



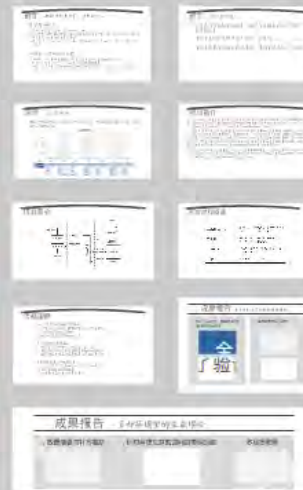
新知识发现与分类系统

北京邮电大学 大数据与智能信息处理实验室

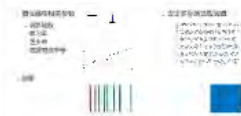
2018年11月

目录

- 前言
- 项目概况及考核指标
- 开发过程概述
- 成果报告
- 项目考核完成情况



成果报告



项目考核完成情况

考核项目	考核完成度
项目总体完成情况	100%
项目经费使用情况	100%
项目成果产出情况	100%
项目考核综合评价	100%

谢谢各位专家的聆听!



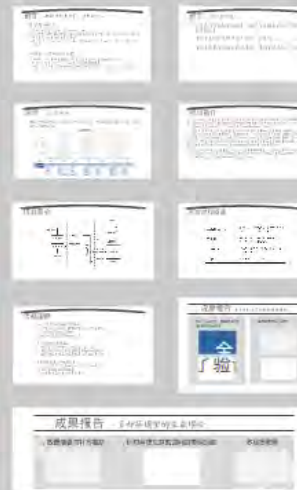
新知识发现与分类系统

北京邮电大学 大数据与智能信息处理实验室

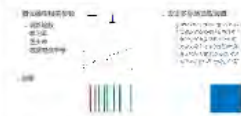
2018年11月

目录

- 前言
- 项目概况及考核指标
- 开发过程概述
- 成果报告
- 项目考核完成情况



成果报告



项目考核完成情况

项目考核项目	考核完成时间
项目总体考核指标	100%
项目考核指标完成率	100%
项目考核指标完成率	100%
项目考核指标完成率	100%

谢谢各位专家的聆听!



新知识发现与分类系统

北京邮电大学 大数据与智能信息处理实验室

2018年11月

目录

前言 知识发现与分类系统

1. 知识发现与分类系统

1.1 知识发现与分类系统的定义
1.2 知识发现与分类系统的组成
1.3 知识发现与分类系统的研究现状
1.4 知识发现与分类系统的研究意义

前言 知识发现与分类系统

1. 知识发现与分类系统的定义
2. 知识发现与分类系统的组成
3. 知识发现与分类系统的研究现状
4. 知识发现与分类系统的研究意义

1.1 知识发现与分类系统的定义
1.2 知识发现与分类系统的组成
1.3 知识发现与分类系统的研究现状
1.4 知识发现与分类系统的研究意义

目录

- 前言
- 项目概况及考核指标
- 开发过程概述
- 成果报告
- 项目考核完成情况



前言

新知识与多标签、分类的关系

• 目前学科特点

- 基础学科：覆盖全面，且基础知识完备。
- 新生学科：对于基础学科的深入性研究及应用性探索，通常产生学科分支或融合各个学科，重新组合成新生学科。
- 新学科产生特点：随着应用需求和研究成果产生，体现在相关期刊和科技文献的发表上。

• 多标签、新知识及分类

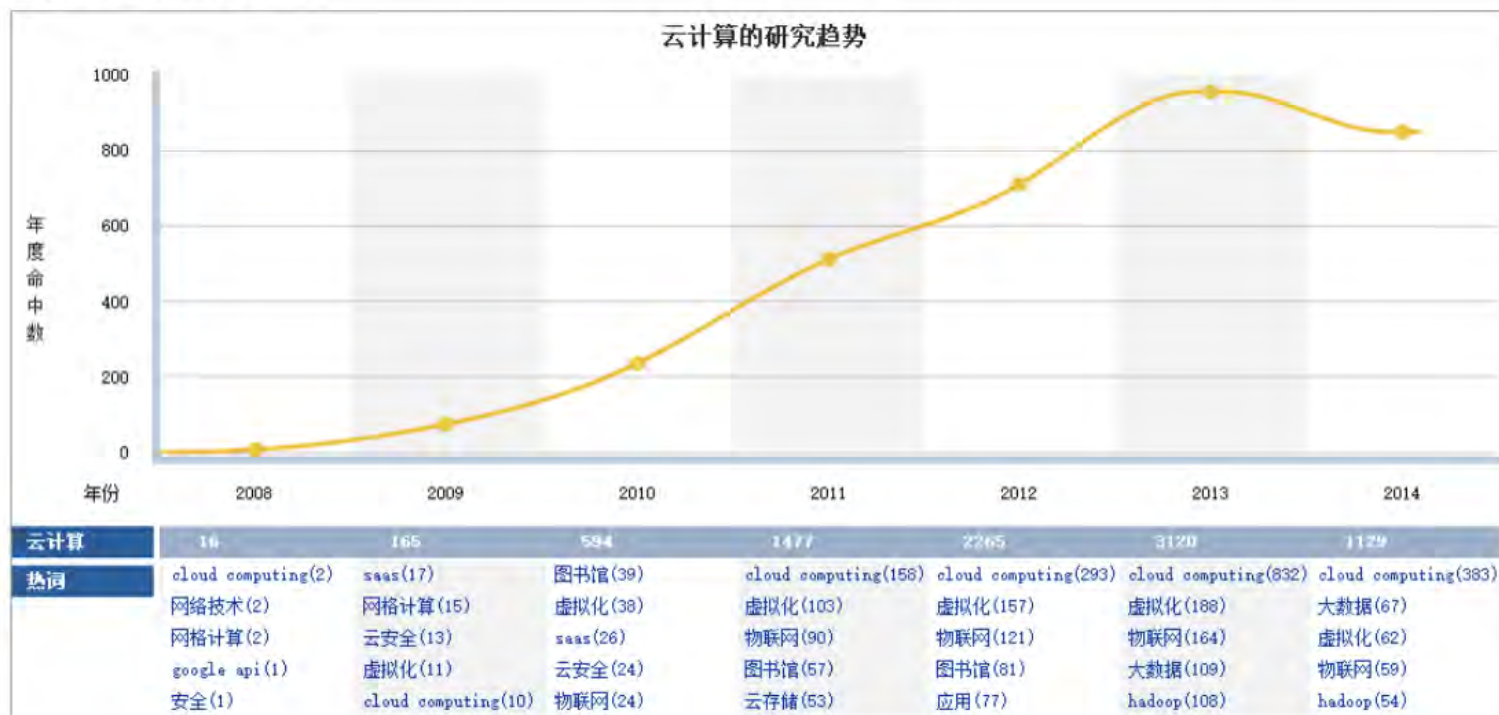
- 多标签：一个基础学科为一个标签，而多个标签组合为新生学科。
- 新知识：学科交叉产生，过去没有过的多标签组合。
- 分类：当某个新的多标签组合逐渐变多并趋于稳定时，产生一个新的学科分类。

