**数字经济领域省科学技术奖（自然科学）推荐项目公示表**

**一、项目名称：大规模复杂问题的非平行支持向量机数据挖掘方法研究**

**二、推荐单位及推荐意见**

|  |  |
| --- | --- |
| 推荐单位 | 绍兴市人民政府 |
| 推荐意见（限150字） |
| 该项目属数据挖掘与机器学习领域国际前沿性、基础性研究方向，取得了创新性成果。该项目重点围绕在非平行支持向量机在大规模数据挖掘中存在的结构化学习问题展开深入研究。7篇代表性论文在国际权威学术期刊上发表，其中5篇影响因子大于3，SCI他引213余次。推荐该项目为省自然科学奖 二 等奖。 |

**三、项目简介**

|  |
| --- |
| 项目主要研究内容、科学发现点、科学价值、同行引用及评价等（限1000字） |
| 数据挖掘(Data Mining)是人工智能、计算机科学和应用数学等领域的核心交叉学科。支持向量机(Support Vector Machines, SVM)是当前公认最有效的数据挖掘方法之一。非平行支持向量机(NSVM)将SVM构造的“平行超平面”推广至“非平行超平面”，是近几年来SVM方法的新突破。它在处理“交叉型”和“异方差噪声型”数据问题上突显出了良好的泛化能力，且具有较少的训练时间，因此受到众多知名学者的关注。然而，与 SVM 相比，NSVM的研究还很不充分，尤其是其最优化理论和模型方面的研究还很欠缺。OPTIMAL机器智能优化科研团队是国内首支研究非平行支持向量机的团队。该团队重点围绕在非平行支持向量机(NSVM)在大规模数据挖掘中存在的结构化学习问题，从最优化模型的构建、样本结构信息的提取、结构学习算法的设计到快速求解算法等展开深入研究，主要内容如下。1、NSVM最优化理论与模型：首次提出面向多标签问题的NSVM学习模型；基于聚类拟合思想，首次提出训练/预测一致性的NSVM学习框架；基于结构风险最小化原则，提出特征值分解支持向量机模型。2、结构信息化学习：使用聚类技术，捕获样本潜在的簇结构信息，并将其作为正则项引入到NSVM学习框架中；针对不均衡分类问题，根据样本分布的局部结构加权信息，提出一个有效的NSVM不均衡学习模型。3、NSVM模型求解算法：提出基于超松弛(SOR)算法的NSVM模型求解算法；针对大规模分类问题，引入加权线性损失，提出WLTSVM模型及快速求解算法。本课题将为NSVM的研究和应用提供理论、方法和技术支持。上述相关研究成果发表在包括Pattern Recogn.、Inform. Sciences.、Knowl-Based Syst.、Neurocomputing等在国内外数据挖掘权威期刊上，已发表学术论文35余篇，其中被科学索引SCI论文20余篇，其中二区SCI检索8篇，共计引用次数400余次(Google学术)，获批国家自然科学基金2项(项目批准号：61603338和11426202)，浙江省自然科学基金3项(项目批准号：LQ13F030010、Y15F030044和LQ17F030003)，2016年浙江省自然科学基金委对该项目的阶段性成果进行专题报道。 |

