

求职意向	行业：互联网、金融、医疗、制造业 职位：AI 算法研发						
个人优势	<ul style="list-style-type: none"> ◇ 机器学习算法理论与实践经验都很丰富，拥有全栈开发能力。 ◇ 有优异的国际 AI 算法竞赛成绩和公司业绩。 ◇ 能研发，也能同时带技术团队。 						
教育背景							
2007/9-2010/7 研究生 (统招)	北京大学 计算机科学技术系，计算机系统结构专业						
2003/9-2007/7 本科 (统招)	北京信息科技大学 计算机信息系，信息管理与信息系统专业						
兴趣爱好	街头健身，长跑，游泳，物理学，哲学						
专业技能							
◇ 专业 竞赛 经历：							
年份	赛事名称	举办单位	应用领域	队伍数	名次/TOP	参赛形式	涉及算法
2023 辞职 期间	Google: asl-fingerspelling	Google kaggle	手语识别	1315	第 16 名 Top:1%	SOLO	CV&NLP Seq2Seq
2021	Jane Street Market Prediction	Jane Street Kaggle	量化交易	4254	第 10 名 Top:2‰	SOLO	RNN
2016	Di-Tech 算法大赛	滴滴	供需预测	1000+	第 2 名 亚军 +最具潜力大奖	队长 (组员 5 人)	DNN
◇ 技能：							
<ul style="list-style-type: none"> ● 熟悉传统机器学习算法、深度学习算法和强化学习(如 GP, LR, SVM, MF, GBDT,W2V, RNN, CNN, Transformer, DQN, PPO 等等) ● 熟悉AIGC, 大模型相关技术、平台和趋势(DALL-E 2, GPT, Hugging Face 等) ● 熟练使用深度学习相关算法库/框架 (如: Tensorflow、Keras、JAX 等) ● 熟练模型量化优化技术、tf-serving、tflite、混合精度训练、gpu/tpu 分布式训练、docker 环境使用 ● 熟悉大数据相关和流行中间件技术 (Kafka, Hadoop, Storm, Spark, Redis, ES 等) ● 精通 python/java/c++ 等编程语言，熟悉常用数据结构、算法、网络编程、多线程技术 							
项目经验							

2009/2-2010/6 (研三期间)

北京大学 网络与分布式系统实验室 天网 Maze 资源推荐系统 项目负责人

研发了基于聚类和索引结构的等价增量聚类方法
研发了基于文本和基于用户行为的混合推荐算法

2010/7-2012/7

腾讯 数据平台部 精准推荐中心 研究员

- **通用推荐系统**研发人员。主要负责游戏(LOL,CF)道具、手 Q 吃喝玩乐、QQ 空间商城等等产品的推荐系统算法设计及实现。职责：设计并实施算法模型，配合平台研发人员进行线上算法实验，分析算法效果，调整参数改进模型。
- **灯塔项目**核心研发人员之一，负责实时 ETL 和实时计算功能研发。

