



人工智能：自然语言处理技术 (选学)

饶洋辉
计算机学院,
中山大学

raoyangh@mail.sysu.edu.cn

<http://cse.sysu.edu.cn/node/2471>

课件来源：中国科学院大学宗成庆教授、中国人民大学赵鑫教授、
伊利诺伊大学香槟分校韩家炜教授、台湾大学李宏毅副教授等

基本概念

- 定义1-1：语言和语言学 (linguistics)
 - 语言：人类所特有的用来表达意思、交流思想的工具，是一种特殊的社会现象。自然/机器
 - 语言学是指对语言的科学研究。
 - 戴维·克里斯特尔，《现代语言学词典》，1997
 - 研究语言的本质、结构和发展规律的科学。
 - 商务印书馆，《现代汉语词典》，1996
 - 语音和文字是语言的两个基本属性。

基本概念

- 定义1-2: 语音学 (phonetics)
 - 研究人类发音特点，特别是语音发音特点，并提出各种语音描述、分类和转写方法的科学。
 - 包括: (1)发音语音学(articulatory phonetics), 研究发音器官如何产生语音; (2)声学语音学(acoustic phonetics), 研究口耳之间传递语音的物理属性; (3)听觉语音学 (auditory phonetics), 研究人通过耳、听觉神经和大脑对语音的知觉反应。
—戴维•克里斯特尔, 《现代语言学词典》, 1997

基本概念

- 定义1-3: 计算语言学 (Computational Linguistics)
 - 通过建立形式化的计算模型来分析、理解和生成自然语言的学科，是人工智能和语言学的分支学科。计算语言学是典型的交叉学科，其研究常常涉及计算机科学、语言学、数学等多个学科的知识。与内容接近的学科自然语言处理相比较，计算语言学更加侧重基础理论和方法的研究。

《计算机科学技术百科全书》(常宝宝)

基本概念

- 定义1-4：自然语言理解 (Natural Language Understanding, NLU)
 - 自然语言理解是探索人类自身语言能力和语言思维活动的本质，研究模仿人类语言认知过程的自然语言处理方法和实现技术的一门学科。它是人工智能早期研究的领域之一，是一门在语言学、计算机科学、认知科学、信息论和数学等多学科基础上形成的交叉学科。

《计算机科学技术百科全书》(宗成庆)

基本概念

- 关于“理解”的标准
 - 如何判断计算机系统的智能？
 - 计算机系统的表现 (act) 如何？
 - 计算机系统的反应 (react) 如何？
 - 计算机系统的相互作用 (interact) 如何？



与有意识的个体（人）比较如何？

图灵设计的“模仿游戏” — 图灵测试 (Turing test)

基本概念

- 定义1-5: 自然语言处理 (Natural Language Processing, NLP)
 - 自然语言处理是研究如何利用计算机技术对语言文本（句子、篇章或话语等）进行处理和加工的一门学科，研究内容包括对词法、句法、语义和语用等信息的识别、分类、提取、转换和生成等各种处理方法和实现技术。

《计算机科学技术百科全书》(宗成庆)

基本概念

- 三个不同的语系
 - 屈折语 (fusional language / inflectional language): 用词的形态变化表示语法关系, 如英语、法语等。
 - 黏着语 (agglutinative language): 词内有专门表示语法意义的附加成分, 词根或词干与附加成分的结合不紧密, 如日语、韩语、土耳其语等。
 - 孤立语 (analytic language) (分析语, isolating language): 形态变化少, 语法关系靠词序和虚词表示, 如汉语。

基本概念

汉语：汉族的语言，是我国的主要语言。

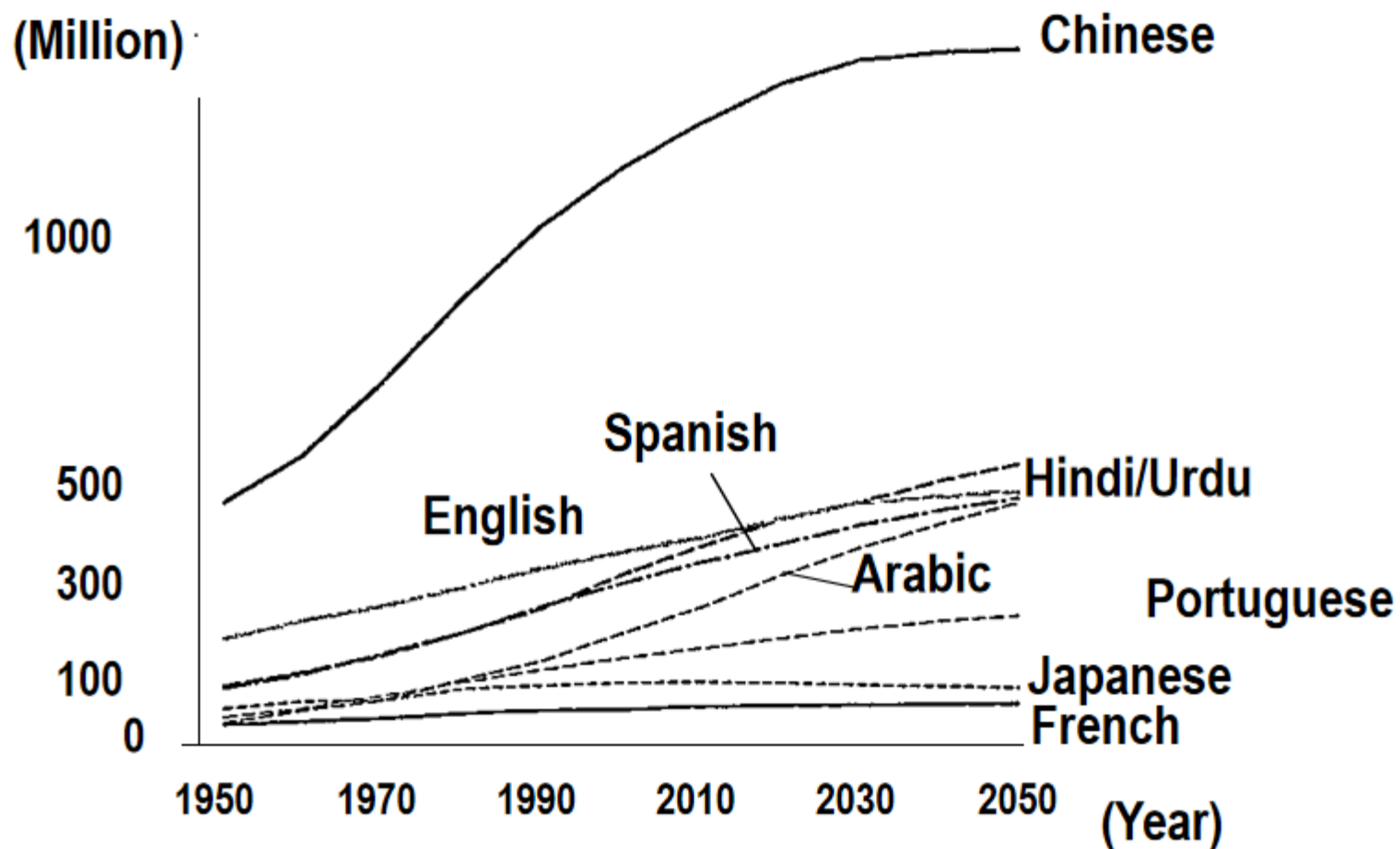
中文：中国的语言文字，特指汉族的语言文字。

— 《现代汉语词典》，1996

- 定义1-6：中文信息处理 (Chinese Information Processing)

针对中文的自然语言处理技术。

基本概念



基本概念

汉语已经不再只是中国人自己使用和关注的语言，无论外国人喜欢与否，但都没有人敢藐视它！针对汉语的处理技术早已经成为国际学术界和企业界共同关注的问题，汉英两大强势语言的自动翻译问题则是人类语言技术中最具挑战性的研究课题。

基本概念

近几年来，自然语言处理技术迅速发展成为一门相对独立的学科，倍受关注，而且该技术不断与语音技术相互渗透和结合形成新的研究分支，因此，很多人在谈到“计算语言学”、“自然语言处理”或“自然语言理解”这些术语时，往往默认为同一个概念。甚至有专著[刘颖，2002]干脆直接解释为：

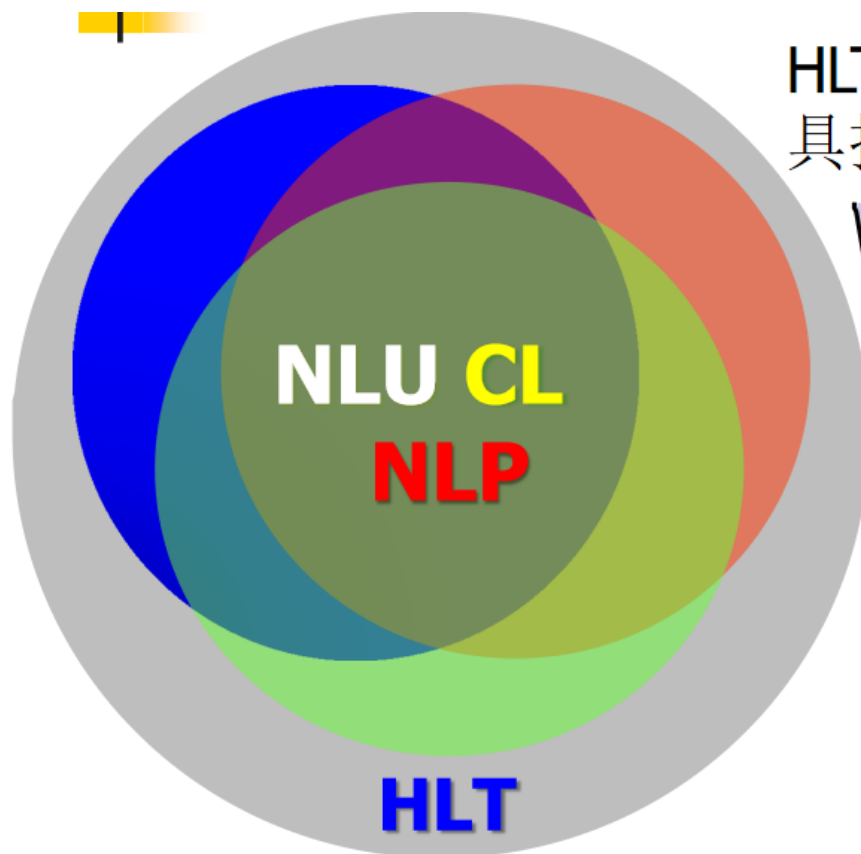
计算语言学也称自然语言处理或自然语言理解

人类语言技术
(Human Language Technology, HLT)

