# 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。

1. **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** |

1. **跨模态阅读理解模块（LayoutXLM模型）**

使用[BERT](https://so.csdn.net/so/search?q=BERT&spm=1001.2101.3001.7020" \t "https://blog.csdn.net/sinat_34072381/article/details/_blank)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](https://arxiv.org/pdf/2104.08836.pdf" \t "https://aistudio.baidu.com/bd-gpu-01/user/2402983/4276813/lab/tree/_blank) 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。

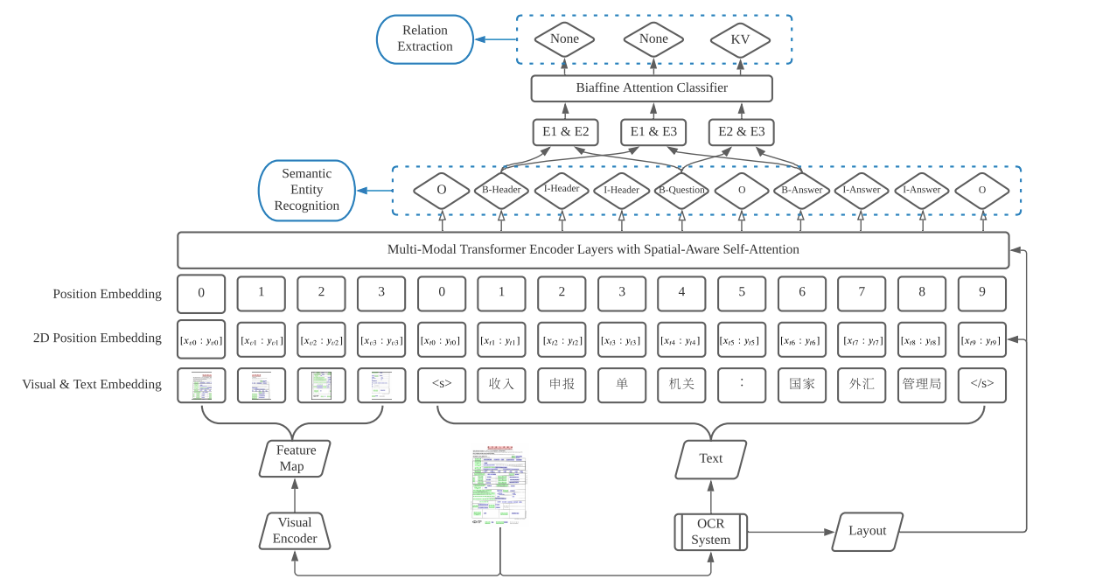


图1 排序模型

1. **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。

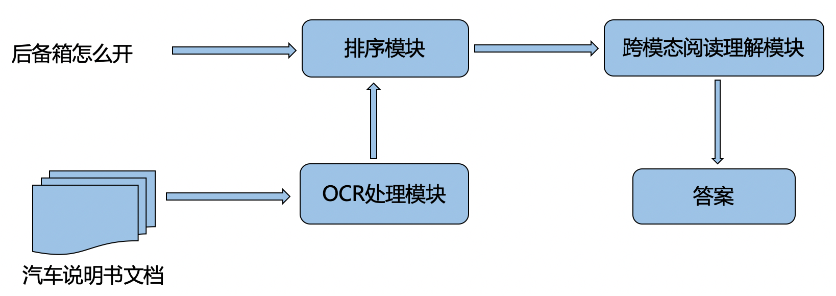


图2 任务完整流程示意图

1.**环境准备**

