

实验二：汽车说明书智能问答

20123101 李昀哲

一、实验题目

基于跨模态文档理解技术，完成自动对用户提出的相应问题给出手册中解答的位置并高亮便于查看。

二、实验内容

本任务主要基于跨模态文档理解技术，深度解析非结构化文档中排版复杂的内容。同时结合此前“简历信息提取”中的 OCR 方法，对说明书进行解析。

这种复杂的排版不仅出现于汽车说明书中，更广泛存在于科研论文、文献中。这种跨模态的理解技术，可以极大减轻科研人员对文档处理时的难度，尤其是在 NLP 领域。

解决痛点：

- 用户：用户没有耐心查阅说明书，打客服电话需要等待；
- 售后客服：需要配置大量客服人员，且客服专业知识培训周期长；
- 构建问题库：需要投入大量人力整理常见问题库，并且固定的问题库难以覆盖灵活多变的提问。

对于用户来说，汽车说明书问答系统能够支持通过车机助手/APP/小程序为用户提供即问即答的功能。对于常见问题，用户不再需要查阅说明书，也无需打客服电话，从而缓解了人工客服的压力。

对于客服来讲，汽车说明书问答系统帮助客服人员快速定位答案，高效查阅文档，提高客服的专业水平，同时也能够缩短客服的培训周期。

1. OCR (Optical Character Recognition) 处理模块

文字识别 OCR 提供在线文字识别、图片文字提取服务，将图片、扫描件或 PDF、OFD 文档中的文字识别成可编辑的文本。OCR 文字识别支持证件识别、票据识别、定制模板识别、通用表格文字识别等。文字识别可提取图片中的文字、识别图片转文字。

2. RocketQA

该方法针对模型训练中存在的问题，通过跨批次负采样（cross-batch negatives）、去噪的强负例采样（denoised hard negative sampling）与数据增强（data augmentation）等技术，大幅提升了 对偶式检索模型的效果。

“端到端问答”摒弃了传统系统中繁杂的构件，系统复杂性大大降低，并且其中每个模块（段落检索和答案定位）都是可学习的，这样的设计能够让整个系统实现端到端训练。这意味着问答系统可以基于用户实时的反馈实现在线训练，而不是只在封闭的数据集上闭门造车。这是智能问答技术的发展趋势，可能会引发问答系统的新一代技术变革。RocketQA 在在谷歌 Natural Questions 数据集上段落检索的效果对比如表 1 所示。

表 1 各方法在谷歌 Natural Questions 数据集上段落检索结果对比

| Model | EM |
|---------------------|-------------|
| BM25+BERT | 26.5 |
| HardEM | 28.1 |
| GraphRetriever | 34.5 |
| PathRetriever | 32.6 |
| ORQA | 33.3 |
| REALM | 40.4 |
| DPR | 41.5 |
| GAR | 41.6 |
| RocketQA+DPR reader | 42.0 |

3. 跨模态阅读理解模块（LayoutXLM 模型）

使用 BERT 作为 backbone，加入 2-D 绝对位置信息，图像信息，分别捕获 token 在文档中的相对位置以及字体、文字方向、颜色等视觉信息。

LayoutXLM 专门用于处理多语言文档理解的跨模态预训练模型，如图 1 所示，其在建模过程中同时利用了视觉信息和文本信息，通过中两类模态信息增益文档理解，提高模型的文档理解能力。