2012/7-2018/12

百分点科技

(第一阶段) 推荐引擎 核心研发人员

- 研发了增量 Item CF 推荐算法
- 研发了基于热传导和概率扩散相结合的离线推荐算法
- 研发了基于关联规则的离线推荐算法

(第二阶段) 互联网业务事业部 研发部 研发 leader

- **推荐算法引擎 Proxima**：完全自主研发了一套实时分布式推荐算法引擎，使得公司在海量数据（亿级用户、千万级物品、十亿级消息）冲击下能高效的提供 SaaS 服务。
【细节：以上推荐算法引擎核心以及主体代码均由本人编写完成，本系统上线后，使得公司的推荐服务能力有了质的突破：解决了单机版推荐引擎对大客户无法服务的问题；**实时性**由天级别**提升到分钟**级别；推荐效果大幅提升。并且新框架**节省了**原有**一半以上的服务器资源。**】
- **用户拉通系统 Gemini**：利用流行的机器学习算法对现有的数据资产进行变现，包括：用户拉通系统 Gemini 的研发、用户画像系统优化。
【细节：采用 GBDT+DNN 模型产生预测拉通的概率结果，再结合社区聚类的方法输出拉通的子图，打破了跨屏用户拉通的障碍，增加了用户的全局唯一标示，为用户画像和数据分析面临的信息零散化、信息缺失化的困境打开了局面，给数据变现带来了新的合作机会和更好的合作效果：如 dsp **用户匹配率提高 3 倍**左右，用户画像的**标签覆盖度提升 1 倍**以上。】
- **网页关键信息抽取系统 PanGu**：实现了一套对全网新闻网页关键信息(KIE)(标题、时间、来源、正文)进行自动化抽取的系统。在关键技术采用了双向递归神经网络(Bi-RNN)，对已知来源新闻网站的 **KIE 平均准确度达到 99%**，对未知来源的新闻网站平均准确度达到 80%。
- **交付项目**：分别完成**华为**公司用户拉通与用户画像项目，**国家信息安全中心**用户拉通项目的交付，获得 2015 年度**最佳项目交付团队奖**。
获得公司 **2014 年度（唯一）技术攻关奖**。
2016 年为公司申请了两个**专利**一个**软著**。

2019/1-2021/1

云之梦科技 研发中心 研发 leader

- **数据平台**: 完成数据平台搭建 (从 0 到 1)。
整合了公司各种异构数据, 分离线和实时处理部分, 支持实时数据分析。
- **推荐系统**: 完成推荐系统研发 (从 0 到 1)。
利用公司丰富的异构数据, 提取图片、行为、标签等特征, 采用 **wide&deep model** 训练试衣时长预测模型, 并**创新**研发了深度学习模型中的**特征权重评估**方法。
- **自动化建模**: 完成深度学习自动化建模(AutoML)产品的研发。

2021/5-2023/3

万兴科技 算法开发 (专家算法工程师 T5.1)

- 负责公司**推荐系统**从 0 到 1 的研发。
- 负责公司**搜索系统**优化。
- **搜索推荐统一模型(CV+NLP+Transformer+MMOE+DPP)**
- 以图搜图, 任意图像、视频**风格迁移**应用系统研发。

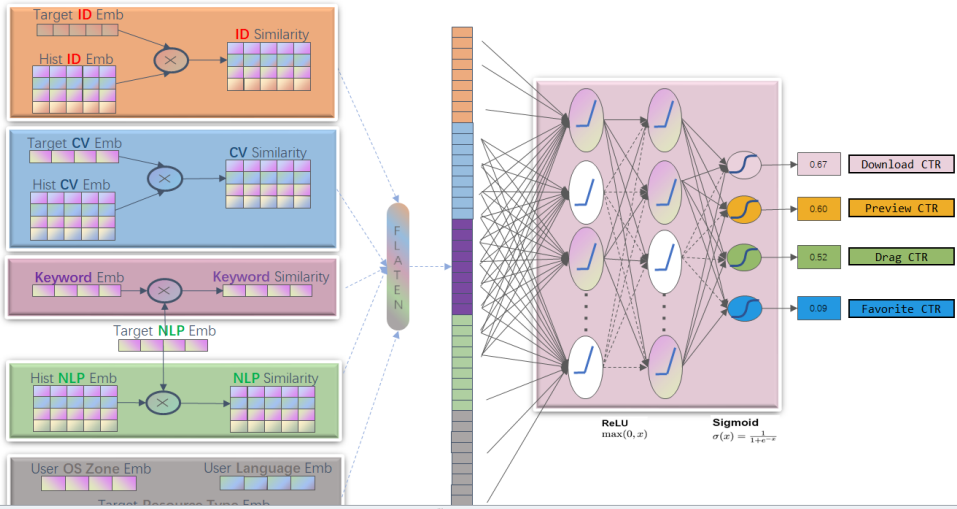
由于推荐效果提升幅度很大, 3 个月试用期后直接晋级。在职期间, 搜索推荐**核心业务指标翻 3 倍以上**。
具体见附件。

