CCMT 2022

第十八届全国机器翻译大会

The 18th China Conference on Machine Translation

8月6日-8月10日 西藏·拉萨

主办单位：中国中文信息学会

承办单位：西藏大学

会议手册

赞助商

钻石

火山翻译 GTCOM

白金

小牛翻译 NiuTrans.com　有道 youdao　微信AI　HUAWEI Translate

金牌

oppo　MAGIC DATA Data Set Your Mind　Baidu 百度　云译科技 CloudTranslation　mi

银牌

新译科技 NewTranx Tech

目录

一、会议信息

1. 会议介绍

第十八届全国机器翻译大会（The 18th China Conference on Machine Translation, CCMT2022）将于 2022 年 8 月 6 日（星期六）至 8 月 10 日（星期三）在西藏拉萨举行。本届会议由中国中文信息学会主办，西藏大学承办。CCMT 旨在为国内外机器翻译界同行提供一个交互平台，加强国内外同行的学术交流，召集各路专家学者针对机器翻译的理论方法、应用技术和评测活动等若干关键问题进行深入的研讨，为促进中国机器翻译事业的发展，起到积极的推动作用。

会议已连续成功召开了十七届（前十四届名为全国机器翻译研讨会 CWMT）。其中，共组织过十一次机器翻译评测，一次开源系统模块开发（2006）和两次战略研讨（2010、2012）。这些活动对于推动我国机器翻译技术的研究和开发产生了积极而深远的影响。因此，CCMT 已经成为我国自然语言处理领域颇具影响力的学术活动。

除学术论文报告外，本次会议将会邀请国内外知名专家进行特邀报告，面向学生和青年学者举行专题讲座，邀请学界和产业界专家举行专题讨论会，面向研究者和用户进行系统展示等，通过丰富多彩的形式和与会者互动探讨机器翻译最炽热的研究论点，揭示机器翻译最

前沿的蓝图。同时，CCMT2022 继续组织机器翻译评测，包括机器翻译双语翻译（汉英、英汉、维汉、藏汉和蒙汉），多语言翻译（汉、日、英），低资源语言翻译（中俄、俄中、中泰、泰中、中越、越中），自动译后编辑（汉英、英汉）和翻译质量自动评估（汉英、英汉）等多个任务。会上也会就评测工作进行学术交流和专题讨论。

2. 程序委员会成员介绍

大会主席： 赵安邦（西藏大学） 陈家骏（南京大学）

大会副主席： 陈天禄（西藏大学） 平措达吉（西藏大学）
格桑多吉（西藏大学）

程序委员会主席： 肖桐（东北大学） Juan Pino (Meta AI)

评测委员会主席： 杨雅婷（中科院新疆理化所）
杨沐昀（哈尔滨工业大学）

组织委员会主席： 尼玛扎西（西藏大学）
张家俊（中科院自动化所）

大会秘书长： 程建（西藏大学，电子科技大学）

讲习班主席： 黄书剑（南京大学） 涂兆鹏（腾讯）

学生论坛主席： 李茂西（江西师范大学） 周浩（字节跳动）

前沿趋势论坛主席： 王瑞（上海交通大学） 何中军（百度）

产业应用论坛主席： 冯冲（北京理工大学） 黄国平（腾讯）

出版主席： 谭知行（清华大学）
陈科海（哈尔滨工业大学（深圳））

赞助主席： 杨浩（华为） 李响（小米）

宣传主席： 李军辉（苏州大学） 魏勇鹏（语智云帆）
高定国（西藏大学）

程序委员会委员：（按姓名首字母排序）

曹海龙（哈尔滨工业大学） 陈科海（哈尔滨工业大学（深圳））

陈毅东（厦门大学） 陈钰枫（北京交通大学）

杜金华（华为伦敦研究所） 杜权（小牛翻译）

段湘煜（苏州大学） 冯骁骋（哈尔滨工业大学）

高盛祥（昆明理工大学） 贡正仙（苏州大学）

郭俊良（微软亚洲研究院） 胡博杰（腾讯）

黄国平（腾讯） 黄辉（澳门大学）

黄书剑（南京大学） 李军辉（苏州大学）

李良友（华为诺亚方舟实验室） 李茂西（江西师范大学）

李响（小米 AI 实验室） 李亚超（西北民族大学）

刘乐茂（腾讯） 刘群（华为诺亚方舟实验室）

刘洋（清华大学） 毛存礼（昆明理工大学）

孟凡东（腾讯） 米海涛（腾讯美国）

苏劲松（厦门大学）

谭旭（微软亚洲研究院）

涂兆鹏（腾讯）

王龙跃（腾讯）

王明轩（字节跳动）

王强（同花顺）

王星（腾讯人工智能实验室）

邬昌兴（华东交通大学）

吴双志（字节跳动）

谢军（阿里巴巴达摩院）

徐金安（北京交通大学）

许鸿飞（郑州大学）

杨宝嵩（阿里巴巴达摩院）

杨沐昀（哈尔滨工业大学）

于恒（虾皮公司）

张飚（爱丁堡大学）

张春良（东北大学）

张大鲲（Systran）

张家俊
（中国科学院自动化研究所）

张文（小米 AI 实验室）

朱慕华（美团）

Toshiaki Nakazawa
（The University of Tokyo）

Yves Lepage （Waseda University）

3. 会议线上直播信息

本次 CCMT 会议采用线下会议加线上同步直播的形式。线上直播平台有：腾讯会议、b 站和微博三种，具体时间安排见会议日程。

腾讯会议：会议号：508-888-802

会议主题：第十八届全国机器翻译大会（CCMT 2022）

会议时间：2022/08/08 08:00-20:00 (GMT+08:00) 中国标准时间 - 北京

重复周期：2022/08/08-2022/08/10 08:00-20:00, 每天

b 站直播间：UID:1397003059

b 站二维码

微博直播间：

微博二维码

二、会议日程

会议地点：拉萨阳光酒店多功能厅

前沿技术讲习班

2022 年 8 月 8 日

9:30-12:00 讲习班 1：自然语言处理中的神经网络结构设计与学习（肖桐 李垠桥 李北）

15:30-17:00 讲习班 2：机器翻译中的信息增强（王龙跃 王星）

21:00-22:30 专委会会议（地点：拉萨阳光酒店大会议室）

第十八届全国机器翻译大会

2022 年 8 月 9 日 - 2022 年 8 月 10 日

2022 年 8 月 9 日

10:00-10:40 开幕式

10:40-10:50 合影

10:50-11:50 特邀报告 1（主持人：肖桐）

AI 赋能内容创作 （段楠）

11:50-12:10 中间休息

12:10-13:25 中文论文报告（主持人：李军辉）

13:25-15:00 午间休息

15:00-16:50 机器翻译前沿趋势论坛（主持人：王瑞、何中军）

16:50-17:20 中间休息、poster 展示（主会论文）

17:20-19:10 机器翻译产业应用论坛（主持人：黄国平、冯冲）

2022 年 8 月 10 日

10:30-12:00 评测报告（主持人：杨雅婷）

12:00-12:30 中间休息、poster 展示（评测论文）

12:30-13:45 英文论文报告（主持人：黄书剑）

13:45-15:00 午间休息

15:00-16:00 特邀报告 2（主持人：张家俊）

神经机器翻译语言迁移和预训练中语言资源的极致利用（刘群）

16:00-16:30 中间休息、poster 展示（评测论文）

16:30-18:30 机器翻译学生论坛（主持人：李茂西、周浩）

18:30-18:40 中间休息

18:40-19:00 闭幕式

三、特邀报告

特邀报告 1：AI 赋能内容创作

主讲人：段楠

报告时间：2022 年 8 月 9 日 10:50-11:50

内容摘要：在内容即核心竞争力的时代，少数专业的内容创作者已经很难满足人们对多样化和个性化内容的巨大需求。如何为内容创造者赋能、降低内容创作的门槛和开销、提升内容创作者的生产力和创造力，已经成为人工智能领域中一个重要的前沿课题。本报告将介绍微软亚洲研究院在 AI 赋能内容开发和创作上的若干最新研究成果。通过自然语言处理、编程语言处理以及多模态内容理解和生成等研究，我们希望人工智能技术能够极大地降低内容创作的门槛，从根本上改变内容创作的方式，使得人人都有机会成为优质内容的高效开发者和

创作者。

主讲人介绍：段楠博士，微软亚洲研究院高级研究经理，中国科学技术大学兼职博导，天津大学兼职教授，CCF 杰出会员，主要从事自然语言处理、编程语言处理、多模态人工智能、机器推理等研究，多次担任 NLP/AI/ML 相关国际会议评测主席、高级领域主席和领域主席，发表学术论文 100 余篇，持有专利 20 余项，多项研究成果用于微软各类产品。

特邀报告 2：神经机器翻译语言迁移和预训练中语言资源的极致利用

主讲人：刘群

报告时间：2022 年 8 月 10 日 15:00-16:00

内容摘要：语言资源的缺乏是神经机器翻译面临的关键问题之一。现有的各种语言迁移和多语言预训练方法均无法充分利用所有各种不同形式的语言资源。本报告将介绍我们最近几篇论文的工作，着重研究了在从高资源语言到低资源语言迁移以及多语言预训练等场景下如何最大限度地利用各种语言资源的问题。我们针对双重语言迁移、三角语言迁移、多语言预训练等不同场景，提出了多种语言迁移和预训练方法，最大限度地利用了各种不同形式的语言资源，以提高目标系统的性能。实验表明，我们提出的方法在这些场景下均显著超越了已有方法，取得了最好的训练效果并达到了最优的系统性能。

