重庆邮电大学研究生学位论文修改情况表

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 学号 | S200303078 | 姓名 | 黄子恒 | 导师姓名 | 李鹏华 |
| 专业/领域 | 控制科学与工程 | | | | |
| 学位论文题目 | 面向车载嵌入式设备的智能语音对话方法研究 | | | | |
| 学位论文修改前存在的主要问题（论文评阅专家提出）  【专家1总体意见】：本文针对基于深度编—解码器的自动语音识别模型参数量庞大的问题，提出了一种基于残差分组线性变换的解码器结构。同时，针对基于显式联合建模的自然语言理解模型交互能力不足的问题，提出了一种基于标签感知的图交互模型。最后，此两种算法都在车载嵌入式平台上进行了语音对话，具有一定的工程应用参考价值。  论文书写格式较为规范，查阅了一定量的中英文文献，工作量基本饱满。同意答辩。  【专家1修改意见】：1:缺图目录，表目录，缩写目录，建议修改论文。  Q2:论文第三章3.1-3.4小节都是介绍性的内容，建议精简，同时增加一个章节内容，说明车载环境下语音对话的特点特征，结合这些特点，选用了数据集，建议修改论文。  Q3:论文第4章也存在类似问题，此类方法是针对在车载场景下那种情况下的语音对话特征特点进行数据集选取，对话内容选取的，在此部分中增加一些内容进行补充说明一下。  Q4:第5章在车载设备上的测试，建议最后算法分析，给出对比图表，关键对标指标，这样更具说服力和效果，建议修改论文。  【专家2总体意见】：论文研究了基于Transformer的自动语音识别方法和基于标签感知图交互的自然语言理解方法，首先在自动语音识别上针对于Transformer参数量庞大的问题，提出了一个基于残差分组线性变换的特征缩放单元并将该单元嵌入Transformer，设计出一个轻量的自动语音识别模型，在小幅度降低性能的情况下，实现模型参数量和计算复杂度的降低；然后在自然语言理解中提出标签映射模块获取原始话语和标签语义之间的相关性以提供丰富的先验知识，然后提出全局图交互模块对语句级别的意图—槽位交互过程进行建模以提供全局优化，从而提升模型性能；最后设计一个面向车载嵌入式设备的本地智能语音对话系统，在Nvidia Jetson TX2上部署模型以实现对话系统的离线运行。论文运用的理论知识、研究方法和技术路线较为合理，实验设计较合理，方法和数据较为正确可靠，说明作者在本学科已具有较好的专业理论基础，能够应用所学知识和通过阅读文献解决具体的问题。论文层次较清楚，有较好的逻辑性，文字图表较规范。  【专家2修改意见】：1.文章3.4节和4.2节中，提到了砖石型缩放单元和标签映射模块，但没有充分说明这些方法提出的技术背景信息和动机，为什么要使用砖石型缩放单元和标签映射模块？它们与现有方法有什么区别？  2.文章3.4节中，文章在构建轻量化的Transformer模型时，将原Transformer的前馈网络层换为轻量级的前馈网络层，它主要是通过改变全连接层的输出通道数降低参数运算，这样的替换是否有精度损失？相对于原通道数扩张的做法在这样改进是否可以达到精度与参数量的平衡？  3.文章提出了两种轻量级的神经网络模型，分别用于自动语音识别和自然语言理解，并在多个数据集上对提出的模型进行了广泛的实验评估，并给出了详细的结果和分析但文章都是基于原大模型的框架进行对比，没有与现有的轻量化模型或者其他的适合嵌入式部署的模型进行比较，缺乏说服力。  4.文章5.4节，对模型进行移植时，没有介绍Jetson TX2的软件环境版本信息，例如CUDA版本等，同时是否采用了其支持的硬件加速框架TensorRT？是否是在PyTorch框架下部署移植的？建议可以给出两种框架在Jetson TX2的对比结果。还有是否考虑输入音源时的存在的噪音问题？ | | | | | |
