

中译语通机器翻译系统介绍

1、系统简介

中译语通自 2014 年开始进行机器翻译技术研究，是国际上继谷歌、微软和百度之后最早进行机器翻译技术研发并实现将机器翻译技术商业化的企业之一。目前已经研制成功的机器翻译产品支持 38 个国际主要语种，约 1332 个语言方向。2016 年底全面转型神经网络机器翻译后，依托高质量的数据资源实现了中英、中俄、中西、中葡、中韩等十余个语言方向的机器翻译准确度达到了世界先进水平，**位列行业第一梯队。**

中译语通机器翻译创造性结合统计机器翻译技术和神经网络机器翻译技术，将统计机器翻译系统对于短句的翻译优势，同神经网络机器翻译系统对于长句和复合句，精准流畅翻译优势有机结合。

中译语通机器翻译系统训练语料规模达 **51 亿+平行语料，200 亿+单语语料**，采用最新的 Transformer 神经网络结构进行训练，通过自注意力层，考虑到句子的全局信息，使得译文既流畅又准确。

多项国际评测成绩领先：中译语通在中对外的多个语言方向，在国际性的比赛中取得了领行业瞩目的成绩。在 2017 年 12 月第 14 届国际口语机器翻译评测比赛（IWSLT）中取得了 20 个语言方向中 16 个语言方向第一名的成绩，是中国历届参与该评测的最好成绩；在 2018 年国际机器翻译评测大赛（WMT）中取得了英中方向自动评测第一名的成绩。

2、技术创新点

- 由于神经网络机器翻译与统计机器翻译在结构上的不同，存在着词表的限制，故使用字节对编码来减小词表，使得词表更加干净。
- 使用了最新的 transformer 模型结构。该结构在词向量的基础上加入了词的位置信息，更多地考虑了词与词之间的关系，即句法上的联系。而基于自注意力层的编码和解码结构，在训练时结合了各个位置的词之间的关系。同时由于结构的特殊性，并行训练速度更高，相比较与传统的循环神经网络结构，训练和解码更加高效。
- 模型结构中还加入了 Dropout 技术，即随机断开神经网络，防止由于模型内

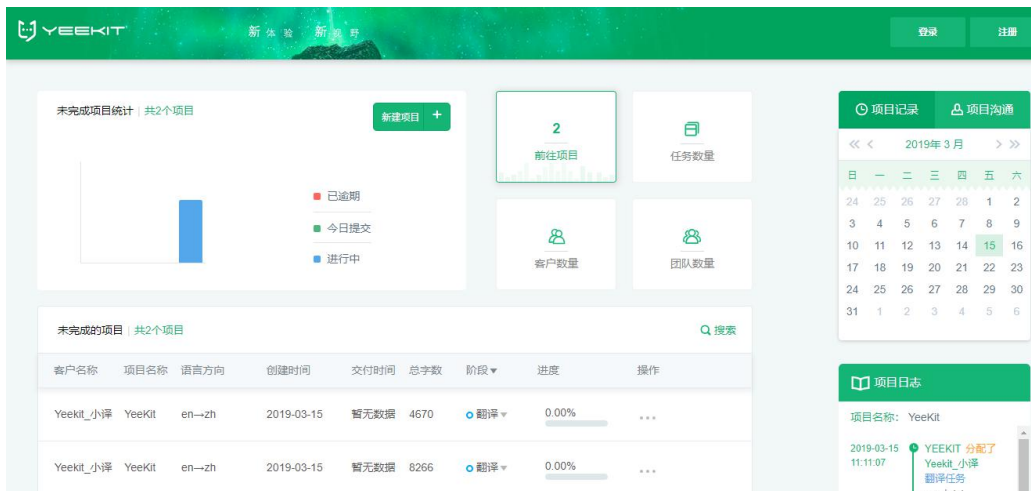
部参数数量庞大导致模型泛化能力较差的问题，以解决翻译与训练集的分布差异较大的句子，得到的译文往往不理想的问题。