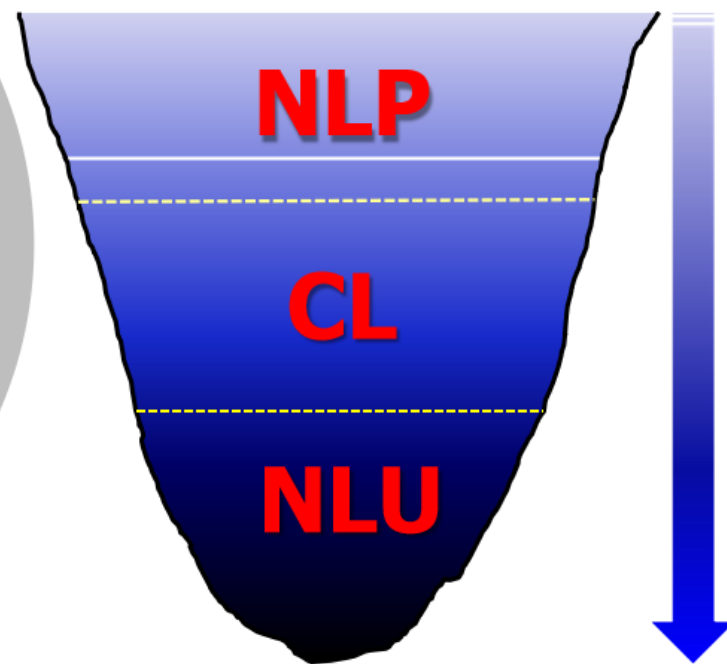
基本概念

- **自然语言理解** (Natural Language Understanding, NLU)是人工智能最重要的研究方向之一，是当今“**人工智能皇冠上的明珠**”。
- **计算语言学** (Computational Linguistics, CL)
 - 1960S，形成相对独立的学科。1962年国际计算语言学学会(ACL)成立，1965年国际计算语言学委员会(ICCL)成立，1966年“计算语言学”首次出现在美国国家科学院ALPAC报告里。
- **自然语言处理** (Natural Language Processing, NLP)
 - 1980S，面向计算机网络和移动通信，从系统实现和语言工程的角度开展语言信息处理方法的研究。专门针对中文的语言信息技术研究成为中文信息处理。

基本概念

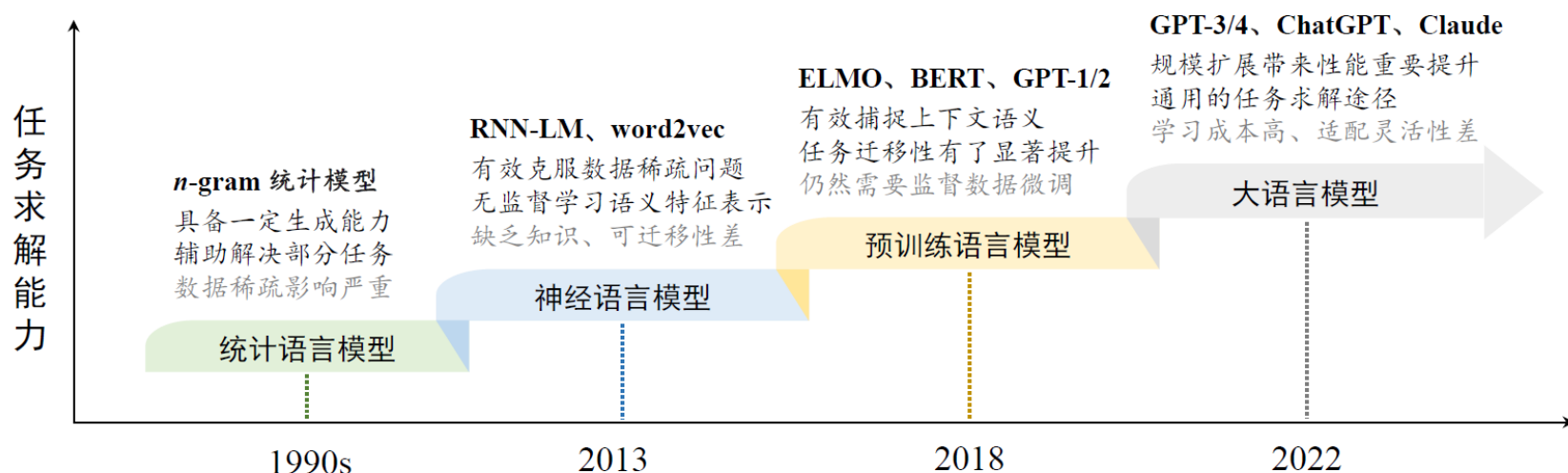


HLT 是当前人工智能领域最具挑战性的研究方向之一。



语言模型

- 语言模型 (Language Model, LM) 旨在对人类语言的内在规律进行建模，从而准确预测词序列中未来（或缺失）词或词元 (Token) 的概率。
- 语言模型是提升机器语言智能的主要技术途径之一。
- 基于任务求解能力的四代语言模型的演化过程：



Wayne Xin Zhao et al. A Survey of Large Language Models. arXiv: 2303.18223 (2023).

统计语言模型

统计语言模型：自然语言处理的基础

- 句子一：“美联储主席昨天告诉媒体7000亿美元的救助资金将借给上百家银行、保险公司”
- 句子二：“美联储主席昨天7000亿美元的救助资金告诉媒体将借给上百家银行、保险公司”
- 句子三：“联主美储席将保险借上媒百诉给银行家昨天亿美元的救资金告体、7000司公”

——吴军《数学之美》

统计语言模型

- **Unigram LM** based on only **single words** (unigrams), considers **no context**, and assumes **independent generation** of words

$$P(\langle t_1, \dots, t_m \rangle) = \prod_{i=1}^m P(t_i)$$

- **Bigram LM** conditions on the preceding term

$$P(\langle t_1, \dots, t_m \rangle) = P(t_1) \prod_{i=2}^m P(t_i | t_{i-1})$$

- ***n*-Gram LM** conditions on the preceding (*n*-1) terms

$$P(\langle t_1, \dots, t_m \rangle) = P(t_1) P(t_2 | t_1) \dots \prod_{i=n}^m P(t_i | t_{i-n+1} \dots t_{i-1})$$

神经语言模型

- 神经词嵌入(word embedding): 将单词表示为数值向量
 - 利用输入语料中邻近的词进行训练
 - 包括Word2Vec (T. Mikolov et al., 2013)等代表方法
- 给定充足的数据、用法与上下文, Word2Vec等神经词嵌入模型可以根据词语在过去出现的情况, 对词语的含义进行高精度的估计。
- 应用: 分析文本、基因、代码、喜好、符号序列等
- 自Word2Vec以来, 众多强大的嵌入方法逐渐出现
 - GloVe (Global Vectors for Word Representation) *J. Pennington, R. Socher, C. D. Manning (Stanford U.) EMNLP 2014*
 - fastText: 将每个词表示为字母的n-gram, 以对前缀和后缀进行理解
 - 网络嵌入 (Network embedding) : 在信息网络中得到每个节点的嵌入

神经语言模型

- Word2Vec和Glove等方法学习到的是**上下文无关的 (context-free)** 词嵌入：
 - 词与词嵌入是一一对应的关系，忽略了词在不同特定语境下的不同含义；
 - 例如，“bank”是一个**多义词**，但它只有唯一的嵌入表示。



- 通过学习词与句子的上下文关系，BERT等预训练语言模型能够克服这个问题，它们可以得到**上下文 (contextualized)** 词嵌入。

预训练语言模型

- 常见的预训练语言模型 (PLM) :

