

大创经验分享(科研向)

软件学院 张柏荣 2024.3.10

关于开展2023年度吉林大学“大学生创新创业训练计划”立项工作的通知

2023年05月09日 08:26 教务处

校内各教学单位：

为激发大学生的创新激情，提升大学生创新思维能力，学习创新方法，掌握创新手段，增强实践动手能力，根据《教育部高等教育司关于开展2023年国家级大学生创新创业训练计划立项和结题验收工作的通知》及《吉林大学大学生创新创业训练计划管理办法》（校发〔2022〕94号）的立项有关规定，学校将于2023年5月开展2023年度吉林大学“大学生创新创业训练计划”项目立项工作，现将具体事宜通知如下：

一、项目类型

大学生创新创业训练计划项目包括创新训练项目、创业训练项目和创业实践项目三类。

1. 创新训练项目：本科生团队在导师指导下，自主完成创新性研究项目设计、准备研究条件和项目实施、撰写研究报告、成果（学术）与学术交流等工作。

2. 创业训练项目：本科生团队在导师指导下，团队中每个学生在项目实施过程中扮演一个或多个具体的角色，完成商业计划书编制、可行性研究、企业模拟运行、撰写创业报告等工作。

3. 创业实践项目：学生团队在学校导师和企业导师共同指导下，采用创新训练项目或创新性实验等成果，提出具有市场前景的创新性产品或服务，以此为基础开展创业实践活动。

二、实施范围及申报要求

1. 实施范围

吉林大学“大学生创新创业训练计划”主要面向全校在籍二年级本科生。

2. 申报要求

吉林大学“大学生创新创业训练计划” 创新训练项目申请书

←

←

项目名称 _____ ←

项目负责人 _____ ←

所在学院、年级、专业 _____ ←

联系电话 _____ ←

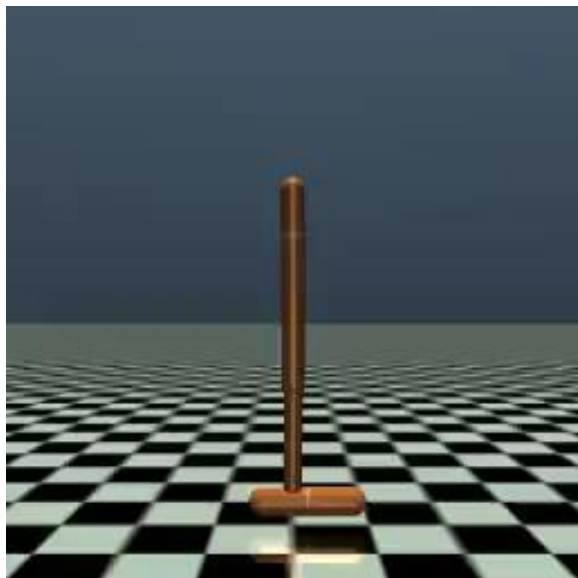
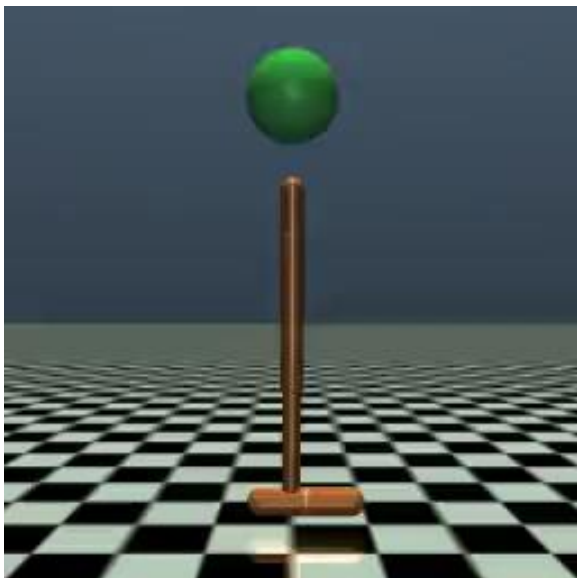
电子邮箱 _____ ←