附件: _____
—



统一模型

细节(8/8)



点击率整体趋势(4/4)



应用效果

ABTest二季度, 阶段3

召回模型中NLP embedding采用Bert, 多语言统一匹配、向量融合

数据来源: ODS, Filmore-Win 不限定版本

S1: 1月1-3月31日						S2: 6月17日-6月23日						点击率 ABTest提升值 季度同比		
ABTest	算法名	搜索拖拽 点击数	搜索收藏 点击数	搜索预览 点击数	搜索下载 点击数	ABTest 提升	ABTest	算法名	搜索拖拽 点击数	搜索收藏 点击数	搜索预览 点击数		搜索下载 点击数	ABTest 提升
B	title_tag_match_both-dnn_uhist_solo-none	0.012	0.0008	0.0196	0.0172	40.0%	B	title_tag_match_both_bert_lcs-dnn_uhist_solo_bert-none	0.0174	0.00133	0.0379	0.0271	110.4%	85.5%
	title_tag_match_both-dnn_uhist_solo-mir	0.0118	0.0009	0.0195	0.0177	43.7%		title_tag_match_both_bert-dnn_uhist_solo_bert-none	0.0173	0.00158	0.039	0.0277	121.2%	69.7%
	title_tag_match_both-dnn_uhist-none	0.0119	0.0009	0.0208	0.0178	46.8%		title_tag_match_both_bert_lcs-dnn_uhist_bert-none	0.0168	0.00138	0.0375	0.0277	110.7%	67.6%
	title_tag_match_both-dnn_uhist-mir	0.0117	0.0008	0.0189	0.0174	38.0%		title_tag_match_both_bert-dnn_uhist_bert-none	0.0161	0.0015	0.0371	0.0269	110.2%	69.0%
A	title_tag_match_both-none-none	0.0079	0.0007	0.0123	0.0128		A	title_tag_match_both-none-none	0.0076	0.00075	0.0167	0.013		10.2%

点击率=行为数/搜索曝光资源数

S1: 1月1-3月31日						S2: 6月17日-6月23日						点击率 ABTest提升值 季度同比		
ABTest	算法名	搜索下载 点击数	搜索预览 点击数	搜索收藏 点击数	搜索拖拽 点击数	ABTest 提升	ABTest	算法名	搜索下载 点击数	搜索预览 点击数	搜索收藏 点击数		搜索拖拽 点击数	ABTest 提升
B	title_tag_match_both-dnn_uhist_solo-none	3.4603	3.9358	0.1581	2.4234	3.5%	B	title_tag_match_both_bert-dnn_uhist_bert-none	4.4479	6.1321	0.248	2.6645	23.7%	37.8%
	title_tag_match_both-dnn_uhist_solo-mir	3.3551	3.6937	0.1657	2.2361	-0.1%		title_tag_match_both_bert-dnn_uhist_solo_bert-none	4.396	6.1693	0.2493	2.7309	24.4%	42.6%
	title_tag_match_both-dnn_uhist-none	3.5146	4.1043	0.1758	2.3433	6.7%		title_tag_match_both_bert_lcs-dnn_uhist_bert-none	4.2926	5.8176	0.2146	2.603	16.2%	24.3%
	title_tag_match_both-dnn_uhist-mir	3.3371	3.6199	0.1571	2.2444	-1.9%		title_tag_match_both_bert_lcs-dnn_uhist_solo_bert-none	4.2367	5.9272	0.2072	2.7247	17.0%	36.0%

转正前业绩:

◆ 主要业绩

5月28-8月18 完成:

- 完成资源推荐系统设计
- 推荐系统+语义搜索系统上线

推荐系统线上效果:

用户行为全流程

推荐相关事件	人工推荐	算法推荐	与人工推荐整体变化	说明
全局推荐请求	3.26801	4.7796	46%	总推荐次数/有行为的用户数
推荐请求	3.30408	4.8738	48%	总推荐次数/有推荐行为的用户
推荐成功率	90.6%	92.7%	2%	
推荐失败率	9.4%	7.3%	-23%	最影响用户体验的指标
翻页	0.88215	1.5645	77%	影响用户体验的指标
下载	2.38364	4.7256	98%	
预览	1.65746	4.8317	192%	
拖拽	1.02689	4.6275	351%	
收藏	0.03434	0.1391	305%	

◆ 主要业绩

语义搜索线上效果:

用户行为全流程

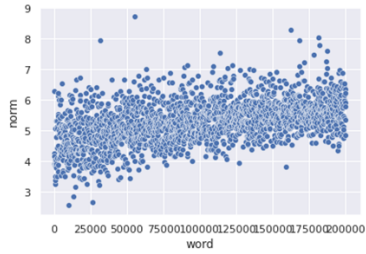
搜索相关事件(人均)	ES搜索	语义搜索	与ES搜索相对变化	说明
全局搜索请求	0.13	0.13	-2%	总搜索次数/有行为的用户
搜索请求	3.35	3.13	-7%	总搜索次数/有搜索行为的用户
搜索成功率	72%	92%	28%	最影响用户体验的指标
搜索失败率	28%	8%	-73%	
翻页	0.0188	0.0747	297%	影响用户体验的指标
下载	0.0678	0.0946	40%	
预览	0.0754	0.1004	33%	
拖拽	0.0780	0.0847	9%	
收藏	0.0037	0.0044	20%	

性能:

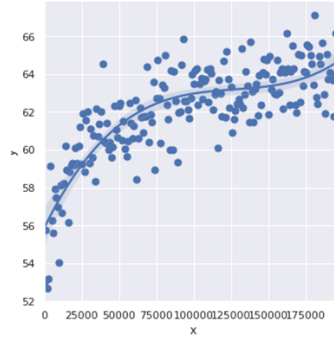
算法服务器: 两台4核32GB内存, 支撑当前推荐30%流量, 搜索50%流量, cpu消耗不到10%。

语言识别技术创新-原理

words norm analysis



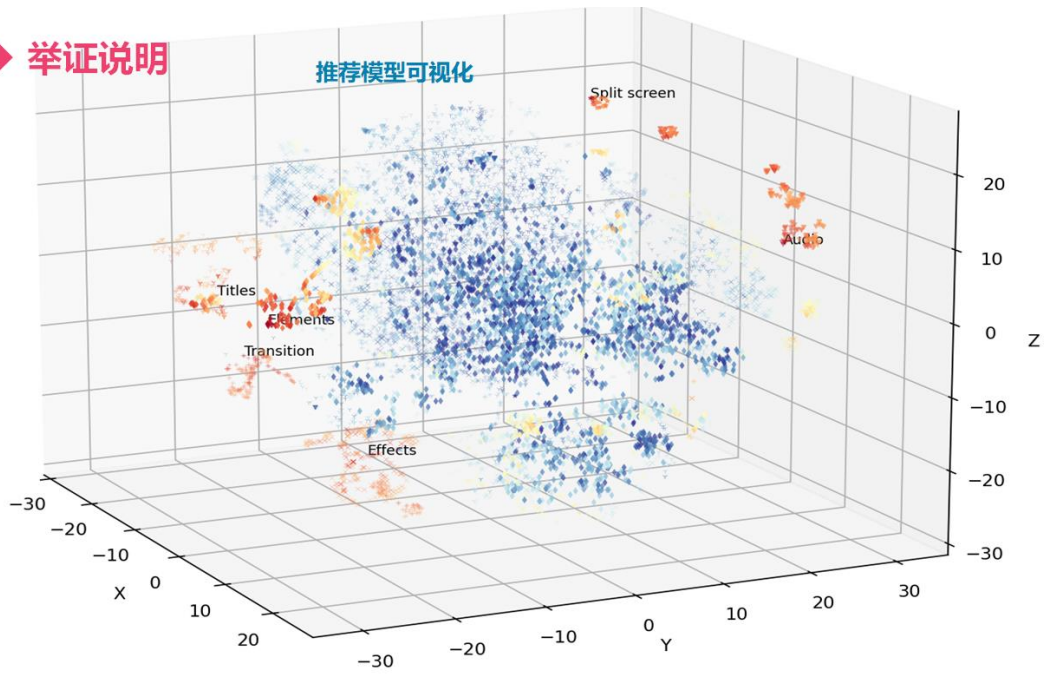
Rare words Gram matrix's Determinant analysis