PTMs	Architecture [†]	Input	Pre-Training Task	Corpus	Params
ELMo [14]	LSTM	Text	BiLM	WikiText-103	
GPT [15]	Transformer Dec.	Text	LM	BookCorpus	117M
GPT-2 [58]	Transformer Dec.	Text	LM	WebText	117M ~ 1542M
BERT [16]	Transformer Enc.	Text	MLM & NSP	WikiEn+BookCorpus	110M ~ 340M
InfoWord [55]	Transformer Enc.	Text	DIM+MLM	WikiEn+BookCorpus	=BERT
RoBERTa [43]	Transformer Enc.	Text	MLM	BookCorpus+CC-News+OpenWebText+ STORIES	355M
XLNet [49]	Two-Stream Transformer Enc.	Text	PLM	WikiEn+ BookCorpus+Giga5 +ClueWeb+Common Crawl	≈BERT
ELECTRA [56]	Transformer Enc.	Text	RTD+MLM	same to XLNet	335M
UniLM [44]	Transformer Enc.	Text	MLM+ NSP	WikiEn+BookCorpus	340M
MASS [41]	Transformer	Text	Seq2Seq MLM	*Task-dependent	
BART [50]	Transformer	Text	DAE	same to RoBERTa	110% of BERT
T5 [42]	Transformer	Text	Seq2Seq MLM	Colossal Clean Crawled Corpus (C4)	220M ~ 11B
ERNIE(THU) [76]	Transformer Enc.	Text+Entities	MLM+NSP+dEA	WikiEn + Wikidata	114M
KnowBERT [77]	Transformer Enc.	Text	MLM+NSP+EL	WikiEn + WordNet/Wiki	253M ~ 523M
K-BERT [78]	Transformer Enc.	Text+Triples	MLM+NSP	WikiZh + WebtextZh + CN-DBpedia + HowNet + MedicalKG	=BERT
KEPLER [80]	Transformer Enc.	Text	MLM+KE	WikiEn + Wikidata/WordNet	
WKLM [57]	Transformer Enc.	Text	MLM+ERD	WikiEn + Wikidata	=BERT
CoLAKE [81]	Transformer Enc.	Text+Triples	MLM	WikiEn + Wikidata	=RoBERTa

- Transformer, GPT和BERT的架构:

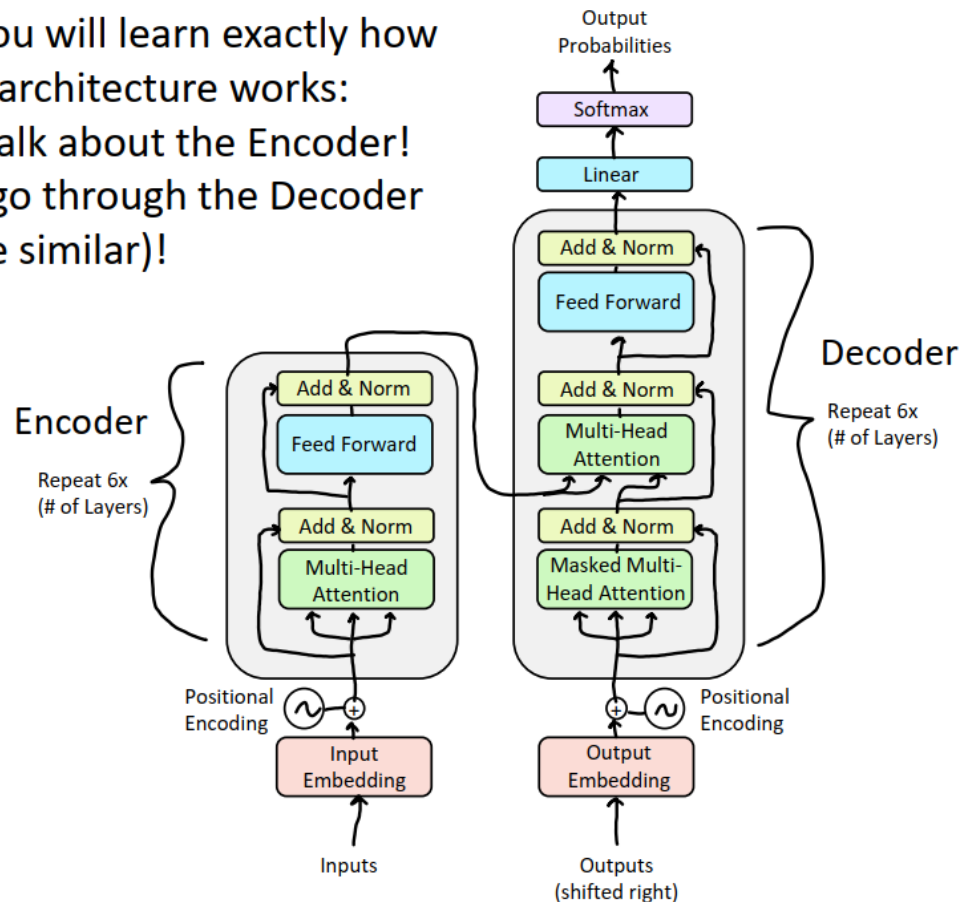


Transformer

The Transformer Encoder-Decoder [Vaswani et al., 2017]

In this section, you will learn exactly how the Transformer architecture works:

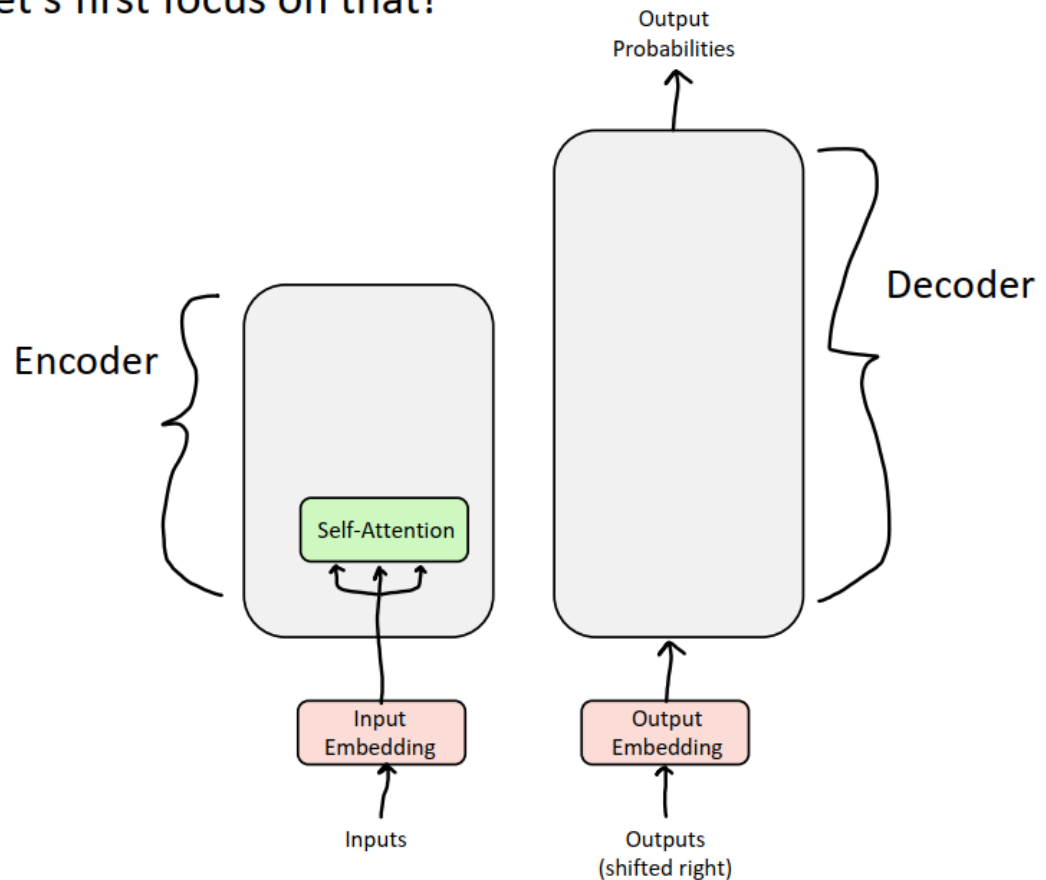
- First, we will talk about the Encoder!
- Next, we will go through the Decoder (which is quite similar)!



Transformer

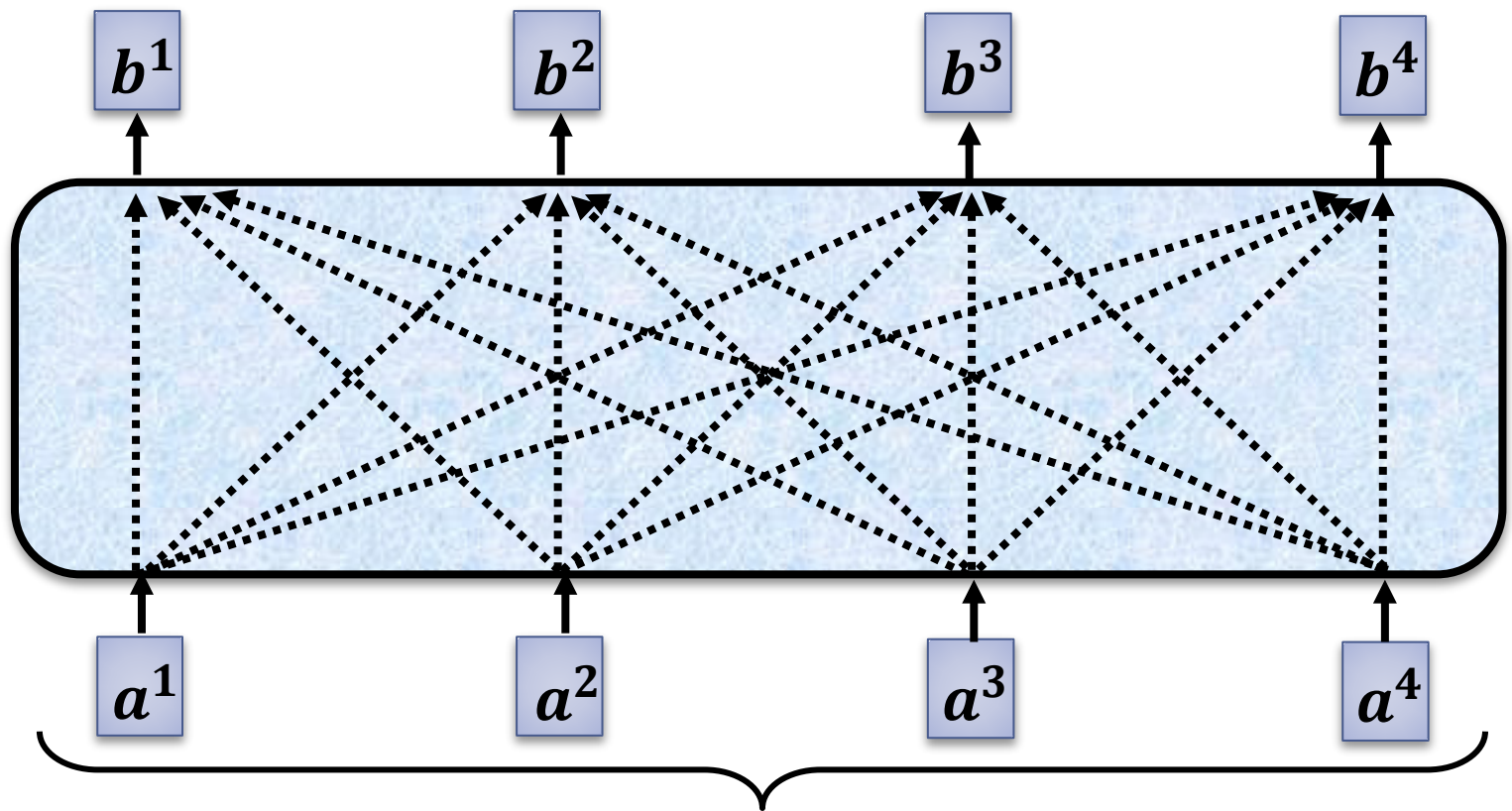
- Encoder: Self-Attention (<https://b23.tv/pTSqhws>)

