

分论点排序和打分

零、目录

分论点排序和打分

一、任务背景

二、总体方案设计

三、数据集构造

四、模型介绍

4.1 排序模型

4.2 打分模型

4.3 积分梯度增加得分可解释性

多策略召回

一、总体方案

二、详细数据

一、任务背景

1、材料作文，总分为各段分数相加，分论点分数和总分强相关

共享、共建、共担 创新社区治理

“三事分流”工作法，是由部门负责解决“大事”，即管辖事项及公共服务；村（居）委会为主导，社区组织等共同协商解决“小事”，即村（居）公共事项及公益服务；由群众寻求市场服务或自行解决“私事”，即个人事务和市场服务。

开头

这是社区治理的创新，是政府治理和社会调节、居民自治良性互动的方式，是在社区这个“小世界”中，打造一个最优的时代，实现“大事”能快办，“小事”能共办，“私事”能自办，共享共建共担。

过渡段

社区的“大事”要快办，需落实责任，联动各方，提升效率。“民生无小事，枝叶总关情”，只要是涉及群众切身利益的公共事项及公共服务，都是“大事”，是最基本、最迫切的民生需求。“大事”的解决不是某个干部、某个部门的一家之责，需要协调各部门，加强联动，以此精简环节、提高效率。如北京实施的“街乡吹哨、部门报到”，实现群众需求与部门职能的精准对接。群众有所呼，政府有所应，切实提升群众生活的幸福感和获得感。

分论点

社区的“小事”要共办，需调动社区群众参与积极性，共同协商。过去，政府为群众办实事办好事，却大包大揽，基层干部疲于奔命，群众产生“等靠要”的依赖思想。发展为了人民，更要依靠人民，社区治理也要坚持群众路线，听民意、汇民智、集民力。如“人居环境整治义务督察员”及时发现、反馈问题，共建美好环境。既有政府的投入与权力赋予，也有群众的责任共担，才能求得自上而下精细管理与自下而上广泛参与的“最大公约数”。

社区的“私事”要自办，既要发挥群众的自主性，也要注重教育引导。我们常说“清官难断家务事”，“外界力量”有时并不能弄清“私事”的全貌。与此同时，“私事”也涉及到隐私，若处理不当，把握不好“度”，反而可能弄巧成拙，激化矛盾。因此，对于家庭私事，倡导“自办”。但这并不意味着政府、社区的袖手旁观、隔岸观火，需要为其提供必要的心理、家庭、教育、法律等普惠性服务，注重教育和引导。

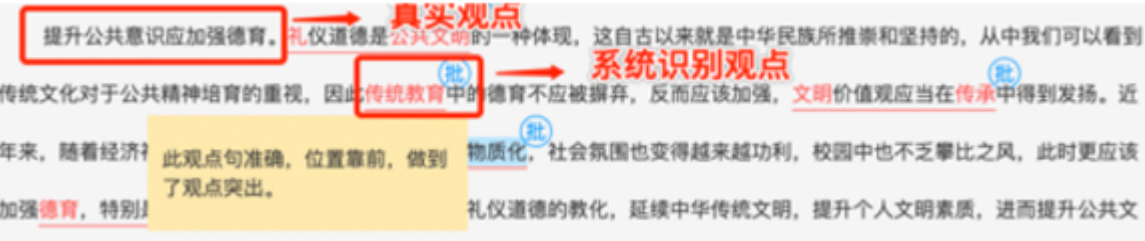
3个论述段

总之，要让“三事分流”工作法真正落地、见实效，需要系统性、全面性、全方位的过程把控。要深入群众，准确把握群众所思所想所盼；要精准识别，合理分类，明确权责；要全面反馈评估，信息公开，接受监督，真正做到准确“找事”、合理“分事”、高效“办事”、全面“反馈评估”。

