**新疆大学**

**专业学位研究生学位论文**

**中期考核报告**

攻读学位层次： □博士 ☑硕士

培养方式： ☑全日制 □非全日制

姓 名： 刘董事

学 号： 107552103296

专业学位类别及领域： 应用统计专业硕士

学 院： 数学与系统科学学院

论文题目： 基于专家知识的树宽有界

贝叶斯网结构学习

校内指导教师： 孙 毅

联合培养兼职指导教师：

填表日期： 2023 年 11 月 03 日

新疆大学研究生院

1. 已完成的主要工作

|  |  |
| --- | --- |
| **1.开题报告通过时间：**2022 年 11 月 20日 | |
| **2. 课程学习、政治思想及日常表现情况** | |
| 是否通过培养计划中规定课程的考核 | ☑是 □否 |
| 政治思想表现情况：（考核学习思想政治理论、坚持党的基本路线、拥护社会主义制度，遵守中华人民共和国宪法、法律、法规，遵守学校各项规章制度，思想品质等方面的表现）。  在研究生期间，思想方面，按时参加学校及学院组织的各项政治理论学习活动，遵守学校各项规章制度，思想端正，未曾做任何违法乱纪的事情。学习方面，认真完成考核任务， 未曾挂科，已经达到毕业所要求的学分要求，并且顺利通过了英语学业水平考试。日常表现方面，积极参加宿舍组织的志愿者活动，参与核酸检测等志愿活动; 另外随时保持寝室干净整洁，与室友关系融洽。综合来说，各方面表现良好。 | |
| **3. 论文研究进展** | |
| 从理论分析或计算部分、实验实践（或实证）工作等方面进行总结（可续页）。  （1）理论部分：  在贝叶斯网络结构学习中，随着网络中变量的数量呈指数增长，贝叶斯结构学习变得越来越具有挑战性。可能的有向无环图（DAGs）数量随着网络中变量的增加呈指数增长。这种指数增长在搜索所有可能的DAGs空间以找到最适合数据的最佳结构时带来了困难。然而，已经发现，在学习过程中引入基于先前知识和有限树宽的约束可以显著缓解这个问题。为此，在本文中，我们提出ET-con（约束hc-ET）方法。它是在 hc-ET 的基础上引入了如下的专家约束;  1.如果 x→y , 将相应的边一一加入到 G 中, 若在加边后 G 的树宽超过给定的树宽, 取消本次加边.  2.如果 x ↛ y , 对于相应的 y, 设置禁止父集（在爬山算法中不产生这条边）.  3.如果 x - y , 判断两个方向的得分, 选择得分高的方向, 将此边加入 G 中.  4. 对于祖先约束（X ⇝ Y ）：在爬山的每次迭代中, 我们选择 G 的一个改进 G′, 用 G′ 代替 G. 即当且仅当 φ (G′) > φ(G) 或者 φ(G) = φ (G′) ,s(G′) < s(G). （φ(G) 表示 DAG G 所满足的祖先约束X ⇝ Y 的数量,s(G) 表示 G 的得分.）  该方法同时满足三个要求：有界树宽、专家约束和可伸缩性。其通过采用基于评分的结构学习搜索并整合先验信息，使搜索空间得到了显著减少。  算法设计如下：   |  | | --- | | **算法1**: 约束消树算法 (ET-con) | | **输入**：X–一组属性变量; D– 关于 X 的数据; s– 评价函数; G0– 初始网络结构; C –边和路 径的约束集; 树宽界限–tb | | **输出**：BN 结构 G | | 1. G ← G0 ;θ ← G 的参数最大似然估计; oldScore ←s(G,θ|D); | | 2. flag ← true | | 3. for flag do | | 4. newScore ←−∞ ; G∗ ← null ; θ∗ ← null; | | 4. c1,...,cl 是被应用于 G 的局部改变 (即, 加边、减边、转边) ; 且 c1,...,cl 均不在满足 x ↛ y 的约束集合中; | | 5. for d = 1,...,l do | | 6. 让G′ 为应用cd 于G的结果. | | 7. 为 G′ 搜索低宽度的消去序 π′ | | 8. if width (G′, π′) ≤ tb and φ (G′) >= φ(G) then | | 9. θ′ ← G′ 的参数最大似然估计; tempScore←s(G′,θ′ |D); | | 10. if (tempScore > newScore): then | | 11. newScore ← tempScore; G∗ ← G′ ; θ∗ ← θ′  end if | | 12. end if  end for | | 13. if newScore > oldScore: then | | 14. G ← G∗;θ ← θ∗ ; oldScore ← newScore; | | 15. end if 16. if newScore <= oldScore: then | | 17. Return (G,θ)  end if  19.end for | |  |   （2）实验模拟部分：  数据集：我们选择的基准数据集由包含在 bnlearn中的 8 个不同大小的贝叶斯网采样而来, 包括: 2 个小型网络（包含 0-20 个变量）, 1 个中型网络 (包含 20-50 个变量), 2 个大型网络 (包含 50-100个变量), 5 个超大型网络 (包含 100-1000 个变量), 以及 1 个巨型网络 (大于 1000 个变量).  