



# 江苏省计算机学会通讯

## COMMUNICATIONS OF THE JSCS

2022年第2期

### 江苏省计算机学会常务理事单位

## 德讯科技股份有限公司

德讯科技股份有限公司（简称“德讯科技”）成立于2005年，是国内领先的数字化服务提供商，是数字化服务领域的先行者。德讯科技以大数据综合服务及应用开发、物联网、区块链、数据中心运营管理等四大板块业务为主体，积极布局数据智能业务，为客户提供贯穿IT建设全生命周期的“一站式”数字化服务。



德讯科技拥有业界顶尖的专业服务团队，与南京大学、中科院等建立深度合作，实现人才汇聚、协同创新、产学研用。迄今，德讯科技已申请获得80余项发明专利、120余项软件著作权，多项产品和技术获得国家各类荣誉奖项，已成功为金融、电信运营商、能源、政府、企事业单位等2000多家行业用户提供系统方案与服务。作为专业的数字化服务提供商，德讯科技勇于创新，不断超越，致力于用数字化服务赋能客户业务的转型和升级，在数字经济时代与客户共创价值。



- 后Hadoop时代大数据计算系统发展的新方向
- 机器学习中的标记增强理论与应用研究
- 多类云服务随机请求的异构资源排队性能分析与调度
- 江苏省第十五届信息安全高层论坛暨“问道网安·科技沙龙”在宁召开



封面 | 江苏省计算机学会副理事长单位

## 扬州大学信息工程学院

### 人工智能学院简介

扬州大学信息工程学院（人工智能学院）所属相关学科专业已有60多年办学历史，为国家培养了大批高素质的信息技术人才。

学院现设有计算机科学与技术、软件工程、电子信息工程、通信工程、自动化、人工智能等6个本科专业。其中，软件工程为国家级一流本科专业建设点，计算机科学与技术专业被列入教育部卓越工程师教育培养计划，为省级特色专业，电子信息工程专业被列入省卓越工程师培养计划。学院被江苏省教育厅、商务厅、财政厅等部门批准为江苏省国际服务外包人才培训基地、江苏省软件类专业卓越工程师培养计划试点单位、江苏省地方高校培养服务外包人才试点单位。《微机原理及应用》为首批国家级一流本科课程。

学院现有软件工程一级学科博士点，软件工程博士后流动站；信息计算科学、农业信息化技术、兽医生物信息学3个二级学科博士点；计算机科学与技术、软件工程、控制科学与工程、管理科学与工程4个一级学科硕士学位点；电子信息、交通运输、工程管理3个专业硕士学位授权点。计算机学科ESI排名进入全球大学和科研机构前1%；计算机科学与工程学科进入“2020软科世界一流学科排行榜”前200名。学院拥有江苏省知识管理与智能服务工程研究中心、省级示范教学实验中心-电工电子教学实验中心，还有计算机信息工程研究中心、汽车电子信息技术研究中心、农业大数据技术重点实验室等教学研究机构。

学院现有专任教师140人，其中教授23人，副教授50人，博、硕士生导师84人，具有博士学位91人。拥有中组部“千人计划”国家特聘专家1人，特聘教授3人，享受政府特殊津贴专家1人，全国优秀教师1人，全国师德先进个人1人，省突出贡献中青年专家2人，省高校教学名师1人，省“青蓝工程”中青年学术带头人2人，省“六大人才高峰”培养对象5人，省“333”工程培养对象4人。近年来，学院先后承担了国家863、973、国家自然科学基金、省部级科研项目以及各类横向科研项目；多次获得省部级科技奖励和教学成果奖励；每年公开发表科研论文200多篇，其中被SCI、EI、ISTP收录论文数居学校前列。

学院现有普通全日制本科生1600多人，博、硕士研究生500多人，成人学历教育学生600多人。学院坚持以生为本，不断深化教学改革，大力推进素质教育，形成了学生基础扎实、知识面宽、创新能力强的人才培养特色。近几年，学生在全国挑战杯、互联网+、创青春、电子设计大赛、大学生数学建模竞赛、智能汽车竞赛、大学生工业自动化挑战赛，以及ACM国际大学生程序设计竞赛等国家级、省级赛事中获奖400多项次，获奖学生600多人次；每年获得大学生学术科技创新基金项目50多项。学院学生除享受学校各类奖助学金外，社会各界人士还为学院学生设立了郭画奖学金、东华测试奖助学金、美嘉奖助学金等。近年来，本科毕业生考研录取率20%以上，全院毕业生就业率一直稳定在100%，人才培养质量受到用人单位的普遍好评。

学院注重国际合作与交流，先后与美国、德国、法国、比利时、日本、英国、澳大利亚、韩国等国外高校进行合作，互派学者访问进行合作研究、学术交流和学生联合培养等。



# 江苏省计算机学会 通讯

Communications of the JSCS



#### 顾问委员会

主任：周志华

副主任：武港山 耿新 胡江溢

李斌 夏士雄 李凡长

陈兵 詹永照 程光

委员：罗军舟 肖亮 申富饶

陶先平 吉根林 胡孔法

张道强 黄强 邓建明

金莹

#### 编委会

主编：路通

副主编：聂长海 张洁

编委：钱柱中 游辉敏 石克

地址：中国江苏省南京市栖霞区仙林大道163号

邮编：210023

电话：025-89680909

邮箱：jscs@nju.edu.cn

#### 专委会专栏

1 后Hadoop时代大数据计算系统发展的新方向 | 王肇康、顾荣

11 数字取证教育部工程研究中心

#### 学术交流

13 机器学习中的标记增强理论与应用研究 | 徐宁、耿新

17 多类云服务随机请求的异构资源排队性能分析与调度 | 王爽、李小平

#### 政策解读

23 江苏省制造业智能化改造和数字化转型三年行动计划解读

25 政策解掌握元宇宙“命门”，在相关前沿领域破冰领航

27 科技向善造福人类——解读《关于加强科技伦理治理的意见》

#### 会员风采

29 在探索图像子空间的科研道路上奋勇向前 | 李峻

32 “教机器”“见微知著” | 魏秀参

#### 科学普及

35 Web2.5已至，Web3.0还会远吗？

39 互联网的下一站，大概率是能源

#### 科创成果

44 X光安检图像智能识别系统研发及产业化应用 | 支洪平、金博伟、许琢、鲁盈悦  
王凯、王涛、郑伟伟、李玉笛、吴子扬、谢小环、李小兵

52 软件工程数据解析与智能释放 | 孙小兵、李斌、李必信、李云、杨辉、许聪颖

#### 学会动态

10 江苏省第十五届信息安全高层论坛暨“问道网安 科技沙龙”在宁召开

16 江苏省计算机学会两个调研课题被省科协立项公示

22 江苏省计算机学会职业院校产教融合专家委员会换届选举会议暨产教融合专题研讨会在线成功召开

24 江苏省计算机学会两个2022年度“科创江苏”专项行动计划立项项目

26 宿迁学院计算机科普教育基地与苏州大学共同打造国产图形化编程平台

34 我会五个学科竞赛项目获2022年全省普通高校本专科生学科竞赛省级赛事认定

38 热烈祝贺大陆杯2022年江苏省大学生计算机设计大赛圆满落幕

43 2022区块链技术论坛在线上成功举办

48 省计算机学会工控安全专委会走进中国（南京）软件谷信创示范区

50 JSCS老科学家委员会召开学习和工作研讨会议

# 后Hadoop时代大数据计算系统发展的新方向

王肇康<sup>1</sup> 顾荣<sup>2</sup>

<sup>1</sup> 南京航空航天大学计算机科学与技术学院/人工智能学院, 南京, 211106

<sup>2</sup> 软件新技术国家重点实验室(南京大学), 南京, 211106

摘要: 目前大数据计算的相关技术已经形成了一套综合的大数据技术栈。但随着数据量的持续积累, 大数据应用对性能的要求日益提高。基于Hadoop的经典大数据技术栈在计算性能与维护开销等方面, 已难以满足当下大数据处理的需要。为应对后Hadoop时代大数据应用的新性能需求, 大数据技术栈的资源层、存储层与计算层分别出现了新的技术与发展方向, 本文简要回顾了大数据计算系统近些年出现的新方向, 总结并展望了大数据技术发展的内生动力与发展方向。

关键词: 大数据 资源管理 并行计算 分布式数据处理

## 1 引言

大数据已经成为信息爆炸时代的一个标志。据国际数据公司(IDC)的观测, 现在全球数据总量以每2年翻一番的速度在迅速增长, 2020年产生的数据总量已突破64.2ZB<sup>[1]</sup>。随着数据的快速积累, 大数据所蕴含的价值已经成为学术界、企业界乃至各国政府的关注焦点<sup>[2]</sup>。美国白宫于2012年成立了“大数据高级指导小组”并发布了《大数据研究和发展计划》, 我国国务院于2015年发布的《促进大数据发展行动纲要》给出了大数据发展的国家顶层设计与总体部署。

大数据相关的计算技术已经形成了一套由资源层、存储层、计算层、分析算法层与应用层组成的大数据技术栈(如图1所示)。在整个大数据技术栈中, 资源层、存储层与计算层主要涉及大数据处理系统, 这些系统为上层的大数据分析算法与大数据应用提供了数据存储、处理与分析能力的支撑, 是本文关注的重点。

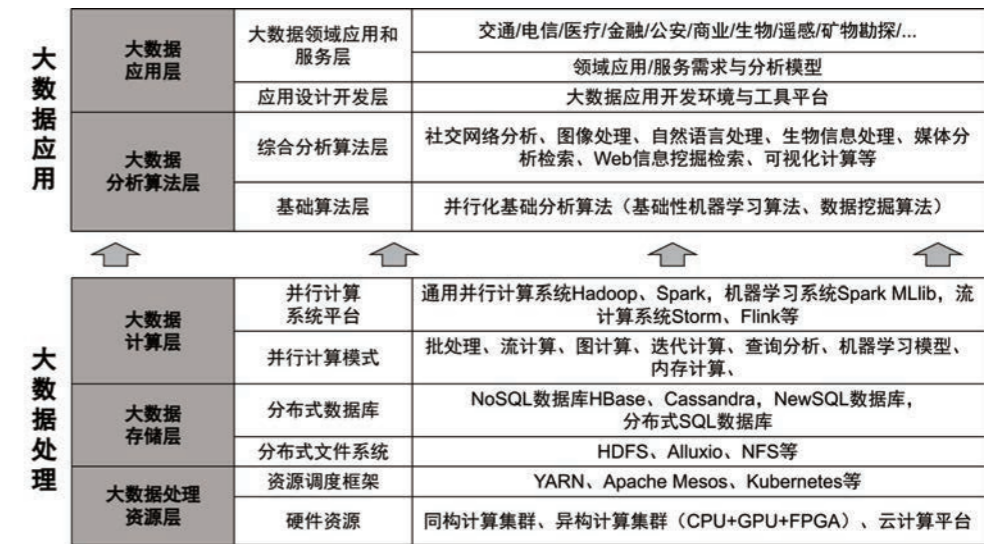


图1 大数据技术栈

随着Google MapReduce的提出与对应开源大数据处理系统Hadoop的出现, Hadoop已成为大数据处理技术的代名词。以Hadoop内核项目为中心, Hadoop项目推出的大数据处理系统(例如HDFS、Hadoop MapReduce、YARN等)以及基于Hadoop衍生的各类系统(例如Apache Giraph), 最终形成了整个如图2所示的Hadoop数据生态系统<sup>[3]</sup>。近些年来, 随着数据量的持续积累, 大数据应用对大数据处理系统的性能要求日益提高, 基于Hadoop的大数据技术栈的计算性能、运行成本与维护开销逐渐难以满足要求。



图2 Hadoop大数据生态系统及新技术方向

为应对后Hadoop时代大数据应用的新性能需求, 大数据技术栈的不同层次陆续出现了如图2所示的新技术方向。在大数据处理资源层, 以Kubernetes(K8s)为代表的新一代基于容器的弹性集群资源管理框架逐渐替代传统的YARN框架, 实现更灵活的资源管理并支持更广泛的处理系统。在大数据存储层, 以HDFS等分布式文件系统为基础, 数据湖进一步实现了对多源、多样、多类型文件数据的统一管理, 数据编排技术则通过统一大数据存储抽象, 实现了大数据计算系统与大数据存储实现的解耦分离; 分布式数据库OLTP与OLAP界限正在逐渐模糊, 越来越多的分布式数据库



系统开始提供混合事务分析处理的能力。在大数据计算层，流批一体的大数据计算系统逐渐成为通用大数据处理的主流系统，其以同一套计算引擎同时支撑流式与批式大数据处理任务；而随着深度学习的成熟与广泛应用，分布式深度学习系统已成为下面面向特定领域的大数据计算系统研究的焦点，引起了学术界与工业界的广泛关注。面对日新月异的大数据技术栈，本文将对上述大数据技术发展新方向进行简要介绍与回顾，总结并展望了大数据技术发展的内生动力与未来方向。

## 2 大数据资源层：基于容器的弹性集群资源管理调度

Linux容器（Container）<sup>[4]</sup>是一种操作系统级的虚拟化机制。一个容器包含了核心服务程序及其依赖的操作系统环境，容器可以在不同节点之间迁移而无需担心软件依赖问题，解决了现有分布式程序部署中存在的软件依赖环境配置难、维护开销高的问题。容器使软件开发者在开发分布式程序时不再受部署集群环境的具体软件版本限制，而保持程序部署环境与程序开发环境一致，降低集群软件环境的配置难度与维护开销。同时，基于操作系统虚拟化技术的容器，在保持程序间执行环境隔离的基础上，相比基于Hypervisor的全栈虚拟化技术的管理开销与执行开销显著降低。这使得容器可以实现秒级的启动与停止，相比传统的全栈虚拟化的启停时间降低一个数量级，为实现分布式环境下秒级弹性缩扩容提供了技术基础。YARN中也提出了一种名为容器的资源调度的单元，其提供了一种包装了CPU、内存、GPU等资源在内的任务调度单元，而操作系统中的容器则可视为一种包含了程序及其依赖的软件环境封装。YARN的容器并不具体解决程序的软件环境及其依赖的问题，需要程序员利用YARN的编程接口与机制实现软件打包与依赖分发等。而操作系统中的容器因为包含了软件依赖环境，可直接分发与部署。

在分布式环境下基于容器实现分布式计算需要由专门的容器编排（Container Orchestration）框架<sup>[5]</sup>进行调度与分配。容器编排框架（例如Kubernetes<sup>[6]</sup>）可以看做一类集群资源调度框架。相比于经典的集群资源调度管理框架（例如YARN、Mesos），容器编排框架以Linux容器为核心，以容器为单位进行资源分配、调度、监控与管理。容器编排框架负责根据用户提交的分布式计算任务请求在集群环境中部署、运行与维护容器。容器编排框架通常应提供基础的集群资源调度功能，提供资源限制控制、容器调度、负载均衡、健康监测与故障容错等基础能力。

容器编排框架借助容器技术所具有的秒级启动与停止能力，可以细粒度的实现高灵敏的自动缩扩容能力。容器编排框架可以更细的粒度响应计算资源请求量的波动情况，并根据当前集群资源负载情况（例如CPU、内存的使用率）或根据应用自身的请求，在集群中增加与减少相应的容器，以应对短时间内突发而来的计算请求或及时释放空闲的计算资源，减少无效的资源占用，从而降低用户的运行成本。现有常用的容器编排框架包括Kubernetes、Docker Swarm等；基于公有云云环境的容器编排框架则有Google Container Engine、Amazon Elastic Container Service、阿里云的弹性容器实例等。

## 3 大数据存储层

### 3.1 大数据文件存储：数据湖与数据编排

大数据的多样性一直是大数据文件存储所面临的关键挑战之一。大数据的多样性可体现在两方面：数据类型的多样性与数据存储后端的多样性。针对这两类多样性问题，数据湖技术通过多类型数据的元信息挖掘与统一管理，降低大数据应用访问多类型数据的难度，数据编排技术则通过引入统一数据访问抽象，实现大数据存储层与计算层的解耦，提升计算存储分离环境下数据访问效率。

#### （1）数据湖

在大型企业中，大量数据集散落在不同的信息系统、存储系统中，数据集的元信息可能存在缺失与错误，数据集之间的关联缺失，难以跨数据集挖掘有价值信息。数据湖（Data Lake）即为了解决大型企业中海量多样异构数据集的统一管理与共享问题而提出。数据湖是由大量来源不同、格式不同、存储于不同存储系统的数据集构成的数据集集合<sup>[7]</sup>。数据湖收集与存储来自不同部门、不同存储系统的大量原始数据集（数据从各种业务系统中“流入”数据湖），从原始数据集中提取元信息，并基于元信息发掘数据集之间的关联，从而将分散在多个系统的数据集整理为统一的数据仓库（Repository），用户可以通过数据湖系统探索与使用企业中积累的大量数据资源（数据从数据湖“流出”到各类分析应用中），挖掘大数据中蕴藏的价值。

数据湖管理系统的典型架构如图3所示。与数据仓库相比，数据集在进入数据湖时会提取数据集本身的元信息（包括数据模式等）以实现数据治理。与数据仓库不同的是，数据湖中的数据集以原始数据的形态存在，而不需要根据预定义的数据模式进行抽取与转换。用户根据具体的数据分析需求从原始数据集中提取关键数据。数据湖通常也包括多种数据治理工具，包括数据结构提取工具、跨数据集元信息匹配工具、数据质量管理工具等。数据湖通过数据治理工具产生的元信息，为各类数据集构建出统一的元数据索引与数据目录（Catalog），降低数据探索与利用的难度，使数据湖中积累的数据可以被有效利用，实现数据从数据湖中的有价值流出。

