低成本普惠型智能家居方言指令转换器的研究

陶理

May 24, 2024

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，方言处于濒临消失的边缘。
- 语音识别领域着重于普通话的识别，而对于方言，特别是小众方言的识别受到了忽略。
- 使用方言进行智能家居控制的精度受限。

必要性分析

国内外研究

有关方言保护

- 吴永焕[3]就方言的保护在必要性和紧迫性的方面进行了论证，强调了要抢记方言资料，尽可能延缓方言特征消失速度；
- 黄涛[9]强调了方言在文化价值上的重要性；相应的保护政策也相继出台，因而保护方言是很有必要的。
- 武瑞丰[7]提出了人工智能在方言建档中有很大的作用和优点

国内外研究

有关语音识别和方言识别

- 王岐学、钱盛友[8]：MFCCs，差分特征，高斯混合模型
- 杨波[6]：RNN，桂柳方言
- 张宇聪[4]：HMM，LSTM
- 余陆峰[2]：Tensorflow，客家方言
- 杨奭喆[5]和彭煦潭等[1]：无监督跨语言词向量

研究过程

1. 学习
   - 数学基础
   - 程序基础
2. 应用
   - 收集数据
   - 提出、打磨、推翻再提出模型
   - 应用并验证模型

研究方法

- MFCCs (梅尔频率倒谱系数)
- SVD (奇异值分解)
- OLS (最小二乘法)
- 回归算法

梅尔频率倒谱系数

Mel刻度，反映人耳对频率的感知:

$$Mel(f) = 2595 \log_{10}(1 + f/700) \tag{1}$$

加窗，目的为消除边界干扰:

$$w(n) = H(n) = 0.54 - 0.46 \cos\left(\frac{2\pi n}{N-1}\right) \tag{2}$$

短时傅里叶变换

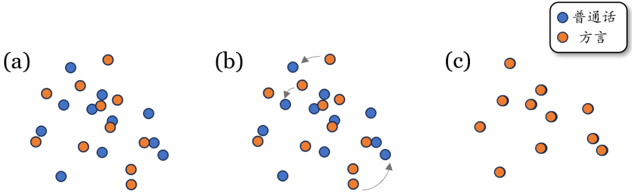
$$X(m,\omega) = \sum_{n=-\infty}^{\infty} x(n) \cdot w(n-m) \cdot e^{-j\omega n} \tag{3}$$

离散余弦变换

$$f_m = \sum_{k=0}^{n-1} x_k \cos[\frac{\pi}{n} m(k+\frac{1}{2})] \tag{4}$$

映射

Figure: 映射关系示意图



为方便表示，定义运算符$\otimes$，称其为对一个矩阵X作元素层级上的函数矩阵操作F:

$$X \otimes F \tag{5}$$

$$X = \begin{bmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,N} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,N} \\ \vdots & \vdots & \ddots & \vdots \\ x_{L,1} & x_{L,2} & \cdots & x_{L,N} \end{bmatrix}, F = \begin{bmatrix} f_{1,1} & f_{1,2} & \cdots & f_{1,N} \\ f_{2,1} & f_{2,2} & \cdots & f_{2,N} \\ \vdots & \vdots & \ddots & \vdots \\ f_{L,1} & f_{L,2} & \cdots & f_{L,N} \end{bmatrix} \tag{6}$$

$$X \otimes F = \begin{bmatrix} f_{1,1}(x_{1,1}) & f_{1,2}(x_{1,2}) & \cdots & f_{1,N}(x_{1,N}) \\ f_{2,1}(x_{2,1}) & f_{2,2}(x_{2,2}) & \cdots & f_{2,N}(x_{2,N}) \\ \vdots & \vdots & \ddots & \vdots \\ f_{L,1}(x_{L,1}) & f_{L,2}(x_{L,2}) & \cdots & f_{L,N}(x_{L,N}) \end{bmatrix} \tag{7}$$