指导教师姓名 _____ 职称 _____ ←

关于我的大创

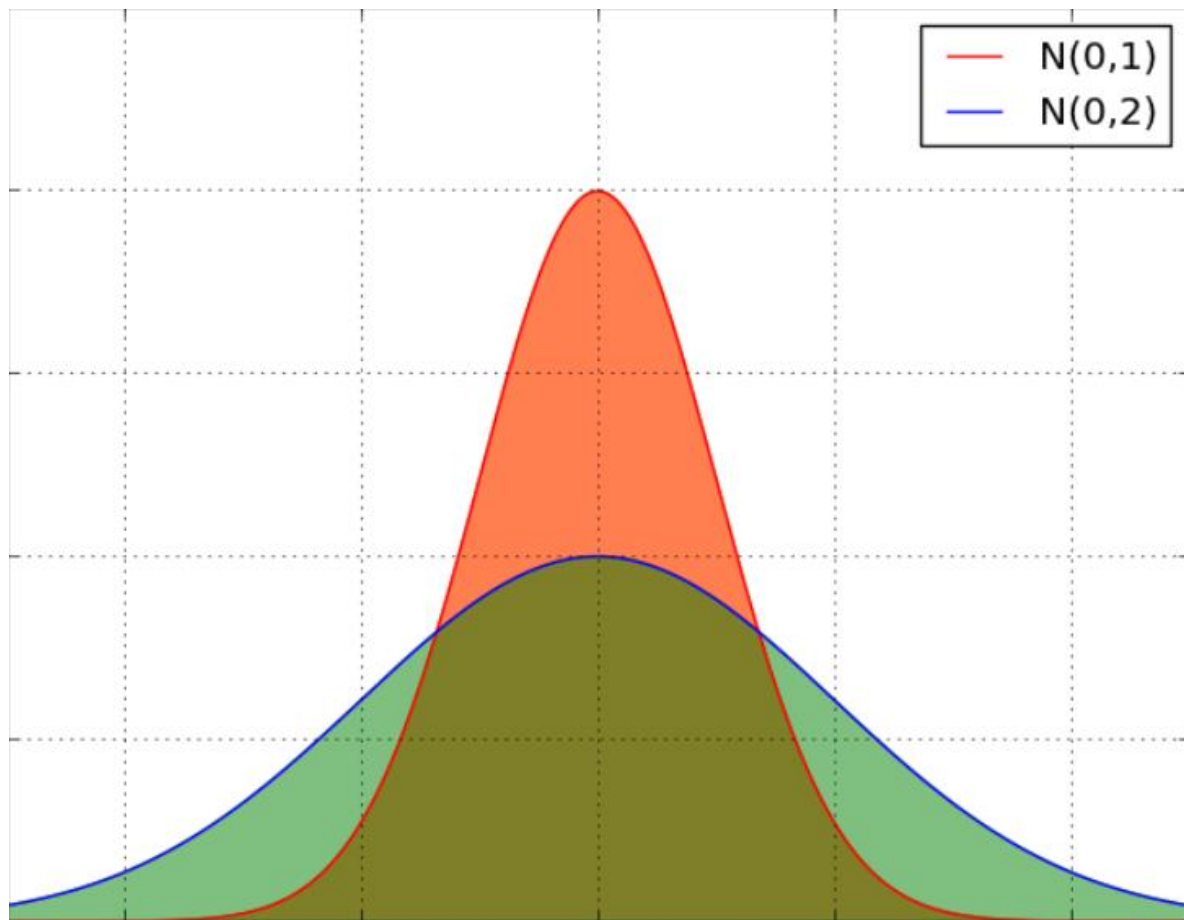
基于策略梯度安全强化学习算法的样本效率研究

- 2022年11月7日确定开始
- 2023年6月12日立项答辩
- 2023年7月4日确定国家级立项， 第一负责人



开始前

- 观点的方差
- 多听、多看、多问
- 知行合一



开始前

一些参考资料

- [中国计算机学会推荐国际学术会议和期刊目录](#)
- [计算领域高质量科技期刊分级目录](#)
- [中国计算机学会推荐中文科技期刊目录](#)
- 参考并借鉴了一些之前的分享，在此感谢：
 - 吕帅老师于2022年4月15日的分享（侧重于科研规范，筹备、规划项目的整套细致流程）
 - PPT链接：<https://pan.baidu.com/s/1j-GP42QXVMgN1l55xtSQVg> 提取码：8zfw
 - 郎骐玮同学于2023年4月21日的分享（分享了很多个人经验以及资源）
 - PPT链接：<https://pan.baidu.com/s/1SGXeXSuD0ErwbpnPE8R1zA> 提取码：iq4d

大创概述

背景

- 大创全名：**大学生创业创新训练计划**，是教育部决定在“十二五”期间实施的国家级计划。
- 课外培养计划之一，**有课外学分**
- 包括：创新训练项目、创业训练项目、创业实践项目
- 主要面向大二的本科生（其它年级也可以参与）
- 必须组队申报（3-5人），支持鼓励跨专业跨学科组队
- **每人只能同时参加一个项目**，项目完成时间基本为1-2年

(from郎骐玮同学)

大创概述

时间节点

- 4-5月（一般为5月中旬前后）：立项评审，确定立项级别【国家级、省级、校级】
- 9-10月：中期检查/定级评审【提交中期检查报告】
- 第二年4-5月、9-10月：进度检查/结题验收
- 结题项目遴选优秀候选项目（不超过结题数的20%、10%，取整），上报学部 PK确定优秀结题

