

难题1：超大规模无线网络能效与体验均衡鲁棒调控技术

出题组织：体验Lab 接口专家：耿新力 gengxinli@huawei.com;

技术背景

无线网络能耗高，面向绿色双碳演进挑战大。随着站点结构演进（CRAN、室内改室外等），站点级能量转换率变高（通信设备能耗/总能耗>0.8）。硬件电路升级使得设备能耗转向强依赖于网络业务负载动态变化（待机静态功耗 $EC_{static} \approx 0\text{ watt}$ ）[1, 2]。基于此的网络新形态下，平衡网络性能&体验与节能收益的网络级调控成为节能关键焦点和突破口。



问题与价值：全量用户业务需求分布存在空时不均匀性和动态时变性，通过对网络参数 θ 动态调控，满足需求约束（如速率 $g(\cdot)$ 、覆盖率 $c(\cdot)$ ）下合理分配资源，实现**网络总能耗 $EC_{network}$ 最小化**；

优化目标：满足约束下最小化 $EC_{network}$

优化参数 θ ：

- ✓ 关断 y_i ：关断状态 $y_i \in \{0,1\}$ ，每时段
- ✓ 功率 p_i ：功率调整 $p_i \in [-3, 3]$ ，整数
- ✓ 互操作 h_i ：互操作调整 $h_i \in [-3, 3]$ ，整数

核心问题：多参数、高维、非凸混合整数规划

$$\min_{\theta} EC_{network}(\theta) = f(Module_{EE}, Traffic_{dis}, Radio_{quality})$$

$$= \min_{m_{i,j}, p_i, y_i} \sum_i EE_{Air}(p_i) \cdot \sum_j m_{i,j} + \sum_i EC_{Static,i} \cdot y_i$$

s.t. $\theta_L \leq \theta \leq \theta_U$,

$\sum_i m_{i,j} = thp_j$ 流量分配约束

$\sum_j m_{i,j} = TH_{thp}$ 负载驻留约束

$g(rate_{vj \in J}) \geq TH_{rate}$ 体验达标率约束

$c(coverage_{vj \in J}) \geq TH_{coverage}$ 良好覆盖率约束

空口能耗: EC_{Air} 小区: cell $i: 1 \leq i \leq I$ 用户: User group $j: u_j, 1 \leq j \leq J$

能效与体验联合调控技术

数学建模、寻优求解

网络参数 θ ↓ ↑ 目标/约束 g, c

网络能效SRCON仿真引擎

需求 thp_j 、负载迁移、能耗/体验机理

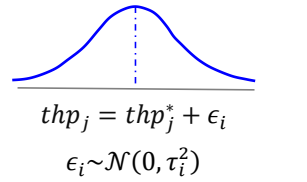
技术挑战

- 为实现**低损伤代价下最大化节能**，考虑**多维参数、细颗粒网络特征的精准、动态调控**是关键。
- 挑战1：**细颗粒度 thp_j 、特征变量（ $Rate_i$ 等）存在**不确定性**， thp_j 变异系数>25%，已有确定性调优技术鲁棒性不足，优度难保障 [3]；
 - 挑战2：**解空间随参数增多爆炸式增长、非凸且机理模型黑盒化，难建模求解。



技术现状

- 鲁棒调控策略：**当前多通过离线寻优叠加3~5轮次现网调控，存在梯度预估偏差，累计收益难保障；业界已有贝叶斯鲁棒调优技术（如Quasi-Monte Carlo） [3]，以N倍计算复杂度估计随机特征，计算复杂度高
- 分簇建模求解：**万级小区网络，联立全量 $\langle I, K, M \rangle$ 建模求解不可行，当前通过基站组簇（如以 a 个小区一簇）、参数分离、局部邻域搜索求解，复杂度 $(I/a) * M * C(a)$ ，因未充分考虑簇、参数、用户群间的关联关系，优度受限