**四、第三方评价**

|  |
| --- |
| 评价结论等（限2400字） |
| 浙江省自然科学基金委报道：http://www.zjnsf.gov.cn/h/01/news\_view.aspx?appid=GUDB66564C022D3371B0142A60793B4291B浙江工业大学陈伟杰博士在浙江省自然科学基金项目“基于最大间隔和结构信息的非平行超平面SVM模型及算法研究”(项目批准号：LQ13F030010)的资助下，围绕非平行支持向量机(NSVM)中的结构化学习问题，从最优化模型的构建、样本结构信息的提取及模型求解算法等方面展开深入研究，取得了一系列阶段性成果。在基金的支持下取得如下最新成果有：针对多标签学习问题，首次提出多标签的非平行SVM学习模型(MLTSVM)；基于聚类拟合思想，提出非平行SVM的训练/预测一致的学习框架，并在此基础上提出NHSVM模型；针对GEPSVM训练/预测度量不一致问题，提出了一致的拟合超平面分类方法(PCC)；针对半监督学习问题，在GEPSVM优化模型中引入结构信息正则项，首次提出半监督GEPSVM学习模型(MPSVM)；针对PU问题，通过构造正类与未标签样本局部近邻图，首次提出非平行SVM的半监督学习模型(LUHC)；针对不均衡分类问题，根据样本分布的局部结构加权信息，提出一个有效的非平行SVM不均衡学习模型(WLTSVM)。上述相关研究成果发表在包括Pattern Recogn.、Inform. Sciences.、Knowl-Based Syst.、Neurocomputing等在内的国际权威期刊上，其中被科学索引SCI论文10篇(4篇TOP SCI、1篇ESI高引用论文)；共计引用次数200余次(Google学术)；并作为负责人后续获批国家自然科学基金2项(项目批准号：61603338和11426202)。 |

**五、主要完成人员情况**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **排名** | **姓名** | **行政****职务** | **技术****职称** | **现从事****专业** | **工作单位** | **二级****单位** | **完成****单位** | **对本项目主要科学发现的贡献** |
| 1 | 陈伟杰 | 无 | 副高 | 计算机科学与技术 | 浙江工业大学之江学院 | 信息工程学院 | 浙江工业大学之江学院 | 本人在主要学科提出非平行SVM及弱监督学习的最优化模型构建(第1和第2点)，主要贡献为对数据挖掘方法NSVM的训练与预测学习框架给出统一的“间隔”定义；首次提出多标签问题下的NSVM最优化模型及其几何解释；将局部结构化信息引入学习框架中，提出一种基于局部“簇粒度”的NSVM学习模型；提出大规模数据问题的快速求解算法等。参与发表论文专著代表作1-6。 |
| 2 | 张聚 | 无 | 正高 | 计算机科学与技术 | 浙江工业大学之江学院 | 信息工程学院 | 浙江工业大学之江学院 | 本人在主要学科提出了非平行SVM在医学图像中的应用(第3点)，主要贡献为针对医学图像分析中存在异构噪声问题，首先通过小波函数提取图像特征，然后再将NSVM模型推广到医学图像分析识别中，最后设计了大量的实验验证了该方法的可行性。参与发表论文专著代表作7。 |
| 3 | 邵元海 | 无 | 副高 | 金融大数据 | 海南大学 | 经济与管理学院 | 浙江工业大学之江学院 | 本人在主要学科提出了非平行SVM及弱监督学习的最优化模型构建(第1和第2点)，主要贡献为针对现有非平行SVM的训练与预测模型不一致问题，基于聚类拟合思想，导致学习器的性能受到较大影响，提出一个统一的训练/预测学习框架；为解决分类问题中的不均衡问题，通过构造类内相似图和类间判别图来挖掘数据分布结构信息，然后在Lagrangian TWSVM模型中引入加权偏置项来解决样本间的不均衡。参与发表论文专著代表作1-6。 |
| 4 | 李春娜 | 无 | 副高 | 信息与计算科学 | 浙江工业大学之江学院 | 理学院 | 浙江工业大学之江学院 | 本人在主要学科提出了非平行SVM快速学习算法(第3点)，主要贡献为针对大规模分类问题，引入加权线性损失，提出基于加权线性损失的TWSVM(WLTSVM)。该模型的解仅需要求解简单的线性方程组，计算复杂度低；大量的实验验证该模型在解决大规模分类问题上的有效性。。参与发表论文专著代表作5。 |