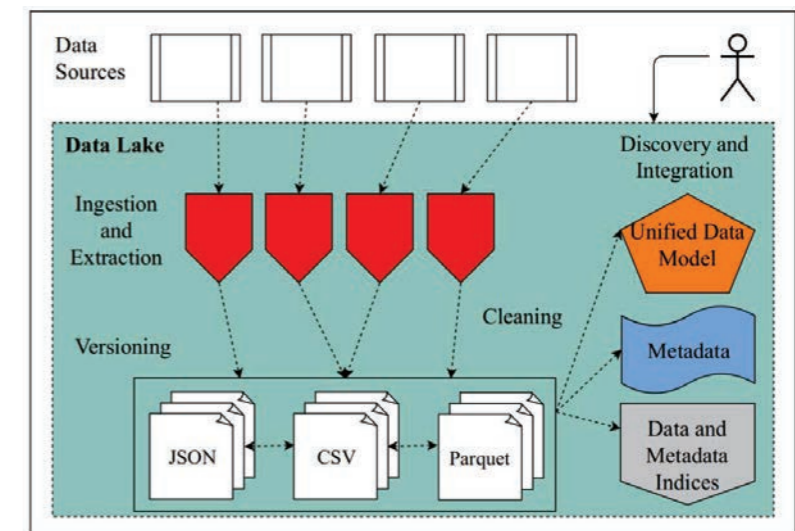


图3 数据湖管理系统的典型架构<sup>[8]</sup>

## (2) 数据编排

数据编排 (Data Orchestration) [9] 平台扮演了大数据处理系统与大数据存储系统之间的桥梁。数据编排平台对不同大数据存储系统提供的数据库访问接口进行抽象与封装, 从而向上层大数据处理系统提供标准化的I/O接口。大数据处理系统可通过数据编排平台透明地访问多种大数据存储系统。数据编排平台以一种统一的路径规则, 将不同存储系统内的数据映射至一套全局的虚拟文件命名空间, 进一步屏蔽不同存储系统之间的差异。为了加速数据访问速度, 数据编排平台可透明地将热点数据缓存至大数据处理系统的本地服务器中, 从而获得内存级访问加速。借助数据编排平台, 大数据处理系统 (计算) 与大数据存储系统 (存储) 通过统一的数据I/O接口实现分离解耦, 使两者可独立部署与扩展, 具备了向云原生等新兴云计算环境迁移的技术基础。

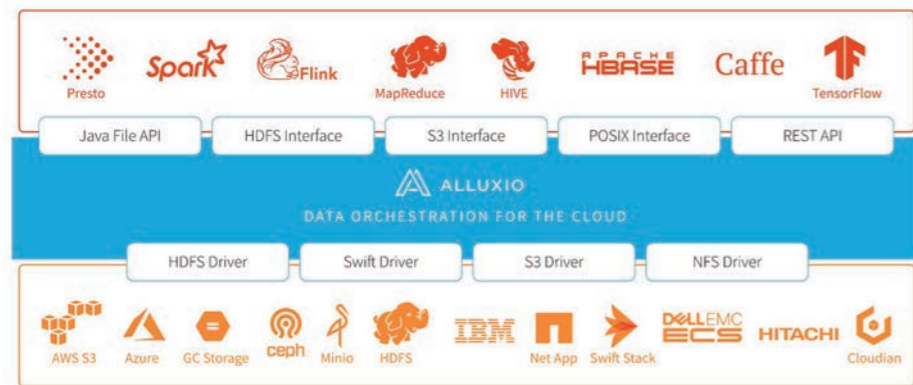


图4 基于数据编排平台Alluxio实现存储与计算解耦 [9]

但现有数据编排平台在迁移至基于容器的云原生环境时, 面临着难以感知云上架构变化的问题。为了进一步获取云原生环境下数据节点以及

上层应用的动态变化信息, 支持数据编排平台中的静态缓存等中间件动态、灵活地调动起来, Fluid项目[10]面向云原生环境提出了抽象数据集的概念, 将数据集的编排转为对承载数据的计算/存储引擎进行编排和管理, 通过对数据引擎进行合理的扩容、缩容和调度操作, 从而实现对数据的迁移与缓存以及数据在容器调度管理框架 (例如Kubernetes) 下灵活调度的管理和变化。

### 3.2 大数据查询处理: 混合事务分析处理

实际应用中部署的大数据应用通常同时需要具备联机事务处理 (OLTP) 与联机分析处理 (OLAP) 需求。传统大数据应用架构中OLTP部分通常由关系数据库系统负责, OLAP部分由数据仓库或大数据处理系统负责, 存储在关系数据库中的事务数据定期 (例如按天或按周) 导入到OLAP系统中进行处理。但随着大数据应用的性能要求与实时性要求逐渐提高, 传统的部署方式逐渐展现出分析实时性差、维护开销高的问题。受限于两类系统之间数据同步的时间差, OLAP系统中的数据更新存在滞后现象, 难以满足实时性强的数据分析需求 (例如平台运营监控等)。维护两套系统以及确保两套系统之间的正确数据同步则增加了额外的维护成本。

面对OLTP与OLAP共存的大数据分析需求, Gartner提出了混合事务分析处理 (Hybrid Transactional/Analytical Processing, HTAP) 的技术设想[11], 近些年逐渐在PolarDB[12]、TiDB[13]等分布式数据库中实现。混合事务分析处理的目标是在一套系统中基于同一套数据源同时支持联机事务处理与联机分析处理, 并且能够基于实时或近实时的数据得到新鲜 (Fresh) 的分析处理结果。根据底层数据存储的组织方式, HTAP系统的常见技术路线可分为统一存储与分离存储两类[14]。采用统一存储技术路线的分布式数据库 (例如PolarDB) 为OLTP和OLAP分别设计了不同的查询引擎, 但两类引擎底层均采用统一数据存储引擎, OLTP引擎可以对数据进行修改而OLAP引擎则仅有只读权限, 从而避免双引擎造成的数据竞争问题, 保证OLTP的事务正确性。采用分离存储技术路线的分布式数据库 (例如MemSQL) 为OLAP和OLTP分别独立设计了查询引擎和存储引擎以适应两类需求。

## 4 大数据计算层

### 4.1 通用大数据计算: 流批一体的大数据处理系统

在经典的大数据处理架构下, 来自原始数据源的数据被同时推送至流处理系统 (例如Apache Storm或Apache Flink) 与批处理系统 (例如Hadoop MapReduce或Apache Spark) 分别处理, 整个大数据技术栈同时维护流处理与批处理两条数据处理链路, 以分别满足在线处理与离线处理的性能需求。这种流批分离的系统架构可能会带来冗余的维护开销。在部分企业业务场景中, 对于同一关键业务指标需要同时获取其在线实时情况与离线统计情况, 这导致批处理与流处理的计算任务存在大量重叠。在流批分离的系统架构下, 需要将同一套业务逻辑分别在两类系统中实现, 并维持两套系统之间的一致性, 维护开销高昂。

针对流批分离架构带来的冗余维护问题, Google提出了流批一体的大数据处理模型Dataflow[15], 利用数据流编程模型, 将算法逻辑描述与大数据处理系统解耦, 用户编写的程序可以同时流处理系统与批处理系统中运行。根据底层计算引擎的工作方式划分, 流批一体的大数据处理系统的技术路线可分为分离引擎与统一引擎两种。在分离引擎的技术路线下, 流处理和批处理采用不同的计算引擎, 但上层向用户提供统一的基于数据流的编程接口, 由大数据处理系统根据用户指定的计算时延上限、计算成本限制与计算精度要求确定具体的执行引擎。此类技术路线的代表性系统是Google提出的Apache Beam项目[16], 该项目以Dataflow编程模型为统一的编程接口, 底层接入Spark、Flink、Google Cloud Dataflow等作为具体的计算引擎。基于统一引擎的技术路线则采用统一的计算引擎同时支撑流处理与批处理, 代表性系统包括Apache Spark Streaming和Apache Flink。Spark Streaming将数据流视为一系列的微批 (micro batch) [17], 利用Spark自身的批处理计算引擎实现流计算, 该方式能充分复用Spark计算引擎已有的优化技术实现高吞吐性能, 但批处理引擎无法确保实时性要求。Flink[18]则与Spark相反, 其将批处理视为一种输入长度有限的流处理过程, 利用Flink的统一DAG调度、流批一体Shuffle、流批一体容错等机制, 以一套执行引擎同时支撑高效的流处理与批处理。



#### 4.2 面向特定领域的大数据计算：分布式深度学习系统

随着深度学习技术的发展，基于海量数据训练得到的大规模深度学习模型（例如GPT系列、文心ERNIE系列、悟道系列等）逐渐展现出了优异预测性能。在TB级规模数据集上训练参数规模达千亿的深度学习模型面临着数据规模与模型规模的双重挑战，因此需要分布式深度学习训练系统提供有效支撑。目前主流的深度学习框架训练系统（例如TensorFlow、PyTorch、PaddlePaddle等）均支持了分布式深度学习模型训练<sup>[19]</sup>。

实现高效的分布式深度学习模型训练是一项系统工程，涉及并行计算模式、模型参数同步框架与通信优化等多种技术，需针对不同的训练场景选择相应的技术组合。可供选择的并行计算模式包括数据并行、模型并行、数据与模型结合的流水线并行与混合并行等多种方式。为协调多机并行训练产生的模型更新，模型参数同步框架可采用基于参数服务器（Parameter Server）的集中式架构或基于allreduce操作的分布式架构，以实现分布式参数管理；可选择同步更新、有界异步更新、全异步更新等同步机制，以确定模型参数在多机之间的同步与更新时机。在通讯优化方面，分布式深度学习系统可采用降低模型精度、压缩模型更新、更新优先调度等技术降低通讯开销，以尽量少的通讯轮数加速模型收敛。

## 5 总结与展望

随着大数据应用的日益落地普及，其对底层大数据处理系统的性能和功能方面的要求也逐步变高，传统的大数据系统技术栈不断面临新的瓶颈挑战。为了解决这些问题，近年来整个大数据处理技术也取得了长足的进步。

在提升大数据处理的计算性能方面，大数据资源层出现了以Linux容器为核心的容器编排框架，降低了大数据计算系统的部署难度，并以秒级启停能力实现了。大数据存储层的数据编排框架则通过数据存储抽象层，实现了大数据处理系统与大数据存储系统的解耦分离，使两者可独立按需扩展，并利用缓存与调度实现了更高效的大数据I/O。大数据计算层则面向超大规模数据集上的分布式深度学习模型训练，研究开发了专用的分布式深度学习系统，并提出了针对性的并行计算模式、模型参数同步与通信方法。

在降低大数据处理的维护开销方面，大数据存储层的数据湖管理系统实现了多源多类型数据集的统一管理。基于混合事务分析处理的分布式数据库以同一套数据同时支撑OLTP与OLAP类需求，避免了维护两套数据库系统以及数据一致性问题。在大数据计算层流批一体的大数据处理系统以同一套编程接口与编程模型同时支撑在线流处理与离线批处理请求，避免了相同计算逻辑的重复编写与维护，提升了数据分析结果的一致性。

纵观大数据处理技术栈的新技术方向与发展趋势，后Hadoop时代大数据计算系统的发展主要由更高的计算性能与更低的维护开销两大需求驱动。一方面，大数据应用产生和处理的数据规模日益增长，需要大数据计算系统能以更快的速度、更低的成本处理海量数据。另一方面，大数据应用的数据来源与计算逻辑日趋复杂，但对计算结果的准确性与实时性要求则日渐提高，需要大数据计算系统能以更低的维护开销、

更低的学习成本，支撑多样化数据的在线与离线处理需求。

未来，预期大数据系统技术会继续沿着更高的计算性能与更低的维护成本路线深入发展。具体技术路线上，与服务器无感知计算（Serverless）等新型云计算模式、代码生成等查询引擎优化技术、基于人工智能的系统优化（AI for System）等系统技术进行更深入的结合使用，从而达成更高效、更好用的目标。

## 参考文献

- [1] IDC. Data Creation and Replication Will Grow at a Faster Rate than Installed Storage Capacity, According to the IDC Global DataSphere and StorageSphere Forecasts[EB/OL].(2021-03-24)[2022-05-17].
- [2] 程学旗, 靳小龙, 杨婧, 等. 大数据技术进展与发展趋势[J]. 科技导报, 2016, 34(14): 49-59. DOI:10.3981/j.issn.1000-7857.2016.14.006.
- [3] 黄宜华, 苗凯翔. 深入理解大数据：大数据处理与编程实践[M]. 机械工业出版社, 2014.
- [4] RED HAT INC. Understanding Linux containers[EB/OL]. (2019-12-10)[2022-05-23]. <https://www.redhat.com/en/topics/containers>.
- [5] CASALICCHIO E. Container Orchestration: A Survey[M]//PULIAFITO A, TRIVEDI K S. Systems Modeling: Methodologies and Tools. Cham: Springer International Publishing, 2019: 221-235. DOI:10.1007/978-3-319-92378-9\_14.
- [6] Kubernetes: Production-Grade Container Orchestration[EB/OL]. [2022-05-23].
- [7] HAI R, GEISLER S, QUIX C. Constance: An Intelligent Data Lake System[C]//Proceedings of the 2016 International Conference on Management of Data. San Francisco, California, USA: ACM, 2016: 2097-2100. DOI:10.1145/2882903.2899389.
- [8] NARGESIAN F, ZHU E, MILLER R J, 等. Data Lake Management: Challenges and Opportunities[J]. Proceedings of the VLDB Endowment, 2019, 12(12): 1986-1989. DOI:10.14778/3352063.3352116.
- [9] LI H. Data Orchestration: The Missing Piece in the Data

### 参考文献

World[EB/OL]. (2019-05-06)[2022-05-17].  
<https://www.alluxio.io/blog/data-orchestration-the-missing-piece-in-the-data-world/>.

[10] Fluid, elastic data abstraction and acceleration for BigData/AI applications in cloud[EB/OL]. [2022-05-17].  
<https://fluid-cloudnative.github.io/>.

[11] PEZZINI M, FEINBERG D, RAYNER N, 等. Hybrid Transaction/Analytical Processing Will Foster Opportunities for Dramatic Business Innovation[R/OL]. Gartner, 2014.  
<https://www.gartner.com/en/documents/2657815>.

[12] CAO W, ZHANG Y, YANG X, 等. PolarDB Serverless: A Cloud Native Database for Disaggregated Data Centers[C]//Proceedings of the 2021 International Conference on Management of Data. Virtual Event China: ACM, 2021: 2477-2489. DOI:10.1145/3448016.3457560.

[13] HUANG D, LIU Q, CUI Q, 等. TiDB: a Raft-based HTAP database[J]. Proceedings of the VLDB Endowment, 2020, 13(12): 3072-3084. DOI:10.14778/3415478.3415535.

[14] ZCAN F, TIAN Y, T Z N P. Hybrid Transactional/Analytical Processing: A Survey [C]//Proceedings of the 2017 ACM International Conference on Management of Data. New York, NY, USA: ACM, 2017: 1771-1775. DOI:10.1145/3035918.3054784.

[15] AKIDAU T, BRADSHAW R, CHAMBERS C, 等. The dataflow model: A practical approach to balancing correctness, latency, and cost in massive-scale, unbounded, out-of-Order data processing[J]. Proceedings of the VLDB Endowment, 2015, 8(12): 1792-1803. DOI:10.14778/2824032.2824076.

[16] APACHE SOFTWARE FOUNDATION. Apache Beam[EB/OL]. [2022-5-23].

[17] ZAHARIA M, DAS T, LI H, 等. Discretized streams: Fault-tolerant streaming computation at scale[C]//Proceedings of the twenty-fourth ACM symposium on operating systems principles. New York, NY, USA: ACM, 2013: 423-438. DOI:10.1145/2517349.2522737.

[18] CARBONE P, KATSIFODIMOS A, EWEN S, 等. Apache Flink: Stream and Batch Processing in a Single Engine[J]. Bulletin of the IEEE

Computer Society Technical Committee on Data Engineering, 2015, 36(4): 28-38.

[19] MAYER R, JACOBSEN H A. Scalable deep learning on distributed infrastructures: Challenges, techniques, and tools[J]. ACM Computing Surveys, 2020, 53(1): 3:1-3:37. DOI:10.1145/3363554.