有特征矩阵$M_{\mathbb{D}}$和$M_{\mathbb{M}}$：

$$M_{\mathbb{D}} = \begin{bmatrix} d_{1,1} & d_{1,2} & \cdots & d_{1,N} \\ d_{2,1} & d_{2,2} & \cdots & d_{2,N} \\ \vdots & \vdots & \ddots & \vdots \\ d_{L,1} & d_{L,2} & \cdots & d_{L,N} \end{bmatrix} M_{\mathbb{M}} = \begin{bmatrix} m_{1,1} & m_{1,2} & \cdots & m_{1,N} \\ m_{2,1} & m_{2,2} & \cdots & m_{2,N} \\ \vdots & \vdots & \ddots & \vdots \\ m_{L,1} & m_{L,2} & \cdots & m_{L,N} \end{bmatrix} \tag{8}$$

寻找函数矩阵F：

$$F = \begin{bmatrix} f_{1,1} & f_{1,2} & \cdots & f_{1,N} \\ f_{2,1} & f_{2,2} & \cdots & f_{2,N} \\ \vdots & \vdots & \ddots & \vdots \\ f_{L,1} & f_{L,2} & \cdots & f_{L,N} \end{bmatrix} \tag{9}$$

使得满足下式：

$$M_{\mathbb{D}} \otimes F = M_{\mathbb{M}} \tag{10}$$

$$M_{\mathbb{D}} \otimes F = M_{\mathbb{M}} \tag{11}$$

即：

$$\begin{bmatrix} m_{1,1} & m_{1,2} & \cdots & m_{1,N} \\ m_{2,1} & m_{2,2} & \cdots & m_{2,N} \\ \vdots & \vdots & \ddots & \vdots \\ m_{L,1} & m_{L,2} & \cdots & m_{L,N} \end{bmatrix} = \begin{bmatrix} f_{1,1}(d_{1,1}) & f_{1,2}(d_{1,2}) & \cdots & f_{1,N}(d_{1,N}) \\ f_{2,1}(d_{2,1}) & f_{2,2}(d_{2,2}) & \cdots & f_{2,N}(d_{2,N}) \\ \vdots & \vdots & \ddots & \vdots \\ f_{L,1}(d_{L,1}) & f_{L,2}(d_{L,2}) & \cdots & f_{L,N}(d_{L,N}) \end{bmatrix} \tag{12}$$

将映射关系限定为线性关系，则对矩阵中每个元素考虑线性回归：

$$\hat{y} = \beta_0 + \beta_1 x \tag{13}$$

$$\begin{cases} \beta_1 = \dfrac{\sum_{k=1}^{L}(\mathbf{V}_{ij_k} - \overline{\mathbf{V}_{ij}})(\mathbf{U}_{ij_k} - \overline{\mathbf{U}_{ij}})}{\sum_{k=1}^{L}(\mathbf{V}_{ij_k} - \overline{\mathbf{V}_{ij}})^2} \\ \beta_0 = \overline{\mathbf{U}_{ij}} - \beta_1 \overline{\mathbf{V}_{ij}} \end{cases} \tag{14}$$

于是我们有线性映射关系矩阵，由映射关系权重矩阵Ω和映射关系常数误差矩阵E构成。

$$\Omega=\begin{bmatrix}\omega_{1,1}&\omega_{1,2}&\cdots&\omega_{1,N}\\\omega_{2,1}&\omega_{2,2}&\cdots&\omega_{2,N}\\\vdots&\vdots&\ddots&\vdots\\\omega_{L,1}&\omega_{L,2}&\cdots&\omega_{L,N}\end{bmatrix},E=\begin{bmatrix}\varepsilon_{1,1}&\varepsilon_{1,2}&\cdots&\varepsilon_{1,N}\\\varepsilon_{2,1}&\varepsilon_{2,2}&\cdots&\varepsilon_{2,N}\\\vdots&\vdots&\ddots&\vdots\\\varepsilon_{L,1}&\varepsilon_{L,2}&\cdots&\varepsilon_{L,N}\end{bmatrix} \tag{15}$$

Ω和E即等价于映射函数矩阵F

$$F=\begin{bmatrix}f_{1,1}&f_{1,2}&\cdots&f_{1,N}\\f_{2,1}&f_{2,2}&\cdots&f_{2,N}\\\vdots&\vdots&\ddots&\vdots\\f_{L,1}&f_{L,2}&\cdots&f_{L,N}\end{bmatrix} \tag{16}$$

