

基于图的半监督学习方法综述

韩灵珊

(重庆师范大学计算机与信息科学学院 重庆 401331)

摘 要 半监督学习是机器学习中结合监督学习和无监督聚类方法的一类学习方法。基于图的半监督学习凭借其直观性得到了半监督学习领域专家的青睐。本文对常用的半监督学习方法进行了介绍和阐述,介绍了基于图的半监督学习的发展现状,并对未来基于图的半监督学习的发展做出展望。

关键词 基于图的半监督分类 机器学习 图方法

中图分类号: TP181

文献标识码: A

0 引言

基于图的半监督学习凭借其直观性也逐渐被更多的学者所研究和应用。本文主要介绍了目前使用较多的基于图的半监督学习的方法分类;介绍了基于图的半监督学习目前的研究成果及现状;最后给出基于图的半监督学习下一步更待研究的方向。

1 基于图的半监督学习方法分类

1.1 图的构造及正则化框架

首先利用样本集 X 构造一个无向加权图 G 。图当中的每个顶点代表了样本集中的样本,图当中边的权值表示了样本对和之间的相似度;构造完图之后,基于图的学习方法通常假设样本标签在图中的分布是平滑的,并由此根据边的连接情况使已标记样本的类别标签在整个图上不断传播并达到最终完成对未标记样本的类别标签的预测。通常,样本对之间的相似度采用高斯核函数来计算。

图模型构造好后,基于图的半监督学习算法需要定义一个函数 f 。我们将基于图的学习方法规范化,提出基于图学习的正则化框架。对于已标记样本,令 (f) 为损失函数,用来调节函数 f 分类标签时预测标签与真实标签值之间的损失或误差;令 (f) 为目标函数的调整项,使标签分布在整个图上并且有足够的平滑性,通常采用引入正则项的方法来确保。一般而言,基于图的学习方法通常都利用图的拉普拉斯性质作为目标函数的调整项,以确保标签能够平滑的在整个图上传递。

1.2 基于图的半监督学习方法分类

1.2.1 标签传播算法

在标签传播算法中,使用的损失函数为,其中表示预测标签概率,表示已标记样本的真实标签值,损失函数表示在标签传递的过程中应当使预测的已标记样本的标签与真实标签类别相同;在调整项中使用 (f) 作为保障标签在整个图上的分布具有平滑性的调整项。

1.2.2 图的最小分割方法

图的最小分割方法(graph mincut algorithm)是由 Blum A 在 2002 年提出的。它的主要思想是:在二分类问题中定义正标记样本作为源点(source),负标记样本作为汇点(sink),目标是:找到一个边集,使得删除该边集之后能够隔绝任意从源点到汇点的流量,并且最终找到的这个边集为最小边集。那些与源点连接的点被标记为正类,与汇点连接的点则被标记为

负类。

1.2.3 调和函数方法

基于高斯域(Gaussian fields)和调和函数(harmonic function)的方法,简称为调和函数方法,针对在图的最小分割方法中未考虑样本的分类概率的硬划分(hard classification)的问题,采用了软划分(soft classification)的方法,将样本的类别用取值连续的变量表示。

1.2.4 局部全局一致性算法

Zhou 等人在标签传播算法和调和函数方法的启发下,提出基于局部与全局一致性的方法(learning with local and global consistency),简称 LGC 算法。LGC 算法的调整项采用了对称拉普拉斯矩阵,提高了分类的精度。保持局部一致性的目标就是要使该调节项最小。与调和函数的目标函数不同,LGC 算法的损失项允许预测标签与真实标签之间有一定的误差,并会使这种误差最小化,使用这样的方式保持样本集的全局一致性。

2 基于图的半监督学习方法研究现状

国外学者对基于图的半监督学习研究起步较早。Yang 等人在 2007 年时首次提出了利用 LPA 算法进行英汉双语信息检索;Raghavan U 则在同年用图方法进行网络社区发现,用空手道俱乐部网和美国大学橄榄球网的实验证明了其良好的检测效果;此外,在降维研究方面也有不少较为成熟的成果:2004 年,Argyriou 等采用 kd 树方法构造稀疏图,通过线性系统的迭代计算加速分类学习的速度,Delalleau 等通过基于所选样本集的子集进行标记传播并利用所选样本与剩余样本的关联降低图拉普拉斯矩阵的大小提出了一种无参数且支持直推式学习的算法。

我国在基于图的半监督学习的研究方向上起步较晚,但发展迅速取得了不少成果。一方面,对算法本身进行了深入研究和改进。例如:王雪松等人在原算法基础上提出了一种简洁的优化算法,通过使用 k 近邻图代替全连接图并且简化目标函数,减少了参数造成的误差影响;李明等人利用一种基于密度的快速聚类的方法对样本数据先聚类后进行标签传递,通过实验最终证明在分类效果上该算法与原算法相比速度大幅提高;Wang 等人利用线性近邻传递思想,构建邻接矩阵,提高分类效果并取得了好的成果。

另一方面,基于图的半监督学习在其他学科领域发挥了

支柱作用:丁宇新等研究中采用了局部全局一致性学习方法,以“人人网”数据作为实验数据,对用户的兴趣与毕业学校进行预测;新浪微博也同样利用了标签传播算法作为其背后的核心算法之一,进行更精准的广告投放和内容投送。

3 结论

通过对目前基于图的半监督学习取得的进展和成果了解分析,从研究内容来看:基于图的半监督学习的基础理论研究已经成熟,并且其成果已经应用于许多实际问题中。如今,如何利用图论知识优化构图,寻找提高学习算法效率,减少计算开销的新思路成为基于图的半监督学习的热点,也为今后的学习研究提供了大的发展空间;同时,如何将基于图的半监督学习方法联系到实际情况中,利用该方法对实际问题进行更好地挖掘和探索,从而利用隐含信息获得知识。

参考文献

- [1] Chappelle O,Scholkopf B.Semi-supervised learning[M].Cambridge:MIT Press,2006:193-196.
- [2] Zhu X.J.and Ghahramani Z. Learning from labeled and unlabeled data with label propagation [R].Technical Report CMU-CALD-02-107.Carnegie Mellon University,2002:1-8.
- [3] Blum A,Chawla S.Learning from labeled and unlabeled data using graph min-cuts.Proceedings of the 18th International Conference on Machine Learning.

