Markov Models

notes for

CSCI-GA.2590

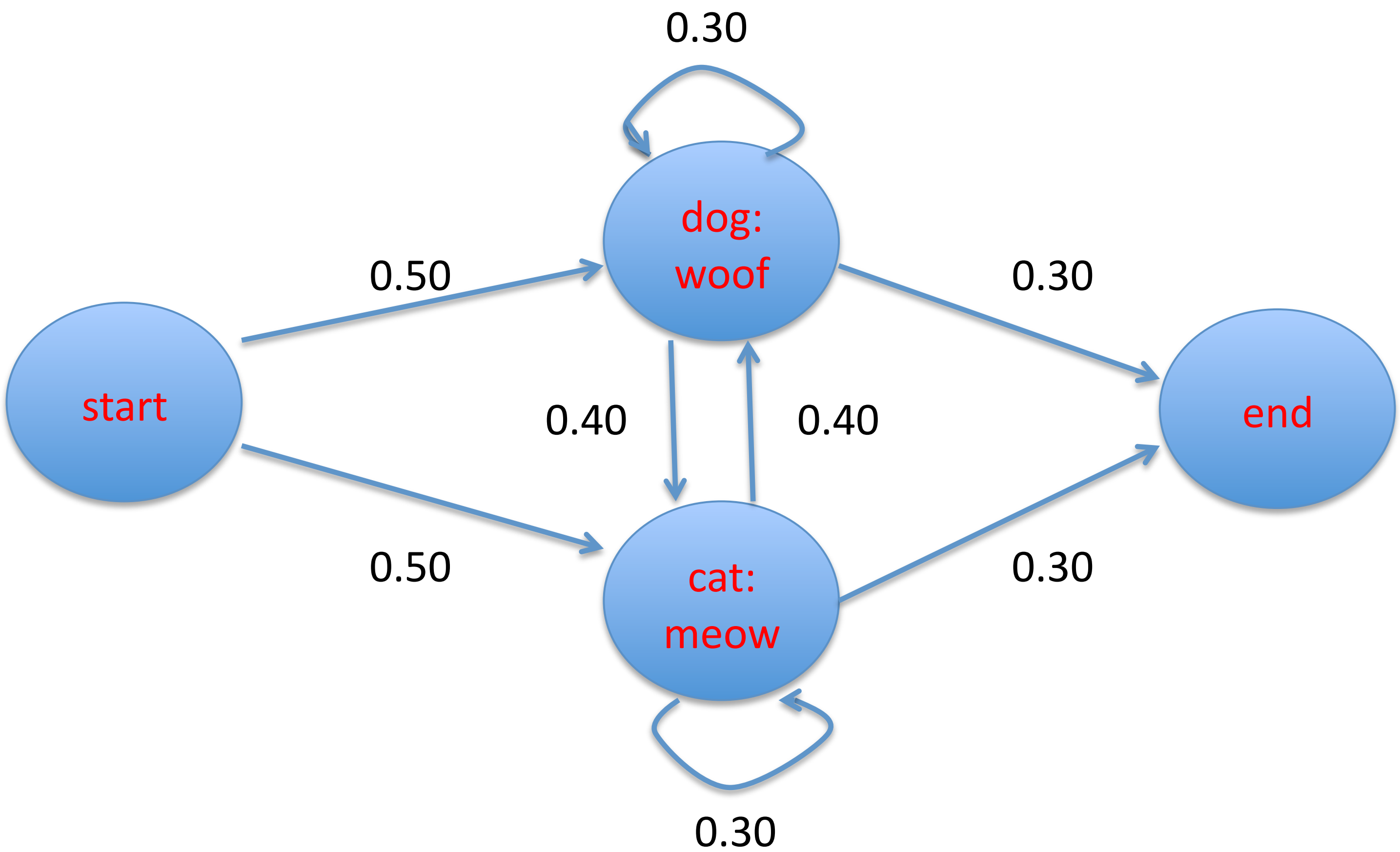
Prof. Grishman

Markov Model

- In principle each decision could depend on all the decisions which came before (the tags on all preceding words in the sentence)

- But we'll make life simple by assuming that the decision depends on only the immediately preceding decision
  - [first-order] Markov Model
  - representable by a finite state transition network
  - T_{ij} = probability of a transition from state i to state j