**六、主要完成单位情况**

|  |  |  |
| --- | --- | --- |
| **排名** | **单位名称** | **对本项目的支撑作用情况** |
| 1 | 浙江工业大学之江学院 | 该项目依托于我单位的浙江省一流学科B类(计算机科学与技术)、浙江工业大学之江学院优势学科(计算机科学与技术)和大数据研究中心。我单位为该项目的完成提供灵活的长期考核机制和宽松自由的学术氛围，提供了良好的科研环境（实验室条件）和良好的科研实验设备，提供国内外数据挖掘学术交流、培训机会，积极推荐课题组参与各种数据挖掘相关的学术组织任职和活动的机会，并对课题组所在学科和实验室给予政策上的倾斜和经费支持，促进其所在学科的快速发展。 |
|  |  |  |
|  |  |  |
|  |  |  |

**七、代表性论文专著目录（不超过10篇）**

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **序号** | **论文专著名称/刊名** | **影响****因子** | **年卷期****页码** | **发表****时间（年、月）** | **通讯****作者** | **第一****作者** | **所有作者** | **SCI****他引次数** | **他引****总次数** | **是否省内完成** |
| 1 | MLTSVM: A novel twin support vector machine to multi-label learning / PATTERN RECOGNITION | 3.965 | 2016,52:61-74 | 2016-04 | 邵元海 | 陈伟杰 | Wei-Jie Chen,Yuan-Hai Shao, Chun-Na Li, Nai-Yang Deng | 24 | 24 | 是 |
| 2 | Nonparallel hyperplane support vector machine for binary classification problems / INFORMATION SCIENCES | 4.305 | 2014,263:22-35 | 2014-04 | 邓乃扬 | 邵元海 | Yuan-Hai Shao, Wei-Jie Chen and Nai-Yang Deng | 59 | 59 | 是 |
| 3 | Laplacian smooth twin support vector machine for semi-supervised classification / INTERNATIONAL JOURNAL OF MACHINE LEARNING AND CYBERNETICS | 2.692 | 2014,5(3):459–468 | 2014-06 | 陈伟杰 | 陈伟杰 | Chen, Wei-Jie; Shao, Yuan-Hai; Hong, Ning | 30 | 30 | 是 |
| 4 | An efficient weighted Lagrangian twin support vector machine for imbalanced data classification / PATTERN RECOGNITION | 3.965 | 2014,47(9):3158-3167 | 2014-09 | 邓乃扬 | 邵元海 | Yuan-Hai Shao, Wei-Jie Chen, Jing-Jing Zhang, Zhen Wang and Nai-Yang Deng | 49 | 49 | 是 |
| 5 | Laplacian least squares twin support vector machine for semi-supervised classification / NEUROCOMPUTING | 3.241 | 2014,145:465-476 | 2014-12 | 陈伟杰 | 陈伟杰 | Wei-Jie Chen, Yuan-Hai Shao, Nai-Yang Deng, Zhi-Lin Feng | 11 | 11 | 是 |
| 6 | Weighted linear loss twin support vector machine for large-scale classification / KNOWLEDGE-BASED SYSTEMS | 4.396 | 2015,(73): 276-288 | 2015-01 | 邓乃扬 | 邵元海 | Shao, Yuan-Hai; Chen, Wei-Jie; Wang, Zhen; Li, Chun-Na; Deng, Nai-Yang | 26 | 26 | 是 |
| 7 | Speckle filtering of medical ultrasonic images using wavelet and guided filter / ULTRASONICS | 2.377 | 2016,(6):177-193 | 2016-02 | 张聚 | 程芸 | Zhang, Ju; Lin, Guangkuo; Wu, Lili; Cheng, Yun | 14 | 14 | 是 |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |

**承诺：**第七部分的论文、专著用于报奖的情况，已征得未列入项目完成单位或完成人的作者同意。

**八、完成人合作关系说明（**附签字扫描件**）**

**无**

承诺：本人作为项目第一完成人，对本项目完成人合作关系及上述内容的真实性负责，特此声明。

 第一完成人签名： 

**九、知情同意证明（附签字扫描件）**

