

分类号 _____

密级 _____

UDC _____

编号 _____

中国科学院研究生院 工学博士学位论文

针对连续状态系统最优控制的在线强化学习算法研究

朱圆恒

指导教师 _____ 赵冬斌 研究员

_____ 中国科学院自动化研究所

申请学位级别 工学博士 学科专业名称 控制理论与控制工程

论文提交日期 2014年10月 论文答辩日期 2014年10月

培养单位 _____ 中国科学院自动化研究所

学位授予单位 _____ 中国科学院研究生院

答辩委员会主席 _____

Using Online Reinforcement Learning Algorithms to Solve Optimal Control Problems of Continuous-State Systems

Yuanheng Zhu

Supervisor:

Prof. Dongbin Zhao

Institute of Automation
Chinese Academy of Sciences

October, 2014

*Submitted in total fulfilment of the requirements for the degree of Doctor of
Engineering
in Control Theory and Control Engineering*

摘 要

最优控制不论是在控制理论研究还是在控制工程实践中一直是一个重要的研究课题。强化学习(Reinforcement Learning, RL)作为解决最优控制的一种有效方法一直受到研究者的关注。尤其是在线RL方法在系统模型未知的情况下能够通过在线学习得到最优策略,摆脱了对模型的依赖。然而,当前RL方法在理论研究和方法实现上还一直存在着许多未解决的问题。很多理论证明的缺乏和算法实现上的缺陷制约着RL方法的应用。为此,本项课题以在线RL方法作为主要研究对象,解决连续状态系统最优控制问题。同时补充离线RL方法的理论体系。本文的主要章节包含以下工作和贡献:

1. 首次给出了近似策略迭代(Approximate Policy Iteration, API)的收敛性证明。策略迭代作为离线RL方法的一种,在求解连续状态系统最优控制时需要采用逼近器去近似连续函数。因而不可避免地引入了逼近误差,算法收敛性受到了误差的干扰。经过本课题的研究分析,发现当逼近器满足一定条件时策略评估阶段的性能指标函数误差可以约束在一定范围内。通过策略提升后的计算结果与精确解是完全相等。证明了在误差干扰下API仍然能够收敛到最优解。

2. 提出一种在线策略迭代算法,以多项式作为基函数,使用线性参数化逼近器近似性能指标和策略函数。由于使用在线学习的方式,算法利用在线采集的数据训练逼近器参数,摆脱了对系统模型的依赖。将算法应用在连续状态系统上,经过在线学习可以得到令人满意的连续控制策略。

3. 针对当前在线RL算法具有运行时间不明确、数据利用率低、结果非最优的缺陷,首次提出一个完全不需要系统模型,可以在有限运行时间内通过在线学习得到连续状态系统近似最优策略的在线RL算法。使用状态离散化的方法对连续状态空间进行划分并存储在线采集的数据。定义迭代操作符和区分观测结果的方法。通过理论分析证明算法满足概率近似正确(Probably Approximately Correct, PAC)的原理。常见实例的仿真结果验证了算法的有效性。同时,把算法应用在汽车巡航控制的汽车动力学模型数据采集上,得到的结果与随机采集方法相比,该算法可以采集到状态范围更为广泛的数据。

4. 为了避免上面提出的算法对状态空间离散化的依赖,提出新的算法直接

使用采集的数据进行学习，提高数据的利用率。引入kd树技术对数据进行存储和管理，给出新的迭代操作符和区分观测结果的定义。经过理论分析证明新算法依然满足PAC原理，即在有限运行时间内算法可以在线学习到连续状态系统的近似最优策略。同时完全不需要系统模型。经过仿真实例的验证，与上面的算法相比新算法可以在更短的时间内学到相同性能的近似最优策略。因而具有更高的运行效率。

关键词： 最优控制，强化学习，连续状态系统，逼近策略迭代，收敛性，在线学习，概率近似最优，kd树

Abstract

This paper is a thesis template of Chinese Academy of Sciences. Besides that the usage of the \LaTeX document class `CASthesis`, a brief guideline for writing the thesis is also included.