主讲人介绍：刘群，博士，教授，ACL Fellow，华为语音语义首席科学家，负责语音和自然语言处理研究。原爱尔兰都柏林城市大学教授、爱尔兰 ADAPT 中心自然语言处理主题负责人、中国科学院计算技术研究所研究员、自然语言处理研究组负责人。分别在中国科学技术大学、中科院计算所、北京大学获得计算机学士、硕士和博士学位。研究方向主要是自然语言理解、语言模型、机器翻译、问答、对话等。研究成果包括汉语词语切分和词性标注系统、基于句法的统计机器翻译方法、预训练语言模型的训练、压缩与应用等。承担或参与多项中国、爱尔兰和欧盟大型科研项目。在国际会议和期刊发表论文 300 余篇，被引用 12000 多次。培养国内外博士硕士 50 多人。获 Google Research Award、ACL Best Long Paper、钱伟长中文信息处理科学技术奖一等奖、国家科技进步二等奖等奖项。

四、前沿技术讲习班

讲习班 1：自然语言处理中的神经网络结构设计与学习

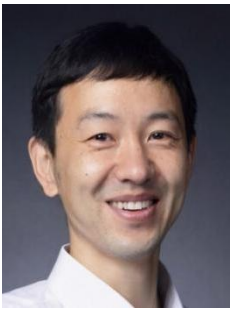
主讲人：肖桐 李垠桥 李北

讲习班时间：2022 年 8 月 8 日 9:30-12:00

内容摘要：人工神经网络方法已经成为了自然语言处理中最重要的范式之一。但是，这类方法大量依赖人工设计的神经网络结构，导致自然语言处理领域的发展很大程度依赖于神经网络结构上的突破。由于神经网络结构设计大多源自研究人员的灵感和大量经验性尝试，如何挖掘这些神经网络结构背后的逻辑，如何系统化的思考不同神经网络结构之间的内在联系，是使用这类方法时所需要深入考虑并回答的问题。甚至，可以想象，让计算机自动设计神经网络架构，也可以成为进一步突破人类思维限制的方向之一。在讲习班中，我们将会对上述问题进行回答，对神经网络架构的基本发展脉络、常用的神经网络架构的设计理念进行分析，同时对神经网络架构的自动设计方法进行整理。这些内容可以为相关研究者提供模型架构设计上的一些思路，以及实践中的参考。同时，我们也希望通过这次讲习班的内容，呼吁更多地以系统化的思考方式来看待神经网络方法在自然语言处理中的应用，而非简单像“黑盒”一样使用它们。

主讲人介绍：

1. 肖桐



肖桐，博士，东北大学教授、博士生导师，东北大学计算机学院人工智能系系主任，东北大学自然语言处理实验室主任，小牛翻译（NiuTrans）联合创始人。于东北大学计算机专业获得博士学位。2006—2009 年赴日本富士施乐、微软亚洲研究院访问学习，并于 2013—2014 年赴英国剑桥大学开展博士后研究。主要研究领域包括自然语言处理、机器翻译等。在国内外相关领域高水平会议及期刊上发表学术论文 70 余篇，并撰写专著《机器翻译：基础与模型》。作为项目技术负责人，成功研发了 NiuTrans、NiuTensor 等开源系统，在 WMT、CCMT/CWMT、NTCIR 等国内外评测中多次获得冠军。2014 年获得中国中文信息学会首届优秀博士论文提名奖，2016 年获得中国中文信息学会“钱伟长中文信息处理科学技术奖”一等奖，2021 年获得中国计算机学会 CCF-NLP 青年新锐奖。任 ACL、EMNLP、AAAI

等国际著名会议及期刊的领域主席、高级程序委员会委员，并多次获得 ACL、NAACL 等会议的 Outstanding Reviewer、Outstanding Action Editor。

2. 李垠桥

李垠桥，东北大学自然语言处理实验室博士生。于东北大学计算机专业获得硕士学位。在 ACL、IJCAI 等国内外相关领域高水平会议期刊上发表学术论文十余篇，参与研发 NiuTensor 开源系统。参与 WMT、CCMT/CWMT 等机器翻译评测任务，作为哈萨克-英语、英语-哈萨克语赛道负责人在 WMT19 评测自动评价中获得双向第一名。任中国中文信息学会青年工作委员会学生委员。多次担任 ACL、AAAI、ICML 等国内外会议期刊审稿人。主要研究方向为神经网络结构搜索、语言建模以及机器翻译等。

3. 李北

李北，东北大学自然语言处理实验室三年级博士生。于东北大学计算机专业获得学士与硕士学位。研究方向涉及机器翻译、篇章级翻译、多模态建模等任务，对复杂网络结构设计与优化抱有浓厚兴趣，以第一作者在 ACL、EMNLP、ICML、AAAI、WMT 等领域内高水平会议上发表多篇学术论文，参与实验室多项开源系统的研发，目前在微软亚研院 NLC 组进行文本图像生成相关的研究。曾多次参与 WMT、CCMT/CWMT 等机器翻译评测任务，取得多项翻译赛道第一名成绩。多次担任 ACL、EMNLP、AAAI、ICML、Neurips 等领域内顶会审稿人，曾荣获 EMNLP2021 杰出审稿人。

讲习班 2：机器翻译中的信息增强

主讲人：王龙跃 王星

讲习班时间：2022 年 8 月 8 日 15:30-17:00

内容摘要：近年来，基于神经网络的机器翻译研究飞速发展，翻译模型性能得到显著提升。但是，落地系统仍面临着资源不足、知识缺失等一系列问题。值得注意的是，信息增强方法可以有效缓解以上问题，在模型演进过程中发挥了重要作用。本报告首先从翻译系统的实际问题展开，分别从数据增强和知识融合两个方面介绍近年来的研究进展。

主讲人介绍：

1. 王龙跃

王龙跃，腾讯人工智能实验室资深研究员。2018 年获都柏林城市大学博士学位，并获欧洲机器翻译学会最佳博士论文奖。主要从事

机器翻译、篇章分析、预训练、自然语言处理等方向研究。在 ICLR、ACL 等国际期刊和会议上发表论文四十余篇，参加 WMT 等国际学术比赛十余次并取得优异成绩。担任 2021 腾讯 AI Lab 犀牛鸟专项研究计划项目负责人、中文信息学会青年工作委员会委员、ACL 领域主席。

2. 王星

王星，腾讯公司人工智能实验室（Tencent AI Lab）高级研究员，2018 年博士毕业苏州大学，主要从事机器翻译相关的研发工作。在 ACL、EMNLP、NAACL、AAAI 等自然语言处理相关会议和刊物上发表论文约三十篇，在 WMT 国际翻译评测多个赛道获得评测比赛第一名，任中国中文信息学会青年工作委员会委员。

五、主会计划

2022 年 8 月 9 日 - 2022 年 8 月 10 日，共两天。

1. 日程总览

2022 年 8 月 9 日		
10:00-10:40	开幕式	
	10:00-10:10	学会领导致辞
	10:10-10:20	承办单位领导致辞
	10:20-10:30	专委会领导致辞
	10:30-10:35	大会主席致辞
	10:35-10:40	程序委员会介绍会议情况
10:40-10:50	合影	
10:50-11:50	特邀报告 1（主持人：肖桐） AI 赋能内容创作 （段楠）	
11:50-12:10	中间休息	
12:10-13:25	中文论文报告（主持人：李军辉 ）	
	12:10-12:25	融合 La 格虚词语义信息的藏文 La 格分类模型（班玛宝，慈祯嘉措，张瑞，才让加）
	12:25-12:40	军政领域社交媒体英汉双语平行语料库构建方法（夏榕璟，张克亮，唐亮，李铭）
	12:40-12:55	同源语料增强的低资源神经机器翻译（王琳, 刘伍颖）
	12:55-13:10	基于降噪原型序列的汉越神经机器翻译（杨汉清，赖华，于志强，余正涛）

	13:10-13:25	神经机器翻译质量与译后编辑实证研究（裘白莲，王明文，罗琪，李茂西）
13:25-15:00	午间休息	
15:00-16:50	机器翻译前沿趋势论坛（主持人：王瑞、何中军）	
	15:00-15:25	无监督机器翻译及其应用 （刘树杰）
	15:25-15:50	低资源机器翻译的研究现状 （陈科海）
	15:50-16:50	圆桌讨论：低资源机器翻译（主持人：王瑞、何中军 嘉宾：刘树杰、陈科海、王明轩、王星）
16:50-17:20	中间休息、poster 展示 1（主会论文）	
	1	PEACook: Post-Editing Advancement Cookbook （Shimin Tao, Jiaxing Guo, Yanqing Zhao, Min Zhang, Daimeng Wei, Minghan Wang, Hao Yang, Miaomiao Ma, and Ying Qin）
	2	Hot-start Transfer Learning combined with Approximate Distillation for Mongolian-Chinese Neural Machine Translation （Pengcong Wang, Hongxu Hou, Shuo Sun, Nier Wu, Weichen Jian, Zongheng Yang, and Yisong Wang）
	3	Life Is Short, Train It Less: Neural Machine Tibetan-Chinese Translation Based on mRASP and Dataset Enhancement （Hao Wang, Yongbin Yu, Nyima Tashi, RinchenDongrub, Ekong Favour, Mengwei Ai, Kalzang Gyatso, Yong Cuo, and Qun Nuo）

	4	藏文虚词 BPE 的藏汉机器翻译方法研究（严松思，珠杰，汪超，刘亚姗，许泽洲，徐泽辉）
	5	自然场景下藏文检测识别数据集与方法（侯琴，胡永祥，刘思宇，尼玛扎西，程建）
	6	基于语料库的古藏文文献字符统计研究（三智多杰，祁坤钰，久仙加）
	7	多引擎机器翻译译文重排序与融合研究（李铭，张克亮，唐亮，夏榕璟）
	8	基于 LSTM 和 CRF 的藏文分词模型（于永斌，陆瑞军，尼玛扎西，群诺，王昊，唐倩，彭辰辉，吕诗仪，项秀才让）
17:20-19:10	机器翻译产业应用论坛（主持人：黄国平、冯冲 嘉宾：陶仕敏（华为）、李响（小米）、王明轩（火山翻译）、张春良（小牛翻译）、孟凡东（微信翻译）、薛征山（OPPO）、宗浩（中译语通）、何中军（百度）、丁丽（云译）、刘劲松（新译）、于超（Magic Data））	
	17:20-18:10	圆桌讨论 1：机器翻译下一阶段的产业落地——瓶颈和突破 （主持人：黄国平）
	18:10-18:20	中场休息
	18:20-19:10	圆桌讨论 2：从工业界角度看研究生培养——需求与建议 （主持人：冯冲）

