中图分类号:学校代码:10055UDC:密级:公开

有阁大 字 硕 士 学 位 论 文

一种基于近边界数据的模型所有权推断方法研究 Research on Model Ownership Inference Based on Near-boundary Data

论文作者 杨宗稳	指导教帅 浦凌君副教授
申请学位工学硕士	培养单位 南开大学
学科专业 计算机科学与技术	研究方向 模型的知识产权保护
答辩委员会主席	评 阅 人

南开大学研究生院

二〇二三年四月

南开大学学位论文使用授权书

本人完全了解《南开大学关于研究生学位论文收藏和利用管理办法》关于南开大学(简称"学校")研究生学位论文收藏和利用的管理规定,同意向南开大学提交本人的学位论文电子版及相应的纸质本。

本人了解南开大学拥有在《中华人民共和国著作权法》规定范围内的学位论文使用权, 同意在以下几方面向学校授权。即:

- 1. 学校将学位论文编入《南开大学博硕士学位论文全文数据库》,并作为资料在学校图 书馆等场所提供阅览,在校园网上提供论文目录检索、文摘及前 16 页的浏览等信息服务;
- 2. 学校可以采用影印、缩印或其他复制手段保存学位论文; 学校根据规定向教育部指 定的收藏和存档单位提交学位论文;
 - 3. 非公开学位论文在解密后的使用权同公开论文。

本人承诺:本人的学位论文是在南开大学学习期间创作完成的作品,并已通过论文答辩;提交的学位论文电子版与纸质本论文的内容一致,如因不同造成不良后果由本人自负。 本人签署本授权书一份(此授权书为论文中一页),交图书馆留存。

学位论文作者暨授权人(亲笔)	签字:_			
	20	年	月	日

南开大学研究生学位论文作者信息

论文题目	一种基于近边界数据的模型所有权推断方法研究							
姓 名	杨宗稳	学号	学号 2120200439			答辩日期		
论文类别	博士 🗆	学历硕	<u>i</u> ± v	专业学	位在	页士 口	同等	等学力硕士 🗆 划团选择
学院 (単位)	计算机学院			学科/专业(专业学位)名称 计算机科学与			计算机科学与技术	
联系电话	13102257615			电子邮箱		yzv	v@ma	nil.nankai.edu.cn
通讯地址 (邮编): 300000								
非公开论文编				备注	=			

注:本授权书适用我校授予的所有博士、硕士的学位论文。如己批准为非公开学位论文,须向图书馆提供批准通过的《南开大学研究生申请非公开学位论文审批表》复印件和"非公开学位论文标注说明"页原件。

南开大学学位论文原创性声明

本人郑重声明: 所呈交的学位论文,是本人在导师指导下进行研究工作所取得的研究成果。除文中已经注明引用的内容外,本学位论文的研究成果不包含任何他人创作的、已公开发表或者没有公开发表的作品的内容。对本论文所涉及的研究工作做出贡献的其他个人和集体,均已在文中以明确方式标明。本学位论文原创性声明的法律责任由本人承担。

学位论文作者签名:	年	月	日

非公开学位论文标注说明

(本页表中填写内容须打印)

根据南开大学有关规定,非公开学位论文须经指导教师同意、作者本人申请和相关部门批准方能标注。未经批准的均为公开学位论文,公开学位论文本说明为空白。

论文题目								
申请密级	□限制((≤2年)		〕秘密 (≤10 억	F) [] 机密 (<	€20年)	
保密期限	20	年	月	月 日至 20		月	日	
审批表编号				批准日期	20	年	月	日

南开大学学位评定委员会办公室盖章 (有效)

注: 限制 ★2 年 (可少于 2 年); 秘密 ★10 年 (可少于 10 年); 机密 ★20 年 (可少于 20 年)

摘要

这里输入中文摘要。 **关键词:** 毕业论文;模板

Abstract

This is the abstract.

