# 1. 选题依据与价值

## 1.1 课题背景

随着自然语言处理(Natural Language Processing, NLP)领域研究的不断深入，学者们的研究已经不只停留在词法解析、句法解析等表层信息的处理上了，让机器理解语义正渐渐成为学者们的研究目标，而自然语言推理作为自然语言理解的重要组成部分，也成为自然语言处理的一项研究热点。

自然语言推理(Natural Language Inference, NLI)通常是指识别文本间的蕴含关系，所以通常也称为文本蕴含关系识别(Recognize Textual Entailment, RTE)。

## 1.2 应用价值

文本间的蕴含关系普遍存在于自然语言文本之间，所以文本蕴含关系识别在许多自然语言处理领域都存在着实际应用价值。本节将列出一些文本蕴含关系识别的应用[1]。

**问答系统** 一个理想的文本蕴含系统可以作为问答系统的答案验证模块，用来判断一个从知识库种抽取的候选答案是否能够推断出目标问题或是目标问题的陈述句形式。例如，某人提问“港珠澳大桥的总工程师是谁？”，通过信息检索技术从知识库中检索出候选答案“林鸣，自2010年12月起，担任港珠澳大桥岛隧工程项目总经理、总工程师”，利用文本蕴含系统可以验证是否可以从候选答案推断出目标问题。为检验文本蕴含系统的这种能力，欧洲跨语言评测平台(Cross Language Evaluation Forum, CLEF)在2006-2008举办了三届答案验证评测(Answer Validation Exercise, AVE)。在评测中，每一条数据由一个问题、一个候选答案和一篇支持文本组成，文本蕴含模型的目标就是在给定支持文本的情景下，针对目标问题，判断候选答案是否正确。

**语义检索** 语义检索的目标是基于检索语句的语义从大量的文档库中检索出目标文档的目标语句。如果将文本蕴含系统运用在语义检索中，它可以用来识别目标问题与目标源文档中各个语句的语义相似性。例如，用户搜索“人们示威反对自由贸易”，现存的基于关键字的搜索引擎将会返回包含“示威”、“自由贸易”等关键字的文档，如果文档中仅存在像“游行人员高举‘反对解除贸易壁垒协议’的标语”这样不含关键字的语句，那将不会被检索到，而造成信息遗漏，利用文本蕴含识别可以避免上述问题。

**自动摘要** 在自动摘要任务中，一个主要的挑战是怎样消除冗余。冗余现象在多文档摘要中特别明显，多文档摘要是融合了多篇源文档的内容，而多篇文档之间的内容可能存在语义相似的，从而造成自动生成的摘要中存在多条语句表达相似语义的问题。文本蕴含识别系统用来判断是否存在语句所表达的语义可以由其他摘要语句推理得到。自动摘要的另一个挑战就是正确性，也就是摘要应该准确地反映单个源文档或多个源文档的内容。文本蕴含识别系统可以通过判断源文档是否可以推理出摘要，以确保摘要的正确性。

**机器翻译的自动评估** 文本蕴含识别的一个相对比较新的应用就是自动评估机器翻译的性能。目前，机器翻译研究者主要利用BLEU作为翻译好坏的指标，BLEU是通过衡量机器翻译语句与人类翻译语句之间的重合程度来，并没有考虑语义层次的相似性。而文本蕴含识别可以判断机器翻译的语句是否可以推断出人类翻译的语句，若可以推断出，则可以认为这是一个好的翻译，即使机器翻译的句子与人类翻译的句子之间没有一个字是一样的。

# 2. 国内外研究现状

## 2.1 文本蕴含识别研究现状

## 2.2 深度学习在自然语言理解领域的研究现状

**神经语言模型(Neural Language Model)** 深度学习在自然语言处理领域的第一次影响深远的运用是利用神经网路构建语言模型。语言模型的任务是在已知一句话的前几个词情况下，来预测下一个词。语言模型在自然语言处理领域具有非常重要的地位，在一些简单的自然语言处理实践中被广泛应用，例如：智能拼写、拼写纠错，邮件回复建议。传统的语言模型主要是基于n元语法，并利用数据平滑算法对未登录词进行处理[2]。2001年 Bengio 等人[3]第一次提出利用前馈神经网络构建语言模型，模型结构如图1所示：