Self-Attention is the core building block of Transformer, so let's first focus on that!



Transformer

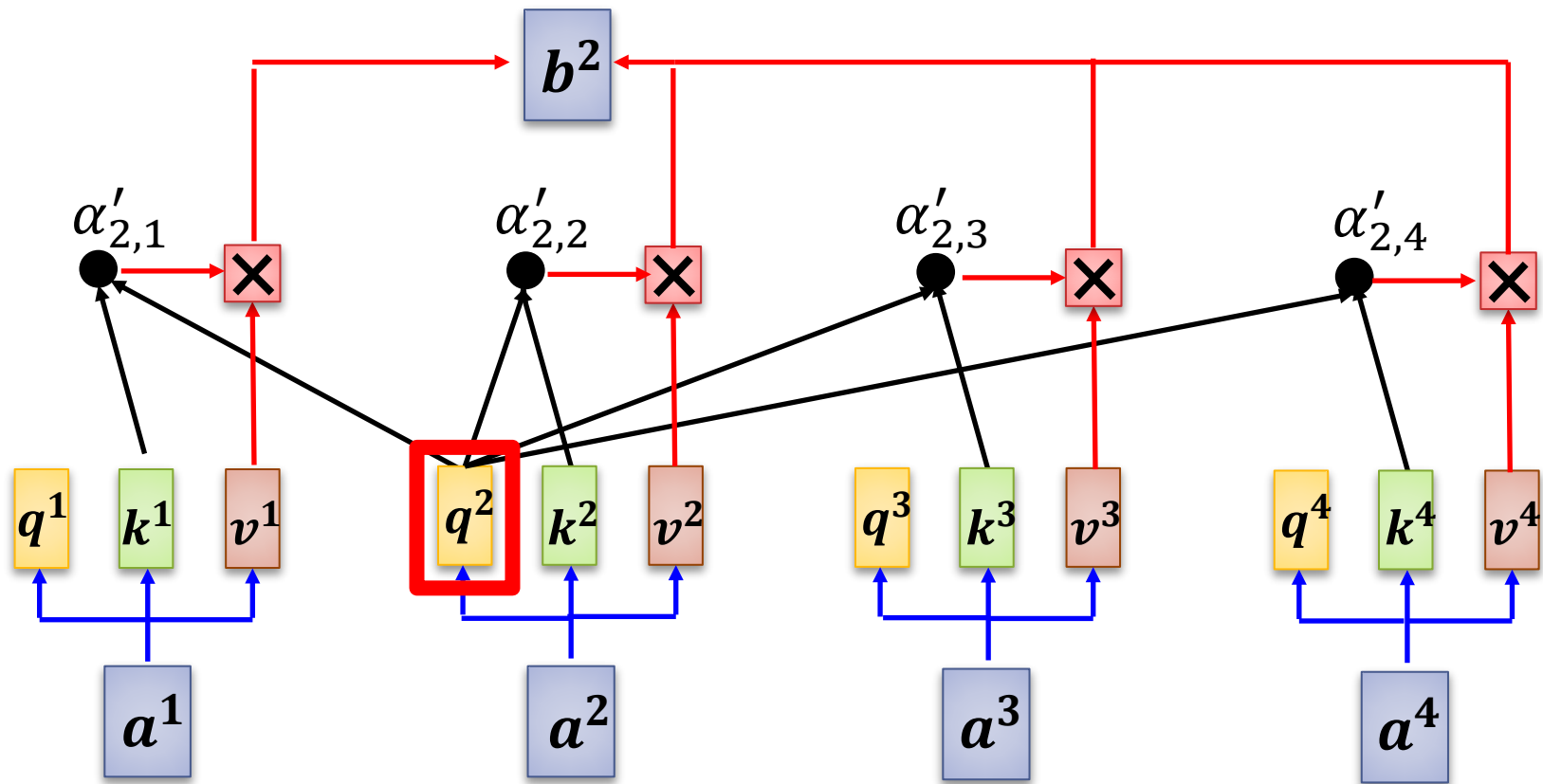
- Encoder: Self-Attention (<https://b23.tv/pTSqhws>)



Can be either **input** or a **hidden layer**

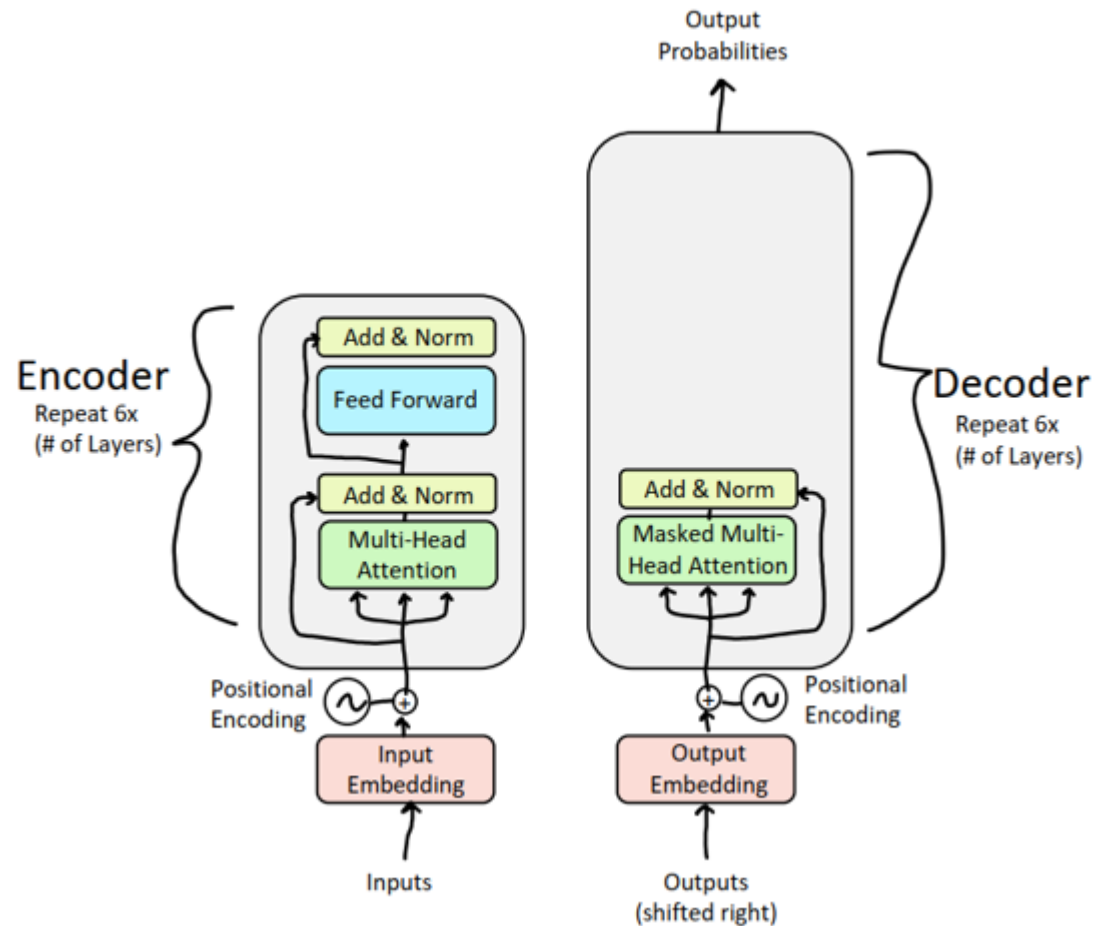
Transformer

- Encoder: Self-Attention (<https://b23.tv/pTSqhws>)