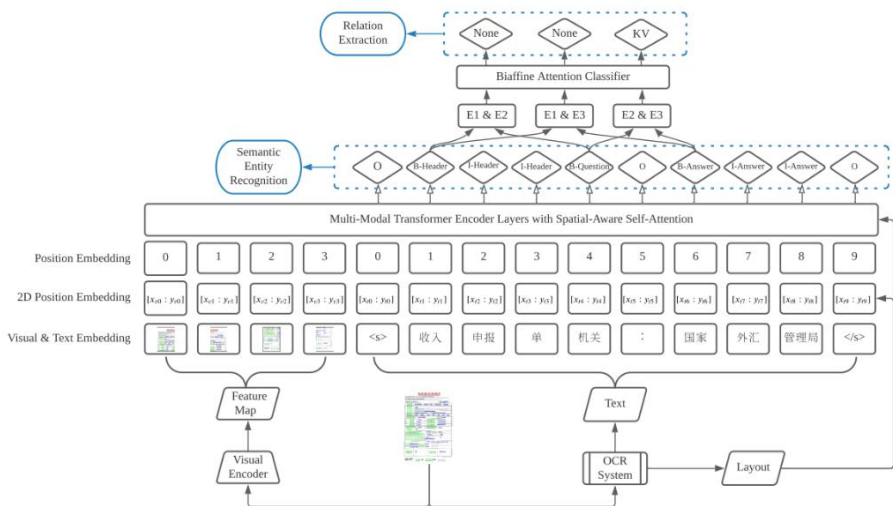


图 1 排序模型

三、实验流程

完整的流程如图 2 所示，需要包含“OCR 处理模块”、“排序模块”、“跨模态阅读理解模块”，分别处理识别和检测不同的功能。

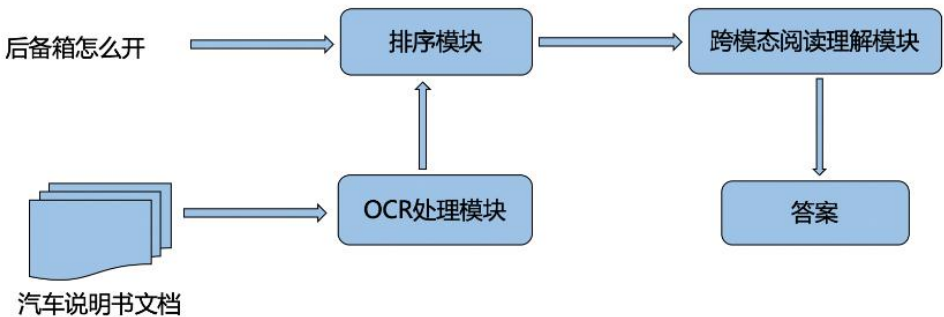


图 2 任务完整流程示意图

1. 环境准备

安装 PaddleNLP 和 PaddleOCR 模块

2. 数据位置

./OCR_process/demo_pics 保存了 10 张说明书的图片；

./OCR_process/demo_ocr_res.json 保存了 OCR 对说明书图片的解析结果

./Rerank/data/train.tsv 保存了 140 条汽车说明书相关的训练样本

3. OCR 处理模块

通过此前“简历信息提取”中的 OCR 方法，对汽车说明书进行识别，记录汽车说明书上的文字和文字布局信息，以方便后续使用计算机视觉和自然语言处理方面的技术进行问答任务。处理得到的结果如下所示。

```
“document”:[..., “车辆”, “解”, “锁”, ...]  
“document_bbox”:[..., [284, 463, 289, 382], [463, 552, 289, 382],  
[552, 642, 289, 382], [279, 341, 447, 526], ...]
```

- "document":OCR 模块识别出的文字（token）；

- "document_bbox":每个 token 在汽车说明书上对应的坐标，分别对应(x_min, x_max, y_min, y_max)。

4. 排序模块

基于 RocketQA 预先训练好的基线模型 base_model。根据用户提出的问题对汽车说明书的不同图片进行打分排序，这样便可以获取和问题最相关的图片，并使用跨模态阅读理解模块在该问题上进行抽取答案。

训练:

数据集中，每条样本包含 3 列数据，分别是：用户问题，文档，标签。其中标签为 1 表示用户问题的答案出现在了该文档中，标签为 0 表示用户问题和该文档无关。

钥匙使用距离 解锁车辆解锁车外解锁智能钥匙解锁进入车辆之前,您需要使用智能钥匙解锁车辆。当智能钥匙电量显示为蓝色或绿色(电量高于20%)时,智能钥匙的有效使用范围为车外10米以内。智能钥匙的按钮及功能如下:1.车辆解锁按钮车辆为p档时,短按此按钮可解锁车辆,解锁后车辆两侧转向灯将闪烁三次以提示解锁成功,解锁成功后车门外把手自动弹出。通过智能钥匙解锁四个车门后,无需钥匙即可从外部打开尾门。先短按解锁按钮,在3秒内进行第二次长按(保持1.5秒)可同时打开全部车窗、天窗及遮阳帘;车窗、天窗及遮阳帘运动过程中放开解锁按钮则车窗、天窗及遮阳帘停留在当前位置。您可以通过中控屏设置智能钥匙的解锁模式;在中控屏下方控制栏最左侧进入设置页面,点击门锁>钥匙解锁模式,选择“全车解锁”,则只需按一次钥匙解锁按钮即可同时解锁全车车门;选择“主驾车门解锁”,则第一次按下钥匙解锁按钮仅解锁驾驶员侧车门,再次按下解锁按钮可解锁其余三个车门。2.车辆上锁按钮车辆挂p档,关闭所有车门(包括前盖及尾门)后,短按此按钮可实现全车上锁,全车上锁后两侧转向灯将闪烁一次并且喇叭鸣响一次以提示上锁成功,上锁成功后车门外把手收回,外后视镜自动折叠(在中控屏下方控制栏最左侧进入设置页面,点击驾驶>后视镜>锁车自动折叠,设置外后视镜自动折叠)。上锁后尾门仅可通过智能钥匙从外部打开。您可以通过中控屏设置喇叭提示音的打开/关闭,在中控屏下方控制栏最左侧进入>设置页面,点击声音>提示音>锁车提示音。1
钥匙使用距离 解锁人详情以及被授权的功能(如媒体、视频、安全箱等)。若当前授权已生效,还可更改授权功能或停止授权。被授权用户可在授权的功能范围内进行操作,不可进行授权管理。设置访客模式。车辆上锁后,您可以使用智能钥匙解锁并进入车辆,自动重新登录为车主账号。说明出于安全考虑,当被授权人正在行驶中,停止授权需等待被授权人停车并锁车后解除授权才可正式生效。说明若添加的被授权用户为蔚来账号用户时,则添加授权过程完成后授权即时生效;若添加的被授权用户为非蔚来账号用户时,需等待此用户注册好蔚来账号后授权方可生效。被授权用户解锁>被授权用户可通过nfc钥匙或远程手机app方式解锁进入车辆,点击中控屏个人头像即可显示被授权用户的账户信息与授权功能;nfc解锁:开启本机nfc钥匙后将手机贴近左侧车身中部区域。访客模式若您需通过借智能钥匙的方式借车给别人时,您可以在中控屏点击个人头像设置访客模式,保护自己的隐私(如导航历史、电话本、视频、照片等),访客模式下仅可使用空调、天气、导航(无历史、收藏等个人信息)等默认车辆功能。当您设置访客模式后,用车人使用智能钥匙解锁进入车辆后将呈现访客模式下的功能;若要退出访客模式,需要输入车辆手密码解除访客模式设置。注意1.仅在非驾驶状态可设置访客模式。主账号,并以车主权限操作车辆。服务授权您或授权用户可通过手机app向蔚来汽车服务中心发起请求,服务中心通过后台管理服务人员并授权其临时使用车辆进行对应的服务(如一键加电服务),服务完成后蔚来汽车服务中心将收回对应的服务权限。12 0

图 3 排序模型的训练样本

基于 `base_model` 在汽车说明书的排序数据集上进行微调训练。

```
!bash run_train.sh ./data/train.tsv ./checkpoints/base_model 50 1
```

- 参数依次为训练数据地址，`base_model` 地址，训练轮次，节点数。
- 在训练过程中，会将模型存放至目录 `./Rerank/output` 下。训练结束后选择合适的模型移动至 `./checkpoints/` 目录下，并正式重命名为 `ranker`。

模型测试：

用户问题和 PaddleOCR 识别结果的文本进行拼接，生成测试集。然后默认会加载 `./Rerank/checkpoints/ranker` 模型，并基于该模型进行测试。最终测试结果将保存至 `./data/demo.score` 文件中。

5. 跨模态阅读理解模块

用户问题和 PaddleOCR 识别结果的文本进行拼接，生成测试集。然后默认会加载 `./Rerank/checkpoints/ranker` 模型，并基于该模型进行测试。最终测试结果将保存至 `./data/demo.score` 文件中。

训练：

提供了 28 条汽车说明书相关的训练样本，用于跨模态阅读理解模型的训练，同时也提供了一个预先训练好的基线模型 `base_model`。本模块可以使用 `base_model` 在汽车说明书训练样本上进一步微调，增强模型对汽车说明书领域的理解。

具体来讲，设定了 4 类标签：`O`, `B-ans`, `I-ans` 和 `E-ans`，用于从给定文档中抽取答案。即首先会将用户问题和文档进行拼接，然后传入 `LayoutXLM+CRF` 模型进行阅读理解任务，模型会对输入文本序列预测相应的标签序列，然后便可以根据标签序列从给定文档中进行抽取答案了。

数据说明：

跨模态阅读理解模型的训练集文件为 `./Extraction/data/train.json`，其中重要的信息包括：`"question"`, `"document"`, `"document_bbox"`和`"labels"`。

- `"question"`: 用户提问的问题
- `"document"`: OCR 模块识别出的文字（token）

- "document_bbox":每个 token 在汽车说明书上对应的坐标，分别对应(x_min, x_max, y_min, y_max)
- "labels": 阅读理解任务的标签数据