图1：神经语言模型(Bengio 等人于2001年发表)

该模型将一段文本的前n个词在词表中的索引作为输入，然后在词表C中查找索引所对应的词向量，这种技术现在输出为一个与词表长度一样长的向量，向量的每一位代表对应词是第n+1个词的概率。之后，Mikolov 等人[4]尝试利用循环神经网代替全连接神经网为语言模型建模，Graves 等人在此基础上又将循环神经网替换为长短记忆神经网（Long short-term memory neural network，LSTM）[5]。直到现在，虽然神经语言模型一直在发展，但是基于LSTM的建模方式依然具有很强的竞争力[6]。

神经语言模型虽然功能很简单，但是对后续的很多自然语言处理领域的发展具有深远意义。在词向量研究方面，词向量的训练过程就是一个语言模型的简化。在序列到序列模型中，模型的输出是每个时刻预测一个词，最后形成一串序列，这样的形式也是受到语言模型的启发。在之后的预训练语言模型中，这些预训练的模型的预训练过程都是在训练一个语言模型，以求能够从中理解语义，然后进行迁移学习。

**多任务学习**

**词嵌入(Word Embedding)**

  自然语言处理领域在过去很长时间里都是应用词袋模型，利用稀疏向量对文本进行表示。2001年，学者们开始利用稠密向量表示词。2013年，Mikolov 等人[7][8]提出两种利用神经网络训练词向量的模型，分别是CBOW和Skip-gram，这两种模型可以更加有效地训练词向量。CBOW模型的基本思想是通过中心词的上下文预测中心词，而Skip-gram模型是通过中心词来预测上下文。两个模型的结构如图2所示：

图1：神经语言模型(Bengio 等人于2001年发表)

  集成这两种模型的开源工具word2vec在很多自然语言处理应用中取得很好的效果。利用word2vec工具，Mikolov 等人[7][8]在大数据集上训练后，发现词向量之间存在着一定的语义关系，例如：词向量"king"减"man"加"woman"可以得到词向量"queue"。   2014年，Pennington 等人[9]提出Glove模型，利用矩阵分解的方法训练词向量。相比于word2vec中集成的两个模型，Glove模型引入了全局信息，通过全局矩阵分解和局部的上下文窗口两种方式来提升词向量对语义的表达能力。同时，通过矩阵分解训练的词向量[9][10][11]也可以获得类似word2vec训练出的词向量所具有的语义关系。

**神经网络在自然语言处理中的应用**

循环神经网络(Recurrent Neural Network)、卷积神经网络(Convolution Neural Netword)及递归神经网络(Recursive Neural Network)

**序列到序列(Sequence to Sequence) 模型框架**

  在自然语言处理领域，有很多序列到序列的任务，例如：机器翻译、语音识别、摘要自动生成等任务。2014年， Sutskever 等人[12]首次提出序列到序列的模型，这是一个将一个序列映射到另一个序列的通用模型框架，模型结构如图所示：

图：序列到序列模型(Sutskever 等人于2014年发表)

  序列到序列模型由一个编码器和解码器构成，编码器将输入序列压缩成一个定长的稠密向量，解码器将编码器输出的稠密向量解码成任务所需要的输出序列。编码器和解码器的网络结构主要是基于循环神经网络，在此基础上也发展出一些变体。目前，序列到序列模型的编码器网络结构有长短记忆神经网络(LSTM)[13]、卷积神经网络(CNN)[14][15]、Transfomer[16]以及一些混合结构[17]，解码器部分的网络结构与编码器类似，不过鲜有论文采用卷积神经网络结构。

**注意力机制(Attention)**

  注意力机制(Attention Mechanism)是一个能够有效改善模型效果的机制，其在计算机视觉及自然语言处理领域均有广泛应用。2015年，Bahdanau 等人[18]首次应用在机器翻译中Attention机制，这个关键创新使得基于神经网络的机器翻译在性能上远远超过传统模型。