(from郎骐玮同学)

软件学院本科生素质类项目加分细则

类型	加分项目		加分对象	认定加分值	备注
一、科研成果	大学生创新创业计划项目	国家级优秀结题项目	负责人	0.1 GPA	学院认定，排名要与立项、中期检查、结题环节一致
			第二名	0.05 GPA	
	学术论文 (计算机类)	中国计算机学会推荐国际学术会议和期刊目录(最新) 中国科学院文献情报中心期刊分区(最新)	● 业界公认学科综合顶级期刊	0.4 GPA	学院认定，必须是第一作者，署名第一单位必须是吉林大学软件学院，加分前要求答辩，由推免工作小组组织相关专家认定。
			● CCF-A 期刊 ● CCF-A 会议长文 ● CCF-B 会议的最佳论文/最佳学生论文	0.2 GPA	
			● CCF-B 期刊论文 ● 中科院期刊分区一区论文 ● CCF C 会议的最佳论文/最佳学生论文	0.1 GPA	
			● CCF-C 期刊论文 ● CCF-B 会议长文 ● 中科院期刊分区二区论文 ● 计算机学报、软件学报发表的学术论文	0.05 GPA	
			● 影响因子非 0 的 SCI 检索期刊论文 ● CCF-C 会议长文 ● 中国科学：信息科学、计算机研究与发展、计算机辅助设计与图形学学报、电子学报中文版、自动化学报发表的学术论文	0.02 GPA	
			● EI 检索期刊	0.01 GPA	

大创概述

大创成果

- 论文
 - 会议（文章较短，投稿时间固定，审稿快）
 - 期刊（文章较长，全年可投稿，审稿周期长）
- 软著
- 专利
 - 实用新型专利（6个月左右）
 - 发明专利（一年半左右）

做大创之前

目的

- 个人成长：

快速熟悉新领域、上手新工具，代码能力，工程思维，掌握第一性原理…

- 功利导向：

论文，保研加绩点，丰富简历（保研/出国），课外学分…

- 别人都做我也做

- …

第一性原理

- 课程学习：

掌握具体的知识，细节要完全准确无误。

- 实际项目：

知识细节服务于项目目标，只需要了解必要的细节和大的框架。

- ...