模型测试：

在模型训练完成后，便可以开启模型测试。

具体来讲，本节会分析排序模块输出的打分结果，然后获取与用户问题最相关的汽车说明书，并将两者相关信息传入至 跨模态阅读理解 模型中，模型将会分析并给出预测结果。

接下来，可以根据预测结果中布局信息，对于该汽车说明书中的答案进行高亮显示，最终预测结果将保存至./answer.png 文件中。

三、实验结果

对于排序模块使用 RocketQA 的方法进行训练，在不同的 epoch 下，随着不断迭代，模型准确度将会区域稳定且处于较高的准确度，如图 4 所示。

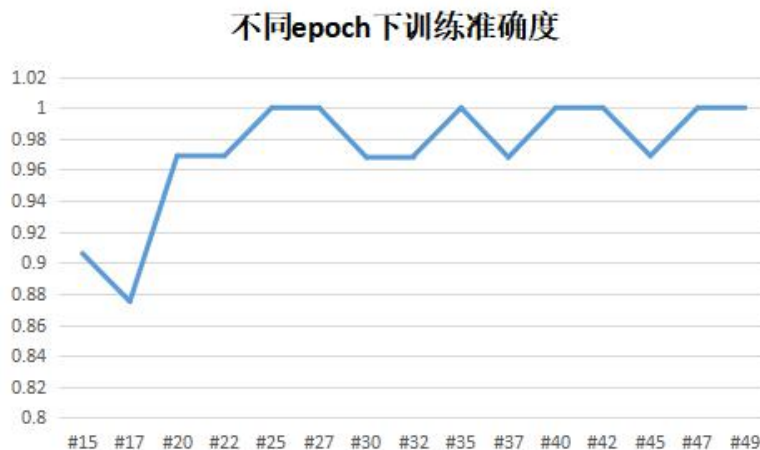


图 4 不同迭代次数下模型准确度

测试集的结果如图 5 所示，对于较为具体的指向性问题，如“后备箱怎么开”精确指明了问题的关键在于“后备箱”、“怎么”和“开”，得到的结果较好。而对于指向不明确的问题“车借给别人”，人类理解上往往倾向于去找到借给别人使需要对钥匙等进行的处理，而机器则并不能做到如此智能，因此得到的结果就较为不尽人意。

| | |
|--|---|
| 先短按上锁按钮, 在 3 秒内进行第二次长按（保持 1.5 秒）可同时关闭全部车窗、天窗及遮阳帘；车窗、天窗及遮阳帘运动过程中放开上锁键则车窗、天窗及遮阳帘停留在当前位置。 | |
| 3. 尾门打开按钮 | 尾门关闭状态时, 先短按尾门按钮, 在 3 秒内进行第二次长按（保持 1.5 秒）可打开尾门; 尾门打开状态时, 再次重复上述操作可关闭尾门。 |
| 4. 智能钥匙电量指示灯 | 智能钥匙的电量可以通过环形指示灯的颜色显示: 蓝色表示电量大于 70%, 绿色表示电量为 20%~70%, 红色表示电量小于 20%。 |
| 5. 寻车按钮 | 车辆处于非驾驶状态且有效智能钥匙距离车辆 10 米以内时, 您可以按下此按钮并保持 3 秒钟以上（或者 3 秒钟内连续按下两次）, 将启动寻车功能, 转向灯及喇叭均发出提示。5 秒钟后再次按下此按钮将解除寻车提示, 否则此寻车功能将在 10 秒钟后自动关闭。 |
| 警告 | 在您离开车辆时如果有人或宠物留在车内, 请务必携带智能钥匙, 钥匙留在车内可能造成车内人员意外伤害。 |
| 注意 | <ul style="list-style-type: none">智能钥匙为电子零件, 请避免撞击、拆卸或放置于高温、潮湿及强烈震动的场所。若您将有效的智能钥匙留在车内, 下车后使用另一把有效的智能钥匙、手机 APP 或 NFC 从车外对车辆进行上锁后, 车内的智能钥匙将失效。再次使用有效的智能钥匙或手机 APP 解锁车辆后, 车内智能钥匙的功能即可恢复。如果在解锁后 30 秒内未打开任何车门或尾门, 所有车门及尾门都自动重新上锁。若钥匙遗失或损坏, 请立即联系蔚来汽车服务中心, 携带当前所有钥匙进行钥匙的解/授权操作, 申领新的钥匙。 |
| 说明 | 为防止儿童玩耍钥匙解锁功能, 若在 10 秒钟内连续按压上锁或解锁按钮达到 9 次, 15 秒内智能钥匙仅可解锁一次车辆且无法上锁车辆。 |
| 智能钥匙与账户绑定 | 首次激活车辆并通过校验时, 智能钥匙默认绑定车主账户。后续使用钥匙解锁车辆时将自动加载车主账户数据。 |

(a) “后备箱怎么开” 问题结果

(b) “车借给别人” 问题结果

图 5 测试问题结果