Transformer

- Decoder: Masked Multi-Headed Self-Attention

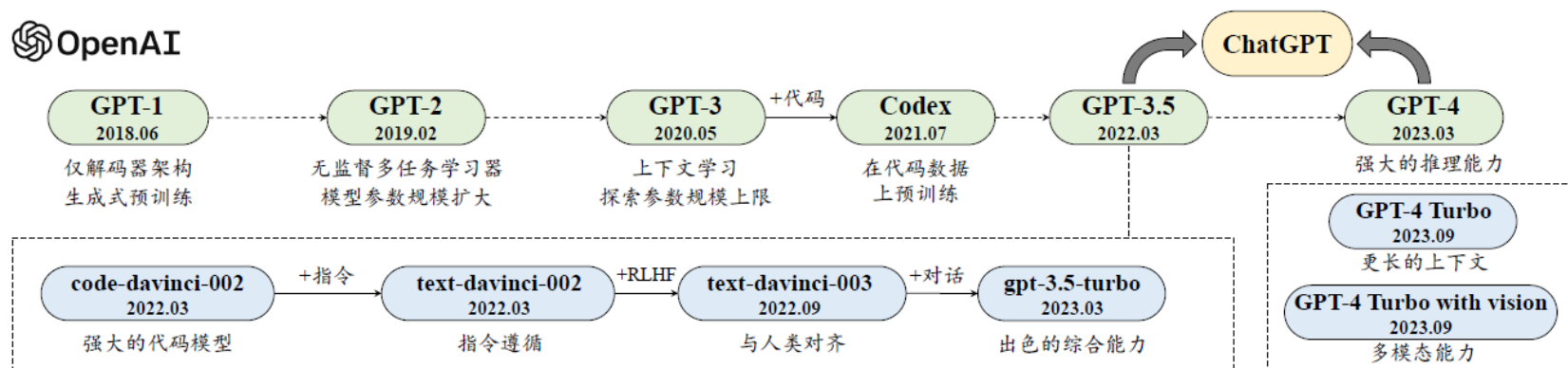


Transformer

- 当Google在2017年推出基于自注意力 (Self-Attention) 机制的 Transformer 模型后，OpenAI团队迅速洞察到了其潜在的优越性，认为这种模型可能是一种大规模可扩展训练的理想架构。
- Transformer架构能够拓展到百亿、千亿甚至万亿参数规模，并且将预训练任务统一为预测下一个词这一通用学习范式。此外，高质量数据、超大规模数据成为大语言模型的关键基础。

GPT

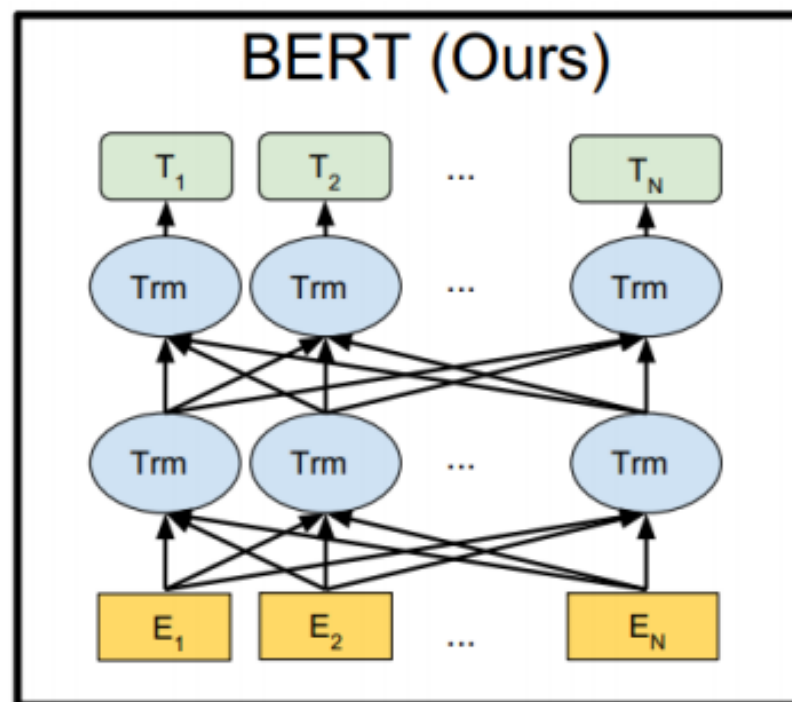
- GPT-1: 首个预训练语言模型，来自OpenAI (2018年)。它采用了仅有解码器 (Decoder)的Transformer架构，以及基于下一个词元预测的预训练任务进行模型的训练。
- GPT系列模型的技术发展历程：



Wayne Xin Zhao et al. A Survey of Large Language Models. arXiv: 2303.18223 (2023).

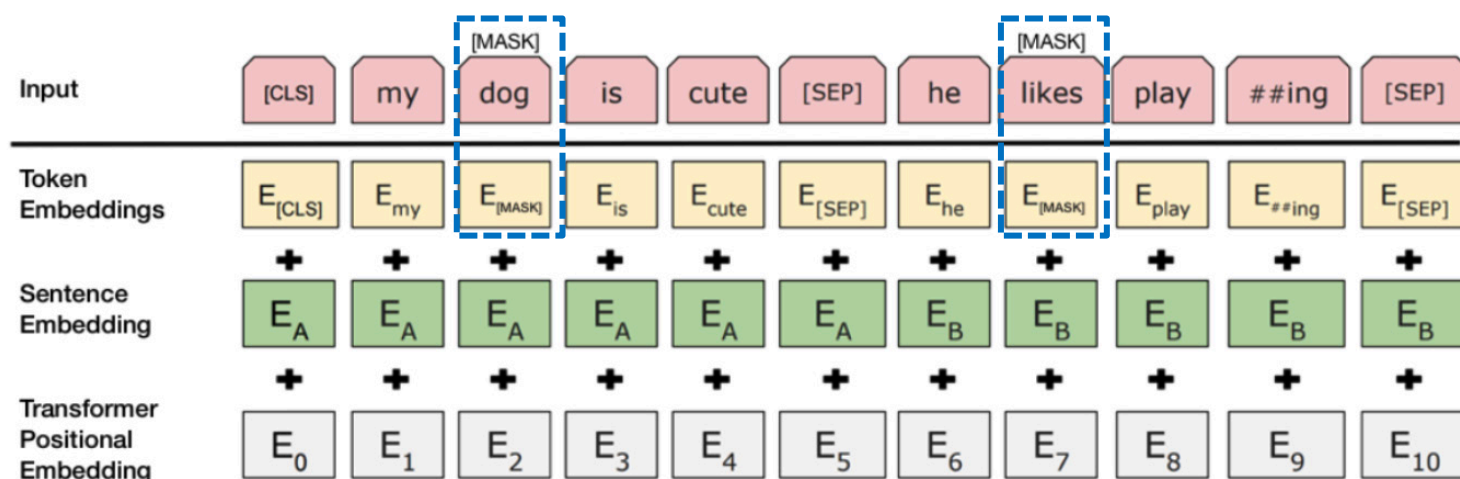
BERT

- 从解码器 (Decoder)到编码器 (Encoder), 来自Google (2019年)。
- B: Bidirectional (双向)
- E: Encoder (编码器)
- R: Representations
- T: Transformers



BERT: 深度双向Transformer

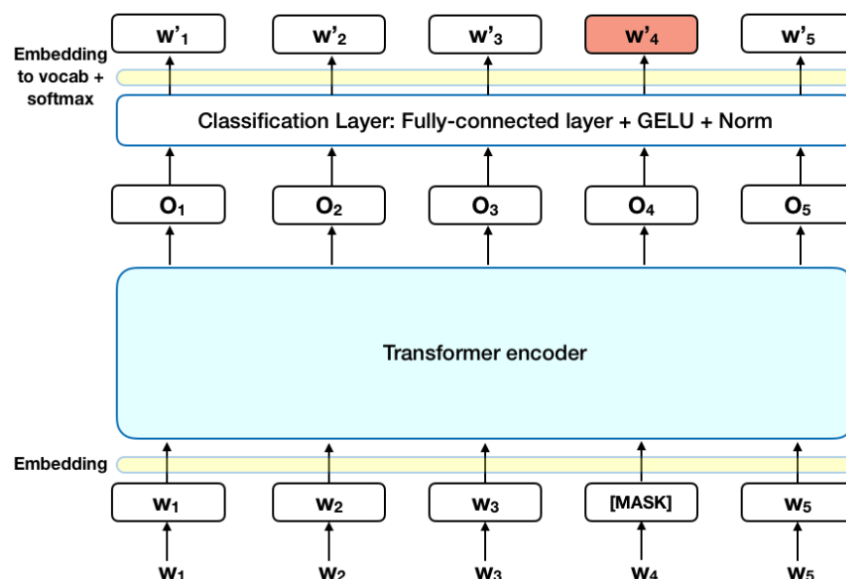
- 双向：BERT通过Masked LM学习来实现真正的双向训练；
- Masked LM任务：随机遮盖15%的词，模型通过学习双向上下文信息来预测被遮盖的词。



Devlin, Jacob, et al. "Bert: Pre-training of deep bidirectional transformers for language understanding." *NAACL* (2019).

