New York, USA, 2009: 633-640

[38] Tian Xun, Wang Xi-Li. Semi-supervised support vector machine based on clustering label mean. Computer Engineering and Science, 2018, 40(12): 2265-2272(in Chinese)
(田勋, 汪西莉. 基于聚类标签均值的半监督支持向量机. 计算机工程与科学, 2018, 40(12): 2265-2272)

[39] Li Y, Kwok J, Zhou Z. Cost-sensitive semi-supervised support vector machine//Proceedings of the 24th AAAI Conference on Artificial Intelligences. Atlanta, USA, 2010: 500-505

[40] Li Y, Tsang I, Kwok J, et al. Scalable and convex weakly labeled SVMs. Journal of Machine Learning Research, 2013, 14: 2151-2188

[41] Geng X, Gu B, Li X, et al. Scalable semi-supervised SVM via triply stochastic gradients//Proceedings of the 28th International Joint Conference on Artificial Intelligence. Macao, China, 2019: 2364-2370

[42] Wang E, Wang Z, Wu Q. One novel class of Bézier smooth semi-supervised support vector machines for classification. Neural Computing and Applications, 2021, 33(16): 9975-9991

[43] Li Y, Zhou Z. Towards making unlabeled data never hurt. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(1): 175-188

[44] Wang Y, Meng Y, Fu Z, et al. Towards safe semi-supervised classification: Adjusted cluster assumption via clustering. Neural Processing Letters, 2017, 46(3): 1031-1042

[45] Huang Hua, Zheng Jia-Min, Qian Peng-Jiang. Adjusted cluster assumption and pairwise constraints jointly based semi-supervised classification method. Journal of Computer Applications, 2018, 38(11): 3119-3126(in Chinese)
(黄华, 郑佳敏, 钱鹏江. 调整聚类假设联合成对约束半监督分类方法. 计算机应用, 2018, 38(11): 3119-3126)

[46] Li Tao, Wang Xi-Li. Semi-supervised SVM classification method based on cluster kernel. Application Research of Computers, 2013, 30(1): 42-45, 48(in Chinese)
(李涛, 汪西莉. 一种基于聚类核的半监督支持向量机分类方法. 计算机应用研究, 2013, 30(1): 42-45, 48)

[47] Tao Xin-Min, Cao Pan-Dong, Song Shao-Yu, et al. The SVM classification algorithm based on semi-supervised gauss mixture model kernel. Information and Control, 2013, 42(1): 18-26(in Chinese)
(陶新民, 曹盼东, 宋少宇等. 基于半监督高斯混合模型核的支持向量机分类算法. 信息与控制, 2013, 42(1): 18-26)

[48] Melacci S, Belkin M. Laplacian support vector machines trained in the primal. Journal of Machine Learning Research, 2011, 12: 1149-1184

[49] Qi Z, Tian Y, Shi Y. Successive overrelaxation for Laplacian support vector machine. IEEE Transactions on Neural Networks and Learning Systems, 2015, 26(4): 674-683

[50] Yang Z, Xu Y. A safe screening rule for Laplacian support vector machine. Engineering Applications of Artificial Intelligence, 2018, 67: 309-316

[51] Khan I, Roth P, Bais A, Bischof H. Semi-supervised image classification with huberized Laplacian support vector machines// Proceedings of the 2013 IEEE 9th International Conference on Emerging Technologies. Islamabad, Pakistan, 2013: 6743545

[52] Dong H, Yang L, Wang X. Robust semi-supervised support vector machines with Laplace kernel-induced correntropy loss functions. Applied Intelligence, 2021, 51(2): 819-833

[53] Pei H, Lin Q, Yang L, et al. A novel semi-supervised support vector machine with asymmetric squared loss. Advances in Data Analysis and Classification, 2021, 15(1): 159-191

[54] Liu Zhen, Yang Jun-An, Liu Hui, et al. Laplacian support vector machine by local behavioral similarity. Journal of Chinese Computer Systems, 2016, 37(12): 2749-2754(in Chinese)

(刘振，杨俊安，刘辉等. 基于局部行为相似性的拉普拉斯SVM半监督学习算法. 小型微型计算机系统，2016，37(12)：2749-2754)

[55] Sun Y，Ding S，Guo L，et al. Hypergraph regularized semi-supervised support vector machine. Information Sciences，2022，591：400-421

[56] Dufrenois F，Noyer J. One class proximal support vector machines. Pattern Recognition，2016，52：96-112

[57] Li G，Yang L，Wu Z，et al. D. C. programming for sparse proximal support vector machines. Information Sciences，2021，547：187-201