2022 年 8 月 10 日		
10:30-12:00	评测报告（主持人：杨雅婷）	
	10:30-10:50	2022 年度评测情况汇报（杨雅婷）
	10:50-11:05	维沃移动通信有限公司
	11:05-11:20	华为文本机器翻译实验室
	11:20-11:28	新疆大学计算机系
	11:28-11:36	哈尔滨工业大学（深圳）
	11:36-11:44	潍坊北大青鸟华光照排有限公司
	11:44-11:52	苏州大学
	11:52-12:00	中国科学院计算技术研究所
12:00-12:30	中间休息、poster 展示 2（评测论文）	
	1	西藏大学藏文信息技术人工智能重点实验室 CCMT2022 评测论文
	2	流利说 CCMT2022 评测论文
	3	新疆大学信息科学与工程学院 CCMT2022 评测论文
	4	北京理工大学 CCMT2022 评测论文
	5	内蒙古大学 CCMT2022 评测论文
	6	南京大学 CCMT2022 评测论文
	7	广西大学 CCMT2022 评测论文
	8	广东工业大学 CCMT2022 评测论文
12:30-13:45	英文论文报告（主持人：黄书剑）	
	12:30-12:45	Review-based Curriculum Learning for Neural Machine Translation （Ziyang Hui, Chong Feng and Tianfu Zhang）
	12:45-13:00	Target-side Language Model for Reference-free Machine Translation Evaluation （Min Zhang, Xiaosong Qiao,

		Hao Yang, Shimin Tao, Yanqing Zhao, Yinlu Li, Chang Su, Minghan Wang, Jiaxin Guo, Yilun Liu, and Ying Qin）
	13:00-13:15	Improving the Robustness of Low-Resource Neural Machine Translation with Adversarial Examples（ Shuo Sun, Hongxu Hou, Nier Wu, Zongheng Yang, Yisong Wang, Pengcong Wang, and Weichen Jian）
	13:15-13:30	Dynamic Mask Curriculum Learning for Non-Autoregressive Neural Machine Translation （Yisong Wang, Hongxu Hou, Shuo Sun, Nier Wu, Weichen Jian, Zongheng Yang, and Pengcong Wang）
	13:30-13:45	Dynamic Fusion Nearest Neighbor Machine Translation via Dempster–Shafer Theory （Zongheng Yang, Hongxu Hou, Shuo Sun, Nier Wu, Yisong Wang, Weichen Jian, and Pengcong Wang）
13:45-15:00	午间休息	
15:00-16:00	特邀报告 2（主持人：张家俊 ） 神经机器翻译语言迁移和预训练中 语言资源的极致利用 （刘群）	
16:00-16:30	中间休息、poster 展示 3（评测论文）	
	1	西藏大学信息科学技术学院 CCMT2022 评测论文
	2	郑州大学 CCMT2022 评测论文
	3	西北民族大学 CCMT2022 评测论文

	4	中国科学技术信息研究所 CCMT2022 评测论文
	5	北京航空航天大学 CCMT2022 评测论文
	6	湖南大学 CCMT2022 评测论文
	7	昆明理工大学信息工程与自动化学院 CCMT2022 评测论文
16:30-18:30	机器翻译学生论坛（主持人：李茂西、周浩）	
	16:30-16:45	科研路上的“屠龙刀”们 （鲍宇）
	16:45-17:00	从业务需求中发现科研问题：记我的一段实习经历 （刘鑫）
	17:00-17:15	我在科研道路上走过的弯路 （许晨）
	17:15-17:30	NMT 可解释性：科研“蓝海”的思考和摸索 （卢宇）
	17:30-17:45	研究初期的个人学术品牌建设 （王龙跃）
	17:45-18:00	我科研历程中的“弯路” （王明轩）
	18:00-18:30	Panel：博士生涯如何“standing out”（嘉宾：鲍宇、刘鑫、许晨、卢宇、王龙跃、王明轩）
18:30-18:40	中间休息	
18:40-19:00	闭幕式	

2. 中文论文报告

报告时间：2022 年 8 月 9 日 12:10-13:25

融合 La 格虚词语义信息的藏文 La 格分类模型

报告时间：2022 年 8 月 9 日 12:10-12:25

班玛宝，慈祯嘉措，张瑞，才让加

内容摘要：采用深度学习方法实现藏文 La 格(ལ་དོན།)分类是一项具有挑战性和重要研究意义的藏语自然语言处理任务。藏文 La 格的自动分类更加依赖于上下文语义信息和特征的时序性，该文通过分析 La 格虚词的语义特征及用法，在设计 La 格虚词语义信息标记算法的基础上，提出一种融合 La 格虚词语义信息的藏文 La 格分类模型。该模型首先以每个音节及对应 La 格虚词或其它音节的语义特征嵌入作为输入，丰富嵌入向量的语义信息，增加输入特征的多样性；然后采用一维卷积融合并学习每个音节及对应 La 格虚词或其它音节语义信息的局部特征向量，提高卷积层的空间特征学习能力；其次使用双向 LSTM 学习时序特征，提高时序特征的学习能力；最后使用注意力机制对双向 LSTM 层每一时刻的输出特征进行加权融合，充分利用每一时刻的输出特征，以提高最终文本表示的特征质量。经在 TLD 数据集上实验显示，该模型的分类效果比基线模型和仅用藏文音节嵌入的模型均有所提升，在测试集上的分类准确率为 93.10%。

军政领域社交媒体英汉双语平行语料库构建方法

报告时间：2022 年 8 月 9 日 12:25-12:40

夏榕璟，张克亮，唐亮，李铭

内容摘要：高质量的机器翻译系统的研究离不开高质量的平行语料库。随着机器翻译技术的发展，平行语料库研究的重要性也愈加凸显。构建军政领域社交媒体语料库，对于军政领域社交媒体语言特点和军政领域社交媒体机器翻译任务具有重要意义。通过分析现有机器翻译引擎对军政领域社交媒体语料的翻译中存在的用户名和标签的不译、网络非正规表达的不译等翻译问题，提取军政领域社交媒体语料特征，在利用翻译引擎的基础上，优化了流畅度，BLEU 值提升 1.37 个百分点，最终构建规模约为 20 万句对、用于机器翻译任务的军政领域社交媒体英汉双语句对齐平行语料库。

同源语料增强的低资源神经机器翻译

报告时间：2022 年 8 月 9 日 12:40-12:55

王琳, 刘伍颖

内容摘要：缺少平行句库的低资源机器翻译面临跨语言语义转述科学问题。围绕具体的低资源印尼语-汉语机器翻译问题，探索了基于同源语料的语言资源扩建方法，并混合同源语料训练出神经机器翻译模型。这种混合语料模型在印尼语-汉语机器翻译实验中取得了 20.30

的 BLEU4 评分。对实验结果的人工简单随机抽样分析发现混合语料模型的机器翻译效果与同时期的谷歌翻译效果相当。对于某些句子混合语料模型的译文质量甚至更优。实验结果证明同源语料能够有效增强低资源神经机器翻译，而这种有效性主要是源于同源语言之间的形态相似性和语义等价性。

基于降噪原型序列的汉越神经机器翻译

报告时间：2022 年 8 月 9 日 12:55-13:10

杨汉清，赖华，于志强，余正涛

内容摘要：原型序列旨在用目标端语言信息指导机器翻译，已有的工作主要是利用相似性翻译作为目标端原型序列,提升神经机器翻译的性能。然而在汉越低资源场景下,平行语料匮乏,原型序列蕴含庞杂的信息,直接使用会增加翻译模型训练的难度，甚至引入噪声。针对此问题，本文提出了一种基于降噪原型序列的汉越神经机器翻译方法。首先，利用跨语言检索得到原型序列；其次，基于实体词典对原型序列中的噪声信息进行掩盖，再综合稀有词词频及语义相似度，得到原型序列的参考价值；最后使用额外的编码器接收原型序列，并允许解码器到两个编码器间建立注意力机制。实验结果表明,相比基线模型,本文所提出的方法能够有效提升汉越神经机器翻译的性能。

神经机器翻译质量与译后编辑实证研究

报告时间：2022 年 8 月 9 日 13:10-13:25

裘白莲，王明文，罗琪，李茂西

内容摘要：随着神经机器翻译质量的提高，机器翻译+译后编辑的工作模式在翻译行业越来越普遍。本文针对神经机器翻译质量对译后编辑过程的影响进行了研究，考察机器翻译质量、源文特征对译后编辑努力的影响，同时观察、比较职业译员和学生译员译后编辑过程的异同。实证研究发现机器译文质量影响译后编辑时间努力和技术努力，而源文句长对时间努力没有影响。职业译员译后编辑速度快于学生译员，其所做的译后编辑操作少于学生译员。汉英语言对神经机器翻译译后编辑实证研究有助于促进用户角度的机器翻译研究，深化对译后编辑过程的了解，同时为译后编辑教学提供有效参考。

3. poster 展示 1——主会论文

展示时间：2022 年 8 月 9 日 16:50-17:20

PEACook: Post-Editing Advancement Cookbook

Shimin Tao, Jiaxing Guo, Yanqing Zhao, Min Zhang, Daimeng Wei, Minghan Wang, Hao Yang, Miaomiao Ma, and Ying Qin

Abstract. Automatic post-editing (APE) aims to improve machine translations, thereby reducing human post-editing efforts. Training on APE models has made a great progress since 2015; however, whether APE models are really performing well on domain samples remains as an open question, and achieving this is still a hard task. This paper provides a mobile domain APE corpus with 50.1 TER/37.4 BLEU for the En-Zh language pair. This corpus is much more practical than that provided in WMT 2021 APE tasks (18.05 TER/71.07 BLEU for En-De, 22.73 TER/69.2 BLEU for En-Zh). To obtain a more comprehensive investigation on the presented corpus, this paper provides two mainstream models as the Cookbook baselines: (1) Autoregressive Translation APE model (AR-APE) based on HW-TSC APE 2020, which is the SOTA model of WMT 2020 APE tasks. (2) Non-Autoregressive Translation APE model (NAR-APE) based on the well-known Levenshtein Transformer. Experiments show that both the mainstream models of AR and NAR can effectively improve the effect of APE. The corpus has been released in the CCMT 2022 APE evaluation task and the baseline models will be open-sourced.

