# 从 ACL'22 投稿情况，速览当下 NLP 研究热点！

原创 Yimin_饭煲 夕小瑶的卖萌屋 2021-12-07 12:05

**文 | Yimin_饭煲**
**编 | 小轶**

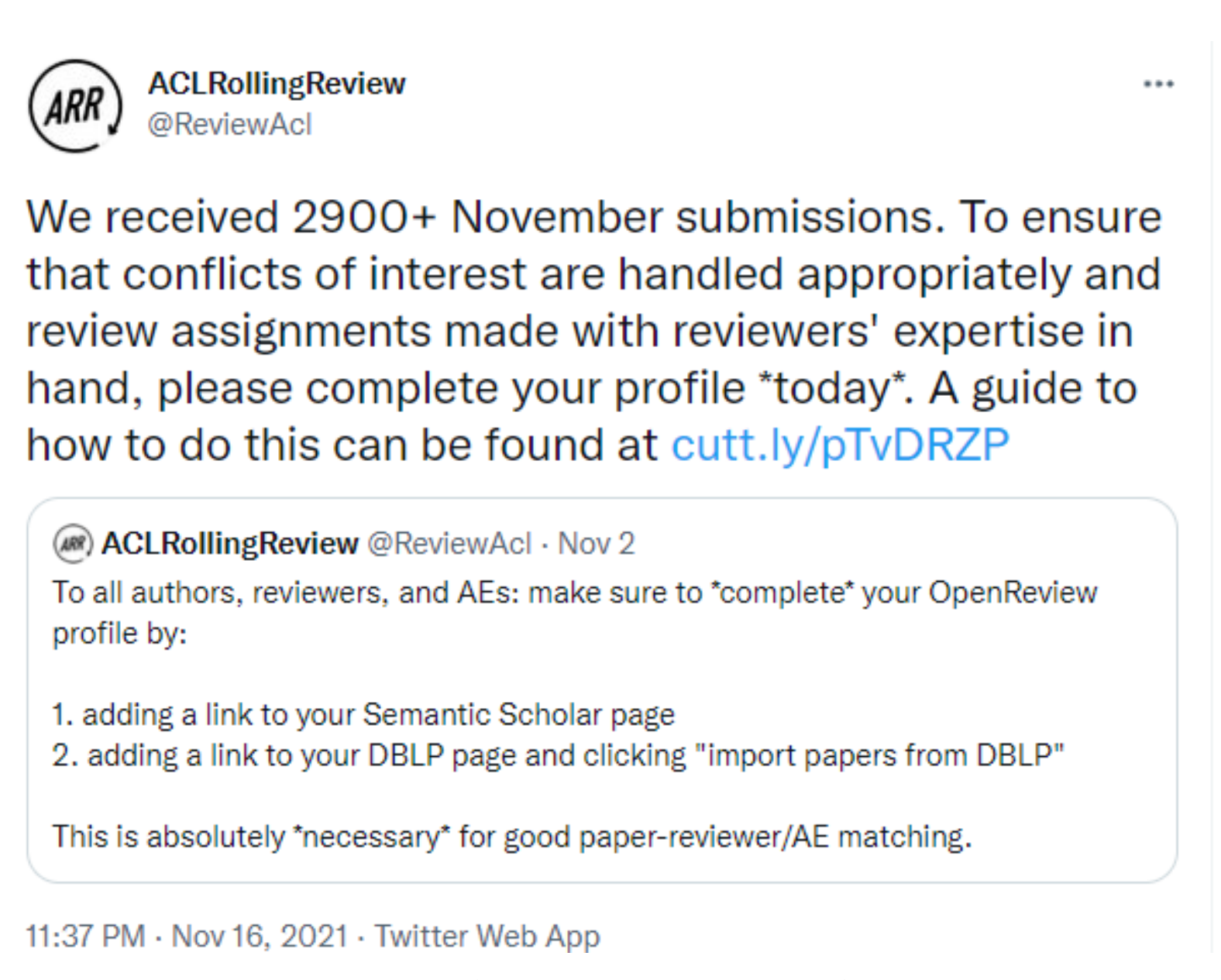
卖萌屋的作者们，最近可真是忙秃了头~，不仅要苦哈哈地赶 ACL 2022 提前了两个月的 Deadline，还要尽心尽力为读者们提供高质量的内容。如果大家心疼卖萌屋的作者们的话，还请多多一键三连:)