技术诉求

- 效率倍增：**基于现网数据和无线网络机理，考虑多网元资源竞争和协作博弈关系，探索结合数据+机理的最优组簇、数学建模以及求解技术。实现万级小区小时级求解，同等优度下效率倍增。
- 节能收益提5%+（同等体验损伤下）：**推导不确定性鲁棒调控理论模型，通过离线仿真求解+3~5轮现网迭代调控，考虑未来设备演进下，实验室环境累计节能增益提升5%+。

参考文献：

[1] Huawei to Unveil 10 New Solutions at MWC Barcelona 2023. <https://www.huawei.com/en/news/2023/2/mwc2023-5g-ten-product-solution-all-bands>

[2] C. Andersson, and et al, "Improving energy performance in 5G networks and beyond," in Ericsson Technology Review, vol. 2022, no. 8, pp. 2-11, Aug. 2022.

[3] Letham B, Karrer B, et al. Constrained Bayesian optimization with noisy experiments[J]. 2019.

难题2：面向海量网络问题处理的大模型多任务强化学习技术

出题组织：GTS算法与技术开发部 接口专家：唐爽 tangshuang3@huawei.com;

技术背景

随着技术与标准演进以及业务发展，ICT基础设施走向超复杂现实环境网络与系统，一方面网上问题(含企业)**270W单/年居高不下**，另一方面**重复问题**仍在不断发生，不断挑战网络安全与稳定底线。同时当前问题处理依然基于人与经验，现有**案例60W+/故障树节点7W+等知识与经验内容结构化和非结构化并存**，如何通过大模型使能领域知识智能，降低应用门槛，突破基于人的经验局限，实现80%问题一线闭环，二线效率倍增目标。

AS IS：案例/资料/工具多，应用门槛高、个人经验能力依赖

- **80%前端基础问题**：一线/伙伴处理，主要通过问答/搜索手段和少数工具应用，搜索答案碎片化、准确率不足，重复性问题拦截效率不高，70%问题仍需要建单升级。
- **20%复杂网络问题**：二三线处理，需多类型深度日志诊断工具组合分析，专家个人经验强依赖，故障树图谱等前期辅助诊断存在天花板（50%~60%）。

TO BE：大模型人机接口，交互式问题处理，降低门槛提升效率

问题受理 → 问题澄清 → 分析定位 → 方案提供 → 问题闭环



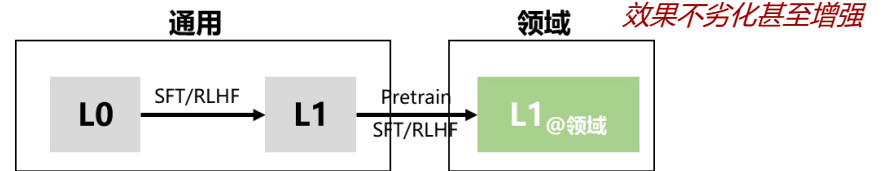
从问题受理到闭环多任务辅助：以大模型为人机接口进行问题澄清、多轮交互、答案生成、准确答复、工具调用等，构建端到端的问题处理辅助能力

参考文献：

- [1]. Kim J, Asai A, Ilharco G, et al. TaskWeb: Selecting Better Source Tasks for Multi-task NLP[J]. arXiv preprint arXiv:2305.13256, 2023.
- [2]. Huang H, Ye D, Shen L, et al. Curriculum-based Asymmetric Multi-task Reinforcement Learning[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022.
- [3]. Lee G, Yang E, Hwang S. Asymmetric multi-task learning based on task relatedness and loss[C]//International conference on machine learning. PMLR, 2016: 230-238.2.