Hot-start Transfer Learning combined with Approximate Distillation for Mongolian-Chinese Neural Machine Translation

Pengcong Wang, Hongxu Hou, Shuo Sun, Nier Wu, Weichen Jian, Zongheng Yang, and Yisong Wang

Abstract. When parallel training data is scarce, it will affect neural machine translation. For low-resource neural machine translation (NMT), transfer learning is very important, and the use of pre-training model can also alleviate the shortage of data. However, the good performance of common cold-start transfer learning methods is limited to the cognate language realized by sharing its vocabulary. Moreover, when using the pre-training model, the combination of general fine tuning methods and NMT will lead to a serious problem of knowledge forgetting. Both methods have some defects, so this paper optimizes the above two problems, and applies a new training framework suitable for low correlation language to Mongolian-Chinese neural machine translation. Our framework includes two technologies: a) word alignment method under hot-start, which alleviates the problem of word mismatch between the transferred subject and object in transfer learning. b) approximate distillation, not only retains the pre-trained knowledge, but also solves the forgetting problem, so that the encoder of NMT has stronger language representation ability. The results show that BLEU is increased by 3.2, which is better than ordinary transfer learning and multilingual translation system.

Life Is Short, Train It Less: Neural Machine Tibetan-Chinese Translation Based on mRASP and Dataset Enhancement

Hao Wang, Yongbin Yu, Nyima Tashi, RinchenDongrub, Ekong Favour, Mengwei Ai, Kalzang Gyatso, Yong Cuo, and Qun Nuo

Abstract. This paper highlights a multilingual pre-trained neural machine translation architecture as well as a dataset augmentation approach based on curvature selection. The multilingual pre-trained model is designed to increase the performance of machine translation with low resources by bringing in more common information. Instead of repeatedly training several checkpoints from scratch, this study proposes a checkpoint selection strategy that uses a cleaned optimizer to hijack a midway status. Experiments with our own dataset on the Chinese-Tibetan translation demonstrate that our architecture gets a 32.65 BLEU score, while in the reverse direction, it obtains a 39.51 BLEU score. This strategy drastically reduces the amount of time spent training. To demonstrate the validity of our method, this paper shows a visualization of curvature for a real-world training scenario.

藏文虚词 BPE 的藏汉机器翻译方法研究

严松思，珠杰，汪超，刘亚姗，许泽洲，徐泽辉

内容摘要：本文针对藏文虚词的文法特点，设计了基于藏文虚词的 BPE 方法，首先通过全部藏文虚词 BPE、过滤兼类虚词 BPE、单音节虚词 BPE 和多音节虚词 BPE，得到四种对应语料，其次将其在

Transformer 模型和 mBART 模型上进行了实验，使用轮数集成和不同网络结构集成来提高最终模型的泛化能力。对比实验证明，藏文虚词 BPE 算法与模型集成策略可以提升藏汉机器翻译的翻译效果，最高可以达到 38.05 个 BLEU。

自然场景下藏文检测识别数据集与方法

侯琴，胡永祥，刘思宇，尼玛扎西，程建

内容摘要：自然场景下藏文文本检测与识别是一项重要且有挑战性的任务。为解决藏文自然图像数据不足的问题，本文构建了一个乌金体藏文检测识别数据集，收集藏区多个城市中包含藏文的 1320 张自然图像，共标注图像中 1979 条文本的位置与内容。同时，本文提出了自然场景下藏文检测与识别方法，使用可微二值化网络实现藏文文本检测，提出基于时序卷积网络的藏文识别框架。为进一步扩充数据，本文提出自然场景藏文图像合成方法，在图像的合适位置渲染藏文文本，生成真实自然的藏文图像。本文在数据集上进行了藏文检测与识别实验，文本检测的 F1 分数达到 86.93%，文本识别的字符识别准确率达到 92.70%，实验表明本文提出的方法有较好的性能。

基于语料库的古藏文文献字符统计研究

三智多杰，祁坤钰，久仙加

内容摘要：本文以敦煌藏文文献为主，构建了古藏文文献标注语料库。在此基础上，应用 python 语言设计出古藏文频率统计软件，对古藏文和现代藏文的元音、辅音、音节频次等方面进行对比分析。归纳出古藏文字符及音节结构的分布特征，以期为古藏文标注语料库构建和藏文文字特征研究提供参考。

多引擎机器翻译译文重排序与融合研究

李铭，张克亮，唐亮，夏榕璟

内容摘要：使用不用的模型、方法、语种、数据构建的机器翻译引擎往往在不同的场景下具有不同的翻译效果。因此，很多研究者都在构建机器翻译引擎时尝试使用多引擎译文融合或多翻译方法融合的方式来利用不同翻译引擎的优点，然而过往的工作没有考虑到如何利用用户在使用多引擎机器翻译所产生的数据来获取存在于用户认知域中对这些引擎译文的评价。本文研究提出了基于六个翻译引擎的多引擎翻译平台。该平台在长期使用中产生了翻译结果、用户特征、人工校译等数据，本文基于以上大规模历史数据构建了翻译模型训练资源库，并结合 Page Rank 算法、贝叶斯公式和 UNQE 方法提出了多引擎机器翻译译文重排序方法，并利用译文重排序的结果与翻译模型训练

资源库中的翻译实例相关数据，进一步使用 Transformer 架构训练了译文融合模型。实验结果表明了本文提出的方法能够融合多引擎优势，提高不同领域的平均译文质量。

基于 LSTM 和 CRF 的藏文分词模型

于永斌，陆瑞军，尼玛扎西，群诺，王昊，唐倩，彭辰辉，吕诗仪，项秀才让

内容摘要：本文提出基于长短时记忆 （Long short-term memory，LSTM）神经网络和条件随机场（Conditional Random Field, CRF）的藏文分词模型。引入注意力机制，获取更多特征信息,提升模型关注上下文信息与当前音节之间联系；提出一种音节扩展方法，获取更多的输入特征信息与语料信息，增强模型单音节特征信息以获取更多语义信息的能力。本文在西藏大学数据集 12261 条的基础上，扩充至 74384 条，形成 Tibetan-News 数据集。实验结果表明，在模型中加入注意力机制并使用音节扩展方法后，模型在 Tibetan-News 数据集上的精确率、召回率和 F1 分别提升 2.9%、3.5%和 3.2%。基于本文模型的分词系统已在工程上应用推广。

4. 评测报告

报告时间：2022 年 8 月 10 日 10:30-12:00

单位：维沃移动通信有限公司

报告时间：2022 年 8 月 10 日 10:50-11:05

vivo AI 研究院 CCMT 2022 评测技术报告

李方圆，郭攀峰，狄亚超，王承之，丁家杰，李春元，滕飞

内容摘要：文本详细介绍了 vivo AI 研究院机器翻译团队参加第十八届全国机器翻译大会机器翻译评测（CCMT 2022）的参赛情况和各项任务采用的技术细节。在本次评测中，vivo 共参加了八个评测任务，分别是汉英新闻领域机器翻译（CE）、英汉新闻领域机器翻译（EC）、蒙汉日常用语领域机器翻译（MC）、藏汉政府文献领域机器翻译（TC）和维汉新闻领域机器翻译 5 个双语翻译评测任务，日英汉专利领域多语言翻译评测任务，以及低资源场景下的中泰、泰中语向机器翻译评测任务。报告将主要阐述本次参评系统采用的算法模型框架、数据处理、数据筛选和数据增强用到的方法，并分别给出在不同的方法配置下参评系统的评测结果对比与分析。

单位：华为文本机器翻译实验室

报告时间：2022 年 8 月 10 日 11:05-11:20

Multi-Strategy Enhanced Neural Machine Translation for Chinese Minority Languages

Zhanglin Wu, Daimeng Wei, Xiaoyu Chen, Ming Zhu, Zongyao Li, Hengchao Shang, Jinlong Yang, Zhengzhe Yu, Zhiqiang Rao, Shaojun Li, Lizhi Lei, Song Peng, Hao Yang and Ying Qin

Abstract. This paper presents HW-TSC' s submissions to CCMT 2022 Chinese Minority Language Translation task. We participate in three language directions: Mongolian → Chinese Daily Conversation Translation, Tibetan→Chinese Government Document Translation, and Uighur→Chinese News Translation. We train our models using the Deep Transformer architecture, and adopt enhancement strategies such as Regularized Dropout, Tagged Back-Translation, Alternated Training, and Ensemble. Our enhancement experiments have proved the effectiveness of above-mentioned strategies. We submit enhanced systems as primary systems for the three tracks. In addition, we train contrast models using additional bilingual data and submit results generated by these contrast models.

