

小爱同学语音识别技术之神经网络语义分析研究

姓名：叶茂林 学号：2021155015

1 应用概述

1.1 本文概述

随着人工智能技术的迅速发展，智能手机和智能音箱等设备的普及，语音识别技术已经广泛应用于人机交互、智能助理、语音搜索等场景中，逐渐成为人机交互领域的重要组成部分。

本文主要对小爱同学语音识别技术中的神经网络语义分析原理进行深入研究和分析，从应用概述、涉及的科学技术问题、相关技术分析及国内外研究现状的角度综述了该技术的进展，并从数学系统建模及算法设计技术角度进行分析和实验验证。

1.2 小爱同学语音识别技术应用

由小米公司开发的智能语音助手小爱同学，作为一个典型的基于语音识别技术的人工智能系统，其语音识别技术的研究具有重要的实践意义。

小爱同学的语音识别技术应用主要可以分为以下几个方面：

- 语音命令控制：**通过语音识别技术，用户可以使用口头命令来控制手机、电视、音箱等智能设备的各种功能，例如打开应用程序、调节音量、播放音乐、设置闹钟等。这种方式比传统的手动操作更加方便快捷，使用户的交互体验更加人性化。
- 智能家居控制：**小爱同学可以与智能家居设备进行连接，通过语音识别技术，用户可以通过对话的方式控制家里的灯光、空调、窗帘、电器等，实现智能家居的自动化管理。用户只需要说出相应的指令，小爱同学就能根据识别结果触发相应的设备操作。
- 实时翻译：**小爱同学还具备实时翻译的功能。通过语音识别技术，它能够将用户说出的中文语音内容迅速转化为其他语言的文字，并通过语音合成技术将翻译结果以听得懂的语言播放出来。这在旅行、学习和跨文化交流中起到了很大的作用。
- 智能问答与知识查询：**用户可以通过语音方式向小爱同学提问，获得相应的回答。小爱同学通过庞大的知识库和自然语言处理技术，可以理解和分析用户的问题，并给出准确的回答。无论是常识性问题、科技问题还是娱乐问题，小爱同学都能够做到即时解答。

2 涉及的科学技术问题

2.1 小爱同学的语音识别过程

小爱同学的语音识别过程主要过程如下：

- 声音采集：**用户使用麦克风等设备进行语音输入，将声音信号转换为电信号。
- 声音信号预处理：**接下来对采集到的声音信号进行预处理，包括去除噪声、降低回声等，以提高后续处理的准确性。

- ③ 特征提取：通过分析声音信号的频谱、时域特性等，提取出表示声音特征参数，如梅尔频率倒谱系数（MFCC）、过零率等。
- ④ 声学模型训练：使用大量标注好的语音数据，采用机器学习方法，例如深度神经网络（DNN）或循环神经网络（RNN），训练声学模型。声学模型用于将声音特征与语音单元（如音素）进行匹配，实现对语音的识别。
- ⑤ 语言模型设计：语言模型用于增强语音识别的语义理解和上下文推测能力。它基于大量文本数据，利用统计语言建模的方法，推测出句子的概率分布，从而提供更准确的语音识别结果。
- ⑥ 解码与识别：在解码阶段，声音特征经过声学模型和语言模型的联合处理，利用搜索算法（如动态规划）找到最可能的识别结果，即最佳匹配的句子。
- ⑦ 输出结果：识别出的文本结果可以通过语音合成技术转化为语音输出给用户，或者直接以文本形式展示在设备屏幕上。

2.2 小爱同学语义分析相关技术

小爱同学语义分析涉及的科学技术问题主要涉及自然语言处理（Natural Language Processing, NLP）和人工智能（Artificial Intelligence, AI）领域的相关技术，加入了使用深度复杂 Unet 的多通道自动语音识别[1]，包括以下几个科学技术问题：

