

# nlp基础概念调研

## 基本概念



知乎 <https://zhuanlan.zhihu.com/p/634689142>

### 什么是自然语言处理？看这篇文章就够了！

1、概述自然语言处理（Natural Language Processing, NLP）是计算机科学、人工智能和语言学领域的一个交叉学科，主要研究如何让计算机能够理解、处理、生成和...

n-gram模型、隐马尔可夫模型（HMM）和条件随机场（CRF）

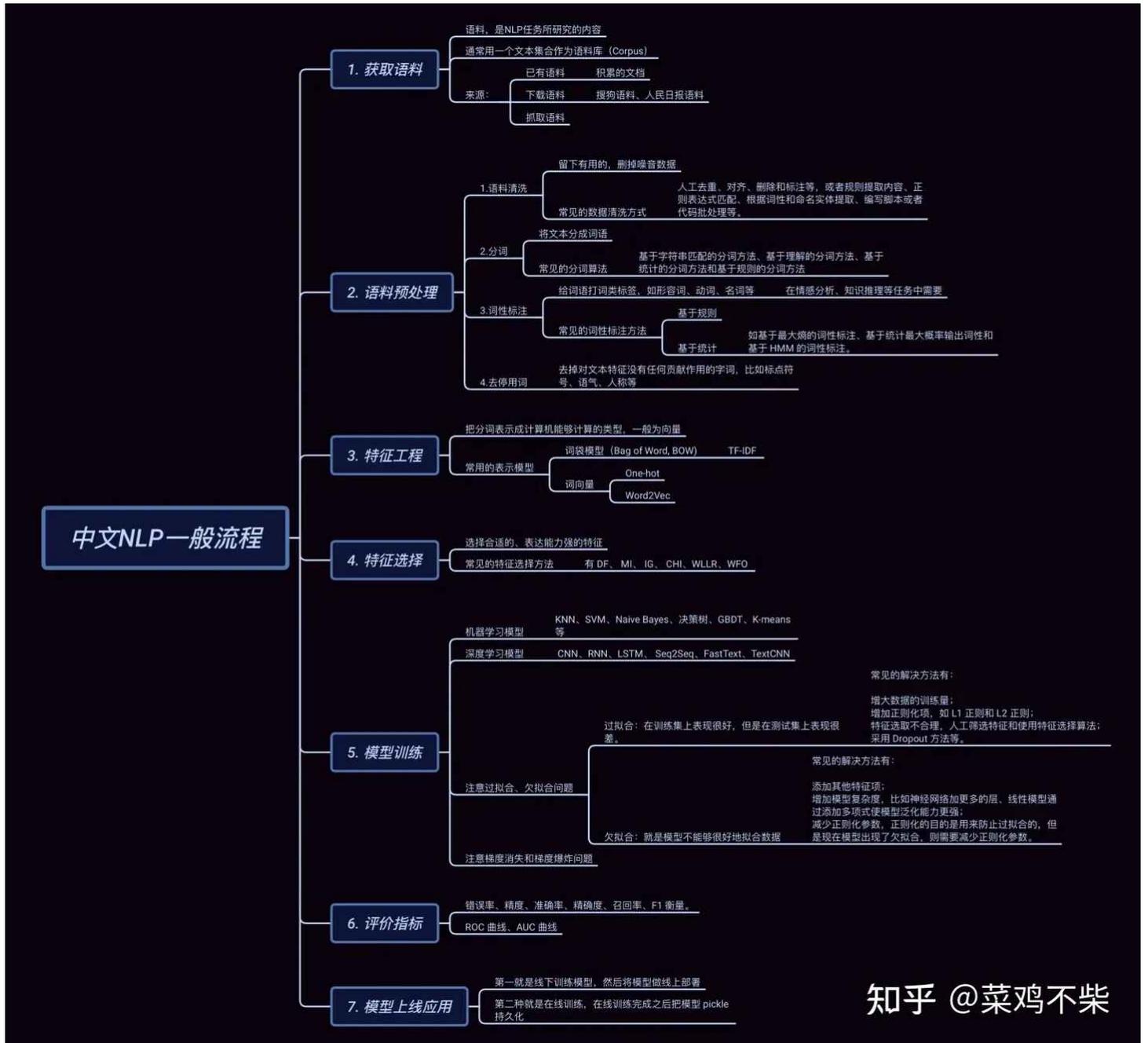
## 自然语言处理的发展阶段：

- a. 规则驱动阶段（1950s - 1990s）：最初的NLP方法主要基于手工编写的规则，试图通过语法和语义规则来处理自然语言。这种方法在处理复杂的语言现象时面临挑战。
- b. 统计方法阶段（1990s - 2010s）：随着计算能力的提升，统计方法开始流行起来。这包括基于概率和统计模型的方法，如隐马尔可夫模型（HMM）和最大熵模型。这一时期也见证了基于大规模语料库的机器学习方法的崛起。
- c. 深度学习阶段（2010s 至今）：随着深度学习技术的兴起，尤其是循环神经网络（RNN）和长短时记忆网络（LSTM）等模型的应用，NLP取得了显著的进展。深度学习模型在各种任务中取得了突破，如语言建模、词嵌入、机器翻译和情感分析。

