## 基于深度学习的语音识别技术在智慧教室中的应用与实现

**摘 要**

基于现代教育信息化的发展需求，以及当前疫情形势的现实因素，智慧教室，在线学堂等线上教育应用越来越成为教育教学场景中的重要组成部分。但是由于当前网络环境不稳定以及学生听课场景复杂等因素影响，会极大降低学生听课效率和使用体验，而在学生终端加入字幕可以很好的解决这类问题，帮助学生更好的理解教学内容。随着当前深度学习技术的深入发展与落地，让基于深度学习的语音识别正确率在特定场景下可以达到人类水平，这就为自动字幕生成奠定了基础。基于上述现实需求和意义，本文将使用基于深度学习的语音识别技术实现在智慧教室场景下的自动字幕生成，同时提高教师教学和学生学习的效率。

本文主要使用基于深度学习的端到端语音识别技术实现智慧教室场景下的自动字幕生成功能。首先使用开源的Wenet语音识别框架和教育场景的语音识别数据集训练基础的语音识别模型，然后基于Libtorch对此模型推理代码进行c++实现，同时提供流式识别和非流式识别两种识别模式，丰富了服务请求的灵活性，且极大提高模型推理的时间性能，从而能够实现教师直播课程的实时字幕生成功能，并且后端加入通用语言模型和热词模型进行辅助解码，极大提高在特定场景下的语音识别准确率。在模型部署上，本文使用Docker容器技术对训练好的语音识别模型进行服务封装部署，并实现了支持多种语言调用的Grpc远程过程调用接口，使语音识别服务端和请求端在语言层面和设计层面解耦，方便模型服务的移植部署。最后在字幕生成阶段，字幕生成模块将智慧教室客户端发来的流式语音信号通过流式模型服务请求获得原始的语音识别结果文本，之后对此原始文本进行预处理，主要包括使用预先训练好的模型进行文本纠错和添加标点操作，进一步提高自动生成字幕的准确率和可读性，然后通过时间戳对齐和字幕缓存技术实时生成字幕帧数据，并将其发送给智慧教室客户端进行最终的渲染和显示。

在上一代智慧教室自动字幕生成模块的基础上，本文主要通过模型算法改进和工程性能优化实现了在线直播课程的实时字幕生成，丰富了自动字幕技术的应用场景，提高了智慧教室项目的功能完备性，也为后续基于语音识别服务的高阶功能扩展开发提供了基础。

关键词：深度学习；语音识别；语言模型；自动字幕；智慧教室

## Application and Implementation of Deep Learning-based Speech Recognition Technology in Smart Classrooms

Fengyu Su(Software Engineering)

Directed by Yanbing Jiang

**Abstract**

Based on the development needs of modern education and the real factors of the current epidemic situation, online education applications such as smart classrooms and online learning classes are becoming an important part of the education teaching scene. However, due to the current unstable network environment and the complexity of students' listening scenarios, it can greatly reduce students' listening efficiency and experience, while adding subtitles to students' terminals can well solve such problems and help students better understand the teaching content. With the deep development and implementation of current deep learning technology, the correct rate of speech recognition based on deep learning can reach human level in specific scenarios, which lays the foundation for automatic subtitle generation. Based on the above realistic needs and significance, this paper will use deep learning-based speech recognition technology to realize automatic subtitle generation in smart classroom scenarios and improve the efficiency of teachers' teaching and students' learning at the same time.

In this paper, we mainly use deep learning-based end-to-end speech recognition technology to implement automatic subtitle generation in smart classroom scenarios. Firstly, we use the open source Wenet speech recognition framework and speech recognition dataset for educational scenarios to train the basic speech recognition model, and then implement this model inference code in C++ based on Libtorch, which enriches the flexibility of service requests and greatly improves the time performance of model inference, thus enabling real-time subtitle generation for teachers' live courses, and the back-end incorporates a common language model and a hot word model to assist in decoding, which greatly improves the speech recognition accuracy in specific scenarios. In terms of model deployment, this paper uses Docker container technology to deploy the trained speech recognition model in a service package, and implements a Grpc remote procedure call interface that supports multiple language calls to decouple the speech recognition server and request side at the language level and design level, which facilitates the porting and deployment of the model service. Finally, in the subtitle generation stage, the subtitle generation module takes the streaming speech signal from the smart classroom client and requests the original speech recognition result text through the streaming model service, and then pre-processes this original text, mainly including error correction and punctuation operations using the pre-trained model to further improve the accuracy and readability of the automatically generated subtitles, and then generates the subtitle frame data in real time through timestamp alignment and Then the caption frames are generated in real time through timestamp alignment and caption caching technology and sent to the smart classroom client for final rendering and display.

Compared to the previous generation of automatic subtitle generation module for smart classroom, this paper mainly implements real-time subtitle generation for online live courses through model algorithm improvement and engineering performance optimization, which enriches the application scenarios of automatic subtitle technology, improves the functional completeness of the smart classroom project, and provides a basis for the subsequent development of advanced functions based on speech recognition services.

**Keywords:** Deeplearning, Speech recognition, Languare model, Automatic subtitle generation, Smart Classroom