社区治理没有终点，只有连续不断的新起点，“治大国如烹小鲜”，要继续提升社区治理水平，真正惠民生、暖民心，让社区成为“人人出力、人人尽责”、共享共建共担的大家庭。

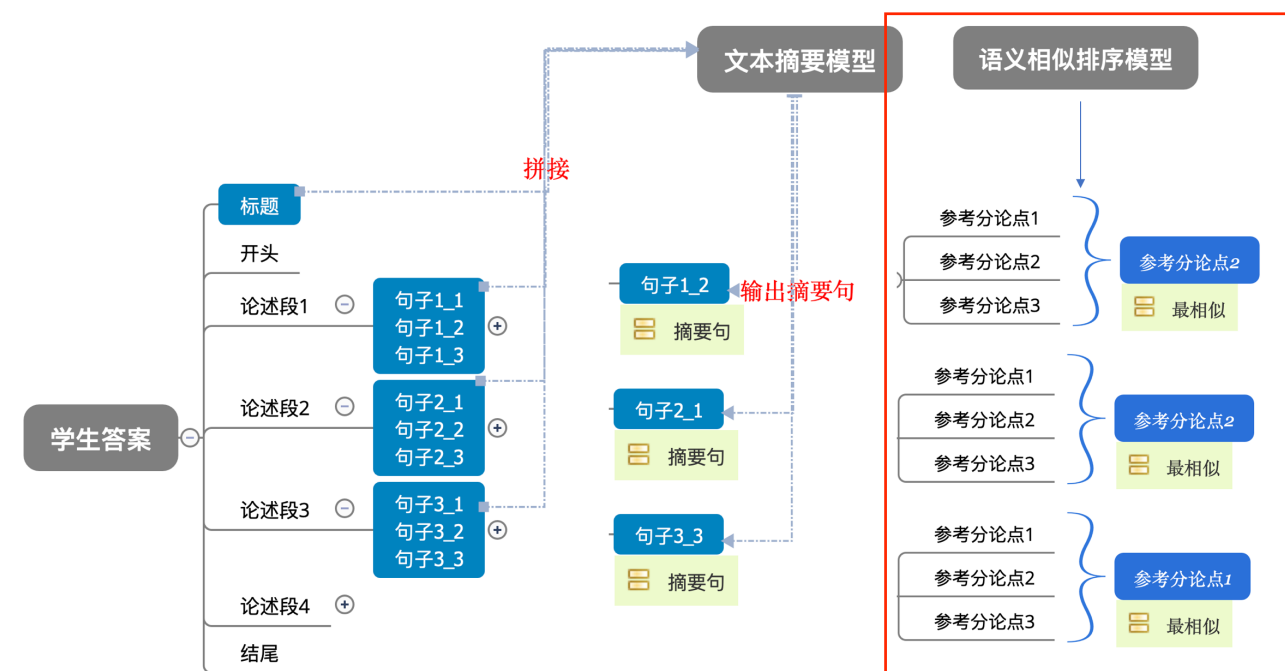
结尾

文章结构图



2 分论点示意图

二、总体方案设计



三、数据集构造

3.1 标注工具制作

3.1.1 <参考分论点123, 学生分论点>，选择最佳匹配参考分论点

2134524	科学具有批判性。它可以被证伪。可以推翻旧的经验。科学的质疑精神能够激发想象力。 艺术具有超越性。它代表美的力量。在批判生活、揭露现实的同时，又追求更美好的生活，艺术的超越精神推动想象力走向远方。 古文化具有神秘性。它包含了未知世界的魅力。需要挖掘与探索。古文化的未知性为想象力带来灵感的碰撞。	分论点 1 分论点 2 分论点 3
	难度解释	完成数/总数 (1482)
	easy:[1.5,2] ; medium:[0.75,1.5];hard:[0.25,0.75]	0
id	学生答案分论点	最佳匹配
2513719_Easy_2273_1_2_1	政府对灾难的治理水平。和人民对灾难的预防工作及把抗灾作为个体的义务和责任能极大的降低灾难的危害。	请选择
2513719_Easy_2286_6_2_1	明确个人的义务与责任，形成政府，企业和社会多方联动，协调应对的观念意识是减少灾难损失，建筑安全，长城的基石。	请选择