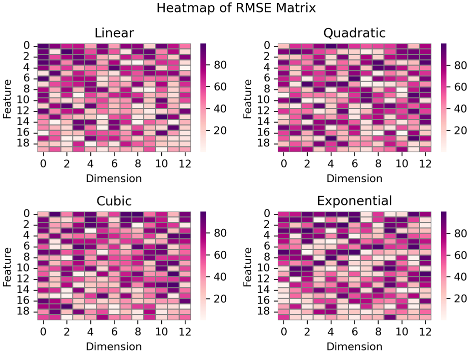
其中

$$F_{i,j}(x)=\omega_{i,j}x+\varepsilon_{i,j} \tag{17}$$

回归函数选择

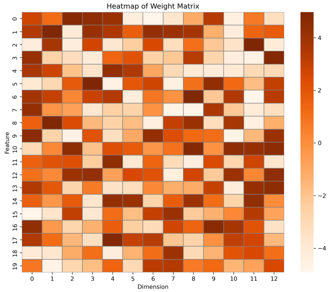
考虑使用不同函数回归结果的均方根误差（RMSE），发现不同的常见函数的结果无明显差异。为避免过拟合，采用最简单的一元线性回归模型。

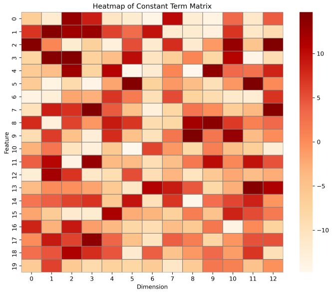
Figure: 不同回归函数均方根误差矩阵热力图



映射关系矩阵

Figure: 映射关系矩阵热力图





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。以下是对研究结果的深入讨论：

- 映射关系矩阵的有效性
- 回归函数的选择

映射关系矩阵的有效性

- **定义**: 将方言指令映射为标准普通话指令的关键组件。
- **重要性**: 决定转换器的准确性和用户体验。

影响因素分析与总结

- **影响因素分析**:
  - **数据覆盖率**: 广泛的方言样本提高了矩阵的适应性。
  - **特征选择**: 使用MFCC和SVD特征提取增强了信号捕捉能力。
- **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关键因素。
  - 实际应用中展示了良好的鲁棒性和稳定性。

通过对梅尔频率倒谱系数（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 Ω 和映射关系常数误差矩阵 E。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，在不同方言之间，映射关系矩阵的准确性存在差异。这表明，不同方言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。

回归函数的选择

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，在比较均方根误差（RMSE）后发现，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。这种选择对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。

参考文献

[1] Xutan Peng, Mark Stevenson, Chenghua Lin, and Chen Li. Understanding linearity of cross-lingual word embedding mappings. *arXiv preprint arXiv:2004.01079*, 2020.

[2] 余陆峰. 基于深度学习的客家方言语音识别. Master's thesis, 华南理工大学, 2019.

[3] 吴永焕. 汉语方言文化遗产保护的意义与对策. 中国人民大学学报, 22(4):5, 2008.

[4] 张宇聪. 基于深度学习的藏语拉萨方言语音识别的研究. PhD thesis, 兰州: 西北师范大学物理与电子工程学院, 2016.

[5] 杨奭喆. 基于映射的无监督跨语言词向量模型研究. Master's thesis, 哈尔滨工业大学, 2020.

[6] 杨波. 基于 rnn 的桂柳方言语音识别系统研究. 现代计算机, (31):6–9, 2019.

[7] 武瑞丰. 人工智能视域下的林州方言建档. 山西档案, (1):62–64, 2018.

[8] 王岐学, 钱盛友, and 赵新民. 基于差分特征和高斯混合模型的湖南方言识别. 计算机工程与应用, 45(35):129–131, 2009.

[9] 黄涛. 语言文化遗产的特性, 价值与保护策略. 中国人民大学学报, 22(4):26–33, 2008.