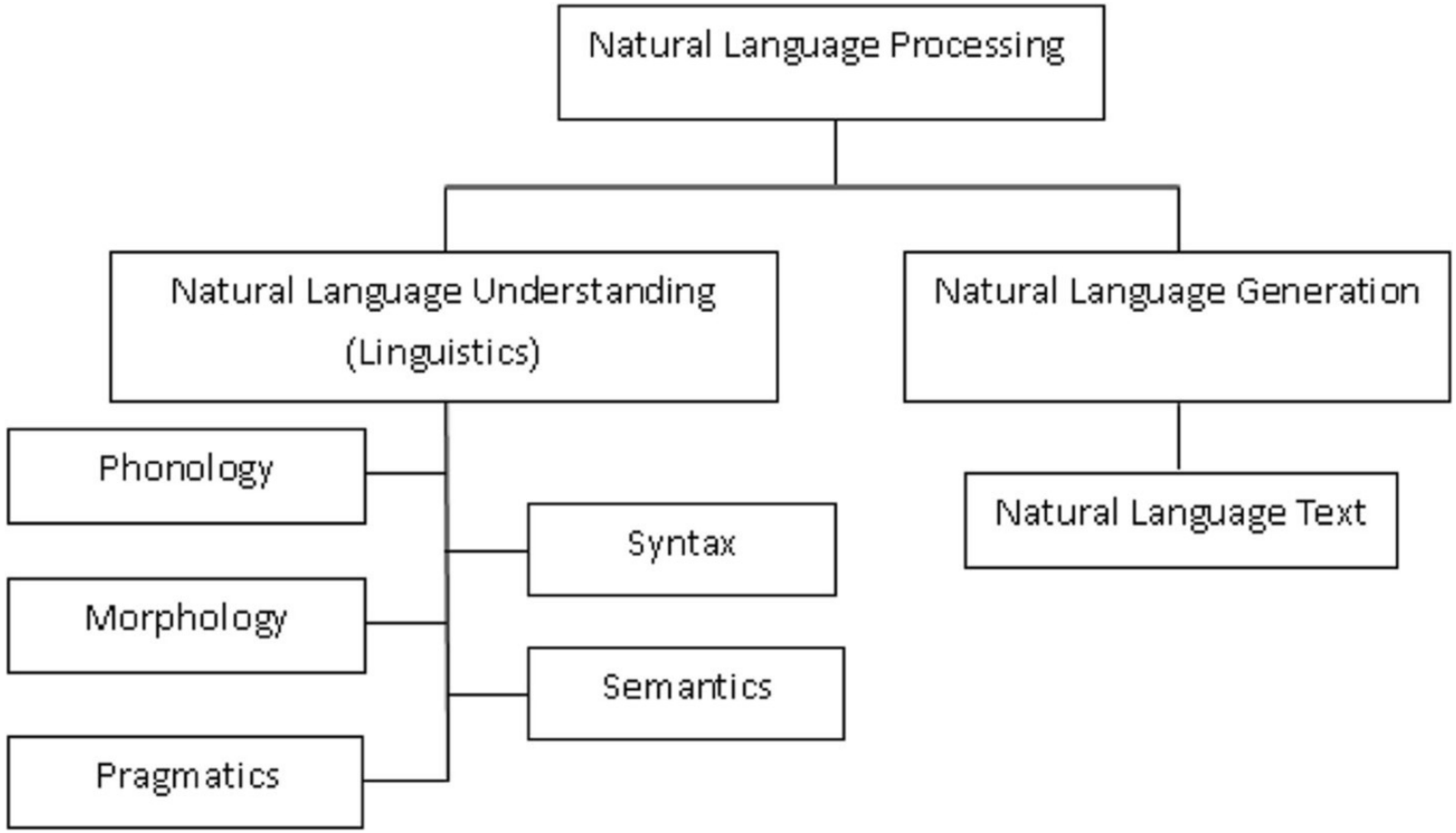
NLP研究热点变迁

第一课

Scott

自然语言处理学科



Natural language processing: state of the art, current trends and challenges

重要的工具库

操作系统:
Linux command line
硬件:
CUDA
编程语言:
Python
Anaconda https://www.anaconda.com/
数据处理:
Numpy https://numpy.org/
Pandas https://pandas.pydata.org/
matplotlib
机器学习:
Scikit learn https://scikit-learn.org/stable/
深度学习:
Torch https://pytorch.org/
自然语言处理:
nltk https://www.nltk.org/
Hugginface Transformer https://github.com/huggingface/transformers