## 学会动态

### 江苏省第十五届信息安全高层论坛暨“问道网安·科技沙龙”在宁召开

2022年5月29日作为2022年江苏省第34届科普宣传周重点活动之一的“江苏省第十五届信息安全高层论坛暨问道网安·科技沙龙”采用线上和线下相结合的方式成功召开。江苏省计算机学会秘书长金莹教授出席开幕式并致辞。论坛分别由江苏省科学技术协会副主席、俄罗斯自然科学院外籍院士、南京理工大学李千目教授，东南大学陈立全教授两位专家主持。

本届论坛特邀国家重点研发计划项目首席科学家、华英青年学者、东南大学网络空间安全学院副院长陈立全教授，教育部青年长江学者、江苏省杰出青年基金获得者、南京信息工程大学计算机学院副院长付章杰教授，江苏省中青年科技领军人才、“辽宁省百千万人才工程”百人层次、南京邮电大学计算机学院王化群教授，江苏省通信管理局网络安全管理处处长马旻正高级工程师，博智安全科技股份有限公司助理总裁王力高级工程师等省内信息安全专家在论坛上分别围绕网络空间安全技术前沿发展分析、数字取证技术与挑战、6G内生安全中的密码技术、工业互联网发展与安全等方面做专题报告和主题讨论。

“问道网安·科技沙龙”围绕“增强网络安全意识保护信息数据安全”主题，采用嘉宾同台、智慧碰撞、对话交流、互动提问的方式进行广泛交流和讨论，线上和线下的提问交流热烈，来自省内外100余所高校的15000余名专家学者分别在腾讯会议和VTC云直播参加了本届论坛，场内外气氛交融，学术气息浓厚。在各方支持和共同努力下，本届论坛取得圆满成功。





# 数字取证教育部工程研究中心

数字取证教育部工程研究中心隶属于南京信息工程大学，是2019年经国家教育部正式批准立项建设（教技函〔2019〕72号）的首家数字取证国家级平台，主要围绕数字水印与版权保护、数字取证、物联网与区块链、人工智能与安全等领域开展科学研究与产学研合作。中心旨在打造国际一流的数字取证“政产学研用”示范基地，构建我国数字取证领域的高素质人才培养与聚集基地、高水平成果产出与转化基地，全面提升我国数字取证相关行业的创新能力和国际竞争力。

取证中心现有专职研发师资40余人，包括教授15人，副教授12人。另有硕博研究生300余人，兼职研发工程师20余人。拥有全国高校黄大年式教师团队、中国气象局特色团队、江苏省高校优秀科技创新团队、江苏省双创计划团队、江苏省“六大人才高峰”等五个省部级及以上创新团队。拥有海外院士1人，国家杰出青年基金获得者1人，教育部“长江学者奖励计划”特聘教授2人，国家优秀青年基金获得者2人。

取证中心目前承担国家级科研项目30多项，省部级项目60余项，在各类高水平期刊发表学术论文500余篇，授权国家发明专利150余项。相关成果荣获江苏省科学技术一等奖、教育部自然科学一等奖等省部级奖10余项。数字取证、数字水印、版权保护、区块链、人工智能等相关研究和开发成果广泛应用于军队、国家安全部门、中国气象系统、能源系统等单位，与江苏警官学院、360公司、蚂蚁金服、中科院软件所、中科实数等国内外100多家企事业单位建立合作关系，近年来承担横向项目100余项，到账科研经费超过2亿元。

好风正劲待扬帆，砥砺奋进正当时。取证中心立足网络安全的大好风口，凝神聚气做一流科研，稳扎稳打研发拳头产品。通过搭建一流的政产学研用一体化平台，制定标准，建立数字取证产业联盟，为我国的数字取证事业培养复合型高端人才，建立成果转化基地，推动数字取证产业飞跃发展，为保障国家知识产权安全 and 产业发展提供技术支撑。



**专利证书:** 取证中心近年获授权国家发明专利150项，主要包括数字水印、数字取证、人工智能安全、物联网与区块链等方向，涉及文本、音频、视频、大数据等各种融合。

序号	标题	专利号	类型	发明人	专利号
1	一种基于区块链的版权保护方法	ZL201810242803.0	发明专利	王德胜, 王德胜, 王德胜	ZL201810242803.0
2	一种基于区块链的版权保护方法	ZL201810242803.0	发明专利	王德胜, 王德胜, 王德胜	ZL201810242803.0
3	一种基于区块链的版权保护方法	ZL201810242803.0	发明专利	王德胜, 王德胜, 王德胜	ZL201810242803.0
4	一种基于区块链的版权保护方法	ZL201810242803.0	发明专利	王德胜, 王德胜, 王德胜	ZL201810242803.0
5	一种基于区块链的版权保护方法	ZL201810242803.0	发明专利	王德胜, 王德胜, 王德胜	ZL201810242803.0
6	一种基于区块链的版权保护方法	ZL201810242803.0	发明专利	王德胜, 王德胜, 王德胜	ZL201810242803.0
7	一种基于区块链的版权保护方法	ZL201810242803.0	发明专利	王德胜, 王德胜, 王德胜	ZL201810242803.0
8	一种基于区块链的版权保护方法	ZL201810242803.0	发明专利	王德胜, 王德胜, 王德胜	ZL201810242803.0
9	一种基于区块链的版权保护方法	ZL201810242803.0	发明专利	王德胜, 王德胜, 王德胜	ZL201810242803.0
10	一种基于区块链的版权保护方法	ZL201810242803.0	发明专利	王德胜, 王德胜, 王德胜	ZL201810242803.0
11	一种基于区块链的版权保护方法	ZL201810242803.0	发明专利	王德胜, 王德胜, 王德胜	ZL201810242803.0
12	一种基于区块链的版权保护方法	ZL201810242803.0	发明专利	王德胜, 王德胜, 王德胜	ZL201810242803.0
13	一种基于区块链的版权保护方法	ZL201810242803.0	发明专利	王德胜, 王德胜, 王德胜	ZL201810242803.0
14	一种基于区块链的版权保护方法	ZL201810242803.0	发明专利	王德胜, 王德胜, 王德胜	ZL201810242803.0
15	一种基于区块链的版权保护方法	ZL201810242803.0	发明专利	王德胜, 王德胜, 王德胜	ZL201810242803.0
16	一种基于区块链的版权保护方法	ZL201810242803.0	发明专利	王德胜, 王德胜, 王德胜	ZL201810242803.0
17	一种基于区块链的版权保护方法	ZL201810242803.0	发明专利	王德胜, 王德胜, 王德胜	ZL201810242803.0
18	一种基于区块链的版权保护方法	ZL201810242803.0	发明专利	王德胜, 王德胜, 王德胜	ZL201810242803.0
19	一种基于区块链的版权保护方法	ZL201810242803.0	发明专利	王德胜, 王德胜, 王德胜	ZL201810242803.0
20	一种基于区块链的版权保护方法	ZL201810242803.0	发明专利	王德胜, 王德胜, 王德胜	ZL201810242803.0
21	一种基于区块链的版权保护方法	ZL201810242803.0	发明专利	王德胜, 王德胜, 王德胜	ZL201810242803.0
22	一种基于区块链的版权保护方法	ZL201810242803.0	发明专利	王德胜, 王德胜, 王德胜	ZL201810242803.0
23	一种基于区块链的版权保护方法	ZL201810242803.0	发明专利	王德胜, 王德胜, 王德胜	ZL201810242803.0
24	一种基于区块链的版权保护方法	ZL201810242803.0	发明专利	王德胜, 王德胜, 王德胜	ZL201810242803.0
25	一种基于区块链的版权保护方法	ZL201810242803.0	发明专利	王德胜, 王德胜, 王德胜	ZL201810242803.0
26	一种基于区块链的版权保护方法	ZL201810242803.0	发明专利	王德胜, 王德胜, 王德胜	ZL201810242803.0
27	一种基于区块链的版权保护方法	ZL201810242803.0	发明专利	王德胜, 王德胜, 王德胜	ZL201810242803.0
28	一种基于区块链的版权保护方法	ZL201810242803.0	发明专利	王德胜, 王德胜, 王德胜	ZL201810242803.0
29	一种基于区块链的版权保护方法	ZL201810242803.0	发明专利	王德胜, 王德胜, 王德胜	ZL201810242803.0
30	一种基于区块链的版权保护方法	ZL201810242803.0	发明专利	王德胜, 王德胜, 王德胜	ZL201810242803.0
31	一种基于区块链的版权保护方法	ZL201810242803.0	发明专利	王德胜, 王德胜, 王德胜	ZL201810242803.0
32	一种基于区块链的版权保护方法	ZL201810242803.0	发明专利	王德胜, 王德胜, 王德胜	ZL201810242803.0
33	一种基于区块链的版权保护方法	ZL201810242803.0	发明专利	王德胜, 王德胜, 王德胜	ZL201810242803.0
34	一种基于区块链的版权保护方法	ZL201810242803.0	发明专利	王德胜, 王德胜, 王德胜	ZL201810242803.0
35	一种基于区块链的版权保护方法	ZL201810242803.0	发明专利	王德胜, 王德胜, 王德胜	ZL201810242803.0
36	一种基于区块链的版权保护方法	ZL201810242803.0	发明专利	王德胜, 王德胜, 王德胜	ZL201810242803.0
37	一种基于区块链的版权保护方法	ZL201810242803.0	发明专利	王德胜, 王德胜, 王德胜	ZL201810242803.0
38	一种基于区块链的版权保护方法	ZL201810242803.0	发明专利	王德胜, 王德胜, 王德胜	ZL201810242803.0
39	一种基于区块链的版权保护方法	ZL201810242803.0	发明专利	王德胜, 王德胜, 王德胜	ZL201810242803.0
40	一种基于区块链的版权保护方法	ZL201810242803.0	发明专利	王德胜, 王德胜, 王德胜	ZL201810242803.0



# 机器学习中的标记增强理论与应用研究

—2021年江苏省计算机学会优秀博士论文

推荐单位：东南大学计算机科学与工程学院PALM实验室  
论文作者：徐宁  
指导老师：耿新

## 论文摘要

标记端多义性是当今机器学习的热点问题。多标记学习中，每个样本都被赋予一组标记子集来表示其多种语义信息。然而，标记强度差异现象在多义性机器学习任务中广泛存在，而既有多标记学习研究中普遍采用的相关/无关两个子集的逻辑划分法几乎完全忽视了这种现象，造成学习过程中不可避免的信息损失。针对这一突出问题，有必要用一种称为标记分布的标注结构来代替逻辑标记对示例的类别信息进行描述。标记分布通过连续的描述度来显式表达每个标记与数据对象的关联强度，很自然地解决了标记强度差异的问题，而在以标记分布标注的数据集上学习的过程就称为标记分布学习。由于描述度的标注成本更高且常常没有客观的量化标准，现实任务中大量的多义性数据仍然是以简单逻辑标记标注的，为此本文提出了标记增强这一概念。标记增强在不增加额外数据标注负担的前提下，挖掘训练样本中蕴含的标记重要性差异信息，将逻辑标记转化为标记分布。本文对标记增强进行研究，主要工作包括：

构建标记增强基础理论框架。该理论框架回答了以下三个问题：第一，标记增强所需的类别信息从何而来？即标记分布的内在生成机制；第二，标记增强的结果如何评价？即标记增强所得标记分布的质量评价机制；第三，标记增强为何有效？即标记增强对后续分类器的泛化性能提升机制。理论分析和实验结果验证了标记增强的有效性。

提出一种面向标记分布学习的标记增强专用算法。以面向标记分布学习的标记增强为目标专门设计的算法十分重要，其关键是如何设计能够充分挖掘数据中隐藏的标记信息的优化目标函数。因此，本文提出一种面向标记分布学习的标记增强方法GLLE。该方法利用训练样本特征空间的拓扑结构以及标记间相关性，挖掘了标记强度信息，从而生成了标

记分布。实验结果验证了GLLE对逻辑标记数据集进行标记增强处理后使用标记分布学习的有效性。

标记增强在其他学习范式上的应用。本文提出了基于标记增强的多标记学习方法LEMLL，该方法将标记增强与多标记预测模型统一到同一学习目标中，使得预测模型可以在更为丰富的监督信息下进行训练，有效地提升了学习效果。本文提出了基于标记增强的偏标记学习方法PLLE，该方法利用标记增强恢复候选标记的描述度，使得后续的学习问题转化为多输出回归问题。在多标记数据集和偏标记数据集上的实验结果显示，相较于对比算法，基于标记增强方法取得了显著更优的表现。

## 论文看点

1. 构建了标记增强基础理论框架。该理论框架研究了以下三个机制：标记分布的内在生成机制、标记增强所得标记分布的质量评价机制、标记增强后学习系统的泛化性能提升机制。
2. 设计了面向标记分布学习的标记增强专用算法。这些方法利用训练样本特征空间的拓扑结构或标记间相关性，挖掘了标记强度信息，从而生成了标记分布；
3. 将标记增强应用于既有学习范式，如多标记学习、偏标记学习等，为解决既有学习问题提供了新思路。

## 作者简介



作者：徐宁，东南大学助理研究员，主要研究方向为机器学习和数据挖掘。  
E-mail: xuning@seu.edu.cn



指导教师：耿新，东南大学教授，主要研究方向为机器学习、模式识别和计算机视觉。  
E-mail: xgeng@seu.edu.cn



//// 学会动态 //////////////////////////////////////

团队简介



PALM (PAttern Learning and Mining) Lab隶属东南大学计算机科学与工程学院，位于东南大学九龙湖校区计算机楼。实验室致力于人工智能、机器学习和模式识别等方面的研究。近年来，实验室导师主持多项国家重点研发计划课题、国家自然科学基金项目和江苏省自然科学基金项目，并获得了国家自然科学基金二等奖、国家级教学成果奖一等奖、教育部自然科学奖一等奖等多项教学、科研奖励。实验室在机器学习理论、计算机视觉、自然语言处理方向的国际顶级会议及期刊（如 NeurIPS、ICML、AAAI、IJCAI、KDD、ACL、EMNLP、CVPR、ECCV、IEEE Transactions on Pattern Analysis and Machine Intelligence、IEEE Transactions on Knowledge and Data Engineering、ACM MM等）发表过多篇文章，部分成果在腾讯、VMware等中外大型企业落地。实验室学生获得中国计算机学会（CCF）优秀博士学位论文、江苏省优秀博士学位论文奖、江苏省计算机学会优秀博士学位论文等。

团队主页：<http://palm.seu.edu.cn/>

江苏省计算机学会两个调研课题被省科协立项公示

5月11日，江苏省科学技术协会对“2022年度江苏省科协调研课题立项资助项目”进行了公示，我会申报的《江苏增强源头供给培育国家战略科技力量的策略研究》、《公共卫生事件应急管理科技支撑体系建设研究》两个课题均在2022年度省科协调研课题拟立项资助项目公示名单中。后续工作中，为充分发挥科技创新在创新发展中的引领作用，我会将进一步提高科技智库建设，团结和服务广大科技工作者履行“争当表率、争做示范、走在前列”光荣使命，为谱写“强富美高”新江苏现代化建设新篇章积极建言献策。

关于2022年度江苏省科协拟立项调研课题的公示

2022-05-11

根据《关于申报2022年度省科协调研课题的通知》（苏科协发〔2022〕23号）要求，经过网上申报、资格审查、选题评审等环节并报请省科协党组会议审议通过，确定下列课题为2022年度江苏省科协调研课题立项资助项目（详见附件），现予以公示，公示时间为2022年5月12日—18日，如有异议请在公示期内书面实名向省科协调宣部反映。

联系人：侯 飞

电话：025—83625043，18915969326

邮箱：[jskxldb@163.com](mailto:jskxldb@163.com)

通信地址：南京市鼓楼区北京西路30号同心大厦10楼1014室，210024

附件：2022年度省科协调研课题拟立项资助项目明细表

江苏省科学技术协会

2022年5月11日

附件

2022年度省科协调研课题拟立项资助项目明细表

(一) 面上调研课题

序号	课题名称	申报单位	负责人
31	公共卫生事件应急管理科技支撑体系建设研究	省计算机学会	朱 琳
32	江苏老年心理健康研究与实践	江苏华瑞老龄服务产业发展研究院	张伟新
33	运用靶向调节技术促进结肠直肠癌转移的应用机制研究	省抗癌协会	朱松君
34	江苏省安宁疗护服务现状调研及政策建议	省护理学会	吴金凤
35	省内重点区域癫痫患者驾驶机动车风险管控对策研究	苏州大学附属第一医院	胡小伟
36	无线感知技术在医疗健康领域的应用研究	省智能助老装备研究会	鲁 超
37	数据要素新形势下江苏省应急管理研究	南京南工应急科技有限公司	钱城江
38	江苏增强源头供给培育国家战略科技力量的策略研究	省计算机学会	施 峰



# 多类云服务随机请求的异构资源排队性能分析与调度

—— 2021年江苏省计算机学会优博论文奖

推荐单位：东南大学计算机科学与工程学院  
论文作者：王爽  
指导老师：李小平

用户请求具有泊松到达、一般随机到达等不同模式，具有不同忍耐度、截止期等约束；服务资源通常为单队列或多队列的异构处理器，其服务方式服从指数或其它随机分布。这些模式和约束组合成多类复杂排队模型，如何为每类供需服务确定合适的处理器数量并合理调度以优化响应时间和功耗、租赁成本、系统能耗、服务提供商收益等不同目标是云计算的关键问题之一。推荐2021年江苏省计算机学会优秀博士学位论文“多类云服务随机请求的异构资源排队性能分析与调度”，针对具有忍耐度、截止期等不同约束的泊松到达用户请求，分别研究单队列和多队列异构处理器的排队性能分析和调度优化方法，结合阿里云数据中心实际数据，验证提出算法的有效性。成果来自东南大学计算机网络和信息集成教育部重点实验室李小平教授团队。

## 专家推荐语

本文面向实际复杂云服务调度问题进行深入研究，主要贡献如下：

- (1) 考虑异构处理器启动、处理和睡眠等多状态，突破传统研究仅局限于处理单状态，提出全面解决响应时间和功耗矛盾的均衡迭代启发式调度策略；
- (2) 考虑实际用户忍耐度不同，突破传统假设用户忍耐度相同的局限，提出最大化收益的多队列调度算法；传统研究面向随机请求在队列长度固定下造成资源浪费，构建随机请求的排队容量动态调整策略，提出能耗最小化的弹性调度算法有效提高资源利用率；
- (3) 针对现有研究只考虑同类用户导致处理器难匹配，提出面向多类用户的租赁成本最优化迭代优化配置策略。文章研究成果突出，是一篇比较好的博士学位论文。

