

## Motivation

- ◆ To capture long-range dependencies in the label sequence;
- ◆ Globally normalized models overcome label bias and exposure bias problems.

### Label bias :

“successors of wrong histories receive the same mass as do the successors of the correct history” [1]

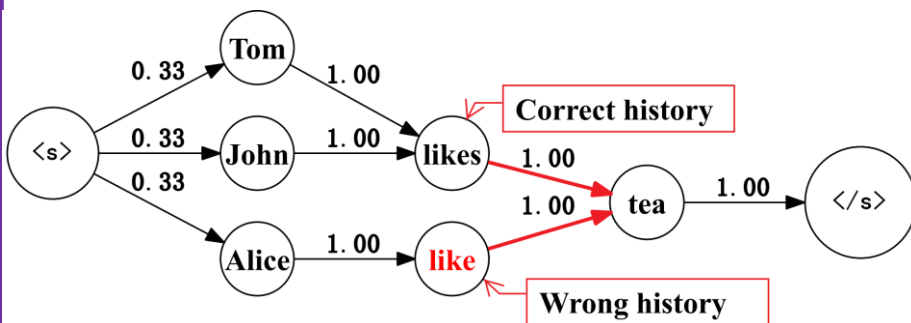
### Exposure bias :

“Only exposed to the training data, instead of its own predictions” [1]

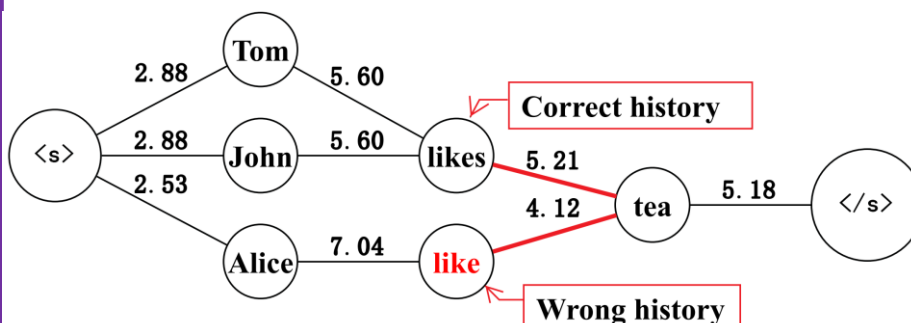
### Training data :

Tom likes tea  
John likes tea  
Alice like tea

### Locally normalized model



### Globally normalized model

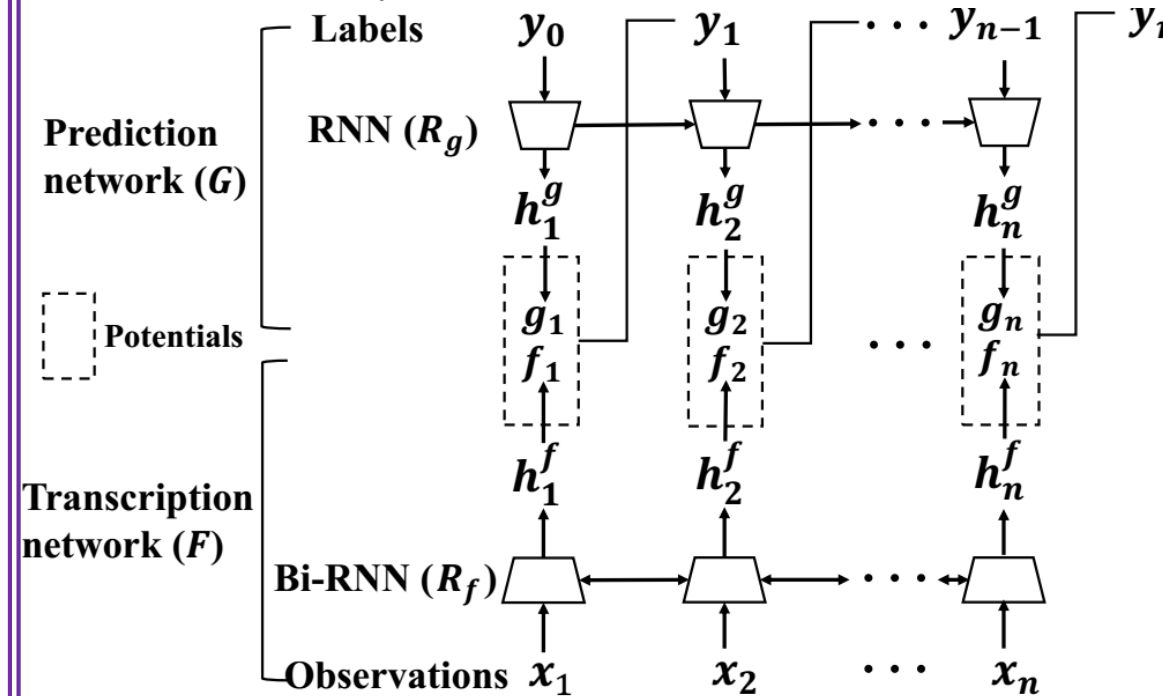


## Model Definition

$$p(y|x; \theta) = \frac{\exp\{u(y, x; \theta)\}}{Z(x; \theta)}$$

$$Z(x; \theta) = \sum_{y'} \exp\{u(y', x; \theta)\}$$

$$u(y, x; \theta) = \sum_{i=1}^n \{\phi_i(y_i, x; \theta) + \psi_i(y_{0:i-1}, y_i; \theta)\}$$



## Potential Design : two choices

$$\phi_i(y_i = k, x; \theta) = f_i^k$$

$$\psi_i(y_{0:i-1}, y_i = k; \theta) = g_i^k$$

$$\phi_i(y_i = k, x; \theta) = [\text{logsoftmax}(f_i)]^k$$

$$\psi_i(y_{0:i-1}, y_i = k; \theta) = [\text{logsoftmax}(g_i)]^k$$

## Training and Decoding

$$L(y^*; \theta) = -u(y^*, x; \theta) + \log Z(x; \theta)$$

### Monte Carlo Method

$$\nabla_{\theta} \log Z(x; \theta) = E_{p(y'|x; \theta)} [\nabla_{\theta} u(y', x; \theta)]$$

### Beam search with early updates [2]

$$L(y_{1:j}^*; \theta) = -u(y_{1:j}^*; \theta) + \log \sum_{y' \in \beta_j} \exp\{u(y'_{1:j}; \theta)\}$$

### Decoding with beam search

## Evaluation

Model \ Task	POS (Accuracy)	Chunking (F1 score)	English NER (F1 score)	Dutch NER (F1 score)	Globally normalized	Long-range dependencies
Linear-chain NCRF	97.52	95.01	91.11	81.53	✓	✗
RNN Transducer [3]	97.50	95.02	91.02	81.59	✗	✓
<b>NCRF Transducer</b>	97.52	<b>95.14</b>	<b>91.40</b>	<b>81.84</b>	<b>✓</b>	<b>✓</b>

Reference:

- [1] Wiseman and Rush, “Sequence-to-sequence Learning as Beam-Search Optimization”, EMNLP 2016.
- [2] Andor, Alberti, et al., “Globally Normalized Transition-Based Neural Networks”, ACL 2016.
- [3] Graves, “Sequence Transduction with Recurrent Neural Networks”, ICML workshop 2012.