前言 云计算为例

- 云计算领域兴起于2008年，由亚马逊推出的云计算服务开始诞生。
- 2010年至2011年被业界所认识，并采用。
- 2012年发展为比较成熟的技术，得到了业界的广泛使用。

前言 云计算为例

- 新生学科发展趋势：随着应用需求产生，伴随着相关期刊和科技文献的发表逐渐成熟。



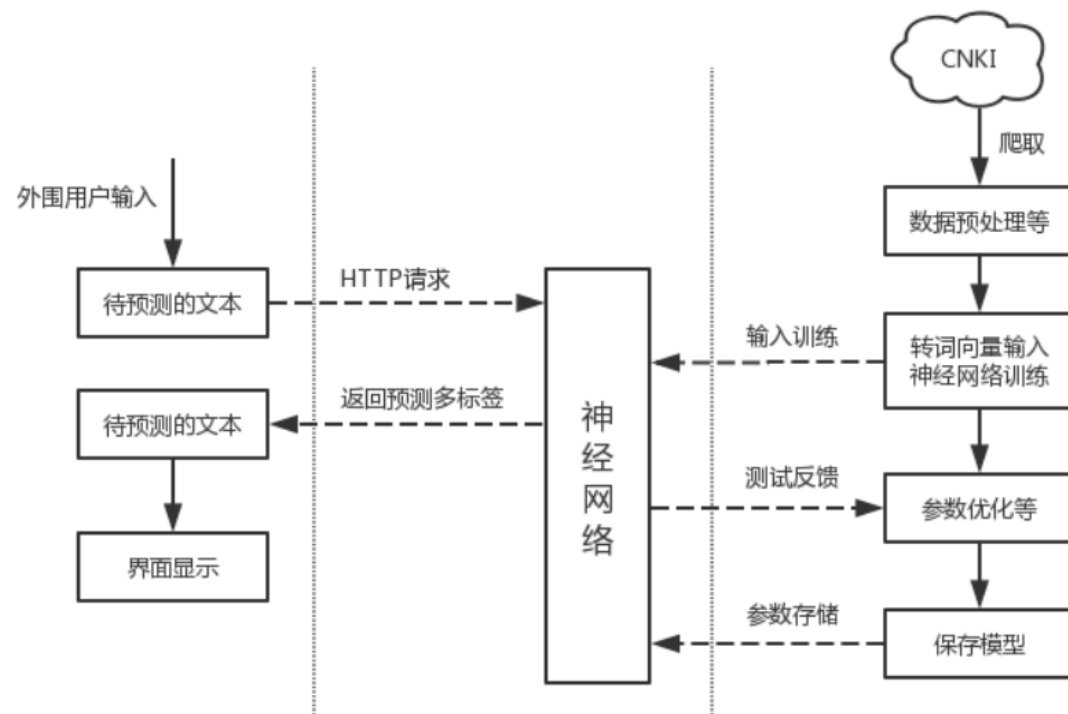
项目简介

新知识的发现与分类离不开多标签技术。随着信息的井喷式增长，我们接触到的不仅仅是已有传统科学分类。很多的新科学新知识伴随着不同学科的交叉应运而生。所以，当不同于以往的标签组合出现时，可能诞生当下新的知识和学科领域。

本项目使用使用分布式爬虫技术对知网(CNKI)所公开的科技文章进行爬取，总计16万余篇、10大类168小类作为原始语料。从原始语料中提取科技文章的标题、摘要和知网标签。通过数据清洗、数据增强等方式作进一步处理。然后基于深度学习网络对每一篇科技文章进行多分类，即多标签。然后将预测的标签组合在过去所有组合中查询是否出现过，若没有出现过则本系统判定发现新知识，进而对新的标签组合进一步进行分类。

通过新知识发现与分类系统，用户只需将文章的摘要输入，即可得到科技文献所匹配的多个知网标签；进而判断该篇科技文献是否可能为新知识，若该组合大量出现时，则为新知识则添加新的分类，完成新知识的发现与分类功能。

项目简介



考核指标

- 基于科技文章的数据处理模块
 - 对科技文章标题、摘要和多标签的快速爬取
 - 快速数据清洗与预处理
 - 文本的规约整理与高效存储
- 多标签模型的生成模块
 - 对已有科技文章文本内容进行数据增强
 - 训练特定针对于中文科技文章的词向量模型
 - 训练神经网络以自动生成科技文章的多标签
- 分类模型的优化和部署模块
 - 调优学习率、数据增强率和丢失率等参数
 - 设定多标签选取的阈值
 - 保存最终模型用于部署与展示