- ① 词法分析与句法分析：语义分析首先需要对输入的文本进行词法和句法分析。词法分析将文本分解为单词或词汇单元，而句法分析则分析句子的结构和组织方式，包括词与词之间的关系。这些分析工具可以帮助理解句子的基本含义和语法结构。
- ② 语义角色标注：语义角色标注是指对句子中的每个成分进行语义角色的标注，如主语、宾语、谓语等。通过标记语义角色，可以更好地理解句子的语义信息，进而进行更深入的语义分析。
- ③ 语义解析：语义解析是指将句子转换为计算机可以理解的形式，例如语义表示形式（Semantic Representation），以便进行进一步的分析和处理。常用的语义表示形式包括逻辑形式、谓词逻辑形式、语义图等。语义解析的目标是使计算机能够准确理解句子的含义，并从中提取出需要的信息。
- ④ 语义关系建模：在语义分析中，需要对句子中的语义关系进行建模。这些关系包括词与词之间的关联、实体与实体之间的关系等。通过建模语义关系，可以帮助理解文本的上下文信息，并对文本进行更细粒度的分析和推理。
- ⑤ 语义推理与知识图谱：语义分析还涉及到语义推理和知识图谱的应用。语义推理是基于逻辑或规则的方法，通过推理来获取额外的语义信息或推导出新的结论。知识图谱是一种将实体、属性和关系组织起来形成的图谱，可以用于存储和查询大量结构化的语义信息。通过利用知识图谱，可以提高语义分析的准确性和效果。
- ⑥ 深度学习与神经网络：在语义分析中，深度学习和神经网络技术被广泛应用。深度学习模型，如循环神经网络（Recurrent Neural Network, RNN）和变换器（Transformer），可以对文本进行编码和表示学习，从而捕获文本的语义信息。通过训练大规模的神经网络模型，可以提高语义分析的性能和泛化能力。

总而言之，小米小爱同学语义分析是基于自然语言处理和人工智能技术的应用，涉及词法分析、句法分析、语义角色标注、语义解析、语义关系建模、语义推理与知识图谱以及深度学习与神经网络等科学技术问题。它的目标是理解和处理人类语言，使得机器能够准确把握用户的意图，并提供相应的智能服务。

3 相关技术分析国内外研究现状

3.1 相关技术分析

小爱同学是基于自然语言处理和人工智能技术开发的智能助手，其中神经网络技术在其语义理解和对话生成等方面发挥了重要作用。神经网络在小爱同学中应用于自然语言理解、对话系统、生成模型和多模态处理等方面，并在国内外的研究中得到了广泛探索和应用。随着深度学习和神经网络技术的不断发展，相信未来在小爱同学中的应用将会有更多的突破和创新。以下是神经网络在小爱同学中应用的相关技术：

- ① 自然语言理解[2] (NLU)：神经网络在自然语言理解中起到了关键作用。语义角色标注、命名实体识别、句法分析等任务可以使用循环神经网络[3] (RNN) 或者卷积神经网络[4] (CNN) 进行建模，并通过学习大量标注数据来捕捉文本的上下文信息和语义特征。此外，由于神经网络具有端到端的优势，还可以使用深度学习模型直接对文本进行编码和表示学习。
- ② 对话系统：神经网络在对话系统中被广泛应用。传统的基于规则或模板的对话系统被深度学习方法取代，通过使用循环神经网络 (RNN)、变换器 (Transformer) 或者生成对抗网络 (GAN) [5] 等模型，可以训练出端到端的对话生成模型。这些模型可以根据上下文理解用户的意图，并生成自然流畅的回复。
- ③ 生成模型：生成模型在自然语言处理中被广泛应用，包括神经机器翻译、文本摘要、文章生成等任务。近年来，基于变换器的模型，如 BERT[6] (Bidirectional Encoder Representations from Transformers) 和 GPT[7] (Generative Pre-trained Transformer) 等，取得了重大突破，可以通过预训练和微调的方式，在多种任务上取得优秀的效果。这些模型在小爱同学中也得到了应用，能够生成自然、准确的回答。
- ④ 多模态处理：除了纯文本信息，神经网络在小爱同学中还可以处理图像、语音和视频等多模态数据。例如，可以使用卷积神经网络 (CNN) 或者预训练的视觉模型进行图像理解和分析，使用循环神经网络 (RNN) 进行语音识别和语音理解。多模态处理可以进一步提升小爱同学对用户需求的理解和交互的灵活性。

