

范玉广

电话 : (86)136-9916-3629 E-mail : fanyuguang924@163.com

教育背景

2015.04 – 2015.07	新加坡国立大学	NExT 搜索研究中心	访问
2012.09 – 2015.07	中国科学院计算技术研究所	移动计算与新型终端北京市重点实验室	硕士
2008.09 – 2012.07	西安电子科技大学	计算机科学与技术专业	学士

工作经验

2021.01 – 现在	蚂蚁集团有限公司, 阿里巴巴	算法专家
--------------	----------------	------

➤ 项目：支付宝服务内容理解

- ◆ 项目介绍：负责支付宝全域服务内容数据的全链路加工和处理工作，建立和提供中台产品，分析与挖掘各种潜在关联，进行深度的语义理解和内容分析，为支付宝多种业务场景提供解决方案。
- ◆ 个人职责：
 - 完备组内基建，建立统一、规范化的模型开发流程，包括：任务评估、数据筛选策略、数据标注策略、基于主动学习的数据迭代策略、模型训练、模型上线，大幅提升数据质量，缩短模型开发周期，在数十个语义挖掘任务上做到 $P > 0.93$, $R > 0.9$ ，平均开发周期为 1 周/模型/人。
 - 基于支付宝海量服务内容，分析数据结构和场景的特性。引入结构化信息，构建基于统一预训练的模型体系，在功能词挖掘($P: 0.960$, $R: 0.907$)、场所类别挖掘($P: 0.942$, $R: 0.953$)、酒店事实类标签挖掘($P: 0.967$, $R: 0.933$)等任务上均优于纯文本结构的挖掘工作。进一步同等精度下应用多任务模型融合节省线上资源开销。
 - 对支付宝服务内容从“行业、类别、组成、元素、字段”，“功能、品牌、品类、场所、地域、时间”等进行多层次深度理解分析，挖掘语义、时空、人群等标签，搭建完备的标签体系，进行服务内容的结构化描述，助力支付宝服务库建设，保证内容质量，为支付宝搜推业务提供丰富的内容供给。
- ◆ 工作成果：深入内容理解中台定位和场景，对内完备组内基建、开发提效。对上游业务保证内容质量，对下游业务提供丰富的内容供给，支持搜推、内容展示等场景，在人工打标基础上扩增了 63% 的标签增益，推荐 feeds 区点击 uv 提升 0.9%，pvctr 提升 0.7%，扩召回长尾内容，带来搜索 pvctr 提升 0.90%，智能文案项目带来 32.43% 的收藏增量。形成了一套完备的内容理解与应用的解决方案。

2017.12 – 2020.12	微软（中国）有限公司	算法专家
-------------------	------------	------

➤ 项目：Smart Lists

- ◆ 项目介绍：在 Bing 广告业务中，将机器学习应用于转化数据，以确定哪些用户最有可能在后续会话中进行转化，并动态管理再营销受众群体，以重点关注这些用户。
- ◆ 个人职责：
 - 结构化广告 UET&Conversion 数据 pipeline，从大规模日志中挖掘目标客户主干节点及用户行为模式。
 - 使用几十种信号(包括事件时间、地理位置、设备、浏览器、referrer、session 时长、浏览页数、行为模式等)，融合图 embedding+深度神经网络挖掘并动态管理 Remarketing list，模型 AUC 0.8399。评估 UET 窗口对 Convert 影响，结果表明在 7 UET window 下 convert 获得最佳性能。

