

本项目得到国家科技创新2030“新一代人工智能”重大项目——“以自然语言为核心的语义理解理论、模型与方法”（项目号：2020AAA0106701）支持。

SpaCE2022 标注

—— task2 异常归因

詹卫东

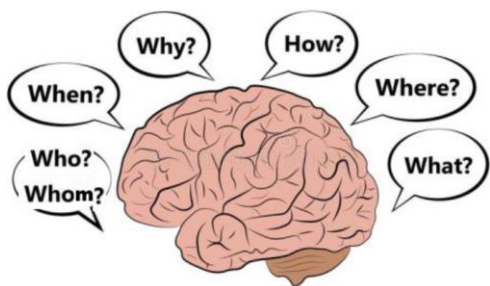
北京大学中文系

zwd@pku.edu.cn

中文空间语义理解评测任务 (**S**patial **C**ognition **E**valuation, 简称SpaCE)

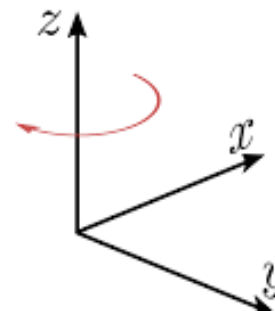


中文空间语义理解评测任务 (Spatial Cognition Evaluation, 简称SpaCE)



人类

空间
认知



计算机



- ✘ 1. 夫妻俩商量了几天，买了一辆农用三轮车。农用三轮车的油门在左把手上，张顺东请人改到左边，他就能用左手操作了。
- ✘ 2. 赵诗人溜回家中吃了午饭，趁着两个女孩在外面逛街，赶紧在沙发旁躺下来睡觉。赵诗人一觉睡到黄昏时刻，两个女孩回来了，见到赵诗人穿着短裤躺在沙发里，几声惊叫才把赵诗人吓醒，他赶紧跳起来，把自己扫地出门。
- ✘ 3. 宋钢看着林红跨外自行车，叮嘱她骑车要小心，林红点点头往西骑车而去。
- ✔ 4. 客人小心翼翼递过去一张字条。贝多芬戴上眼镜专注地凝视了一会儿：“好，你们竟敢到兽穴里来抓老狮子的毛。”

- ✓ ① 夫妻俩商量了几夭，买了一辆农用三轮车。农用三轮车的油门在右把手上，张顺东请人改到左边，他就能用左手操作了。
- ✓ ② 赵诗人溜回家中吃了午饭，趁着两个女孩在外面逛街，赶紧在沙发上躺下来睡觉。赵诗人一觉睡到黄昏时刻，两个女孩回来了，见到赵诗人穿着短裤躺在沙发里，几声惊叫才把赵诗人吓醒，他赶紧跳起来，把自己扫地出门。
- ✓ ③ 宋钢看着林红跨上自行车，叮嘱她骑车要小心，林红点点头往西骑车而去。
- ✗ ④ 客人小心翼翼递过去一张字条。贝多芬戴上眼镜专注地凝视了一会儿：“好，你们竟敢到兽穴外来抓老狮子的毛。”


- ❌ 1. 夫妻俩商量了几天，买了一辆农用三轮车。农用三轮车的油门在左把手上，张顺东请人改到左边，他就能用左手操作了。

问：这段话为什么错？

答：农用三轮车的油门在左把手上，如果要改位置，按常理，就只能改到右把手。但文中说“张顺东请人改到左边”，这显然跟常理冲突。

问题：怎么把上面这个回答讲给计算机听？

用计算机“更好懂”的方式来说明这个错句错在哪儿。



规范化
表达

方法：把空间语义拆解为“空间实体”“空间方位”“空间行为”三部分

Spatial_entity

Place

Event

S

P

E

❌ 1. 夫妻俩商量了几天，买了一辆农用三轮车。农用三轮车的油门在左把手上，张顺东请人改到左边，他就能用左手操作了。

问：这段话为什么错？

答：农用三轮车的油门在左把手上，如果要改位置，按常理，就只能改到右把手。但文中说“张顺东请人改到左边”，这显然跟常理冲突。



序号	异常片段1			异常片段2			异常类型
	S1	P1	E1	S2	P2	E2	
1	油门	左把手上	在	油门	到左边	改	语义冲突
2	油门	在左把手上		油门		改到左边	语义冲突
3	油门	在左把手上		油门	到左边	改	语义冲突

❌ ④ 客人小心翼翼递过去一张字条。贝多芬戴上眼镜专注地凝视了一会儿：“好，你们竟敢到兽穴外来抓老狮子的毛。”

问：这段话为什么错？

答：到兽穴里抓老狮子的毛，才需要说“竟敢”。说“竟敢到兽穴外抓老狮子的毛”，不合常理。

序号	异常片段1			异常片段2			异常类型
	S1	P1	E1	S2	P2	E2	
✓ 1	你们	到兽穴外来	竟敢抓老狮子的毛				不符合常识
❌ 2	你们	到兽穴外	抓老狮子的毛				不符合常识
❌ 3	老狮子的毛	到兽穴外	竟敢抓				不符合常识