对比算法及评价指标：选择 k-MAX、k-greedy、hc-ET与ET-con 算法在网络结构和网络与真实网络的相似度上进行比较, 相对应的评价指标是网络结构的准确率、结构的查全率、BIC 评分、算法效率等  实验结果：通过在小型网络 Asia、中型网络 Alarm、大型网络 Win95pts 和特大型网络 Munin 等网络上对比其他算法, ET-con 算法在效率和算法正确率上提高了 10% 以上, 同时在 BIC 得分上也略微提升。  下面是在树宽为3，5，7时的部分实验结果：  表1: 树宽为3的实验结果   |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | | 数据集 | 算法 | BIC 时间 | EP | EC | EF | 正确边个数 | |  | ET-con | 1338301.46 | 118.906 | 0.551 | 0.467 | 526 | | link | hc-ET |  |  |  |  |  | |  | k-greedy |  |  |  |  |  | |  | k-MAX |  |  |  |  |  | |  | ET-con |  |  |  |  |  | | numin | hc-ET |  |  |  |  |  | |  | k-greedy |  |  |  |  |  | |  | k-MAX |  |  |  |  |  | |  | ET-con |  |  |  |  |  | | numin1 | hc-ET |  |  |  |  |  | |  | k-greedy |  |  |  |  |  | |  | k-MAX |  |  |  |  |  |   表2: 树宽为5的实验结果   |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | | 数据集 | 算法 | BIC 时间 | EP | EC | EF | 正确边个数 | |  | ET-con | 1338301.46 | 118.906 | 0.551 | 0.467 | 526 | | link | hc-ET |  |  |  |  |  | |  | k-greedy |  |  |  |  |  | |  | k-MAX |  |  |  |  |  | |  | ET-con |  |  |  |  |  | | numin | hc-ET |  |  |  |  |  | |  | k-greedy |  |  |  |  |  | |  | k-MAX |  |  |  |  |  | |  | ET-con |  |  |  |  |  | | numin1 | hc-ET |  |  |  |  |  | |  | k-greedy |  |  |  |  |  | |  | k-MAX |  |  |  |  |  |   表3: 树宽为5的实验结果   |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | | 数据集 | 算法 | BIC 时间 | EP | EC | EF | 正确边个数 | |  | ET-con | 1338301.46 | 118.906 | 0.551 | 0.467 | 526 | | link | hc-ET |  |  |  |  |  | |  | k-greedy |  |  |  |  |  | |  | k-MAX |  |  |  |  |  | |  | ET-con |  |  |  |  |  | | numin | hc-ET |  |  |  |  |  | |  | k-greedy |  |  |  |  |  | |  | k-MAX |  |  |  |  |  | |  | ET-con |  |  |  |  |  | | numin1 | hc-ET |  |  |  |  |  | |  | k-greedy |  |  |  |  |  | |  | k-MAX |  |  |  |  |  |   （3）学位论文结构:  第一章: 说明了文章的研究目的和研究意义, 并介绍了贝叶斯网结构学习在国内外研究状况, 以及给出本文的研究内容和论文结构.  第二章: 主要介绍了图模型理论中的一些基本概念并给出本文研究内容相关的理论知识.  第三章: 介绍基于专家知识的树宽有界贝叶斯网结构学习，进行相关的模拟实验和实际数据分析。  第四章: 介绍基于专家知识的结构 EM 算法。进行相关的模拟实验。  第五章: 对本文所研究的内容作出总结并给出未来还可以研究的方向. | |
| **4. 阶段性研究成果** | |
| 按《新疆大学研究生学位论文写作指南》的格式要求分类填写与学位论文相关的阶段性研究成果，例如期刊论文、会议论文、科研获奖、专利、制定标准等，限填本人第一作者或导师为第一作者本人第二作者成果，其中已录用、已投稿或拟投稿的在括号内注明（可续页）。  本人于 2022 年成功参加了第十九届数学建模竞赛，获得了成功参赛奖。  D.S. Liu, Y. Sun Bounded tree width Bayesian network structure learning based on expert knowledge [J]. International Journal of Approximate Reasoning, (拟投稿) | |