第一性原理 实例

```
1  if __name__ == '__main__':
2      eg = ExperimentGrid(exp_name='Example')
3
4      # Set the algorithms.
5      example_policy = ['PPOlag', 'TRPOLag']
6
7      # Set the environments.
8      eg.add('env_id', 'SafetyAntVelocity-v1')
9
10     eg.add('algo', example_policy)
11     eg.add('train_cfgs:torch_threads', [1])
12     eg.add('algo_cfgs:batch_size', [32, 64, 128])
13     eg.add('logger_cfgs:use_wandb', [False])
14     eg.add('seed', [0])
15     # total experiment num is better to be divisible by num_pool
16     # meanwhile, users should decide this value according to their machine
17     eg.run(train, num_pool=6, gpu_id=None)
```

自动化参数调优：在多个参数组成的高维向量空间上找到一个最好(相对)的取值。

问题：

1.自动分析？

人脑分析

2.人脑足够分析吗？

简单实例下可以

3.如何唯一标识一组实验？

把参数字符串整个做哈希

第一性原理 实例

```
SafetyAntVelocity-v1---1f58ce80fc9540b32a925d95694e3f836f80a5511e9e5c834e77195a2e9c3944
SafetyAntVelocity-v1---7a451ea3e08cfb6caf64d05c307be9b6c32a509dc425f758387f90f96939d720
SafetyAntVelocity-v1---7cefb92954e284496a08c3ca087af3971f8a37ba1845242208ef2c6afcaf4d27
SafetyAntVelocity-v1---564ef55d6dac0002b8ecf848a240fe05de8639cc33229b4f773157dd2f828e71
SafetyAntVelocity-v1---9997d3e3b2555d9f0da2703b24b376aa5ddd73d8abaffe95288b23bfd7304779
SafetyAntVelocity-v1---50699a2818176e088a359b124296d67ac6fb130336c5f7b66f356b34f361e356
```

After the experiment is finished, you can use the `/examples/analyze_experiment_results.py` script to analyze the results. For example, to plot the average return/cost of the `SafetyAntVelocity-v1` environment, you can set the `/examples/analyze_experiment_results.py` file as follows:

```
1  # just fill in the path in which experiment grid runs.
2  PATH =
3  if __name__ == '__main__':
4      st = StatisticsTools()
5      st.load_source(PATH)
6      # just fill in the name of the parameter of which value you want to compare.
7      # then you can specify the value of the parameter you want to compare,
8      # or you can just specify how many values you want to compare in single graph at most,
9      # and the function will automatically generate all possible combinations of the graph.
10     # but the two mode can not be used at the same time.
11     st.draw_graph(parameter='algo_cfgs:batch_size', values=None, compare_num=3, cost_limit=None
```

自动化参数调优

问题：

3.如何设计一个自动呈现简单判断的程序交互方式？

- ① 每次只对比一个参数上的不同取值，例如学习率=0.1,0.01,0.001的不同情况，绘制在同一张图上。
- ② 支持指定每次显示在同一张图上的最大数量，例如学习率有12种取值，但是人脑最多同时看6条曲线，那就绘制所有6条曲线的组合。

做大创之前

能力&自我评估

- 准备好付出什么？

时间（会占用课业、娱乐时间），精力（免不了熬夜、焦虑），挫败（不一定有成果）...

- 是否能持之以恒？

有没有钻研的热情，是否对科研或工程感兴趣。

- 是否准备好了持续学习？

可能你过去学的每一门课程都与大创项目无关。

- 成绩跟大创是否有必然联系？

大创更侧重实践，需要探索性地解决问题、而非掌握既定模式。

- ...