Key Words: Thesis; template

目录

摘要		I
Abstract		II
第一章	绪论	1
第一节	研究背景与意义	1
第二节	相关研究现状	1
第三节	本文主要工作	1
第四节	本文主要架构	1
第二章	技术背景	2
第一节	深度神经网络	2
第二节	对抗性攻击	2
第三节	对抗生成网络	2
第四节	深度神经网络模型窃取攻击	3
第五节	深度神经网络模型的知识产权保护	3
第六节	本章小结	3
第三章	基于对抗生成网络特征提取的近边界数据研究	4
第一节	近边界对抗性样本	4
第二节	CW 生成近边界对抗性样本	6
第三节	近边界数据私有化	6
第四节	本章小结	6
第四章	基于近边界数据的模型所有权推断方法研究	7
第一节	数据集推断	7
第二节	近边界数据推断模型所有权	7
第三节	本章小结	7
第五章	基于近边界数据的模型所有权推断方法分析	8
第一节	实验设置	8
5.1.1	什么是 TeX/LATeX, 我是否应该选择它	8

目录

5.1.2	我该用什么编辑器?	9
5.1.3	我该去哪里寻找答案?	9
5.1.4	我应该看什么 IATEX 读物?	9
5.1.5	什么知识会过时?什么不会? 1	0
5.1.6	插图格式 1	2
5.1.7	LATEX 作图 1	2
第二节	生成初始近边界数据的算法选择	2
第三节	数据近边界特性的评估与扩展 1	2
第四节	推断模型所有权 1	2
第五节	微调目标分类边界的影响1	2
第六节	可伸缩性扩展 1	2
第七节	本章小结 1	2
第六章	总结与展望1	3
第一节	工作总结 1	3
第二节	工作展望 1	3
参考文献	t ······· 1	4
致谢		5
个人简用	.	6

第一章 绪论

本模板参照南开大学学位论文写作规范编写,仅仅提供了论文的基本格式,包括章节标题和正文字体、字号等等的设置。

您自愿使用这个模板。提供本模板的目的是为了给您的论文写作带来方便, 然而,作者不保证这个模板完全符合学校的要求,也不对由此产生的任何后果 负责。如果您不同意这些条款,请不要使用这个模板。

第一节 研究背景与意义

- 参考文献的录入请参考??;
- 图片插入参考第六节;
- 分数和公式参考??;
- Latex 绘图工具参考第一节:
- 代码块参考第二节;

第二节 相关研究现状 第三节 本文主要工作 第四节 本文主要架构

第二章 技术背景

第一节 深度神经网络

模板仅在 TeXLive 2016, TeXLive 2018 下测试通过。对于其它 TeX 发行版可能需要做个别修改。

第二节 对抗性攻击

本模板可以使用以下方式编译:

1. X元ATEX[推荐]

例如,

xelatex main

biber main % 处理参考文献

xelatex main % 连续编译两遍以生成正确的文献引用。

本模板用到宋体、楷体、仿宋、黑体四种字体. 若需重新配置字体, 请修改 NKTfonts.cfg. 对于 Linux/Mac 下的 TeX Live 2009, 可能需要设置环境变量 OSFONTDIR, 具体内容请参考 texmf.cnf.

我们建议您使用 X-LAT_EX 编译。与前两种方法相比,X-LAT_EX 编译长文档的速度更快,编译一篇一百多页的论文只需几秒的时间 (SL9400 @ 1.86GHz)。

在改变编译方式前应先删除 *.toc 和 *.aux 文件,因为不同编译方式产生的辅助文件格式可能并不相同。

注意: 使用 X_HT_EX 编译时, X_HT_EX 的版本应不低于 0.9995.0(MiKTeX 2.8 或者 TeXLive 2009)。

第三节 对抗生成网络

引用章节号请参考如下格式: 第二章第三节.