1. 存在的主要问题和解决办法

|  |
| --- |
| 1.未按开题计划完成的研究工作，研究工作存在的原理性、技术性难题以及在实验条件等方面的限制（可续页）。  1.在论文写作过程中，遇到的问题是实际数据分析部分没有在开题计划时间内完成，  2.另外一篇论文的内容不足以完成毕业论文。 |
| 2.针对上述问题采取何种解决办法，对学位论文的研究内容及所采取的理论方法、技术路线和实施方案的进一步调整，以及下一步的研究研究计划（可续页）。  针对上述问题，解决办法如下:  首先，在糖尿病数据上进行实验，并进行分析。其次，将专家约束引入到结构 EM 算法中，以填充毕业论文的内容。  下一步的研究计划:  • 2023 年 11 月-2023 年 12 月，根据导师的意见修改小论文，并将其投稿。  • 2023 年 12 月-2024 年 2 月, 将专家约束引入到结构 EM 算法中，同时进行实验验证和实际数据数据分析工作，完成毕业论文的第三、四章的撰写。  • 2023 年 2 月-2023 年 3 月，完成毕业论文初稿，根据指导老师意见，完善论文的修改工作，为毕业论文预答辩做准备。  • 2024 年 3 月-2023 年 5 月，根据预答辩评审老师的意见，提升毕业论文质量和专业水准，完成毕业论文答辩。 |

1. 中期考核审查意见

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **1.导师（组）对工作进展及研究计划的意见：** | | | | | | |
| 该生学习态度端正，思想积极上进、学习认真刻苦。在硕士论文研究期间，能根据就读专业方向的要 求，有针对性地认真研读文献，掌握本研究课题的基础理论知识，并且能按照要求按时完成论文创作。目 前，论文进展顺利、工作安排得当，能按照计划完成毕业论文写作。  校内导师（组）签字： 2023年 11月 10日  联合培养兼职导师（校外导师）签字： 年 月 日 | | | | | | |
| **2.中期考核专家组意见** | | | | | | |
| 考核日期 | 2023年11月4日 | | 考核地点 | | 数学楼 118 教室(腾讯会议) | |
| 考  核  小  组  成  员 | 姓名 | 职称 | | 现工作单位 | | 本人签名 |
| 于海征 | 副教授 | | 新疆大学 | |  |
| 梅雪晖 | 副教授 | | 新疆大学 | |  |
| 木拉提 | 讲师 | | 新疆大学 | |  |
|  |  | |  | |  |
|  |  | |  | |  |
|  |  | |  | |  |
|  |  | |  | |  |
| 秘 书 |  |  | |  | |  |
| 考核结果 | 合格 3 票 基本合格 票 不合格 票 | | | | | |
| 结 论 | ☑通过 □原则通过 □不通过  **通过：**表决票均为合格。  **原则通过：**表决票中存在不合格或基本合格，且合格、基本合格票数达到三分之二及以上  **不通过：**表决票中合格、基本合格票数不足三分之二。 | | | | | |
| 对学位论文工作进展以及下一步研究计划的建议，是否适合继续攻读学位：  该生学位论文工作进展比较顺利，符合预期计划，具备一定的从事科学研究能力，学术道德较为规范、 端正。  建议:按指导教师安排进入学位论文撰写工作。  考核专家组组长签名：  2023年 11 月 12 日 | | | | | | |
| **3.学位评定分委员会意见：** | | | | | | |
| 学位评定分委员会主席签名： 年 月 日 | | | | | | |