Keywords: Chinese Academy of Sciences (CAS), Thesis, \LaTeX Template

目 录

摘要	i
Abstract	iii
目录	v
第一章 引言	1
1.1 系统要求	1
1.2 下载与安装	1
1.3 宏包定制	2
1.4 问题反馈	2
第二章 数学公式	3
第三章 表格图形	5
3.1 表格	5
3.2 图形	5
3.2.1 浮动图形	5
第四章 常见问题	9
4.1 表格	9
4.2 脚注	9
附录 A 中国科学院研究生院学位论文撰写要求	11
A.1 基本要求	11
A.1.1 硕士学位论文	11
A.1.2 博士学位论文	11

A.2 学位论文的组成部分和排列顺序	12
A.2.1 封面	12
A.2.2 论文摘要	13
A.2.3 论文目录	13
A.2.4 正文	13
A.2.5 参考文献	14
A.2.6 发表文章目录	14
A.2.7 致谢	14
A.3 学位论文的书写、装订要求	15
发表文章目录	17
简历	19
致谢	21

表 格

插图

3.1	中科院数学与系统科学研究院院徽（在页面中间）	5
3.2	中科院数学与系统科学研究院院徽（在页面上方）	6
3.3	中科院数学与系统科学研究院院徽（在页面下方）	6
3.4	中科院数学与系统科学研究院院徽（在独立页面中）	8

第一章 引言

虽然中国科学院研究生院制定了学位论文撰写要求（参见附录 A），但目前中科院学位论文的排版仍然不是很规范。其中的一个原因是多样化的排版工具，有 Word 的也有 L^AT_EX 的。即使都是 L^AT_EX 排版的，但由于没有统一的模板，每个学位论文的排版结果都不一样。CASthesis 宏包是以宏包作者的博士学位论文为基础模板，根据中国科学院研究生院学位论文撰写要求编写的。宏包的另一个目的是简化学位论文的撰写，使得论文作者可以将精力集中到论文的内容上而不是浪费在版面设置上。同时宏包在符合学位论文撰写要求的基础上尽可能地进行了美化，其中还参考了出版界的一些排版规范。

1.1 系统要求

CASthesis 宏包可以在目前大多数的 T_EX 系统中使用，例如 C_T_EX、MiK_T_EX、t_eT_EX、f_pT_EX。

CASthesis 宏包通过 ctex 宏包来获得中文支持。ctex 宏包提供了一个统一的中文 L^AT_EX 文档框架，底层支持 CCT 和 CJK 两种中文 L^AT_EX 系统。最新的 ctex 宏包可以从 <http://www.ctex.org> 网站下载。

此外，CASthesis 宏包还使用了宏包 amsmath、amsthm、amsfonts、amssymb、bm 和 hyperref。目前大多数的 T_EX 系统中都包含有这些宏包。

最新的 C_T_EX 套装（2.4.1 以上版本）中包含了以上列出的各种宏包，用户无需额外的设置即可使用。

1.2 下载与安装

CASthesis 宏包的最新版本可以从 <http://www.ctex.org> 网站下载。C_T_EX 套装每次更新时都将会包含最新版本的 CASthesis 宏包。

CASthesis 宏包包含两个文件：CASthesis.cls 和 CASthesis.cfg。简单方便的安装方法是将宏包文件和学位论文 .tex 文件放在同一目录下。或者将宏包文件放置到 T_EX 系统的 localtexmf/tex/latex/casthesis 目录下，然后刷新 T_EX 系统的文件名数据库。

同时，宏包还提供了一个使用模板，也就是这份说明文档的源文件。用户可以通过修改这个模板来编写自己的学位论文。

用户也可以下载宏包源文件 `CASthesis.dtx` 和 `CASthesis.ins`，然后对 `CASthesis.ins` 文件运行 `latex` 编译命令来得到宏包文件。

关于安装过程的问题可以参考 `CTEX-FAQ` 以及其他 `LATEX` 教材。

1.3 宏包定制

`CASthesis` 宏包的设置都保存在 `CASthesis.cfg` 文件中。用户可以在 `.tex` 中通过宏包提供的命令修改设置。对于常用的设置修改，如培养单位名称、专业名称等，可以直接在 `CASthesis.cfg` 文件中进行。各培养单位可以修改后提供本单位统一的 `CASthesis.cfg` 文件供本单位用户使用。

1.4 问题反馈

用户在使用中遇到问题或者需要增加某种功能，都可以和作者联系：

吴凌云(aloft) aloft@ctex.org

欢迎大家反馈自己的使用情况，使我们可以不断改进宏包。

第二章 数学公式

[?]