3.2 国内的研究现状

在中国，深度学习神经网络的研究得到了迅速的发展，并在各个领域取得了重要的突破，国内的研究机构和科技公司在自然语言处理和对话系统方面积极投入研究，并取得了显著成果。他们在以下方面进行了重要的研究工作：

- ① 图像处理和计算机视觉：图像处理和计算机视觉是深度学习的重要应用领域。国内的研究者在图像分类、目标检测和图像分割等任务上进行了大量的研究。例如，利用卷积神经网络

络（CNN）模型，如 ResNet、VGG 和 GoogLeNet，可以实现高精度的图像分类和目标检测。

- ② 语音识别和自然语言处理：语音识别和自然语言处理是深度学习的重要应用领域之一。国内的研究者在声学模型和语言模型方面进行了大量的研究。例如，基于循环神经网络（RNN）、长短时记忆网络（LSTM）和注意力机制的模型，在语音识别和机器翻译等任务上取得了显著的进展。
- ③ 增强学习和自主控制：增强学习和自主控制是深度学习的热门研究领域之一。国内的研究者在强化学习算法和机器人控制方面进行了重要研究。例如，通过深度强化学习算法，可以实现智能机器人的自主导航和决策能力。
- ④ 预训练和微调：通过大规模的预训练和微调，基于变换器的模型在中国的智能助手中得到了广泛应用。这些模型可以实现文字生成、文本分类和情感分析等任务。
- ⑤ 多轮对话系统：国内的研究者开展了多轮对话系统的研究，包括对话生成、对话管理和对话评估等方面。他们致力于提高对话系统的鲁棒性和交互效果。
- ⑥ 情感分析：国内的研究者研究了基于神经网络的情感分析方法，旨在识别和理解文本中的情感信息。他们的研究对于情感分析在社交媒体和舆情监测等领域的应用具有重要意义。

3.3 国外的研究现状

在国际上，许多研究机构和公司也在深度学习神经网络领域做出了重要贡献，在自然语言处理和对话系统领域推动着相关技术的研究进展，并取得了一些突破性成果：

- ① 图像生成和风格迁移：图像生成和风格迁移是深度学习的前沿研究方向之一。国外的研究者提出了一系列创新的生成模型，如生成对抗网络（GAN）和变分自编码器（VAE），可以生成逼真的图像和进行风格转换。
- ② 自动驾驶和智能交通：自动驾驶和智能交通是深度学习的热门应用领域。国外的很多科技公司都在开发利用深度学习技术的自动驾驶系统，通过学习和识别道路标志、车辆和行人等信息来实现自动驾驶。
- ③ 医疗诊断和药物研发：深度学习在医疗诊断和药物研发领域也取得了一些重要的突破。国外的研究者利用深度学习神经网络对医疗影像进行分析和诊断，并在药物研发中应用深度学习模型来预测药效和副作用。
- ④ GPT 系列模型的引起关注：OpenAI 发布的 GPT（Generative Pre-trained Transformer）系列模型在自然语言生成和对话系统方面引起了广泛关注。这些模型通过大规模的预训练和微调，在文本生成和对话生成任务上取得了令人瞩目的表现。
- ⑤ BERT 模型的重大突破：Google 的 BERT（Bidirectional Encoder Representations from Transformers）模型在预训练和微调方面取得了重大突破。该模型采用了双向的 Transformer 网络结构，能够更好地捕捉文本的语义和上下文信息，成为自然语言处理领域的新里程碑。