HW-TSC Submission for CCMT 2022 Translation Quality Estimation Task

Chang Su, Miaomiao Ma, Hao Yang, Shimin Tao, Jiaxing Guo, Minghan Wang, Min Zhang, and Xiaosong Qiao

Abstract. This paper presents the method used by Huawei Translation

Services Center (HW-TSC) in the quality estimation (QE) task - sentence-level post-editing effort estimation - in the 18th China Conference on Machine Translation (CCMT) 2022. This method is based on a predictor-estimator model. The predictor is an XLM-RoBERTa model pre-trained on a large-scale parallel corpus and extracts features from the source language text and machine-translated text. The estimator is a fully connected layer that is used to regress the post-editing distance scores using the extracted features. In the experiment, it is found that pre-training the predictor with the semantic textual similarity (STS) task in the parallel corpus and using augmented training data constructed by different machine translation (MT) engines can improve the prediction effect of the Human-targeted Translation Edit Rate (HTER) in both Chinese-English and English-Chinese tasks.

单位：新疆大学计算机系

报告时间：2022 年 8 月 10 日 11:20-11:28

新疆大学 CCMT2022 英汉机器翻译评测任务技术报告

宜年，艾山·吾买尔，汪烈军

内容摘要：本文主要介绍新疆大学信息科学与工程学院在第十八届全国机器翻译大会机器翻译评测项目中参赛的基本情况。在本次机器翻译评测中，参加了英汉新闻领域机器翻译评测项目。本文主要阐述本次参赛的英汉神经机器翻译系统采用的模型框架、数据预处理过程、数据增强以及模型微调和集成等方法。最后给出不同方法和模型在测

试集上的性能表现，并进行对比和分析。

单位：哈尔滨工业大学（深圳）

报告时间：2022 年 8 月 10 日 11:28-11:36

An improved Multi-task Approach to Pre-trained Model Based MT Quality Estimation

Binhuan Yuan, Yueyang Li, Kehai Chen, Hao Lu, Muyun Yang and Hailong Cao

Abstract. Machine translation (MT) quality estimation (QE) aims to automatically predict the quality of MT outputs without any references. State-of-the-art solutions are mostly fine-tuned with a pre-trained model in a multi-task framework (i.e., joint training sentence-level QE and word-level QE). In this paper, we propose an alternative multi-task framework in which post-editing results are utilized for sentence-level QE over an mBART-based encoder-decoder model. We show that the post-editing sub-task is much more informative and the mBART is superior to other pre-trained models. Experiments on WMT2021 English-German and English-Chinese QE datasets showed that the proposed method achieves 1.2%-2.1% improvements in the strong sentence-level QE baseline.

单位：潍坊北大青鸟华光照排有限公司

报告时间：2022 年 8 月 10 日 11:36-11:44

华光机器翻译系统 CCMT2022 评测技术报告

殷建民，郑文亮，王淑珍

内容摘要：本文描述了华光机器翻译系统参加 2022 年全国机器翻译大会（CCMT 2022）翻译评测任务的情况。本次评测中，我们参与了蒙汉综合领域、藏汉综合领域和泰中综合领域的 3 个评测项目，提交了测试集翻译结果。本文对评测所用系统、技术路线以及系统运行软硬件环境进行了详细的介绍。

单位：苏州大学

报告时间：2022 年 8 月 10 日 11:44-11:52

基于多策略优化的译文质量评估

雷涛，叶恒，蒋宇龙，姜云卓，徐文斌，贡正仙

内容摘要：针对 CCMT 2022 中英译文质量评估任务（QE），本文探索了基于“预训练-评估器”模型的 QE 系统性能提升的多策略优化方法。本文以 Transquest 模型作为 QE 的基础系统架构。第一个策略是改进 QE 系统依赖的 XLM-R 预训练模型，使用 CCMT 2022 官方提供的中英平行语料，利用 TLM 掩码策略对 XLM-R 预训练模型进一步训练，增强 XLM-R 的双语关联能力。其次，在 QE 模

型训练过程中，分别探索了在输入端扩展伪 PE 数据的策略和在 XLM-R 输出端融入依存句法特征的策略。实验表明，相对基准系统，这些策略都能不同程度地提升 QE 系统的预测能力。

单位：中科院计算所

报告时间：2022 年 8 月 10 日 11:52-12:00

中科院计算所 CCMT 2022 蒙汉翻译技术报告

房庆凯，马铮睿，桂尚彤，张倬诚，伍烜甫，黄浪林，张民，冯洋

内容摘要：本文介绍了中国科学院计算技术研究所参加第十八届全国机器翻译大会（CCMT 2022）蒙汉机器翻译评测的技术实现。本次参评系统使用深层 Transformer 作为基线模型，通过启用 dropout 的反向翻译进行数据增广，结合真实平行语料和伪平行语料一起训练模型。在此基础上，引入正则化方法 PD-R 辅助模型训练。在解码阶段，将不同数据训练得到的模型的神经网络输出层概率进行平均。实验结果表明，本文的方法在翻译质量上相比基线系统取得了显著的提升。

中科院计算所 CCMT 2022 低资源神经翻译技术报告

张倬诚，伍烜甫，黄浪林，房庆凯，马铮睿，桂尚彤，张民，冯洋

内容摘要：本文介绍了中国科学院计算技术研究所参与第十八届全国机器翻译大会的总体情况和技术细节。在本次评测中我们参加了泰语

到汉语和汉语到泰语低资源翻译两个任务。针对两个任务的不同特点，本文给出了我们获取并处理语料、模型架构以及针对性的训练策略等方面的技术细节。实验表明，本次评测我们采用的策略和方法有效地提高了低资源情况下的翻译质量。

5. poster 展示 2——评测论文

展示时间：2022 年 8 月 10 日 12:00-12:30

单位：西藏大学藏文信息技术人工智能重点实验室

藏文信息技术人工智能重点实验室 CCMT2022 评测报告

杨丹，唐超超，仁青卓玛，拥措，尼玛扎西

内容摘要：本文详细介绍了西藏自治区藏文信息技术人工智能重点实验室（西藏大学）参加第十八届全国机器翻译大会（CCMT 2022）（China Conference on Machine Translation, CCMT）的藏汉机器翻译评测情况。本次测评采用了谷歌 Transformer 神经网络机器翻译架构作为基线翻译模型，主要利用数据增强的方式对藏汉平行语料进行扩充、优化藏汉神经机器翻译所用到的词表并探索跨语言预训练模型中的联合词表对翻译性能的影响，最终使用了一种融合跨语言预训练模型 mRASP 与改进后的绿色联合词表的方法提交了藏汉神经机器翻译的主系统。对比系统分别是基于 ALBERT 预训练语言模型和基于 transformer-big 的翻译系统。与传统的基于预训练语言模型的藏汉神经机器翻译相比，该方法可以有效提高藏汉机器翻译性能。

单位：流利说

LAIX CCMT 2022 翻译质量评估任务测评报告

余勇宏，王永杰，王川，李若冰，林晖

内容摘要：本报告介绍了我们在 2022 年第十八届全国机器翻译大会（CCMT 2022）机器翻译评测翻译质量评估（Quality Estimation，QE）任务中的中-英和英-中两个赛道中所采用的主要方法。本方法基于现有主流的预测器-评估器双阶段模型的框架，其中预测器我们采用了大规模中英双语平行语料和中英文语法纠错（Grammatical Error Correction，GEC）数据合成的伪 QE 数据对 Cross-lingual Language Model（XLM）、XLM-RoBERTa（XLM-R）、XLM-RoBERTa-Large（XLM-R-LARGE）和 Multilingual BERT（mBERT）等预训练语言模型进行继续训练，并对原文句子、译文句子以及参考翻译进行特征提取，评估器通过输入的上述特征对 HTER（Human-targeted Translation Edit Rate）值进行回归预测。在实验过程中，通过引入 GEC 数据合成的伪 QE 数据，我们在中-英和英-中两个赛道的 QE 任务预测效果都有显著提升。在最终的 CCMT 2022 QE 离线测试结果中，我们的集成系统在中-英赛道排在第二，英-中赛道排在第四。

单位：新疆大学信息科学与工程学院

新疆大学信息学院 CCMT2022 维-汉机器翻译系统评测技术报告

买日旦·吾守尔，斯拉吉丁，艾斯卡尔·艾木都拉

内容摘要：本文主要介绍新疆大学信息学院研究团队参与第十八届全国机器翻译大会机器翻译测评中的维汉翻译测评基本情况。在此次评测中，本研究团队参与了维-汉机器翻译任务，并在该任务上提交了评测系统。本技术报告主要介绍参赛的维汉神经网络机器翻译系统选用的不同模型对比分析、不同粒度选择、数据增强使用方法以及该系统在开发集上的性能。

单位：北京理工大学

北京理工大学 CCMT2022 技术报告

依西降参，简林圳，朱晓光，史树敏，鉴萍

内容摘要：本文详细介绍了北京理工大学参加第十八届全国机器翻译大会（CCMT 2022）评测的情况。在本次评测中，我们参加了其中的3个翻译任务，分别是维汉新闻领域机器翻译、蒙汉日常用语机器翻译和藏汉政府文献领域机器翻译。本系统分别采用了掩码语言模型预训练、反向翻译、滑动参数平均、集成学习等方法提升翻译效果。实验表明，相对于基线系统，本系统采用的方法可以显著提升模型的翻译效果。

单位：内蒙古大学

内蒙古大学 CCMT2022 蒙汉翻译评测技术报告

杨宗恒，侯宏旭，孙硕，乌尼尔，菅伟辰，王翌松，王鹏聪

内容摘要：本文主要介绍内蒙古大学计算机学院蒙古文信息处理技术重点实验室在第十八届全国机器翻译大会（CCMT2022）机器翻译评测项目中参赛的基本情况。在本次机器翻译评测中，我们参加了蒙汉综合领域双语翻译评测项目的在线评测和离线评测。本文采用基于自注意力网络的 Transformer 架构训练了一个基线蒙汉翻译模型，并使用非参数方法使用训练数据构建了一个外部记忆模块，在模型解码时通过检索与匹配从外部知识中获得指导以提升模型能力。本文主要介绍了该模型所采用的具体方法和实验细节，并通过实验表明了该模型在蒙汉双语翻译任务上性能的提升，并对此进行了深入的分析和研究。

单位：南京大学

NJUNLP's Submission for CCMT 2022 Quality Estimation Task

Yu Zhang, Xiang Geng, Shujian Huang, and Jiajun Chen

Abstract. Quality Estimation is a task aiming to estimate the quality of translations without relying on any references. This paper describes our submission for CCMT 2022 quality estimation sentence-level task for English-to-Chinese (EN-ZH). We follow the DirectQE framework, whose target is bridging the gap between pre-training on parallel data and

fine-tuning on QE data. We further combine DirectQE with the pre-trained language model XLM-RoBERTa (XLM-R) which achieves outstanding success in many NLP tasks in order to improve performance. With the purpose of better utilizing parallel data, several types of pseudo data are employed in our method as well. In addition, we also ensemble several models to promote the final results.