第三章 表格图形

3.1 表格

3.2 图形

3.2.1 浮动图形

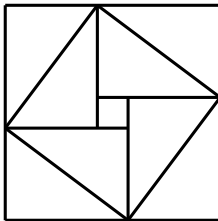
[illegible]

图 3.1: 中科院数学与系统科学研究院院徽 (在页面中间)

[illegible][illegible]

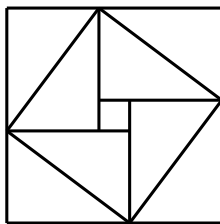
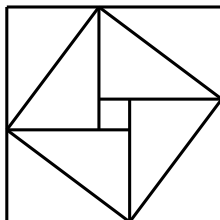
[illegible][illegible][illegible]

图 3.3: 中科院数学与系统科学研究院院徽 (在页面下方)

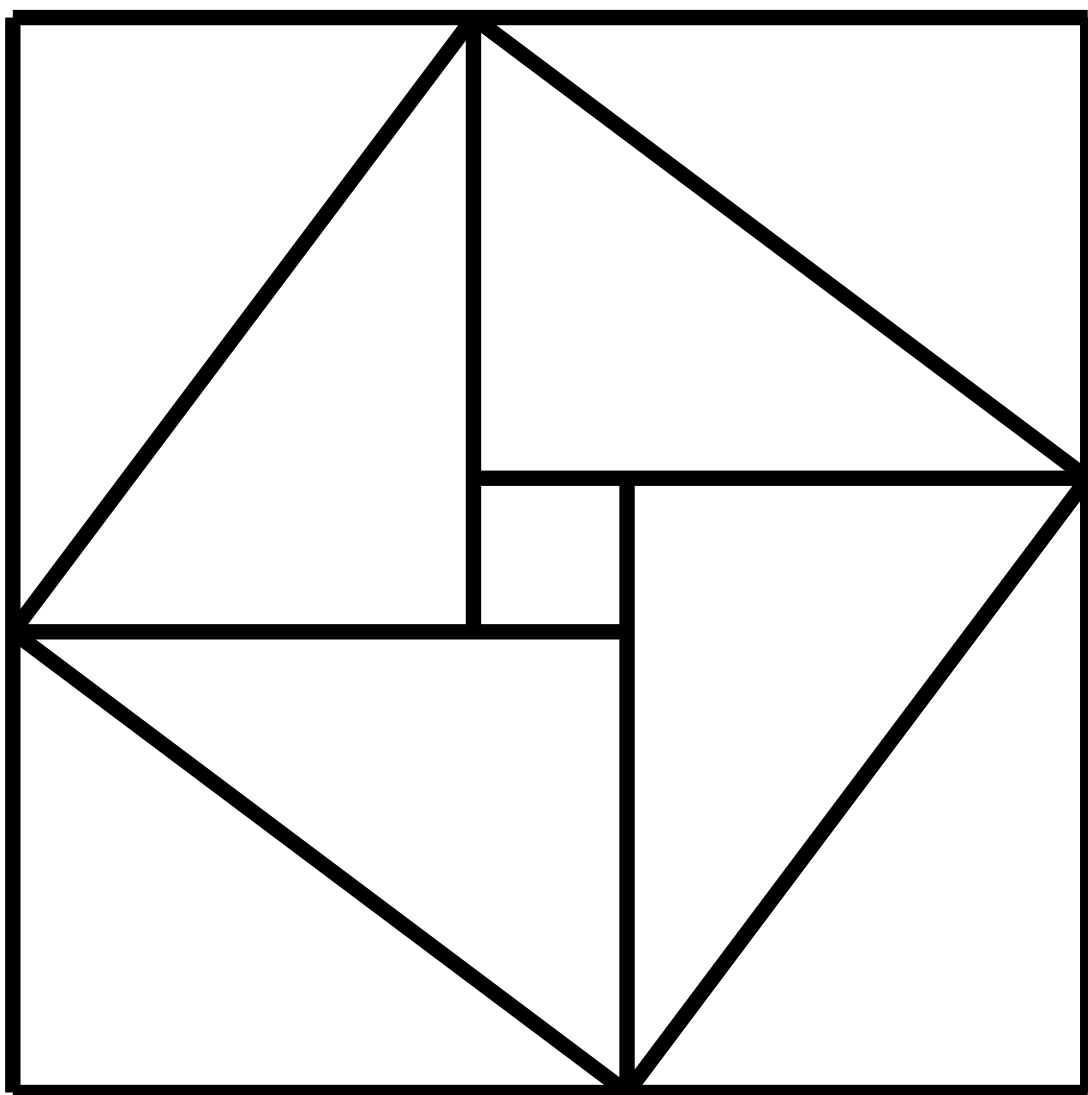


图 3.4: 中科院数学与系统科学研究院院徽（在独立页面中）

第四章 常见问题

4.1 表格

问题 1. 页眉里论文题目和各章标题中的字母均为大写，不能实现大小写的区别，而我写的论文需要在页眉中出现的标题中区分英文字母的大小写比如：YBaCuO而不是YBACUO。