- ⑥ RoBERTa 模型的优化改进: Facebook 的 RoBERTa[8] (Robustly Optimized BERT approach) 模型对 BERT 模型进行了一系列的优化改进, 包括更大的训练数据规模、更长的训练时间和更多的预训练任务。这些改进使模型在多项自然语言处理任务上取得了更好的性能。
- ⑦ 学术研究机构 and 公司的推动: 除了 OpenAI、Google 和 Facebook 等知名机构之外, 国际上的其他学术研究机构 and 科技公司也在不断推动深度学习神经网络相关技术的研究进展。例如, 斯坦福大学、麻省理工学院、微软研究院等都在该领域具有重要的贡献。

4 系统建模

对于深度学习神经网络在语音识别系统中语义识别的应用, 我们取一个深度学习神经网络在分析文本情感的应用作为例子进行系统建模。

4.1 数据准备

- ① 数据收集: 首先需要收集与问题或任务相关的数据。这可能涉及到从各种来源获取大量的训练数据, 如文本、图像、语音等。
- ② 数据预处理: 对采集到的数据进行预处理是很重要的一步。这包括数据清洗、去除噪声、特征提取和标准化等操作。预处理的目标是使输入数据具有可处理性, 并且能够更好地与神经网络模型相匹配。

4.2 网络结构设计

- ① 输入层: 根据具体任务的特点, 确定输入层的节点数, 以适应输入数据的特征。例如, 在图像分类任务中, 输入层节点数可能对应于图像的像素数。
- ② 隐藏层: 通过多个隐藏层的叠加, 可以将低级特征逐渐转化为高级抽象特征。隐藏层的节点数和层数是设计深度神经网络时需要考虑的重要因素。
- ③ 输出层: 输出层的节点数与任务相关, 例如在分类任务中, 输出层节点数可能对应于类别的数量。
- ④ 激活函数: 激活函数是在每个神经元上引入非线性特性的函数。常用的激活函数有 sigmoid、ReLU、tanh 等, 根据任务选择适当的激活函数。

4.3 参数优化

- ① 损失函数: 损失函数用于衡量模型预测值与真实值之间的差异。针对不同的任务, 可以选择合适的损失函数, 如均方误差 (MSE)、交叉熵损失函数等。
- ② 反向传播算法: 通过反向传播算法, 可以计算模型中各个参数对损失函数的梯度, 然后利用梯度下降方法更新参数, 使得损失函数最小化。
- ③ 优化算法: 常用的优化算法包括随机梯度下降 (SGD)、动量法、Adam 等。这些算法能够帮助神经网络更快地收敛并找到更好的参数配置。

4.4 模型训练和评估

- ① 模型训练：使用准备好的数据集，将其划分为训练集和验证集，通过迭代训练过程，逐渐调整网络权重和偏置，使模型适应训练数据，并能够对未知数据进行泛化预测。
- ② 模型评估：在训练过程中，使用验证集对模型进行评估和调优。可以使用各种指标（如准确率、精确率、召回率、F1 值等）来评估模型的性能。
- ③ 过拟合和欠拟合处理：对于过拟合问题，可以采取正则化方法（如 L1/L2 正则化）或者提前停止训练等方法进行处理。对于欠拟合问题，可以调整网络结构、增加隐藏层或节点数等方式来改进。

5 算法设计

5.1 IMDB数据集

IMDB 数据集是一个常用的电影评论数据集，它来源于互联网电影数据库（Internet Movie Database，简称 IMDb）。这个数据集用于情感分析任务，即判断电影评论中的情感是正面的还是负面的。

IMDB 数据集包含了来自不同用户对电影的大量评论，这些评论被标注为正面或负面。数据集一共包括 50,000 条评论，其中 25,000 条用于训练模型，另外 25,000 条用于测试模型。每个数据集都平均分为正面和负面评论各一半。

在 IMDB 数据集中，每个评论是以原始文本的形式存在的，没有经过预处理。评论的长度不一，有短的几个词的评论，也有长达数百词的评论。评论中可能包含各种语言、拼写错误、缩写、斜体字等。

IMDB 数据集通常用于自然语言处理和情感分析的研究。研究人员可以使用这个数据集来训练模型，通过对新的电影评论进行情感分类，判断评论是正面还是负面的。这对于电影工业和社交媒体等领域来说具有重要意义，因为可以帮助理解用户对电影的评价和观点。