开发过程概述

11月2日-11月5日

科技文章的数据爬取，实现数据清洗与预处理，完成规约与存储。

11月6日-11月13日

用大量的规约数据对神经网络进行训练，并完成初步的测试。

11月13日-11月16日

对深度网络中的参数进行调优，调整阈值等完成模型优化。

11月16日-11月22日

界面编写、测试和优化。

成果报告

基于科技文章的数据处理模块

- 对科技文章标题、摘要和多标签的快速爬取与存储



- 快速数据清洗与预处理



对科技文章标题、摘要和多标签的快速爬取与存储

快速数





北京市科学技术情报研究所

Beijing Institute of Science and Technology Information

admin

后端管理页面

- 用户
- 数据
- 数据管理**

用户管理 | 用户日志管理 | **数据管理**

基础学科	工程科技 I 级	工程科技 II 级	农业科技	医药卫生科技	哲学与人文科学	社会科学 I 辑	社会科学 II 辑	信息科技	经济与管理科学											
自然科学与方法	数学	化学	无机化工	工业通用技术	机械工业	农业基础科学	农业工程	医药卫生	医学教育与预防医学	文艺理论	世界文学	马克思主义	中国共产党	社会科学理论	社会学及统计学	无线电电子学	通信技术	宏观经济管理	经济理论与发展	
力学	物理学	工程	工业	航空航天	武器工业与工程	园艺	林业	预防医学与中医学	外国语言文学	音乐舞蹈	思想政治教育	行政学及国家行政学	人才学	与劳动教育理论	及计算机应用	自动化技术	经济统计	农业经济		
天文学	生物学	材料科学	矿业工程	学与工程	军事技术	畜牧与动物医学	蜜蜂与野生	卫生学	中医学	戏剧电影与艺术	育	行政管理	人才学	与劳动教育理论	及计算机应用	自动化技术	经济统计	农业经济		
自然地理学	气象学	金属学及冶金工程	冶金工业	铁路运输	公路与水路运输	医学	动物保护	中西医结合	基础医学	电视艺术	政党及群众组织	军事	动科学	教育管理	用	新闻与传媒	工业经济	交通运输经济		
和测绘学	海洋学	陶瓷工艺	轻工业手工	业	汽车工业	船舶工业		临床医学	感染性疾病	美术书法	地理	组织	公安	学前教育	初等教育	出版	图书情报与	经济		
地质学	地理物理学	环境科学	与资源利用	安全科学与工程	水利电力工程	建筑科学与工程		心血管系统	呼吸系统疾病	史学理论	世界历史	行政法及地方法制	民法	职业教育	成人教育	档案及博物	文化经济	信息经济与		
资源科学		环境科学	与资源利用	安全科学与工程	水利电力工程	建筑科学与工程		消化系统	内分泌腺及病	中国通史	中国民族与地方史志	经济法	诉讼法与司法制度	特殊教育	馆	邮政经济	服务业经济	贸易经济		
		环境科学	与资源利用	安全科学与工程	水利电力工程	建筑科学与工程		病	全身性疾病	中国古代史	中国近现代史	国际法				财政与税收	金融	投资	会计	
		环境科学	与资源利用	安全科学与工程	水利电力工程	建筑科学与工程		外科学	泌尿科学	考古	人物传记					证券	保险	审计	市场研究与	
		环境科学	与资源利用	安全科学与工程	水利电力工程	建筑科学与工程		妇产科学	儿科学	哲学	逻辑学					信息		管理学	领导学与决	
		环境科学	与资源利用	安全科学与工程	水利电力工程	建筑科学与工程		神经病学	精神病学	伦理学	心理学					策学		科学研究管	理	
		环境科学	与资源利用	安全科学与工程	水利电力工程	建筑科学与工程		肿瘤学	眼科与耳鼻喉科	美学	宗教									
		环境科学	与资源利用	安全科学与工程	水利电力工程	建筑科学与工程		口腔科学	皮肤病与性病											
		环境科学	与资源利用	安全科学与工程	水利电力工程	建筑科学与工程		特种医学	急救医学											
		环境科学	与资源利用	安全科学与工程	水利电力工程	建筑科学与工程		军事医学	药学											
		环境科学	与资源利用	安全科学与工程	水利电力工程	建筑科学与工程		卫生	生物医学工											
3				1976年龙陵地震诱发滑坡的影响因子敏感性分析				作为1976年龙陵地震诱发滑坡的影响因子。利用GIS强大的空间分析能力,结合滑坡确定性系数(CF)的方法,对1976年龙陵地震诱发滑坡的诸影响因子进行敏感性分析,确定了该区域内各因子最利于地震滑坡发育的数值区间,为进一步区域地震滑坡稳定性评价奠定基础。						天文学、地球科学	天文学、地球科学			地球物理学	水文地质学与工程地质学	
								在整合区域海水入侵历史观测数据的基础上,结合现代遥感综合观测手段,开展莱州湾南岸海水入侵时空动态变化过程和海岸线历史												





北京市科学技术情报研究所

Beijing Institute of Science and Technology Information

admin

后端管理页面

用户

数据

数据管理

用户管理 | 用户日志管理 | 数据管理

基础学科	工程科技 I 级	工程科技 II 级	农业科技	医药卫生科技	哲学与人文科学	社会科学 I 级	社会科学 II 级	信息科技	经济与管理科学
ID	标题	摘要	类别1	类别2					
1	2.6GHz高速CMOS环形振荡器设计	文章提出了一种偶数级环形振荡器的设计方案,中心频率为2.3GHz,利用起振电路使其能够快速起振,当环形振荡器的控制电压为1.2~2.0V时,其线性调谐范围为1.9~2.6GHz;电路设计采用TSMC0.18μm 1P6M混合信号生产工艺;利用Cadence Spectre RF进行仿真。结果显示,在中心频率为2.3GHz、偏移载波频率为10MHz的情况下,环形振荡器的相位噪声为-112.9dBc/Hz。该电路可用于高速锁相环的设计中。	工业技术	无线电电子学、电信技术					
2	2018 IEEE EMC&APEMC国际学术研讨会——最佳文章(二)	<正>Enforcing Correlation Between Conducted and Radiated Susceptibility Test Setups for Aerospace Involving Shielded Cables摘要:设计通过了大电流注入方法的传导敏感度试验程序,重现航天设备中屏蔽电缆标准辐射敏感度的测试过程。传导敏感度和辐射敏感度之间的相关性由“过测试”来解释。过测试是一个随机变量,作者通过其累积分布函数来表征,并用实验进行了验证。在实现大电流注入与辐射敏感度试验之间相关性的基础上,进行了相关分析和讨论。	工业技术 航空、航天	无线电电子学、电信技术 航天(宇宙航行)					
3	2018全球半导体产值首破5000亿美元	<正>过去两年中由于内存、闪存芯片涨价,半导体行业产值接连创新高,这一趋势在2018年还将继续。来自IC Insights的消息称,2018年全球电子系统产值将达到1.62万亿美元,其中半导体市场产值将达到5091亿美元,也是全球半导体产值首次突破5000亿美元大关,同比增长14%,原因跟DRAM内存、NAND闪存均价大幅增长有关。根据IC Insights的预测,今年全球电子系统的产值为1.622万亿美元,而	经济	工业经济					
4	2μm波段单频拉曼光纤放大器中受激布里渊散射研究	2μm波段高功率单频拉曼光纤放大器(SF-RFA)的性能通常会受到受激布里渊散射(SBS)效应的影响。本文考虑SBS效应下拉曼光纤放大器的速率方程,建立了SBS限制下的SF-RFA理论模型,计算了SF-RFA的能量分布和功率输出。讨论了泵浦功率、光纤长度等因素对光纤放大器中SBS效应和输出功率的影响,总结了如何在提高放大器输出功率的同时有效抑制SBS效应的方法。本文数值模拟结果为实验研究高功率SF-RFA中抑制SBS效应提供一定的指导。	工业技术	无线电电子学、电信技术					
5	8英寸绝缘栅双极型晶体管(IGBT)关键技术研究	基于8英寸绝缘栅双极型晶体管(IGBT)生产线的建设,重点解决了8英寸IGBT先进工艺技术、第四代高压双扩散金属氧化物半导体(DMOS+)IGBT技术和第五代沟槽栅IGBT技术等关键技术问题,实现了高压IGBT芯片制造从6英寸到8英寸的技术突破,自主开发的1600 A/1.7 kV与1500 A/3.3 kV IGBT模块已经被成功制造并通过考核,现已应用于轨道交通牵引系统。	工业技术	无线电电子学、电信技术					



快速数据清洗与预处理





北京市科学技术情报研究所

Beijing Institute of Science and Technology Information

bigdata v

- 前端展示界面
- 数据展示
- 数据处理
- 导入数据
- 切词分词处理
- 词相似度展示
- 类别预测

- 标签分布
- 数据总览
- 导入数据

输入需要查询的文章摘要 提交成功! 请在下面步骤查看结果, 或者重新提交新数据

摘要 生成树协议是网络核心层建设中经常采用的解决冗余链路的技术,但由于网络链路的复杂性及生成树协议的工作原理,导致在配置STP时经常会出现一些隐蔽性较强的故障,本文分析了生成树协议的常见故障并对故障的解决提出了一些建议。

