

BAI

智源大会企业成果展示——百度

2020年6月

公司名称 百度

成果名称 ERNIE：知识增强的语义理解技术及平台

成果文字介绍

知识增强的语义理解技术 ERNIE (Enhanced Representation from kNowl-edge IntEgration) 是百度研发的 NLP 预训练模型，其创新性的将大数据预训练与多源丰富知识相结合，通过持续学习技术，不断吸收海量文本数据中词汇、结构、语义等方面的新知识，实现模型效果不断进化，在 16 个中英文典型 NLP 任务上显著超越谷歌 BERT 等世界最好效果。

2019 年 12 月，ERNIE 在国际权威的通用语言理解评估基准 GLUE 上(包含自然语言推断、语义相似度、问答匹配、情感分析等典型自然语言处理任务)，得分首次突破 90 分，超过人类水平 3 个百分点(人类水平是 87.1)，获得全球第一，领先谷歌、微软、Facebook、普林斯顿、斯坦福等国际科技巨头与著名高校学术单位的相关技术和模型。2020 年 3 月，ERNIE 在全球最大规模之一的语义评测 SemEval2020 上取得了 5 项世界冠军。



百度 ERNIE

知识增强的语义理解技术及平台



ERNIE 的创新成果，不仅体现在基于权威数据集的理论评测上，也体现在面向用户的实际产品应用中。为了让低内存，低功耗的业务场景也能应用到领先的学术成果，百度发布了语义理解模型压缩技术 ERNIE Tiny，大幅降低了 ERNIE 模型落地成本，使得 ERNIE 技术的广泛应用成为可能。同时，为了更好的服务各类行业，推动行业智能化发展，百度还构建了基于行业数据的专业语义理解技术，将 ERNIE 更好的用在了法律文书要素抽取、医疗病例结构化抽取等行业语义理解任务中。

百度围绕 ERNIE 语义理解技术，在语言生成领域进一步取得突破，提出了世界首个基于多流机制生成完整语义片段的预训练语言生成技术 ERNIE-GEN, 该技术在各类语言生成任务上也显著超越国际最好效果。

ERNIE 相关创新成果被国际顶级学术会议 AAAI 2020 和 IJCAI 2020 收录，也被包括《麻省理工科技评论》、德国光谱杂志、韩国 AITimes、日本 AI-SCHOLAR 等多国权威科技媒体官方报道。

百度依托 ERNIE 系列技术的突破构建了文心语义理解平台，该平台广泛应用在搜索引擎、新闻推荐、广告预估、语音交互、智能客服等产品中，大幅提升产品智能化水平，让人们更加方便、快捷、准确地获取信息，为社会带来了显著的经济和社会效益。

目前，ERNIE 已经对业界开源开放，有 120 多家外部企业与机构基于 ERNIE 进行了技术研发，极大促进了 NLP 领域前沿研究发展、人工智能开源平台繁荣及产业智能化变革。

成果介绍网址链接

ERNIE 开源地址: <https://github.com/PaddlePaddle/ERNIE>

ERNIE 平台页面: <http://paddlecloud.baidu-int.com/product/nlp>

ERNIE 介绍页面: <https://www.paddlepaddle.org.cn/ernie>

ERNIE 论文地址: <https://arxiv.org/abs/1907.12412>

<https://arxiv.org/abs/2001.11314>

公司名称 百度

成果名称 百度翻译同传：高质量、低延迟机器同传服务解决方案

成果文字介绍

百度翻译同传是一套**全场景、多模态、跨平台**的智能机器同传解决方案，翻译质量高、时间延迟小、易于部署，为用户提供视听一体、沉浸式同传体验。

百度翻译同传首次提出“语义信息单元驱动的机器同传模型”并研发了业内首个语音到语音的同传系统，通过将实时语音流动态切分为语义单元并结合上下文信息进行解码，突破了翻译质量和时间延迟难以平衡的公认难题；同时在语音识别容错、鲁棒翻译模型等方面取得突破，进一步提升了翻译结果的准确性。获评《麻省理工科技评论》2018年“全球十大突破性技术”之一实时语音翻译领域国内唯一“关键参与者”。

百度翻译同传会议版全面聚焦会议场景：1、全面服务支持线下及在线会议方式；2、全面支持主题演讲，圆桌讨论，远程连线等交流形式；3、全面覆盖会前、会中、会后全流程；4、全面适用于国际会议、讲座培训、内部会议、商务洽谈等会议场景。



百度翻译同传

高质量、低延迟机器同传服务解决方案



百度翻译同传会议版包括同传 PC 客户端、同传 APP、投屏网页、同传小程序/H5 等，支持多种平台；支持同传双语字幕投屏，使用手机扫码即可实时收听同传语音等多种模式；支持中英日韩等多种语言；同时支持行业术语定制、术语百科解释、会议同传记录等多个扩展能力。

除此之外，为辅助个人用户在观看英语视频、线上直播，在线会议交流等场景中可以快速获取跨语言信息，百度翻译还推出了轻量版的百度翻译同传助手，能够通过拾取电脑声音，实时生成同传字幕。同时，针对企业开发者，提供 iOS、Android、JAVA 多端 SDK，赋能开发者满足个性化同传需求。

百度翻译同传为本次 2020 北京智源大会的举行提供了全面技术支持。

成果介绍网址链接

<https://fanyi-api.baidu.com/simultaneous>

公司名称 百度

成果名称 RNA 二级结构预测算法 LinearFold 和 mRNA 疫苗序列设计算法 LinearDesign

成果文字介绍

百度正在利用人工智能技术提升对核糖核酸 (RNA) 这一基本遗传物质载体的认知, 并将研究结果应用于基于 mRNA 的疫苗设计, 以及新冠病毒分析等。

百度提出的 RNA 二级结构预测算法 LinearFold 可大大提升新冠病毒 RNA 结构预测速度。该算法使得整序列、整基因组的 RNA 快速结构预测成为可能, 也是 RNA 结构预测领域 40 年来第一次重大提速, 其可将新型冠状病毒的全基因组二级结构预测从 55 分钟缩短至 27 秒, 提速 120 倍, 节省了两个数量级的等待时间。目前, 该工作已发表于生物信息学顶级会议 ISMB 2019 和生物信息学权威杂志 Bioinformatics, 受到了诸多业内专家的高度评价。

基于此, 百度研发出 mRNA 疫苗序列设计算法 LinearDesign, 该算法可以最快 16 分钟设计出稳定的 mRNA 疫苗序列。LinearDesign 算法获得了国内外多家顶级科研机构、病毒防疫机构和相关公司的关注。



百度 Linear Fold

RNA二级结构预测算法

百度 Linear Design

mRNA疫苗序列设计算法



2020年4月，百度与中国疾病预防控制中心病毒病预防控制所签署战略合作协议，并联合设立“中国CDC应急技术中心-百度基因测序工作站”。将基于百度AI、大数据能力，推动线性二级结构预测算法 LinearFold、全新出炉的 mRNA 疫苗二级结构设计算法 LinearDesign，以及 Covseq 新冠病毒分析溯源可视化平台等前沿技术陆续在工作站投入使用，助力新冠肺炎相关研究。

成果介绍网址链接

<http://rna.baidu.com/>

Beijing Academy of Artificial Intelligence



微信关注
北京智源人工智能研究院