ACL2022 全部转向了使用 ACL Rolling Review(ARR) 投稿，所有的投稿必须提交到 ARR 11 月及其之前的 Rolling Review (每月可以投稿一次)。考虑到大多数 NLPer 都是 DDL 战士，因此大多数投稿都集中在了 ARR 11 月，可以预见 ACL 2022 的大多数投稿都来源于 ARR 11 月投稿。因此，可以通过对 ARR 11 月投稿的分析，来前瞻 ACL 2022 上的研究趋势~此外，这也是NLP第一顶会 ACL 第一次采用 OpenReview.net 作为投稿的网站，允许作者们发布匿名预印本。下面，就让笔者带着大家一起读读ARR 11月这些匿名发布的投稿，**提前两个月预告一下 ACL2022 上的那些热点吧~**。

**11 月 ARR 网址链接**：
*https://openreview.net/group?id=aclweb.org/ACL/ARR/2021/November*

## 总览：984篇匿名公开投稿 2900+总投稿量

首先总览一下 ARR 11 月份的投稿情况。尽管 ACL 2022 全部采用 ARR 的形式让 NLPer 们的 ACL deadline 一下提前了两个月，不过各位 NLPer 的爆肝能力还是很给力的！**总的投稿数目接近 3000 篇**，如果在算上前几个月 Rolling Review 投稿到 ACL 2022 的数目，ACL 2022 的投稿数目应该能够达到 ACL 2021 时 3300 篇左右的水平。



ACLRollingReview
@ReviewAcl

We received 2900+ November submissions. To ensure that conflicts of interest are handled appropriately and review assignments made with reviewers' expertise in hand, please complete your profile *today*. A guide to how to do this can be found at cutt.ly/pTvDRZP

ACLRollingReview @ReviewAcl · Nov 2
To all authors, reviewers, and AEs: make sure to *complete* your OpenReview profile by:

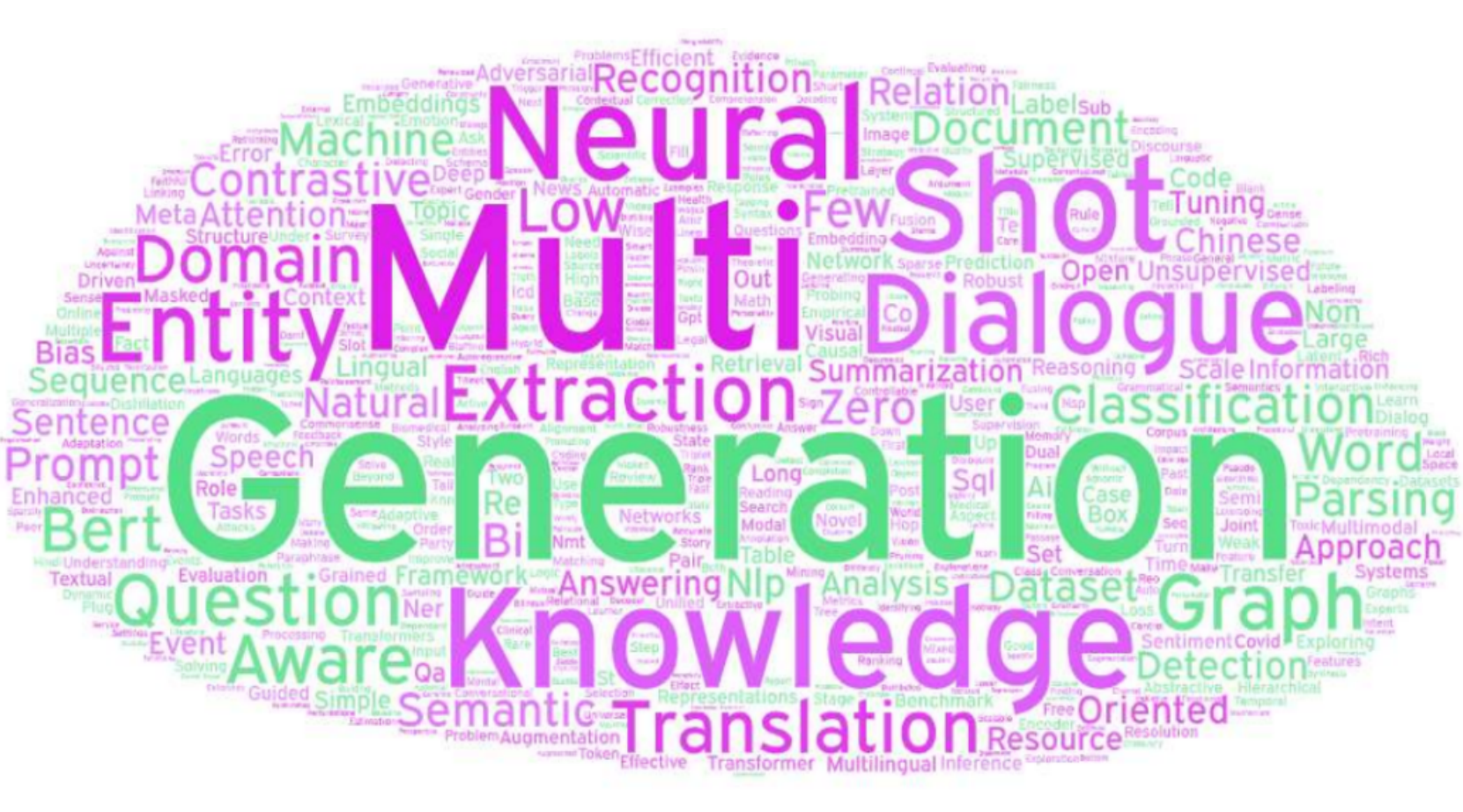
1. adding a link to your Semantic Scholar page
2. adding a link to your DBLP page and clicking "import papers from DBLP"