- 分词空间的秩
- 弃词比例
- 词频加权
- 对应语言资源数权重

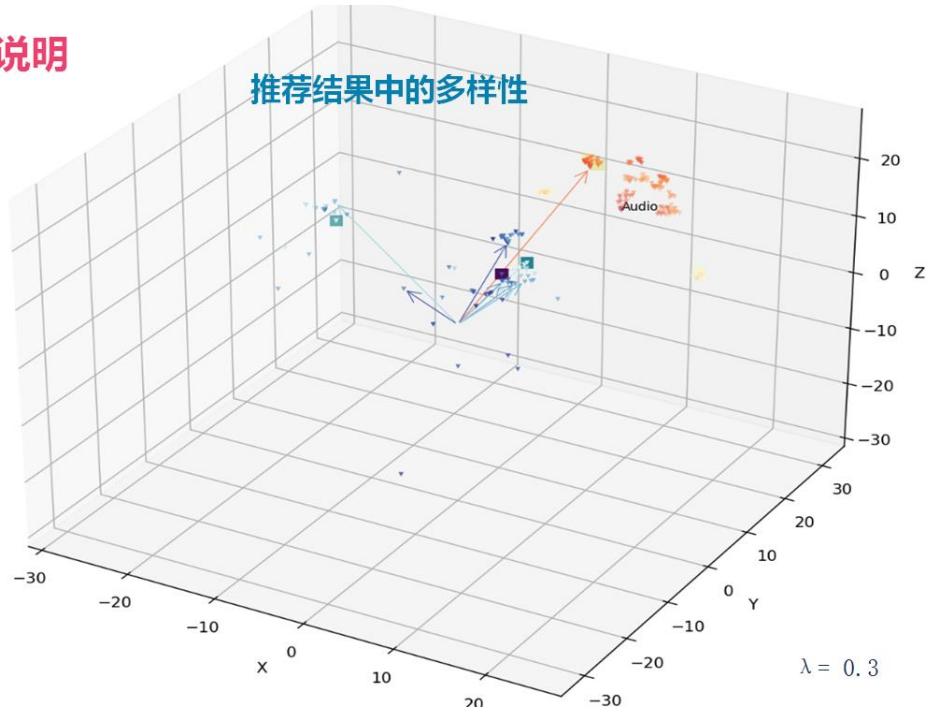
◆ 举证说明

推荐模型可视化



are

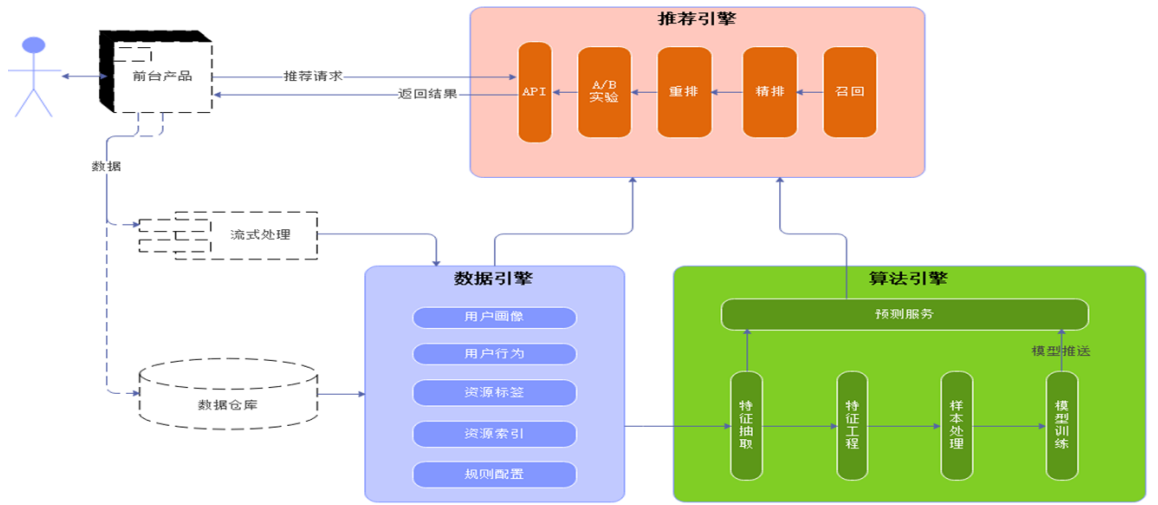
◆ 举证说明



◆ 举证说明 (围绕任职资格认证标准, 结合自身能力及业绩产出进行论证)

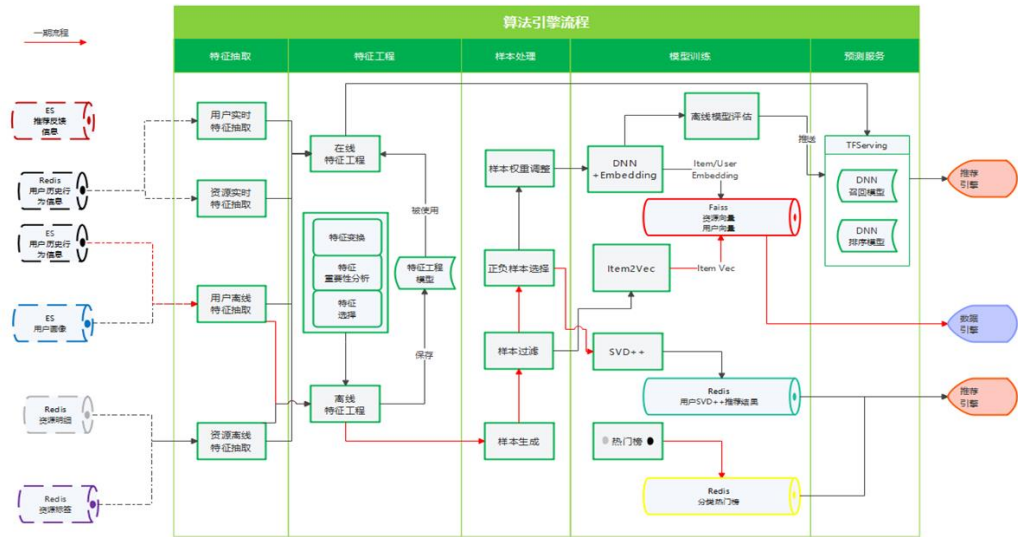
专业技能-专业知识/技能 / 工作成果

概要设计-系统架构



◆ 举证说明 (围绕任职资格认证标准, 结合自身能力及业绩产出进行论证)

专业技能-专业知识/技能 / 工作成果 概要设计-流程设计- (2) 算法引擎



◆ 举证说明 (围绕任职资格认证标准, 结合自身能力及业绩产出进行论证)

专业技能-专业知识/技能 / 工作成果

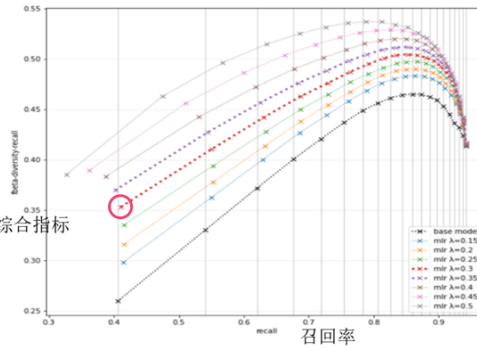
推荐模型中相关参数优化实验

λ 的选择

按 λ 细分, f_β-diversity-recall

$$F_{\beta}(div, recall) = \frac{F_{\beta}(div) \times recall}{(1 + \beta^2) (\beta^2 \times F_{\beta}(div) + recall)}$$

