**Word Embedding (GloVe, Word2Vec)**

Corpus: [Doc]. Token. Vocabulary.

One-hot: sparse, high dim. OOV->UNK(sparse,概率0)  
WordNet: 同义词Synsets集合; L2：距离. Cosine：角度

Word2Vec: Arch: CBOW/SkipGram. Train: NegSample  
CBOW:. SkipGram:   
SG:   
SG: . 计算 Softmax 分母很贵  
NegSample:   
Noise: Freq

TFIDF: 文档，, 单词

Classification

Indep assum:   
Optimise: binary naive bayes. ControlOfNeg: NOT\_like.

RNN:   
CNN: 假设文本是 . Conv2d: , 滑动窗口

CondInd not implies Ind:   
Ind NOT implies CondInd   
Language Modeling: NGram   
Language prob 不包含<s>,包含</s>

CrossEntropy

*Intrinsic Eval: 内部任务 ExtrinsicEval: 下游任务*

Add-1 Smoothing: . Stupid Back-off:   
Interpolation:

也可以使用 FFLM(FeedForwardLM)

BiRNN: , Classification: . Modeling . MultiLayer: .

LSTM: GRU:

Encoder-Decoder (MT): NaiveBiRNN:

MLPAttention:

A table with numbers and letters

AI-generated content may be incorrect.Evaluation: MP-N:假设有MT 和参考 。先分别计算的n-gram.

Chr-F: 字符n-gram 。  
TER: Translation Error Rate: 将MT转为Ref的最少编辑数  
ROUGE-: Ref & MT’s ngram FScore. BERT Score  
ROUGE-L: LCS, Longest Common Subsequence (Underline)  
Hyp: "The cat and the dog" Ref: "The cat is on the mat"  
Precision = |LCS|/|Hyp| = 3/5. Recall = |LCS| / |Ref| = 3/6  
METEOR: R的权重是P的9倍。考虑stemming & synonymy。Chunk: 理想情况 Hyp=1Chunk=Ref. 多个不连续Chunk坏。

Sample: Greedy Decoding, Beam Search, Temperature Sample:

DA: Backtranslation: 训练. 把tgt翻译成src。这样有两份src，同语意，不同结构。Synonym Replacement: Dictionary/Nearest Noughbour&Emd

经验：将相似长度分组在同batch中。先用更简单/更短的seq len训练。（避免pad浪费）

Attention:   
LayerNorm: .   
PointwiseFFN [SOS][EOS]

Sinusoid Emb:

位置 的 或 个分量。

EncBlock:

CrossAttention:

DecBlock:

ElMo(LSTM):

**EncoderOnly:** MLM, NSP. BERTbase[L:12, H:768,h=12,110M], BERT-large[L:24, H:1024,h=16,340M][CLS]. Token Labeling;Sentence Pair Classification;QA. RoBERTa: NoNSP+MoreData. DeBERTa: BetterPE+RBT+MoreData

SpanBERT: Mask Span Not Token; DistilBERT/ALBERT. Big Bird and LongFormer: Sparse Attention; ClinicalBert, MedBert, PubMedBert, BEHRT:DSL. ERNIE: KG. ImageBERT:img as token: MLM, MObjM, PredFeatMaskObj, ClsImgTxtRlted. ViLBERT:FastrRNN+2 BERT. MAE.

ED: BART: TokenMask,SentencePermutation,DocRotation,TextInfilling, TokenDetection. **Replace corrupted spans**:   
Instructional tr(FLAN-T5): Template(premise,hyp)→Option

Dec: zero/one/few-shot. GPT FT (分类器): [START]s1[DELIM]s2[EXTRACT]  
ChainOfThought,RAG,RLHF: 专家生成数据集进行FT。收集对比数据，训练奖励模型。模型回答，依据奖励模型RL。 Hallucinations/Bias/Jailbreaking/MaliciousUse/ self explain.  
Prompt tuning: 引入特殊Token，FT这些TokenEmbd其余冻结。  
Prefix tune: 在每一层的输入前添加可训练的特定任务"前缀"（tokens）。  
Adapters: 插入可训练层。BitFit: 只训练 bias. LoRA.

POS Tagging SOTA UD Ensemble:

HMM : : States. 。Viterbi/HMM。MEMM:

Constituency:CKY.

. CNF:

. Discard LHS: Flatten. Otherwise: \*P (wo NNP). Probabilistic Lexicalised CFG

DependencyPrs: structure DAG . Tree: 1ROOT no income. only 1 income. Uniq path from ROOT to , no cycle.  
Projective:deps no cross. Arcs. ShiftRdc: **MaltParser** (fast, not acc),Spanning: MSTParser(slow, acc). oper: shift,larc, rarc

*A diagram of a tree

AI-generated content may be incorrect.*

A diagram of a diagram

AI-generated content may be incorrect.

A diagram of a multi-head attention

AI-generated content may be incorrect.

A diagram of a process

AI-generated content may be incorrect.

A diagram of a diagram

AI-generated content may be incorrect.