- 搭建 traffic flow 辅助客户了解用户模式，看是否达到了预期效果，辅助用户设置更有效的 conversion goal。
- ◆ 工作成果：提出用户行为模式挖掘并应用于 Remarketing list，在 Bing top 客户 walmart 上取得 0.7% cvr 提升。
- 项目：Bid Area
 - ◆ 项目介绍：Microsoft Advertising 通过对搜索、点击、转化的深度分析来帮助客户管理出价及在预算下有效出价。
 - ◆ 个人职责：
 - 追加 fetch/view/apply 等日志记录用户访问到调整出价策略的完整流程。
 - 修正 Bid landscape revenue 计算方式，在 listing bid landscape 和 adgroup bid landscape 两个场景分别砍掉 41%、20% 的数据，使 revenue 计算仅受组内 bid suggest 工作影响，保证收益计算精准。
 - 设计统一的 Bid landscape metrics 结构，计算出价访问量/应用量/访问率/应用率/收益，给出 T+1 更新的 metrics 报表，支持 listing/adgroup/campaign/bsc bid landscape, inline bid suggestion 等场景。
 - ◆ 工作成果：搭建 Bid Area 收益报表，了解用户调整出价后效果指标。
- 项目：汽车领域自然语言处理平台
 - ◆ 项目介绍：搭建汽车领域私有化部署的自然语言处理平台，满足车企内部当前和未来大量自然语言处理相关需求，提供自然语言处理能力，平台支持模型训练及服务部署。当前提供词向量、分词、命名实体识别、情感分析等服务。
 - ◆ 个人职责：
 - 平台基于 S3+DockerRegistry+OpenShift，其中 S3 用于统一知识库管理及数据托管。
 - 规范分词标准，基于 190w 汽车领域原始评论数据训练分词模型，提供分词服务，并支持用户词表。
 - 从大量汽车领域评论数据中抽取情感六元组，提供情感分析服务，建立行业质量报告。较当前的整句情感分析及片面情感分析任务，情感六元组任务更加精细、复杂、可提供更多信息。
 - 设计统一标注平台，支持但不限于分词、命名实体识别、意图匹配、情感分析。支持任务分配及交叉校验。
 - ◆ 工作成果：负责自然语言处理平台项目，已在销量 Top3 某车企内部上线使用。
- 项目：Dialog Engine 智能客服系统
 - ◆ 项目介绍：开发智能问答系统，让机器以准确而自然的语言回答用户所提出的问题，系统对接企业客户。
 - ◆ 个人职责：
 - 负责问答系统 Dialog Engine 中的分词、命名实体识别、实体标准化、Word Embedding 等自然语言处理模块的构建。以 gRPC+Docker+Nuket 包的结构搭建 NLU 微服务，服务供 Dialog Engine 项目及组内其他项目调用。
 - 对接企业客户，包括梳理业务需求，整理并分析企业数据，推动项目发展。给出实体识别模块在该场景下的可行性方案。企业 POC 阶段，针对企业数据给出分词，命名实体识别支持。
 - ◆ 工作成果：参与智能问答系统项目，负责其中分词、命名实体识别等模块的搭建。在模块的灵活性上适应了 2B 业务的特点。参与多个企业客户的对接，深耕智能问答系统在多领域的实践，如音乐、汽车，银行，保险等领域。

2016.08 – 2017.12

小米移动软件有限公司

算法工程师

- 项目：MIUI9 系统传送门
 - ◆ 项目介绍：在 MIUI 9 系统中，通过对文字区域的长按操作，触发系统对内容的智能分析，并把相关信息即刻呈现给用户，提高了跨应用获取信息的效率，为各类信息提供特定的直达服务。

- ◆ 个人职责：
 - 分词阶段，整理豆瓣电影、电视、书籍词条数据共 80 万条，包括词条名规整，词条冗余处理，白名单、黑名单挑选等。
 - Tensorflow 深度学习框架下实现双向 LSTM 对数据进行命名实体识别训练，Word2vec 在训练阶段同步进行，尝试加入词性，BLSTM+CRF 等方法。
 - 提供 Tensorflow Serving grpc 服务，编写 Tensor 流的数据处理，通过 Java Client 提供 API 调用。
- ◆ 工作成果：在 11w 训练集，GPU 迭代 20w 次，双向 LSTM 与 CRF 方法取得同等精度，Movie 准确率 0.942，召回率 0.847，F 值 0.892，生产服务中，Movie 准确率 0.911，召回率 0.865，F 值 0.887，单次请求 serving 响应时间约 18ms。词条的整理将词典召回率由 0.833 提升至 0.9266，对服务召回率有约 0.1 的提升，保证服务质量。

➤ 项目：小米生活

- ◆ 项目介绍：日常购物平台 App，与电商平台合作，优选天猫、淘宝商品。
- ◆ 个人职责：
 - 完成从商品图片中筛选出封面图片的自动化操作，Caffe 深度学习框架下采用 Inception 网络对商品图片进行二分类训练，内部标注训练集规模 1w，对内提供 HTTP 服务。
 - Caffe 环境部署及其繁琐，将上述环境打包成 Docker 镜像，内部通过镜像进行筛选模型的 Train/Predict。
 - 在 Tensorflow 深度学习框架下对上述进行重现，并同时提供 Tensorflow Serving Java API 及 HTTP 服务。
- ◆ 工作成果：分类准确率 0.982，HTTP 单次请求响应时间 0.7s。当前服务已稳定为小米生活 App 提供商品封面筛选服务，作为自动化工具，极大节省人力封面筛选成本。