**记忆网络(Memory-based Networks)**

**预训练模型**

# 3. 相关理论和技术

## 循环神经网络(Recurrent Neural Network, RNN)

### 长短记忆神经网络(Long Short-term Memory, LSTM)

### 双向长短记忆神经网络(Bidirectional LSTM, BiLSTM)

## Attention机制

## 神经网络训练优化算法

## ELMo词嵌入模型

  ELMo全称Embeddings from Language Models，是一种词嵌入模型。预训练文本表示在自然语言理解领域具有重要地位，其中Word2vec[7][8]和Glove[9]对预训练词表示起到了关键作用。但是Word2vec和Glove都只是将一个词用一个向量表示，而词在不同的语境中会有不同的语义，而用一个固定的向量来表示一个词则无法体现词的多义性。

## Transformer模型

# 研究内容与实施方案

## 研究内容

### 课题描述

### 数据集介绍

#### SNLI数据集

#### MultiNLI数据集

  MultiNLI (Multi-Genre Natural Language Inference) 数据集[19]是目前最大的文本蕴含数据集，是Stanford所推出的自然语言推理数据集SNLI的升级版本。MultiNLI数据集总共有433000组句子样本，包含10种不同的领域，风格上包括文本和口语两种英文风格。

## 实施方案

  鉴于Chen 等人所建立的模型ESIM[20]在文本蕴含数据集SNLI与MutiNLI上的成功，本文将选择ESIM模型作为基准模型，并希望能够对模型的并行计算能力与数据性能指标进行改善。下文将对ESIM模型的主要部分进行描述并提出改进计划。

### ESIM模型

  ESIM模型全称Enhanced Sequential Inference Model，曾在snli数据集上取得最高的性能指标，并被选为MultiNLI数据集的官方基准模型，具有很高的参考价值。ESIM模型主要由三个部分组成：输入编码模块(Input Encoding), 局部推理模块(Local Inference Modeling)和组合推理模块(Inference Composition)。

#### 输入编码模块(Input Encoding)

  文本蕴含问题是判断前提(premise)与假设(hypothesis)这两句话之间是否存在语义的蕴含关系。所以，模型首先需要对两个文本序列进行处理。本文规定 与 分别表示前提序列和假设序列，其中 分别是一个维度为n的向量，分别表示前提序列的第i个词与假设序列的第j个词，这些词向量一般通过预训练的词向量进行初始化。然后模型利用双向长短记忆网络分别对前提序列和假设序列进行编码，编码过程可由如下公式表示：

  其中 分别是一个h维的向量，分别是前提序列的第i个词向量和假设序列的第j个词向量在经过双向长短记忆网络后输出的隐状态向量。双向长短记忆神经网络每个时刻输出的隐状态向量不仅具有序列在该时刻词的信息，并且还包含该时刻词的上下文信息。

#### 局部推理模块(Local Inference Modeling)

  对输入序列进行编码之后，前提序列与假设序列分别由 和 表示，每个被编码的序列元素包含着对应时刻词的信息和其上下文的信息，但两个序列之间的联系并没有描述，模型需要对两个序列的关系进行建模描述。局部推理模块利用软注意力(soft attention)机制对两个被编码的序列的各个元素之间的关系进行建模。为了判断假设序列中的每个词是否可以从前提序列中推理出来，软注意力机制首先计算两个序列各个元素之间的相似度，然后通过softmax操作得到一个权重向量，再对前提序列中的每个元素进行加权求和，得到前提序列针对假设序列上下文向量。同理可得，假设序列针对前提序列的上下文向量。上述过程可由如下公式表示：

  其中表示前提序列的第i个元素与假设序列的第j个元素相关程度，计算元素之间关联程度的方式有很多，本文通过最简单的方式，计算两序列中各个词向量之间的内积来衡量两词之间关联程度，公式如下：

  实验表明[20]，其他一些更加复杂的衡量词之间关联程度的方式并没有改善实验性能。对于两个序列的各个元素之间的相似度计算可利用张量(Tensor)运算进行并行计算。   然后模型对序列的各方面的信息进行融合。针对前提序列，本模型对编码后的矩阵 和假设序列的上下文矩阵 做了一些操作后，之后进行合并，过程如下公式所示：

  其中表示矩阵中对应位置的元素相乘。同理可得，假设序列的信息融合过程如下公式所示：

#### 组合推理模块(Inference Composition)   局部推理信息形成之后，模型需要对局部推理信息进行进一步的组合，使得每个时刻的元素都包含上下文的信息。本模型

  最后的推理模块利用池化与合并操作将处理后的两条序列信息转化为一个固定长度的向量，并且通过分类器对两序列的蕴含关系进行分类。池化操作主要是将平均池化和最大池化相结合，然后将池化后的向量进行合并操作，形成一个长度固定的向量。向量的计算过程如下：