安装PaddleNLP和PaddleOCR模块

1. **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

1. **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)。

1. **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。

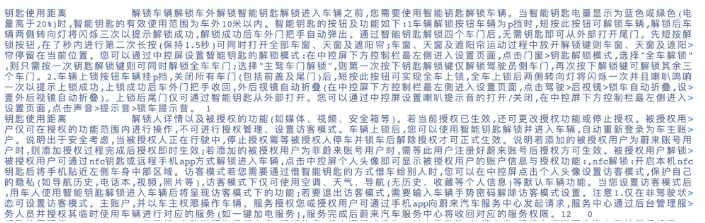


图3 排序模型的训练样本

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。

1. **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。

1. **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。

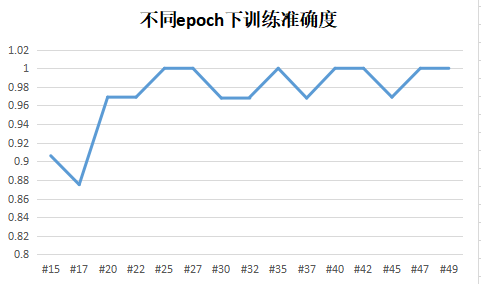
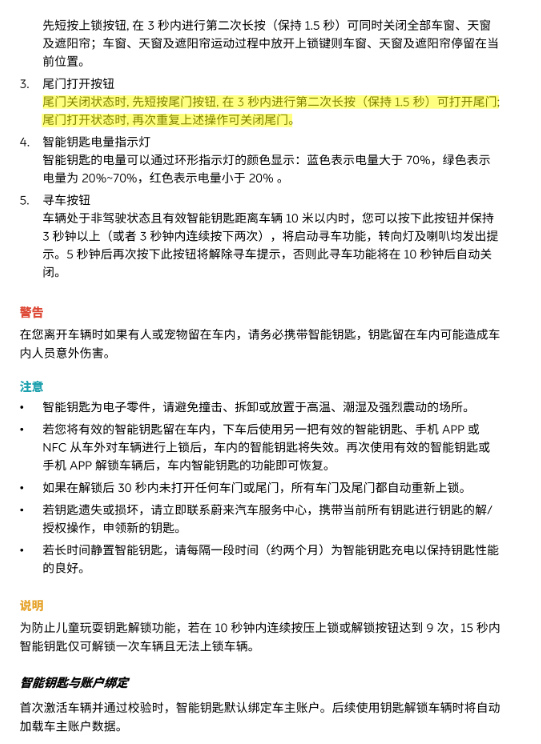
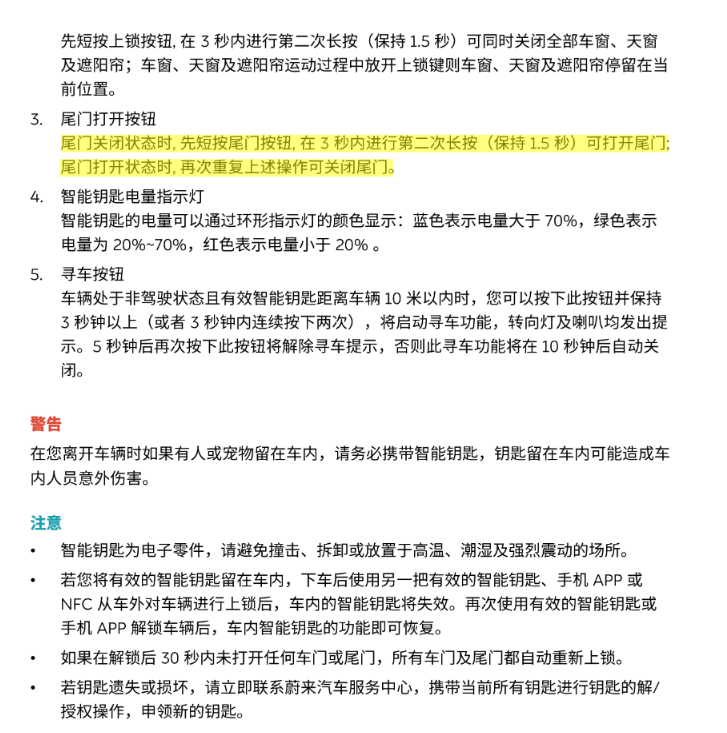


图4 不同迭代次数下模型准确度

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。



(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