回答 1. 在CASthesis.cfg 文件中加上

```
\renewcommand\title[2][\CAST@value@title]{%
  \def\CAST@value@title{#2}
  \def\CAST@value@titlemark{#1}}
\def\chaptermark#1{\markboth {{\ifnum \c@secnumdepth>\m@ne
  \if@mainmatter\CTEXthechapter \quad\fi
  \fi #1}}{}}%
\def\sectionmark#1{\markright{{\ifnum \c@secnumdepth >\z@
  \CTEXthesection \quad \fi #1}}}
```

4.2 脚注

问题 2. 如果在章节标题中加入注脚，则不仅会出现在本章首页的页脚，也会出现在目录的页脚，不知是否能够让其不要出现在目录的页脚中。

回答 2. 可以使用如下的命令来定义章节的标题

```
\chapter[出现在目录和页眉的标题]{出现在正文的标题\footnote{这个不会
出现在目录中。}}
```

section、 subsection 等命令也有类似的用法。

附录 A 中国科学院研究生院学位论文撰写要求

学位论文是为申请学位而撰写的学术论文，是评判学位申请者学术水平的主要依据，也是学位申请者获得学位的必要条件之一。为规范和统一我院研究生学位论文的写作，根据《中华人民共和国学位条例暂行实施办法》的有关规定，提出以下要求：

A.1 基本要求

学位论文必须是一篇（或由一组论文组成的一篇）系统的、完整的学术论文。学位论文应是学位申请者本人在导师的指导下独立完成的研究成果，不得抄袭和剽窃他人成果。学位论文的学术观点必须明确，且逻辑严谨，文字通畅。

A.1.1 硕士学位论文

硕士学位论文要注意在基础学科或应用学科中选择有价值的课题，对所研究的课题有新的见解，并能表明作者在本门学科上掌握了坚实的基础理论和系统的专门知识，具有从事科学研究工作或独立担负专门技术工作的能力。

硕士学位论文工作一般在硕士生完成培养计划所规定的课程学习后开始，应包括文献阅读、开题报告、拟定并实施工作计划、科研调查、实验研究、理论分析和文字总结等工作环节。硕士学位论文必须有一定的工作量。在论文题目确定后，用于论文工作的时间一般不得少于一年半。

A.1.2 博士学位论文

博士学位论文要选择在国际上属于学科前沿的课题或对国家经济建设和社会发展有重要意义的课题，要突出论文在科学和专门技术上的创新性和先进性，并能表明作者在本门学科上掌握了坚实宽广的基础理论和系统深入的专门知识，具有独立从事科学研究工作的能力。博士学位论文工作是培养博士学位研究生最重要的环节，其工作时间一般不应少于两年。博士研究生入学后，要在导师指导下确定科研方向，收集资料，阅读文献，进行调查研究，选择研究

课题。一般在第二学期，最迟在第三学期通过开题报告并制定论文工作计划，之后根据论文工作计划分阶段报告科研和论文工作进展情况。

A.2 学位论文的组成部分和排列顺序

学位论文一般由以下几个部分组成：封面、论文摘要、论文目录、正文、参考文献、发表文章目录、致谢等。

A.2.1 封面

根据原国家标准局《科学技术报告、学位论文和学术论文的编写格式》（国家标准GB7713-87）的封面要求，特规定中国科学院研究生院研究生学位论文的封面格式（见样张1和样张2），并提出以下具体要求：

A.2.1.1 分类号

必须在封面左上角注明分类号。一般应注明《中国图书资料分类法》的类号，同时注明《国际十进分类法UDC》的类号。

A.2.1.2 编号

各培养单位自定。

A.2.1.3 密级

论文必须按国家规定的保密条例在右上角注明密级（如系公开型论文则可不注明密级）。

A.2.1.4 论文题目

学位论文题目应当简明扼要地概括和反映出论文的核心内容，一般不宜超过20个字，必要时可加副标题。

A.2.1.5 指导教师

指导教师必须是被批准上岗的指导教师。

A.2.1.6 申请学位级别

填硕士学位或博士学位。

A.2.1.7 学科、专业名称

按国家颁布的学科、专业目录中的名称填写。

A.2.1.8 论文提交日期和论文答辩日期

按实际提交和答辩日期填写。

A.2.1.9 培养单位

填写培养单位全称。

A.2.1.10 学位授予单位

填写“中国科学院研究生院”。

A.2.2 论文摘要

论文摘要应概括地反映出本论文的主要内容，主要说明本论文的研究目的、内容、方法、成果和结论。要突出本论文的创造性成果或新见解，不要与引言相混淆。中文摘要力求语言精炼准确，字数在500字左右。英文摘要内容要与中文摘要内容一致。并在英文题目下面第一行写研究生姓名。专业名称用括号括起后，置于姓名之后。研究生姓名下面的一行写导师姓名，格式为：Directed by.....。无论中英文摘要都必须在摘要页的最下方另起一行，注明本文的关键词（3~5个）。

