


Projects

在此输入您的封面副标题

Projects


https://www.biendata.com




En/中 我要办赛 Models 讨论区 登录 注册


重磅比赛

全部 >

【人工智能大赛】

正在进行


Name Disambiguation
OAG 2019




OAG-WholsWho 赛道二

智源 - AMiner 姓名排歧大赛 赛...


主办方: 智源、AMiner


奖励: ¥50,000

时间轴: 2019-09-30 ~ 2019-12-02

【人工智能大赛】

正在进行


Name Disambiguation
OAG 2019




OAG-WholsWho 赛道一

智源 - AMiner 姓名排歧大赛 赛...

主办方: 智源、AMiner

奖励: ¥50,000

时间轴: 2019-09-30 ~ 2019-12-02




已结束

DigSci 科学数据挖掘大赛 ...
DigSci 科学数据挖掘大赛 2019

主办方: AMiner · Microsoft · biendata

奖励: ¥40,000

时间轴: 2019-10-02 ~ 2019-10-12


FALSE NEWS DETECTION
智源·假新闻识别
MCG-FNeWS'19

正在进行

智源&计算所-互联网虚假...

主办方: 智源研究院·中科院计算所

奖励: ¥100,000

时间轴: 2019-08-30 ~ 2019-10-24

人工智能大赛

正在进行


BAAI-ZHIHU CUP
智源·看山杯
FindExp



智源 - 看山杯 专家发现算...


“达观杯”
文本智能信息抽取
挑战赛

已结束

“达观杯”文本智能信息抽...



已结束

Objects365

旷视 

Objects365 图片物体检测



已结束

CrowdHuman

旷视 

CrowdHuman 人体检测...

Projects

- **每人**从以下介绍的7个Projects中选择一个作为你课程的Project; Project的具体描述、数据集、评测标准请在相关链接中查看;
- **请及时创建账号报名感兴趣的任務 (Project) , 并且在任务的截止日期之前提交代码**, 保存在线的评测结果, 作为最终分数的依据之一;
- Project Report
 - 按照论文的格式写, 包括Abstract, Introduction, Related Work, Approach, Experiments, Conclusion等部分;
 - Report中英文皆可以, 鼓励大家用英文写, 且用:
 - The ACM Proceedings Format, using either the sample-sigconf.tex or ACM_SigConf.docx template provided at <https://www.acm.org/publications/proceedings-template> for LaTeX.
- Project Slides:
 - 做一个15分钟左右的slides, 介绍自己的工作;
- 截止时间: **12月31日** (代码、Report、Slides)

Project1: 论文的冷启动消歧

(Name Disambiguation from Scratch)

- 任务描述：<https://biendata.com/competition/aminer2019/>
 - 给定一堆拥有同名作者的论文，要求返回一组论文聚类，使得一个聚类内部的论文都是一个人的，不同聚类间的论文不属于一个人。最终目的是识别出哪些同名作者的论文属于同一个人。
- 参考方法：
 - 解决这一问题的常用思路就是通过聚类算法，提取论文特征，定义聚类相似度度量，从而将一堆论文聚成的几类论文，使得聚类内部论文尽可能相似，而类间论文有较大不同，最终可以将每一类论文看成属于同一个人的论文。
 - [1] 是一篇经典的使用聚类方法的论文，它使用了原子聚类的思想，大致思路是首先用较强的规则进行聚类，例如：两篇论文如果有两个以上的共同作者，那么这两篇论文属于同一类，这样可以保证聚类内部的准确率，随后用弱规则将先前的聚类合并，从而提高召回率。
 - 有些工作考虑了传统特征的局限性，所以利用了低维语义空间的向量表示方法，通过将论文映射成低维空间的向量表示，从而基于向量使用聚类方法 [2]。
 - [1] Name Disambiguation Using Atomic Clusters. Web-Age Information Management, WAIM '08.
 - [2] Name Disambiguation in AMiner: Clustering, Maintenance, and Human in the Loop.(KDD'18).

Project2:论文的增量消歧 (Continuous Name Disambiguation)

- 任务描述：
 - 线上系统每天会新增大量的论文，如何准确快速的将论文分配到系统中已有作者档案，这是线上学术系统最亟待解决的问题。所以问题抽象定义为：给定一批新增论文以及系统已有的作者论文集，最终目的是把新增论文分配到正确的作者档案中。
- 参考方法：
 - 增量消歧任务与冷启动消歧的任务不同，它是基于有一定作者档案的基础，对新增论文进行分配。所以，容易直接想到的方法就是将已有的作者档案与新增论文进行比较，提取合作者，单位机构或者会议期刊之间相似度的传统特征，随后利用svm之类的传统分类器进行分类。还可以利用基于低维空间的向量表示方法，通过将作者与论文表示成低维向量，使用监督学习方法进行特征提取及模型训练。