✘ 3. 宋钢看着林红跨外自行车，叮嘱她骑车要小心，林红点点头往西骑车而去。

问：这段话为什么错？

答：“跨外自行车”很别扭，没有这种说法，不知道这是在说什么。

序号	异常片段1			异常片段2			异常类型
		text1			text2		
✓ 1		跨			外		搭配不当
✘ 2		跨外			自行车		搭配不当
✘ 3		林红			跨外自行车		搭配不当

核心问题	答案
为何标注	(1) 描述句子的空间信息有什么问题 (2) 为计算机空间语义理解提供负例数据
如何标注	(1) 通读 ：一句一句通读全段 (2) 找错 ：找到有 信息异常点 的小句（标点句） (3) 分类 ：对异常情况进行分类 (4) 转写 ：把异常点用S、P、E要素方式呈现出来 (5) 验证 ：看看（4）是否符合（2）的原意
如何判断	(1) 基于 个人自然语感 的判断（不自然、费解，即异常） (2) 基于 日常生活常识 的推理（合乎常情，无关严密逻辑）
如何优化	(1) 形式整齐 优于 形式杂乱 (2) 多原因 优于 单原因 (3) “归入正常” 优于 “为异常 强找理由 ”
怎分高下	(1) 空间信息描述清晰准确是最高标准 (2) 不一定能做到最好，但千万别做 最差 ：避免严重的信息描述错误（跨句远距离选错要素成分，填入“的”“的话”等助词性成分，弄错S、P、E要素的填写位置，等等，均为严重错误）

分类	含义	示例
搭配不当	句中有 两个表示空间信息 的语言成分 不能组合	把头放在 地面边
语义冲突	句中有 两处空间信息 存在 矛盾、冲突	孩子们睡在 蒙古包里 ，他们一醒来，就跳起来 跑进蒙古包 找妈妈
不符合常识或背景信息	句中有一处文本片段的空间语义违反常识或者违反句子的背景信息	南川河、北川河、西川河从北、北、西三方汇入
并无异常	一段话空间语义正常	圆明园在北京西北郊，是一座举世闻名的皇家园林。
不适合用于计算机空间语义理解	一段话有 错字 、句子整体 无空间方位义 、句中存在 敏感信息	由谁去充任那些苦难的角色？又有谁去体现这世间的幸福，骄傲和快乐？

空间语义三要素	S（空间实体）	P（空间方位）	E（空间事件）
含义	一个实体是空间实体，当且仅当句中描述了该实体的空间方位信息	空间实体方位是句中描述空间实体S的空间方位信息的语言成分	空间义相关事件是表达空间实体S位于P的方式、目的等的核心谓词成分
示例1	奖状	在墙上	贴
示例2	戒指里面刻着他的名字 [_P 戒指里面][_E 刻] 着 [_S 他的名字]		
示例3	她从不坐贵宾席 [_S 她] 从不 [_E 坐] [_P 贵宾席]		
示例4	大客车未遵守信号灯指示就右转弯由西向东行驶 [_S 大客车] 未遵守信号灯指示就 [_P 右转弯由西向东] [_E 行驶]		

句子	异常类型	异常片段					
她胳膊上少了那只篮子，她把它塞到那个女学生座位四面了	搭配不当	text1			text2		
		塞到			四面		
孩子们睡在蒙古包里，他们一醒来，就跳起来跑进蒙古包找妈妈	语义冲突	S1	P1	E1	S2	P2	E2
		孩子们	在蒙古包里	睡	孩子们	进蒙古包	跑
打开西宁的卫星遥感地图，南川河、北川河、西川河从北、北、西三方汇入，在城区形成湟水河，自西向东穿城而过。	不符合常识或背景信息	S			P	E	
		南川河、北川河、西川河			从北、北、西	汇入	

空间语义要素标注原则	要点	具体要求
原则1: 要素完整原则	勿遗漏信息	(1) S、P、E尽量齐备 (2) S里填写的成分清楚完整；P里填写的成分清楚完整；E里填写的成分清楚完整
原则2: 要素邻近原则	勿舍近求远	(1) S、P、E尽量在一个标点句范围内选取，非必要，不跨出当前标点句。 (2) 如果跨标点句选S、P、E成分，一定是实义成分，而非功能成分（比如：介词不需要也不能跨标点句选取）；一定是离当前标点句最近的成分。
原则3: 语义简明原则	勿冗词赘语	(1) S、P、E能清楚表达空间信息即可，不需要填写修饰性成分。 (2) E中动词词组如果包含表达空间信息的修饰性成分，应该填入。
原则4: 平行比对原则	勿形义错配	(1) 两组<S, P, E> 尽量保持无空间信息冲突的成分形式一样，有空间信息冲突的成分形式有别 (2) S1 – S2, P1 – P2, E1 – E2 并行对照
原则间关系	(1) 如果有多个候选标注方案，那么，符合原则最多的方案最优 (2) 原则的优先级顺序是：原则1 > 原则4 > 原则3 > 原则2	