数学知识

- 线性代数
- 矩阵运算（torch）

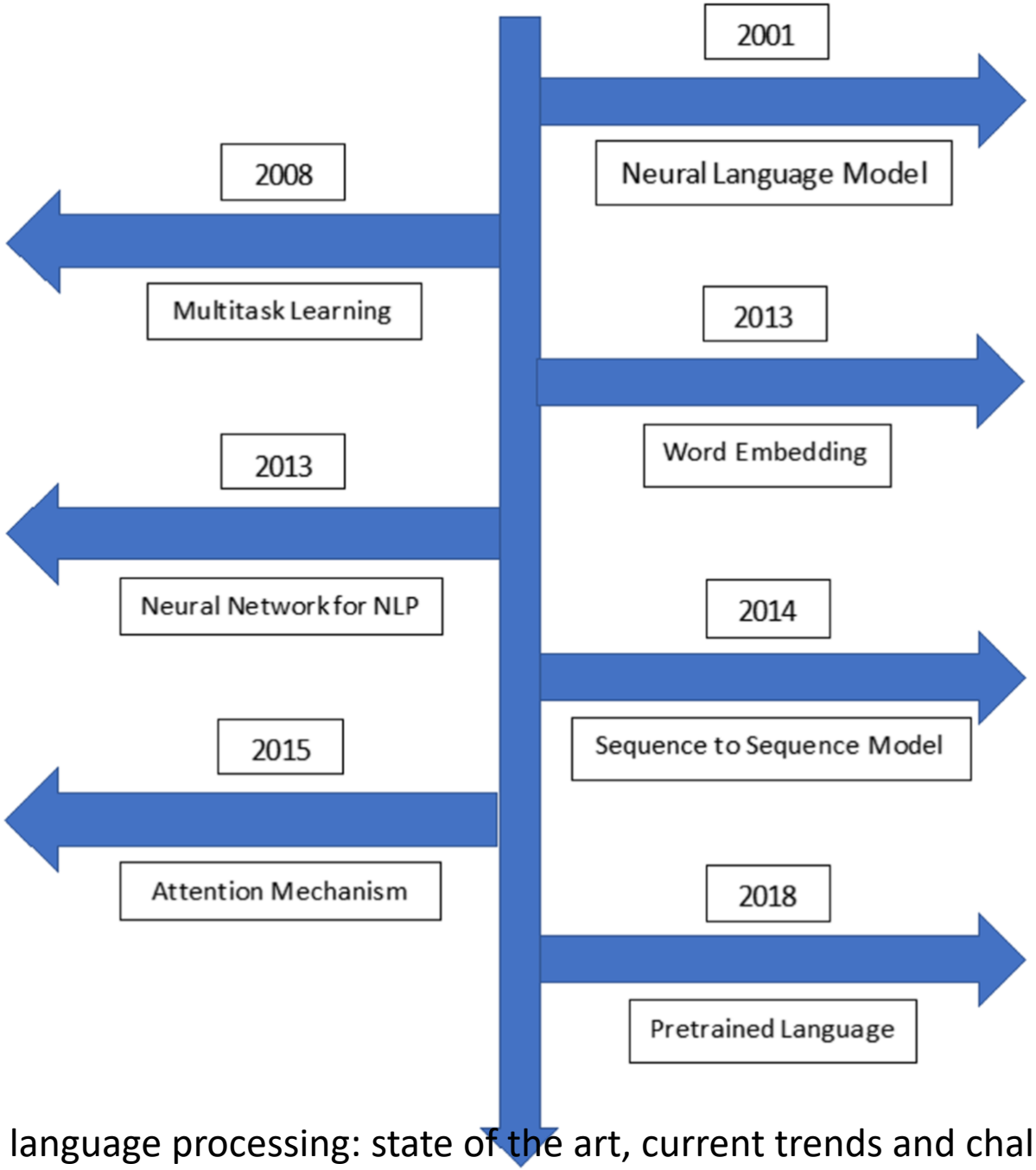
- 微积分
- 导数
- 偏导

- 概率与统计
- 条件概率
- 最大似然估计
- 贝叶斯公式（先验概率/后验概率）
- 联合概率的链式法则/马尔可夫链

信息论
- 熵
- 互信息
- KL散度

凸优化
- 梯度下降
- 线性规划
- 拉格朗日乘子法

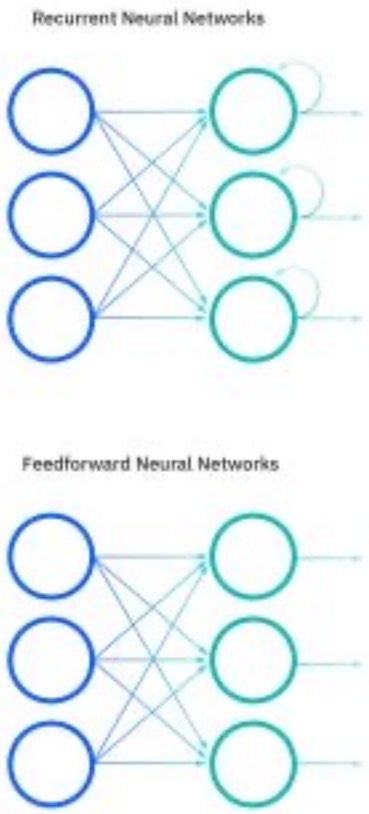
时间线

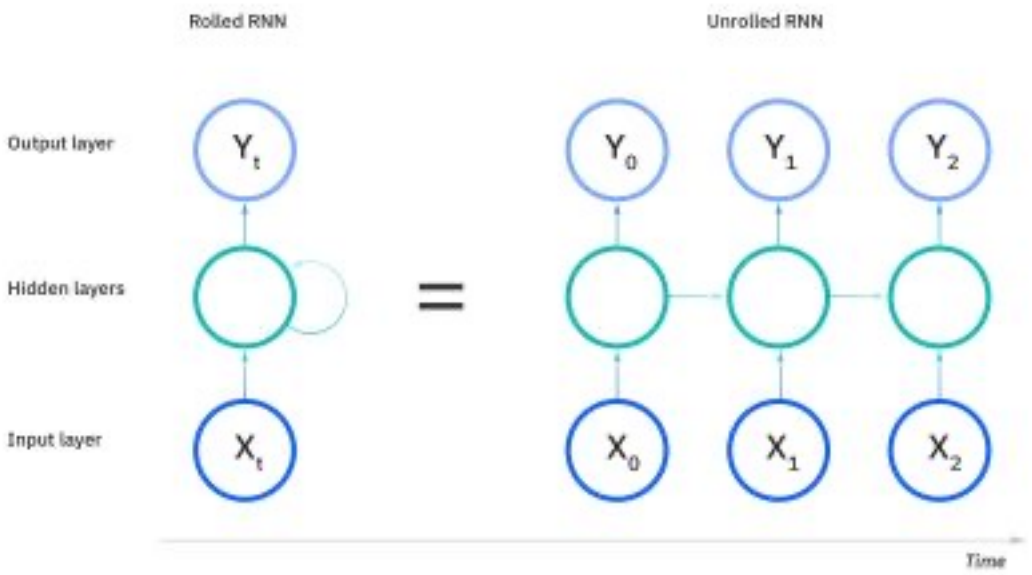


Natural language processing: state of the art, current trends and challenges