标注工具排序图

3.1.2 <参考分论点, 学生分论点>,勾选各自匹配部分

1、参考分论点和学生答案都采用子句选择的方式

2、匹配程度设定：【语义匹配，强/中/弱】【主题匹配 强/弱】，这个分档相比以往的打分制，让老师划档更清晰避免歧义。

18			x			
19	2513719_Hard_22580_1_2	语义匹配	参考答案分论点	x	2 奉献爱心	3 团结互助
20				4 共同抗灾	5 从而展现人性光辉。	
21			学生答案分论点	x	2 每一位受灾群众都不孤单	3 因为我们在一起
22				4 心心相连	5 手手相扣	5 并肩而立。

标注工具打分图

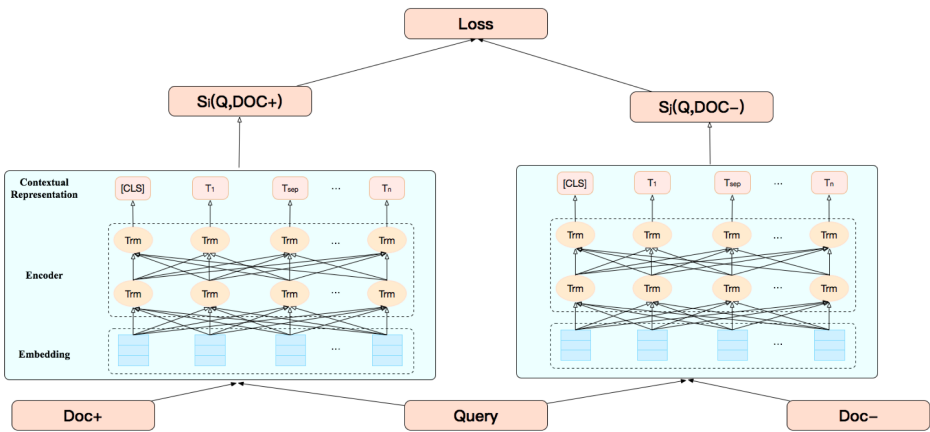
3.2 创新点以及效果

- 1、这种方式虽然标注了**2w**条数据，挖掘正样本、负样本、困难样本可以产生30w-100w的标注数据
- 2、一份标注数据同时用于句向量生成、排序任务、打分模型，一份多用
- 3、python生成excel, 4星期人力--->2星期人力

四、模型介绍

4.1 排序模型

4.1.1 pairwise



$$\text{Loss} = \text{sim}(\text{学生参考答案}, \text{top1参考分论点}) - \text{sim}(\text{学生参考答案}, \text{top[2345]参考分论点})$$

pairwise训练方式

- 1、任务建模：<参考分论点123,学生分论点1>分别拼接，输入模型，获得排序值，最大的值就是最佳匹配的句子
- 2、数据集构建：<anchor,positive,negative>：<学生分论点左右扩展,参考分论点左右扩展，参考分论点同一qid的另外2个分论点>，数据集大小有30w和100w数据集
- 3、模型以及loss：3层bert 模型，loss为triple loss
- 4、模型结果,acc 0.92

4.1.2 simcse句向量模型

- 1、任务建模：计算<参考分论点123句向量,学生分论点句向量>的consine相似度，相似度最高的参考分论点为结果
- 2、数据集构建：构建三元组，<anchor,positvie,negative>,anchor为参考分论点， positive为强匹配的句子， negative为弱匹配的句子，数据集大小为8w
- 3、模型以及loss:模型为bert， loss为交叉熵损失函数， 参考simcse论文
- 4、模型结果， acc 0.82