总结来说，IMDB 数据集是一个包含大量电影评论的数据集，用于情感分析任务，其中评论被标注为正面或负面。这个数据集在自然语言处理和情感分析的研究中得到广泛应用。

5.2 神经网络前向后向传播算法

深度神经网络中的误差逆传播算法（backpropagation, BP 算法）是一种用于训练神经网络的反向传播算法。它利用梯度下降法来调整网络参数，使得网络的输出尽可能接近目标输出。

首先，我们需要定义一个损失函数（loss function）。在训练过程中，我们需要将该损失函数最小化，以使网络输出和目标输出的差异最小化。一般情况下，我们会使用均方误差（MSE）或交叉熵（Cross-Entropy）作为损失函数。

接着，利用前向传播算法计算网络的输出值。前向传播算法是指输入数据从输入层到输出层的正向传播过程，整个过程中的每一个神经元都会进行一次乘加运算。

然后，我们需要根据输出误差对每个参数进行更新。这里使用反向传播算法来计算参数的梯度，从而实现参数更新。反向传播算法的核心思想是将误差从输出层一直传递到输入层，并根据每个参数的贡献程度来分配误差值。具体来说，误差的传递由两部分组成：前向传播和反向传播。

在前向传播时，我们通过正向计算来得到输出值；在反向传播时，我们先计算输出误差，再将误差反向传递，并根据每个参数的贡献程度来分配误差值。最后，我们可以使用梯度下降法来调整网络参数，使得网络的输出尽可能接近目标输出。

(1) 前向传播

记 w_{jk}^l 为第 $l-1$ 层第 k 个神经元到第 l 层第 j 个神经元的权重， b_j^l 为第 l 层第 j 个神经元的偏置， a_j^l 为第 l 层第 j 个神经元的激活值， σ 为激活函数，则有迭代公式：

$$a_j^l = \sigma(\sum_k w_{jk}^l a_k^{l-1} + b_j^l)$$

(2) 后向传播

记 a^l 为第 l 层的输出向量， y 为期望的输出向量，系数矩阵为 W ，偏移向量为 b ，使用均方差作为损失函数来度量损失：

$$J(W, b, x, y) = \frac{1}{2} \|a^l - y\|_2^2$$

而 $a^l = \sigma(W^l a^{l-1} + b^l)$ ，则损失函数变为：

$$J(W, b, x, y) = \frac{1}{2} \|\sigma(W^l a^{l-1} + b^l) - y\|_2^2$$

记 $z^l = W^l a^{l-1} + b^l$ ，求 W 和 b 的梯度：

$$\Delta W^l = \frac{\partial J}{\partial W^l} = \frac{\partial J}{\partial z^l} \frac{\partial z^l}{\partial W^l} = (a^l - y) \circ \sigma'(z^l) (a^{l-1})^T$$

$$\Delta b^l = \frac{\partial J}{\partial b^l} = \frac{\partial J}{\partial z^l} \frac{\partial z^l}{\partial b^l} = (a^l - y) \circ \sigma'(z^l)$$

其中 \circ 为 Hadamard 积，若 $A = \{a_{ij}\}$ 和 $B = \{b_{ij}\}$ 是同阶矩阵，则二者的 Hadamard 积为 $A \circ B = \{a_{ij} \times b_{ij}\}$ 。

则 z^l 的梯度为：

$$\Delta z^l = \frac{\partial J}{\partial z^l} = \frac{\partial J}{\partial z^l} \frac{\partial z^l}{\partial b^l} = (a^l - y) \circ \sigma'(z^l)$$

则 W 和 b 的梯度可以写成：

$$\Delta W^l = \Delta z^l (a^{l-1})^T$$

$$\Delta b^l = \Delta z^l$$

而 $z^{l+1} = W^{l+1}a^l + b^{l+1} = W^{l+1}\sigma(z^l) + b^{l+1}$ ，则 z^l 的梯度又等于：

$$\Delta z^l = \frac{\partial J}{\partial z^l} = \frac{\partial J}{\partial z^{l+1}} \frac{\partial z^{l+1}}{\partial z^l} = \Delta z^{l+1} (W^{l+1})^T \circ \sigma'(z^l)$$