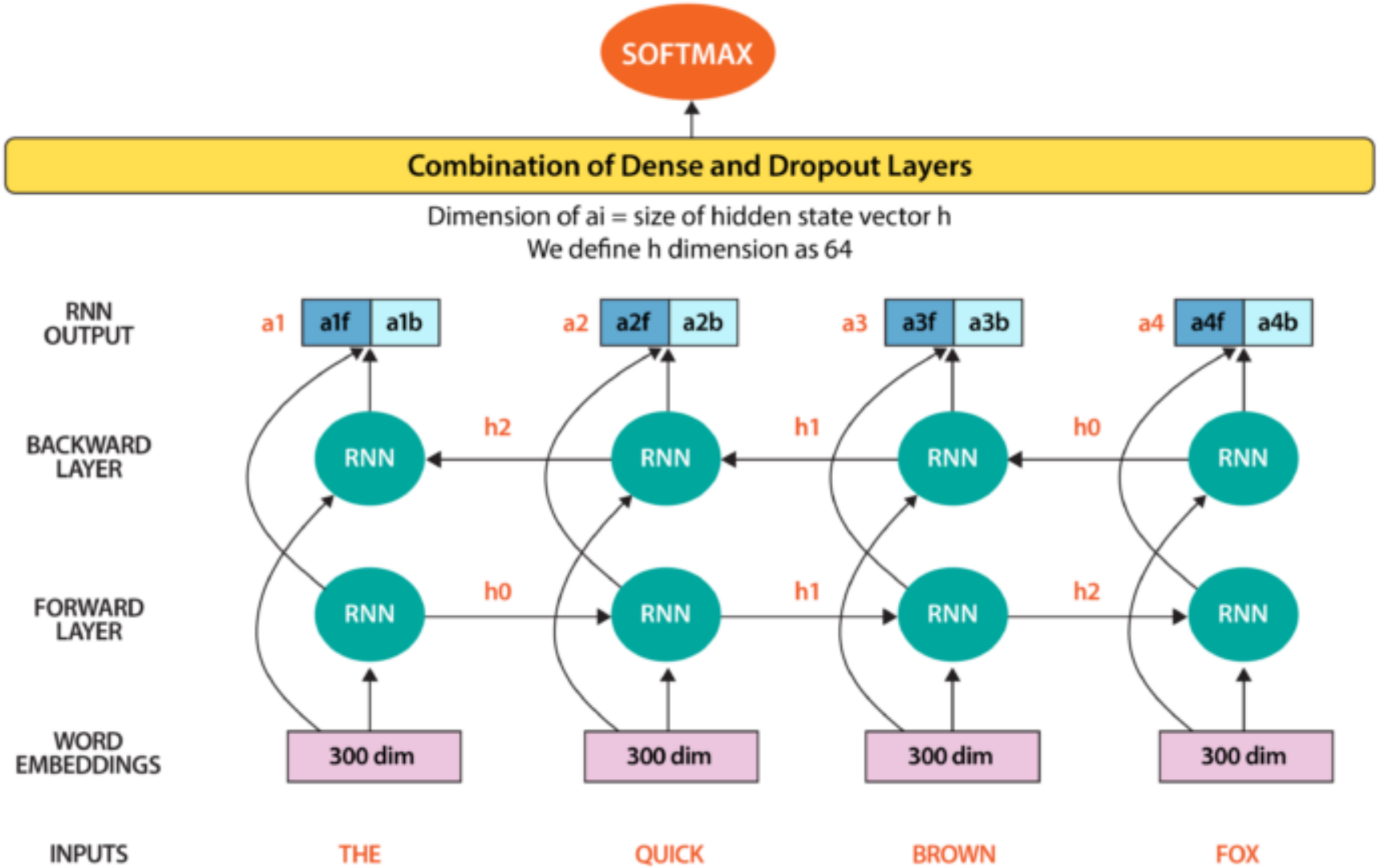
Pre-transformer

- RNN
- LSTM
- GRU



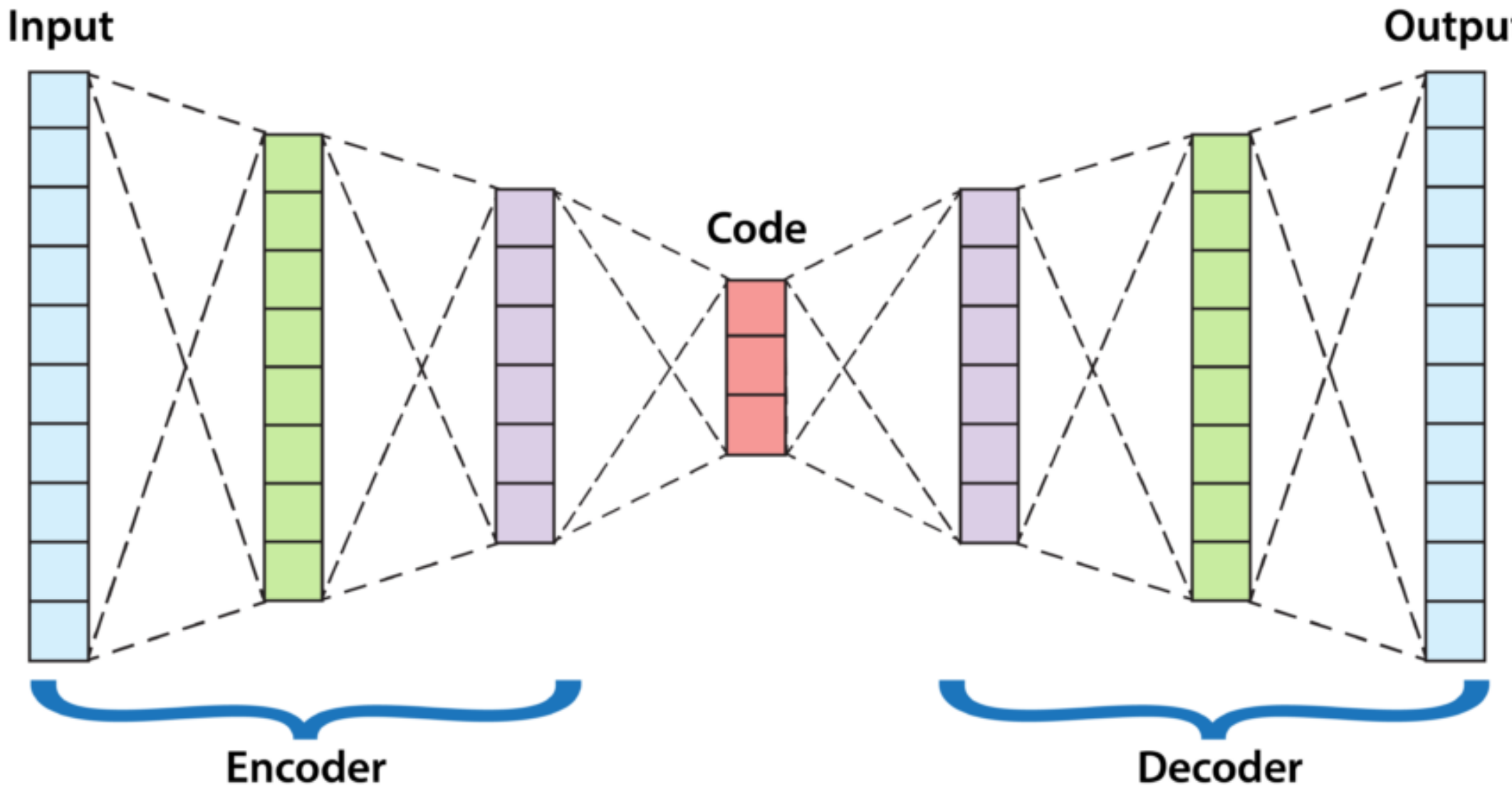


RECURRENT NEURAL NETWORK



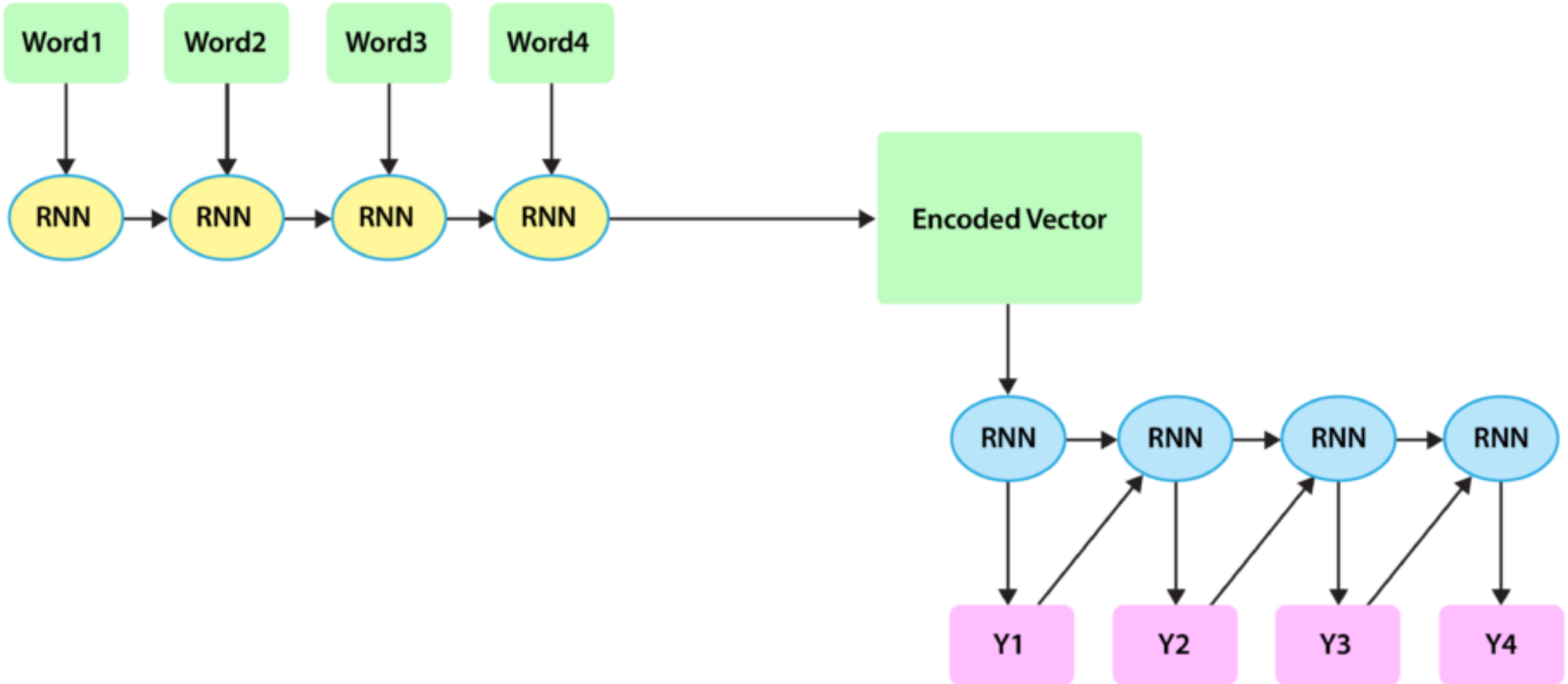
A bidirectional recurrent neural network processes the input both forward and backward to improve the representations it produces.

AUTO-ENCODER