## 论文看点

- 1、针对云服务随机请求的单队列性能分析与调度问题，构建单队列排队模型；根据服务系统的状态空间，采用马尔可夫过程计算系统拒绝率；提出基于二分法的处理器数量和初始类型的决策算法；构建处理器选择策略的评估算法，评估系统的性能；提出系统响应时间和功耗均衡的迭代改进请求调度策略。
- 2、针对带忍耐度的云服务随机请求单队列性能分析与调度问题，构建租赁成本最小化的单队列排队模型；构建系统状态空间；采用马尔可夫过程分析请求的实际等待时间；计算系统拒绝率；提出处理器配置合理性判断策略；提出面向不同忍耐度用户需求处理器迭代优化配置策略，构建租赁成本最小化的调度算法。
- 3、针对截止期约束的云服务随机请求弹性单队列性能分析与调度问题，构建队列长度不定的弹性单队列模型；依据马尔可夫过程证明拒绝率与弹性队列长度的定性变化关系；提出最佳处理器配置和队列长度的策略；提出能耗最小化的云服务随机请求弹性单队列调度算法。
- 4、针对带忍耐度的云服务随机请求多队列性能分析与调度问题，构建多队列排队模型；提出最小化请求响应时间的请求流拆分算法以合理分配请求到多队列；提出各队列请求的最小化响应时间分配算法；提出最大化服务提供商收益的多队列随机请求服务调度算法，确定各队列中最优处理器数量和请求调度方案。

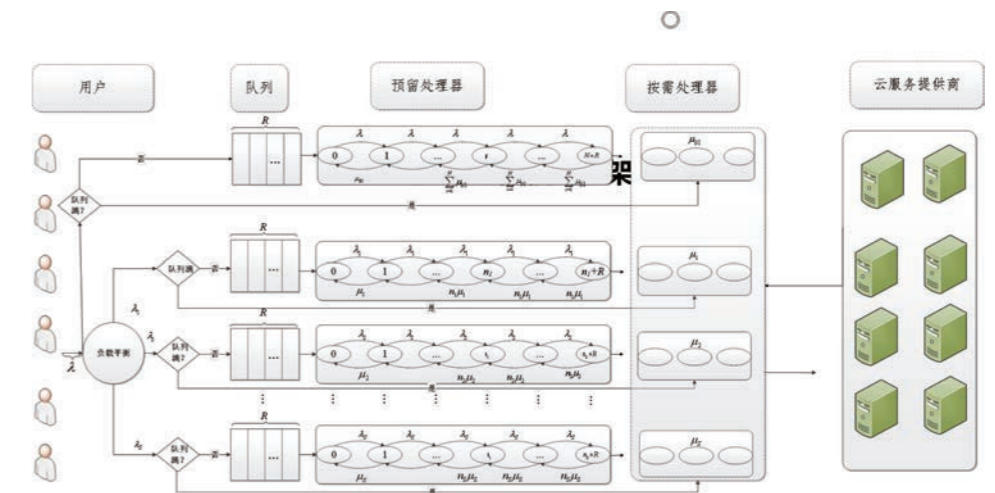


图 1 总体排队模型



本文针对云服务排队模型系统特征总结如下：

- 动态随机请求
- 海量异构资源
- 动态性能分析
- 调度优化

云服务随机请求的单队列性能分析与调度

单队列系统模型：考虑云服务随机请求的单队列性能分析与调度，以系统中处理器功耗和请求的响应时间的平衡为目标，考虑独立随机请求在异构系统中被处理的场景，若系统中有空闲排队位置，请求被分配到排队位置等待异构处理器处理；否则，请求离开系统。云服务系统中请求服从泊松到达，到达速率为 $\lambda$ ，处理器有COLD, SETUP, HOT三种状态。 $N_r(t)$ 、 $N_c(t)$  和 $N_h(t)$ 分别代表系统中请求的数量、系统中处于COLD状态处理器数量和系统中处于HOT状态处理器数量， $R_q$ 表示队列中请求数量， $R$ 表示系统排队容量。系统不同状态的变迁示意图如图2所示。

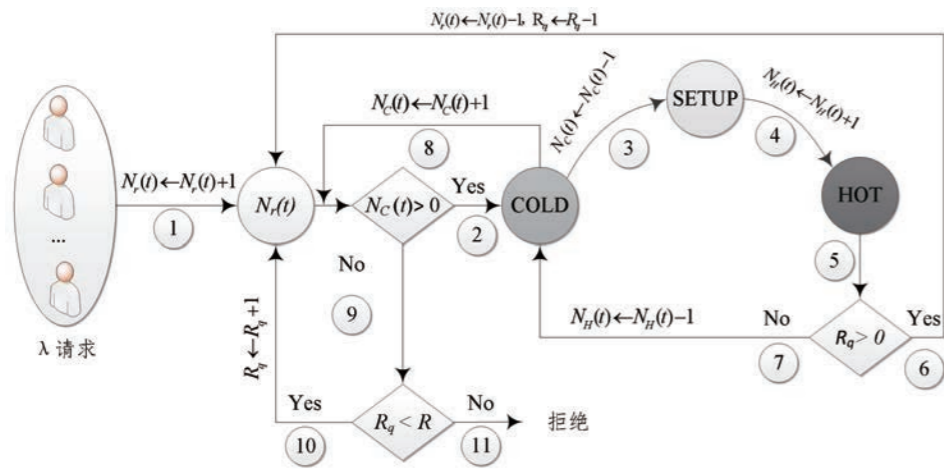


图 2 单队列模型状态变迁示意图

带忍耐度的云服务随机请求多队列性能分析与调度

多队列系统模型：服从泊松分布的请求随机独立到达云服务系统，到达速率为 $\lambda$ 。云提供商提供  $S$  种类型处理器，每种类型处理器视为一个队列，那么系统中有  $S$  个队列，为  $Q_1, \dots, Q_S$ 。处理器处理速率是独立的并服从指数分布，满足  $1 \leq 2 \leq \dots \leq S$ 。根据按需实例或者预留实例租赁模式，服务提供商从云提供商租赁处理器。如果预留实例无法满足用户需要，服务提供商租赁按需处理器改善系统性能。将请求调度到系统分为两个阶段：请求分配到  $S$  个队列和每个队列中的请求分配到合适处理器处理。

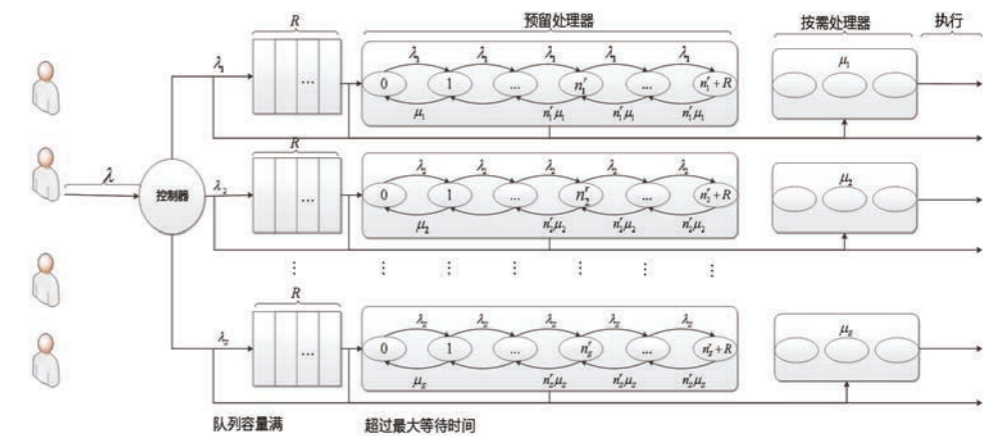


图 3 云服务多队列模型

挑战与展望

1、基于排队模型的调度优化技术：针对云服务随机请求的单队列性能分析与调度问题，如何构建单队列排队模型，并将随机云服务请求调度到异构处理器很关键；针对带忍耐度的云服务随机请求多队列性能分析与调度问题，如何将带有忍耐度的随机请求流分配到多个队列最小化平均响应时间，具有挑战。

2、基于深度学习的性能优化技术：当随机请求到达云系统时，随着处理器数量和类型增加，系统中状态空间是巨大的，这将导致无法计算状态稳态概率，因此，无法评估系统性能。由于处理随机请求时，通常采用马尔可夫决策过程，在对巨大状态空间进行处理时，如何结合深度学习技术将问题分解成有小空间状态构成的子问题，在以后的研究中，是一个很好的解决多类云服务随机请求的异构资源排队性能分析与调度的方法。

作者简介



第一作者：王爽，东南大学讲师，主要研究方向为云计算、人工智能、真值发现。

Email: shuangwang@seu.edu.cn



指导教师：李小平，东南大学特聘教授，主要研究方向为云计算、调度优化、服务计算、云制造、人工智能算法。

E-mail: xpli@seu.edu.cn

//// 学会动态 //////////////////////////////////////

团队简介



东南大学李小平教授团队，研究方向主要包括、云计算、服务计算、大数据、调度优化、人工智能等。实验室承担了国家重点研发计划、国家863计划、国家自然科学基金重点项目、国家自然科学基金等项目。实验室近年来申请发明专利18项（已授权9项）。实验室研究成果丰富，师生已在TC、TPDS、TSE、TSC、ICSOC等云计算、服务计算、调度优化顶级期刊和会议发表论文100多篇，实验室学生已获得江苏省计算机学会优秀博士学位论文、中国计算机学会服务计算优秀博士奖、ACM南京分会优秀博士论文奖等。

团队主页: <http://www.scholat.com/xpli>

江苏省计算机学会职业院校产教融合专家委员会换届选举会议暨产教融合专题研讨会在线成功召开

6月19 日下午，江苏省计算机学会职业院校产教融合专家委员会换届选举会议暨产教融合专题研讨会在线顺利召开，此次会议由江苏省计算机学会职业院校产教融合专家委员会主办、江苏安全技术职业学院承办、北京神州数码云科信息技术有限公司协办，来自省内高校专家以及企业领域100多名代表参加此次在线会议。



会议开幕式由江苏海事职业技术学院陈永副教授主持，江苏省计算机学会秘书长金莹教授、江苏安全技术职业学院院长李桂萍教授出席开幕式并致辞。





# 《江苏省制造业智能化改造和数字化转型三年行动计划》解读

来源：江苏省工业和信息化厅

为贯彻落实省第十四次党代会精神，坚持把数字经济作为江苏转型发展的关键增量，加快推进数字产业化、产业数字化，深化实施先进制造业集群培育和产业强链行动计划，全面推动全省制造业智能化改造和数字化转型（以下简称“智改数转”），省政府办公厅近日印发《江苏省制造业智能化改造和数字化转型三年行动计划（2022-2024年）》，明确了今后三年我省推动“智改数转”的总体要求、主要目标、重点任务和保障措施。

## 一、总体要求和目标

以习近平新时代中国特色社会主义思想为指导，全面贯彻党的十九大和十九届历次全会精神，深入贯彻落实习近平总书记关于制造强国战略重要论述、网络强国重要思想和对江苏工作重要指示精神，完整、准确、全面贯彻新发展理念，以深化新一代信息技术与制造业融合发展为主线，以智能制造为主攻方向，以工业互联网创新应用为着力点，加快推动制造业质量变革、效率变革、动力变革，着力提升产业链供应链现代化水平，为加快建设制造强省和网络强省提供有力支撑。

通过三年的努力，全省制造业数字化、网络化、智能化水平显著提升，新业态、新模式、新动能显著壮大，制造业综合实力显著增强，率先建成全国制造业高质量发展示范区。

到2024年底，全省规模以上工业企业全面实施智能化改造和数字化转型，劳动生产率年均增幅高于增加值增幅；重点企业关键工序数控化率达65%，经营管理数字化普及率超过80%，数字化研发设计工具普及率接近90%。

## 二、重点任务

大力实施“十大工程”，推动“智改数转”各项任务加快落地落实。

一是龙头骨干企业引领工程。对标世界智能制造领先水平，支持行业龙头骨干企业开展集成应用创新，每年认定一批省级智能制造示范工厂、示范车间和工业互联网标杆工厂、5G全连接工厂，加快形成“一行业一标杆”。

二是中小企业“智改数转”推进工程。加快建设江苏省中小企业“智改数转”云服务平台，省、市、县对规上中小工业企业协同开展智能制造免费诊断服务，推行智能制造顾问制度，帮助企业提供解决方案。

三是产业链“智改数转”升级工程。支持“链主”企业基于产业链协作平台开展协同采购、协同制造、协同销售和协同配送等应用。推行数字化交付，带动上下游企业数字化协作和精准对接。

四是工业互联网创新工程。支持综合型、特色型和专业型工业互联网平台建设，每年新认定10个省级重点工业互联网平台，打造20个“5G+工业互联网”融合应用项目，20个工业大数据应用示范项目。

五是领军服务商培育工程。分类制定标准，遴选建立全省“智改数转”生态资源池。建立服务绩效考核评价机制，对考核优秀的服务商给予支持。

六是自主可控工业软件应用工程。支持围绕企业“智改数转”需求开展工业软件技术攻关、产品研发和解决方案集成。发布首版次工业软件应用推广指导目录，鼓励制造业企业运用目录和清单内软件产品推进“智改数转”。

七是智能硬件和装备攻坚工程。分行业梳理智能硬件和装备供给短板，支持企业研发智能制造设备，每年认定智能制造领域首台套重大装备20个以上。

八是工业互联网支撑工程。组织制造业企业与网络运营商对接合作，加快改造企业内网，推动企业外网建设。优化全省数据中心布局，支持企业建设标识解析二级节点及数字运营中心。

九是工业信息安全保障工程。完善工业信息安全风险评估、信息通报、应急处置等制度。建设省级工业互联网安全信息共享与应急服务协同保障平台，培育工业信息安全防护星级企业。

十是优秀解决方案推广工程。总结提炼“智改数转”经验做法，每年征集和遴选100个应用场景、100个实践案例。开展各类供需对接活动，加大优秀方案和实践案例的宣传推广。

## 三、保障措施

（一）加强工作统筹。省制造强省建设领导小组办公室统筹推进制造业“智改数转”工作，成立江苏省制造业“智改数转”战略咨询工作专家组。各地人民政府主要负责人要亲自抓本地区制造业“智改数转”工作。

（二）加大政策支持。省级财政每年安排12亿元专项资金，支持工业企业“智改数转”。鼓励有条件的地方在省级财政补助的基础上，给予一定比例的配套补助，形成政策叠加效应。

（三）强化人才支撑。聚焦“高精尖缺”引进“智改数转”领域的战略科学家、科技领军人才和创新团队。开展制造业“智改数转”人才培养试点，培养制造业“智改数转”卓越工程师及青年科技人才，以及一批高技能人才和大国工匠。

（四）创新金融服务。支持金融机构创新金融产品和服务。引导金融机构增加制造业中长期贷款，支持中小企业设备更新和技术改造。支持企业通过融资租赁方式开展“智改数转”，融资租赁费用可享受同等财政补助政策。

（五）实施跟踪监测。探索建立制造业“智改数转”成效评估指标体系，开展全省制造业“智改数转”监测工作。

（六）营造良好环境。开展制造业“智改数转”环省行、区县行、进园区等活动，宣传制造业“智改数转”典型案例、解读相关政策，扩大示范带动效应。

# 掌握元宇宙“命门” 在相关前沿领域破冰领航

来源：科技日报

关键核心技术是元宇宙最大的“命门”，要始终坚持科技自立自强，充分发挥我国超大规模市场优势和新型举国体制优势，在虚拟现实、人工智能、区块链等元宇宙相关前沿领域破冰领航。

日前，世界各国都在积极布局元宇宙。在我国，上海、成都、武汉等城市也正在积极探索元宇宙产业发展路径。

“元宇宙涵盖了数字孪生、虚拟现实、增强现实、5G、云计算、人工智能、大数据、区块链等多项尖端技术，不仅可以加速各行各业的数字化进程，带动相关信息产业的大力发展，促进数字经济产值的持续提升，最关键的是，率先发展元宇宙，就能掌握未来互联网的高地和话语权。”全国政协委员、佳都科技集团股份有限公司董事长刘伟说。

掌握关键核心技术，需加强国家顶层设计

刘伟认为，元宇宙作为新兴事物，更需要政府带头示范、引导，使其从发轫之初就走上赋能实体产业、壮大数字经济的正确道路，也能更好为中国抢占未来科技的前沿阵地打下基础。因此，他建议直接由政府主导打造“元宇宙中国”数字经济体。

“国家有关部门加强顶层设计，将有力、有序推进元宇宙健康发展，更好地布局数字经济新赛道、抢占数字经济新高地，做强、做优、做大我国数字经济。”全国人大代表、江西省农村信用社联合社党委书记、理事长孔发龙持相同观点。

全国政协委员、上海市经济和信息化委员会副主任张英同样建议，应加快科学布局，促进元宇宙产业健康发展。她建议多部门联合开展“元宇宙+”应用创新试点示范，聚焦工业、商业、教育、医疗等领域，打造一批行业应用标杆。

