A Research on Patent Classification Using Industrial Heterogeneous Graph Network

张鼎

武汉大学计算机学院

Oct. 26 2021





研究背景

1 研究背景

研究背景

- 2 相关工作
- 3 研究方法
- 4 实验设计
- 5 研究计划



Outline

研究背景

•0

- 1 研究背景
- 2 相关工作

- 3 研究方法
- 4 实验设计
- 5 研究计划

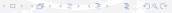




研究背景

研究背景

- 专利指标已经成为国家竞争力和科技创新能力的重要标志。
- 在通常情况下, 专利审查员会根据其具有的领域知识手工将 每个专利手动分为多个类别。
- 专利申请迅速增长,传统的手动操作费力费时,几乎无法满 足需求。
- 迫切需要自动专利分类工具来支持相关服务。
- 专利的自动分类对于提高大规模专利管理和服务的效率具有 重要意义。



Outline

- 1 研究背景
- 2 相关工作

- 3 研究方法
- 4 车验设计
- 5 研究计划



相关工作

目前关于专利自动分类的相关研究大致可以分为两个方面:

- 挖掘有效的特征
- 设计专用的分类器



相关工作

研究背景

目前关于专利自动分类的相关研究大致可以分为两个方面:

- 挖掘有效的特征
 - ◇ 专利文档中含有大量的元特征以及文本信息,通过从中挖掘 有效的特征对专利进行分类。
 - ◇ 文档分割及语法分析
 - ◇ 关键词提取
 - ◇ 文本聚类
- 设计专用的分类器
 - ♦ SVM, KNN, CNN, GRU, BERT 等方法被引入来解决专利 自动分类问题。
 - ♦ DeepPatent 建立了深度卷积神经网络模型,并结合词嵌入对 专利文件进行分类。
 - ♦ PatentBert 利用了功能强大的预训练语言模型 Bert, 然后对 其进行微调以处理多标签专利分类问题。



Outline

研究背景

- 1 研究背景
- 2 相关工作
- **3** 研究方法 当前挑战与问是

创新点 多视图 整体架构







研究背景 相关工作 研究方法 实验设计 研究计划 结语 ○○ ○○ ○●○○○○○○○○○ ○○ ○○○ ○○○○ ○○○○ ○○

当前挑战与问题

Outline

- 1 研究背景
- 2 相关工作
- 3 研究方法 当前挑战与问题

创新点 多视图 整体架构



5 研究计划



研究背景

当前排战与问题

- 现在使用最多的专利分类数据集为 USPTO 数据集、为英文 专利数据集。缺少一个大规模的中文专利分类数据集。
- 现有的专利分类器专注干从专利本身内容抽取特征,忽略了 专利发明人、公司之间的关联信息。
- 专利文本分类方法需要使用到专利全文数据, 涉及到大量的 专有名词,特征提取难度大。且当数据规模量大后,模型的 计算效率会降低, 计算资源、显存消耗量显著变大。

Outline

- 1 研究背景
- 2 相关工作
- **3** 研究方法 当前挑战与问是

创新点 多视图 整体架构

- 4 实验设计
- 5 研究计划





研究背景

创新点

- 以图的视角切入专利分类的问题,不需要用到专利全文数据进行分类。避免了原来专利分类在大数据集上分类速度降低、使用资源显著变大的问题。
- 提出了一种多视图融合的异质图神经网络,可以同时捕捉到局部(专利本身)和全局(公司、产业之间)的语义信息,利用注意力机制、将不同视图下的信息进行有效的聚合。