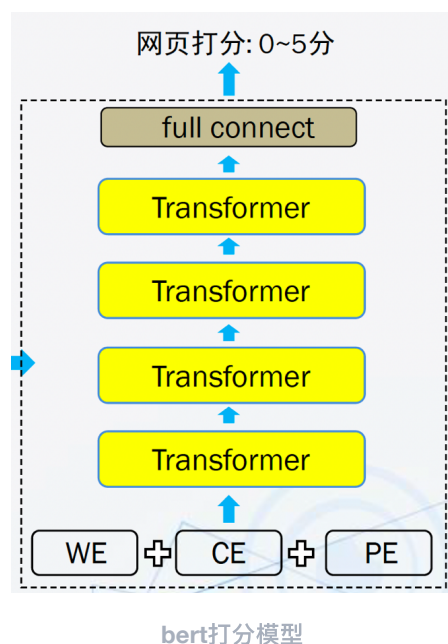
4.1.3 综合结果

结果如下： 3层bert模型 acc 0.92, 耗时1.5ms。效果上远远大于目前公开句向量模型结果（0.76）

	acc
simbert V2	0.76
simCSE	0.82
pairwise	0.92

4.2 打分模型

- 1、任务建模：（输入2个句子的拼接，输出0-1的得分）
- 2、数据集构建:refer（标注的人工参考答案句子各种组合），stu_answer（学生答案各种组合）,score（标注的人工评级），从1.9w条标注数据，构建出30w-60w的训练样本
- 3、模型以及loss:6层bert模型，loss为均方误差
- 4、模型结果：acc 0.9