作为一种多项数字技术的综合集成应用，元宇宙从概念到落地需要实现多种技术的融合创新。但从目前来看，元宇宙仍处萌芽期，关键技术有待突破。

“元宇宙是否可行，关键在基础设施。”全国政协委员、陕西省高级人民法院副院长巩富文直截了当地指出，“我国必须着力构建以5G通信网络为基础、以数据算力设施为核心、以智能互联为突破的现代化数字基建新体系，打牢元宇宙发展根基。”

巩富文认为，关键核心技术是元宇宙最大的“命门”，要始终坚持科技自立自强，充分发挥我国超大规模市场优势和新型举国体制优势，在虚拟现实、人工智能、区块链等元宇宙相关前沿领域破冰领航。

“我们应加快关键技术攻关，打造紧跟全球、引领未来的硬核技术体系。”张英建议，由国家发改委、工业和信息化部、科技部设立一批重大专项，聚焦下一代通信技术、卫星互联网、智能交互等，进一步发挥广大市场主体、科研院所等的战略科技力量，加紧布局研究、加快攻克一批关键核心技术；积极参与相关国家标准以及国际标准的制定，努力掌握国际话语权。

在孔发龙看来，应成立国家级元宇宙研发机构，加大资金、人才和激励支持力度。“应重点攻坚地理空间等元宇宙基础底层技术和关键核心技术，为促进元宇宙相关技术的自主可控发展提供坚实支撑。”他说。

理性看待元宇宙热潮，推动相关产业健康发展

当前，元宇宙行业发展呈现多元化势头，但也存在概念泛化、炒作严重的现象，这并不利于行业健康规范发展。

巩富文指出，要理性看待元宇宙所引发的新一轮数字经济热潮及其社会影响，客观正视资本扎堆、舆论炒作、概念泡沫、监管不足等问题，高度关注元宇宙在伦理、垄断、算力、隐私、知识产权等方面的潜在风险。

加强元宇宙前瞻性立法研究，推进元宇宙法制建设，成了代表委员们的共识。

“加快在国家层面研究制定元宇宙产业发展规划，建立产业标准、规范和体系，明确产业的边界、红线和禁区。”孔发龙呼吁。

刘伟建议，“元宇宙中国”数字经济体不仅是个技术平台，也是面向市民、游客、企业的复杂开放经济体系，需要建立适合虚拟数字经济的民事经济法律制度。他建议，应同步探索建立虚拟经济规则体系，推动数字资产确权、交易、隐私保护等方面的配套法规，打造制度样板，规范虚拟数字经济的运行。

“加快完善数字空间的治理规则，促进元宇宙相关产业健康发展。”张英进一步建议，加强法制建设，完善法律法规和司法解释，创新元宇宙发展监管机制；同时，打击违法犯罪活动，净化元宇宙产业发展环境。

二是中小企业“智改数转”推进工程。加快建设江苏省中小企业“智改数转”云服务平台，省、市、县对规上中小工业企业协同开展智能制造免费诊断服务，推行智能制造顾问制度，帮助企业提供解决方案。

## 学会动态

### 宿迁学院计算机科普教育基地与苏州大学共同打造国产图形化编程平台

党的十九大报告指出“科技兴国在中华民族伟大复兴中起着支撑和引领的作用”，“要培养造就一大批具有国际水平的科技人才”。江苏省计算机学会授牌的江苏省计算机科普教育基地之一——宿迁学院与苏州大学联合组建开发团队，怀着科技兴国提升青少年综合素质的使命，鼓舞中国青少年以国为傲，种下“国产化”的种子，推出了拥有自主知识产权的国产图形化编程平台。这是一个用拖拽图标替代编程语言图形化编程系统，它降低了青少年编程、特别是硬件编程的门槛，培养青少年对编程的兴趣，提升青少年解决问题的综合素质能力。



# 科技向善造福人类 ——解读《关于加强科技伦理治理的意见》

来源：中华人民共和国科学技术部

科技伦理是开展科技活动需要遵循的价值理念和行为规范，是促进科技事业健康发展的重要保障。近日，中办国办印发《关于加强科技伦理治理的意见》，对加强科技伦理治理作出系统部署，是我国首个国家层面的科技伦理治理指导性文件。意见的出台有何考虑？将推出哪些具体举措？就这些热点问题，记者进行了相关采访。

## 我国首个国家层面的科技伦理治理指导性文件

新一轮科技革命和产业变革正深刻改变世界发展的面貌和格局，科学新发现、技术新突破在造福人类的同时，伦理风险和挑战也相伴而生。科技部部长王志刚表示，科技部坚决贯彻落实习近平总书记的重要指示精神，把科技伦理治理工作放在事关科技创新工作全局的重要位置，在国家科技伦理委员会的指导下，牵头会同有关部门研究起草了意见。

“此次印发的意见是我国首个国家层面的科技伦理治理指导性文件，着力解决我国科技伦理治理体制不健全、制度不完善、发展不平衡等突出问题，对科技伦理治理作出顶层设计和系统部署。”王志刚说。

意见要求强化底线思维和风险意识，把科技伦理要求贯穿于科学研究、技术开发等科技活动全过程，推动科技向善，确保科技活动风险可控，科技成果造福于民。

王志刚表示，意见的出台对加强科技伦理治理工作具有重大意义，将极大促进我国科技伦理治理能力的系统提升，为实现高水平科技自立自强、推动构建人类命运共同体提供有力支撑。

## 坚持以人民为中心，推动科技向善

科技伦理原则是科研人员需要遵守的总体性、根本性的伦理要求。很多国家、地区和国际组织以及学术机构等均制定了相应的科技伦理原则。

科技部副部长相里斌表示，意见提出的五项科技伦理原则，充分吸收借鉴了国际科技伦理相关规则和共识，同时结合我国科技发展的历史阶段及社会文化特点，由我国科技界、伦理学界、法学界的专家学者集体研究、充分论证形成，并听取了广大科研人员

的意见建议。科技伦理原则反映了各科技领域伦理的共性要求，具有普遍适用性。开展科技活动，要坚持将科技伦理原则贯穿始终。

科技伦理原则的提出，一方面明确了我国对科技活动的基本伦理要求，以有效防范科技发展可能带来的不确定风险，促进科技向善；另一方面也阐明了我国政府及科技界对科技伦理治理的立场和态度，对加强国际科技交流、推进全球科技伦理治理具有重要意义。”相里斌说

## 多方参与、协同共治，不断提升科技伦理治理能力和成效

此次印发的意见提出了加强科技伦理治理的五项基本要求。即：伦理先行，推动科技伦理要求贯穿科技活动全过程；依法依规，加快推进科技伦理治理法律制度建设；敏捷治理，快速、灵活应对科技创新带来的伦理挑战；立足国情，建立符合我国国情的科技伦理体系；开放合作，积极推进全球科技伦理治理。同时，意见明确了政府部门、创新主体、科技社团及科研人员的职责和任务，致力于形成多方参与、协同共治的科技伦理治理格局。

加强科技伦理治理，制度是基础。”科技部科技监督与诚信建设司司长戴国庆介绍，意见从制定完善科技伦理规范和标准、建立科技伦理审查和监管制度、提高科技伦理治理法治化水平、加强科技伦理理论研究等方面对制度建设作出具体部署。

审查和监管是保障科技活动符合伦理要求的重要方式和手段。意见明确了科技伦理审查范围和要求，提出了科技伦理监管机制和措施，并对违法违规行为处理作出明确要求。”戴国庆说。

下一步，科技部将在国家科技伦理委员会的指导下，会同各有关部门和地方，切实抓好意见的贯彻落实。确立伦理先行的理念，强调源头治理、注重预防，建立科技伦理监管体制机制，对科技伦理高风险科技活动实行更严格的监管措施，对科技伦理（审查）委员会和科技伦理高风险科技活动依规进行登记，加强科技计划项目的科技伦理监篇.加强对国际合作研究活动的科技伦理监篇。

## 学会动态

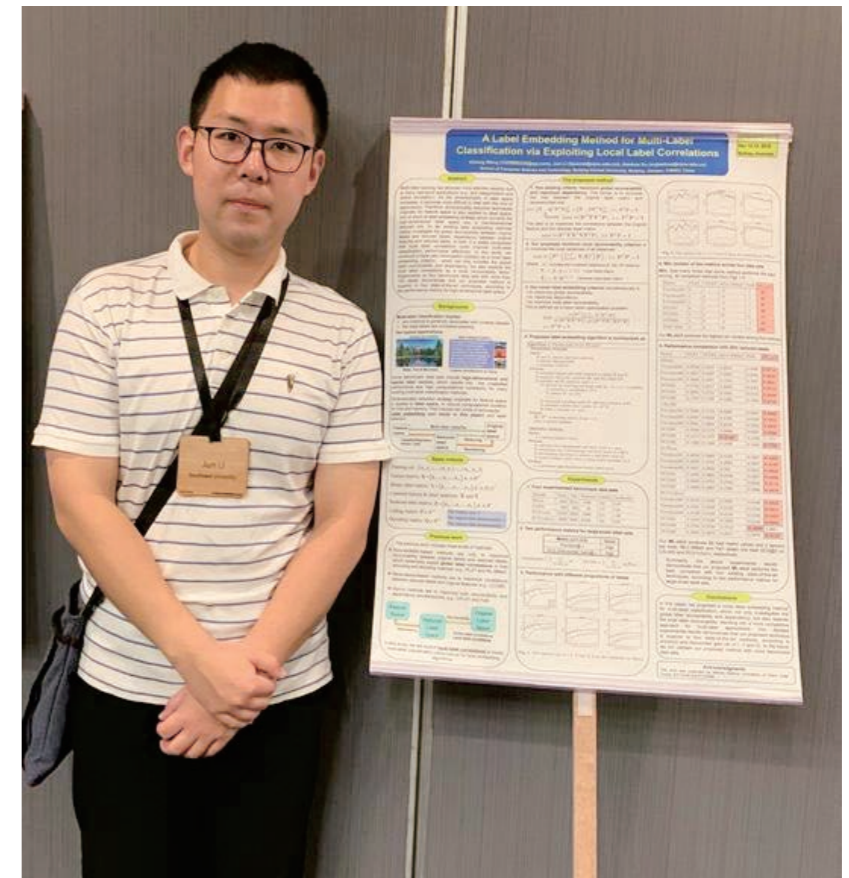
### 江苏省计算机学会两个2022年度“科创江苏”专项行动计划立项项目

5月11日，江苏省科学技术协会对“2022年度江苏省科协调研课题立项资助项目”进行了公示，我会申报的《江苏增强源头供给培育国家战略科技力量的策略研究》、《公共卫生事件应急管理科技支撑体系建设研究》两个课题均在2022年度省科协调研课题拟立项资助项目公示名单中。后续工作中，为充分发挥科技创新在创新发展中的引领作用，我会将进一步提高科技智库建设，团结和服务广大科技工作者履行“争当表率、争做示范、走在前列”光荣使命，为谱写“强富美高”新江苏现代化建设新篇章积极建言献策。

# 在探索图像子空间的科研道路上奋勇向前

—— 2021年江苏省计算机学会优秀科技者李峻副教授

李峻，南京师范大学计算机与电子信息学院/人工智能学院，副教授，硕士生导师。博士毕业于东南大学自动化学院，师从孙长银教授。曾在中国科学院自动化研究所、澳大利亚悉尼科技大学、加拿大阿尔伯塔大学等国内外高校和研究所进行访问学习和合作研究。先后获得2020 ACM SIGWEB China分会新星奖和江苏省计算机学会优秀科技工作者称号。现任中国/江苏省计算机学会会员，中国/江苏省自动化学会会员，江苏省人工智能学会模式识别专委会会员，IEEE会员。主要研究方向包括计算机视觉、模式识别、多媒体检索和机器人视觉等。先后主持国家自然科学基金两项、江苏省自然科学基金青年基金项目一项、中国博士后科学基金面上项目一项，迄今在图像处理、多媒体、数据挖掘等相关研究领域共发表论文40余篇，包括主流学术期刊IEEE TIP, IEEE TKDE, IEEE TCSVT, IEEE TITS, PR等。并担任多个知名国际SCI期刊的审稿人，曾先后获得Elsevier期刊Pattern Recognition, Neurocomputing以及Journal of Visual Communication and Image Representation颁发的审稿杰出贡献奖。



李峻在多媒体和模式识别领域做出重要学术贡献，近年来以多视角子空间学习为中心，从判别式鲁棒子空间学习、特权信息感知子空间学习以及轻量化子空间学习三个方面，围绕基于子空间学习的多视角异质特征融合方法展开深入研究，取得了一系列研究成果。发表论文30余篇，包括CCF A/B类论文8篇。获得了2020年ACM SIGWEB中国分会新星奖，并在2021中国图灵大会上做了题为“图像实例检索中的多视角子空间学习研究”的学术报告，获得了同行的一致认可和好评。







### 个人简介

魏秀参，南京理工大学计算机科学与工程学院教授，入选中国科协“青年人才托举工程”、江苏省科协“青年人才托举工程”、南京理工大学“青年拔尖人才”，主持国家重点研发计划青年科学家项目、首届CAAI-华为学术奖励基金等。曾在澳大利亚阿德莱德大学进行访问与合作研究。曾获江苏省计算机学会青年科技奖、南京经开区中青年优秀人才、南京新港高新技术工业园先进个人等荣誉。主要研究领域为计算机视觉与机器学习。目前在相关领域共发表论文50余篇，包括国际一流学术会议NeurIPS、CVPR、ICCV、ECCV、AAAI、IJCAI等，以及国际一流学术期刊IEEE TPAMI、IEEE TIP、IEEE TKDE、Machine Learning等。在电子工业出版社出版编著《解析深度学习：卷积神经网络原理与视觉实践》。研究成果荣获国际细粒度旗舰赛事iNaturalist、iWildcam等在内的共4项世界冠军，力压美国通用动力、Facebook等世界劲旅；相关研发系统入选《科技日报》“年终技术盘点”十项突破之一。

李峻注重人才培养，坚持立德树人，工作以来指导或协助指导10余名博士/硕士研究生，并带领指导本科生参加包括“大学生创新创业训练计划项目”在内的各个科研项目 and 学科竞赛，获得了较好的成绩。此外，李峻还积极参与人工智能方面的科普教育，现担任江苏省计算机学会普及工委副主任，江苏省计算机学会人工智能专委会委员，为人工智能、模式识别等相关领域的科研发展、人才培养以及科普教育持续做出贡献。

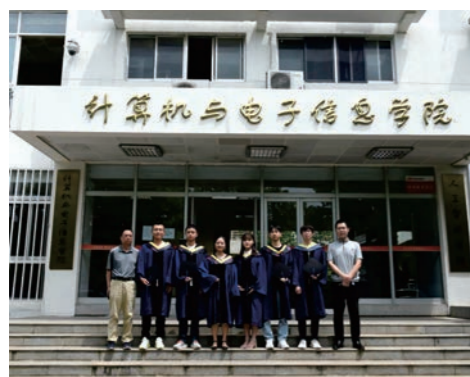


图1 魏秀参与图灵奖得主、中科院院士姚期智教授交流

魏秀参在计算机视觉和机器学习领域有深入研究，尤其是细粒度图像分析、深度学习等相关方面颇有建树，取得了一系列成果，发表论文50余篇，包括CCF A类论文30余篇，多篇入选ESI高被引论文，Google Scholar引用2800余次。魏秀参的研究工作已经在国际上有一定影响，如曾提出首个深度学习下的细粒度图像检索方法，填补了该方向近十年来的研究空白，被国际同行专家作为工作基础，目前单篇引用逾300次。更为难得的是，魏秀参团队的原创研究成果取得了国际计算机视觉权威竞赛评测共4项世界冠军，力压美国通用动力、Facebook等世界劲旅，体现出其研究工作健壮的实用性和广阔的应用前景。

### 细粒度图像分析

- 国家重大需求**  
《新一代人工智能发展规划》
- 国际学术前沿**  
"Fine-grained image analysis is the basic of perceptual embedding."  
--Serge Belongie, New York R Conference, Apr. 2017
- 关键产业应用**
  - 智能新经济
  - 工业互联网
  - 卫生大数据

图 2 图像细粒度分析

### 追求卓越，价值务实

魏秀参曾博士未毕业即被国内人工智能独角兽企业旷视科技聘为旷视南京研究院首任院长，全面负责南京研究院整体工作。博士毕业后，继续担任旷视南京研究院院长，对旷视的相关工作进行总体规划，从零到一创建南京研究院，在极短时间内凭借自身影响力组建算法研究员和研发工程师在内的核心研发团队，其中不乏多位国际计算机视觉竞赛冠军和ACM奖牌得主。

在推进人工智能技术落地应用的过程中，魏秀参带领团队围绕图像细粒度分析相关技术，开展了一系列有显著成效的工作，为旷视城市大脑业务和手机智能业务的应用落地提供原创技术支撑，先后获得国家发明专利近30项。特别在新零售业务中，魏秀参带领团队在算法攻关、数据建设、工程基建等维度从无到有研发旷视商品识别系统，已服务于徐福记、江中集团等国内零售巨头。其中的代表性技术，如“商品智慧巡检”“辅助收银”“智能货架”等受邀参展第二十届中国零售业博览会，尤其是团队自研的食品细粒度智慧收银系统“饕餮”入选《科技日报》2018年度“年终技术盘点”十项突破之一。