An autoencoder uses an encoder to compress an input into a representation and a decoder to reconstruct the input from the representation.

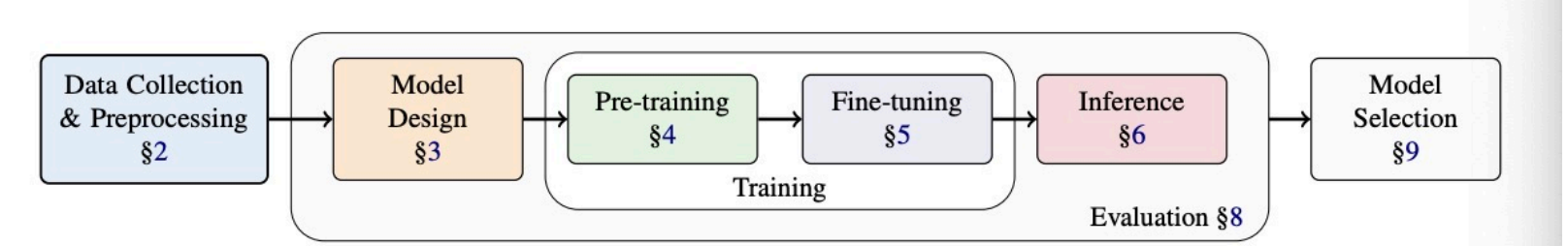
SEQ2SEQ MODEL FOR TRANSLATION

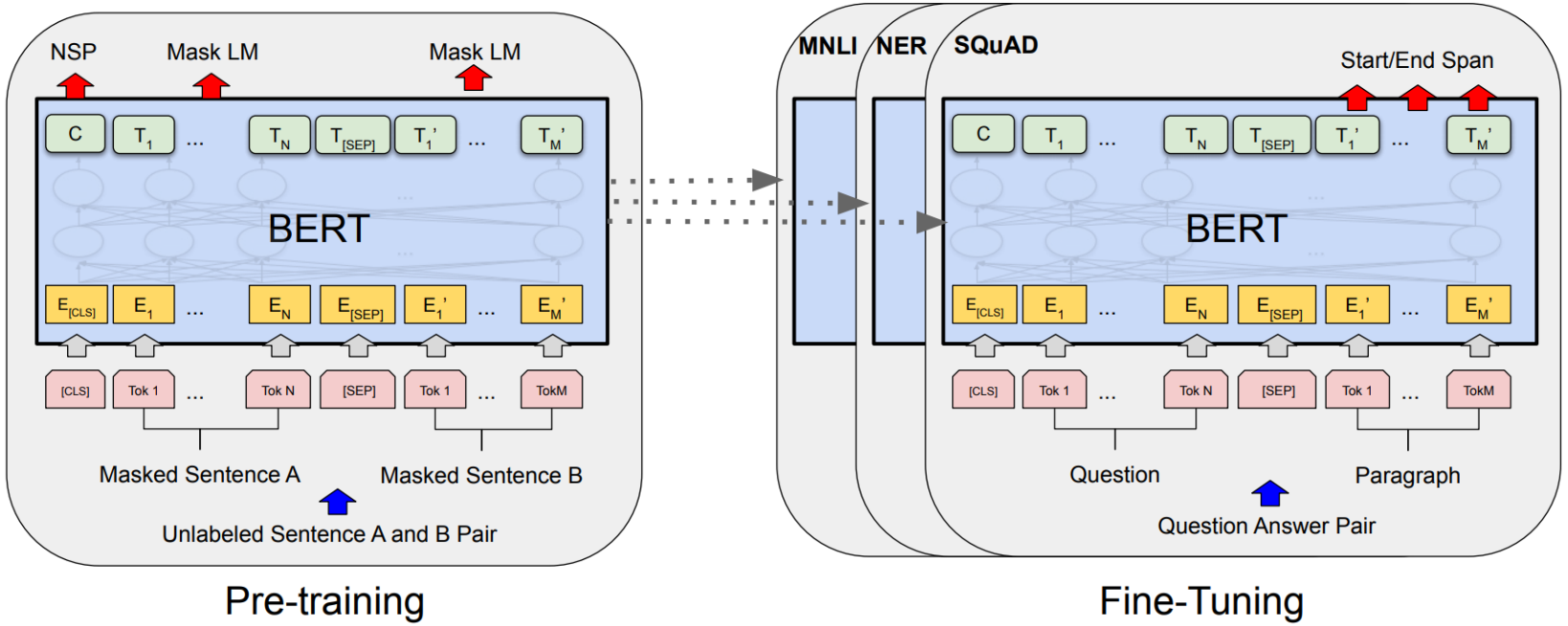


Given a sentence, a Recurrent Neural Network encodes the sentence and then iteratively generates a translation.