| 学位论文内容所做的具体修改和充实情况  1:缺图目录，表目录，缩写目录，建议修改论文。  回复：在最新版的学位论文模板中关于图、表、缩略词目录的批注为：“均并非必要。”论文考虑到涉及到的图、表和缩略词数量，为了方便排版，故未加图、表、缩略词目录。  Q2:论文第三章3.1-3.4小节都是介绍性的内容，建议精简，同时增加一个章节内容，说明车载环境下语音对话的特点特征，结合这些特点，选用了数据集，建议修改论文。  回复：相关内容已精简。在节3.5开头，补充了关于为何选取这两个数据集的说明，同时在第5章中增加了关于车载环境下的专用驾驶数据集收集的说明。  Q3:论文第4章也存在类似问题，此类方法是针对在车载场景下那种情况下的语音对话特征特点进行数据集选取，对话内容选取的，在此部分中增加一些内容进行补充说明一下。  回复：和上一问题一样，相关内容已精简，且关于数据集选取的说明已补充。  Q4:第5章在车载设备上的测试，建议最后算法分析，给出对比图表，关键对标指标，这样更具说服力和效果，建议修改论文。  回复：在节5.5.2增加了实车测试内容，包括测试界面和测试用例等，同时针对一些关键指标（通过率和响应时间）和近年文献作了比较。  1.文章3.4节和4.2节中，提到了砖石型缩放单元和标签映射模块，但没有充分说明这些方法提出的技术背景信息和动机，为什么要使用砖石型缩放单元和标签映射模块？它们与现有方法有什么区别？  回复：在第3章开头和节3.3.2补充了“钻石”型缩放单元所提出的背景、特征及对该方法在ASR上的改进；同样地，在第4章开头、节4.2和节4.3补充了标签映射模块和全局图交互模块所提出的背景和特征。  2.文章3.4节中，文章在构建轻量化的Transformer模型时，将原Transformer的前馈网络层换为轻量级的前馈网络层，它主要是通过改变全连接层的输出通道数降低参数运算，这样的替换是否有精度损失？相对于原通道数扩张的做法在这样改进是否可以达到精度与参数量的平衡？  回复：在节3.4.1补充了轻量级前馈网络层提出的动机和前提，同时在节3.5.5增加了关于前馈网络层类型的消融实验，实验结果表明该做法可以达到精度与参数量的平衡。  3.文章提出了两种轻量级的神经网络模型，分别用于自动语音识别和自然语言理解，并在多个数据集上对提出的模型进行了广泛的实验评估，并给出了详细的结果和分析但文章都是基于原大模型的框架进行对比，没有与现有的轻量化模型或者其他的适合嵌入式部署的模型进行比较，缺乏说服力。  回复：在节3.5.4增加两种轻量级同类模型，即SSAN和HA-Transformer作为对照组，实验结果结果表明针对自动语音识别的改进模型有较好的效果；而针对自然语言理解模型，由于所搭建模型的方法不同，当前研究难点不在于模型参数量而在于预测性能，在节4.5.4展现了详细的对比结果。  4.文章5.4节，对模型进行移植时，没有介绍Jetson TX2的软件环境版本信息，例如CUDA版本等，同时是否采用了其支持的硬件加速框架TensorRT？是否是在PyTorch框架下部署移植的？建议可以给出两种框架在Jetson TX2的对比结果。还有是否考虑输入音源时的存在的噪音问题？  回复：在节5.2补充了关于TX2刷机后的运行环境版本介绍，包括Python版本、PyTorch版本和CUDA版本等；系统运行未使用TensorRT框架，理由有以下三点：首先在尝试将自然语言理解模型运行在TensorRT框架下时，该模块的运行时间仅提高了10ms左右，在实际应用中并不明显；其次若要部署在TensorRT框架下，需要将.pt模型（PyTorch框架下）转换为ONXX通用模型，这一过程会使得模型参数量变大，以尝试的自然语言理解模型为例，.pt模型参数量为2.46M，转换后为52.17M，可能是由于ONXX中有大量重复的算子，这与论文的轻量级需求稍有偏差；最后由于TensorRT是半开源的，只含有一些经典网络模型的文档资料，而论文中涉及到的网络模型类型较多，在开发上难度较大。 | | | | | |
| 硕士生签字 日期  导师签字 日期 | | | | | |