在魏秀参的带领下，旷视南京研究院因在人工智能行业的突出成绩被省经信委重点支持，并入选南京市首批人工智能产业创新集群牵头企业，积极参与“强富美高”新江苏与南京创新名城建设。华人首位图灵奖得主姚期智院士于2018年1月到访南京研究院时高度评价研究院工作，并寄语“要做世界领先的智能技术”。

### 教书育人，热心分享

魏秀参对高校教育和基础研究工作亦具有高度热情，曾牵头与南京大学人工智能学院、南京理工大学ICPC校队等高校单位达成战略合作关系，在履行反哺高校、培养人才的社会责任基础上，充分发挥校企双方优势，在培养人工智能专业人才方面进行有益尝试，探索人工智能人才培养新模式。魏秀参在繁忙的管理工作之余，从实际应用视角思考计算机视觉，后因对基础研究和教书育人工作的高度热情而志愿放弃企业高薪到高校从教，并以“青年拔尖人才”被南京理工大学引进聘为教授。

可贵的是，魏秀参在日常个人工作学习之余，还乐于在网络社区分享专业知识和见解，此外还牺牲个人休息时间，著写《解析深度学习：卷积神经网络原理与视觉实践》一书。值得一提的是，在写就该书的第一时间，魏秀参便率先将该书免费开源，为社区提供了免费电子版供广大读者免费参考学习。后因读者盛情相邀，特与电子工业出版社合作出版，目前该书销量逾万册，已第四次印刷，成为了相关专业学生和业内从业者入门深度学习技术的主要学习教材，受到了学界和业界的广泛好评。此外，面向国家能源安全、人民生命健康等重大需求，魏秀参与中国海洋石油集团、江苏省血吸虫病防治研究所等合作，将其研究成果成功应用于油气管网细粒度安全运维及传染性钉螺细粒度智能识别等任务，对我国能源安全保障和传染病防治工作起到了显著推动作用。

### 学会动态

#### 我会五个学科竞赛项目获2022年全省普通高校本专科生学科竞赛省级赛事认定

6月14日江苏省高等教育学会发布《关于公布2022年全省普通高校本专科生学科竞赛省级赛事认定（培育）结果的通知》苏高教会[2022]17号，我会五个学科竞赛项目获认定。

具体赛事如下：

江苏省大学生计算机设计大赛暨中国大学生计算机设计大赛江苏省级赛

江苏省网络空间安全职业技能竞赛暨世界技能大赛网络安全项目江苏省选拔赛

江苏省大学生程序设计大赛

全国大学生元宇宙设计大赛暨江苏省大学生虚拟（增强）现实技术大赛



# Web2.5已至，Web3.0还会远吗？

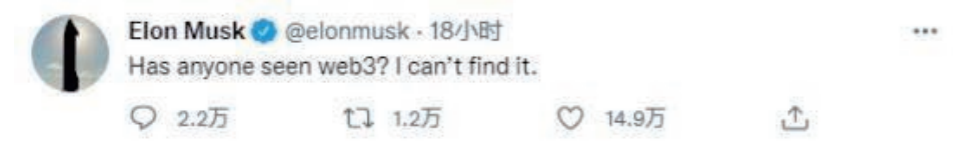
来源科普中国

过去一年，与元宇宙、NFT一样，Web3.0的热度也水涨船高，不仅百度指数Web3.0的搜索量与半年前相比翻了5、6倍，而且Web3.0在谷歌的搜索加权计算也是屡创新高。

事实上，这个概念很早就出现了，但其在去年爆发的“原点”是在12月9日，美国前货币监理所的代理所长Brian Brook在美国国会做演讲时，科普了Web3.0这个概念，并积极推动美国重视Web3.0，希望确保美国在Web3.0时代处于领导地位。



虽然国会议员对Web 3.0抱着积极和正面的态度，但同样有人不看好NFT。特斯拉CEO马斯克就曾连发多条Twitter表达他对Web3.0的困惑，“有人看到Web3了吗？我找不到。”、“我并不是说Web3是真实存在的——它现在看起来更像是一个市场流行词而不是现实”。



那么，被马斯克“唱衰”的Web3.0究竟是什么东西？它与元宇宙、NFT之间存在着怎样的联系？未来Web3.0又是否能够真的实现？

Web1.0到Web3.0，用户的价值逐渐得到体现有3.0自然而然就有1.0和2.0。

Web1.0时期大约从1991年持续到2004年，在这个阶段互联网才刚刚兴起，网络上以静态页面为主，用户仅仅作作为接受一方，被动的接受平台输出的全部内容，这个时候的企业代表以搜狐、新浪、网易、YAHOO、Google等传统的门户网站为主。

2004年之后也就是Web1.0阶段的后期，随着算力、光纤基础设施和搜索引擎等领域的飞速发展，Web2.0应运而生，区别于1.0的“只读”，2.0正式进入到“读写”阶段，每一个用户都开始成为内容的创作者，但平台的权力还是远大于用户，比如现在的Facebook、Twitter、微博、抖音都是这个阶段的产物。

以抖音和快手为例，在这个平台虽然看似每一个用户都可以搭配着不同的BGM创作自己感兴趣的视频，但实际上“生杀大权”是由平台所掌握的，你这个视频是否违规、侵权，有没有抄袭，都是由平台来判定的，甚至不需要你同意就能把你发布的视频“和谐”掉，涉及到更严重的事情，用户还有可能会被平台封号。

当然，这个逻辑在微博、推特也同样适用，毕竟，美国前总统特朗普在Twitter发表对互联网公司不利的言论后，Twitter可以直接封杀美国前总统在社交媒体上的账户。

直到Web3.0的出现，用户才有了真正“翻身做主人”的可能。

一些人看来，Web3.0代表的是未来互联网发展的一种新趋势，涉及去中心化、用户对数据和信息拥有实际的掌控权、线上资产搭建等。通俗点来解释，在这个阶段，通过一系列协议，用户不仅可以生产内容，还可以控制自己的内容。

Web3.0对比Web1.0与Web2.0来看，其优越性的体现或许可以用不久前ConstitutionDAO这个典型的、能够类比Web3.0的事件来进行分析、推演。

全世界仅存13份的1787年《美国宪法》在苏富比进行拍卖，但其高达千万美元的入场券也确实劝退了很大一批人，ConstitutionDAO这个组织却突然出现并在Twitter上发表言论称要众筹竞拍《美国宪法》。

这个想法确实有些离谱，但Constitution DAO领导人在Juicebos表示网友在捐款后将凭借捐款数量获得不同数量的凭证代币“people”，它代表众筹过后分红的比例与投票权。比如说，《美国宪法》将要去哪展览，展览的分红比例都可以通过“people”来决定。截止苏富比拍卖《美国宪法》时，筹款金额到达4700万美元。

综合Web3.0的基本性质以及上述案例的表现，开放是Web3.0时代的重要特点，即用户在不同的互联网应用中有着充分的准入自由，并且不受第三方主体限制。

第二也是最重要的一点则是用户之间的共建、共享、共创。

假设《美国宪法》如果被Constitution DAO成功拍下，那么它就会被所有出资的用户共有；它要用什么样的形式展出、在哪里展出、什么时候展出，都由用户来决定，在这个过程中也将产生大量的自下而上的创新；同时产生的收益也将由全部用户共享。

这或许也是这个项目能够打破“没有一家自治组织能够在短短一周时间里筹集到4700万美元”的原因。

第三，现在的Web2.0侵犯个人隐私，各大平台搜集用户数据，为平台牟利，而在Web3.0阶段，随着应用的去中心化，以及链上数据可查的情况下，用户行为、产生的数据乃至应用协议都能得到隐私保护。

第四，去年与元宇宙、NFT、Web3.0一样引起市场广泛关注的还有互联网大厂的垄断，国内对阿里罚款182亿和对美团34.42亿的反垄断处罚引起了市场的广泛的关注，当人们在讨论应该如何回收互联网大厂的权力，使其不能垄断整个互联网时，Web3.0凭借着其能在“某种程度上收回、控制互联网大厂的权力”也被多次提起，并被寄予厚望。

综合Web3.0全部的特征来看，它与NFT其实都是基于区块链出现的概念，只不过Web3.0是一组新的互联网协议，而NFT是一种数字资产。与此同时，考虑到Web2.0阶段平台意志决定内容导向，单体用户的价值难以体现，所以Web3.0与NFT、物联网、边缘计算和分布式存储一样，都是构筑元宇宙的重要部分。三者之间不存在谁成就谁，但命运却息息相关。

前有狼、后有虎，何时能从Web2.5迈向Web3.0?

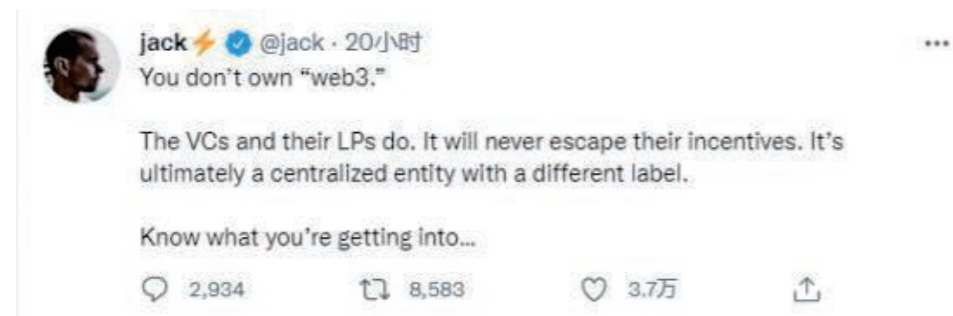
集开放、共建、隐私于一身的Web3.0为何会遭到一部分人的唱衰，在《智能相对论》看来主要有以下三个原因：

首先，对互联网公司而言，如果选择投入其中，需要付出的人力、物力、财力暂且不提，全新的互联网协议就是第一道坎，这意味着公司搭建的平台也不再像Web2.0阶段那样具有足够的权威性和话语权。

当然，现阶段，虽然Web3.0逐渐兴起，但绝大部分人基于便捷和依赖，依旧选择对Web2.0妥协，但互联网公司如果选择固守“老家”，放弃对“新世界”的探索，如果Web3.0真的是大势所趋，互联网公司也就失了先发优势，变得十分被动。在这样的不确定性下，难免陷入左右为难的境地。

其次，对投资者而言，也要谨防被割韭菜。

这也是Twitter前首席执行官多尔西抨击Web3.0的主要原因，他指出“用户并不实际拥有Web3产品，Web3的实际拥有者是项目背后的风投机构（VC）及其有限合伙人（LP），Web3永远不能脱离他们设定的激励机制。最终，Web3将是一个带有不同标签的中心化实体”。



也就是说，在多尔西的观点里，Web3.0并不会促进社会的民主化，而只会将决定用户的权力从Facebook等现有的互联网巨头手中夺走，却又转交给风险投资基金的手中。如果真是如此，那么那些被“共建、共治、共享价值”鼓动一拥而上的用户俨然就成了某些风险投资家手中的“工具人”。

所以，在Web3.0还不够不成熟的背景下，不仅项目本身可能因为不确定性太多而存在风险，而且还有可能被有心之人利用，成为他们圈钱的武器，这都需要投资者们以更加审慎的态度去对待Web3.0。

最后，也是包括元宇宙、NFT等概念都有的通病，那就是背后的监管挑战巨大，

这体现在两个方面。

一方面指的是这个概念对于监管的敏感度较高，毕竟不管是元宇宙、NFT，还是Web3.0都与币圈有着千丝万缕的联系，这也就意味它们对于相关的政策、法律法规有着极高的敏感度，一旦有利空的消息出现，势必将出现大幅的震荡，这一点其实可以从国内监管层对蹭元宇宙热度的企业发监管函，导致“元宇宙概念股”股价一路下行中可以窥见。

另一方面，则是监管本身的难度也不小。像是号称最安全的比特币都被黑客攻击过，累计经济损失数以亿计，也就是说去中心化的加密算法也无法做到真正的安全，这也是监管也难管制的领域；再比如每个用户都可以自己生产内容并控制内容之后，现在都无法规避的谣言等非法内容的传播问题也将更加肆虐，在这一点上进行监管又与Web3.0的本质有所背离。

所以，仅从监管的角度出发，现阶段的Web3.0还是一个比较“幼稚”的概念，走向“成熟”的过程中还需要做更多的准备。

总而言之，目前，已经有越来越多的组织和个人投身到“探索Web3.0的实验”之中，以共建为主要特征之一的Web3.0正借众人之手垒砌数字砖瓦。由于Web3.0还只是在实验中并没有到来，现在我们反而更像是处于“Web2.5”的过渡期。

为了实现平稳的过渡，很多企业对于拥有Web3.0相关技能的工程师的需求逐渐变大，据“币金所快讯”报道，在伦敦，一位熟悉Solidity、Rust等编程语言的入门级区块链开发人员的薪水几乎比传统软件工程师的平均薪水高出22%。在这样的趋势下，区块链技术、边缘计算技术等Web3.0相关的技术或许将率先得到发展，想必在不久的将来，Web2.0时代“羊毛出在猪身上，由狗买单”的局面也将彻底过去。

## 学会动态

### 热烈祝贺新大陆杯2022年江苏省大学生计算机设计大赛圆满结束

2022年5月21日在苏州大学举办了新大陆杯2022年江苏省大学生计算机设计大赛暨 2022年（第15届）中国大学生计算机设计大赛江苏省级赛决赛。

中国大学生计算机设计大赛是全国高校计算机科学学科和大学生计算机领域的一项盛事。自2008年开始至今，已发展成为全国高校影响力最大的大学生赛事之一，入选全国高等教育学会发布的全国普通高校学科竞赛排行榜，也成为全国各高校具有重要影响的年度学生课外实践活动之一。江苏省大学生计算机设计大赛现已举办9届，也已入选全省普通高校本专科生学科竞赛省级赛事项目认定名单，比赛不断完善，规模扩大，形式多样，影响力越来越广。省内各大高校积极响应，踊跃参赛，经过前期的校内选拔和推荐，共收到作品1289件，比去年增长近13%，参赛学校92所，参赛学生3656人次。初评过程中邀请了全省75名专家参加了初评评审工作，优选出429件作品入围决赛。

### “江苏省计算机学会科技产品：职芽线上线下融合招聘会系统解决就业大难题”

本次江苏省2022届高校毕业生“就业促进周”视频面试全省联动招聘会由认定为江苏省计算机学会科技产品的职芽线上线下融合招聘会系统提供全程服务，为高校精心搭建视频面试间，使用面试间的“电脑+摄像头”实时展示面试间的全貌；同时开发上线“E面试平台”，供学生与用人单位视频面试使用，实时同步面试招聘对接数据，展示视频面试服务成效。



# 互联网的下一站，大概率是能源

来源：科普中国

在2018年，互联网公司在巨大的压力下，迎来了一场走向2B的大迁移。三年多时间过去，这场迁移的价值确实得到了社会各界的认同，“信息技术与实体经济融合”成为了多项国家战略的基本行动逻辑。

但客观来看，这场被冠以“产业互联网”之名的行动，却没有给互联网公司们带来想象中的“二次崛起”。迟迟打不开的盈利空间，与传统行业融合的种种困难，都在提醒我们互联网摧枯拉朽的神话，在2B市场上不大可能复刻。

而另一面，随着疫情形势愈加严峻与监管力度的持续加大，互联网公司在C端的生存环境在进一步恶化。裁员、拆分、股市暴跌，等等关键词频繁闪烁在互联网巨头的头顶，从而导致这些企业走向2B的压力进一步加大。



接下来，我们大概率会看到这样一种变化：互联网行业在2B样样都能做，结果样样都做不好的怪圈中发生改变，开始思考能不能先把一个或几个重点行业做好，形成真正的赢利点，然后再谈其他。

而这场从互联网解决一切问题，到互联网服务重点行业的思维收束中，第一站大概率是能源。

赋能or服务：产业互联网的真正难题

这几年，我们已经看到了太多互联网、云计算、AI公司的2B发布会。每次都会用一页PPT列上数十个行业，再用一页打上几十家合作伙伴，再请几位案例级客户来发表证言。

长此以往，大家都很疲倦了，开始意识到这些互联网公司总是能拿出一大堆新概念，但真正的客户却基本听不懂；每次都有若干案例，但却没有真正的规模化营收；每次都展示不少客户，但这些很可能是他们在这个行业仅有的客户。

在互联网巨头开始推进产业互联网的时候，对其定义是“以互联网技术赋能传统行业，帮助传统行业振兴”。但几年来我们和很多实体经济的企业管理者、行业专家交流，大家普遍都有一个感觉：谁用你赋能？到底是你帮我振兴，还是我的钱帮你振兴？

时至今日，企业数字化和IT投入已经是常规选项，上云已经是基本技能。信息技术对各个领域的企业来说都不是秘密。但来自互联网的技术服务商，又很难说清自己与其他云计算、IT厂商提供的价值有什么本质区别。尤其对于那些对社交、电视、企业网络推广没有依赖性的行业来说，名为产业互联网的公司，并没有比其他IT供应商、ISV多长了三头六臂。

与此同时，互联网公司在走进B端时，往往携带的是一种互联网必然能改变一切的经验主义。认为基于云计算和中台，互联网公司什么都能提供。这种认知模式忽略了企业本身的诉求、发展轨迹，以及壁垒森森的行业区别，甚至有几分幼稚。



