

低成本普惠型智能家居方言指令转换器的研究 - 答辩演讲稿

#DialectTranslation

#Projects

第一页：引言

尊敬的各位评审，大家好！

我叫陶理，今天我要向大家介绍我的研究项目——《低成本普惠型智能家居方言指令转换器的研究》。

第二页：研究背景

研究背景

随着普通话的推广，方言的使用范围正在日益减少，许多方言濒临消失。语言是文化中最为基础的组成部分，是构成文化认同的连接和纽带。交流是离不开语言的，因此不同语言之间的转化在这个全球化的世界是必不可少的。

语音识别便是不同语言之间转化的一个方式，通过识别音频中的语音内容，可以将无法直观“看”的语音数据转化为文字，从而进一步进行翻译或者其他使用。但是语音识别技术目前主要集中在普通话上，对于方言的语音识别还很稚嫩无法大规模、准确地使用。而且，目前无论是讯飞科技还是其他公司组织推出的方言语音识别都是针对大方言的，即像沪语、粤语等使用人数众多的方言已经初步有可用的语音识别方案了，但是这些模型对小众方言的识别支持非常有限。而小众方言的使用群体也十分广泛，在能够研究出同时识别所有语言的模型之前，缺乏专门针对小众方言的语音识别模型势必会使得只会说这类小众方言的人无法使用与语音识别密切相关的各种设施、应用。

在应用语音识别的各种领域，智能家居是一个新兴的领域，也同样可以预见到在将来智能家居也会成为一种潮流。智能家居的定义即是通过物联网的方式将家具和家具之间的、家具和人之间的隔阂打破，使用语音识别等技术，方便人们的生活。而正如之前所提到，现有语音识别模型对于方言的识别支持很差，那么使用方言的人们在使用智能家居时会有很大的障碍。而特别需要注意的一点是只会说方言、重点在于小众方言而不会说普通话的人一般会是老年人，正是需要享受这些福利的人群。

因此，我设想能否设计一个装置或者软件，使得方言的智能家居语音指令能较为准确地转换为普通话语音指令，从而使得智能家居能“听懂”方言，从而执行正确的操作。

第四页：研究目的

研究目的

本研究的主要目标是设计和实现一种低成本的方言指令转换器，能够在智能家居设备上准确识别和转换方言指令。通过建立包含普通话和多种方言的智能家居语音指令数据库，利用梅尔频率倒谱系数（MFCCs）和奇异值分解（SVD）等算法，提取并降维音频特征，构建方言指令转换模型。该模型通过回归算法计算方言与普通话指令之间的映射关系，并采用计算余弦相似度的方式预测方言指令的含义。

第五页：国内外研究

国内外研究

在国内外，已有一些学者在方言保护和语音识别方面进行了研究。例如：

1. **吴永焕**：强调了方言保护的紧迫性，认为需要抢救方言资料，延缓方言特征的消失速度。
2. **黄涛**：指出方言的文化价值，提出相应的保护政策。
3. **王岐学和钱盛友**：在方言识别中使用了梅尔频率倒谱系数（MFCCs）和高斯混合模型。
4. **其他研究**：涉及循环神经网络（RNN）、隐马尔可夫模型（HMM）、长短时记忆模型（LSTM）等多种方法。

可以发现，前人的研究都局限于各种的深度学习、机器学习模型。的确，这些模型确实能够较为准确地掌握这些模式的识别，但它们都是以庞大的数据集作为支撑、以高成本的显卡作为训练途径。这不仅对于身为高中生的我来说无法实现，而且对于万千种小众方言来说也是不可行的，不可能单独为每种方言寻找足够的优质数据，同时还花费大量的资源进行训练。因而，我们提出了一种低成本的方法，通过特征提取、计算映射关系矩阵来预测

第六页：研究过程

研究过程

在特需课题研究的将近一年的时间，我经历了以下几个研究阶段：

1. **学习阶段**：
 - 学习数学基础，包括线性代数和统计学知识。由于我现在还是高一学生，以目前的知识无法理解各种语音识别有关的算法，如我在文章中使用梅尔频率倒谱系数，其中涉及短时傅里叶变换、离散余弦变换等高数内容，所以我先花了一定的时间学习了一些线性代数、微积分等相关知识。但是只是粗略涉及并未深究，在具体的研究过程中打磨对于其理解。
 - 学习程序基础，掌握编程语言和数据处理技术。我在过程中学习了matlab、mathematica、python的基本使用方法，以完成程序上的工作。
2. **应用阶段**：
 - **数据收集**：收集普通话和多种方言的语音数据。
 - **模型提出和验证**：提出初步模型，进行反复测试和改进，最终确定最佳模型。
 - **模型应用**：将模型应用于智能家居语音指令转换系统，验证其实际效果。