- 在训练中使用了退火式的 Adam 优化策略。结合 Adam 的快速收敛以及自适应地选择学习速率的优点，以及吸取随机梯度下降优化算法的优势，使用退火式的 Adam 优化策略，减半学习速率并初始化训练，以达到更优的结果。

3、产品介绍

● CAT 辅助翻译系统

CAT 辅助翻译平台是以中译语通基于 40 多年翻译和管理经验，结合目前先进的人工智能技术打造的一款行业领先的计算机辅助翻译平台。它完美的结合了高效的机器翻译、精准的记忆库、术语库、机器翻译工具，并融入翻译生产管理流程，为项目经理、译员提供了完善的生产管理和辅助翻译工具，极大的提高了翻译质量和翻译效率、降低了翻译成本。Yeekit CAT 提供账号服务和本地化部署两种服务模式。



● 文档翻译系统

该系统主要解决具有大量文档需要翻译的用户需求。用户一键上传文件，系统自动翻译，一键原格式译文导出。系统支持的文件类型包括 word、excel、PDF、PPT 等 60 余种。支持引擎类型：通用翻译引擎、领域翻译引擎。

序号	文件名	字数	进度
1	新Web前端.docx	994	<div style="width: 100%;"></div>
2	中译语通公司介绍.docx	1659	<div style="width: 100%;"></div>
3	第一章 概述.docx	-	<div style="width: 0%;"></div>

文档字数 2653 单价 ¥ 1 /千字 支付金额 ¥ 2.65

支持docx(建议)、doc文件格式, 最多10个文件, 单个文件小于20MB

同意并提交订单 我已阅读并同意《文档翻译服务协议》

中译语通科技股份有限公司 Chinese(Simplified)
Copyright 2018 www.yeekit.com All rights reserved. 京ICP备13002826号-6

● 字幕通

字幕通软件是一款针对企业内大量音频和视频文件进行自动翻译的产品。用户可将视频文件上传后系统可自动实现时间轴切分、原文识别、字幕翻译、字幕上屏、成品导出的一站式操作服务。支持 10 种语言：中文，英语，德语，西班牙语，法语，阿拉伯语，意大利语，日语，葡萄牙语，俄语。支持输入音视频格式：mp4、WAV 等，输出视频格式：mp4、mkv、avi、wmv、flv、mpeg、asf，输出的字幕格式：srt、ass、json、txt。

● LanguageBox 智能会议同传终端

LanguageBox 是中译语通（GTCOM）基于语音识别、机器翻译等 AI 技术，面向企业级用户打造的一款会议机器翻译系统。可应用于产品/战略发布会、国际会议/跨国企业会议、外语教学等不同场景；同时还提供人工语音呼叫译员等个性化语言科技解决方案。产品支持会后及时导出会议纪要，可用于演讲、访谈、办公会议等场景下。LanguageBox 提供互联网服务和本地化部署两种服务模式。



● 会议转写系统

会议转写系统是中译语通基于语音识别、分角色识别等 AI 技术打造的一款智能会议转写系统。实现会议发言内容实时转写，形成分角色的会议纪要。转写文本文件分句时间戳对应音频文件，语音文件存储。通过音频处理器区分不同发言人的发言内容，进行角色区分。语音识别完成后，用户可在转写文本内容上进行实时修改校对，对相关领域的词语进行文本训练。会议转写系统提供互联网服务和私有化部署两种服务模式。

● 多语种机器翻译引擎

MerCube 是中译语通推出的一款企业级人工智能一体机，将机器翻译、语音识别、图像识别、NLP 算法、知识图谱等人工智能算法及引擎内置其中，为医疗、金融、科技、政务等各领域提供“**All In One**”的解决方案。在中、英、俄、法、德、日、韩、西、葡、阿等 38 个语种进行语言定制化服务。MerCube 提供专属的私有化部署方式，企业级用户的定制化服务，结合用户数据及自身通用数据，提供专属的机器翻译引擎。