立即提交 重置





- 前端展示界面
- 数据展示
- 数据处理
- 导入数据
- 切词分词处理**
- 词相似度展示
- 类别预测

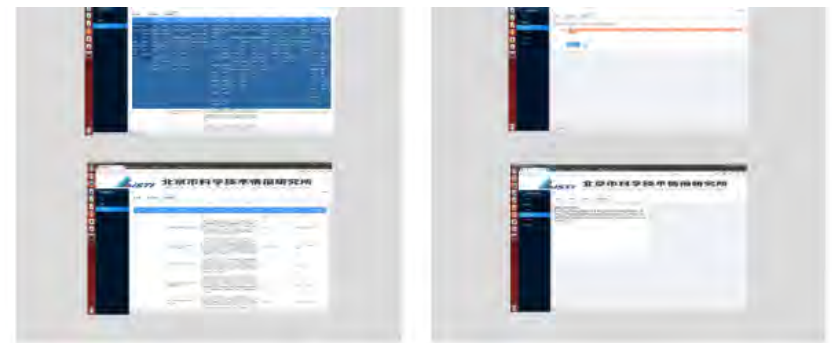
- 标签分布
- 数据总览
- 导入数据
- 切词分词处理**

最新提交数据分词结果：
 生成 树 协议 网络 核心层 建设 中 采用 解决 冗余 链路 技术 网络 链路 复杂性 生成
 树 协议 工作 原理 导致 配置 时 出现 隐蔽性 较强 故障 本文 分析 生成 树 协议 常见
 故障 故障 解决 提出 建议 空空空空空空空空空空空空空空空空空空空空空空
 空空空空空空



· 文本规则约束学习与双向付函

- 多标签模型的生成模块
 - 对已有科技文章文本内容进行数据增强
 - 训练特定针对于中文科技文章的词向量模型
 - 训练神经网络以自动生成科技文章的多标签
- 分类模型的优化和部署模块
 - 调优学习率、数据增强率和丢失率等参数
 - 设定多标签选取的阈值
 - 保存最终模型用于部署与展示



成果报告 多标签模型的生成模块

· 数据增强与补齐截断



· 针对科技文章的词向量模型训练

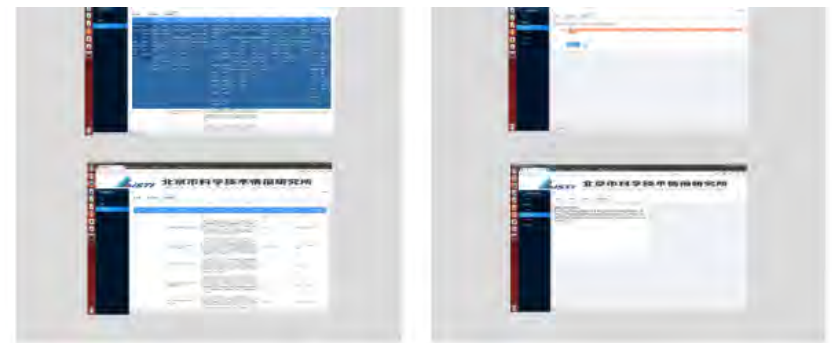


· 多标签预测



· 文本正则表达式与向量化

- 多标签模型的生成模块
 - 对已有科技文章文本内容进行数据增强
 - 训练特定针对于中文科技文章的词向量模型
 - 训练神经网络以自动生成科技文章的多标签
- 分类模型的优化和部署模块
 - 调优学习率、数据增强率和丢失率等参数
 - 设定多标签选取的阈值
 - 保存最终模型用于部署与展示



成果报告 多标签模型的生成模块

· 数据增强与补齐截断



· 针对科技文章的词向量模型训练



· 多标签预测





北京市科学技术情报研究所

Beijing Institute of Science and Technology Information

bigdata

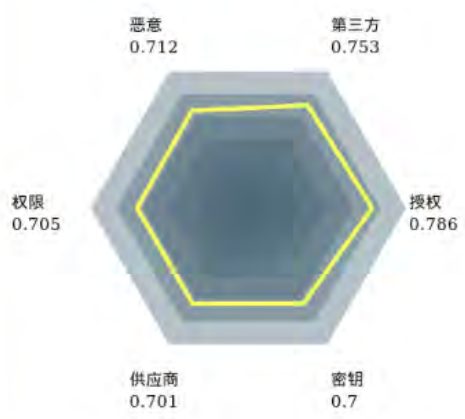
- 前端展示界面
- 数据展示
- 数据处理
- 导入数据
- 切词分词处理
- 词相似度展示**
- 类别预测

- 标签分布
- 数据总览
- 导入数据
- 切词分词处理
- 词相似度展示**

输入需要查询的词汇

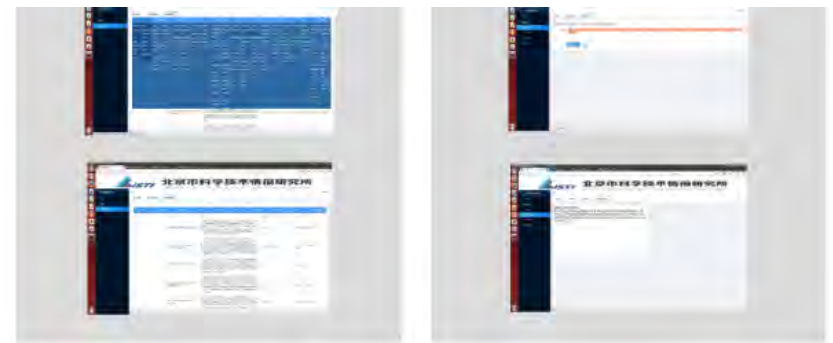
相似词展示 协议

立即提交 重置



· 文本规则约束学习与双向付函

- 多标签模型的生成模块
 - 对已有科技文章文本内容进行数据增强
 - 训练特定针对于中文科技文章的词向量模型
 - 训练神经网络以自动生成科技文章的多标签
- 分类模型的优化和部署模块
 - 调优学习率、数据增强率和丢失率等参数
 - 设定多标签选取的阈值
 - 保存最终模型用于部署与展示



成果报告 多标签模型的生成模块

· 数据增强与补齐截断



· 针对科技文章的词向量模型训练



· 多标签预测





北京市科学技术情报研究所

Beijing Institute of Science and Technology Information

bigdata

- 前端展示界面
- 数据展示
- 数据处理
- 类别预测
- 类别结果预测**

- 标签分布
- 数据总览
- 导入数据
- 切词分词处理
- 词相似度展示
- 类别结果预测**

最新提交数据预测结果：
自动化技术、计算机技术

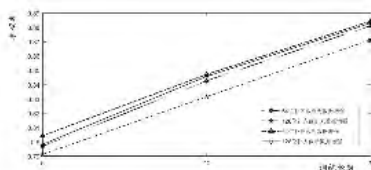


成果报告

分类模型的优化与部署模块

· 调优模型相关参数

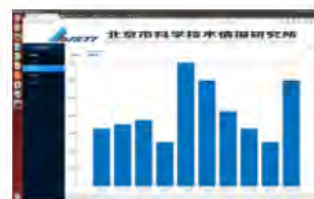
- 训练轮数
- 学习率
- 丢失率
- 数据增强率等



· 设定多标签选取阈值



· 部署



```
Fitted to 70244 vectors
Fitted to 70199 vectors
Fitted to 70155 vectors
Fitted to 71484 vectors
Fitted to 70082 vectors
Fitted to 69845 vectors
Fitted to 69919 vectors
Fitted to 69919 vectors
Fitted to 70174 vectors
Fitted to 69701 vectors
Fitted to 70391 vectors
Fitted to 69830 vectors
Fitted to 70193 vectors
Fitted to 70095 vectors
Fitted to 70535 vectors
Fitted to 70530 vectors
Fitted to 70480 vectors
Fitted to 70405 vectors
Fitted to 70509 vectors
Fitted to 70301 vectors
Fitted to 70570 vectors
Fitted to 70452 vectors
Fitted to 30807 vectors
```