单位：广西大学

Effective Data Augmentation Methods for CCMT 2022

Jing Wang and Lina Yang

Abstract. The purpose of this paper is to introduce the specific situation in which Guangxi University participated in the 18th China Conference on Machine Translation (CCMT 2022) evaluation tasks. We submitted the results of two bilingual machine translation (MT) evaluation tasks in CCMT 2022. One is Chinese-English bilingual MT tasks from the news field, the other is Chinese-Thai bilingual MT tasks in low resource languages. Our system is based on Transformer model with several effective data augmentation strategies which are adopted to improve the quality of translation. Experiments show that data augmentation methods have a good impact on the baseline system and aim to enhance the robustness of the model.

单位：广东工业大学

基于正则泛化的中泰机器翻译系统

林楠铠，林晓钿，黄锦荣，蒋盛益

内容摘要：基于 Transformer 的神经机器翻译模型在高资源语言上取得了成功，然而在低资源语种上则模型效果较差。针对中泰机器翻译任务，本文梳理了该任务的研究现状，并尝试将正则泛化技术应用到该任务上，通过不同的 dropout 方式得到不同的概率标签，通过要求不同的输出之间的差异尽可能小，从而提高模型的泛化能力，减少训练和推理中存在的不一致性。实验结果表明，本文采用技术的有效性，该方法可以有效提升低资源场景下的中-泰与泰-中机器翻译效果。

6. 英文论文报告

报告时间：2022 年 8 月 10 日 12:30-13:45

Review-based Curriculum Learning for Neural Machine Translation

报告时间：2022 年 8 月 10 日 12:30-12:45

Ziyang Hui, Chong Feng and Tianfu Zhang

Abstract. For Neural Machine Translation (NMT) tasks with limited domain resources, curriculum learning provides a way to simulate the human learning process from simple to difficult to adapt the general NMT model to a specific domain. However, previous curriculum learning methods suffer from cata-strophic forgetting and learning inefficiency. In this paper, we introduce a re-view-based curriculum learning method, targetedly selecting curriculum ac-cording to long time interval or unskilled mastery. Furthermore, we add gen-eral domain data to curriculum learning, using the mixed fine-tuning method, to improve generalization and robustness of translation. Extensive experi-mental results and analysis show that our method outperforms other curriculum learning baselines across three specific domains.

Target-side Language Model for Reference-free Machine Translation Evaluation

报告时间：2022 年 8 月 10 日 12:45-13:00

Min Zhang, Xiaosong Qiao, Hao Yang, Shimin Tao, Yanqing Zhao, Yinlu Li, Chang Su, Minghan Wang, Jiaxin Guo, Yilun Liu, and Ying Qin

Abstract. With the rapid progress of deep learning in multilingual language processing, there has been a growing interest in reference-free

machine translation evaluation, where source texts are directly compared with system translations. In this paper, we design a reference-free metric that is based only on a target-side language model for segment-level and system-level machine translation evaluations respectively, and it is found out that promising results could be achieved when only the target-side language model is used in such evaluations. From the experimental results on all the 18 language pairs of the WMT19 news translation shared task, it is interesting to see that the designed metrics with the multilingual model XLM-R get very promising results (best segment-level mean score on the from-English language pairs, and best system-level mean scores on the from-English and none-English language pairs) when the current SOTA metrics that we know are chosen for comparison.

Improving the Robustness of Low-Resource Neural Machine Translation with Adversarial Examples

报告时间：2022 年 8 月 10 日 13:00-13:15

Shuo Sun, Hongxu Hou, Nier Wu, Zongheng Yang, Yisong Wang, Pengcong Wang, and Weichen Jian

Abstract. Weak robustness and noise adaptability are major issues for Low-Resource Neural Machine Translation (NMT) models. That is, once some tiny perturbs are added to the input sentence, the model will produce completely different translation with high confidence. Adversarial example is currently a major tool to improve model robustness and how to generate an adversarial examples that can degrade the performance of the model and ensure semantic consistency is a

challenging task. In this paper, we adopt reinforcement learning to generate adversarial example for low-resource NMT. Specifically, utilizing the actor-critic algorithm to modify the source sentence, the discriminator and translation model in the environment are used to determine whether the generated adversarial examples maintain semantic consistency and the overall deterioration of the model. Furthermore, we also install a language model reward to measure the fluency of adversarial examples. Experimental results on low-resource translation tasks show that our method highly aggressive to the model while maintaining semantic constraints greatly. Moreover, the model performance is significantly improved after fine-tuning with adversarial examples.

Dynamic Mask Curriculum Learning for Non-Autoregressive Neural Machine Translation

报告时间：2022 年 8 月 10 日 13:15-13:30

Yisong Wang, Hongxu Hou, Shuo Sun, Nier Wu, Weichen Jian, Zongheng Yang, and Pengcong Wang

Abstract. Non-autoregressive neural machine translation is gradually becoming a research hotspot due to its advantages of fast decoding. However, the increase of decoding speed is often accompanied by the loss of model performance. The main reason is that the target language information obtained at the decoder side is insufficient, and the mandatory parallel decoding leads to a large number of mistranslation and missing translation problems. In order to solve the problem of

insufficient target language information, this paper proposes a dynamic mask curriculum learning approach to provide target side language information to the model. The target side self-attention layer is added in the pre-training phase to capture the target side information and adjust the amount of information input at any time by way of curriculum learning. The finetuning and inference phases disable the module in the same way as the normal NAT model. In this paper, we experiment on two translation datasets of WMT16, and the BLEU improvement reaches 4.4 without speed reduction.

Dynamic Fusion Nearest Neighbor Machine Translation via Dempster–Shafer Theory

报告时间：2022 年 8 月 10 日 13:30-13:45

Zongheng Yang, Hongxu Hou, Shuo Sun, Nier Wu, Yisong Wang, Weichen Jian, and Pengcong Wang

Abstract. kNN-MT has been recently proposed, uses a token-level knearest neighbor approach to retrieve similar sentences, obtaining knowledge guidance from an external memory module, and then combined with the prediction results of the translation model, which greatly improves the accuracy of machine translation. However, kNN-MT uses simple linear interpolation in the fusion of retrieval probability and translation probability, which can not dynamically adjust the fusion ratio according to the matching degree of the retrieved sentences. Moreover, different fusion ratios need to be explored in different translation scenarios, and the translation effect will be affected when the retrieved

sentences have a low matching degree or contain noise. In this paper, we propose an approach via Dempster–Shafer theory(DST) to dynamically fuse different probability distributions to suit different scenarios. We demonstrate that our approach is more significantly improved and more robust than the traditional kNN-MT, and we explore the application of kNN-MT in low-resource translation scenarios for the first time.

7. poster 展示 3——评测论文

展示时间：2022 年 8 月 10 日 16:00-16:30

单位：西藏大学信息科学技术学院

CCMT2022 低资源藏汉机器翻译评测报告

严松思，汪超，珠杰

内容摘要：本文介绍了实验室参加第十八届全国机器翻译大会（CCMT2022）机器翻译评测中的藏汉日常用语机器翻译评测项目的情况。本报告主要介绍参赛的藏汉神经网络机器翻译系统以及该系统所采用的方法及该系统在开发集和测试集上的性能。

单位：郑州大学

Optimizing Deep Transformers for Chinese-Thai Low-Resource Translation

Wenjie Hao, Hongfei Xu, Lingling Mu, and Hongying Zan

Abstract. In this paper, we study the use of deep Transformer translation model for the CCMT 2022 Chinese↔Thai low-resource machine translation task. We first explore the experiment settings (including the number of BPE merge operations, dropout probability, embedding size, etc.) for the low-resource scenario with the 6-layer Transformer. Considering that increasing the number of layers also increases the regularization on new model parameters (dropout modules are also introduced when using more layers), we adopt the highest performance

setting but increase the depth of the Transformer to 24 layers to obtain improved translation quality. Our work obtains the SOTA performance in the Chinese-to-Thai translation in the constrained evaluation.

单位：西北民族大学

西北民族大学 2022 年全国机器翻译大会机器翻译评测技术报告

李亚超，江涛，加羊吉，胡阿旭，吕世良，祁坤钰

内容摘要：本文介绍了西北民族大学参加的第十八届全国机器翻译大会中的藏汉、蒙汉以及维汉等三个翻译任务。由于藏语、维吾尔语测试集所包含的源语言输入句子较长，甚至是整篇文档，我们采用了针对性的处理方法，将较长句子切分成和训练语料长度符合的较短句子，并予以翻译。最后，我们提交了相应的翻译结果。

单位：中国科学技术信息研究所

ISTIC’s Thai-to-Chinese Neural Machine Translation System for CCMT’2022

Shuao Guo, Hangcheng Guo, Yanqing He, and Tian Lan

Abstract. This paper introduces technical details of Thai-to-Chinese neural machine translation system of Institute of Scientific and Technical Information of China (ISTIC) for the 18th China Conference on Machine Translation (CCMT’ 2022). ISTIC participated in a low resource evaluation task: Thai-to-Chinese MT task. The paper mainly illuminates

its system framework based on Transformer, data preprocessing methods and some strategies adopted in this system. In addition, the paper evaluates the system performance under different methods.