多样性和召回率综合指标



Kaggle 比赛结果截图

1, Jane Street 量化交易大赛

23:54



大小

kaggle.com



Overview Data Code Discussion **Leaderboard**

Public Leaderboard

Private Leaderboard

Preliminary standings up to 2021-03-15

Raw Data

Refresh

#	Team Name	Score ?
1	Robert Manolache	6242.530
2	float	6088.694
3	the best jane	5969.461
4	Vyacheslav Kotrachev	5873.745
5	一剑风吼	5865.595
Your Best Entry ↑ Your submission scored 4400.797, which is not an improvement of your best score. Keep trying!		
6	nice	5852.109
7	Cat Laundering Money...	5777.615
8	all luck no skill	5695.689
9	codefluence	5683.408
10	Martin BB	5661.822
11	*	5645.413



2, google 手语识别大赛

20:15

Overview Data Code Discussion Leaderboard

Rank	Participant	Status	Score
7	forest	🏆	0.787
8	bestfitting	🏆	0.787
9	Rob-Gilles-Rafi 🙌	🏆	0.787
10	keetar	🏆	0.783
11	Baohao & Shaomu	🏆	0.782
12	greySnow	🏆	0.779
13	pudae	🔄	0.777
14	Passiflora edulis	🔄	0.776
15	Selim Seferbekov	🔄	0.775
16	一剑风吼	🔄	0.774
17	🤔🤔🤔🤔.	🔄	0.774
18	Jobayer Hossain	🔄	0.772
19	Kelva Ferrat	🔄	0.771

大小 kaggle.com

3, 滴滴算法大赛结果报道

滴滴全球Di-Tech算法大赛落幕 中国选手夺得10万美元大奖 转载

2016-08-15 16:36:58



chvalrous 码龄12年

关注

7月20日消息，滴滴出行首届全球Di-Tech算法大赛今日正式落幕，经过数轮激烈角逐后，由三名成员组成的中国团队“inferrr”以绝对优势夺得冠军并收获10万美元奖金，“一剑风吼”和“blitz”两团队则分获二、三名。

滴滴出行CEO程维在算法大赛颁奖现场发表演讲，他表示：“互联网的下半场是人工智能。很庆幸，我们有世界上最难的题目，打造交通的AI。第一天创业的时候没想到我们在算法上这么难，我们要考虑的维度、复杂性、实时性超越了其他的行业。也很庆幸，最难的题目才需要最聪明的人工智能引擎，才有可能诞生未来最强大的大数据算法团队。我们希望建设全球顶级的技术团队，去面对这样的挑战。算法大赛我们会一直办下去，第一届十万美元奖金，将进一步扩大规模和影响力，吸引越来越多的人才参与。

APP内打开

算法 特征工程 GBDT

写评论

2 收藏 点赞 踩 分享

大小

blog.csdn.net



18:22



群聊设置



di-tech—一剑风吼队



群主很懒，还没有群介绍哦~

群聊成员

查看 6 名群成员 >



乐凡



刘洋



(*^_^*...



NotOnl...



邀请

群应用

管理 >



群管理



文件



相册



精华消息



收集表



签到



接龙统计



作业



解忧杂货铺



更多

群聊信息

群聊名称

di-tech—一剑风吼队 >

群号和二维码

560190026 >

群公告

百度云
向延超
董洋



我的本群昵称

未设置 >

群聊备注

未设置 >

获奖答辩 ppt 截图



团队成员介绍



专利链接

跨屏用户标识归一的方法及其系统 CN106528777A

<https://patents.google.com/patent/CN106528777A/zh>

<https://patents.google.com/patent/CN106202503B/en?q=cn106202503b>