开始训练

```
WARNING:tensorflow:From D:\Anaconda3\lib\site-packages\keras\backend\tensorflow_backend.py:1062: calling reduce_prod (from tensorflow.python.ops.math_ops) with
keep_dims is deprecated and will be removed in a future version.
```

Instructions for updating:

keep_dims is deprecated, use keepdims instead

```
WARNING:tensorflow:From D:\Anaconda3\lib\site-packages\keras\backend\tensorflow_backend.py:1123: calling reduce_mean (from tensorflow.python.ops.math_ops) with
keep_dims is deprecated and will be removed in a future version.
```

Instructions for updating:

keep_dims is deprecated, use keepdims instead

Epoch 1/2

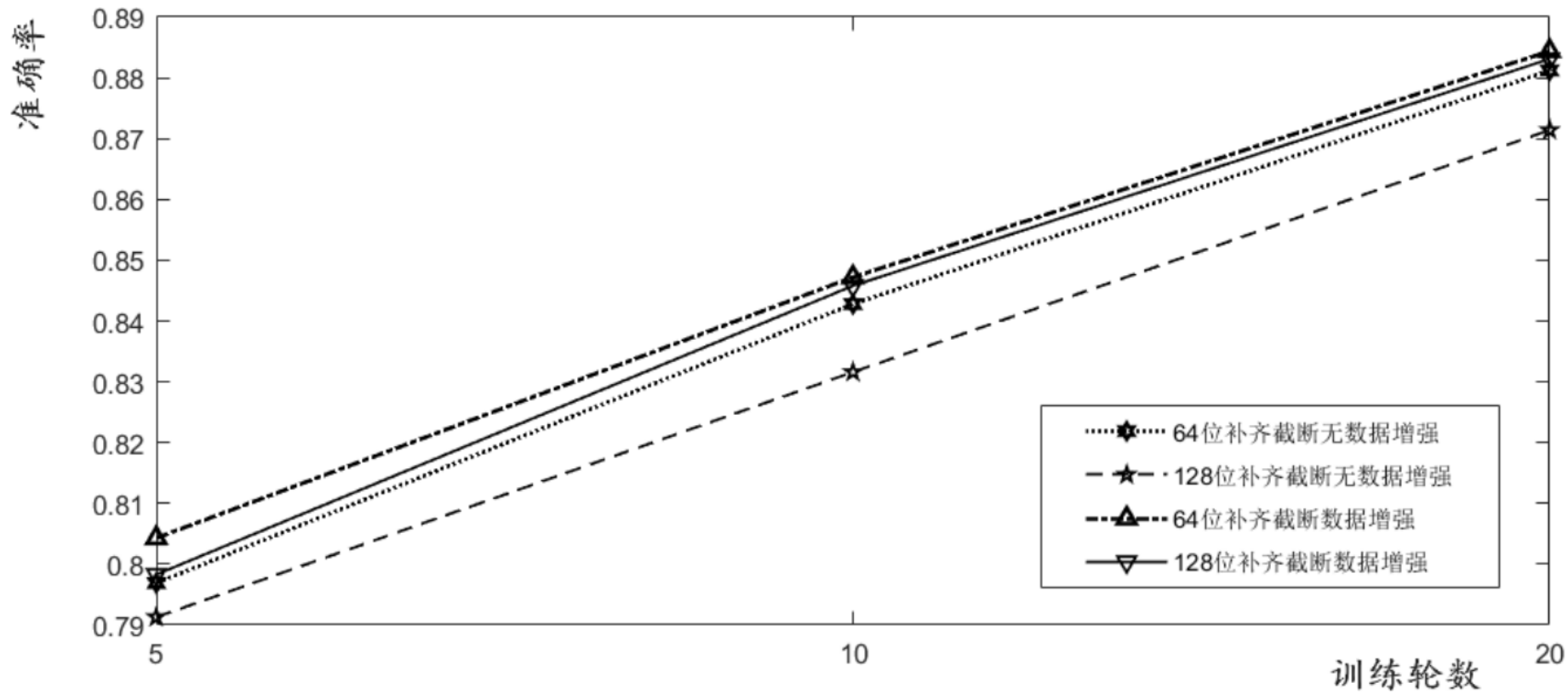
```
1575/1575 [=====] - 1560s - loss: 0.0220 - top_k_categorical_accuracy: 0.5631
```

Epoch 2/2

```
1575/1575 [=====] - 1544s - loss: 0.0172 - top_k_categorical_accuracy: 0.7053
```