有周大學

图 2.1 南开大学

第四节 深度神经网络模型窃取攻击

使用 X_TLAT_EX 编译时,会自动在中英文转换时添加必要的空格。使用 [PDF]LAT_EX 编译时仅忽略中文之间的空格,而中英文之间的空格予以保留。因此,不管何种编译方式,您都不需要在中英文间添加~以获得额外的空格。例如,

这是 English 中文 x = y 测试

这是 English 中文 x = y 测试

可以看出,以上两行用 X-TAT-X 编译的结果是相同的。

第五节 深度神经网络模型的知识产权保护

NKThesis 已经调用以下宏包,您无须重新调用。

表 2.1 NKThesis 预调用的宏包

编译方式	调用的宏包
Xaralex	xeCJK, CJKnumb, graphicx, mathptmx, amssymb,mathpazo,pgf,tikz

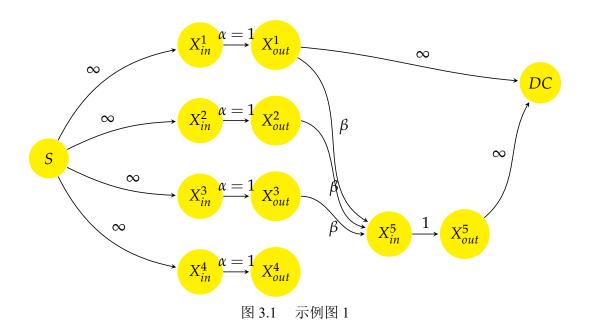
第六节 本章小结

插图的例子:

第三章 基于对抗生成网络特征提取的近边界数据研究

第一节 近边界对抗性样本

The PDF package, where "PDF" is supposed to mean "portable graphics format" (or "pretty, good, functional" if you prefer...), is a package for creating graphics in an "inline" manner. It defines a number of TEX commands that draw graphics. For example, the code \tikz \draw (0pt,0pt) -- (20pt,6pt); yields the line and the code \tikz \fill[orange] (1ex,1ex) circle (1ex); yields .



In a sense, when you use PDF you "program" your graphics, just as you "program" your document when you use TEX. You get all the advantages of the "TEX-approach to typesetting" for your graphics: quick creation of simple graphics, precise positioning, the use of macros, often superior typography. You also inherit all the disadvantages: steep learning curve, no WYSIWYG, small changes require a long recompilation time, and the code does not really "show" how things will look like.

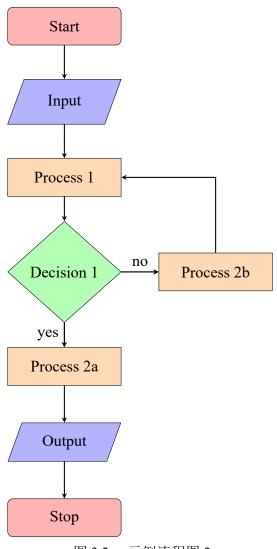


图 3.2 示例流程图 2

第二节 CW 生成近边界对抗性样本

python 代码可以直接使用 python 环境

```
1 def fibonacci(n):
2  # Fibonacci number
3  if n < 0:
4    return False
5  if n <= 1:
6    return n
7  return fibonacci(n-2) + fibonacci(n-1)</pre>
```

代码 3.1 斐波那契 Python

C/C++ 代码可以直接使用 cpp 环境

```
1 unsigned long Fibonacci(int n)
2 {
3
      // Fibonacci start from 0
      if (n <= 1)
5
      {
6
          return n;
      }
8
      else
9
      {
10
         return Fibonacci(n - 1) + Fibonacci(n - 2);
11
12 }
```

代码 3.2 斐波那契 C++

其他代码,使用 lstlisting 指明 language 即可,如 matlab 代码

```
1 function a = factorial(n)
2 % return n!
3    if n==0
4        a=1;
5    else
6        a=n * factorial(n-1);
7    end
```

代码 3.3 Matlab 代码

第三节 近边界数据私有化 第四节 本章小结

第四章 基于近边界数据的模型所有权推断方法研究

第一节 数据集推断 第二节 近边界数据推断模型所有权 第三节 本章小结

第五章 基于近边界数据的模型所有权推断方法分析

第一节 实验设置

以下内容是 milksea@bbs.ctex.org 撰写的关于 T_EX 的简单介绍。注意这不是一个入门教程,不讲 T_EX 系统的配置安装,也不讲具体的 LAT_EX 代码。这里仅仅试图以一些只言片语来解释:进入这个门槛之前新手应该知道的注意事项,以及遇到问题以后该去如何解决问题。