[58] An Yue-Xuan，Ding Shi-Fei，Hu Ji-Pu. Twin support vector machine：A review. Computer Science，2018，45(11)：29-36 (in Chinese)
(安悦瑄，丁世飞，胡继普. 孪生支持向量机综述. 计算机科学，2018，45(11)：29-36)

[59] Che Z，Liu B，Xiao Y，et al. Twin support vector machines with privileged information. Information Sciences，2021，573：141-153

[60] Liu X，Wang N，Molina D，et al. A least square support vector machine approach based on bvRNA-GA for modeling photovoltaic systems. Applied Soft Computing Journal，2022，117：108357

[61] Wang G，Zhang G，Choi K，et al. Output based transfer learning with least squares support vector machine and its application in bladder cancer prognosis. Neurocomputing，2020，387：279-292

[62] Qi Z，Tian Y，Shi Y. Laplacian twin support vector machine for semi-supervised classification. Neural Networks，2012，35：46-53

[63] Zhang Z，Zhen L，Deng N，et al. Manifold proximal support vector machine with mixed-norm for semi-supervised classification. Neural Computing and Applications，2015，26(2)：399-407

[64] Yang Z，Xu Y. A safe sample screening rule for Laplacian twin parametric-margin support vector machine. Pattern Recognition，2017，84：1-12

[65] Rastogi R，Sharma S. Fast Laplacian twin support vector machine with active learning for pattern classification. Applied Soft Computing Journal，2019，74：424-439

[66] Yu G，Ma J，Xie C. Hessian scatter regularized twin support vector machine for semi-supervised classification. Engineering Applications of Artificial Intelligence，2023，119：105751

[67] Xie X，Sun F，Qian J，et al. Laplacian Lp norm least squares twin support vector machine. Pattern Recognition，2023，136：109192

[68] He K，Zhao W，Xie X，et al. Synergistic learning of lung lobe segmentation and hierarchical multi-instance classification for automated severity assessment of COVID-19 in CT images. Pattern Recognition，2021，113：107828

[69] Li J，Zhang L，Shu X，et al. Multi-instance learning based on spatial continuous category representation for case-level meningioma grading in MRI images. Applied Intelligence，2023，53(12)：16015-16028

[70] Gao Z，Mao A，Wu K，et al. Childhood leukemia classification via information bottleneck enhanced hierarchical multi-instance learning. IEEE Transactions on Medical Imaging，2023，42(8)：2348-2359

[71] Liu B，Xiao Y，Hao Z. A selective multiple instance transfer learning method for text categorization problems. Knowledge-Based Systems，2018，141：178-187

[72] Wan F，Ye Q，Yuan T，et al. Multiple instance differentiation learning for active object detection. IEEE Transactions on Pattern Analysis and Machine Intelligence，2023，45(10)：12133-12147

[73] Wang B，Zhao Y，Li X. Multiple instance graph learning for weakly supervised remote sensing object detection. IEEE Transactions on Geoscience and Remote Sensing，2022，60：5613112

[74] Gao W，Wan F，Yue J，et al. Discrepant multiple instance learning for weakly supervised object detection. Pattern Recognition，2022，122：108233

[75] Han Hai-Yun. Research on Multi-Instance Learning Algorithm Based on Instance Selection[M. S. dissertation]. Xidian University，Xi'an，2021(in Chinese)
(韩海韵. 基于示例选择的多示例学习算法研究[硕士学位论文]. 西安电子科技大学，西安，2021)

[76] Andrews S，Tsochantaridis I，Hofmann T. Support vector machines for multiple-instance learning//Proceedings of the Advances in Neural Information Processing Systems. Cambridge，USA，2003，15(2)：561-568

[77] Astorino A，Fuduli A，Gaudioso M. A Lagrangian relaxation approach for binary multiple instance classification. IEEE Transactions on Neural Networks and Learning Systems，2019，3(9)：3566-3577

[78] Avolio M，Fuduli A. A semiproximal support vector machine approach for binary multiple instance learning. IEEE Transactions on Neural Networks and Learning Systems，2020，32(8)：1-12

[79] Yuan M，Xu Y. Instance elimination strategy for non-convex multiple-instance support vector machine. Applied Soft Computing，2022，129：109564

[80] Li Y F，Kwok J，Zhou Z. A convex method for locating regions of interest with multi-instance learning//Proceedings of the 2009th European Conference on Machine Learning and Knowledge Discovery in Databases. Berlin，Germany，2009：15-30

[81] Xiao Y，Liu B，Hao Z，et al. A similarity-based classification framework for multiple-instance learning. IEEE Transactions on Cybernetics，2014，44(4)：500-515

[82] Cheung P，Kwok J. A regularization framework for multiple-instance learning//Proceedings of the 23rd International Conference on Machine Learning. Pennsylvania，USA，2006：193-200