技术挑战

- **通用大模型领域适配问题**：大模型领域化过程中，基于通用模型领域化预训练和指令微调面临着灾难性遗忘、负面知识迁移等挑战，如何选择有助于学习领域新任务的源任务，保证模型效果领域化过程中不劣化？



- **大模型多任务Instruction Tuning和RLHF问题**：领域任务复杂程度普遍高于通用任务，1) 如何/是否考虑从简单的任务开始、并逐渐去学习更加复杂的任务；2) 同时，在统一的LLM框架、但用户多维度反馈内涵下，是否需要/如何考虑多任务学习、基于反馈多任务强化学习？

技术现状

针对上述问题：1) 基于**任务对迁移**^[1]的任务选择方法，通过研究任务之间的相互关系，使用TASKSHOP算法选择有效源任务和构建较小的多任务训练集，实现对多任务学习的改进，在实验验证中取得了显著的性能提升；2) 基于**课程学习的非对称多任务（强化）学习**^{[2][3]}，根据学习进度在并行单任务和基于课程的非对称多任务学习(AMTL)模式之间切换，实现适应从易到难的多任务学习。

技术诉求

研究面向端到端问题处理辅助**问题处理的大模型多任务强化学习技术**，领域化模型通用任务指令准确率持平，领域**指令准确率平均提升10%**，以大模型为人机接口嵌入问题处理作战各个环节，支撑网上**80%重复出现的常见问题辅助解决率达到50%**。

难题3：基于领域知识库和工程师反馈的事实性增强技术

出题组织：GTS算法与技术开发部 接口专家：张豪哲 zhanghaozhe7@huawei.com;

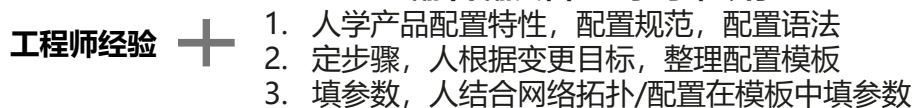
技术背景

全球运营商网络日常存在大量变更操作，比如网络搬迁、特性开通、参数优化等，以我司为例，**22年100万次变更单**，其中5G云核12万单、IP网络10万单（涉及超1百万网元），**高危操作次数超过8万次**，一次变更涉及**上万行命令行**。网络变更风险高，责任重大，华为**十几万工程师“负重前行”**；厚重的变更工作流在保证网络变更可靠性的同时也带来了效率的问题。**如何通过大模型技术给工程师“减负”、提高效率是个重要的课题。**

以命令行编写为例：

AS IS：人拉肩扛，强依赖经验，周期长，难复用

翻书/翻文档 -> 手写命令行



TO BE：人机协同，大模型辅助生成脚本，仿真验证，代际提效

工程师意图



由于概率语言模型的技术本质，大模型的生成结果具有随机性，可能包含虚假信息或者逻辑错误^[1]。因此，如何基于领域知识库和工程师反馈提高大模型在领域任务上的事实性成为关键。

参考文献：

[1] Ji, Ziwei, et al. "Survey of hallucination in natural language generation." ACM Computing Surveys 55.12 (2023): 1-38.

[2] Nakano, R., Hilton, J., Balaji, S., Wu, J., Ouyang, L., Kim, C., ... & Schulman, J. (2021). Webgpt: Browser-assisted question-answering with human feedback. arXiv preprint arXiv:2112.09332.

[3] Peng, Baolin, et al. "Check your facts and try again: Improving large language models with external knowledge and automated feedback." arXiv preprint arXiv:2302.12813 (2023).

技术挑战

- **如何基于领域知识库增强大模型：**工程师作业中沉淀了大量的服务经验（如60多万篇案例、50多万篇高质量MOP方案、1000多棵故障树等）和产品知识（含1.5亿+原子知识的产品文档、100+张产品图谱）。不同于通用语料，领域知识库中的文档有大量的表格（如产品规格、注意事项）、图片（如拓扑图、流程图）、代码（如MML命令行）等信息，知识和经验的密度高，如何将信息对齐、联合训练用于增强大模型是一大挑战。
- **如何基于工程师反馈增强大模型：**十几万工程师可以为大模型生成的结果提供可靠及时的反馈信息，然而ICT服务领域评分反馈往往多维度，奖励模型难以建模，而且强化学习PPO算法稳定收敛困难，经常陷入局部最优解，优化算法挑战大。