This is absolutely *necessary* for good paper-reviewer/AE matching.

11:37 PM · Nov 16, 2021 · Twitter Web App

▲ARR 在 11 月收到了 2900+ 份投稿

今年也是(据笔者所知) ACL 主会第一次使用 OpenReview 网站进行投稿。不过并不要求所有投稿必须公开匿名的预印本，而是可以选择公开或者不公开。同时，所有的 Review 也并不会被公开(避免了公开处刑的尴尬~)。**有 984 篇投稿选择了公开匿名预印本**，因此大家可以先对这些论文一睹为快~



▲ARR 公开投稿标题词云

笔者可视化了984篇论文投稿的标题云图，可以从图中看出

- 标题含有 "**Multi**" 的论文占有很大的比重，不管是 "Multi-lingual", "Multi-Modal" 还是 "Multi-domain" 都频频出现。可见 NLP 领域正在朝着更通用、更全面的方向发展。
- **传统的 NLP 方向，如对话、问答、命名实体识别、分类等领域仍然占有重要的地位**
- **Prompt 异军突起**，在图中左侧颇为显眼
- **Contrastive Learning** 风头仍劲，依然占有一席之地（图中左上角）
- 三年过去，**BERT** 依然稳居流量高位

下面，笔者将为大家一一解读在已公开投稿中笔者认为有趣的点~和大家**聊聊ARR 11月投稿中体现出的NLP发展趋势**！

## "Multi-X" 研究火热，通用智能研究正当其时

在 984 篇匿名公开的投稿预印本中，有接近 100 篇投稿中的标题都涉及到 "Multi"，**达到了 10% 的极高比例**。其中，有 30 篇左右的工作关注 "**Multi-Modal**"（多模态），有 25 篇左右的工作关注"**Multi-lingual**"（多语言），有 10 篇左右的工作关注 "**Multi-task**"(多任务)。随着更通用的深度学习架构的发展(例如Transformers)，各种模态、各种语言的数据都可以被同一个模型进行处理，同时，通用的Backbone模型也可以为各类丰富的下游任务提供强大的表示，因此具有强大的多任务能力。未来，面向多语言、多模态、多任务的通用深度学习模型必然会快速发展，不仅能做到“一个模型走天下”，降低训练多个模型的开销，更能通过多个模态之间的借鉴和补充提升每个模态和任务上的性能。

## Prompt 异军突起，将成2022年NLP界最大热点

如果要评选NLP界2021年的关键词，**Prompt Learning**一定是许多人的选择（当然，2021的最后一个月是否会出现新的爆点也尚未可知）。CMU 关于 Prompt Learning 的综述，把这一领域送入了许多人的视线。卖萌屋团队对 Prompt Learning 这一领域一直保持了高度的关注，欢迎大家关注卖萌屋的相关推文：《格局打开，带你解锁 prompt 的花式用法》，《一文跟进Prompt进展！综述+15篇最新论文逐一梳理》，《别再Prompt了！谷歌提出tuning新方法，强力释放GPT-3潜力！》。