bert模型	pearson	acc
12层	0.86	0.93
6层	0.83	0.91
3层	0.53	0.77

4.3 积分梯度增加得分可解释性

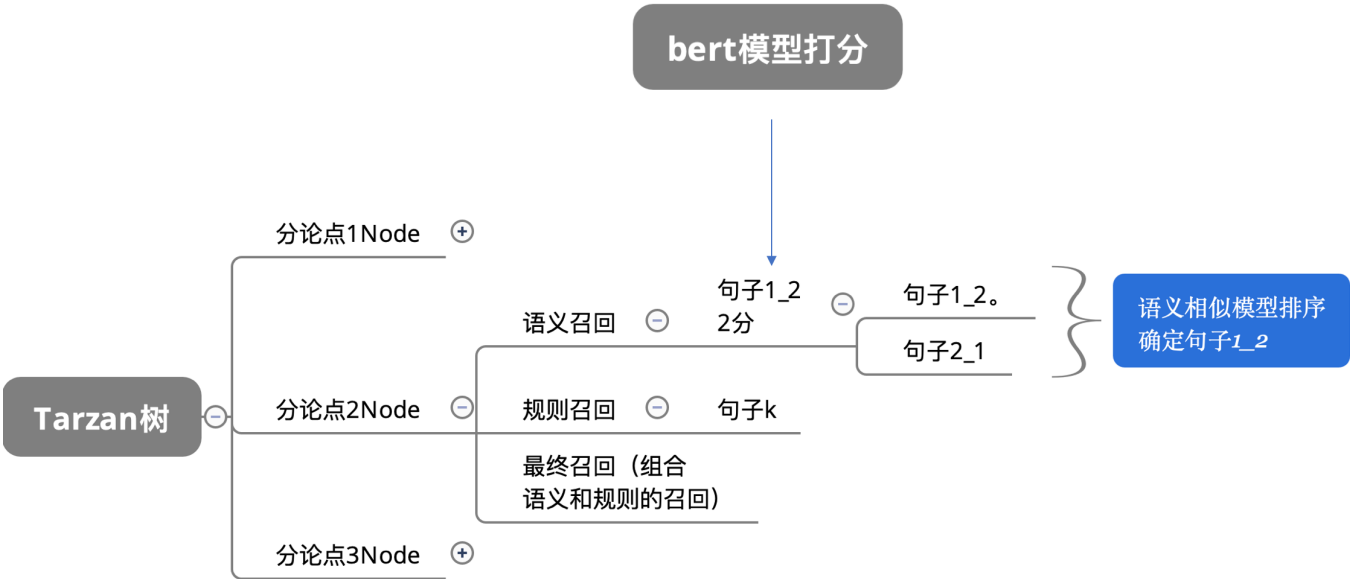
Integrated Gradients: 找出输入部分对得分影响比较大的词

结果：由于速度太慢，最后用的是其他规则

在全方位收集数据过程中[MASK]必然会面临信息泄露等问题[MASK]信用体系建设[MASK]这不仅考验企业技术能力[MASK]也考验企业责任感和担当感[MASK]更需要相关立法和监管紧密跟上[MASK]个人征信牵涉到个人隐私[MASK]引入民营征信机构是否会泄露个人信息[MASK]是民众最大的担忧[MASK]我们要提防信息收集和记录变为一场隐私泄露的闹剧[MASK]就必须补上法律漏洞[MASK]严格审核民营机构的资质[MASK]制定严格准入标准[MASK]提高机构质量的同时也更好保障信用信息的准确性[MASK]避免出现造假作弊[MASK]信息良莠不齐[MASK]真伪难辨[MASK]为民众信息安

多策略召回

一、总体方案



二、详细数据

2.1、匹配评估

参考分论点1:

- 未召回任何句子，标记为0
- 召回标注句子，标记为1
- 召回其他句子，标记为2

评估类别，3个类别，1,2为正类别

数据集	策略	P	R	F1
8000篇数据集	线上规则	0.37	0.40	0.38

	4月线上融合	0.56	0.69	0.61
	7月线上融合	0.58	0.70	0.63
	7月线上融合+ 位置特征	0.62	0.68	0.65
	8月调语义匹配阈值			0.66
190个段落 第2批	线上规则	0.56	0.56	0.56
	线上融合	0.61	0.68	0.64
	线上融合+ 位置特征	一样	一样	一样

2.2 得分评估

2.2.1 参考分论点1得分，与标注的得分

		MSE	RMSE	MAE	R2	pearson 系数	spearson系数
online 线上规则		1.20	1.09	0.81	-1.76	0.32	0.33
merge 线上规则+语义	4月上线	0.95	0.97	0.72	-1.20	0.42	0.43
	7月线上	0.94	0.97	0.71	-1.16	0.43	0.46
	7月线上+特征	0.88	0.94	0.68	-1.02	0.42	0.44
	8月调语义匹配阈值	0.87	0.93	0.67	-1.00	0.43	0.45

2.2.2 3个参考分论点的和，与标注得分

--	--	--	--	--	--	--	--

		MSE	RMSE	MAE	R2	pearson系数	spearson系数
online_sum 3个分论点求和 线上规则		6.93	2.63	2.08	-2.51	0.40	0.40
merge_sum 3个分论点求和 线上规则+语义	4月上线	5.64	2.37	1.9	-1.86	0.50	0.50
	7月线上	5.4	2.32	1.85	-1.73	0.51	0.51
	7月上线+特征	4.15	2.04	1.60	-1.1	0.52	0.52
	8月调语义匹配阈值	3.99	2.00	1.56	-1.02	0.53	0.53

2.3、勾画评估

参考分论点命中句子，与标注句子的勾画是否为同一句，不同于匹配，勾画不区分被哪个参考分论点匹配到

		P	R	F1
线上规则		0.49	0.27	0.35
语义匹配		0.84	0.54	0.66
融合	4月上线	0.74	0.52	0.61
	7月线上	0.75	0.53	0.62
	7月线上+特征	0.76	0.52	0.62
	8月调语义匹配阈值			0.62