做大创之前

队友

- 什么样的队友适合你？

大创能让你更好地认识自己和队友。

- 什么样的队友是合格的队友？

沟通顺畅、能够独当一面、可以一起讨论，提供见解。

- 如何进行团队沟通？

多多换位思考，目标一致。

做大创之前

老师

- 如何找老师？
 - 途径：自己上的课 / 学长或同学推荐 / 官网找
 - 方向：尽量和你想做的方向一致或相近
 - 要找对学生负责的老师
- 如何跟老师沟通？
 - 邮件礼仪
 - 信息密度
 - 换位思考

做大创之前

老师

- 老师的职责是？
 - 把控整体方向。
 - 提供探索方向的指导。
 - 针对立项书和论文提供反馈，指出问题。
 - ...

但**绝对**不是手把手一对一辅导。

- 想象中的老师跟实际当中的老师？
 - 老师很忙，你需要主动跟老师交流。
 - 要对老师的时间负责，不要进行无谓的沟通。
 - 要向老师展现出自驱力。

做大创之前 选题

兴趣，但不完全兴趣。——很多“感兴趣”源于不了解。

挑战，但不完全挑战。——难度太大容易放弃。

- 开发？

相对简单，学习既有模式，经验驱动。

- 工程？

需要有一定创新，相比于开发更加“独一无二”。

- 科研？

需要创新，为社区提供独到的见解。

做大创之前 选题

- 如何评价一个方向
 - 兼听则明
 - 从自身角度出发，去思考这个过程当中你能学到什么、获得什么。
 - 这个方向的学习资料是不是已经很成熟，你能不能建立起知识壁垒？
 - 实际应用的前景、难度——本科生很难做非常基础(困难)的研究。

大创立项

立项前-选择具体题目

- 自己想做的：自由度较大，但老师接收难度也较大
 - 目前的热点方向。
 - 自己想做的东西（如开发等）。
- 老师的课题：自由度较小，但更容易找
 - 子课题。

(from郎骐玮同学)

大创立项

立项前-前期准备

- 学习前置知识
 - 学科知识（如概率论）、编程语言等固定的工具。
- 对领域进行广泛调研
 - 综述（读英文文献，语言是最不重要的因素）。
 - 近几年的顶刊顶会论文(特指CCF-A分类当中的顶刊顶会)。
 - 努力克服畏难心态，并不是更差的论文就更容易懂，越好的论文越希望读者可以读懂。
 - 在各大视频网站，资源分享网站看知名学者的tutorial & talk。
 - 多与人交流。

大创立项

立项前-前期准备

- 多做一些工作，并且展示出你比别人更有能力完成高质量项目。
 - 调研深入程度。
 - 基础的代码实现。
 - 前期实验。
 - 问题建模、对于理论的理解/推导。

前期工作的质量和项目预期产出的数量是**决定项目等级的关键**。

——**预期产出意味着结项难度**

大创当中

立项后-科研过程

- 永远没有一个完全准备好的状态、随时准备迎接挑战。
- 方向确定好之后，少说，多做。
- 科研绝对不是一个让人心情激动的过程：
挫败感，自我怀疑，长期付出的疲劳感，延迟的回报（or没有功利性回报），队友之间的矛盾。
- 你做的东西很可能看起来是没有价值的（跟想象中比起来）。
- 细节决定成败（也决定了暴涨的工作量）。
- 如果一件事你自己没有把握只要花时间一定能做好，不要期待你的队友可以胜任。
- Arxiv\在投\有审稿意见\中稿的区别？

大创结项

具体形式

- 软著：周期在一个月到三个月不等，相对较水
 - 申请步骤：软团团教程，搜索“软著”（基本流程，细节可能有所不同）
- 论文：周期在3-7天、三个月、半年、一年不等
 - EI / SCI
 - EI 水刊水会：3-7天给结果
- 常规会议及期刊：会议 deadline 明确，周期相对短；期刊随时可投，周期相对长（看实际情况）
 - CCF A/B/C
 - 产出CCF A的周期正常在一年左右
- 专利：周期基本为一年