Process finished with exit code 0





['伦理学(道德哲学)', 0.9495048], ['中国政治', 0.11589816], ['社会科学现状及发展', 0.042934787], ['经济计划与管理', 0.021217635], ['美学', 0.016545633], ['工业经济', 0.006914209], ['自动化技术、计算机技术', 0.0045790104], ['高等教育', 0.0044955728], ['社会学', 0.003077357], ['宗教', 0.00285954], ['信息与知识传播', 0.0027452712], ['经济学', 0.002354216], ['职业技术教育', 0.0021726484], ['管理学', 0.0017242201], ['法律', 0.001560161], ['贸易经济', 0.001255112], ['一般工业技术', 0.00071523443], ['世界各国经济概况、经济史、经济地理', 0.00070402917], ['内科学', 0.0006592507], ['中国文学', 0.0006581223], ['中等教育', 0.0006049858], ['中国人物传记', 0.0005353089], ['工人、农民、青年、妇女运动与组织', 0.0005285489], ['中国共产党', 0.0005283288], ['财政、金融', 0.00049267896], ['世界各国文化与文化事业', 0.0004837525], ['物理学', 0.00041951353], ['中国哲学', 0.00036185488], ['科学、科学研究', 0.00035391483], ['无线电电子学、电信技术', 0.00033406448], ['语言学', 0.0003048893], ['逻辑学(论理学)', 0.00028289558], ['交通运输经济', 0.00027441935], ['中国军事', 0.00025456236], ['自然科学理论与方法论', 0.00020981164], ['临床医学', 0.00018396796], ['化学', 0.0001822868], ['神经病学与精神病学', 0.00016658955], ['航天(宇宙航行)', 0.00014129988], ['外交、国际关系', 0.00013456853], ['农业经济', 0.00013062112], ['艺术理论', 0.00012989178], ['教育', 0.00012162757], ['一般理论', 9.4960065e-05], ['环境质量评价与环境监测', 9.200258e-05], ['政治理论', 8.97607e-05], ['地理', 7.798227e-05], ['化学工业', 7.2577524e-05], ['中国史', 7.124617e-05], ['畜牧、动物医学、狩猎、蚕、蜂', 6.516095e-05], ['武器工业', 6.429606e-05], ['废物处理与综合利用', 6.186415e-05], ['水路运输', 5.757935e-05], ['预防医学、卫生学', 5.5687553e-05], ['公路运输', 5.4708617e-05], ['世界各国艺术概况', 5.1250317e-05], ['人才学', 4.8280544e-05], ['汉语', 4.4588538e-05], ['建筑科学', 4.3913657e-05], ['社会与环境', 4.2244756e-05], ['机械、仪表工业', 4.1085765e-05], ['能源与动力工程', 4.039077e-05], ['绘画', 4.0152197e-05], ['电工技术', 3.6318415e-05], ['中国教育事业', 3.6307265e-05], ['环境保护管理', 3.587324e-05], ['环境污染及其防治', 2.6274973e-05], ['金属学与金属工艺', 2.3317509e-05], ['社会科学理论与方法论', 2.022043e-05], ['统计学', 1.9963658e-05], ['邮电经济', 1.8840656e-05], ['心理学', 1.8811677e-05], ['基础医学', 1.8629544e-05], ['特种医学', 1.722242e-05], ['体育', 1.71839e-05], ['自然科学现状及发展', 1.611007e-05], ['轻工业、手工业', 1.5797907e-05], ['肿瘤学', 1.5633883e-05], ['初等教育', 1.47442815e-05], ['史学理论', 1.30325825e-05], ['戏剧艺术', 1.2162655e-05], ['世界文学', 1.1846312e-05], ['哲学理论', 1.1519617e-05], ['文化理论', 1.0955755e-05], ['工艺美术', 1.0437116e-05], ['力学', 1.0271053e-05], ['水产、渔业', 1.01110545e-05], ['测绘学', 9.807546e-06], ['环境科学基础理论', 9.363385e-06], ['成人教育、业余教育', 9.216937e-06], ['自然科学机构、团体、会议', 8.849664e-06], ['风俗习惯', 8.467902e-06], ['原子能技术', 8.113196e-06], ['文学理论', 6.8014565e-06], ['安全科学', 6.460122e-06], ['园艺', 6.3961534e-06], ['外科学', 5.537938e-06], ['中国医学', 5.084372e-06], ['大气科学(气象学)', 4.9151863e-06], ['世界史', 4.518425e-06], ['运筹学', 4.379279e-06], ['林业', 4.2671013e-06], ['冶金工业', 4.2324127e-06], ['马克思主义、列宁主义、毛泽东思想、邓小平理论的学习和研究', 4.220647e-06], ['药学', 4.1337075e-06], ['海洋学', 4.083071e-06], ['数学', 3.697977e-06], ['系统科学', 3.2626424e-06], ['妇产科学', 2.9785513e-06], ['劳动科学', 2.4050494e-06], ['民族学', 2.3361392e-06], ['中国文物考古', 2.313787e-06], ['学前教育、幼儿教育', 2.219311e-06], ['低温物理学', 2.148447e-06], ['航空', 2.0386553e-06], ['地球物理学', 1.8175358e-06], ['欧洲哲学', 1.6729451e-06], ['天文学', 1.4456116e-06], ['皮肤病学与性病学', 1.4262456e-06], ['眼科学', 1.3912405e-06], ['水利工程', 1.2725255e-06], ['矿业工程', 1.2194733e-06], ['植物保护', 1.0771859e-06], ['耳鼻咽喉科学', 1.0622813e-06], ['农业基础科学', 1.0382408e-06], ['石油、天然气工业', 9.706669e-07], ['战略学、战役学、战术学', 9.504493e-07], ['世界政治', 7.934852e-07], ['农作物', 6.3568973e-07], ['植物学', 6.3311956e-07], ['电影、电视艺术', 5.6848575e-07], ['人口学', 5.6209046e-07], ['特殊教育', 4.686335e-07], ['军事理论', 4.5974159e-07], ['地质学', 4.3236858e-07], ['自然地理学', 3.23398e-07], ['军事技术', 3.2045472e-07], ['航空、航天技术的研究与探索', 2.9296123e-07], ['生物工程学(生物技术)', 2.2749772e-07], ['生物物理学', 1.6487576e-07], ['分子生物学', 1.6302235e-07], ['自然科学研究方法', 1.6021782e-07], ['社会教育', 1.5417817e-07], ['雕塑', 1.3446902e-07], ['世界军事', 1.3388552e-07], ['概率论与数理统计', 1.2793343e-07], ['文物考古', 1.00250574e-07], ['儿科学', 9.110942e-08], ['音乐', 8.621066e-08], ['矿床学', 7.7874745e-08], ['铁路运输', 6.8876446e-08], ['农学(农艺学)', 6.4493975e-08], ['亚洲史', 5.8270004e-08], ['控制论、信息论(数学理论)', 5.696754e-08], ['美洲哲学', 4.7901253e-08], ['医学研究方法', 3.4004024e-08], ['欧洲史', 2.9226937e-08], ['家庭教育', 2.908796e-08], ['摄影艺术', 2.6925475e-08], ['原子核物理学、高能物理学', 2.58436e-08], ['世界哲学', 2.4640434e-08], ['水文地质学与工程地质学', 2.3971914e-08], ['生物科学教育与普及', 1.9972024e-08], ['农业工程', 1.9967379e-08], ['航天系统工程', 1.7795532e-08], ['世界文物考古', 1.6662952e-08], ['常用外国语', 1.6621781e-08], ['环境保护宣传教育及普及', 1.2045033e-08], ['自然科学丛书、文集、连续性出版物', 1.0540006e-08], ['航天用燃料(推进剂)及润滑剂', 9.5241175e-09], ['地球物理勘探', 9.284489e-09], ['师范教育', 8.778347e-09], ['现状与发展', 6.3212626e-09], ['灾害及其防治', 6.316743e-09], ['社会科学研究方法', 5.7209237e-09], ['农业科学技术现状与发展', 4.7211333e-09], ['美洲史', 2.740305e-09], ['计算数学', 2.6169199e-09], ['仿生学', 2.5566759e-09], ['动物学', 2.5566759e-09], ['传记', 2.3282297e-09], ['毛泽东著作', 2.1222219e-09], ['社会科学机构、团体