Williamstorn,USA:Morgan Kaufmann Publisher,2001:19-26.

- [4] Zhu X,Ghahramani Z,Lafferty J.Semi-supervised learning using Gaussian fields and harmonic functions[C].Proceedings of the 20th International Conference on Machine Learning.Washington:[s.n.],2003:912-919.
- [5] Zhou D,Bousquet O,Lal T N,et al.Learning with local and global consistency [C].Thrun S,Saul L,Scholkopf B,et al.Advances in Neural Information Processing Systems 16.Cambridge: MIT Press,2004:321-328.
- [6] YANG L P,JI D H.Information Retrieval Using Label Propagation based ranking[C].In: Proceedings of NTCIR-6 Workshop Meeting,Tokyo,Japan,2007: 140-144.
- [7] RAGHAVAN U N,ALBERT R. KUMARA S. Near linear time algorithm to detect community structures in large-scale networks[J].Physical Review E. 2007(76):1-12.
- [8] Argyriou A.Efficient Approximation Methods for Harmonic Semi-supervised Learning [D].University College London,2004.
- [9] Delalleau O,Bengio Y.and Roux N L Efficient Non-parametric Function Induction in Semi-supervised Learning[A].Proceedings of the 10th International Workshop on Artificial Intelligence and Statistics [C].New Jersey,USA:Society for Artificial Intelligence and Statistics,2005:96-103.
- [10] 王雪松,张晓丽,等.一种简洁局部全局一致性学习[J].控制与决策,2011,26 (11):1726-1734.
- [11] 李明.基于局部聚类与图方法的半监督学习算法[J].自动化学报,2010,36 (12):1655-1660.
- [12] Wang,Zhang.Label propagation through linear neighborhood[J].IEEE Trans on Knowledge and Data Engineering,2008,20(1):55-67.
- [13] 丁宇新,肖晓,等.基于半监督学习的社交网络用户属性预测[J].通信学报, 2014,35(8):15-22.

(上接第139页)发广大员工为企业安全稳定发展作贡献的积极性,树立起强大的治安安全思想防线;采取妥善有力的思想政治工作方法,协调各单位、各部门的工作,调动全员参与到维护治安稳定的大局中来,齐心协力解决影响企业治安稳定的难点、热点、重点问题,进而把企业建设成一个安全、稳定、文明、和谐的先进单位。平衡好经济效益和企业安全保卫和思想政治氛围的培养,做到业务工作和思想工作的双方面的重视。

2.2 建立健全思想政治工作机制

正是因为武装保卫部门工作的特殊性,因此企业必须要加强武装保卫的队伍建设,要将思想政治工作渗透到武装保卫工作的各项机制中。同时,要加强武装保卫部门的组织建设,及时了解和关心武装保卫人员的思想动态,针对每个企业武装保卫人员进行思想政治修养的培养。当然,建立健全思想政治工作机制还要加强对于武装保卫人员的监督和考核,加强思想政治教育,形成针对武装保卫人员的考核制度,通过考核赏罚分明,激发武装保卫人员工作的热情和积极性。

2.3 充分发挥激励功能,调动武保人员工作积极性

思想工作首先要“尊重”上下功夫。一靠素质。先从自身做起,尊重基层工作的武保人员。任何时候都要摆正自己的位置,严格要求自己,认真约束自己的行为,虚心听取意见、建议,尊重他们的权利、他们的想法,并要端正态度,平等待人,以诚相待,有针对性地为他们分析某种不被社会认可的现象,帮助他们转变观念,树立职业自信心;二靠理解。要

学会换位思考,设身处地为对方着想,急其所急,想其所想,帮其所需。多看别人的长处、优点,信任他们、肯定他们,多交流、多通气,在思想的沟通中逐渐达成共识,减少误会,增进了解和友谊,顺利地贯彻执行党的方针、政策,做好工作;三靠激励。激励是对人的价值和工作的认可与肯定。通过各种活动、比赛,让广大队员认识到自己的价值,再通过单位的各种物资、精神奖励,激发他们的工作积极性,工作热情,从而推动武装保卫工作向前发展。

3 结语

随着企业改革的深化,企业武装保卫工作对于企业的发展 and 人民生命财产安全的作用有越来越突出。因此,在新的历史形势下,必须要通过加强武装保卫人员的思想政治教育,建立健全相关的监督和考核体制,相关领导从根本上提高对于企业武装保卫人员思想政治工作的重视,提高企业武装保卫人员的素质,充分调动起工作的积极性,在各方面的共同努力下,保证企业稳定、和谐、有序的发展。

参考文献

- [1] 李积峰.发挥思想政治工作优势推动企业武装保卫工作[J].现代企业文化, 2013(26).
- [2] 郭卫军.新形势下政工干部如何做好基层武装保卫系统的思想政治工作 [J].东方企业文化,2014(6).
- [3] 郝志军.企业武装保卫思想政治工作探讨[J].企业文化(中旬刊),2015(5).
- [4] 张志英.浅谈如何加强煤炭企业武装人员的思想政治工作[J].东方企业文化,2013(22).