即有迭代公式：

$$\Delta z^l = \Delta z^{l+1} (W^{l+1})^T \circ \sigma'(z^l)。$$

5.3 情感分析模型的构建

本文使用 python 的 TensorFlow 和 Keras 建立一个基于 IMDB 电影评论数据集的情感分类模型。

首先进行数据处理，因为电影评论文本长度不一，而我们构建的神经网络需要统一固定长度的向量作为模型的输入，所以对评论文本进行裁剪和填充，并将数据分为训练集和测试集。

然后开始构建我们的神经网络结构，在这里，我们使用的是 Sequential 模型，适用于顺序层堆叠的神经网络结构，添加一层 Embedding 层将单词映射为固定长度的向量表示，将文本数据转换为神经网络可以处理的数值数据。通过学习这些向量的权重，可以捕捉到单词之间的语义关系。接着使用 GlobalAveragePooling1D 层对序列维度进行平均池化操作，用于减少文本序列的维度，并从中抽取总体信息。最后使用 Dense 层进行二分类预测，输出一个 0 到 1 之间的概率值。使用 sigmoid 激活函数是因为它可以将输出限制在 0 到 1 之间，便于进行二分类概率预测。

对于参数的优化，我们选择了 adam 优化器，这样可以结合动量优化和 RMSprop 算法的优点，能够更快地收敛并且适用于大多数深度学习任务。

在模型训练的过程中使用了批量梯度下降的方法，让模型能够充分学习数据的特征，并且在训练过程中应用了验证集进行模型性能的监控，以避免过拟合或欠拟合。

6 数据及实验结果分析

训练过程如图 1 所示。



图1 IMDB 电影评论情感分析训练过程

具体数据如表 1 所示。

表 1 IMDB 电影评论情感分析训练数据

IMDB电影评论：嵌入层+平均池化层+全连接sigmoid										
训练轮次	1	2	3	4	5	6	7	8	9	10
识别率	0.7272	0.7764	0.7992	0.8236	0.8388	0.8485	0.8569	0.8626	0.866	0.8697
损失	0.6653	0.611	0.5522	0.4985	0.4549	0.4205	0.3941	0.3729	0.3562	0.3426