A.2.3 论文目录

论文目录是论文的提纲，也是论文各章节组成部分的小标题。

A.2.4 正文

正文是学位论文的主体和核心部分，不同学科专业和不同的选题可以有不同的写作方式。正文一般包括以下几个方面：

A.2.4.1 引言

引言是学位论文主体部分的开端，要求言简意赅，不要与摘要雷同或成为摘要的注解。除了说明研究目的、方法、结果等外，还应评述国内外研究现状和相关领域中已有的研究成果；介绍本项研究工作前提和任务，理论依据和实验基础，涉及范围和预期结果以及该论文在已有的基础上所解决的问题。

A.2.4.2 各具体章节

A.2.4.3 结论

结论是学位论文最终和总体的结论，是整篇论文的归宿。应精炼、准确、完整。着重阐述作者研究的创造性成果及其在本研究领域中的意义，还可进一步提出需要讨论的问题和建议。

A.2.5 参考文献

学位论文的撰写应本着严谨求实的科学态度，凡有引用他人成果之处，均应按论文中所引用的顺序列于文末。参考文献的著录均应符合国家有关标准（按照GB7714-87《文后参考文献著录格式》执行）。

1. 文献是期刊时，书写格式为：序号 作者. 文章题目. 期刊名, 年份(期数):起止页码
2. 文献是图书时，书写格式为：序号 作者. 书名. 版次. 出版地：出版单位, 年份. 起止页码

A.2.6 发表文章目录

指学位申请者在学期间在各类正式刊物上发表或已被接受的学术论文。

A.2.7 致谢

表达作者对完成论文和学业提供帮助的老师、同学、领导、同事及亲属的感激之情。

A.3 学位论文的书写、装订要求

(一) 中国科学院研究生院研究生学位论文必须用中文书写

1. 论文“题目”：黑体小三号
2. 论文“章”：黑体四号
3. 论文“节”：黑体小四号
4. 正文：宋体小四号

5. 为美观方便起见，要有页眉，奇数页上注明每一章名称，偶数页上注明论文题目。

为了便于国际合作与交流，学位论文亦可有英文或其它文字的副本。

(二) 文中的图表、附注、参考文献、公式一律采用阿拉伯数字连续（或分章）编号。如图1，表1，附注：1，文献（1），公式（1）。图序及图名置于图的下方；表序及表名置于表的上方；论文中的公式编号用括弧括起来写在右边行末，其间不加虚线。

(三) 文中所用单位一律采用国务院发布的《中华人民共和国法定计量单位》，单位名称和符号的书写方式，应采用国际通用符号。

(四) 学位论文封面采用全院统一格式，封面用纸为150克花纹纸，博士学位论文封面颜色为红色，硕士学位论文封面颜色为蓝色（见样张博士学位论文封面、硕士学位论文封面）。

(五) 学位论文一律用A4打印纸装订。

发表文章目录

- [1] Ji-Hong Zhang, Ling-Yun Wu and Xiang-Sun Zhang. Reconstruction of DNA sequencing by hybridization. Bioinformatics (SCI), vol 19(1), pages 14-21, 2003.
- [2] 吴凌云. C_TE_X FAQ (常见问题集). 2004.

简 历

基本情况

吴凌云，男，福建省屏南县人，1975 年 11 月出生，未婚，中国科学院数学与系统科学研究院在读博士研究生。

教育状况

1993 年 9 月至 1997 年 7 月，武汉大学数学系，本科，专业：应用数学。

1997 年 9 月至 2002 年 7 月，中国科学院数学与系统科学研究院，硕博连读研究生，专业：运筹学与控制论。

工作经历

无。

研究兴趣

运筹学及其应用，人工神经网络，组合优化，生物信息学。

联系方式

通讯地址：北京市 2734 信箱，中科院数学与系统科学研究院应用数学所

邮编：100080

E-mail: aloft@ctex.org

致 谢

值此论文完成之际，谨在此向多年来给予我关心和帮助的老师、同学、朋友和家人表示衷心的感谢！

.....

谨把本文献给我最敬爱的父亲！