5.1.1 什么是 TFX/IATFX, 我是否应该选择它

T_EX 是最早由高德纳(Donald Knuth)教授创建的一门标记式宏语言,用来排版科技文章,尤其擅长处理复杂的数学公式。T_EX 同时也是处理这一语言的排版软件。L^AT_EX 是 Leslie Lamport 在 T_EX 基础上按内容/格式分离和模块化等思想建立的一集 T_EX 上的格式。

 T_{EX} 本身的领域是专业排版(即方正书版、InDesign 的领域),但现在 T_{EX}/LaT_{EX} 也被广泛用于生成电子文档甚至幻灯片等, T_{EX} 语言的数学部分偶尔也在其他一些地方使用。但注意 T_{EX} 并不适用于文书处理(MS Office 的领域,以前和现在都不是)。

选择使用 TFX/LATFX 的理由包括:

- 免费软件;
- 专业的排版效果;
- 是事实上的专业数学排版标准;
- 广泛的西文期刊接收甚或只接收 LaTeX 格式的投稿;

.

不选择使用 TFX/IATFX 的理由包括:

- 需要相当精力学习;
- 图文混合排版能力弱;
- 仅流行于数学、物理、计算机等领域:
- 中文期刊的支持较差;

• • • • • •

请尽量清醒看待网上经常见到的关于 T_EX 与其他软件的优劣比较和口水战。 在选择使用或离开之前,请先考虑 T_EX 的应用领域,想想它是否适合你的需要。

5.1.2 我该用什么编辑器?

编辑器功能有简有繁,特色不一,从简单的纯文本编辑器到繁复的 Emacs,因人而易。基本功能有语法高亮、方便编译预览就很好了,扩充功能和定制有无限的可能。初学者可以使用功能简单、使用方便的专用编辑器,如 TeXWorks、Kile、WinEdt 等,或者类似所见即所得功能的 LyX; 熟悉的人可以使用定制性更强的 Notepad++、SciTE、Vim、Emacs 等。这方面的介绍很多,一开始不妨多试几种,找到最适合自己的才是最好的。

另外提醒一句,编辑器只是工作的助力,不必把它看得太重。一些编辑器有极为繁杂的功能,一些编辑器常常会引来黑客们的论战(如 Emacs 与 Vim)。为工作,别为这些浪费太多精力,适用即可。

5.1.3 我该去哪里寻找答案?

- 0、绝对的新手,先读完一本入门读物,了解基本的知识。
- 1、无论如何, 先读文档! 绝大部分问题都是文档可以解决的。
- 2、再利用 Google 搜索,利用 (bbs.ctex.org) 版面搜索。
- 3、清楚、聪明地提出你的问题。

5.1.4 我应该看什么 IAT_EX 读物?

这不是一个容易回答的问题,因为有许多选择,也同样有许多不合适的选择。这里只是选出一个比较好的答案。更多更详细的介绍可以在版面和网上寻找(注意时效)。

近两年 T_EX 的中文处理发展很快,目前没有哪本书在中文处理方面给出一个最新进展的合适综述,因而下面的介绍也不主要考虑中文处理。

- 1. 我可以阅读英文
 - (a) 我要迅速入门: ltxprimer.pdf (LaTeX Tutorials: A Primer, India TUG)
 - (b) 我要系统学习: A Guide to LaTeX, 4th Edition, Addison-Wesley 有机械工业出版社的影印版(《LaTeX 实用教程》)
 - (c) 我要深入学习:要读许多书和文档,TeXbook 是必读的

- (d) 还有呢? 去读你使用的每一个宏包的说明文档
- (e) 还有许多专题文档,如讲数学公式、图形、表格、字体等

2. 我更愿意阅读中文

- (a) 我要迅速入门: Inotes.pdf (LaTeX Notes, 1.20, Alpha Huang)
- (b) 我要系统学习:《LaTeX2ε 科技排版指南》,邓建松(电子版)如果不 好找,看《LaTeX 入门与提高》第二版,陈志杰等
- (c) 我要深入学习: TeXbook 0.pdf (特可爱原本, TeXbook 的中译, xianxian)
- (d) 还有呢? 英语,绝大多数 TeX 资料还是英文的

5.1.5 什么知识会过时?什么不会?