https://biendata.com/competition/aminer2019_2/

Project1-2: 参考文献清单

- [1]. Jie Tang, Jing Zhang, Limin Yao, Juanzi Li, Li Zhang, and Zhong Su. ArnetMiner: Extraction and Mining of Academic Social Networks. In Proceedings of the Fourteenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (SIGKDD'2008). pp.990-998.
- [2]. Yutao Zhang, Fanjin Zhang, Peiran Yao, and Jie Tang. Name Disambiguation in AMiner: Clustering, Maintenance, and Human in the Loop. In Proceedings of the Twenty-Forth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'18).
- [3]. Jie Tang, A.C.M. Fong, Bo Wang, and Jing Zhang. A Unified Probabilistic Framework for Name Disambiguation in Digital Library. IEEE Transaction on Knowledge and Data Engineering (TKDE), 2012, Volume 24, Issue 6, Pages 975-987.
- [4]. Xuezhi Wang, Jie Tang, Hong Cheng, and Philip S. Yu. ADANA: Active Name Disambiguation. In Proceedings of 2011 IEEE International Conference on Data Mining (ICDM'11), pages 794-803.
- [5]. <https://biendata.com/competition/scholar2018/data/>
- [6]. The Microsoft Academic Search Dataset and KDD Cup 2013
- [7]. Wang, F. , Li, J. , Tang, J. , Zhang, J. , & Wang, K. . (2008). Name Disambiguation Using Atomic Clusters. Web-Age Information Management, 2008. WAIM '08. The Ninth International Conference on.

Project3:专家发现/问题路由 (Expert Detection/Question Routing)

- 提供知乎的问题信息、用户画像、用户回答记录，以及用户接受邀请的记录，**要求选手预测这个用户是否会接受某个新问题的邀请。**
 1. 问题信息。包括<问题id、问题创建时间、问题的话题、问题的文本、问题的描述>等。
 2. 用户的回答。包括<回答id、问题id、作者id、回答的文本、回答时间、点赞数、收藏数、感谢数、评论数>等。
 3. 用户画像数据。包括<用户id、性别、活跃频次、关注话题、长期兴趣、盐值>等。
 4. <topic、token(词)、单字 64维 embedding >数据。
 5. 最近一月的邀请数据包括<问题id、用户id、邀请时间、是否回答>。

<https://biendata.com/competition/zhihu2019/>

Project4： 虚假新闻文本检测

- 文本是新闻信息的主要载体，对新闻文本的研究有助于虚假新闻的有效识别。
- 具体任务为：**给定一个新闻事件的文本，判定该事件属于真实新闻还是虚假新闻。**

<https://biendata.com/competition/falsenews/>

Project5： 虚假新闻图片检测

- 虚假新闻图片是指虚假新闻中的配图。在虚假新闻中，新闻配图往往也包含着丰富的信息，有助于虚假新闻的判别。
- 具体任务为：给定一张图片，要求参赛者判断该图片是虚假新闻图片还是真实新闻图片。

https://www.biendata.com/competition/falsenews_2/

Project6： 虚假新闻多模态检测

- 随着多媒体技术的发展，新闻当中通常都包含着文本与图片等多模态信息，不同模态之间既存在着增强关系，同时也包含着互补信息。如何充分利用多模态信息进行虚假新闻检测仍是一项具有挑战的工作，为此，我们设立虚假新闻多模态检测子任务以促进该领域研究。
- 具体任务为：给定一条新闻的多模态内容，包括文本、配图、用户特征等，要求参赛者判断该新闻属于虚假新闻还是真实新闻。

https://www.biendata.com/competition/falsenews_3/

Project4-6: 参考文献清单

- Jin Z, Cao J, Zhang Y, et al. News verification by exploiting conflicting social viewpoints in microblogs. AAAI 2016.
- Jin, Z., Cao, J., Zhang, Y., Zhou, J., & Tian, Q. Novel visual and statistical image features for microblogs news verification. TMM, 19(3).
- Jin, Z., Cao, J., Guo, H., Zhang, Y., & Luo, J. Multimodal fusion with recurrent neural networks for rumor detection on microblogs. MM 2017.
- Guo H, Cao J, Zhang Y, et al. Rumor detection with hierarchical social attention network. CIKM 2018.
- Qi P, Cao J, Yang T, et al. Exploiting Multi-domain Visual Information for Fake News Detection. ICDM 2019.
- Guo C, Cao J, Zhang X, et al. Exploiting Emotions for Fake News Detection on Social Media. arXiv:1903.01728.

Project7：疾病问答迁移学习

- 迁移学习是自然语言处理中的重要一环，其主要目的是通过从已学习的相关任务中转移知识来改进新任务的学习效果，从而提高模型的泛化能力。
- 任务的主要目标是针对中文的疾病问答数据，进行病种间的迁移学习。具体而言，给定来自5个不同病种的问句对，**要求判定两个句子语义是否相同或者相近**。所有语料来自互联网上患者真实的问题，并经过了筛选和人工的意图匹配标注。
- 示例

问句1：糖尿病吃什么？	问句1：乙肝小三阳的危害？	https://biendata.com/competition/chip2019/
问句2：糖尿病的食谱？	问句2：乙肝大三阳的危害？	http://cips-chip.org.cn/evaluation
label:1	label:0	
- 报名时间：9月2日—10月14日
- 报名方式：参赛队伍由1~5人自由组队，通过邮件提交报名信息，发送报名表至920625863@qq.com，邮件标题为CHIP2019评测2+参赛队伍名称，报名信息包含参赛队名，成员姓名，联系邮箱，手机号码和参赛队伍名称 此外，请已报名队伍加比赛官方QQ群：729245835。后续比赛信息会在本群公布。联系人：骆迅,920625863@qq.com