[83] Melki G，Cano A，Ventura S. MIRSVM：Multi-instance support vector machine with bag representatives. Pattern Recognition，2018，79：228-241

[84] Cheplygina V, Tax D M, Loog M. Multiple instance learning with bag dissimilarities. Pattern Recognition, 2015, 48(1): 264-275

[85] Chen Y, Wang J. Image categorization by learning and reasoning with regions. The Journal of Machine Learning Research, 2004, 5: 913-939

[86] Chen Y, Bi J, Wang J. MILES: Multiple-instance learning via embedded instance selection. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2006, 28(12): 1931-1947

[87] Li W, Yeung D. MILD: Multiple-instance learning via disambiguation. IEEE Transactions on Knowledge and Data Engineering, 2010, 22(1): 76-89

[88] Fu Z, Robles-Kelly A, Zhou J. MILIS: Multiple instance learning with instance selection. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2011, 33(5): 958-977

[89] Huang S, Liu Z, Jin W, et al. Bag dissimilarity regularized multi-instance learning. Pattern Recognition, 2022, 126: 108583

[90] Zhu Y, Wu Q. Elementary discourse units with sparse attention for multi-label emotion classification. Knowledge-Based Systems, 2022, 240: 108114

[91] Lin N, Fu S, Lin X, et al. Multi-label emotion classification based on adversarial multi-task learning. Information Processing and Management, 2022, 9(6): 103097

[92] Deng J, Ren F. Multi-label emotion detection via emotion-specified feature extraction and emotion correlation learning. IEEE Transactions on Affective Computing, 2023, 14(1): 475-486

[93] Zhang Y, Wu J, Cai Z, et al. Multi-view multi-label learning with sparse feature selection for image annotation. IEEE Transactions on Multimedia, 2020, 22(11): 2844-2857

[94] Maltoudoglou L, Paisios A, Lenc L, et al. Well-calibrated confidence measures for multi-label text classification with a large number of labels. Pattern Recognition, 2022, 122: 108271

[95] Ma Y, Liu X, Zhao L, et al. Hybrid embedding-based text representation for hierarchical multi-label text classification. Expert Systems with Applications, 2022, 187: 115905

[96] Song R, Liu Z, Chen X, et al. Label prompt for multi-label text classification. Applied Intelligence, 2023, 53(8): 8761-8775

[97] Chauhan V, Tiwari A, Joshi N, et al. Multi-label classifier for protein sequence using heuristic-based deep convolution neural network. Applied Intelligence, 2022, 52(3): 2820-2837

[98] Chen T, Wu T, Pan D, et al. TransRNAm: Identifying twelve types of RNA modifications by an interpretable multi-label deep learning model based on transformer. IEEE/ACM Transactions on Computational Biology and Bioinformatics, 2023, 20(6): 3623-3634

[99] Boutell M, Luo J, Shen X, et al. Learning multi-label scene classification. Pattern Recognition, 2004, 37(9): 1757-1771

[100] Hsieh C, Si S, Dhillon I. A divide-and-conquer solver for kernel support vector machine//Proceedings of the 31st International Conference on Machine Learning. Beijing, China, 2014: 855-870

[101] Liu Jing, Guo Zhong-Wen, Sun Zhong-Wei, et al. Research on fast multi-label SVM classification algorithm using divide-and-conquer strategy. Periodical of Ocean University of China, 2020, 50(12): 160-166(in Chinese)
(刘竞, 郭忠文, 孙中卫等. 采用分而治之策略的快速多标签支持向量机分类算法研究. 中国海洋大学学报(自然科学版), 2020, 50(12): 160-166)

[102] Read J, Pfahringer B, Holmes G, et al. Classifier chains for multi-label classification//Proceedings of the Machine Learning and Knowledge Discovery in Databases. Berlin, Germany, 2009: 254-269

[103] Read J, Pfahringer B, Holmes G, et al. Classifier chains for multi-label classification. Machine Learning, 2011, 85(3): 333-359

[104] Furnkranz J, Hullermeier E, Mencia E, et al. Multilabel classification via calibrated label ranking. Machine Learning, 2008, 73(2): 133-153

[105] Tsoumakas G, Vlahavas I. Random *k*-labelsets for multilabel classification. IEEE Transactions on Knowledge and Data Engineering, 2011, 23(7): 1079-1089

[106] Yapp E, Li X, Lu W, et al. Comparison of base classifiers for multi-label learning. Nerocomputing, 2020, 394(21): 51-60

[107] Elisseeff A, Weston J. A kernel method for multi-labelled classification//Proceedings of the Advances Neural Informational Processing Systems. Cambridge, USA, 2001: 681-687