①	句子	异常类型	异常片段			
她胳膊上少了那只篮子，她把它塞到那个女学生座位四面了	搭配不当	text1		text2		
		塞到		四面		
	不符合常识或背景信息	S	P		E	
		它	到那个女学生座位四面		塞	

②	句子	异常类型	异常片段			
核桃被他从左扔到左，从右扔到左	搭配不当	text1		text2		
		从左		扔到左		✓
	不符合常识或背景信息	S	P		E	✓
		核桃	从左到左		扔	

③	句子	异常类型	异常片段		
许多村民是坐在草垛上、在村口的路灯上	搭配不当	text1		text2	✖
		坐		在村口的路灯上	
	不符合常识或背景信息	S	P	E	✔
		村民	在村口的路灯上		

附录：错误标注举例

entry#49060



这时她已经洗完衣服了，本来要站起来了，听了小关剪刀的一番话，她又蹲在了那里，她出神地看了一会儿河面，才端起脸盆里_里的衣服起身走上了台阶。小关剪刀一直跟随她走到家门口，看着她把衣服晾在_在绳子_外，小关剪刀又说了一遍：

标注 #96682

审批



詹卫东审批不通过「S：衣服，P：在绳子外，E：晾 不符合常识。“绳子”跟“外”是可以搭配的」标注者尚未处理

0 text1 绳子 text2 外 搭配不当

- (1) 语料 …… 场子边缘拉起绳子，所有的观众都在 绳子 外。他们多半是中国人，也有不少白人记者……
- (2) 选项说明力度：“搭配不当”并不能完全说明这个句子的异常

附录：错误标注举例

entry#48938



可是有一天，他们俩为了一点小事吵了一架。山顶上再也看不到飘荡的风筝，小溪旁再也看不到漂流的纸船了。

标注 #97764

审批



詹卫东审批不通过「应该是：S: 纸船，P: 小溪旁，E: 漂流 不符合常识； S: 漂流的纸船，P: 小溪旁，E: 看不到 不符合常识」
标注者尚未处理

0 S 纸船 P 小溪旁 E 看不到 与常识或背景信息不符

- 纸船

小溪旁

漂流

✓
- 漂流的纸船

小溪旁

看不到

✓

附录：错误标注举例

entry#50210



饕林少城老味道餐馆老板何旭戴着口罩，腰束围裙，手拿菜单，站在门口招徕过往客人。见有人停步，他赶紧上前招呼：“坐屋里还是坐左面？”

标注 #98693

审批



暂无批示

0 text1 坐 text2 左面 搭配不当



(1) “坐 左面” 确实别扭，少见

“坐屋里” “坐左面” 搭配不当

(2) 用“坐 左面” - “搭配不当”，是不是能完全说明这个句子的异常呢？

“坐屋里” “坐左面” 语义冲突

S: 客人, P: 屋里还是左面, E: 坐 不符合常识

附录：错误标注举例

entry#48670



他感到非常难为情。他把头藏到翅膀外面，不知道怎么办才好。他感到太幸福了，但他一点也不骄傲，因为一颗好的心是永远不会骄傲的。他想起他曾经怎样被人迫害和讥笑过，而现在他却听到大家说他是美丽的鸟中最美丽的一只。紫丁香在他面前把枝条垂到水里去。太阳照得很温暖，很愉快。

标注 #97254

审批



暂无批示

这段话不存在异常

(1) “他把头藏到翅膀外面” 有异常

(2) “把头” “藏到翅膀外面” 搭配不当

S: 头, P: 到翅膀外面, E: 藏 不符合常识

附录：正确标注举例

entry#48881



飞船攀升俯冲，左拐右转，四周的惊险、恐怖情景不断变幻，坐在座位下的观众如身临其境，无不吓得大呼小叫。

标注 #97255

审批



詹卫东审批通过

0 S 观众 P 在座位下 E 坐 与常识或背景信息不符

观众能不能坐在座位下？观众会不会坐在座位下？

并非绝对没有可能……但是，这种情况常见吗？合情合理吗？不合情理。

附录：正确标注举例

entry#47975



稻草堆里，显微镜下，细菌过滤器和生物培养箱后，迎来晨辉初绽，送走晚霞余红，提取出一种种稻草病菌，培养出一批批菌种新苗。

标注 #97108

审批



暂无批示

0 S 病菌 P 细菌过滤器后 E 提取出 与常识或背景信息不符

1 S 新苗 P 生物培养箱后 E 培养出 与常识或背景信息不符

Q & A

zwd@pku.edu.cn

<https://2030nlp.github.io/Sp22AnnoOL/menu>