后transformer时代

- 自然语言处理发展进程（后transformer时代）范式转移与统一范式





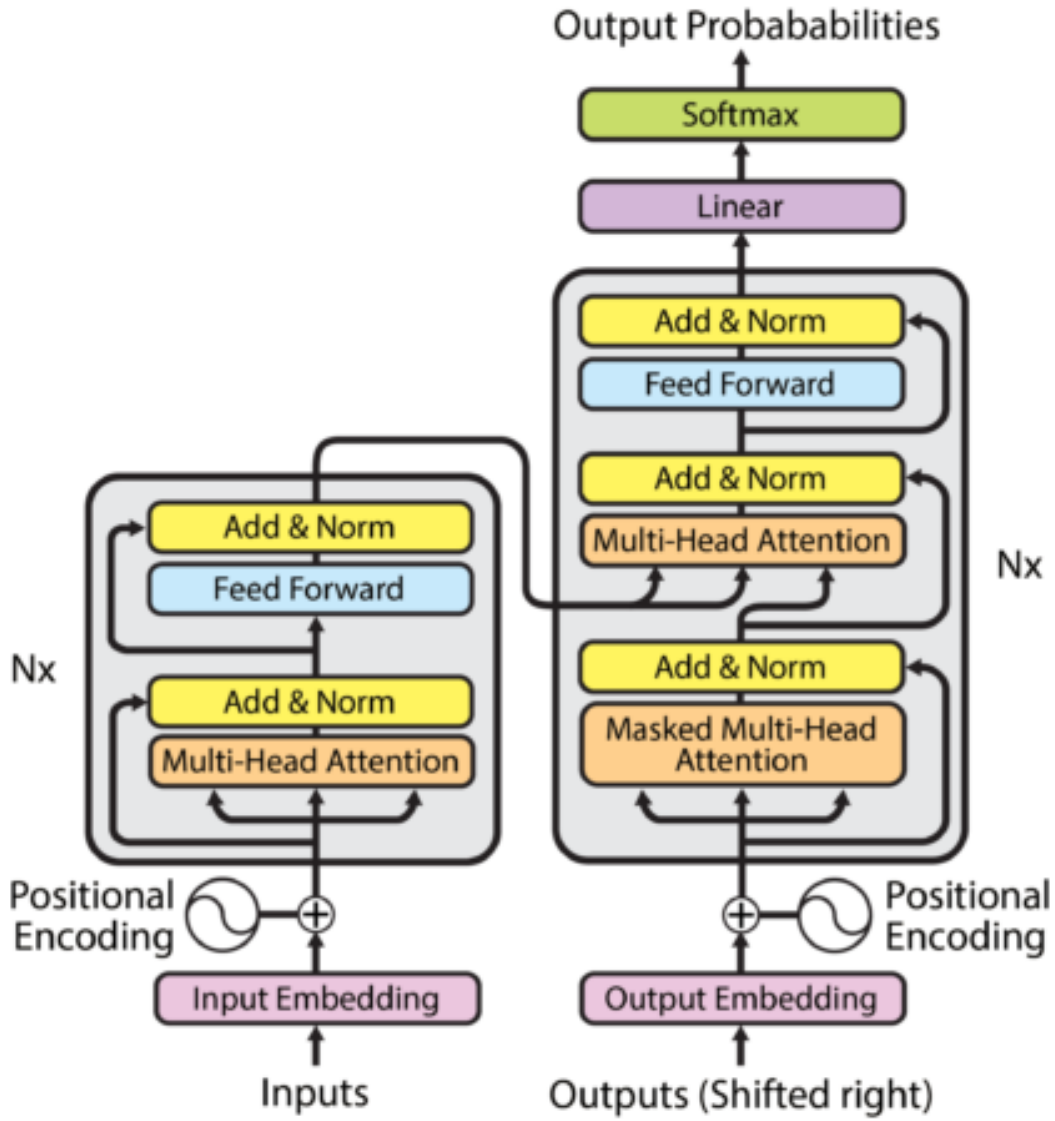
Transformers

- Self-attention

Attention is all you need

https://github.com/huggingface/transformers

TRANSFORMER



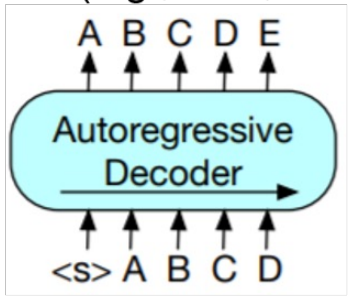
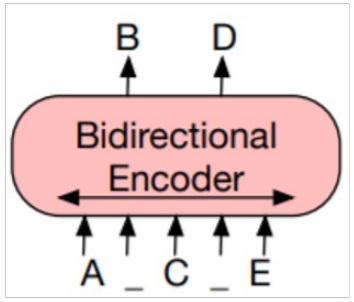
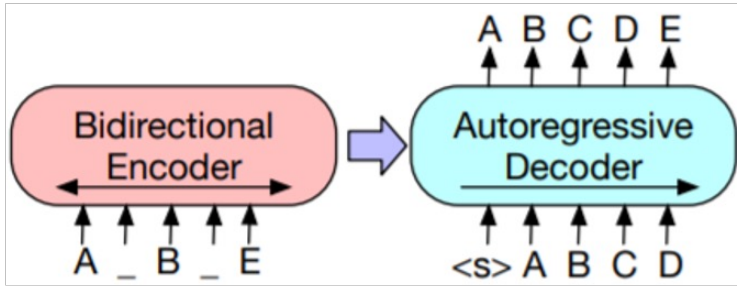
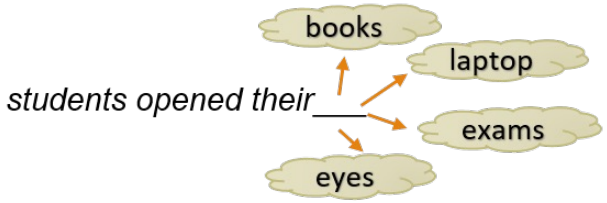
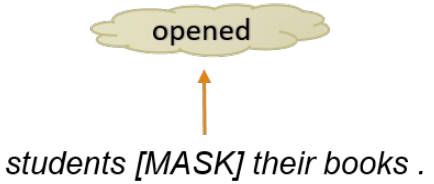
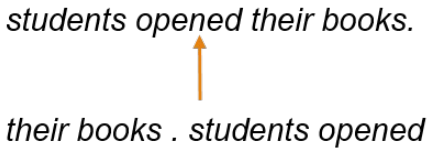
The encoder-decoder transformer used for translation. Encoder on the left, decoder on the right. Note that the decoder takes in its previously generated words during generation.