单位：北京航空航天大学

北京航空航天大学 CCMT2022 评测报告

唐子安，巢文涵，徐贝凝

内容摘要：本报告介绍了本单位参加第十八届全国机器翻译大会（CCMT 2022）评测的情况。在本次评测中，我们参加了其中的 3 个翻译任务，分别是蒙汉综合领域机器翻译、藏汉综合领域机器翻译和维汉新闻领域机器翻译。以上翻译任务的主要问题为：维汉新闻领域语料资源稀缺。针对此问题，本系统采用了反向翻译方法提升翻译效果。实验表明，相对于基线系统，本系统采用的方法可以显著提升模型的翻译效果。

单位：湖南大学

A Multi-tasking and Multi-stage Chinese Minority Pre-Trained Language Model

Bin Li, Yixuan Weng, Bin Sun, Shutao Li

Abstract. The existing multi-language generative model is considered as an important part of the multilingual fifield, which has received extensive attention in recent years. However, due to the scarcity of Chinese

Minority corpus, developing a well-designed translation system is still a great challenge. To leverage the current corpus better, we design a pre-training method for the low resource domain, which can help the model better understand low resource text. The motivation is that the Chinese Minority languages have the characteristics of similarity and the adjacency of cultural transmission, and diffffferent multilingual translation pairs can provide the pre-trained model with suffiffifficient semantic information. Therefore, we propose the Chinese Minority Pre-Trained (CMPT) language model with multi-tasking and multi-stage strategies to further leverage these low-resource corpora. Specifically, four pre-training tasks and two-stage strategies are adopted during pre-training for better results. Experiments show that our model outperforms the baseline method in Chinese Minority language translation. At the same time, we released the fifirst generative pre-trained language model for the Chinese Minority to support the development of relevant research3.

单位：昆明理工大学信息工程与自动化学院

昆明理工大学 CCMT2022 机器翻译评测报告

王振晗，叶俊杰，陈蕊，朱志国，高盛祥，毛存礼

内容摘要：本文详细介绍了昆明理工大学云南省人工智能重点实验室参加 2022 年全国机器翻译(CCMT2022) 评测任务的情况。本次评测我们参加泰语-汉语受限域及非受限域两个测评任务。我们的系统采用基于深度神经网络的 Transformer 模型和基于回译的数据增强方法

进行，分别在受限及非受限两种方式下进行模型学习和训练，受限方式即训练数据完全来自于评测方提供的训练数据，非受限方式即在评测方提供的训练数据的基础上，增加实验室收集的双语平行语料，双语词典。

六、机器翻译论坛

1. 机器翻译前沿趋势论坛

报告 1：无监督机器翻译及其应用

主讲人：刘树杰

报告时间：2022 年 8 月 9 日 15:00-15:25

内容摘要：随着深度学习的发展，神经机器翻译在近几年取得了巨大的进步。然而对于一些低资源的翻译任务，比如小语种翻译，或者某些特定的领域，大规模的双语句对并不是特别容易获得，从而导致翻译质量不尽如人意。如何基于少量的双语数据或者在没有任何双语数据的情况下设计和实现机器翻译系统越来越受到研究者的重视。在本报告中，我们将介绍在不利用任何双语数据的情况下训练和搭建机器翻译系统的常用方法，即无监督机器翻译，包括双语词典的构建，基

于去噪自编码器和预训练方法的翻译模型初始化，基于无监督统计机器翻译的初始化，以及基于联合训练的迭代优化方法等。在此基础上，进一步介绍无监督机器翻译在代码智能以及文本风格转换中的应用。

主讲人介绍：刘树杰，微软亚洲研究院高级研究员和研究经理。于2012 年从哈尔滨工业大学博士毕业。研究兴趣包括自然语言处理，语音处理和深度学习相关技术。在自然语言处理以及语音处理各顶级期刊和会议（包括 CL, JSTSP, ACL, ICASSP, AAAI, EMNLP, NAACL, INTERSPEECH 等）上发表论文 100 余篇，并合著《机器翻译》一书，参与编写《人工智能导论》一书。研究成果被广泛应用于 Microsoft Translator、Skype Translator、Microsoft IME、微软小冰和微软语音服务（包括语音生成，语音分离和识别）等微软重要产品中。

报告 2：低资源机器翻译的研究现状

主讲人：陈科海

报告时间：2022 年 8 月 9 日 15:25-15:50

内容摘要：本报告将系统地介绍低资源机器翻译的研究方法和发展趋势。主要内容包括数据扩增、多语言、迁移学习、小样本以及方法适应性等算法及其应用；并且会介绍低资源机器翻译的未来发展趋势，包括数据集构建、开源工具与框架、提示学习、可解释性建模、先验知识引入、模型鲁棒性等。

主讲人介绍：陈科海，哈尔滨工业大学（深圳）计算机科学与技术学院助理教授，主要研究方向是机器翻译和自然语言处理。2018 年 10 月博士毕业哈尔滨工业大学，获 2020 年中国中文信息学会优秀博士学位论文奖（全国共 5 人），在 2018 年 11 月至 2021 年 12 月就职于

日本国立情报通信研究机构从事博士后研究，共发表 CCF-A/B 会议和国际期刊论文 30 余篇。先后担任 AAAI-2021/AAAI-2022 高级程序委员、ACL-2022 机器翻译共同领域主席、中国机器翻译年会 CCMT-2022 的共同出版主席、中国中文信息学会青年工作委员会和中国中文信息学会自然语言生成与智能写作专委会的委员。

圆桌讨论：低资源机器翻译

嘉宾：刘树杰、陈科海、王明轩、王星

时间：2022 年 8 月 9 日 15:50-16:50

嘉宾介绍：

1. 王明轩

单位：字节跳动

个人简介：字节跳动科技有限公司机器翻译业务负责人，研究方向主要为机器翻译和自然语言处理。在机器翻译领域，发表包括 ACL、EMNLP 等顶级会议论文超过 40 篇，多次拿到 WMT 等国际翻译评测比赛第一。同时还担任 EMNLP2022 赞助主席，和 NeurIPS 2022、NLPCC 2022、AACL2022 等会议领域主席。

2. 王星

单位：腾讯公司人工智能实验室

个人简介：腾讯公司人工智能实验室（Tencent AI Lab）高级研究员，2018 年博士毕业于苏州大学，主要从事机器翻译相关的研发工作。在 ACL、EMNLP、NAACL、AAAI 等自然语言处理相关会议和刊物上发表论文约三十篇，在 WMT 国际翻译评测多个赛道获得评测比赛第一名，任中国中文信息学会青年工作委员会委员。

2. 机器翻译产业应用论坛

圆桌讨论 1：机器翻译下一阶段的产业落地——瓶颈和突破

主持人：黄国平

时间：2022 年 8 月 9 日 17:20-18:10

圆桌讨论 2：从工业界角度看研究生培养——需求与建议

主持人：冯冲

时间：2022 年 8 月 9 日 18:20-19:10

嘉宾：陶仕敏（华为）、李响（小米）、王明轩（火山翻译）、张春良（小牛翻译）、孟凡东（微信翻译）、薛征山（OPPO）、宗浩（中译语通）、何中军（百度）、丁丽（云译）、刘劲松（新译）、于超（Magic Data）

嘉宾介绍

1. 陶仕敏

单位：华为

个人简介：华为 2012 文本机器翻译实验室技术专家，北京团队负责人。主要负责研究、创新工作及对外技术合作等。

2. 李响

单位：小米

个人简介：小米 AI 实验室机器翻译业务负责人，博士毕业于中国科

学院计算技术研究所，研究方向为机器翻译，带领团队研发的机器翻译相关技术和产品已广泛赋能小米手机，小爱翻译，小爱同学等小米软硬件产品，月活跃用户超千万。

3. 王明轩

单位：火山翻译

个人简介：字节跳动科技有限公司机器翻译业务负责人，研究方向主要为机器翻译和自然语言处理。在机器翻译领域，发表包括 ACL、EMNLP 等顶级会议论文超过 40 篇，多次拿到 WMT 等国际翻译评测比赛第一。同时还担任 EMNLP2022 赞助主席，和 NeurIPS 2022、NLPCC 2022、AACL2022 等会议领域主席。

4. 张春良

单位：小牛翻译

个人简介：张春良，小牛翻译总裁、创始合伙人，负责政府事务、海外市场、小牛翻译云平台运营等业务。主持/参与包括“科技冬奥”“新一代人工智能 2030”等科技部重点项目课题在内等国家及省部级机器翻译项目 12 项；2016 年获国内自然语言处理领域最高科技奖——钱伟长中文信息处理科技一等奖。

5. 孟凡东

单位：微信翻译

个人介绍：腾讯专家研究员、技术 Leader，博士毕业于中科院计算所，研究方向为机器翻译。负责微信翻译业务，包括算法研究、系统研发与应用，服务微信生态内多个场景，以及公司内、外多家合作伙伴。在 ACL、EMNLP、NAACL、AAAI、AIJ 等国际顶会/刊上发表论文 60 余篇，获 ACL 2019 最佳长论文奖，多次获得 WMT 国际机器翻译评测比赛第一。

6. 薛征山

单位：OPPO

个人介绍：薛征山，OPPO 机器翻译团队负责人，研究方向为机器翻译和自然语言处理。带领团队研发的机器翻译引擎应用在 OPPO 各业务场景中，在国内外机器翻译竞赛 WMT、IWSLT 等获得多次第一名。