训练出的电影评论情感分析模型在测试集上的准确率和损失随训练的轮次的变化如图 2 所示。

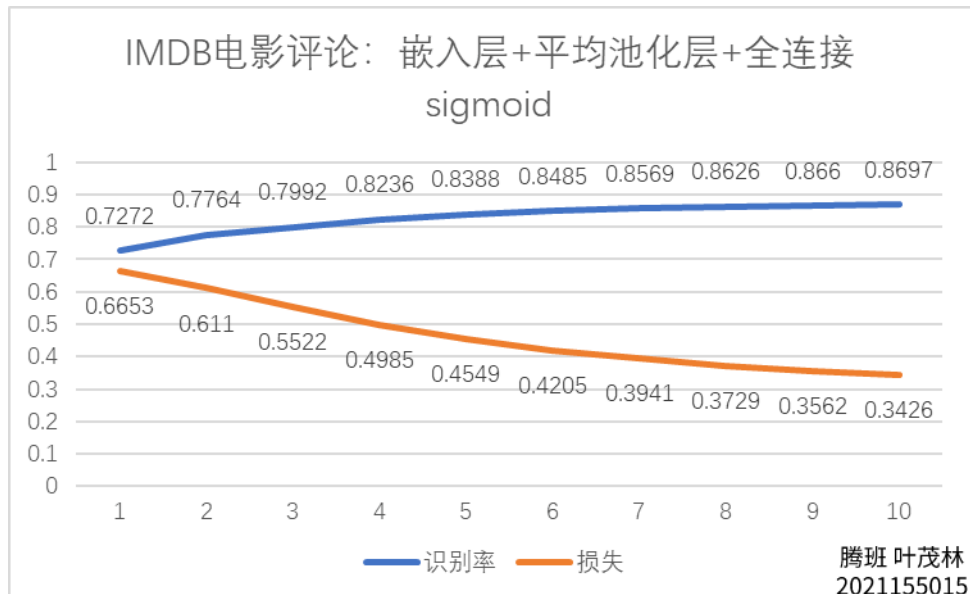


图2 情感分析准确率与损失

由结果可以知道，我们训练出来的电影评论情感分析模型，其数据的拟合效果和测试的泛化效果都比较理想，基本上能够正确的分析出电影评论的情感是正面的还是负面的。

7 总结与展望

近年来,随着人工智能技术的快速发展,语音识别作为人机交互的重要方式之一,得到了广泛应用。而在语音识别领域中,基于神经网络的语义分析扮演着关键角色,利用深度学习中的神经网络模型对语音信号进行处理和分析,以提取其中蕴含的语义信息。

通过将语音信号转化为数字化的特征向量表示,神经网络语义分析能够进行模式识别和语义理解,实现语音指令的准确识别和语义分析。在本次研究中,我们通过构建一个简单的基于深度学习神经网络的电影评论情感分析模型来探索神经网络语义分析的实际应用表现,我们通过分析训练过程和结果表现,有以下几点发现。

首先,数据集的质量和规模对神经网络语义分析的效果有着直接影响。为了提高语音识别的准确性和鲁棒性,需要构建大规模的标注数据集来训练神经网络模型。

其次,神经网络模型的结构和参数配置也对语音识别效果起着重要作用。在小爱同学的研究中,我们通过不断优化和调整模型结构、损失函数、激活函数等关键因素,成功提升了语音识别的性能。

此外,传统的语音识别系统通常由多个组件组成,包括声学特征提取、语音识别、语义分析等。然而,神经网络语义分析技术能够实现端到端的语音识别,简化了系统流程并提高了整体性能。

展望未来,语音识别技术可以在以下几个方面继续拓展:首先是多语种支持。随着人们对跨语言通信需求的增加,多语种语音识别变得越来越重要。因此,我们将进一步研究和开发支持多语种的神经网络语义分析模型,以满足用户的多样化需求。

其次是增强学习的应用。增强学习在语音识别领域具有广阔的应用前景。通过引入增强学习技术,我们可以通过与用户的交互不断优化语音识别模型,提升系统性能和用户体验。

最后是多模态融合。将语音识别技术与其他感知模态如图像、文本等进行融合,将进一步丰富语义分析的能力,提升整体系统的性能和适用范围。

综上所述,神经网络语义分析在语音识别领域的应用前景广阔,通过不断优化模型、扩展语种支持以及与其他感知模态的融合,我们将不断向着更高效准确的语音识别系统迈进。

参考文献:

- [1] 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.
- [2] Collobert, Ronan and Jason Weston. "A unified architecture for natural language processing: deep neural networks with multitask learning." *International Conference on Machine Learning* (2008).
- [3] Sherstinsky, A. (2018). Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) Network. ArXiv, abs/1808.03314.
- [4] Girshick, Ross B.. "Fast R-CNN." 2015 IEEE International Conference on Computer Vision (ICCV) (2015): 1440-1448.
- [5] Ledig, Christian et al. "Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network." 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2016): 105-114.
- [6] Charoenkwan, P., Nantasenamat, C., Hasan, M.M., Manavalan, B., & Shoombuatong, W. (2021). BERT4Bitter: a bidirectional encoder representations from transformers (BERT)-based model for improving the prediction of bitter peptides. *Bioinformatics*.

- [7] Luo, R., Sun, L., Xia, Y., Qin, T., Zhang, S., Poon, H., & Liu, T. (2022). BioGPT: Generative Pre-trained Transformer for Biomedical Text Generation and Mining. *Briefings in bioinformatics*.
- [8] Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., & Stoyanov, V. (2019). RoBERTa: A Robustly Optimized BERT Pretraining Approach. *ArXiv*, abs/1907.11692.