基于transformer的预训练模型

Model	Pre-Training Sources	Size of Pre-Training Corpus	# Model Parameters
(1) **English Monolingual Models**			
BERT (BASE/LARGE) [31]	Wiki, books	3.3B tokens (13 GB data)	110M/340M
ROBERTA [112]	Wiki, books, web crawl	161 GB data	340M
XLNET [203]	Wiki, books, web crawl	142 GB data	340M
GPT [144]	Web crawl	800M tokens	117M
GPT-2 [145]	Web crawl	8M documents (40 GB data)	1.5B
GPT-3 [39]	Wiki, books, web crawl	300B tokens	175B
GPT-J [186]	Wiki, books, papers, web crawl	~275B tokens (825 GB data)	6B
GOPHER [38]	Books, news, code, web crawl	300B tokens	280B
BART [96]	Wiki, books	3.3B tokens	~370M
T5 [147]	Web crawl	200B tokens (750 GB data)	11B
(2) **Multilingual Models**			
MBERT [31]	Wiki	21.9B tokens	172M
XLM-R(BASE/LARGE) [22]	Web crawl	295B tokens	270M/550M
MT5 (LARGE/XXL) [147]	Web crawl	6.3T tokens	1.2B/13B

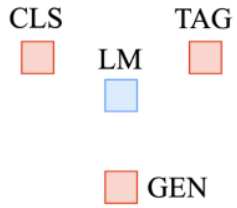
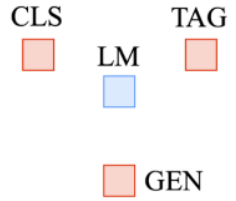
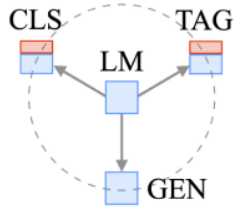
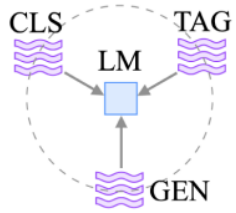
Recent Advances in Natural Language Processing via Large Pre-trained Language Models: A Survey

预训练模型分类

	Autoregressive language model (e.g., GPT, GPT-2/3)	**Masked** language model (e.g., BERT, RoBERTa, XLM-R)	**Encoder-Decoder** (e.g., BART, T5)
Model & illustration			
Training objective	Predicting what word comes next given previous words	Predicting masked words given other words in the sequence	Corrupting a sequence and then predicting the original sequence
Example			

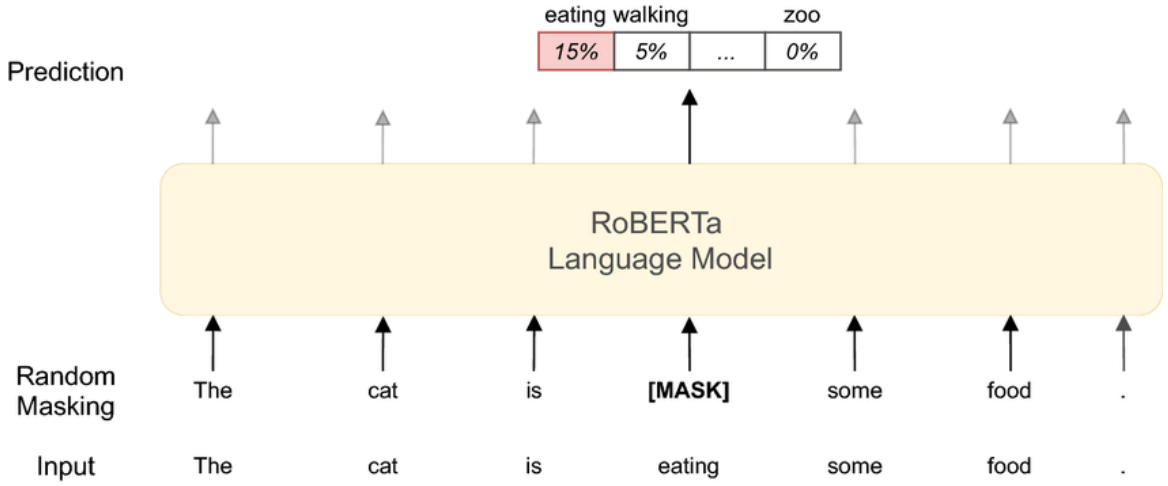
Recent Advances in Natural Language Processing via Large Pre-trained Language Models: A Survey

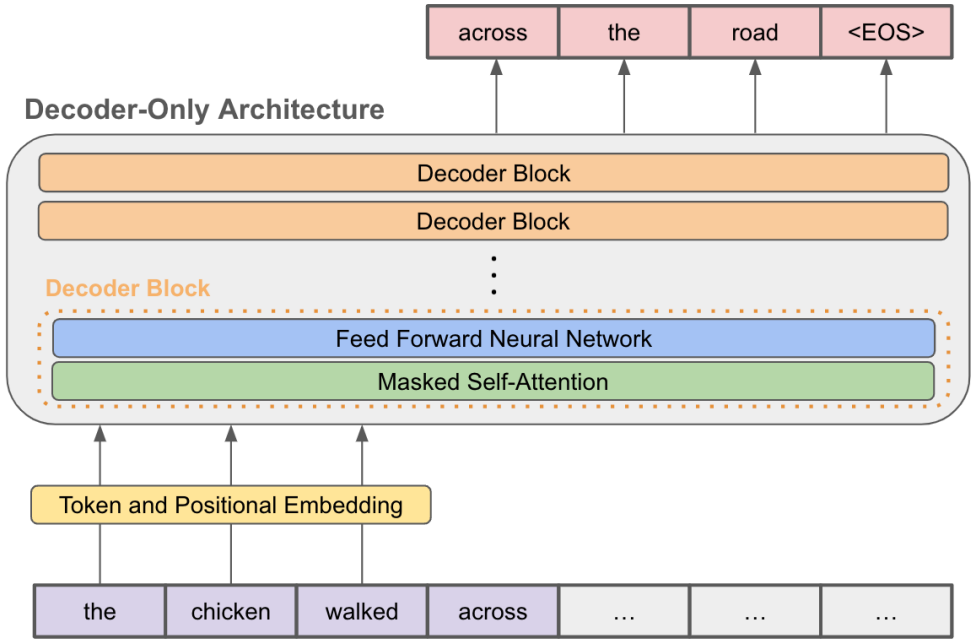
NLP研究范式

Paradigm	Engineering	Task Relation
a. Fully Supervised Learning (Non-Neural Network)	Feature (e.g. word identity, part-of-speech, sentence length)	
b. Fully Supervised Learning (Neural Network)	Architecture (e.g. convolutional, recurrent, self-attentional)	
c. Pre-train, Fine-tune	Objective (e.g. masked language modeling, next sentence prediction)	
d. Pre-train, Prompt, Predict	Prompt (e.g. cloze, prefix)	

Recent Advances in Natural Language Processing via Large Pre-trained Language Models: A Survey

Transfer learning





- BERT - masked language modeling (MLM), next sentence prediction (NSP)
- GPT - predicting the next word
- BART - denoising (masking, sentence permutation, token deletion, document rotation)

