小米智能音箱“小爱同学”涉及的语音识别技术

姓名：郑雨婷 学号：2021150122

# 系统建模

在针对语音识别进行句子情感分析这个问题时，我们进行系统建模，遵循以下五个步骤：

①问题定义：

我们在拿到一段语音，并且将其正确地转变为一端文本文字之后，我们的目标是去判断它所包含的是正面情感还是负面情感。

②数据收集与准备：

我们使用了IMDB电影评论数据集。IMDB数据集通常指的是其中一个子集，即IMDB电影评论数据集（IMDB Movie Review Dataset）。这个数据集由大量来自IMDB网站的电影评论组成，被广泛应用于情感分析和自然语言处理任务中。IMDB电影评论数据集包含了一万多条英文电影评论，每个评论都有相应的情感标签，表示为正面情感（positive）或负面情感（negative）。这些评论通常是与对应电影观众的真实评论相对应的，并且经过了人工标注的处理。

③特征工程：

我们对预处理后的语音数据进行特征向量转换，以便模型使用。我们使用TF-IDF（词频-逆文档频率）向量化方法将文本转换为数值特征表示。这能够捕捉到不同词汇的重要性和频率，例如语调、韵律和能量变化等特征，可以通过基于信号处理的方法进行提取。

④标注数据：

对语音数据中的情感类别和强度进行人工或自动标注，将其标记为正面或负面。

⑤模型训练与优化：

首先，选择适合的模型架构。在这个问题中，我们选择了支持向量机（SVM）作为我们的模型。SVM是一种常用的监督学习算法，在分类问题中表现出色。之后，我们将数据集划分为训练集和测试集。训练集用于模型的训练，而测试集则用于评估模型的性能。这样能够更好地了解模型在未见过的数据上的表现。接下来，我们使用训练集来对模型进行训练，并使用验证集进行参数调整。通过不断调整模型的参数，我们可以找到合适的参数配置，使得模型在验证集上的表现最佳。在系统建模过程中，我们对选定的算法进行训练和优化。训练阶段，我们使用标记好的数据来调整模型的参数，以使其在未见过的数据上具有更好的泛化能力。通过训练和优化，我们希望模型能够从训练数据中学到有效的模式和规律，并能够准确地推广到新的数据中。

⑥模型评估与验证：

我们使用准确率作为主要评估指标来衡量模型的性能。准确率是通过计算预测结果与真实标签一致的样本数量与总样本数量之间的比例得出的。准确率较高，说明模型在训练过程中取得了良好的成果，并且能够对新的样本做出较为准确的预测。通过对模型进行评估和验证，可以帮助我们判断模型是否具有良好的泛化能力，即在未见过的数据上的表现是否仍然可靠。这有助于确定模型的可信度和可用性。

# 参考文献:

1. Help People who are Blind or Partially Sighted. https://www.orcam.com/en/
2. Hewett R, Douglas G, Keil S. Young people, visual impairment and preparing to live independently[J]. Visual Impairment Centre for Teaching and Research, University of Birmingham, 2015.
3. Bigham J P, Jayant C, Ji H, et al. Vizwiz: nearly real-time answers to visual questions[C]//Proceedings of the 23nd annual ACM symposium on User interface software and technology. 2010: 333-342.
4. https://blog.csdn.net/Rmwcf/article/details/125663393
5. [小米AI音箱\_百度百科 (baidu.com)](https://baike.baidu.com/item/%E5%B0%8F%E7%B1%B3AI%E9%9F%B3%E7%AE%B1/22046214#:~:text=%E5%B0%8F%E7%B1%B3AI%E9%9F%B3%E7%AE%B1%E5%8F%AF%E4%BB%A5%E5%92%8C%E7%B1%B3%E5%AE%B6APP%E5%BD%93%E4%B8%AD%E7%BB%91%E5%AE%9A%E7%9A%84%E9%83%A8%E5%88%86%E5%B0%8F%E7%B1%B3%E6%99%BA%E8%83%BD%E5%AE%B6%E5%B1%85%E4%BA%A7%E5%93%81%E8%BF%9B%E8%A1%8C%E6%8D%86%E7%BB%91%EF%BC%8C%E9%80%9A%E8%BF%87%E6%8E%88%E6%9D%83%E4%B9%8B%E5%90%8E%E5%8F%AF%E4%BB%A5%E9%80%9A%E8%BF%87%E9%9F%B3%E7%AE%B1%E7%9A%84%E8%AF%AD%E9%9F%B3%E4%BA%A4%E4%BA%92%E8%BF%9B%E8%A1%8C%E6%8E%A7%E5%88%B6%E3%80%82,%E5%8F%AF%E6%8E%A7%E5%88%B6%E7%94%B5%E8%A7%86%E3%80%81%E7%9B%92%E5%AD%90%E3%80%81%E6%89%AB%E5%9C%B0%E6%9C%BA%E5%99%A8%E4%BA%BA%E3%80%81%E7%94%B5%E9%A5%AD%E7%85%B2%E3%80%81%E7%A9%BA%E6%B0%94%E5%87%80%E5%8C%96%E5%99%A8%E3%80%81%E7%94%B5%E9%A3%8E%E6%89%87%E3%80%81%E6%99%BA%E8%83%BD%E7%81%AF%E7%AD%89%E5%B0%8F%E7%B1%B3%E5%8F%8A%E7%94%9F%E6%80%81%E9%93%BE%E8%AE%BE%E5%A4%87%E3%80%82)
6. [小米官方揭秘小米AI技术 一文看懂小爱同学进化的奥秘\_ZNDS资讯](https://news.znds.com/article/52204.html)
7. Y. Kong et al., "Multi-Channel Automatic Speech Recognition Using Deep Complex Unet," 2021 IEEE Spoken Language Technology Workshop (SLT), Shenzhen, China, 2021, pp. 104-110, doi: 10.1109/SLT48900.2021.9383492.