25HS_NLP_List

Lecture Notes

目录 / Contents

1 Core Concept Checklist

L6. CRF

Topic	Key Point
Semiring 定义	五元组 + 4 条公理，distributivity 最关键
常见 semirings	Boolean/Real/Tropical/Viterbi/Log 的 $\oplus, \otimes, \mathbf{0}, \mathbf{1}$
代数推导 DP	Distributivity 将 $O(\|T\|^N)$ 变 $O(N\|T\|^2)$
3-gram 依赖	complexity 变 $O(N\|T\|^3)$——多一层循环
Forward/Backward	同一 algo，换 semiring 换语义
Viterbi	Backward + Viterbi semiring + backpointers
Log-sum-exp	$\log(e^x + e^y) = x + \log(1 + e^{\{y-x\}})$
Idempotency	$a \oplus a = a$；max/min 满足，+ 不满足
Structured Perceptron	CRF + $T \to 0$ (hard max)
WFSA path sum	有 cycle 可能发散；收敛条件类 geometric series
为何 transliteration 不用 CRF	对齐非一对一
FST composition	类 matrix multiplication + marginalization

l7. WFSA

Topic	Key Point
Topological sort	Acyclic graph 可做；非唯一；$O(\|E\|)$
Closed semiring	有 Kleene star 满足两条公理
Geometric series 公理化	$a^* = 1 + a \cdot a^* = 1 + a^* \cdot a$
Matrix 收敛条件	Largest eigenvalue < 1
Lehmann's algorithm	$O(\|Q\|^3)$; 通用 cyclic WFSA path sum
Floyd-Warshall 为何无 star	Shortest path 不走 cycle，$a^* = 0$
Kleene's algo	Lehmann + regex semiring = FSA $\to$ regex
Constituent	作为 single unit 的 word group
Syntactic ambiguity	同一 string 多个 parse trees
CFG 四元组	$\langle \mathcal{N}, S, \Sigma, \mathcal{R} \rangle$
为何叫 context-free	Rule 应用不依赖 context
CFG 是 model	非 ground truth；解释 data 的工具
PCFG	Locally normalized over RHS
WCFG Z 可能发散	Cyclic rules 导致 infinite trees
CNF	Binary branching + terminal emission
CNF 保证	Finite trees；fixed tree size $2N - 1$
Catalan number	$O(4^N)$ trees for length-N string

l8. CKY

Topic	Key Point
CKY 命名	Cocke-Kasami-Younger，1960s 独立发明
CKY complexity	$O(N^3 \ \|\mathcal{R}\|)$
为何需要 CNF	Binary branching $\to$ 固定 N^3；otherwise $N^{\{k+1\}}$

CKY 的 topological order	按 span length 递增；同 length 内任意
Earley's contribution	对任意 CFG（非 CNF）也能 $O(N^3 \ \|G\|)$
Dependency tree 定义	Directed spanning tree + root constraint
Projective vs non-projective	Arcs 是否 crossing
Edge factorization	Score 分解到 edges；最强的 tractable assumption
为何不能 second-order	Can encode Hamiltonian path (NP-hard)
Matrix-Tree Theorem	$Z = \det(\boldsymbol{L}), O(N^3)$
为何不能 semiringify MTT	需要 subtraction
Kruskal 为何不 work	Directed case 有 incoming degree constraints
CLE algorithm 核心	Greedy + cycle contraction + reweighting
CLE complexity	$O(N^2)$ (Tarjan's version)

l9. CCG

Topic	**Key Point**
Truth-conditional semantics	Meaning = 何时为 true
Quantifier scope ambiguity	Semantic ambiguity（非 lexical/syntactic）
Logical form	可执行的 meaning representation
Compositionality	Complex meaning = function of parts
Lambda calculus terms	Variables, abstraction ($\lambda x.M$), application (MN)
Free vs bound	Bound = 在某 λ scope 内
α-conversion	Rename bound variable（不改 meaning）
β-reduction	$(\lambda x.M)N \to M[x := N]$（考试重点 !）
Non-termination	$\Omega = (\lambda x.xx)(\lambda x.xx)$
Combinators	$\boldsymbol{I}x = x$; $\boldsymbol{K}xy = x$; $\boldsymbol{S}xyz = xz(yz)$
CCG categories	Encode argument structure via slashes
CCG rules	Application (>, <), composition ($\boldsymbol{B}$), type-raising ($\boldsymbol{T}$)
CCG 优点	Tight syntax-semantics interface

lc.10 Trsf

1. **Attention** = soft hash table：$\boldsymbol{c} = \sum_i \text{softmax}(\text{score}(\boldsymbol{q}, \boldsymbol{k}_i))\boldsymbol{v}_i$
2. **Self-attention**：Q, K, V 来自同一序列，捕获 intra-sequence dependencies
3. **Transformer** = attention + positional encoding + residual + layer norm，**可并行**
4. **Decoding**：beam search（deterministic）或 sampling（stochastic），无 exact solution
5. Engineering-heavy：很多设计是"empirically works"，理论滞后

lc.11 Modelling Pipeline Checklist

1. **Problem Definition**：Classification? Structured prediction? data 特性?
2. **Model Choice**：Probabilistic vs deterministic? 有无 independence assumptions? Interpretability 需求?
3. **Loss Function**：MLE（默认）? Hinge（SVM）? 需满足 convexity + differentiability
4. **Regularization**：L1/L2? Dropout? Early stopping?
5. **Evaluation**：F1（分类）? Task-specific metric? Intrinsic vs extrinsic?
6. **Model Selection**：Cross-validation + statistical testing，注意 multiple testing correction

l12. Axes

Statistical Testing：

- NLP 常用 non-parametric tests（McNemar, permutation）
- 多重比较需 Bonferroni correction
- 5×2 CV t-test 解决 fold 依赖问题

Domain Adaptation：

- Importance sampling：用 density ratio 重加权
- Feature augmentation：shared + domain-specific features

Bias in NLP：

- 来源多样：data, task design, model, deployment
- Debiasing：identify bias subspace → neutralize → equalize
- Linear assumption 在实践中足够有效

Occam's Razor：在性能相近时，选择更简单的模型。

Tutorial 补充考点

Topic	Key Point
Monoid 判定	Closure + associativity + identity
Semiring 判定陷阱	$\mathbf{0} = \mathbf{1}$ 必失败；检查 distributivity 方向
非交换例子	String concat, language concat
Forward vs Backward 差异	初始化不同；Forward 需遍历最后一列
Dijkstra 失效	负权边；sum semiring 无加速
CKY chart 索引	$\boldsymbol{C}[i, k, X]$：span $[i, k)$, non-terminal X
CKY 遍历顺序	按 span length 递增；同 length 内任意
Catalan number	$C_N \approx \frac{4^N}{N^{\frac{3}{2}}}$：binary trees 数量
CNF 转换	消 ε, 消 unit, binary 化, 分离 terminals
CRF as CFG	Right-recursive grammar，$O(\lvert T\rvert^2)$ rules
Log semiring 用途	数值稳定；用 `logsumexp`
Masking 技巧	Padding tokens 需 mask 掉