互联网，到底是要赋能传统行业，还是服务传统行业？这两个词互联网人看来可能没有区别，并且能滔滔不绝地讲述为什么没有区别；但在行业用户耳中，区别可大了。

总体而言，产业互联网发展到如今，主要有三个问题需要克服：

一、概念太多太杂，各说各话，缺乏行业公司。每家公司都认为自己做的是真理，用户迫切需要。但客户却很难像购买IT服务和企业网络服务一样在产业互联网的市场中找到共识和标准。

二、做了太多案例，但这些案例定制化程度太高，投入成本过大。其中很多甚至是互联网大佬用自己的面子和酒局争取过来的，缺乏可复制性和规模化价值。

三、行业的壁垒森严，很多产品和解决方案是难以在不同行业之间

通用的。产业互联网提供商往往认为这些实体行业都一样，我提供的价值可以通用。但在实体企业看来，我和隔壁工厂都是天差地别。

在这三个问题面前，互联网企业或许需要学会做减法。改变以互联网为中心，以平台化为基础的思维方式，走向以行业为中心，以服务为基础的新思维逻辑。

然后就是千万别想着互联网是万能钥匙，而是耐心寻找可以深入的产业空间。将一个或者几个重点行业的空间做大，可以通过快速复制获得利润，同时实现相对标准化的供需匹配。

这种很可能到来的改变中，互联网行业需要的是一个市场足够大，能够容纳互联网巨头的大船，同时也有足够改变意愿，需要更多前沿技术的行业。

在“双碳”目标之下，能源是一个亟待成长的B端风口。

能源互联网的定义、价值与可行性

在产业互联网被提出之后，接下来大家能听到的细分概念是工业互联网。但工业互联网本身也是一个非常笼统，各说各话的行业概念。虽然有政府支持，和一些行业标准化工作的推进，但其性质决定了很难实现高度标准化。

而“双碳”目标兴起之后日渐火爆的能源互联网，则在很大程度上符合标准化和概念清晰的要求。一般来说，能源互联网的定义主要有两部分组成，一是对能源体系进行数字化、智能化改造，二是搭建和更新能源相关的网络信息系统，从而实现低碳节能。



相比于目前其他实体经济和实体产业来说，能源行业与互联网企业的最大结合点就在于商业空间。能源行业资金充沛，并且各地“双碳”目标下达之后，能源相关的数字升级需求空前加大。尤其各地、各行业的“双碳”往往会层层加码，这导致很多地方和企业急于完成能效改造，继而打破了传统行业产业升级相对滞后的惯性。

与此同时，可以看到能源行业本身标准化程度较高，细分下来的垂直领域不算太多，一种新的数字化、网络化产品可规模化复制的程度很高。

而另一边，互联网公司云计算作为底座，对数据中心、网络节能的能源调配有天然的经验积累与优势，同时能源上云也是大势所趋。另一方面，互联网公司对AI、物联网等新一代信息技术的掌握程度更好，这符合能源行业以新技术换“双碳”空间的诉求；与此同时，互联网公司的软件开发能力更强，有高效率的产业需求满足能力，这也会比能源行业传统的ISV更有竞争优势。

而从可行性上看，互联网公司提供的技术，可以在多方面渗透到能源行业，以及与能源高度相关的产业空间。比如谷歌早在2016年就宣布，使用AI技术使谷歌数据中心能耗下降了15%，实现了数据中心PUE（电能使用效率）达到1.12。这相当于目前东数西算工程对新型数据中心的最严格PUE要求。

再比如，最近几年以南方电网为代表的电力系统，正在大幅推进用“摄像头+无

人机”的模式代替传统巡检。这其中，机器视觉技术、云管理系统与边缘计算体系都有着广阔的市场空间，符合互联网公司提供的业务能力。

事实上，从2018年走向产业互联网开始，能源一直是互联网巨头关注的方向。比如早在2019年的腾讯全球数字生态大会上，腾讯云就发布“智慧能源解决方案”，2020年又推出了综能工场、能源认知大脑、企业电像、智慧加油站等产品。而阿里云则面向分布式发电、微电网、电动汽车、家庭能源管理等10个重点能源行业提供了云计算和大数据支持。百度智能云则推出了能源AI中台解决方案。



想让能源互联网深度发展，互联网企业需要解决三个问题：

1.模式的问题。

能源行业是标准的国家主导，互联网公司只能做服务商与供应商。除了电池和光伏，剩下能切入的非常少。这就需要互联网公司理清自己的位置与发展空间，尤其需要讲述清楚自己与其他IT企业、能源数字化服务企业的本质区别是什么。既要学会服务，又要总结出自己的新服务模式。

同时，安全可靠是能源行业永远无法绕开的问题。这个层面是互联网企业和互联网云计算厂商的弱势。如何打造可信的技术供应体系，增强产品的信创率，是互联网公司必须解决的问题。

2.技术的问题。

在今天，能源行业是相对信赖新技术，愿意找寻新技术的行业。但互联网厂商依旧要理解行业诉求，讲清楚云计算、AI、网络化这些技术给行业用户带来的价值是什么，发展增量是什么。只有用新技术驱动市场改变，互联网公司才能逐渐获得话语主导权。

3.思维方式的问题。

说回赋能还是服务的思维问题，能源行业的主要特性是央企主导创新。往往用户的选择权非常大，以用户为中心形成生态空间。这与互联网的思维是正好反过来的。互联网公司能不能扮演好服务者，而不是以自我为中心形成平台化，将是产业互联网向前发展的最大方法论挑战。

另一方面，能源是创新幅度很大、技术迭代非常快的行业。互联网公司总是希望用户理解互联网思维，但自己又可曾去理解能源思维、工业思维呢？懂得行业诉求，理解行业知识，绝不能仅仅停留在一些沟通交流的层面。这是互联网走向产业几年来



都没有做好的工作。

只有改变自己，理解行业，尊重供需，才能真正让互联网与新一代信息技术走向产业端。没有互联网和千行百业，只有千行百业其中之一的互联网。按下了这个开关，能源互联网的风口才能打开，更多B端市场亦是如此。



## 学会动态

### 2022区块链技术论坛在线上成功举办

4月22日，首届网络空间内生安全发展大会暨第四届先进计算与内生安全国际学术会议分论坛区块链技术论坛在线上成功举办。本次论坛由紫金山实验室主办，南京邮电大学及江苏省计算机学会区块链专委会承办，中国计算机学会区块链专业委员会给予了支持。来自省内外相关高校、科研院所、企事业单位、政府等3000余名专家学者和相关专业技术人才参加了本次论坛，论坛由江苏省计算机学会区块链专委会副主任、南京邮电大学孙国梓教授主持。共有6位专家作了精彩的主题报告。



上海交通大学区块链研究中心首席科学家、中国计算机学会区块链专委会主任王明教授作了题为《区块链安全机理与拟态防御》主题报告，报告从区块链安全机理、区块链系统面临的安全威胁，拟态现象与拟态防御、网络空间拟态防御等方面进行了介绍，并给出了区块链系统安全性分析与拟态解决思路。



# X光安检图像智能识别系统研发及产业化应用

—2021年江苏省计算机学会科学技术奖一等奖

### 基本情况

项目名称：X光安检图像智能识别系统研发及产业化应用

完成人：支洪平、金博伟、许琢、鲁盈悦、王凯、王涛、郑伟伟、李玉笛、吴子扬、谢小环、李小兵

完成单位：科大讯飞（苏州）科技有限公司

### 项目简介

随着轨道交通、民航、快递物流业和公共事业的快速发展，公共安全面临严峻挑战，安全防范压力也在不断增加。2020年我国铁路营业里程达到14.63万公里，2021年“双11”期间国内快递和邮政企业共运送快递包裹68亿件，2022年春运全国发送旅客达10.5亿人次。大客流量和海量包裹给安检工作带来前所未有的压力，安检管理面临巨大挑战。当前X光安检基本依赖人工，判图速度和行李过检速度缓慢造成拥堵，影响乘客出行体验；同时由于安检员工作强度高、易疲劳以及注意力涣散等原因，往往会发生错漏检现象，造成安全隐患。传统安检业务管理还普遍停留在手工记录纸质台账阶段，禁限带品难以溯源，安检行业信息化水平较低，存在数据孤岛，安全管理者无法及时有效获得全局安检态势信息。安检行业亟需能够解决以上痛点的产品和解决方案。

近年来，党中央、国务院、工信部、公安部等多部门相继发布系列政策，支持人工智能赋能安检行业，明确提出智能安检是人工智能创新应用重要推广领域。

本项目将新一代信息和人工智能技术与安检行业融合，通过所研发的X光安检图像智能识别系统，持续提升识别准确率和效率，配合安检员进行人机耦合判图。推动高铁、地铁等行业领域安检设备升级改造，建立安检大数据云平台，强化安检全面监控和统筹管理，改善安检管理效率，同步推动行业智能化和信息化升级，充分体现一

线安检的最后屏障作用，对服务国家反恐战略，保障社会治安稳定，改善人们出行体验具有重要意义。

### X光安检图像智能识别系统

本项目的核心产品——X光安检图像智能识别系统包含值机台、开包台、云平台和安检员培训考试四大模块，如下图所示：



图1 X光安检图像智能识别系统组成

现有系统组成如下：

(1) 智能识别系统：基于深度学习的X光伪彩色图像禁限带品识别系统，本系统可与X光安检机通过视频接口对接或植入X光安检机设备内部，实现X光安检禁限带品检测。检测结果在安检机同屏界面上显著提示，提供外接声光报警等多种方式，以人机耦合方式辅助安检员智能判图。自研行李分割算法、禁限带品检测和跟踪算法，从头打造了阅光宝盒硬件识别终端。

(2) 开包台系统：通过整合人包图关联、禁限带品收缴处置等多个业务信息化模块，改变原有人工记录纸质台账的业务管理模式，实现了安检业务全流程的信息化转型。

(3) 云平台：支持对全网内上传的安检大数据的实时分析和展示，并根据预设模型进行危情预警，也实现远程设备管理、模型下发、配置管理等功能，还有对抗巡检、集中判图等特色管理功能。

(4) 培训考试系统：支持个性化学习模式，用于行业内安检员以及标定人员的能力评估与资质考试。支持在线百余名考生同时考试，支持千余个用户同时在线。解决行业培训难、认证难的痛点问题。

产品可实现的功能如下：

(1) 精准识别：系统可实现104种以上违禁物品的精准识别，可疑液体、易燃易爆物品、刀具等多种禁限带品和危险品一览无遗，识别准确率超过95%。

(2) 同屏显示：系统采用同屏显示技术，智能识别出的物品被标记在值机员原有X光安检画面上，不需要另外部署屏幕，值机员注意力更聚焦。

(3) 兼容性强：系统可兼容国内外多品牌安检机，针对安检机的输出视频特点进行优化，可对接不同品牌、型号的单、双视角安检机。

(4) 禁限带品处置管理：系统自动将含禁限带品的包裹实时照片及对应X光图片传送到开包台，工作人员可快速定位目标包裹，禁限带品的照片、种类、携带者信息等只需要安检员动动手指就可在触摸屏上一键记录，实现了无纸化办公、全流程数据化监管，提高禁限带品处置效率。

(5) 安检勤务管理：系统可将安检人员当值信息即时登记到系统中，与安检设备运行状况、禁限带品检出信息在时间轴上进行同步，建立多维度安检信息数据联动，可溯源论功追责，发挥安检管理效能价值。

(6) 安保等级自定义：系统支持配置多级安保勤务，配置范围包括预警禁限带品种类及识别阈值。安保勤务等级越高，预警的禁限带品种类越多，识别阈值越低。系统可一键下发勤务等级到各安检点位，满足不同安保等级要求。

(7) 移动端信息查询：系统可将安检信息同步手机移动web端，包括现场过包信息，禁限带品信息，以及安检员在岗工作情况实时更新，提高管理者对现场安检情况的掌控能力，实时掌握全局信息。

(8) 远程巡检：系统支持发送测试禁限带品数据至前端安检机，一键获悉全网安检站点工作状态及反馈情况，快速完成全网远程巡检工作，扫除巡检盲区、提升巡检工作效率。

(9) 集中判图：系统支持将各安检通道实时画面传输至后端进行集中判图，在保障安检效率的同时，合理优化安检人员配置，提高安检员工作效能。

(10) 安检培训考试：系统可以联动使用实际工作中采集的禁限带品过包真实数据，实现培训考试和业务实战的紧密贴合，全面提高安检员工作能力，进一步降低安全风险。

本项目取得了如下成果：

(1) 拥有多项专利和荣誉：已申请29件专利（其中发明专利18件，实用新型和外观专利11件），已授权专利21件，获得软件著作权证书11件，注册商标19件，已获得21项各类省市和行业内荣誉。

(2) 可准确识别多达104种治安领域常见禁限带物品：包括枪支、各类刀具、烟花爆竹、打火机、压力容器罐等等，也含疫情期间铁路行业关心的酒精喷雾等，平均识别准确率达95%，是目前行业的最高水平。

(3) 在各类评测和算法比赛中取得佳绩：在10多次现场PK中夺冠完胜。例如：于2019年3月参加阿里天池X光算法比赛脱颖而出，获得两轮初赛第一名；2020、2021年成功组织了两届X光安检图像智能检测大赛，全国各高校、研究所和企业累计超过1500支队伍参赛。在2020年11月，在铁科院组织的行业评测中，获得图片评测、实物评测和实战评测三项第一。

(4) 行业领先的算法和解决方案能力：吸引了业界大量的合作伙伴，目前已经和X光安检机厂家和行业解决方案商建立了合作，实现了从前装到后装市场的双轮驱动。顺利在武汉地铁、苏州地铁、南京地铁、合肥地铁、苏州中院、芜湖中院、南通中院、武汉中院、浙江二监、三监等地合计部署了数十个点位，稳定运行超过3年。



于2020年春运期间支持杭州客运站的安检工作，得到了CCTV13 新闻报道。产品涵盖机场、地铁、高铁、法院、邮政等各个场景，并将在全国高铁18个路局全面推广。

### 主要创新技术

系统基于X光安检伪彩色图像禁限带品检测的需求，基于深度学习SOTA技术成果，研究业界领先的检测和跟踪算法。在同等规模标定样本数量的前提下，相比竞品可提供更好的识别效果。对于新增禁限带品品类，具备快速响应能力，可高效迭代算法，满足安检现场业务需求。产品主要创新点如下：

#### (1) 算法：融合各种先进算法，构建全新的训练和推理网络框架。

针对X光安检机的伪彩色图像特点，全新自主研发检测算法。该算法以二阶段深度学习目标检测框架为基础，辅以多尺度特征提取和特征融合、半监督学习、模型蒸馏、图像融合、结构搜索等先进技术方法，设计并优化各阶段的损失函数，训练专门用于X光伪彩色图像的目标检测模型。模型能适应多厂商间的数据分布，对不同厂商都有一定识别能力，具备泛化到不同厂商和机型的最佳技术实践路线。

##### ① 特征融合技术

和常规图像相比，X光图像中各类禁限带品目标尺度差异较大且交叠遮挡现象严重。每一幅图像，深度网络将通过级联映射获得在某一形上的表征，和原图像相比，该表征更具有计算机视角下的语义性。

算法在特征融合方面做出深入研究，研发尝试多种特征融合方式，在特征提取网络提取不同层级特征，包括纹理细节丰富的底层特征和语义信息丰富的高层特征；不同层级的特征融合后可提供同时具备细节和语义信息的特征图，提高了模型对细粒度目标特别是多目标在交叠遮挡时的检测能力，达到检测任务的分类和定位两种要求；

通过在FPN结构引入Domain Attention block、Non-local block等方法，能有效提取X光伪彩色图像的特征，解决了交叠情况下提取特征的难题；针对不同厂商数据的尺度变化大的问题，为了适应不同物理尺度的目标、适应在不同分辨率图像中的不同尺度的目标，充分利用了多尺度特征提取和特征融合技术，使得模型可以识别常见尺寸下的禁限带品。

##### ② 半监督学习技术

针对大量无标注数据没有被充分利用的问题，利用无标注数据辅助模型学习潜在的数据分布特征，提升模型泛化性能；同时减少模型训练对大规模标注数据集的依赖，降低数据标注的成本。在充分研究mean teacher、最大期望EM、pseudo label等多种无监督学习方法的基础上，探索出无标注数据的最佳应用时间路线。

在X光图像目标检测任务中，尝试使用mean teacher、最大期望EM、pseudo label等多种无监督学习方法。本项目所应用的是pseudo label伪标签方法，通过高精度模型，对大量无标注数据进行多尺度预测，提升数据的伪标签质量。

##### ③ 自监督学习技术

受限于标注人力成本和时间成本，未标注的数据数量远远小于精标数据数量，这

是当前业界面临的一大难题。为了解决更好地利用海量数据，以自监督学习的方法利用大规模无标注数据产生某种形式的目标任务来训练神经网络，从而挖掘出更加丰富的特征表示。

通过组合多种数据增强操作构建正负样本对，构建对比损失函数完成自监督，引入可学习的非线性变换模型结构保留重要特征信息，以滑动平均方式动态更新特征编码器参数。相对于使用ImageNet监督训练的模型参数初始化模型，使用BYOL在ImageNet上自监督训练模型的参数进行初始化时，X光模型的检测精度得以有效提升，取得了业界领先的成果。