  最后，模型分类器利用多层感知机(Multi-Layer Perceptron)对句子蕴含关系进行分类。本模型的多层感知机只有一个隐藏层，激活函数是tanh，最后输出层利用softmax进行数据归一化。分类的计算过程如下：

  其中和分别是全连接层的权重和偏置，输出向量中的每一个元素表示相对应蕴含关系的概率。

### 实验细节

### 部分实验结果

### 改进计划

#### 文本表示

#### 序列编码

#### 序列交互与特征抽取

# 参考文献

[1] MACCARTNEY B. Natural Language Inference[D]. 2009: 179.

[2] KNESER R, NEY H. Improved Backing-off for N-gram Language Modeling[C]//International Conference on Acoustics, Speech, and Signal Processing. 1995: 181–184 vol.1.

[3] BENGIO Y, DUCHARME R, VINCENT P. A Neural Probabilistic Language Model[C]//2001.

[4] INTERSPEECH 2010, 11th Annual Conference of the International Speech Communication Association, Makuhari, Chiba, Japan, September 26-30, 2010[M]. KOBAYASHI T, HIROSE K, NAKAMURA S. ISCA, 2010.

[5] GRAVES A. Generating Sequences With Recurrent Neural Networks[J]. CoRR, 2013, abs/1308.0850.

[6] MELIS G, DYER C, BLUNSOM P. On the State of the Art of Evaluation in Neural Language Models[J]. CoRR, 2017, abs/1707.05589.

[7] MIKOLOV T, CHEN K, CORRADO G, 等. Efficient Estimation of Word Representations in Vector Space[J]. 2013: 1–12.

[8] MIKOLOV T, SUTSKEVER I, CHEN K, 等. Distributed Representations of Words and Phrases and their Compositionality[J]. 2013: 1–9.

[9] PENNINGTON J, SOCHER R, MANNING C D. GloVe: Global Vectors for Word Representation[J]. 2014.

[10] LEVY O, GOLDBERG Y. Neural Word Embedding as Implicit Matrix Factorization[C]//Advances in Neural Information Processing Systems 27: Annual Conference on Neural Information Processing Systems 2014, December 8-13 2014, Montreal, Quebec, Canada. 2014: 2177–2185.

[11] LEVY O, GOLDBERG Y, DAGAN I. Improving Distributional Similarity with Lessons Learned from Word Embeddings[J]. TACL, 2015, 3: 211–225.

[12] SUTSKEVER I, VINYALS O, LE Q V. Sequence to Sequence Learning with Neural Networks[J]. CoRR, 2014, abs/1409.3215.

[13] WU Y, SCHUSTER M, CHEN Z, 等. Google’s Neural Machine Translation System: Bridging the Gap between Human and Machine Translation[J]. CoRR, 2016, abs/1609.08144.

[14] KALCHBRENNER N, ESPEHOLT L, SIMONYAN K, 等. Neural Machine Translation in Linear Time[J]. 2016.

[15] GEHRING J, AULI M, GRANGIER D, 等. Convolutional Sequence to Sequence Learning[J]. 2017.

[16] VASWANI A, SHAZEER N, PARMAR N, 等. Attention Is All You Need[J]. 2017.

[17] CHEN M X, FIRAT O, BAPNA A, 等. The Best of Both Worlds: Combining Recent Advances in Neural Machine Translation[J]. 2018.

[18] BAHDANAU D, CHO K, BENGIO Y. Neural Machine Translation by Jointly Learning to Align and Translate[J]. CoRR, 2014, abs/1409.0473.

[19] WILLIAMS A, NANGIA N, BOWMAN S. A Broad-Coverage Challenge Corpus for Sentence Understanding through Inference[C]//Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers). New Orleans, Louisiana: Association for Computational Linguistics, 2018: 1112–1122.

[20] CHEN Q, ZHU X, LING Z, 等. Enhanced LSTM for Natural Language Inference[J]. Research in Immunology, 2016, 141(4): 431–440.