技术现状

增强大模型事实一致性的现有前沿工作有**检索增强**和**事实检测器**：1) WebGPT^[2]通过克隆“人搜索网页查找答案”的行为序列，来提高GPT-3模型利用知识库文档回答开放域问题的能力；2) 通过LLM-Augmenter事实检测框架^[3]纳入外部知识和自动反馈，在不影响回答流畅性和信息性的前提下减少了幻觉。

技术诉求

设计一套利用领域知识库增强基础大模型的事实一致性与可靠性的算法方案，支撑ICT服务领域大模型的构建与应用，在领域文本生成、领域DSL生成、阅读理解、多轮问答、逻辑推理等5个领域任务上的事实准确度指标达到GPT-3.5 130%分位数，支撑网络配置翻译、云核MML生成等场景在准确率不变的前提下专家投入时长降低50%。

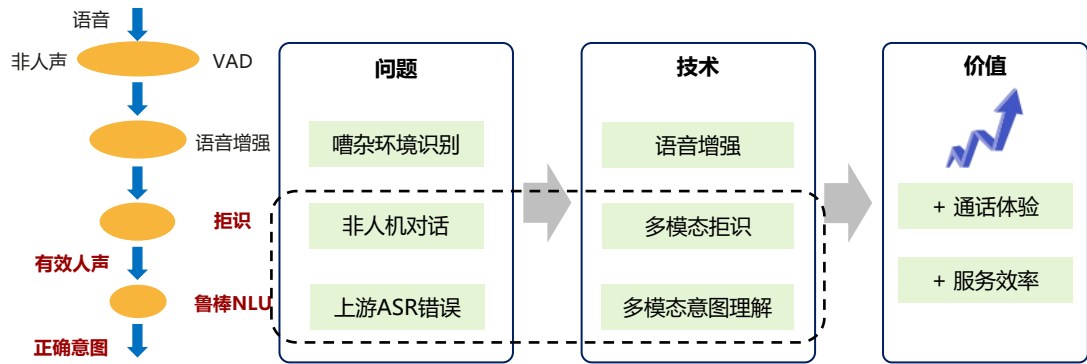
难题4：复杂场景下语音机器人无效声拒识技术

出题组织：软件业务公共开发部 接口专家：揭扬 jieyang.xu@huawei.com

技术背景



- 在客服语音外呼机器人场景下，用户与机器人实时语音对话过程中存在外部噪声干扰、非人机对话、用户口语化表达等无效音问题，往往导致外呼机器人的无效响应。
- 外呼机器人应拒未拒导致的无效响应不仅严重破坏用户通话体验，也使得对话机器人的服务效率降低。
- 期望通过**多模态拒识**、**文本语音表征意图理解**等技术手段，解决无效音误识问题，提升对话机器人体验。



参考文献：

- [1] BUILDING ROBUST SPOKEN LANGUAGE UNDERSTANDING BY CROSS ATTENTION BETWEEN PHONEME SEQUENCE AND ASR HYPOTHESIS
- [2] Pre-training for Spoken Language Understanding with Joint Textual and Phonetic Representation Learning
- [3] Towards an ASR Error Robust Spoken Language Understanding System
- [4] 科大讯飞2022年1024全球开发者节

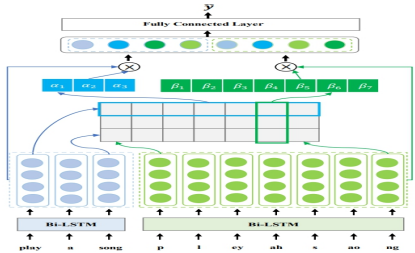
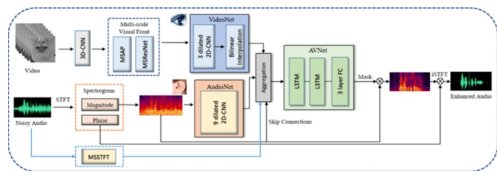
技术挑战