Outline

- 1 研究背景
- 2 相关工作
- 3 研究方法 当前挑战与问题

创新点 多视图 整体架构

4 实验设计

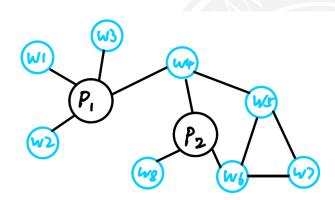
5 研究计划





背景 相关工作 **研究方法** 实验设计 研究计划 结语 ○○○ ○○○○**○○○○○○○○**○○○○ ○○○ ○○○ ○○○○ ○

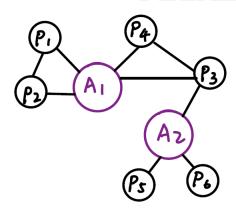
_{多视图} 多视图





多视图

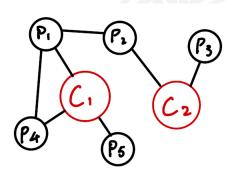
多视图





多视图

多视图

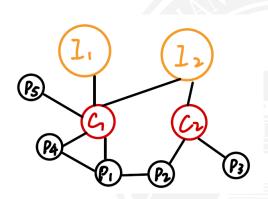




背景 相关工作 **研究方法** 实验设计 研究计划 结语 ○○○ ○○○○○**○○○○○○**○○○○ ○○○ ○○○ ○○○○ ○

多视图

多视图





整体架构

Outline

- 1 研究背景
- 2 相关工作
- **3** 研究方法 当前挑战与问题

创新点 多视图 整体架构

- 4 实验设计
- 5 研究计划



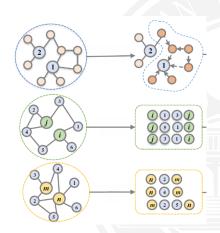


 相关工作
 研究方法
 实验设计
 研究计划
 结语

 ○○○
 ○○○○
 ○○○○
 ○○○○
 ○○○○

研究背景 ○○ 整体架构

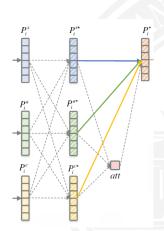
整体架构





整体架构

整体架构



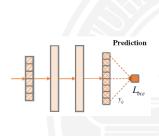




相关工作 研究方法 实验设计 研究计划 结语 ○○○ ○○○○○○○○○○○○○ ○○ ○○○○○ ○

整体架构

整体架构





Outline

研究背景

- 1 研究背景
- 2 相关工作

- 3 研究方法
- 4 实验设计
- B 研究计划





研究背景 相关工作 研究方法 **实验设计** 研究计划 结语 ○○ ○○ ○○ ○○○○○○○○○○○○○ ○●**○** ○○○○○○ ○

实验设计

Outline

- 1 研究背景
- 2 相关工作

- 3 研究方法
- 4 实验设计实验设计
- 5 研究计划



1 实验数据

研究背景

实验设计

- 130w+条专利数据与公司产业数据,构建产业异质图数据 集。
 - 异质图节点包含: 专利、发明人、公司、产业
- 2 评价指标
 - Precision@K
 - Recall@K
 - Precision Recall 曲线 AUC 值
- 3 对比模型
 - 基于文本分类: Bi-LSTM
 - 基于文本分类: DeepPatent
 - 基于文本分类: PatentBert
 - 基于图: Deepwalk
 - 基于图: Node2vec
 - 基于图: GCN
- 4 对比实验
 - 验证多视图的有效性



Oct. 26 2021

Outline

研究背景

- 1 研究背景
- 2 相关工作
- 3 研究方法

4 实验设计

5 研究计划

论文提纲

Outline

- 1 研究背景
- 2 相关工作
- 3 研究方法

4 实验设计

5 研究计划 论文提纲 时间安排



 相关工作
 研究方法
 实验设计

 000
 00000000000000
 000

论文提纲

研究背景

论文提纲

- 1 绪论
- 2 专利分类相关理论与技术
- 3 基于产业的图结构编码
- 4 基于产业异质图神经网络的专利分类
- 5 实验与结果分析
- 6 总结与展望



研究计划

研究背景

论文提纲

- 1 绪论
 - 1.1 研究背景与意义
 - 1.2 国内外研究现状
 - 1.3 研究内容与预期 目标
 - 1.4 论文结构安排
- 2 专利分类相关理 论与技术
 - 2.1 文本分类
 - 2.2 图网络

- 2.3 异质图网络
- 3 产业异质图构建 方法
 - 3.1 基于元路径的构 建方法
 - 3.2 基于元图的构建 方法
- 4 基于产业异质图 网络的专利分类
 - 4.1 单视图网络表征

- 4.2 多视图融合模块 4.3 专利分类
- 5 实验与结果分析
- 5.1 数据与预处理
 - 5.2 实验设置
 - 5.3 消融实验
 - 5.4 结果分析
- 6 总结与展望
 - 6.1 全文总结
 - 6.2 未来工作展望

研究背景 相关工作 研究方法 实验设计 **研究计划** 结语 ○○ ○○ ○○○○○○○○○○○○○ ○○○ ○○○●○ ○○○

时间安排

Outline

- 1 研究背景
- 2 相关工作
- 3 研究方法

- 4 实验设计
- 5 研究计划 论文提纲 时间安排





研究背景 ○○ 时间安排

时间安排

时间安排

时间	安排
2021年10月2021年11月-2022年01月	论文选题,查阅相关文献资料,撰写开题报告。 设计模型细节,数据预处理。
2022年01月-2022年02月2022年04月202279年047022704702270270270270270202022702702022702702270270	编写代码,实现模型。 验证模型的性能,做对比实验,完善相应算法。
2022 年 04 月-2022 年 05 月 2022 年 05 月	根据实验结果撰写论文初稿。 论文修改、定稿,参加答辩。



研究计划 ○○○○○●

感谢聆听 敬请各位老师批评指导

答辩人: 张鼎



研究背景