Conditional Random Field (CRF) Model

Problem with Maximum Entropy Models

Per-state normalization:

All the mass that arrives at a state must be distributed among the possible successor states

Problem with Maximum Entropy Models

Per-state normalization:

All the mass that arrives at a state must be distributed among the possible successor states

This gives a ‘label bias’ problem:

Let’s see the intuition

What do you mean by normalization at each state

Noun $e^1/2e^1$	**Adjective** $e^2/e^1 + e^2$		**Adjective**
	Verb	**Det**	**Verb**
The	*Light*	*The*	*Light*

Noun-Verb

Noun-Adjective

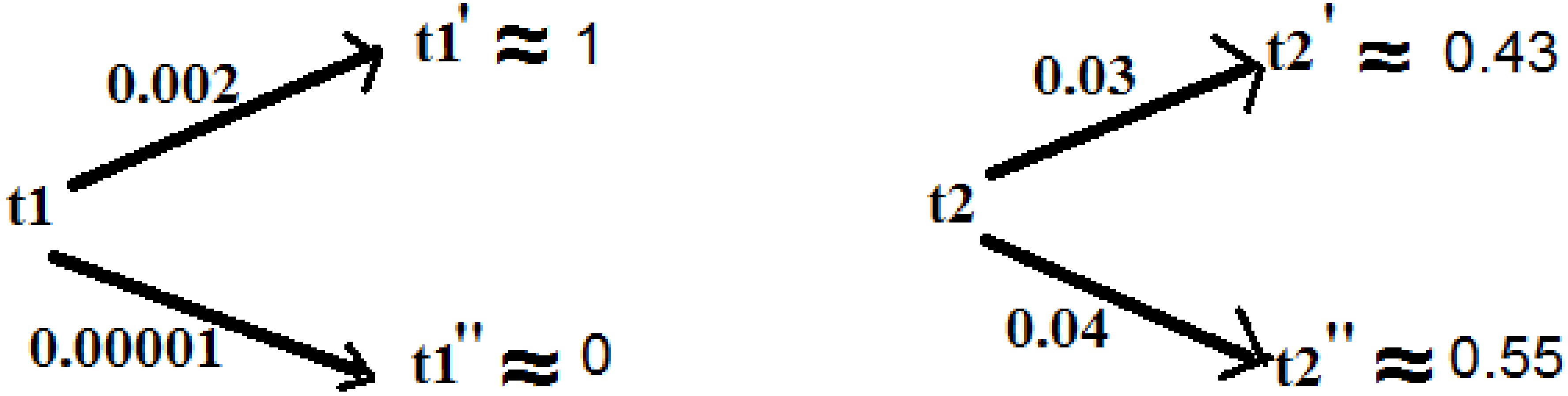
Similarly,

Det – Adjective

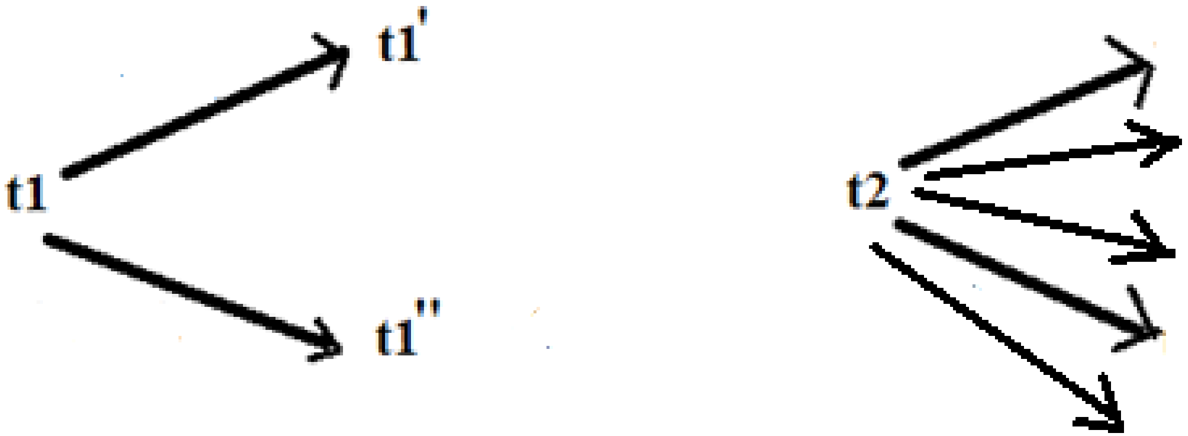
Det - Verb

Hypothetical Example

Hypothetical Example



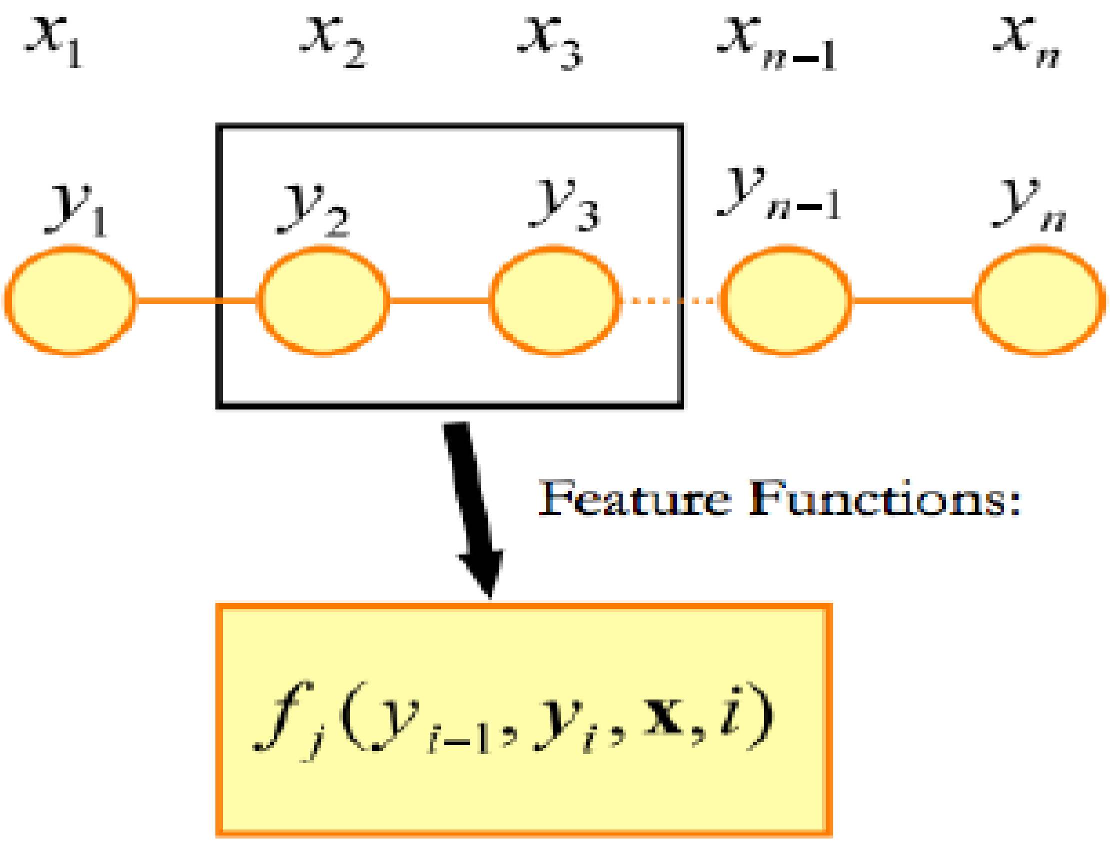
Problem of MEMM



CRF

- CRFs are conditionally trained, undirected graphical models.
- Let’s look at the linear chain structure

CRF: Feature Functions



Feature functions

Express some characteristics of the empirical distribution that we wish to hold in model distribution

$$f_j(y_{i-1}, y_i, \mathbf{x}, i) = \begin{cases} 1 & \text{if } y_{i-1} = IN \text{ and } \\ & \quad y_i = NNP \text{ and } \\ & \quad x_i = September \\ 0 & \text{otherwise} \end{cases}$$

Conditional Random Fields: Distribution

Label sequence modelled as normalized product of feature functions

$$P(\mathbf{y} \mid \mathbf{x}, \lambda) = \frac{1}{Z(\mathbf{x})} \exp \sum_{i=1}^{n} \sum_{j} \lambda_j f_j(y_{i-1}, y_i, \mathbf{x}, i)$$

$$Z(\mathbf{x}) = \sum_{\mathbf{y} \in Y} \sum_{i=1}^{n} \sum_{j} \lambda_j f_j(y_{i-1}, y_i, \mathbf{x}, i)$$

CRFs

1. Have the advantages of MEMM but avoid the label bias problem
2. CRFs are globally normalized, whereas MEMMs are locally normalized.
3. Widely used and applied. CRFs have been (close to) state-of-the-art in many sequence labeling tasks.