(from郎骐玮同学)

本科生科研的时间线探讨

- 本科的学习内容

 - 大一、大二：打基础

 - 大三：对于计算机领域有更完整的认知

- 本科的发展阶段

 - 大一、大二：学习课程内容，保证成绩、参加竞赛

 - 大三上下：开始了解不同的方向（考研的话大三下要开始考研了，保研的话，大三下已经在忙着找老师了）

 - 大四上：考研结束，保研后做老师项目。

 - 大四下：考研复试，保研后做老师项目。

个人杂谈

- 老师看重什么特质？

- 实际能力（基础）
- 沟通习惯（必要条件）
- 洞察力（加分项）

成绩是敲门砖，解决问题的综合能力是核心竞争力。

- 做题跟实际解决问题之间的gap？

- 做题是学习一个确定的过程，科研很可能是从0到1，你提出的每一个方案或见解都会遭到无数人的质疑。

- 选择的局部最优性？

- 多了解、多积累。

个人杂谈

- 合作伙伴的重要性？

- 只有很少的工作是不需要动脑，按照需求就能做的。
- 天才的代码也需要一行一行写。
- 建立可持续的、良好的合作关系对于完成项目和未来发展都至关重要。

（要把自己的个人统计量尽可能广的传播出去，最应该依靠的是由弱连接关系形成的人际关系网络。因此，那些只是认识你的人怎么看待你，对你来说最为重要。如果你在他们之中建立起了良好的口碑，你的声誉就会被广泛传播。而口碑的建立，靠的是自己踏踏实实的做好每一件事。你要知道，你在做每一件事情的时候，都有很多人看在眼里，记在心里。把每件小事都做好，就赢得了周围这些人的认可，就能逐步在弱连接网络中树立起自己的口碑。有了这样的口碑，不愁机会不上门。——徐高（经济学家）

- 学生思维？

- 觉得可以努力够一够的事情就去尝试，很多事情不知道最优解是什么，没有办法定义何时你准备好了。

个人杂谈

- 说话，写材料时遵循务实的态度
 - 如果一个事情你说不清楚，就不要写上去/说出来。
 - 说出来的一定是精确的，可以立即get到的表达。
 - 学术论文有固定的表达规范，要去遵循，不要标新立异（详情参考吕帅老师去年的PPT）。
 - 一定要避免使用假大空的词语和表达方式。
 - 例如（from吕帅老师的分享）：
 - 避免笼统概括，尽可能量化
 - 学习成绩优异 -> 平均绩点3.5/4.0 V 专业排名20/380
 - 避免常识性表述或冗余内容
 - 学习程序设计基础、离散数学、数据结构 -> CSP 300分
 - 避免不切实际的形容
 - 熟练掌握C++ -> LeetCode刷题***道
 - 去掉与项目无关的表述
 - **部副部、**班**干部；与项目不相关的文体爱好、社交能力等
 - 避免使用试图、尝试、力争、尽可能等前缀用词

个人杂谈

- 完美主义?
 - 科研就是在做trade off。——王恩老师
 - 人生也是在做trade off, 能往前走一点就往前走一点, 不要被完美主义缚住手脚!
- 如果你想发顶刊顶会, 马上开始!
 - 审稿周期长 (3-4个月) 。
 - 中稿难度大。
 - 本科阶段课业繁重, 时间线紧张。

个人杂谈

- 不要重复造轮子
 - 没有必要从第0行代码开始写，高效可靠地利用前人搭建好的工具/代码等资源可以大幅度减小你的工作量。
- 纸上得来终觉浅
 - 光读论文你会觉得都是对的，最好动手实操，复现。同时要思考，为什么这篇工作要这么做，必须这么做吗？搞清楚它存在的必要性，你才能对自己的方法有比较扎实的思考。

Q&A