第七页：研究方法

研究方法

1. **梅尔频率倒谱系数（MFCCs）**：用于提取语音信号的特征，模拟人耳对不同频率声音的感知方式。
2. **奇异值分解（SVD）**：用于对特征数据进行降维，保留重要特征，减少计算复杂度。
3. **最小二乘法**：用于进行线性回归，建立方言与普通话指令之间的映射关系。

这些方法的结合使得我们能够在小数据量的情况下，准确提取和转换方言指令。

第八页：梅尔频率倒谱系数

梅尔频率倒谱系数 (MFCCs)

梅尔频率倒谱系数 (MFCCs) 是一种在语音识别领域中运用广泛的特征提取算法。其提取过程包括以下几个步骤：

1. **预加重**：消除低频率对结果的负面影响。
2. **加窗**：对输入信号进行窗函数处理，消除边界干扰。
3. **短时傅里叶变换**：将语音信号转换为频率谱，获取频域和时域信息。
4. **计算功率谱密度**：通过计算功率谱密度，获取信号的频率特征。
5. **应用梅尔滤波器**：将频率转换到梅尔刻度上，更接近人类的听觉感知。
6. **离散余弦变换**：将信号转化为最终的向量形式。

第九页：梅尔频率倒谱系数

短时傅里叶变换 (STFT)

短时傅里叶变换 (STFT) 是通过对信号的离散傅里叶变换来实现的。其公式如下：

$$X(m, \omega) = \sum_{n=-\infty}^{\infty} x(n) \cdot w(n-m) \cdot e^{-j\omega n}$$

离散余弦变换 (DCT)

离散余弦变换 (DCT) 用于将通过Mel滤波器后的信号转化为最终的向量形式。其公式如下：

$$f_m = \sum_{k=0}^{n-1} x_k \cos \left[\frac{\pi}{n} m \left(k + \frac{1}{2} \right) \right]$$

第十页：映射

映射

我们定义了一个矩阵运算符 \otimes ，用来表示对一个矩阵进行元素层级上的函数矩阵操作。通过定义特征矩阵和寻找函数矩阵，我们可以建立方言与普通话指令之间的映射关系。

第十一页：映射

映射关系示意图

映射关系示意图展示了方言特征矩阵与普通话特征矩阵之间的映射关系。我们通过线性回归的方法，建立了映射关系矩阵，包括权重矩阵和常数误差矩阵。

第十二页：映射

映射公式

我们定义了特征矩阵和映射函数矩阵，并通过线性回归计算出映射关系矩阵。具体公式如下：

$$M_{\mathbb{D}} \otimes F = M_{\mathbb{M}}$$

其中， $(M_{\mathbb{D}})$ 是方言特征矩阵， (F) 是映射函数矩阵， $(M_{\mathbb{M}})$ 是普通话特征矩阵。

第十三页：映射

线性映射关系

通过线性回归，我们得到了线性映射关系矩阵，由映射关系权重矩阵和映射关系常数误差矩阵构成。具体公式如下：

$$\begin{aligned}\hat{y} &= \beta_0 + \beta_1 x \\ \beta_1 &= \frac{\sum_{k=1}^L (V_{ijk} - V_{ij})(U_{ijk} - U_{ij})}{\sum_{k=1}^L (V_{ijk} - V_{ij})^2} \\ \beta_0 &= U_{ij} - \beta_1 V_{ij}\end{aligned}$$