智能客服领域全双工外呼语音对话场景面临新的挑战：

- 终端设备限制，大多数智能客服机器人只能获取到声音和文本相关的信息，考虑如何充分利用两种模态信息提高拒识准确率；
- 电话信道影响，8KHz采样率/语音压缩编码对ASR识别额外更多的精度损失，除提升ASR自身鲁棒性外，下游意图理解模块也要有更高的容错能力；
- 现有数据约束，利用无标注的历史通话录音和转写文本来提升拒识效果。

技术现状

- 复杂场景下周边人声识别在业界没有非常成熟的解决方案，科大主要通过多模态语音增强和识别框架(声音+视觉)的方案，从嘈杂背景音中提取目标人声。其它厂商采用语音特征和文本特征融合方式
- 针对增强意图理解模块的鲁棒性问题，业界部分厂商通过结合音素的语义理解模块，将文本表征和音素表征进行融合来对文本信息做增强表示，Snips验证当ASR错误率 > 60%，意图识别准确率 > 88%



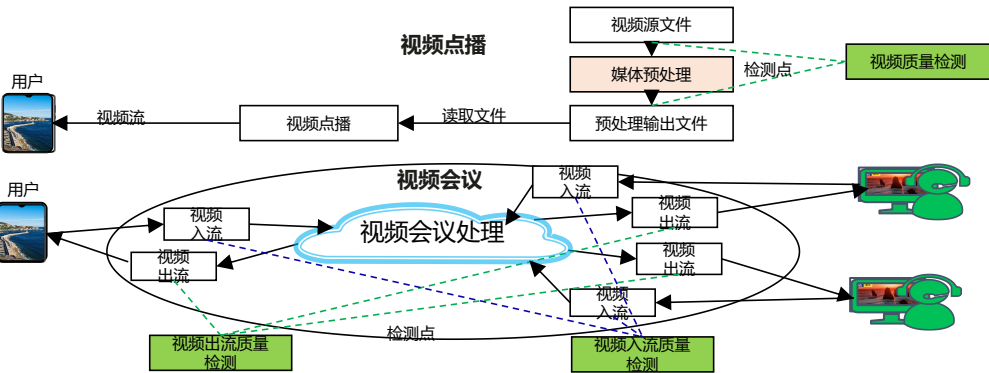
技术诉求

在客服电话信道场景下，通过语音语义多模态理解技术，**识别无关对话改善任务型对话引擎拒识问题**，期望复杂场景下无效音拒识准确率 > 95%，意图识别准确率 > 90%，时延 < 100ms，支持CPU推理。

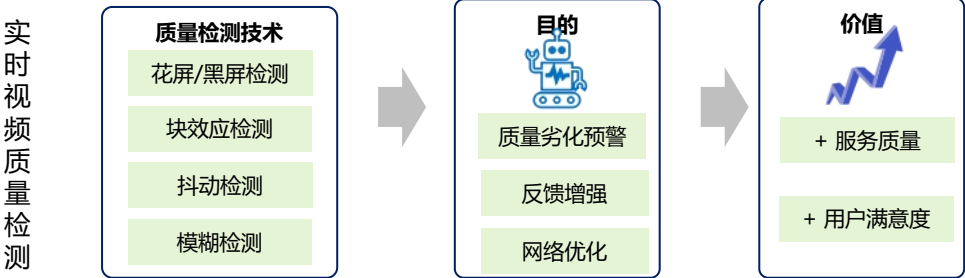
难题5：高性能低时延视频质量检测技术

出题组织：软件业务公共开发部 接口专家：吕梁 steve.lvliang@huawei.com

技术背景



- 随着视频业务普及，视频点播(彩铃)、视频会议(AICC)等场景中，需具备对视频入/出流实时质量检测能力，一方面作为客户端的体验评价或故障检测，另一方面作为技术人员的算法优化性能参考。入流检测可判断视频输入质量，可提示用户网络质量不佳；出流检测可反向控制编码输出码率、并针对低质量图像做增强(如去噪、块效应平滑等)；
- 典型视频质量问题如：块效应、花屏、黑屏、卡屏、抖动、运动模糊等，传统算法计算开销小，但误检率过高，AI算法精确度高但计算开销大、难以应用；
- 希望通过一套算法集实现典型视频质量问题高精度检测，并控制算力开销在合理范围，应用到实时系统中。
- 业务价值：MCU视频会议媒体路径长，通过入出流位置检测，可准确判断异常位置、快速定界，同时可反向控制编码侧调整输出码率等。此通用技术可应用在端侧SDK、SFU会议等场景。