2015.07 – 2016.06

中国科学院软件研究所

实习研究员

➤ 项目：数据交换池系统

- ◆ 项目介绍：建立出版社单位的统一数据交换平台。支持业务系统之间对象数据、资源数据、业务数据、管理数据等多种数据类型的数据交换、数据存储和统一管理。系统采用两级架构，由中心系统和前置机系统组成。
- ◆ 个人职责：
 - 开发数据处理模块，提供数据读取、写入、压缩解压缩、加解密等接口，支持适配器自动更新。其中加解密采用 AES 128。开发用户管理模块。
- ◆ 工作成果：支持适配器自动加载，支持本地文件系统内或本地文件系统与 FTP 服务器间文件传输。

2015.04 – 2015.07

新加坡 ViSenze 公司

实习

➤ 项目：ViRecognize 识别服务系统

- ◆ 项目介绍：基于 ViSenze 可视化知识库，提供一套从图像和视频媒体文件中提取信息的通用解决方案。
- ◆ 个人职责：
 - 对视频中识别对象，开发完成标签 Merge 算法模块，给定标签时间量级 1s，MongoDB 中每 5 分钟增量更新标签结果，将 Merge 结果可视化展示。
- ◆ 工作成果：将视频序列中以帧为单位的标签生成以段为单位的新标签，算法具有增量性。

2015.04 – 2015.07

新加坡国立大学

访问

➤ 项目：基于 CNN+RNN 的图像描述研究

- ◆ 项目介绍：借鉴 RNN 在机器翻译中取得的优秀成果，将图像作为输入，通过深度卷积神经网络 CNN 识别图像中物体，将提取的特新信息作为递归神经网络 RNN 的输入，并最终输出场景描述。
- ◆ 个人职责：
 - 研究 Caffe 深度学习框架和开源项目 NeuralTalk，训练集为 Flickr8K，Flickr30K，采用 VGG-16 + LSTM 网络结构训练 Image to Caption。
- ◆ 工作成果：Tornado + Caffe + NeuralTalk 搭建 Web 站点用于图像描述生成。

2012.09 – 2015.07

中国科学院计算技术研究所

硕士研究生

➤ 项目：面向高清视频的帧率上转换技术研究

- ◆ 项目介绍：帧率上转换用于提高视频的时间分辨率，以满足不同视频源间的格式转换需求。在传统的基于运动补偿的帧率上转换算法基础上，面向高清视频应用，提出一种基于自适应块划分和结构优化的帧率上转换方法。
- ◆ 个人职责：
 - 提出自适应块划分方法和改进的块匹配搜索方法。
 - 提出时空优化的运动矢量后处理技术、前向后向帧插值、结构优化的运动补偿技术。
 - 解决高清视频帧内缺少细节信息导致块匹配搜索不准确问题，改善空洞、咬合、模糊、块效应等问题。
- ◆ 工作成果：相比于基于非重叠块运动补偿的帧率上转换算法，恢复帧 PSNR 值提升约 2.17。

Publication

Paper : Yuguang Fan, Yiqiang Chen, and Wen Ji. "An adaptive smoothing method for collaborative HD video communications." In the 5th International Conference on Game Theory for Networks (GAMENETS), pp. 1-5, 2014.

Meiyu Huang, Yiqiang Chen, Wen Ji, Xiaojuan Ma, Jing Zhang, Yuguang Fan, Lianjun Dai. "CaneFitter: investigation on appropriate cane selection and proper cane use for the elderly." UbiComp/ISWC Adjunct 2015.

Patent : Yizhan Lu, Gang Chun, Yuguang Fan, inventors; Beijing Xiaomi Mobile Software Co Ltd, assignee. "Information display method and device." U.S. Patent No. 10,810,253. 20 Oct. 2020.

个人技能

- ◆ 熟悉基本数据结构、常用算法及机器学习基本方法。
- ◆ 熟悉深度学习网络结构，有 Caffe/Tensorflow 框架下处理 CV/NLP 问题相关经验。