北京市科学技术情报研究所

Beijing Institute of Science and Technology Information

bigdata v

- A 前端展示界面
- ✓ 数据展示
- 标签分布
- 数据总览
- 数据处理
- 类别预测





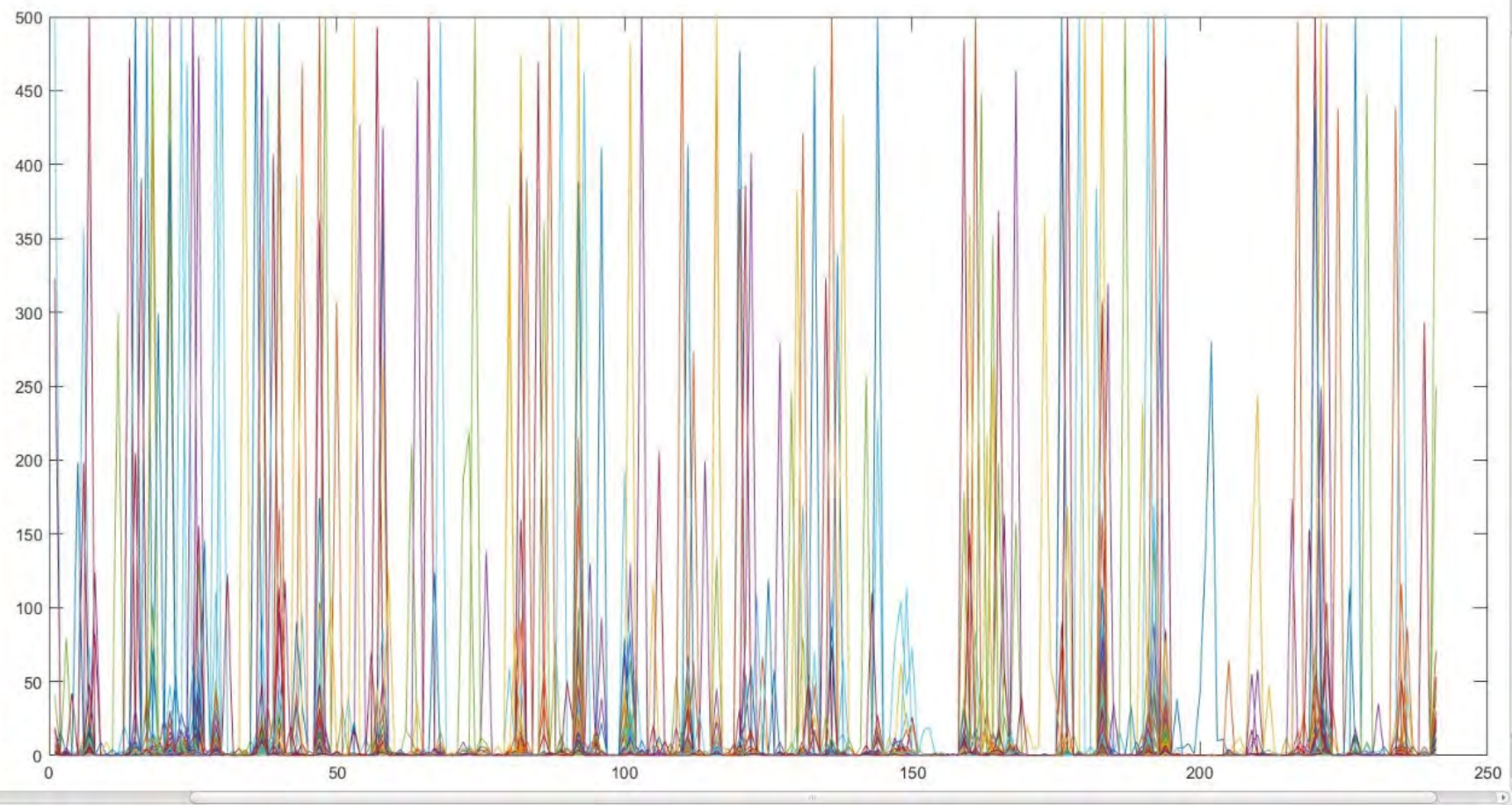
北京市科学技术情报研究所

Beijing Institute of Science and Technology Information

bigdata ▾

- 前端展示界面
- 数据展示
- 标签分布**
- 数据总览
- 数据处理
- 类别预测

标签分布





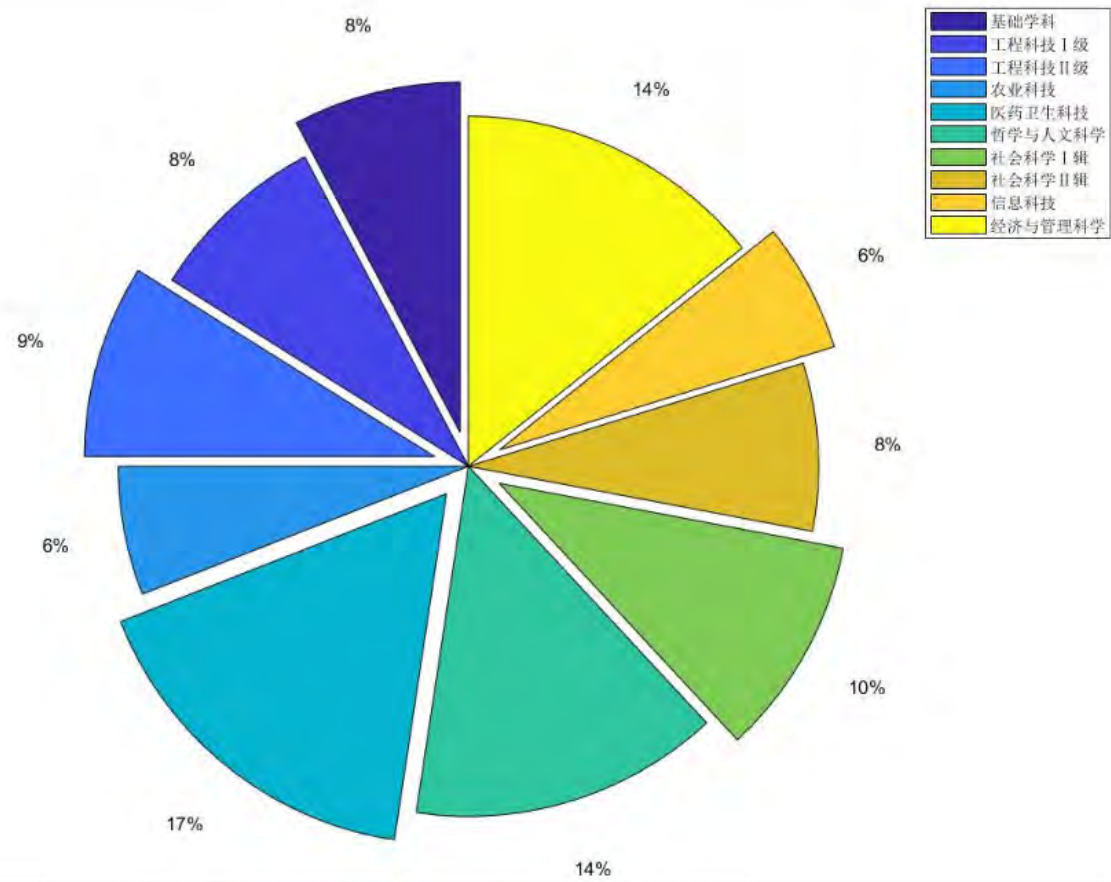
北京市科学技术情报研究所

Beijing Institute of Science and Technology Information

bigdata

- 前端展示界面
- 数据展示
- 标签分布
- 数据总览**
- 数据处理
- 类别预测

标签分布 数据总览



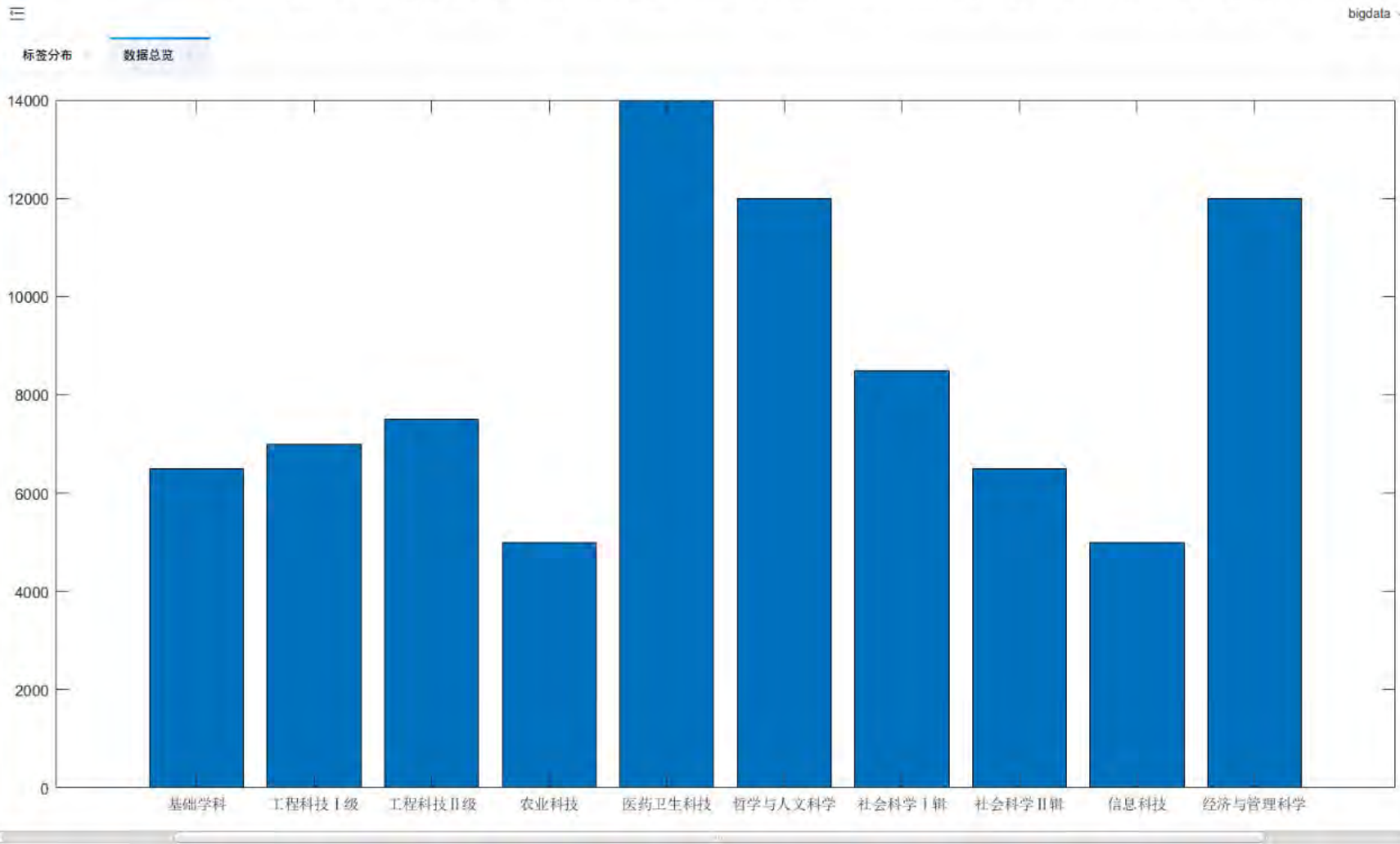


北京市科学技术情报研究所

Beijing Institute of Science and Technology Information

bigdata

- 前端展示界面
- 数据展示
- 标签分布
- 数据总览**
- 数据处理
- 类别预测





北京市科学技术情报研究所

Beijing Institute of Science and Technology Information

admin

后端管理页面

用户

用户管理

用户日志管理

数据

用户管理

ID	用户名	性别	邮箱	描述	操作1	操作2
1	1	man	11	112	编辑	删除
2	name_8	8	8	8	编辑	删除
3	name_9	9	9	9	编辑	删除
4	name_10	10	10	10	编辑	删除
5	name_11	11	11	11	编辑	删除
6	name_12	12	12	12	编辑	删除
7	name_13	13	13	13	编辑	删除
8	name_14	14	14	14	编辑	删除
9	name_15	15	15	15	编辑	删除
10	name_16	16	16	16	编辑	删除
11	name_17	17	17	17	编辑	删除