TeX 是排版语言,也是广泛使用的软件,并且不断在发展中;因此,总有一些东西会很快过时。作为学习 TeX 的人,免不了要看各种各样的书籍、电子文档和网络论坛上的只言片语,因此了解什么知识会迅速过时,什么知识不会是十分重要的。

最稳定的是关于 Primitive T_EX 和 Plain T_EX 的知识,也就是 Knuth 在他的《The TeXbook》中介绍的内容。因为 T_EX 系统开发的初衷就是稳定性,要求今天的文档到很久以后仍可以得到完全相同的结果,因此 Knuth 限定了他的 T_EX 语言和相关实现的命令、语法。这些内容许多年来就没有多少变化,在未来的一些年里也不会有什么变化。Primitive T_EX 和 Plain T_EX 的知识主要包括 T_EX 排版的基本算法和原理,盒子的原理,底层的 T_EX 命令等。其中技巧性的东西大多在宏包设计中,初学者一般不会接触到很多;而基本原理则是常常被提到的,譬如,T_EX 把一切排版内容作为盒子(box)处理。

相对稳定的是关于基本 LATeX 2_{ε} 的知识,也包括围绕 LATeX 2_{ε} 的一些核心宏包的知识。LATeX 2_{ε} 是自 1993 年以来的一个稳定的 LATeX 版本,直到最近的一次修订(2005 年)都没有大的变动。LATeX 的下一个计划中的版本 LATeX 3 遥遥无期,在可预见的将来,LATeX 2_{ε} 不会过时。LATeX 2_{ε} 的知识是目前大部分 LATeX 书籍的主体内容。关于 LATeX 的标准文档类(article、report、book、letter、slide等),关于基本数学公式的输入,文档的章节层次,表格和矩阵,图表浮动体,LR 盒子与段落盒子……这些 LATeX 的核心内容都是最常用的,相对稳定的。与LATeX 2_{ε} 相匹配的核心宏包,如 graphics(x)、ifthen、fontenc、doc 等,也同样是相对稳定的。还有一些被非常广泛应用的宏包,如 amsmath 系列,也可以看作

是相对稳定的。

简单地说,关于基本 T_EX/LAT_EX 的语言,都是比较稳定的。与之对应,实现或者支持 T_EX/LAT_EX 语言的软件,包括在 T_EX/LAT_EX 基础上建立的新的宏,都不大稳定。

容易过时的是关于第三方 LAT_EX 宏包的知识、第三方 T_EX 工具的知识,以及新兴 T_EX 相关软件的知识等。T_EX 和 LAT_EX 语言是追求稳定的;但无论是宏包还是工具,作为不断更新软件,它们是不稳定的。容易过时的技术很多,而且现在广泛地出现在几乎所有 LAT_EX 文档之中,因此需要特别引起注意:宏包的过时的原因可能是宏包本身的升级换代带来了新功能或不兼容,也可能是同一功能的更新更好的宏包代替了旧的宏包。前者的典型例子比如绘图宏包 PGF/TikZ,现在的 2.00 版功能十分强大,和旧的 1.1x 版相差很大,和更旧的 0.x 版本则几乎完全不同;后者的典型例子比如 caption 宏包先是被更新的 caption2 宏包代替,后来 caption 宏包更新又使得 caption2 宏包完全过时。——安装更新的发行版可以避免使用过旧的宏包;认真阅读宏包自带的文档而不是搜索得到的陈旧片断可以避免采用过时的代码。

工具过时的主要原因也是升级换代和被其他工具替换。前者的典型例子是编辑器 WinEdt 在 5.5 以后的版本支持 UTF-8 编码,而旧版本不支持;后者的典型例子是中文字体安装工具从 GBKFonts 到 xGBKFonts 到 FontsGen 不断被取代。图形插入是一个在 TEX 实现、宏包与外围工具方面都更新很快的东西。在过去,最常用的输出格式是 PS(PostScript)格式,因此插入的图像以 EPS 为主流。使用 Dvips 为主要输出工具,外围工具有 GhostScript、bmeps 等等,相关宏包有 graphics 等,相关文档如《LaTeX2e 插图指南》。