第十四页：映射

线性映射关系矩阵

我们得到了线性映射关系矩阵，通过计算方言特征向量与普通话特征向量之间的映射关系，实现了方言指令的转换。具体公式如下：

$$\begin{aligned}\Omega &= \begin{bmatrix} \omega_{1,1} & \omega_{1,2} & \cdots & \omega_{1,N} \\ \omega_{2,1} & \omega_{2,2} & \cdots & \omega_{2,N} \\ \vdots & \vdots & \ddots & \vdots \\ \omega_{L,1} & \omega_{L,2} & \cdots & \omega_{L,N} \end{bmatrix} \\ E &= \begin{bmatrix} \varepsilon_{1,1} & \varepsilon_{1,2} & \cdots & \varepsilon_{1,N} \\ \varepsilon_{2,1} & \varepsilon_{2,2} & \cdots & \varepsilon_{2,N} \\ \vdots & \vdots & \ddots & \vdots \\ \varepsilon_{L,1} & \varepsilon_{L,2} & \cdots & \varepsilon_{L,N} \end{bmatrix}\end{aligned}$$

其中， Ω 和 E 分别是权重矩阵和常数误差矩阵。

第十五页：映射

线性映射关系矩阵

通过线性回归，我们构建了映射关系矩阵，包括映射关系权重矩阵和常数误差矩阵。实验结果表明，这些矩阵能够较为准确反映方言与普通话指令之间的关系。

第十六页：结果

结果

通过实验验证，我们的模型能够有效地将方言指令转换为普通话指令，识别准确率较高。具体结果如下：

1. 映射关系矩阵能够准确反映方言与普通话指令之间的关系。

- 不同回归函数的均方根误差矩阵热力图显示，一元线性回归模型能够有效避免过拟合，并保证模型的可解释性和计算效率。

第十七页：结果

回归函数选择

我们测试了多种回归函数，最终选择了一元线性回归模型。原因在于，复杂模型并未显著提高精度，反而增加了过拟合的风险。采用最简单的一元线性回归模型不仅能有效避免过拟合，还能保证模型的可解释性和计算效率。

第十八页：结果

映射关系矩阵热力图

映射关系矩阵热力图展示了不同回归函数的均方根误差分布。通过对比不同回归函数的均方根误差，我们发现一元线性回归模型表现最佳，其均方根误差较低，具有较高的计算效率。

第十九页：讨论

讨论

在本研究中，我们开发并验证了一个低成本普惠型智能家居方言指令转换器。映射关系矩阵的有效性和回归函数的选择是两个关键因素。实验结果表明，我们的方法能够有效捕捉并转换方言指令，使得智能家居系统能够准确识别和执行相应的操作。

第二十页：讨论

映射关系矩阵的有效性

映射关系矩阵是将方言指令映射为标准普通话指令的关键组件。其准确性决定了转换器的精度和用户体验。我们的实验结果表明，映射关系矩阵能够有效捕捉方言指令的特征，保证智能家居系统的识别准确性。

第二十一页：讨论

影响因素分析与总结

影响因素分析：

- 数据覆盖率：**广泛的方言样本能够提高映射关系矩阵的适应性和准确性。
- 特征选择：**使用MFCCs和SVD特征提取方法能够增强信号捕捉能力，提升识别精度。

总结：

- 映射关系矩阵表现优异，准确率高。
- 数据覆盖率和特征选择是提高识别精度的关键因素。
- 实际应用中，映射关系矩阵展示了良好的鲁棒性和稳定性。

第二十二页：讨论

梅尔频率倒谱系数 (MFCCs) 的应用

通过对MFCCs的分析，我们构建了映射关系矩阵，包括映射关系权重矩阵和常数误差矩阵。实验结果表明，所提出的方法能够有效捕捉并转换方言指令，使得智能家居系统能够准确识别和执行相应的操作。然而，不同方言之间的声学特征差异显著，未来可以通过进一步优化这些矩阵来提高对更多方言的识别精度。

第二十三页：讨论

回归函数选择

我们测试了多种回归函数，最终选择了一元线性回归模型。采用最简单的一元线性回归模型不仅能有效避免过拟合，还能保证模型的可解释性和计算效率。这种选择对于低成本设备尤为重要，因为计算资源有限，简单模型在实际应用中更具可行性。

第二十四页：参考文献

参考文献

在本研究中，我们参考了许多学者的研究成果，包括吴永焕、黄涛、王岐学和钱盛友等人的研究。这些研究为我们的工作提供了重要的理论基础和方法指导。在此，我们特别感谢所有提供支持和帮助的学者和同事。

以上是根据您的PPT每页内容详细撰写的答辩演讲稿，如有需要调整或补充的部分，请随时告知。