12	name_18	18	18	18	编辑	删除
13	name_20	20	20	20	编辑	删除
14	name_21	21	21	21	编辑	删除
15	name_22	22	22	22	编辑	删除





北京市科学技术情报研究所

Beijing Institute of Science and Technology Information

admin

后端管理页面

- 用户
- 用户管理
- 用户日志管理
- 数据

用户管理 | 用户日志管理

用户信息

ID	用户名	操作	操作时间
1	1	login	Nov. 13, 2018, 2:16 p.m.
2	1	checkPredictData	Nov. 13, 2018, 2:21 p.m.
3	1	checkDataShow	Nov. 13, 2018, 2:22 p.m.
4	1	checkLabelDistribution	Nov. 13, 2018, 2:22 p.m.
5	1	checkParticipleWord	Nov. 13, 2018, 2:22 p.m.
6	1	checkPredictData	Nov. 13, 2018, 2:22 p.m.
7	1	logout	Nov. 13, 2018, 2:22 p.m.
8	c	register	Nov. 13, 2018, 2:23 p.m.
9	c	checkDataShow	Nov. 13, 2018, 2:23 p.m.
10	0	login	Nov. 13, 2018, 2:28 p.m.
11	1	login	Nov. 13, 2018, 2:28 p.m.
12	2	login	Nov. 13, 2018, 2:28 p.m.
13	3	login	Nov. 13, 2018, 2:28 p.m.
14	4	login	Nov. 13, 2018, 2:28 p.m.
15	5	login	Nov. 13, 2018, 2:28 p.m.
16	6	login	Nov. 13, 2018, 2:28 p.m.
17	7	login	Nov. 13, 2018, 2:28 p.m.
18	8	login	Nov. 13, 2018, 2:28 p.m.



项目考核完成情况

计划工作内容	实际完成情况
科技文章的数据处理模块	完成
多标签模型的生成模块	完成
分类模型的优化和部署模块	完成
界面展示和用户交互模块	完成

谢谢各位专家的聆听！