但凡提及"LATEX 只支持 EPS 图形"的,就是这个过时的时代的产物。事实上 TEX/LATEX 并不限定任何图形格式,只不过是当时的输出格式(PS)和工具(Dvips)对 EPS 情有独钟而已。后来 PDF 格式成为主流,pdfTeX、DVIPDFM、DVIPDFMx、XeTeX 等工具则主要支持 PDF、PNG、JPG 格式的图形,涉及一系列工具如 ImageMagick、ebb 等。

值得特别提出注意的就是,中文处理也一起是更新迅速、容易过时的部分。而且因为中文处理一直没有一个"官方"的"标准"做法,软件、工具、文档以及网上纷繁的笔记也就显得相当混乱。从八十年代开始的CCT系统、天元系统,到后来的CJK方式,到近来的X河XX、LuaTeX方式,中文处理的原理、软件、

宏包、配置方式等都在不断变化中。

5.1.6 插图格式

前面提到,LAT_EX 主要支持 EPS 格式的插图文件,而 PDFLAT_EX 则更喜欢 PDF、PNG、JPG 格式的图形。为解决兼容性,最新版的 PDFLAT_EX 会自动把 EPS 文件转换为 PDF 文件。因此,使用 EPS 格式的插图可能具有最广泛的通用性。

5.1.7 LAT_FX 作图

目前已经有很多优秀的 LATEX 作图宏包,如 pgf/Tikz 和 pstricks,两者都具有强大的作图能力。

第二节 生成初始近边界数据的算法选择 第三节 数据近边界特性的评估与扩展 第四节 推断模型所有权 第五节 微调目标分类边界的影响 第六节 可伸缩性扩展 第七节 本章小结

第六章 总结与展望

任何问题可在 GitHub 上发起 issue github.com/NewFuture/NKThesis/issues/

第一节 工作总结 第二节 工作展望

参考文献

- [0] 陈建军, 车建文, 陈勇. 具有频率和振型概率约束的工程结构动力优化设计. [J]. 计算力学学报, 2001, 18 (1): 74-80.
- [0] NADKARNI MA, NAIR CKK, PANDEY VN, et al. Characterization of alpha-galactosidase from corynebacterium murisepticum and mechanism of its induction. [J]. J Gen App Microbiol, 1992, 38: 223–234.
- [0] 华罗庚, 王元. 论一致分布与近似分析: 数论方法 (I). [J]. 中国科学, 1973 (4): 339-357.
- [0] 竺可桢. 物候学. [M]. 北京: 科学出版社, 1973.
- [0] 霍夫斯塔主编. 禽病学: 下册. 第 7 版. [M]. 胡祥壁, 译. 北京: 农业出版社, 1981: 798-799.
- [0] Nadkarni M A, Nair C K K, Pandey V N, et al. Characterization of alpha-galactosidase from corynebacterium murisepticum and mechanism of its induction. J Gen App Microbiol, 1992, 38: 223 ~ 234.
- [0] 姜锡洲. 一种温热外敷药制备方案: 88105607.3[P]. P (中国)1989-07-26.
- [0] TIMOSHENKO S P. Theory of plate and shells. 2nd ed. [M]. New York: McGraw-Hill, 1959: 17–36.
- [0] 张全福,王里青."百家争鸣"与理工科学报编辑工作.[G] // 郑福寿.学报编辑论丛:第2集.南京:河海大学出版社,1991:1-4.
- [0] 丁光莹. 钢筋混凝土框架结构非线性反应分析的随机模拟分析: [博士学位论文]. 上海: 同济大学, 2001.
- [0] 全国文献工作标准化技术委员会第六分委员会. GB 6447-86 文摘编写规则. 北京: 中国标准出版社,1986.

致谢

感谢您使用本模板。

个人简历

xxx,出生于yyyy年mm月dd日。在20yy年毕业于xx大学XX专业并获得xx士学位。于20xx年至今在南开大学就读xxx研究生。

研究生期间发表论文:

- 周恩来. 周恩来选集 [M]. 人民出版社, 1980.
- 周恩来. 周恩来外交文选 [M]. 中央文献出版社, 1990.