今年的 NLP 各大会议投稿，Prompt Learning 可以算是增长最快的研究热点了，笔者粗略统计了ARR 11月的投稿，有31篇和Prompt Learning有关的工作，**不仅有在信息抽取、语义解析、命名实体识别、事件检测、文本生成等传统NLP领域的应用，还有在多模态、视觉-语言等交叉领域中的尝试**。有趣的是，在卖萌屋的推送《别再Prompt了！谷歌提出tuning新方法，强力释放GPT-3潜力！》的评论区，有两位读者表示希望探索 Prompt Learning 在信息抽取领域的应用，不知 ARR 11 月的几篇关于 Prompt Learning 在信息抽取中应用的投稿，有没有哪篇来自于卖萌屋的读者呢~

JayJay 19
哈哈，感觉还是Prompt-tuning，只不过Prompt上升到task层面了。此外，虽然用更少的参数超过GPT3，但用了更多的数据呀😬
还有一点：为什么好多Prompt-tuning不敢去做信息抽取任务呢😁

张博 4
像T5吧

睡不着的宇航员
Prompt 和 instruction 本质区别在哪里？用起来差不多啊！

Didi
这不就是多任务么？😂
回jayjay：ie任务我们搞起来啊😁

▲往期推送《别再Prompt了！谷歌提出tuning新方法，强力释放GPT-3潜力！》的评论区

## 对比学习应用广泛，或将成为语言表示学习标准范式之一

在ACL 2021中，NLPer 们纷纷将 CV 中研究火热的对比学习迁移到 NLP 领域中。ACL21的接受论文中有21篇论文题目包含了"contrastive"，卖萌屋团队往期推文《我分析了ACL21论文列表，发现对比学习已经...》中，我们详细分析了这些论文的主要研究点。一年过去了，对比学习在NLP领域依然热度不减，今年的 ARR 11 月投稿中有 37 篇涉及到对比学习的工作，几乎涵盖了 NLP 中的所有领域，不仅有**句子表示和文档表示**生成等经典应用场景，更有**拼写纠错、文本总结、事实验证等新兴应用领域**。对比学习能帮助模型生成性质更好的语义表示，对无标注样本有着更好的利用。未来，也许对比学习会成为语言表示学习领域的标准范式之一！

## 数据集价值越发重要，新兴 Benchmark 赋能 NLP 快速发展

数据集的发展和方法的进步互相耦合，共同推进了 NLP 领域的发展。目前在 NLP 领域的方法层面，大多基于 Transformer 结构进行改进，在模型结构和算法方面的创新似乎陷入了“瓶颈期”。于是，**发展更新、更难、应用场景更丰富的数据集**，定义新的任务场景，成为了越来越多工作的着力点。在 ARR 11 月的已公开投稿中，**有 50 篇以上提出新数据集和评测基准的工作**。不仅涉及了**少样本学习、对抗样本检测、因果推理、开放域问答等传统的NLP领域**，同样有许多关注垂直领域自然语言处理应用的数据集，例如**法律文本自然语言处理、生物医药自然语言处理、短视频标题生成、手语识别、科幻文本理解**。在可预见的未来，这些数据集将促进对模型、更全面的比较，加速垂直领域自然语言处理的学研究和产业应用。

## 速览中文领域NLP工作，期待中文研究产生更大国际影响力

知乎上有一个问题“**为什么中文NLP数据集这么少？**”，吸引了刘知远、邱锡鹏老师等国内 NLP 学术界大神回答。尽管在中文领域已经有了许多大模型可供使用，也出现了 CLUE 等中文 NLP 评测基准，但在多模态、问答、文本摘要等细分领域仍然缺少高质量的中文数据集发展，在一些领域的中文数据集甚至是由英文数据集直接翻译而来。此外，由于中文的文字和语言学特征与主流的研究语言英语有着较大的差距，基于英语研究得到的方法未必能迁移到中文上得到好的效果。建立全面高质量的中文 NLP 评测基准，发展面向中文优化的 NLP 方法，不仅能促进中文 NLP 学术界和工业界的研究和应用，更有助于提升中文 NLP 研究在国际学术界的影响力。**11 月的 ARR 投稿中有 27 篇面向中文 NLP 的研究工作**。在数据集方面，有工作提出了**中文领域的小样本关系链接基准、科技类文本数据集、新闻摘要数据集、对话常识知识图谱、生物医药文本理解数据集、短视频标题生成基准**。在方法方面，解决了**中文命名实体识别的优化、拼音输入的优化、少数民族语言预训练、中文拼写检查、古诗情感分类**等一系列极具中文特色的研究问题。期待中文NLP研究继续蓬勃发展，产生更大的国际影响力。

