

曹丛峰 中国科学院大学 遥感与数字地球研究所 硕士

姓名	曹丛峰
地址	北京市海淀区清华同方大厦 A 座 20 层
邮箱	caocf@radi.ac.cn
电话	+86 13021919407
GitHub	https://github.com/frfy
个人主页	https://byfrfy.plus/CV
博客	https://byfrfy.top
性别	男
雅思	6

个人简介

我目前在清华同方大数据技术中心担任算法工程师，主要负责研究自然语言处理，知识图谱和文本挖掘。同时也是算法组的组长负责带领 6 位同事做自然语言处理方面的研究。在研究生和本科阶段我做过很多计算机视觉和数字图像处理方面的研究。在研究生和本科阶段主要学习机器学习，深度学习和模式识别相关课程，对机器学习和深度学习理论有较深的理解。此外，我对深度学习，计算机视觉和自然语言处理的前沿研究成果也很熟悉，也跟高校的博士合作过很多项目。

教育

中国科学院大学 中国科学院遥感与数字地球研究所	2014-2017
信号与信息处理	GPA: 3.58/4.0
山东农业大学 信息科学与工程学院	2010-2014
遥感科学与技术	GPA: 3.7/4.0
计算机科学与技术	GPA: 3.5/4.0

工作/研究经历

清华同方 大数据技术中心	2017-至今
<ul style="list-style-type: none">负责自然语言处理，知识图谱和文本挖掘的算法研究和落地。负责实现项目所需算法的开发以及服务搭建。算法组组长，负责管理算法团队工作进度和分配任务，负责项目管理。	
产品：知识图谱产品	2017-2019
项目经理/算法工程师	
<ul style="list-style-type: none">实现知识图谱的半自动化构建，通过自然语言处理和深度学习实现实体关系的自动识别。	

- 与清华大学人工智能学院针对知识图谱进行学术交流，实现实体和关系抽取更高准确率。
- 产品应用在海关二期攻关项目，为海关构建行人，违禁品和濒危物种知识图谱，辅助海关查验可疑人和可疑货物。
- 应用于重庆交委交通项目，提取路段和司机的事故信息等用于研判。

[链接](#)

产品：大数据分析 and 智能决策平台

2018-至今

算法负责人/算法工程师

- 通过自然语言处理识别文本要素（主要是指标等分析要素）与数据分析的要素进行关联
- 对文本分类，提取标签，构建实体关系模型。
- 对于从文本中识别出的关联要素，例如经济指标，通关数据分析方法进行预测和回归。
- 产品应用在江苏和湖北等地。

[链接](#)

产品：智能政务系统

2019-至今

算法负责人/算法工程师

- 对政务发布文本进行分析，通过自然语言处理实现提取话题，文本摘要提取，文本相似度计算，文本的推荐和挖掘。
- 与中国知网合作，挖掘期刊中的政务信息和人物信息，关联当地政务系统对政务文件和成果报告实现更准确的推荐。

产品：新冠肺炎预防和分析平台

2020

算法工程师/算法设计

- 对行人轨迹和地区进行分析，对人员感染风险进行分类和预警。

中国科学院大学，中国科学院遥感与数字地球研究所

2014-2017

- 负责前沿的机器学习，深度学习和计算机视觉的研究，将研究成果应用于遥感图像拼接，分割和目标识别。
- 负责实现项目所需算法，如图像拼接，目标检测和图像分割等。

项目：大科学工程项目-多模态相机的研制和图像处理

2014-2017

算法研究/参与实验

- 提供相机快视图图像的拼接算法（SURF, SIFT 和相关改进算法）。
- 参与相机研究和飞行检校实验。

项目：北京市科技创新项目-基于窄带滤光片阵列的无人机

2015-2017

多光谱小相机的研制

算法研究/参与实验

- 为相机快视图提供图像拼接算法（ORB, SURF 和相关改进算法）。
- 研究基于滤光片相机的多光谱图像的目标提取和图像分割算法。
- 发表文章 “Development of UAV-borne multispectral camera based on narrow band width filter array”。

项目：探月三期 - 图像处理模块中图像数据的设计与实现

2014

数据库设计实现/参与实验

- 使用数据库存储原始图像和经过算法处理的图像，存储对应的图像质量

评价参数和实验参数等相关数据。

与高校合作项目

项目： 中医药问答项目（北京大学医学院） **2018-2019**

算法研究/系统开发

- 从文本中构建中医药，症状和疾病的知识图谱。
- 基于知识图谱构建关于症状，疾病和相对应药物的问答系统。

项目： 公共舆情分析系统（中国传媒大学） **2018-2019**

算法研究/系统开发

- 通过爬虫抓取微博数据，通过爬取数据构建文本分类，文本聚类，文本相似度计算，话题和情感分析模型，整合这些模型构建舆论分析系统，实时监督制定账号推文对相关推文指数进行预测和判断。

项目： 颅骨 CT 影像下颌检测（北京大学医学院） **2017-2018**

算法研究/系统开发

- 检测人体头颅 CT 影像并且标定头颅中下颌骨的位置。

项目： 基于树莓派的人脸识别和行人检测 **2018**

算法实现/系统开发

- 基于树莓派 USB 摄像头实现人脸识别和行人检测。

会议 & 文章

Congfeng Cao, Junyong Fang, Dong Zhao. Development of UAV-borne multispectral camera based on narrow band width filter array. OPTICAL TECHNIQUE, 2018, 44(1).

荣誉 & 奖金

学业奖学金 **2016-2017**

学业奖学金 **2015-2016**

学业奖学金 **2014-2015**

优秀学生奖学金 **2013-2014**

技能

- 机器学习
 - 研究和工作中使用 Pytorch 两年 TensorFlow 六年
 - 熟悉人工智能理论和常用最新深度学习网络（如：ResNet, ResNeXt, Bert, Transformer, GPT-3, XLNet 等）。
 - 熟悉监督和非监督的机器学习方法，强化学习和贝叶斯方法理论。
 - 熟悉常见的自然语言处理和计算机视觉工具（如：Gensim, NLTK,

OpenCV 等)。

- 编程和工具

- 熟悉 python, 熟悉 python 运行机制 (7 年, 目前主要编程工具)。
- 曾经使用并且熟悉的编程语言有: C/C++, C#, R, Visual Basic, HTML, JavaScript 等
- 数据库四级四级工程师(常用数据库: MongoDB, Oracle, MySQL, Redis, Neo4J, PostgreSQL 等)
- 知识图谱工具(如: Apache Jena, OpenKE, Protege 等)。
- Elasticsearch, Spark 等相关工具。