##### ④ 模型轻量化技术

针对模型部署中大尺度图像的推理速度不够快的问题，应用模型轻量化的方法，在不损失精度的前提下减少模型参数，提高推理效率。从小模型的角度考虑模型轻量化，以轻量化检测框架入手，同时研究剪枝和模型蒸馏的方法，在召回率低下或精确率低下时有不同的蒸馏策略，探索出一套在保持精度的前提下提升推理速度的实践方法。



## 学会动态

### 省计算机学会工控安全专委会走进中国（南京）软件谷信创示范区

4月9日，省计算机学会工控安全专委会（以下简称“工控安全专委会”）走进中国（南京）软件谷信创示范区进行交流学习。

本次活动由博智安全科技股份有限公司（以下简称“博智安全”）牵头发起，中国（南京）软件谷管委会副主任黄敖齐，中国（南京）软件谷管委会副主任刘峰，工控安全专委会主任、南京信息工程大学教授傅德胜，工控安全专委会副主任、江苏航天七零六信息科技有限公司总经理于然，工控安全专委会副主任、中新赛克科技股份有限公司总经理王明意，工控安全专委会秘书长、南京邮电大学教授季一木等十余名专委会常委出席活动。

从结构空间搜索(NAS)的角度考虑模型轻量化,主要搜索Backbone、FPN等模块;从小模型的角度考虑模型轻量化,如MobileNetV2、V3以及Efficient-Net系列;从轻量化检测框架入手,如One-stage检测框架,Yolo-v5等;从剪枝和模型蒸馏的角度考虑模型轻量化。其中,模型蒸馏方法针对不同的检测情况模型蒸馏的方式也不同,召回率不足时主要蒸馏RPN与FPN,精度不足时主要蒸馏的RCNN,当Teacher model与Student model 容量差距太大时可以分步蒸馏。通过运用各种模型轻量化技术,可以有效压缩模型的体积和推理速度,满足了前端的业务需求。

#### ⑤ 模型量化技术

针对模型部署中算力不够的问题,应用量化方法,针对模型个性化定义量化模块,包括量化的精度选择、统计量选择等,在实际应用中能够有效降低推理过程中对存储和算力需求。量化方法分为两种:模型训练后再量化(优势:实施容易/方便快捷,劣势:精度下降的较多,小模型表现明显)和在线量化finetune(优势:精度损失较少,劣势:量化时间较长)。采用了伪量化(quantization-training-aware)方法,可以针对模型个性化定义量化模块。通过运用模型量化技术,可满足不同硬件的定制化需求。

(2)数据:独有的图像融合技术和风格迁移技术,业界领先的数据标注方法和规范。

为拓展数据数量、质量和丰富程度,收集不同厂商和型号,尤其是头部厂商常见型号的运维数据,利用图像处理技术清洗运维数据,对于罕见但业务上关注的禁限带品,通过人造图像和图像融合的方式补充数据,可持续、低成本扩充禁限带品检测种类,满足实战需求。此外通过风格迁移技术可实现不同厂商间的数据均衡,使得单一模型能够适应不同X光厂商数据源。

#### ① 图像融合技术

针对于部分禁限带品数据获取难的问题,通过图像融合技术法进一步降低制造和收集数据的难度,可持续、低成本扩充禁限带品检测种类,满足实战需求。图像融合主要分为三个层次:即数据级、特征级和决策级融合。数据级融合是最低层次的信息融合,也称像素级融合,是目前图像融合研究的重点之一。

项目针对不同单例禁限带品前景数据,根据背景图像的梯度复杂度及颜色分布等统计信息选择适合区域进行了自适应加权融合。

#### ② 图像风格迁移技术

针对于多厂商数据颜色分布偏差大的问题,开发了域自适应数据增强的方法,对不同的X光厂商的数据进行Domain Translation,从而成倍的扩充了多厂商数据量,最终使得单一模型能够适应不同X光厂商数据源。

将各个厂商的ID映射为一个Mask Vector作为condition加入生成模型中,在判别模型中额外的引入了Multi-Domain Classification Task,保证重构图像得真实性及光滑性加入了重构正则及TV L2正则。

不同厂商风格迁移图像通过风格迁移技术有效扩充了训练样本,可解决不同厂商间数据量不均衡的业界难题。

#### ③ 数据标注技术

数据是算法的基石,数据的标注质量值得格外关注。针对安检场景数据特点,开

发了专业数据标定工具包,有一套完善的标注体系。

按照业务上关注的类别,确定标注类别的名称和典型实例图像,确定不同形状物品的标注方法,提取物品最主要特征,制定标注规范;培养专业的数百名标注人员队伍,每一位标注人员接受严格的培训,对安检图像的识别能力普遍超过安检员;抽查、复查标注文件,确保每一张图像的标注质量过关。通过数据标注技术,有效保证了数据标注的质量和速度,为训练样本提供了有效支撑。

(3)软件:开发适合安检机的多目标实时跟踪方法,原创了一种目标尺寸匹配技术。

#### ① 多目标实时跟踪技术

X光安检图像禁限带品跟踪属于典型的多目标跟踪问题,但其图像中目标移动有一些特殊的运动特征,比如目标的定向移动性、运动平缓性和尺度不变性,充分考虑其运动的规律性,不需要采用传统复杂的多目标跟踪方法,而是采用局部模板匹配的方法就可以达到较好的跟踪效果。

采用局部模板匹配的方法,使用待搜索图像的固定大小的图像区域作为模板图像,结合X光图像定向移动的特点,进行单向模板匹配,将当前目标位置的扩大区域作为被搜索图像。利用局部模板匹配的方法,解决了运算量大、时间复杂度高的问题,从而满足了实时性的要求。本方法在1920\*1080分辨率下可以同时输出双路视频的帧率达到120帧/秒,即在8毫秒以内完成所有目标跟踪以及目标框、目标标签等渲染工作。针对目标框的抖动问题,本方法引入投票机制,将多个模板匹配的结果综合考虑,减少模板匹配带来的误差,结合X光安检图像所有目标位移相同的特点,有效地解决了目标框抖动的问题。

通过上述方法,实现了业界领先的违带品多目标跟踪算法,解决了现有安检行业多目标实时跟踪的关键技术难题,与现有技术相比至少有50%的效率提升。

## ////// 学会动态 //////////////////////////////////////

### JSCS老科学家委员会召开学习和工作研讨会议

2022年6月12日,江苏省计算机学会老科学家委员会在南京召开学习和工作研讨会议。参加本次会议和活动的老科学家共有17位,他们分别来自南京大学、东南大学、南京理工大学、南京邮电大学等高校的计算机各领域的老专家们。具体有:王汝传、张云鹤、张福炎、陈世福、陈俊良、杨献春、孙志挥、王绍棣、许满武、刘凤玉、傅德胜、潘金贵、王和珍、蔡瑞英、陈汉武、宗平、秦军等。江苏省计算机学会秘书长金莹教授,秘书处张洁副秘书长和办公室朱中之主任参加了会议。



## ② 目标尺寸估算技术

传统的禁限带品尺寸估算对人力资源的消耗仍然较大，并且处理流程较长，因安检员的主观因素可能造成不准确的问题。对此提出了一种简单高效的X光机安检图像禁限带品尺寸估算方法。

针对现有方案中需要开包员开包确认物品尺寸的问题，利用参照物的固定尺寸，然后通过检出物品的最小外接矩形估算出待测物品的实际尺寸，满足了实时性的要求，提高了工作效率。针对现有方案中单视角机器产生的数据单一，双视角机器一个安检员可能无法同时兼顾的问题，本方法将双视角数据源作为数据集，利用丰富的数据集，并结合的多角度的特性，提高了尺寸的估算结果的准确性。针对现有方案中安检员可能无法适应不同牌、型号的安检机的问题，利用参照物的固定大小，通过不同安检机的自学习过程，可以统计出不同水平角下参照物的尺寸，然后利用比例关系估算尺寸。方法能兼容不同的安检机，具有较强的适应性。针对现有方案中安检员的主观因素可能造成不准确的问题，方法利用数学方法得到的统计数据对尺寸估算，更具有客观性。

### 社会效益

本产品契合了新基建政策要求，满足治安防控体系建设政策中“地铁公交智慧防控”要求，以无感知安检提升民生获得感。武汉地铁用户考核报告显示：使用期间，禁限带品检出数量较试点前4个月时，检出率从40%提升至85%。对抗巡检为轨交公安节省大量警力，解决了传统巡检模式“汗水警务”的问题，以武汉地铁为例，每周约进行600余次暗访、巡检，以单次巡检节约20分钟计算，每周将为地方轨交公安节约200小时的巡检工作时间。集中判图为地铁运营公司实现减员增效，以长沙地铁试点测算，节约安检人力约32.5%，年度节约人力成本234万。

在疫情公共卫生安全事件下，安检管理的重要性和风险控制面临更大挑战。通过X光安检智能识别系统和智慧安检解决方案的应用，能够通过智能辅助安检员识图，大幅度提高安检检出效率，降低安检漏检率，避免危险物品影响公共交通安全。利用铁路车站实际安检数据，通过考试系统进行培训，对于安检人员业务能力进行考核与评估，全面提高安检员业务水平。

项目团队的愿景是：以AI赋能安检行业，推动行业技术和管理变革；助力平安中国建设，开启X光安检新智能时代。

通过上述方法，实现了业界领先的违禁品多目标跟踪算法，解决了现有安检行业多目标实时跟踪的关键技术难题，与现有技术相比至少有50%的效率提升。



# 软件工程数据解析与智能释放

—— 2021年江苏省计算机学会科学技术奖三等奖

## 基本情况

项目名称：软件工程数据解析与智能释放

完成人：孙小兵、李斌、李必信、李云、杨辉、许聪颖

完成单位：扬州大学、东南大学

## 项目简介

软件系统多年的运行和维护中不断产生包括源代码、缺陷、维护及需求变更、软件演化历史等众多。这些软件工程数据中蕴藏了大量关于软件开发和维护的有用知识。但由于这些数据具有噪声数据多、多源异构、动态演化等特点，导致其难以使用，无法满足软件开发人员的不同应用场景诉求。

本项目按照“软件工程数据分析→软件工程信息探索→软件工程实践”的研究思路，利用数据挖掘、信息检索、自然语言处理等技术对软件工程数据进行分析，抽取软件制品相关的有用信息；基于搜索、推荐等技术实现软件工程数据中的智能释放，有效支撑软件工程实践。

(1) 软件工程数据分析：针对软件工程数据中大量的非结构化数据较难处理的问题，设计了一种面向软件工程数据预处理的通用框架和技术。将信息检索技术中的主题模型运用于非结构化的软件数据处理，解析软件工程数据中的语义关联，有效揭示软件代码中的结构与功能特征，改进了软件工程数据分析质量。

(2) 软件工程信息探索：针对现有检索技术不能较好地适配软件工程数据的搜索或推荐问题，全面分析代码中的关联、演化等关系，结合用户行为，研究了基于Wordnet的代码查询扩充技术以及基于开发者行为分析的个性化项目推荐技术，实现了更准确的代码搜索和个性化项目推荐，辅助开发人员更加高效地开展各种软件工程任务，提高了软件开发和维护效率。

(3) 软件维护任务分配：针对软件维护任务要求与开发人员技能难以匹配的问

题，对各类维护任务特征和开发人员分配方式开展了系统的经验分析，揭示了现有开发者推荐技术的有效性关联因素，均衡考虑开发者开发经验、软件维护质量等因素，提出了增强的维护任务分配技术，实现开发人员的分级推荐，并结合相关主题模型推荐参考源代码文件作为参考信息辅助软件维护，提高了软件维护的整体质量和效率。

相关成果在CCF推荐刊物或会议发表论文30余篇，引用超过1000篇次；授权发明专利15项，软件著作权登记12项，3件发明专利进行了转化。

#### 主要创新成果

##### 1. 提出了基于主题模型的软件工程数据处理方法

随着软件持续不断的维护与演化，不断涌现大量多源异构的软件开发数据，如软件开发文档、修改日志、Bug报告、版本演化数据、问答数据等。软件开发人员需要挖掘和处理这些软件开发数据以支持各种软件工程活动，如特征定位、软件信息检索、修改影响分析、软件调试。但是，这些数据由于其结构、特征、使用目的不同，保存在不同的软件仓库中，导致难以有效使用，成为遗留历史数据，没有形成有用的知识以有效支持软件工程各项活动。

为此，提出并设计了一种面向软件工程多源异构数据处理的通用框架和技术。该框架将信息检索技术中的主题模型技术运用于各种非结构化的软件数据挖掘和分析中，对原始的各类多源异构软件数据进行处理，提取软件工程数据中蕴含的主题，基于主题表征将相关的软件工程数据抽取出来作为预处理后的数据源支撑各类软件工程活动，改进软件工程数据的分析质量，提高了各种软件工程活动的精确性和有效性。实验结果显示，该技术在特征定位和修改影响分析这两个典型的软件工程任务场景中，改进效果可以提高10%以上。而且该技术在实际应用中不受数据类型和程序语言约束，因而可以应用于各种软件工程数据源分析中。

##### 2. 提出了特征定位驱动的API推荐技术

在软件维护过程中，开发人员经常需要基于已有系统实现新的功能需求。为了加速开发进程，开发人员大量使用社区中的第三方库及其中的应用程序编程接口（API）辅助代码新功能的实现。然而，现有API推荐技术无法适配软件开发人员的代码开发和维护场景，所推荐的API经常不可用，且不知道在哪里使用，反而浪费开发人员在API理解和程序定位过程的时间。

为此，提出了一种推荐API并定位其在源程序中使用位置的技术。该技术首先使用特征定位技术，利用基于Wordnet的代码检索技术将特征相关的源代码文件标识为API的使用位置，实现API及其使用位置的推荐；然后基于特征相关的源代码文件，融合历史特征库和API库相关信息实现API方法的推荐。实验结果表明，与已有的API推荐技术相比，该技术可以更加准确地推荐出API方法及其使用位置，其中API方法推荐结果在指标Hit@5的准确率值相比现有技术提高了29.7%，该技术相关代码已开源，并得到开源社区部分用户的良好使用反馈。

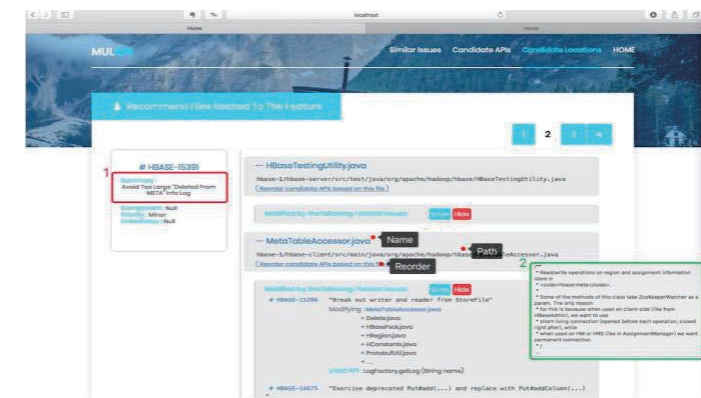
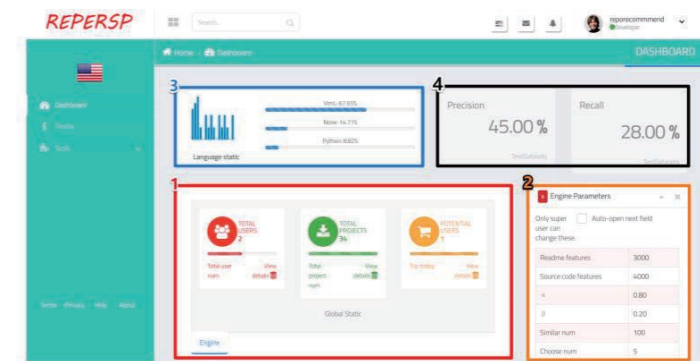
##### 3. 提出了面向GitHub的个性化软件库推荐方法

GitHub是目前开发人员参与最多的共享代码平台，开发人员可基于

该平台相互协作并参与他们感兴趣的项目开发，也可以利用该平台复用其功能和特性辅助开发人员自己项目的快速开发。但是，GitHub上项目众多，开发人员利用该平台自带的搜索引擎难以找到满意的项目参与或复用。

为此，提出了一种面向GitHub的个性化软件库推荐方法。该方法基于项目特性和用户对其他项目的行为，针对每一个开发者给出Top-N的个性化项目库推荐。此外，该系统还整合了用户反馈以进一步提高推荐的准确度。对于项目的特性，该方法分析了项目的源代码和说明文件，计算项目的相似度矩阵。对于用户行为，将开发者在GitHub上的行为，即Create, Fork和Star 提取为用户项目评估矩阵。最后，综合开发人员的行为和项目的特性，向每个开发人员推荐Top-N个与他们工作相关的项目。此外，本推荐方法将整个过程并行化以及使用模拟退火算法来生成最佳参数配置，使得开发者推荐的精度不断优化，实现更加准确的个性化软件库推荐。实验结果显示，该方法的查全率、召回率、准确率值远高于用户协同过滤推荐和项目协同过滤推荐技术。此外，实际的用户调查研究表明，该方法的确能够准确推荐用户所需要的项目，获得良好的反馈。

#### 部分成果图片



#### 社会效益

本项目中软件工程数据解析与智能释放相关技术已给开源社区部分开发者使用，并且获得了积极的反馈与评价。另外，一些技术也在企业项目团队的软件开发和维护过程中进行了部署，从应用情况来看，这些技术能够有效提高开发团队开发和维护效率，节省了开发和维护时间，提高了软件演化效能。