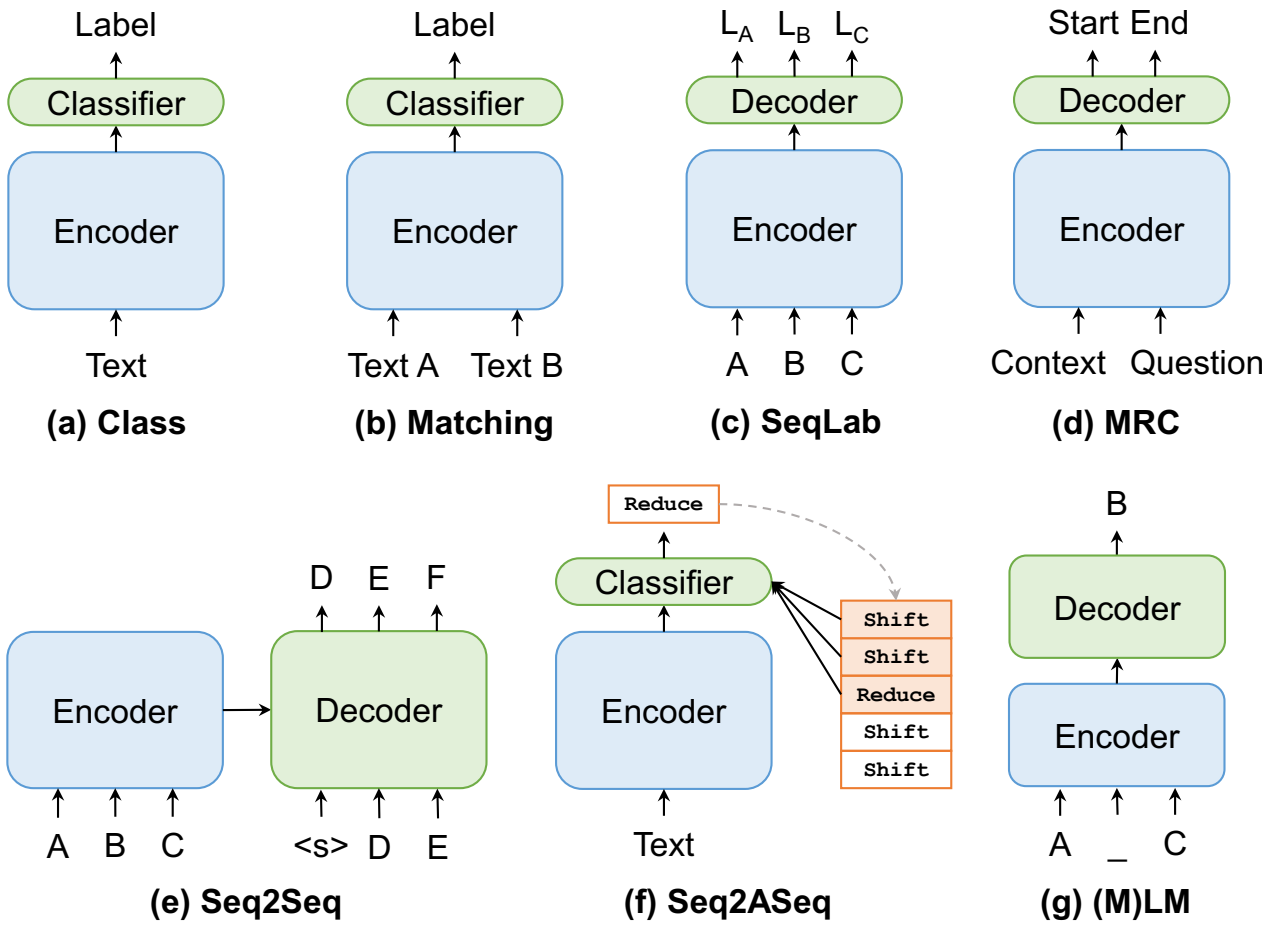
从分类到生成

- Label verbalize 标签描述

Solving classification as a generation task (T5)

Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer

NLP研究的范式转移



Paradigm Shift in Natural Language Processing

Prompt Engineering

Instruction based learning (priming)

en → fr translation

```
Translate English to French:

sea otter => loutre de mer

peppermint => menthe poivrée

plush girafe => girafe peluche

cheese => .....
```

Template based learning

sentiment classification

```
Best pizza ever! It was .....
```

great | bad

topic classification

```
..... News: OpenAI presents a
new model!
```

World | Sports | Tech

textual entailment

```
It's snowing. ....., it's cold.
```

Yes | Maybe | No

Proxy-task based learning

emotion classification

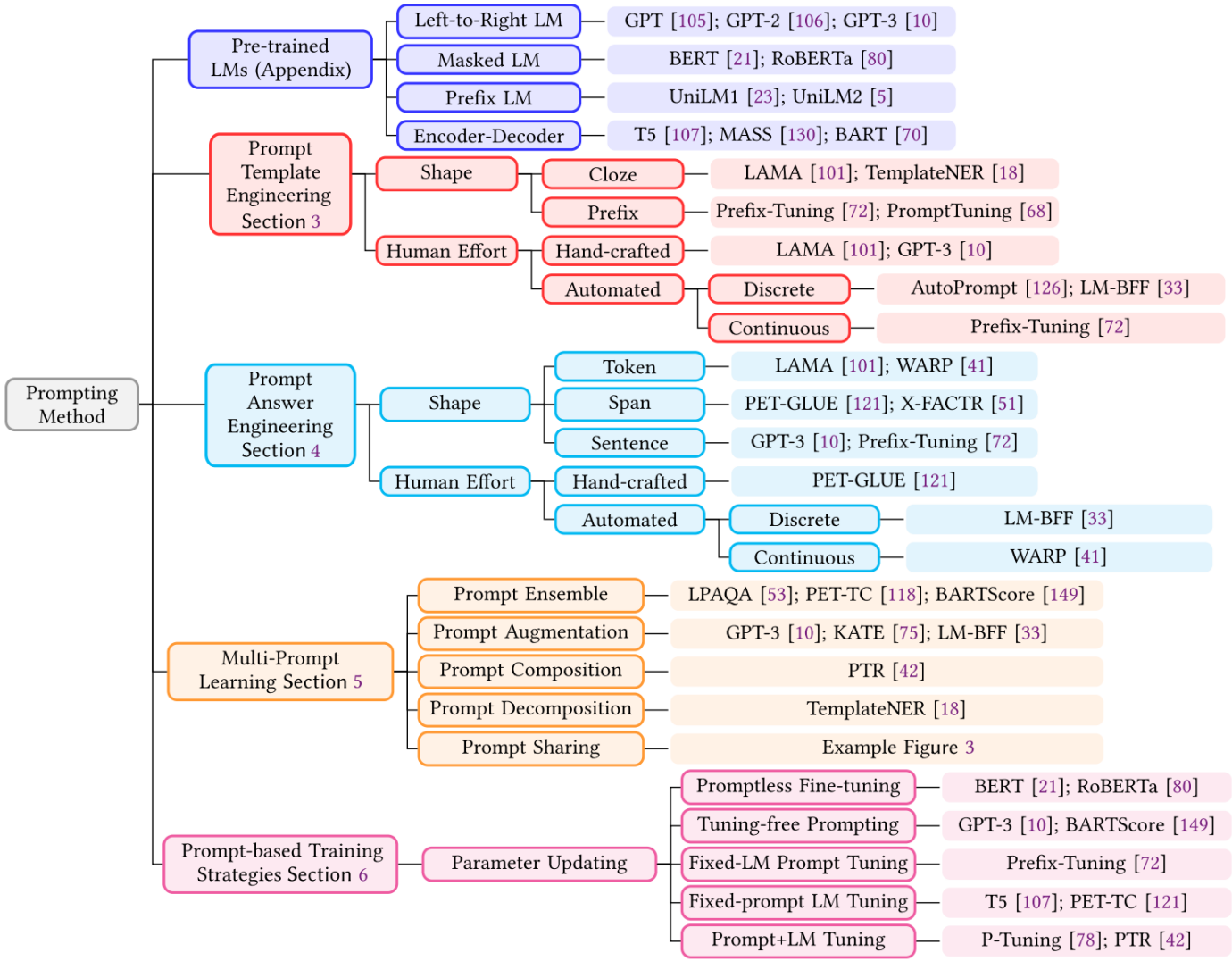
```
premise: I am feeling grouchy.
hypotheses:
        It expresses love.
        It expresses anger.
        It expresses sadness.
```