## 自然语言处理的主要分类：

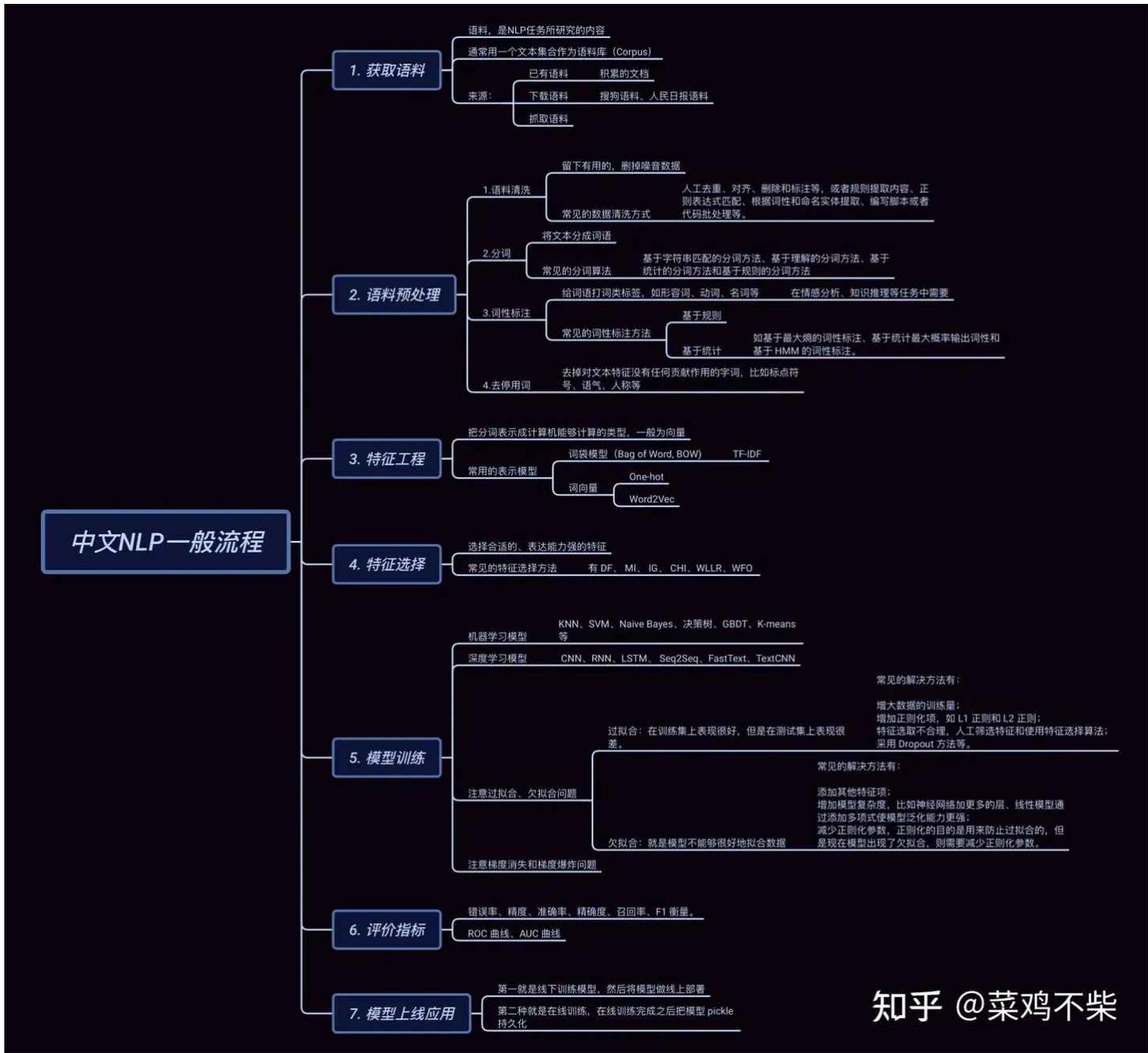
- a. 语言建模（Language Modeling）：通过建立概率模型来预测一段文本中下一个词的可能性，是许多NLP任务的基础，如机器翻译和语音识别。
- b. 信息检索（Information Retrieval）：这包括搜索引擎和文档检索系统，致力于从大量文本中找到相关的信息。
- c. 命名实体识别（Named Entity Recognition, NER）：识别文本中具有特定意义的命名实体，如人名、地名和组织名。
- d. 机器翻译（Machine Translation）：使用计算机自动将一种语言翻译成另一种语言。
- e. 情感分析（Sentiment Analysis）：分析文本中的情感极性，通常包括正面、负面或中性情感。
- f. 问答系统（Question Answering）：设计能够理解并回答用户提出的问题的系统，涉及语义理解和信息检索。
- g. 语音识别（Speech Recognition）：将语音信号转换为文本形式，以便计算机理解和处理。