## 趣谈：NLP 界又出现了多少 All you need?

如果要选取五年来NLP领域中最经典的一篇工作，那么 *Attention Is All You Need* 这篇工作应该是一个大多数人都能够信服的选择。这篇工作提出了现在 NLP 领域的核心结构 Transformer。基于 Transformer 结构的模型已经逐渐在NLP、多模态乃至计算机视觉领域都成为了主流。同样抓人眼球的是这篇文章的标题，一时间在AI圈掀起了一股 “XXX is all you need” 的取名潮流。笔者了解到“All you need"的就包括且不限于 **CNN/Pre-training/Image Augmentation/Depthwise convolution/Bytes/Focus/Channel Attention is All you need**。当然，还有业内人士时常调侃的所谓"Money is All you need"。ARR 11月的投稿中，也有两篇用 All you Need 起名的投稿，分别是 **Multimodal Learning: Are Captions All You Need?** 和 **Tokenization on the Number Line is All You Need**。分别描述了在**视觉语言学习中使用标题替代视觉信号的作用**和**一种优化数字分词的方式**，都是很有趣的短文~推荐一读。

## 趣谈：最长标题和最短标题

大多数的AI论文，标题长度都在5个词到15个词左右，不过总有一些有趣的论文有着很长或者很短的标题。ARR 11月投稿中**标题最长的论文**：**Innovative Measures of Patient and Disease Phenotyping: Optimizing Linguistic and Machine Learning Techniques in the Investigation of Electronic Health Record (EHR) Data**。本文描述了一种通过结合语言特征工程、竞争建模和人类反馈的电子健康数据利用方式。ACL ARR 11月投稿中标题**最短的论文仅有一个词: EventBERT**。本文通过结合事件相关的语义表示使得BERT从基于事件的图结构和语义表示中获益，并在GLUE上验证了有效性。

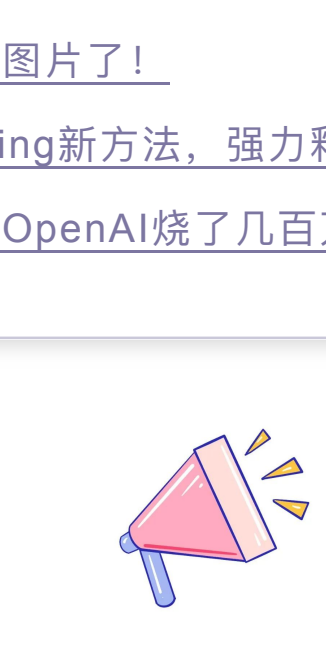
## 趣谈：三年之后，又有多少新BERT?

2018年BERT的横空出世，改变了NLP领域研究的格局，基于BERT架构的模型迅速占据了各大排行榜的前列，出现了各种各样修改版的 "XXBERT"。ARR 11 月的投稿中，也有许多以 “XXBERT” 命名的工作，包括 KNN/Kinya/Bangla/ga/Lord/Mark/Aleph/Cal/Pinyi/PromptBERT 等各式各样的 BERT 出现，面向少数群体语言、Prompt Learning 等任务上基于 BERT 模型得到了优秀的效果。尽管有许多尝试改进 Transformers 结构的工作出现，但是在大多数主流的 NLP 任务中，BERT 还是 YYDS！

## 结语

读完这篇 ACL 2022 投稿前瞻预告后，不知道大家对 ACL 2022 的精彩是不是更加期待了呢~

最后想说，Paper诚可贵，健康价更高。希望卖萌屋的读者们，在辛苦赶完ACL DDL之后，多休息多运动，准备好迎接下一个项目的挑战吧~最后祝大家的 ACL 投稿都有一个好结果，Paper 高中！现在越来越多的会议开始转为 openreview 的形式。如果大家对这篇解读 ACL Rolling Review 的工作感兴趣的话，卖萌屋日后也会努力推出更多的类似解读！



**萌屋作者：Yimin_饭煲**
在微软NLC组搬砖的联培博士生，爱好摄影和运动，希望卖萌屋早日开通视频业务，我来当摄影师！
**作品推荐**

1. 学完文本知识，我就直接看懂图片了！
2. 别再Prompt了！谷歌提出tuning新方法，强力释放GPT-3潜力
3. 大模型炼丹无从下手？谷歌、OpenAI烧了几百万刀，总结出这些方法论…

后台回复关键词【**入群**】
**加入卖萌屋NLP/IR/Rec与求职讨论群**
后台回复关键词【**顶会**】
**获取ACL、CIKM等各大顶会论文集！**