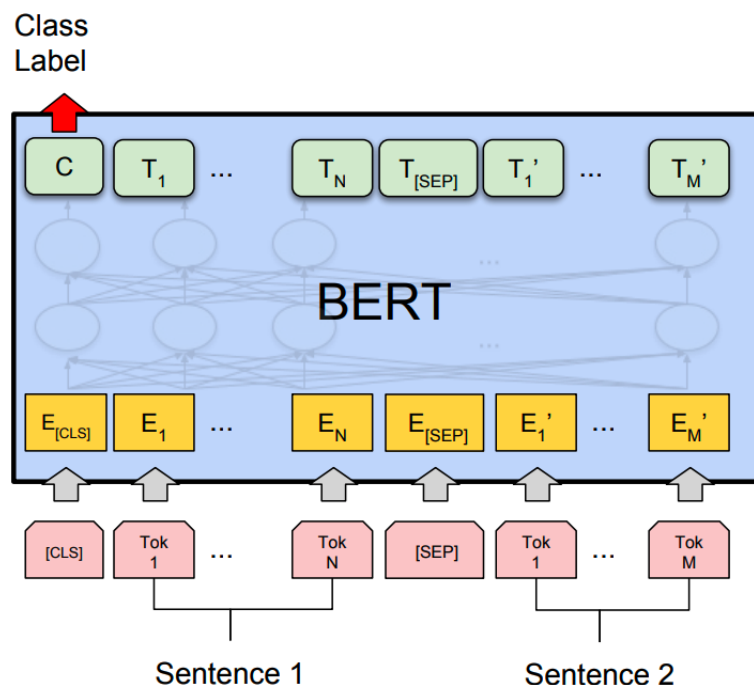
BERT: 深度双向Transformer

- Transformer编码器：一次性读取整个词序列，并据此学习词的上下文。
- Transformer通过注意力机制学习文本序列中词（和子词）的上下文关系。



BERT: 深度双向Transformer

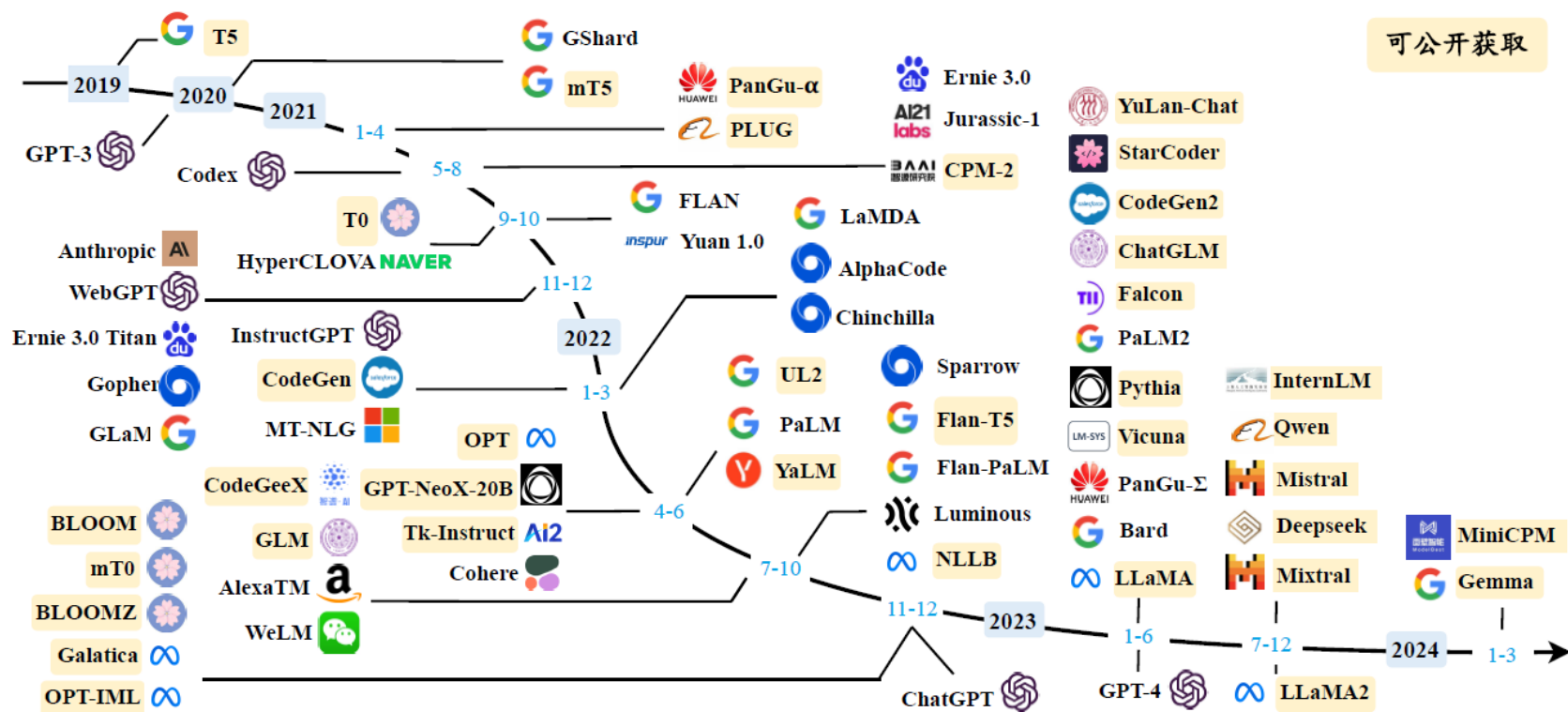
- Next Sentence Prediction任务：输入同一篇文章中的一个句子对，模型预测在原始文档中句子2是否是句子1的下一个句子。



Devlin, Jacob, et al. "Bert: Pre-training of deep bidirectional transformers for language understanding." *NAACL* (2019).

大语言模型

• 大语言模型发展时间线：



Wayne Xin Zhao et al. A Survey of Large Language Models. arXiv: 2303.18223 (2023).

大语言模型

- 大语言模型的主要能力特点：
 - **具有较为丰富的世界知识。**与传统机器学习模型相比，大语言模型经过超大规模文本数据的预训练后能够学习到较为丰富的世界知识。早期的BERT、GPT-1等预训练语言模型的参数规模与数据规模都相对较小，需要使用微调为主要手段来解决下游任务。
 - **具有较强的通用任务解决能力。**大语言模型主要通过预测下一个词元的预训练任务进行学习，虽然并没有针对特定的下游任务进行优化，却能够建立远强于传统模型的通用任务求解能力。

大语言模型

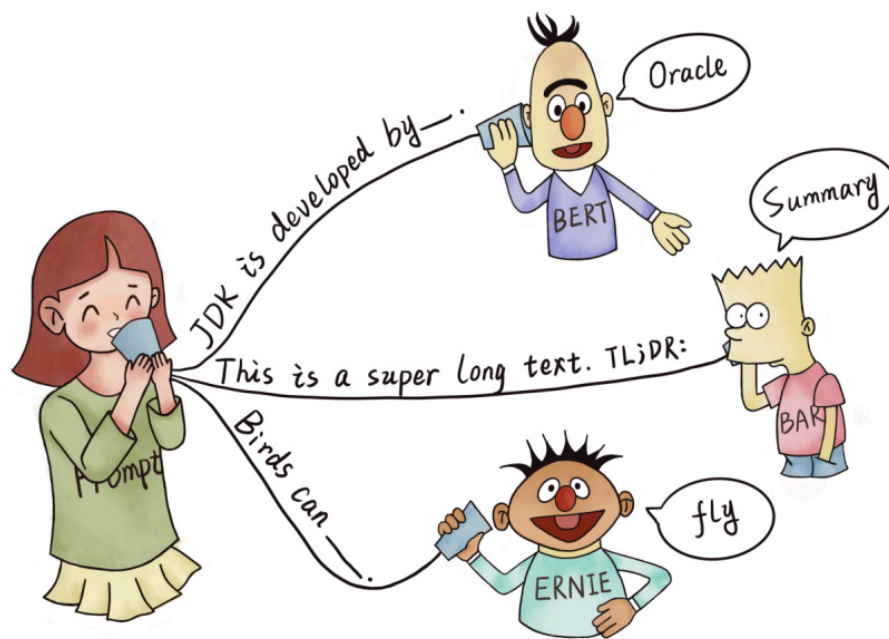
- 大语言模型的主要能力特点：
 - 具有较好的人类对齐能力。由于大语言模型具有出色的性能，如果不进行有效的对齐与监管，将可能带来非常严重的后果。目前广泛采用的对齐方式是基于人类反馈的强化学习技术，通过强化学习使得模型进行正确行为的加强以及错误行为的规避，进而建立较好的人类对齐能力。
 - 具有可拓展的工具使用能力。由于大语言模型具有较为通用的任务求解形式，可以通过微调、上下文学习等方式掌握外部工具的使用，如搜索引擎与计算器。这借鉴了人类使用工具的思路。

大语言模型

- 大语言模型的主要能力特点：
 - 具有较好的复杂任务推理能力。大语言模型能够回答知识关系复杂的推理问题，还可以解决涉及复杂数学推理过程的数学题目。
 - 具有较强的人类指令遵循能力。在早期的对话系统中，指令遵循就是一个受到广泛关注的研究方向。然而，传统模型缺乏通用的任务理解与执行能力，仍然需要依赖人工规则或者先验信息辅助指令理解模块的设计与训练。大语言模型能够直接通过自然语言描述下达任务指令（又称为“提示学习”），使得任务输入与执行结果均可通过自然语言进行表达。