7. 宗浩

单位：中译语通

个人介绍：宗浩，中译语通科技股份有限公司机器翻译负责人，负责中译语通机器翻译技术研发，带领团队获得工信部人工智能首批揭榜优胜单位，参与科技部科技创新 2030—“新一代人工智能”重大项目，在国际机器翻译比赛 IWSLT、WMT 中累计获得第一名达到 28 项。

8. 何中军

单位：百度

个人介绍：何中军，博士，百度人工智能技术委员会主席。长期从事机器翻译研究与开发，研发了全球首个互联网神经网络机器翻译系统及语义单元驱动的机器同传系统。曾获国家科技进步二等奖、中国电子学会科技进步一等奖、北京市科技进步一等奖、中国专利银奖等多项奖励。

9. 丁丽

单位：云译科技

个人介绍：云译科技总经理，联合创始人。安徽师范大学英语教育本科，北京大学 MBA，从事翻译行业二十余年，积累丰富的人工翻译经验翻译管理经验，熟悉人工翻译行业技术及应用，2017 年和厦门大学史晓东教授联合创立云译科技，从事人工智能机器的研发以及机器翻译在人工翻译行业的结合的应用。

10. 刘劲松

单位：新译科技

个人介绍：刘劲松，新译研究院副院长。研究方向为跨语言服务咨询、机器翻译质量评估，跨文化沟通与技术写作。参与申请新译科技为国家首批示范性语言服务出口基地，为国内外知名企业诺基亚、微软、爱立信、三星、华为、联想、阿里巴巴以及保密单位等提供多语言解决方案咨询；Tekom 中国首批技术写作讲师；参与 ISO 国际/GB 国内国际标准修订以及转化工作；参与大型国际化项目若干。

11. 于超

单位：Magic Data

个人介绍：于超，Magic Data 自然语言处理业务负责人，毕业于北京大学，研究方向为自然语言处理。主要客户覆盖阿里巴巴、腾讯集团、字节跳动、微软等数十家世界 500 强企业，为客户提供 AI 数据解决方案咨询与服务。同时带队研发功能丰富的 Annotator6.0 智能化标注平台—NLP 模块，为企业数字化转型降本增效。

3. 机器翻译学生论坛

报告 1：科研路上的“屠龙刀”们

主讲人：鲍宇

报告时间：2022 年 8 月 10 日 16:30-16:45

内容摘要：工欲善其事，必先利其器。在生活中，准确地选择和使用工具往往可以极大地提高我们的效率，让我们事半功倍，在科研工作中亦是如此。本次报告将带来科研学术道路上的“工具篇”，结合个人经验，分享在科研工作中用于文献调研和管理、图表制作、写作润色等环节的工具，希望能够帮助到有需要的同学。

主讲人介绍：鲍宇，字节跳动 AI Lab 研究员。2022 年博士毕业于南京大学，师从陈家骏教授和黄书剑副教授。研究兴趣为深度文本生成模型，重点关注于非自回归式生成模型的研究，相关工作发表于 ACL（2019-2022）、NAACL2021 等。

报告 2：从业务需求中发现科研问题：记我的一段实习经历

主讲人：刘鑫

报告时间：2022 年 8 月 10 日 16:45-17:00

内容摘要：科研与实际业务的结合一直是产业界工作的重点，而从实际业务中抽象出具有科研价值的问题则是其中的重要一环。本报告将分享一次我的实习经历，从个人视角出发讲述我从业务需求中发现和归纳问题的过程，并介绍在此期间发表的两项研究成果——解决生成式模型在预训练、微调范式下的期望子词粒度不匹配问题；根据用户关键词生成符合人类常识句子的生成模型。

主讲人介绍：刘鑫同学在此前为厦门大学的硕士研究生，其导师是苏劲松教授。他的主要研究方向包含文本生成、机器翻译和机器阅读理解。他有着在百度和阿里达摩院自然语言处理部门两段实习经历。在研究生期间，刘鑫共以第一作者/共同第一作者的身份发表了 CCF-A 类文章 4 篇，并将前往密歇根大学继续攻读博士学位。

报告 3：我在科研道路上走过的弯路

主讲人：许晨

报告时间：2022 年 8 月 10 日 17:00-17:15

内容摘要：科研不是坦途，失败才是常态。尤其对于刚进入研究生阶段的同学们来说，在最初踏上科研道路时，往往都会遇到各种挑战，没有 idea、实验结果不理想、写作无从下手等问题，陷入焦虑和自我怀疑等负面情绪中。本次报告根据个人的经历，分享在研究过程中走过的弯路，包括研究方向选择、论文投稿过程以及跳出科研舒适圈等问题，希望可以帮助到有需要的人。

主讲人介绍：许晨，东北大学自然语言处理实验室 19 级博士研究生，导师是朱靖波教授和肖桐教授。研究方向主要包括机器翻译、语音翻译和语音识别。多次参加 WMT 新闻翻译评测、质量评估评测与 IWSLT 语音翻译评测，并参与实验室自研张量计算库 NiuTensor 的开发与书籍《机器翻译：基础与模型》的撰写。曾获 CCL2021 最佳中文论文奖，相关工作发表在 ACL、COLING 等会议，并担任相关会议审稿人。

报告 4：NMT 可解释性：科研“蓝海”的思考和摸索

主讲人：卢宇

报告时间：2022 年 8 月 10 日 17:15-17:30

内容摘要：当深度神经网络的性能不断提升，如何解构和理解这个黑盒模型成为关注的焦点，“NLP 模型的可解释性与分析”也从 2020 年开始成为 ACL 的新赛道。新的领域逐步扩展着研究空间，但同时也提高了科研难度，增加了试错成本。这里，我将结合自己的研究课题，和大家分享在新方向上“摸爬滚打”的经验和教训。

主讲人介绍：卢宇，中科院自动化所模式识别国家重点实验室四年级在读博士，导师为张家俊研究员。研究兴趣为神经机器翻译的可解释性，包括模型和数据不确定性的分析和改进。已发表 CCF-A 类论文 2 篇，B 类期刊 1 篇。

报告 5：研究初期的个人学术品牌建设

主讲人：王龙跃

报告时间：2022 年 8 月 10 日 17:30-17:45

内容摘要：本报告将结合具体研究项目，探讨不同阶段中遇到几种关系问题，如深度与广度、过程与结果、独立与合作、科研与落地，与大家共同探讨如何在研究初期打造个人学术品牌。

主讲人介绍：王龙跃，腾讯人工智能实验室资深研究员。2018 年于都柏林城市大学获计算机应用专业博士学位，欧洲机器翻译学会 2018 年最佳博士论文奖获得者（1 位/年）。其主要从事机器翻译、自然语言处理等方向研究，在 ICLR、ACL 等国际期刊和会议上发表论文四十余篇，累计申请相关专利五十余项。参加国际学术比赛十余次，其中在 WMT 国际机器翻译比赛中共获六次冠军。担任 2021 腾讯 AI Lab 犀牛鸟专项研究计划项目负责人，中文信息学会青年工作委员会委员。在知名期刊和会议中担任编委会成员、领域主席，获 NAACL2022 最佳领域编委。

报告 6：我科研历程中的“弯路”

主讲人：王明轩

报告时间：2022 年 8 月 10 日 17:45-18:00

内容摘要：主要回顾自己过去多年从博士到工作的科研历程，并总结其中在关键节点做出的选择，其中有一些正确的选择。但是，回过头来看，更多的事情可以做的更好。这个分享抽象了讲者在科研道路上所走过的“弯路”，并进行总结，希望可以给听众一些启发，在科研道路上走的更远。

主讲人介绍：王明轩，字节跳动科技有限公司机器翻译业务负责人，研究方向主要为机器翻译和自然语言处理。在机器翻译领域，发表包括 ACL、EMNLP 等顶级会议论文超过 40 篇，多次拿到 WMT 等国际翻译评测比赛第一。同时还担任 EMNLP2022 赞助主席，和 NeurIPS 2022、NLPCC 2022、AACL2022 等会议领域主席。

Panel：博士生涯如何“standing out”

嘉宾：鲍宇、刘鑫、许晨、卢宇、王龙跃、王明轩

时间：2022 年 8 月 10 日 18:00-18:30

评测支持单位

中国外文局翻译院是由中国外文局主管的事业单位，旨在汇聚整合翻译行业资源，承担国家重大翻译任务，满足国家对外翻译高端需要，服务于国际传播事业，促进跨文化交流。

翻译院拥有精通中外双语文字和语音的多语种专家人才资源，并在翻译行业标准化和译文质量评价体系建设上引领着新时代翻译行业的发展。

面对文化产业和国际传播事业的新形势、新任务、新挑战、新机遇，翻译院已启动“智能翻译实验室”筹建，旨在以“开源、开放、共商、共建、共享”的合作机制，以术语库、语料库、人才库及多语种终身学习平台“三库一平台”建设为抓手，依托人工智能、5G、云计算、大数据等新一代信息技术赋能行业标准制定、智库研究服务等，对行业未来发展进行前瞻性技术和资源布局，取得若干重要成果，推动政产学研结合、面向行业开放服务。

腾讯会议：

会 议 号： 508-888-802

会议主题：第十八届全国机器翻译大会（CCMT 2022）

会议时间：2022/08/08 08:00-20:00 (GMT+08:00) 中国标准时间 - 北京

重复周期：2022/08/08-2022/08/10 08:00-20:00, 每天

b站直播间：

b站二维码

微博直播间：

微博二维码

赞助商

钻石

火山翻译 GTCOM

白金

小牛翻译 NiuTrans.com　有道 youdao　微信AI　HUAWEI Translate

金牌

oppo　MAGIC DATA　Baidu 百度　云译科技 CloudTranslation　mi

银牌

新译科技 NewTranx Tech