参考文献：

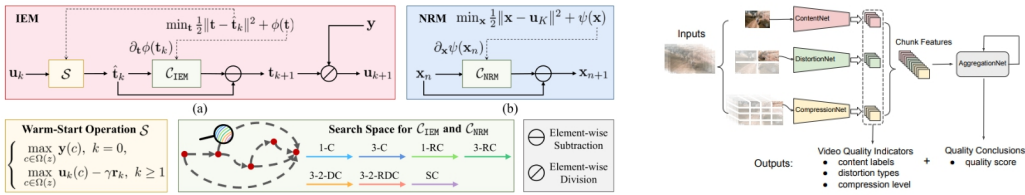
[1] Image Blind Denoising With Generative Adversarial Network Based Noise Modeling
 [2] DeFMO: Deblurring and Shape Recovery of Fast Moving Objects
 [3] Research and System Implementation of Blur and Color Cast in Video Quality Inspection
 [4] AP-BSN: Self-Supervised Denoising for Real-World Images via Asymmetric PD and Blind-Spot Network
 [5] Retinex-inspired Unrolling with Cooperative Prior Architecture Search for Low-light Image Enhancement
 [6] CutPaste: Self-Supervised Learning for Anomaly Detection and Localization

技术挑战

基于视频纹理特征、时序多帧对比等高性能的传统检测算法，可检测典型视频质量问题。算法实现简单，但精度较低、泛化能力弱，需做算法优化；
 基于深度学习的视频质量检测，检测精度高，但鲁棒性不足、实现复杂度高，目前业界研究较少，检索的论文集中在去噪、去雨/去雾、去模糊等场景。

技术现状

- 块效应以及噪点等检测和处理，基于深度学习相关技术，在小模型，小样本，自监督学习等方面做出了较多探索；
- 视频评估从传统方法走向机器学习和深度学习，如全参考型的VMAF视频多评估方法融合；无参考型的VSFAI以及谷歌的CoINVQ，使用多个神经网络分别处理不同纬度的特征；
- 声网建立了Agora-VQA Dataset并训练了Agora-VQA Model，利用深度学习检测清晰度，设计极小端上模型，综合考虑时空相关性/时域相关性/设备相关性等因素，参数量和计算量有几个数量级的下降，准确度达到甚至超过部分大模型。



技术诉求

- 检测精度：**提供高精度视频质量检测算法集，可同时覆盖典型的视频质量问题，期望**准确率>97%，质量问题召回率>95%**
- 算力要求：**通过传统算法和深度学习算法相结合等措施，控制整体算法计算开销，**单流视频质量检测CPU < 0.2vC，实时检测时延<50ms**

华为充分尊重和保护老师的原创，提交给华为的材料会规范管理。请老师通过链接<https://www.chaspark.com/#/questions>“我要揭榜”功能提交材料。

黄大年茶思屋

CHASPARK



鸿蒙/安卓版app下载

汇聚优质内容，粘接精英智慧

黄大年茶思屋帐号注册须知

- 1、黄大年茶思屋科技网站**仅支持华为帐号注册登录**，可通过第三方帐号如qq、微信等授权华为帐号登陆；
- 2、通过华为账号**登录（注册）**后，使用.edu、.ac的邮箱进行认证后可获全平台阅览互动权限；
- 3、用户可通过PC端和APP端进行注册，APP端暂仅支持安卓/鸿蒙版本，iOS尚在开发中；
- 4、可通过华为应用市场直接搜索“黄大年茶思屋”下载。

黄大年茶思屋注册指导:



(一) APP端注册

step1

在个人页面, 点击左上方“点击登录”;



step2

在选择登录方式页面, 点击左下方“华为帐号”



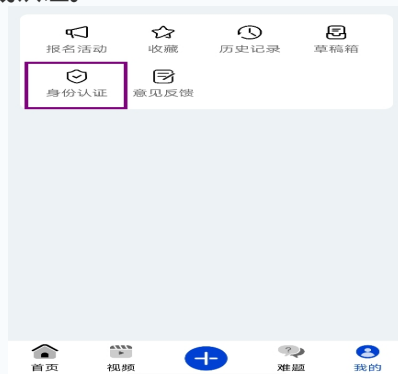
step3

输入手机号、短信验证码即可注册



step4

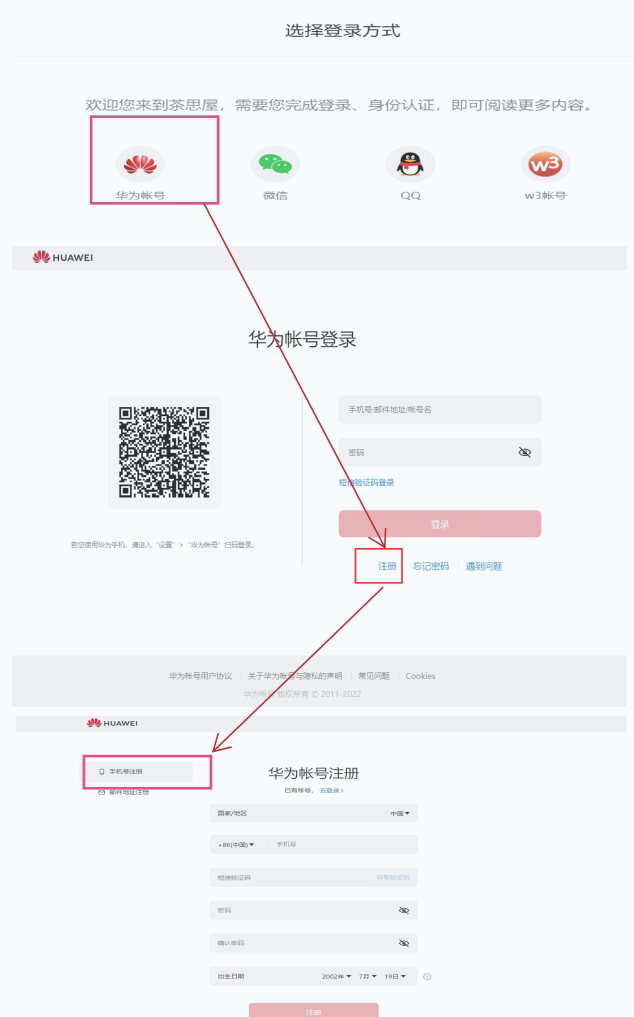
在“我的”, 点击“身份认证”输入.edu、.ac邮箱, 提交后前往邮箱地址完成认证。



(二) PC端注册

step1

登录地址:
<https://www.chaspark.com/#/questions>
选择“华为帐号”并注册。



step2

在顶部导航栏右侧点击个人头像, 进入个人帐号页面;



如因注册中遇到问题无法提交揭榜材料, 请直接把姓名、邮箱 (必须是.edu .ac)、电话发送到: challenges@huawei.com 直接开通账号。

Thank you.

把数字世界带入每个人、每个家庭、
每个组织，构建万物互联的智能世界。

Bring digital to every person, home, and
organization for a fully connected,
intelligent world.

**Copyright©2018 Huawei Technologies Co., Ltd.
All Rights Reserved.**

The information in this document may contain predictive statements including, without limitation, statements regarding the future financial and operating results, future product portfolio, new technology, etc. There are a number of factors that could cause actual results and developments to differ materially from those expressed or implied in the predictive statements. Therefore, such information is provided for reference purpose only and constitutes neither an offer nor an acceptance. Huawei may change the information at any time without notice.