大语言模型

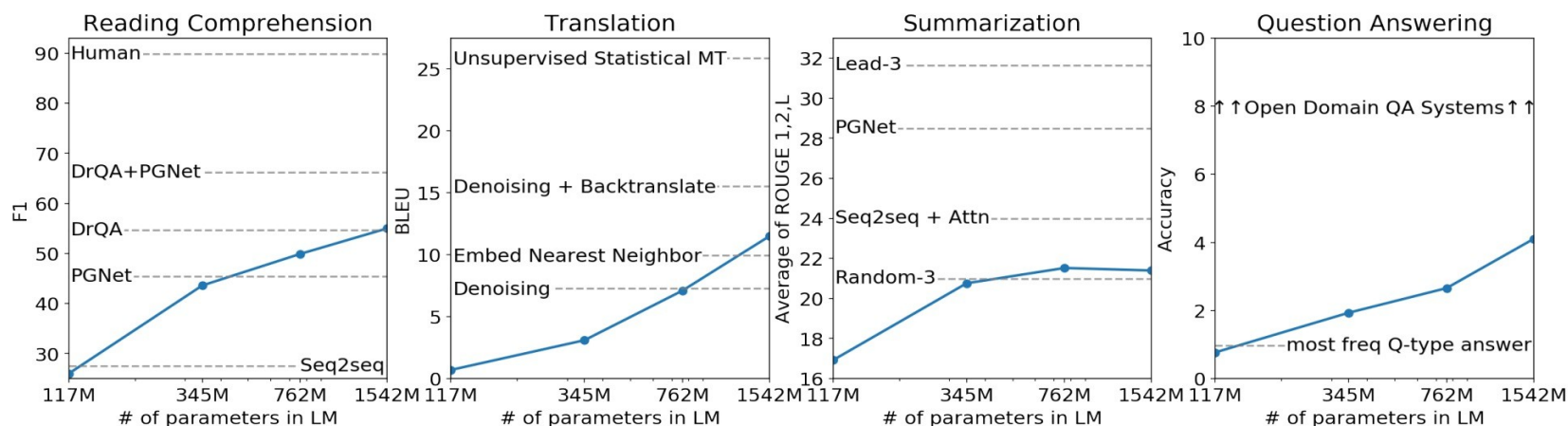
- 在自然语言处理领域，很多传统任务（如分类、翻译、摘要、问答等）都可以采用基于大语言模型的提示学习方法进行解决，而且能够获得较好的任务效果，早期任务特定的解决方案已经被逐步替代。



大语言模型

The authors “demonstrate that language models begin to learn [question answering, machine translation, reading comprehension, and summarization] tasks **without any explicit supervision** when trained on a new dataset of millions of webpages called WebText.”

Across a variety of tasks, they saw this behaviour “emerge” with larger model sizes.



“Language Models are Unsupervised Multitask Learners.” Radford et al. (2019)

大语言模型

- Zero-Shot Learning (Sentiment Classification)

Prompt:

Review: Let there be no question: Alexions owns the best cheeseburger in the region and they have now for decades. Try a burger on Italian bread. The service is flawlessly friendly, the food is amazing, and the wings? Oh the wings... but it's still about the cheeseburger. The atmosphere is inviting, but you can't eat atmosphere... so go right now. Grab the car keys... you know you're hungry for an amazing cheeseburger, maybe some wings, and a cold beer! Easily, hands down, the best bar and grill in Pittsburgh.

On a 1 to 4 star scale, the reviewer would probably give this restaurant a

Model's Output Probabilities:

$P(\text{"1"} \mid \text{prompt}) = 0.03$

$P(\text{"2"} \mid \text{prompt}) = 0.07$

$P(\text{"3"} \mid \text{prompt}) = 0.12$

$P(\text{"4"} \mid \text{prompt}) = 0.78$

大语言模型

- Zero-Shot Learning (Sentiment Classification)

Prompt:

Review: Let there be no question: Alexions owns the best cheeseburger in the region and they have now for decades. Try a burger on Italian bread. The service is flawlessly friendly, the food is amazing, and the wings? Oh the wings... but it's still about the cheeseburger. The atmosphere is inviting, but you can't eat atmosphere... so go right now. Grab the car keys... you know you're hungry for an amazing cheeseburger, maybe some wings, and a cold beer! Easily, hands down, the best bar and grill in Pittsburgh.

On a 1 to 4 star scale, the reviewer would probably give this restaurant a

Model's Output Probabilities:

$P(\text{"1"} \mid \text{prompt}) = 0.03$

$P(\text{"2"} \mid \text{prompt}) = 0.07$

$P(\text{"3"} \mid \text{prompt}) = 0.12$

$P(\text{"4"} \mid \text{prompt}) = 0.78$

Template used:

Review: `<review>`

On a 1 to 4 star scale, the reviewer would probably give this restaurant a `<class_label>`

This template is tried for all possible class labels.

大语言模型

- Zero-Shot Learning (Machine Translation)

文心大模型3.5



请将下述英文翻译为中文: Let there be no question: Alexions owns the best cheeseburger in the region and they have now for decades. Try a burger on Italian bread. The service is flawlessly friendly, the food is amazing, and the wings? Oh the wings... but it's still about the cheeseburger. The atmosphere is inviting, but you can't eat atmosphere... so go right now. Grab the car keys... you know you're hungry for an amazing cheeseburger, maybe some wings, and a cold beer! Easily, hands down, the best bar and grill in Pittsburgh.



毫无疑问: Alexions拥有该地区最好的芝士汉堡, 并且已经持续了几十年。试试意大利面包上的汉堡。服务无可挑剔, 食物令人惊叹, 还有鸡翅? 哦, 鸡翅.....但最重要的还是芝士汉堡。氛围很吸引人, 但你不能吃氛围.....所以现在就出发吧。拿起车钥匙.....你知道你渴望一个美味的芝士汉堡, 也许还有一些鸡翅和一杯冰镇啤酒! 毫无疑问, 这是匹兹堡最好的酒吧和烧烤店。

由文心大模型 3.5 生成

大语言模型

- Few-Shot Learning (Machine Translation)

Prompt:

The dog chased a squirrel at the park. = 那只狗在公园里追一只松鼠。

I was late for class. = 我上课迟到了。

The hippopotamus ate my homework. =

Model Generation:

河马吃了我的家庭作业。



Template Used:

```
<example1_en> = <example1_zh>  
<example2_en> = <example2_zh>  
<query_en> =
```

大语言模型

- Few-Shot Learning (Machine Translation)

Prompt:

The dog chased a squirrel at the park. = 那只狗在公园里追一只松鼠。

I was late for class. = 我上课迟到了。

The hippopotamus ate my homework. =

Model Generation:

河马吃了我的家庭作业。



Template Used:

```
<example1_en> = <example1_zh>  
<example2_en> = <example2_zh>  
<query_en> =
```

Prompt with Alternative Template:

Translate from English to Chinese.

English: The dog chased a squirrel at the park.

Chinese: 那只狗在公园里追一只松鼠。

English: I was late for class. Chinese: 我上课迟到了。

English: The hippopotamus ate my homework. Chinese:

The different templates are called **verbalizers**.

大语言模型

- **Emergence**: when quantitative changes in a system result in qualitative changes (量变到质变) in behavior.
- **Emergent (涌现) behaviors/abilities** : abilities that larger models have and smaller models don't.
- **In-context learning (上下文学习)**: when a language model “learns” how to do a task from a textual prompt containing a natural language instruction for the task, several exemplars of the task, or both.
- **Zero-shot learning (零样本学习)**: In context learning that does not include any exemplars of the task.
- **Few-shot learning (少样本学习)**: In context learning that contains several exemplars of the task.
- **Prompt engineering (提示工程)**: The painstaking process of trying out many different prompts until you find one that works well for your task.
- **Verbalizer**: The template (模板) we wrap an example in order to perform the task.

实验对比

- SemEval数据集 (1221篇新闻标题)
 - 样例： europe retain trophy with big win, **joy**
 - 训练集： 598； 验证集： 311； 测试集： 312
 - 标签集： **anger, disgust, fear, joy, sad, surprise**
- 20News数据集 (18314篇新闻文本)
 - 样例： how about those toneau covers \? i 've been thinking of building one from chipboard for roadtrips any comment on how they affect mileage in highway travel \? charles, **rec.autos**
 - 训练集： 8791； 验证集： 2197； 测试集： 7326
 - 标签集： **alt.atheism, comp.graphics, comp.os.ms-windows.misc, comp.sys.ibm.pc.hardware, comp.sys.mac.hardware, comp.windows.x, misc.forsale, rec.autos, rec.motorcycles, rec.sport.baseball, rec.sport.hockey, sci.crypt, sci.electronics, sci.med, sci.space, soc.religion.christian, talk.politics.guns, talk.politics.mideast, talk.politics.misc, talk.religion.misc**

实验对比

- 朴素贝叶斯分类模型

```
import numpy as np
from sklearn.naive_bayes import MultinomialNB
from sklearn.feature_extraction.text import CountVectorizer

def load_data():
    ...
    cv = CountVectorizer()
    ...

# 文本表示
x_test, y_test = load_data()

m_clf = MultinomialNB()
m_clf.fit(X = x_train, y = y_train)

y_pred = m_clf.predict(x_test)
acc = accuracy_score(y_test, y_pred)
```

实验对比

- 朴素贝叶斯分类模型

```
import numpy as np
from sklearn.naive_bayes import MultinomialNB
from sklearn.feature_extraction.text import CountVectorizer

def load_data():
    ...
    cv = CountVectorizer()
    ...

# 文本表示
x_test, y_test = load_data()

m_clf = MultinomialNB()
m_clf.fit(X = x_train, y = y_train)

y_pred = m_clf.predict(x_test)
acc = accuracy_score(y_test, y_pred)
```

SemEval准确率: 训练集
96.15%、验证集44.37%、
测试集53.85%

20News准确率: 训练集
69.55%、验证集53.94%、
测试集54.35%

实验对比

- 卷积神经网络

```
import numpy as np
import CNN
import torch
import torch.nn as nn
import torchtext.vocab as vocab

# 加载词汇表和预训练的Glove词嵌入
def load_vocab_and_embeddings():
    with open(GLOVE_FILE, 'r', encoding='utf-8') as f:
        ...
    return word_to_idx, weights

# 定义卷积神经网络
class TextCNN(nn.Module):
    ...
```


实验对比

- 卷积神经网络

```
import numpy as np
import CNN
import torch
import torch.nn as nn
import torchtext.vocab as vocab
```

```
# 加载词汇表和预训练的Glove词嵌入
```

```
def load_vocab_and_embeddings():
    with open(GLOVE_FILE, 'r', encoding='utf-8') as f:
        ...
    return word_to_idx, weights
```

```
# 定义卷积神经网络
```

```
class TextCNN(nn.Module):
    ...
```