event argument-extraction

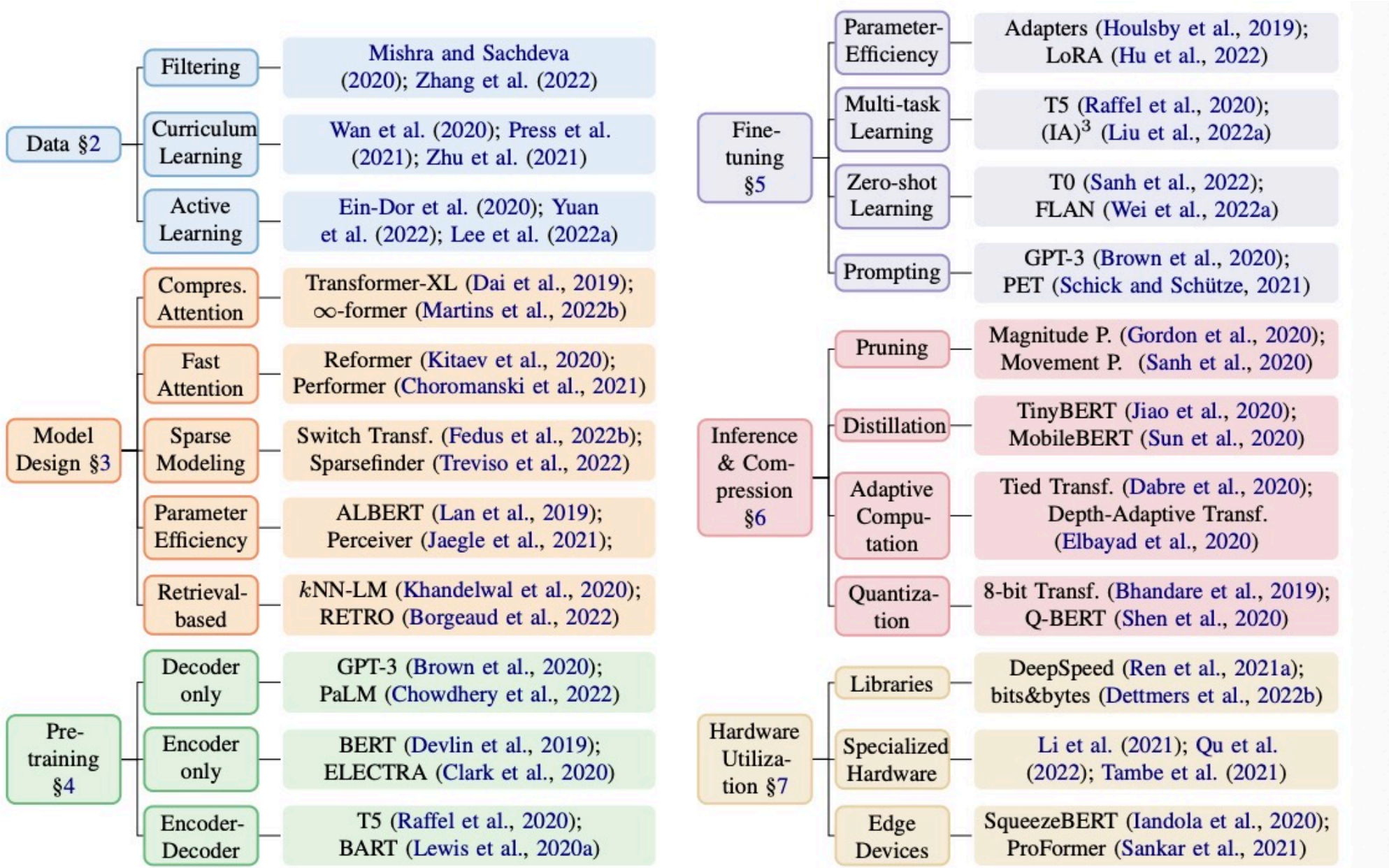
```
C: China has purchased two
nuclear submarines from Russia
last month.

Q: Who bought something?
A: China
Q: What is bought?
A: Two nuclear submarines.
```

Recent Advances in Natural Language Processing via Large Pre-trained Language Models: A Survey

Prompting Method



Pre-train, Prompt, and Predict: A Systematic Survey of Prompting Methods in Natural Language Processing



Efficient Methods for Natural Language Processing: A Survey

Conference and journals

- AAAI
- IJCAI
- ICML
- NeurIPS
- ACL

- ICLR
- WWW
- SIGIR

- EMNLP
- COLING
- NAACL

- NLPCC
- CONLL

- TASLP
- TAC

- KBS
- NC
- NLE

Conference List
https://conferencelist.info/upcoming/

Deadline
https://ccfddl.github.io/

CoT

- **Tree of Thoughts: Deliberate Problem Solving with Large Language Models**
- https://arxiv.org/abs/2305.10601
- **Automatic Chain of Thought Prompting in Large Language Models**
- https://arxiv.org/abs/2210.03493
- **Chain-of-Thought Prompting Elicits Reasoning in Large Language Models**
- https://arxiv.org/abs/2201.11903
- **Self-Consistency Improves Chain of Thought Reasoning in Language Models**
- https://arxiv.org/abs/2203.11171
- **Large Language Models Are Human-Level Prompt Engineers**
- https://arxiv.org/abs/2211.01910
- **Distilling Reasoning Capabilities into Smaller Language Models**
- https://arxiv.org/abs/2212.00193
- **Analysing Mathematical Reasoning Abilities of Neural Models**
- https://arxiv.org/abs/1904.01557
- **Large Language Models are Zero-Shot Reasoners**
- https://arxiv.org/abs/2205.11916
- **Selection-Inference: Exploiting Large Language Models for Interpretable Logical Reasoning**
- https://arxiv.org/abs/2205.09712
- **ART: Automatic multi-step reasoning and tool-use for large language models**
- https://arxiv.org/abs/2303.09014

Baseline
https://github.com/kyegomez/tree-of-thoughts