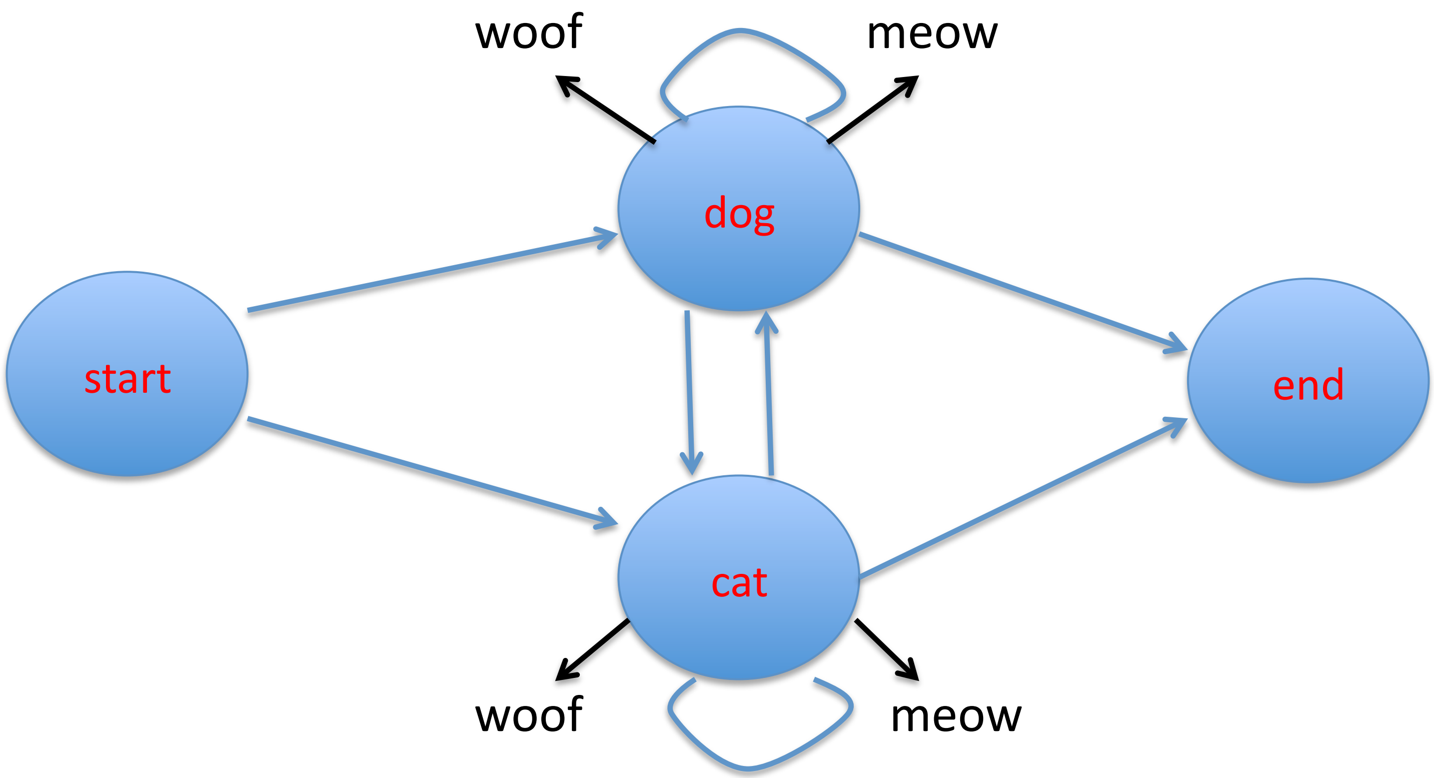
Finite State Network



Our bilingual pets

- Suppose our cat learned to say “woof” and our dog “meow”
- … they started chatting in the next room
- … and we wanted to know who said what

Hidden State Network



- How do we predict
    - When the cat is talking: t_i = cat
    - When the dog is talking: t_i = dog

- We construct a probabilistic model of the phenomenon

- And then seek the most likely state sequence S

$$S = \underset{t_1 \ldots t_n}{\arg\max} \, P(t_1 \ldots t_n \mid w_1 \ldots w_n)$$

Hidden Markov Model

- Assume current word depends only on current tag

$$S = \underset{t_1 \ldots t_n}{\arg\max}\, P(t_1 \ldots t_n \mid w_1 \ldots w_n)$$

$$= \underset{t_1 \ldots t_n}{\arg\max}\, P(w_1, \ldots, w_n \mid t_1, \ldots, t_n) P(t_1, \ldots, t_n)$$

$$= \underset{t_1 \ldots t_n}{\arg\max} \prod_{i=1}^{n} P(w_i \mid t_i) P(t_i \mid t_{i-1})$$