[108] Xu J. An efficient multi-label support vector machine with a zero label. Expert Systems with Applications, 2012, 39(5): 4796-4804

[109] Xu J. Random block coordinate descent method for multi-label support vector machine with a zero label. Expert Systems with Applications, 2014, 41(7): 3418-3428

[110] Du K, Ruan C, Sun X. On the convergence of a randomized block coordinate descent algorithm for a matrix least squares problem. Applied Mathematics Letters, 2022, 124: 107689

[111] Wang H, Xu Y. Sparse elastic net multi-label rank support vector machine with pinball loss and its applications. Applied Soft Computing Journal, 2021, 104: 107232

[112] Sharma S, Rastogi R, Chandra S. Large-scale twin parametric support vector machine using pinball loss function. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2021, 51(2): 987-1003

[113] Wu G, Zheng R, Tian Y, et al. Joint ranking SVM and binary relevance with robust low-rank learning for multi-label classification. Neural Networks, 2020, 122: 24-39

[114] Zhong P, Xu Y. Subspace screening rule for multi-label estimator with sparsity-inducing regularization. Neurocomputing, 2023, 527: 143-154

[115] Xu J. Fast multi-label core vector machine. Pattern Recog-

nition, 2013, 6(3): 885-898

[116] Xu J. Multi-label Lagrangian support vector machine with random block coordinate descent method. Information Sciences, 2016, 329: 184-205

[117] Chen W, Shao Y, Li C, et al. MLTSVM: A novel twin support vector machine to multi-label learning. Pattern Recognition, 2016, 52: 61-74

[118] Azad-Manjiri M, Amiri A, Sedghpour A. ML-SLSTSVM: A new structural least square twin support vector machine for multi-label learning. Pattern Analysis and Applications, 2020, 23: 295-308

[119] Ray S, Page D. Multiple instance regression//Proceedings of the 18th International Conference on Machine Learning. San Francisco, USA, 2001: 425-432

DING Shi-Fei, Ph. D., professor, Ph. D. supervisor. His research interests include intelligent information processing, pattern recognition, machine learning, data mining, and granular computing.

SUN Yu-Ting, Ph. D. candidate. Her research interests include support vector machine and weakly supervised learning.

LIANG Zhi-Zhen, Ph. D., associate professor. His research interests include image processing, machine learning, and data mining.

GUO Li-Li, Ph. D., lecturer. Her research interests include deep learning and multimodal emotional analysis.

ZHANG Jian, Ph. D., lecturer. His research interests include deep learning and multi-label learning.

XU Xiao, Ph. D., lecturer. Her research interests include data mining and density peaks clustering.

Background

Support vector machine (SVM) is a machine learning method based on VC dimension and structural risk minimization in statistical learning theory. In comparison to artificial neural networks, SVM effectively avoids the issues of local optimal solutions and the curse of dimensionality. With its unique advantages in solving nonlinear, few shot and high dimensional problems, it has gained significant attention and widely applied in various fields. As a classical machine learning method, SVM requires the labels of training samples to be complete, accurate and unique, so as to ensure the generalization ability of the learner. However, meeting these requirements simultaneously often proves challenging in practical application scenarios, and manual labeling is expensive and time-consuming, especially in certain specific fields where professional knowledge is also needed for labeling. Consequently, the labels of a large number of samples in many practical scenarios are lacking, inaccurate, and even ambiguous. Standard SVM in such scenarios which can be uniformly called weakly supervised scenarios struggles to achieve satisfactory performance. To effectively model SVM in weakly supervised scenarios using rough labeled samples with less supervisory information and lower labeling costs, researchers have proposed many methods based on semi-supervised learning, multiple instance learning, and multi-label learning, which have alleviated the shortcomings of traditional SVM that heavily rely on manual labeling and achieved more efficient practical applications.

In this paper, the current research status and developments of SVM algorithms in weakly supervised scenarios are summarized. Firstly, the concept of weakly supervised scenarios and the basic mathematical principle of SVM are briefly introduced. Secondly, the existing SVM algorithms in weakly supervised scenarios are divided into three categories according to different learning paradigms, namely, the semi-supervised learning based methods, the multiple instance learning based methods, and the multi-label learning based methods. The core ideas, characteristics and shortcomings of the representative methods within these categories are systematically introduced, providing a basis for selecting different SVM methods in various task scenarios. Then some representative algorithms are evaluated on some public datasets. Finally, the future research directions are discussed, including data imbalance, weakly supervised regression, mixed weakly supervised learning, large-scale deep-level tasks and learning problems for open.

This work is supported by the National Natural Science Foundation of China (62276265, 61976216, 62206297, 62206296).