- h. 文本生成 (Text Generation) : 利用模型生成自然语言文本, 如聊天机器人和摘要生成。
- i. 语义分析 (Semantic Parsing) : 将自然语言文本映射到形式化的表示, 以便计算机理解其含义。
- j. 语言生成 (Natural Language Generation, NLG) : 创建计算机生成自然语言文本的系统, 通常涉及到将结构化数据转换为自然语言。



知乎 @菜鸡不柴

一般流程



知乎 @菜鸡不柴

## 语音识别:

<https://zhuanlan.zhihu.com/p/635372844>  
 zhuanlan.zhihu.com

## LLM大模型:

<https://zhuanlan.zhihu.com/p/646385704>  
 zhuanlan.zhihu.com

## 其他文档链接：

<https://blog.csdn.net/u014636245/article/details/134451109>

### 【AI视野·今日NLP 自然语言处理论文速览 第六十四期】 Fri, 27 Oct 2023-CSDN博客

文章浏览阅读565次。AI视野·今日CS.NLP 自然语言处理论文速览Fri, 27 Oct 2023Totally 80 papers👉上期速览👈更多精彩请移步主页Daily Computation and Language Paperstorchdistill Meets Hugging Face Libraries for Reproducible, Coding-Free Dee...

## chatgpt给的链接：

### a. Hugging Face Transformers:

- GitHub 链接: <https://github.com/huggingface/transformers>
- Hugging Face Transformers 提供了许多预训练的自然语言处理模型，包括用于情感分析的模型，如BERT、GPT等。

### b. TextBlob:

- GitHub 链接: <https://github.com/sloria/TextBlob>
- TextBlob 是一个用于处理自然语言文本的简单库，提供了易用的 API 和一些基本的情感分析功能。

### c. DeepMoji:

- GitHub 链接: <https://github.com/bfelbo/DeepMoji>
- DeepMoji 是一个基于深度学习的情感分析工具，使用卷积神经网络（CNN）对表情符号进行训练。

### d. BERT:

- GitHub 链接: <https://github.com/google-research/bert>
- BERT (Bidirectional Encoder Representations from Transformers) 是一个预训练的 Transformer 模型，可用于各种自然语言处理任务，包括情感分析。

### e. Sentiment140:

- GitHub 链接: <https://github.com/kaz3w/sentiment140>
- Sentiment140 包含一个用于情感分析的数据集，该数据集包含从 Twitter 提取的带有情感标签的推文。

### f. Sentiment Analysis in Python:

- GitHub 链接: <https://github.com/AdalbertoCq/sentiment-analysis-in-python>
- 这个仓库包含用 Python 进行情感分析的示例代码和教程。

请注意，GitHub 上还有许多其他优秀的仓库和项目，您可以根据自己的需求在 GitHub 上搜索并找到适合您的工具和代码。