SemEval准确率: 训练集
83.11%、验证集50.80%、
测试集56.09%

20News准确率: 训练集
86.49%、验证集52.07%、
测试集52.42%

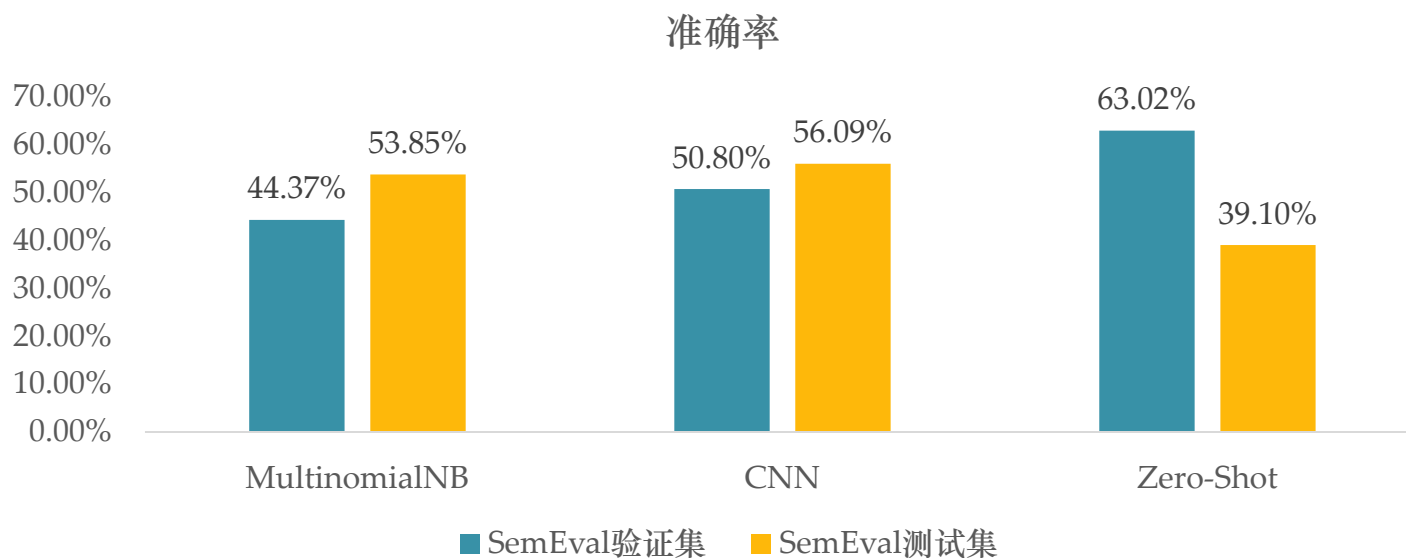
实验对比

- 大语言模型：
Qwen1.5-
7B-Chat
- 设备：
RTX3090
- 方法：
Zero-Shot
& Few-Shot
Learning

```
base_url = "http://localhost:8000/v1"
client = OpenAI(api_key="EMPTY", base_url=base_url)
results = []
#具体不同任务的prompt
prompt = ""
# 迭代每一行数据
for description in data['Text']:
    messages = [
        {'role': 'system', 'content': prompt},
        {'role': 'user', 'content': description}
    ]
    # 调用模型获取结果
    response = client.chat.completions.create(
        model="Qwen1.5-7B-Chat",
        messages=messages,
        stream=False,
        max_tokens=2048,
        temperature=0.9,
        presence_penalty=1.1,
        top_p=0.8)
    # 检查响应并处理结果
    if response:
        content = response.choices[0].message.content
        results.append(content)
    else:
        results.append(f"Error: {response.status_code}")
```

实验对比

- Zero-Shot Prompt for SemEval: Given the emotion labels of anger, disgust, fear, joy, sad, and surprise, you should choose the most appropriate label describing the following sentence, you can only choose from the six labels that I have mentioned: anger, disgust, fear, joy, sad, and surprise.
- 准确率：训练集58.03%、验证集63.02%、测试集39.10%



实验对比

- **Few-Shot** Prompt for SemEval: Perform the following actions:
1 - Identify the emotion of the following sentence. 2 - Given the emotion labels of anger, disgust, fear, joy, sad, and surprise, and choose the most appropriate label describing the following sentence.
3 - Only response with the label in at most 1 word.

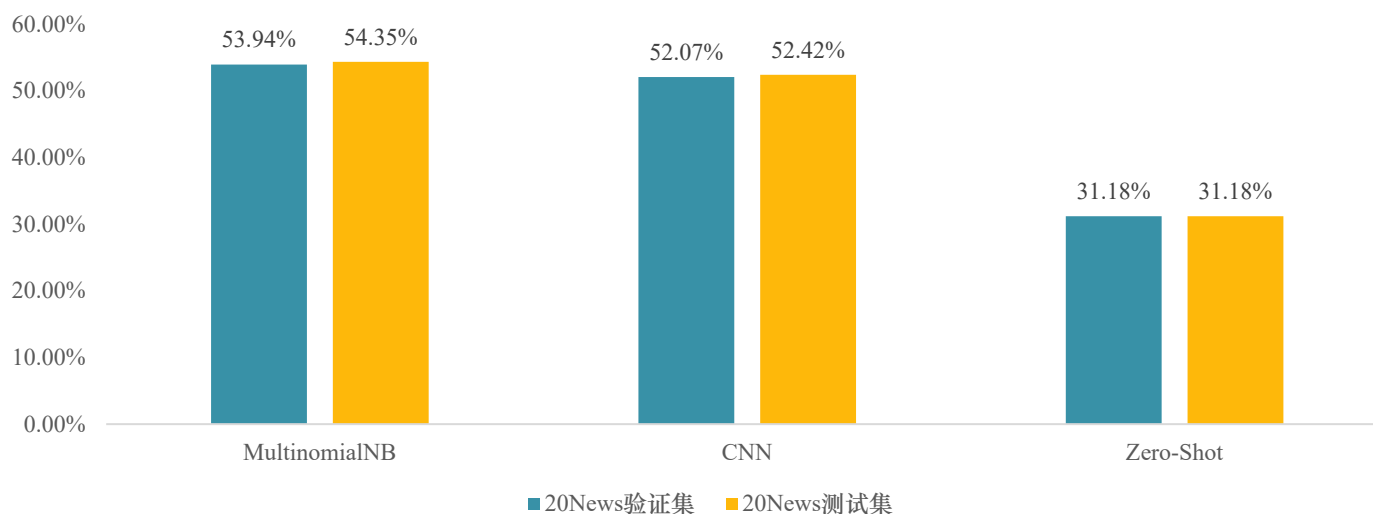
senate votes to revoke pensions --> sad

the amounts you have to pay for a bomb scare --> fear

- 准确率：训练集59.20%、验证集**58.52%**、测试集**45.83%**
- 验证集和测试集上的平均准确率：
 - 朴素贝叶斯：49.11%
 - 卷积神经网络：53.45%
 - Few-Shot Learning：52.18%

实验对比

- Zero-Shot Prompt for 20News: Given the text in the content, identify the most relevant topic from the predefined list of topics. Each piece of text comes from a diverse range of subjects, and you should categorize it based on the main subject matter discussed. Topics: alt.atheism comp.graphics comp.os.ms-windows.misc comp.sys.ibm.pc.hardware comp.sys.mac.hardware comp.windows.x misc.forsale rec.autos rec.motorcycles rec.sport.baseball rec.sport.hockey sci.crypt sci.electronics sci.med sci.space soc.religion.christian talk.politics.guns talk.politics.mideast talk.politics.misc talk.religion.misc Please choose the most appropriate topic number that best fits the text and only response with the topic without number.
- 准确率：训练集29.46%、验证集31.18%、测试集31.18%



实验对比

- **Few-Shot Prompt for 20News:** Given the text in the content, identify the most relevant topic from the predefined list of topics: alt.atheism comp.graphics comp.os.ms-windows.misc comp.sys.ibm.pc.hardware comp.sys.mac.hardware comp.windows.x misc.forsale rec.autos rec.motorcycles rec.sport.baseball rec.sport.hockey sci.crypt sci.electronics sci.med sci.space soc.religion.christian talk.politics.guns talk.politics.mideast talk.politics.misc talk.religion.misc.

Text: ``since when does atheism mean trashing other religions \? there must be a god of inbreeding to which you are his only son`` Topic: alt.atheism

Text: ``battery powered devices like the powerbook are sometimes more sensitive to serial port weirdness i had trouble with connecting my mac plus to an hp 95lx handheld everything else worked okay on that port , but not the hp \ (it runs on two penlite batteries \) it turned out that the plus \ (by accident or by design flaw \? \) was putting a 4 volt bias on the serial port that was doing weird things to the hp \ (which has only 3v dc ! \) the hp worked fine when connected to the printer port does your pb screen get dim or anything when connected to the device \? have you tried using the printer port \? good luck Jamie`` Topic: comp.sys.mac.hardware

Text: ``how does one read the betting spreads for baseball \? they tend to be something like 8 9 which means it must not be runs ! that spread means you bet 5 on the underdog to win 8 , or 9 on the favorite to win 5`` Topic: rec.sport.baseball

Text: ``true rumor fact ! a big three way deal ! eric lindros going to ottawa senators and senators get 15mill from montreal montreal gets alexander daigle \ (the first round pick from senators \) philly gets damphousse , bellow , patrick roy and a draft pick Topic: rec.sport.hockey

Text: ``if they had n't killed the atf people in the original raid , i think i would laugh my ass off \ (actually , to be honest , i still might Topic: talk.politics.guns

- 平均准确率: 朴素贝叶斯54.15%、卷积神经网络52.25%、Few-Shot Learning34.87%