



2018 机器阅读理解技术竞赛

(2018 NLP Challenge on Machine Reading Comprehension)

2018 机器阅读理解技术竞赛由[中国中文信息学会](#) (CIPS) 和[中国计算机学会](#) (CCF) 联合主办, 百度公司、中国中文信息学会评测工作委员会和计算机学会中文信息技术专委会联合承办。竞赛将于 2018 年 3 月 1 日正式开启报名通道, 获胜团队将分享总额 10 万人民币的奖金, 并将在第三届“[语言与智能高峰论坛](#)”举办技术交流和颁奖。在此, 诚邀学术界和工业界的研究者和开发者参加本次竞赛!

竞赛背景

[中国中文信息学会](#)和[中国计算机学会](#)于2016年和2017年联合发起了两届“[语言与智能高峰论坛](#)”, 邀请了国内外相关领域、学术界和工业界的知名专家学者, 共同探讨语言与智能领域的新发展和新技术。第三届“[语言与智能高峰论坛](#)”将于2018年7月28日在北京召开, 除向社会公众介绍国际语言与智能及相关领域的发展趋势和创新成果外, 本届会议还将举办机器阅读理解技术竞赛, 进一步推动语言与智能领域的技术交流和发展的。

机器阅读理解(Machine Reading Comprehension) 研究近年来受到广泛关注, 任务通常定义为: 让机器阅读文本, 然后回答和阅读内容相关的问题。阅读理解涉及到语言理解、知识推理、摘要生成等复杂技术, 极具挑战。该任务的研究对于智能搜索、智能推荐、智能交互等人工智能应用具有重要意义, 是自然语言处理和人工智能领域的重要前沿课题。为了促进阅读理解技术的发展, 本次竞赛将提供面向真实应用场景的大规模中文阅读理解数据集, 为研究者提供学术交流平台, 旨在进一步提升阅读理解的研究水平, 推动语言理解和人工智能领域技术研究和应用的发展。

竞赛任务

■ 任务描述

对于给定问题 q 及其对应的文本形式的候选文档集合 $D=d_1, d_2, \dots, d_n$, 要求参评阅读理解系统自动对问题及候选文档进行分析, 输出能够满足问题的文本答案 a 。目标是 a 能够正确、完整、简洁地回答问题 q 。

输入/输出:

◇ 输入: 问题 q 及其对应的候选文档集合 D

◇ 输出： 满足用户问题 q 的文本答案 a

■ 数据集

竞赛数据集包含 30 万来自百度搜索的真实问题，每个问题对应 5 个候选文档文本，以及人工撰写的优质答案。数据集划分为 28 万的训练集，1 万开发集和 1 万测试集。该数据集中包含了 [DuReader](#) 中已发布的 20 万问题数据，可自由下载（[下载地址](#)）用于预训练和测试。竞赛报名团队将获得新增的 10 万问题数据集。

■ 评价方法

基于测试集的人工标注答案，采用 ROUGH-L 和 BLEU 作为评价指标。全部测试集结果（即主任务）作为最终评价结果。

■ 基线系统

竞赛将提供两个开源的阅读理解基线系统，基线系统的实现及结果评价请参考：[开源系统](#)和数据集[论文](#)。

奖项设置

竞赛将评出一等奖 1 名，二等奖 2 名，三等奖 3 名，主办方中国中文信息学会和中国计算机学会（CCF）将为获奖者提供荣誉证书认证。同时，百度将为获奖者提供奖金和参会交流赞助。

- 一等奖： 50,000 +荣誉证书
- 二等奖： 20,000 +荣誉证书
- 三等奖： 3,000 +荣誉证书

重要时间

- 2018/3/1: 启动竞赛报名，竞赛平台开放，发放样例数据；
- 2018/3/31: 报名截止，对报名者发放全部训练数据；
- 2018/4/23: 发放测试数据；
- 2018/4/30: 系统结果提交截止；
- 2018/5/15: 公布竞赛结果，接收系统报告和论文；
- 2018/7/28: 在“语言与智能高峰论坛”上交流和颁奖；

竞赛组织

- 主办方： 中国中文信息学会，中国计算机学会
- 组织方

- 百度公司
- 中国中文信息学会评测工委会
- 中国计算机学会中文信息技术专委会

■ 指导委员会

- 孙 乐 中国科学院软件技术研究所
- 周 明 微软亚洲研究院
- 杨尔弘 北京语言大学
- 赵东岩 北京大学
- 吴 华 百度公司

■ 评测委员会

- 吕雅娟 百度公司
- 韩先培 中国科学院软件研究所
- 万小军 北京大学
- 刘 凯 百度公司

注册报名

预报名：即日起至 2018 年 2 月 28 日，有意向参加的单位和个人可以扫描下方二维码进行预报名。预报名的团队在竞赛网站正式报名开通后会收到邮件通知，正式报名后各团队将会获得大赛纪念礼品一份。



正式报名：竞赛网站将于 2018 年 3 月 1 日正式开放注册报名，请及时关注。正式报名并最终提交有效结果的团队每位成员均将获得一件大赛定制纪念 T 恤。

联系我们

任何与本次技术竞赛相关的问题，请随时联系竞赛会务组。

竞赛会务组邮箱：MRC2018@126.com