CMPT 413: Computational Linguistics

HMM1: Introducing Hidden Markov Models

Anoop Sarkar

http://www.cs.sfu.ca/~anoop

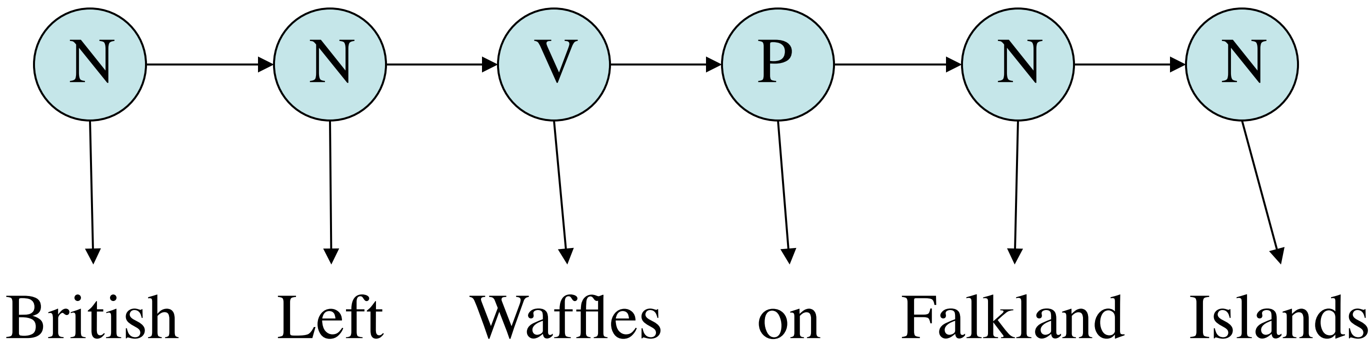
Sequence Learning / Tagging

- British Left Waffles on Falkland Islands
  - (N, N, V, P, N, N) *parts of speech: N, V, P*
  - (N, V, N, P, N, N)
- Segmentation 中国十四个边境开放城市经济建设成就显著
  - (b, i, b, i, b, b, i, b, i, b, i, b, i, b, i, b, i, b, i)

中国	十四	个	边境	开放	城市	经济	建设	成就	显著

China 's 14 open border cities marked economic achievements

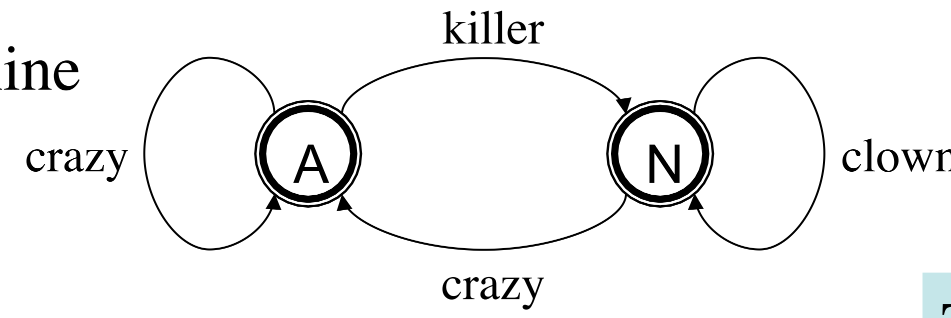
Sequence Learning



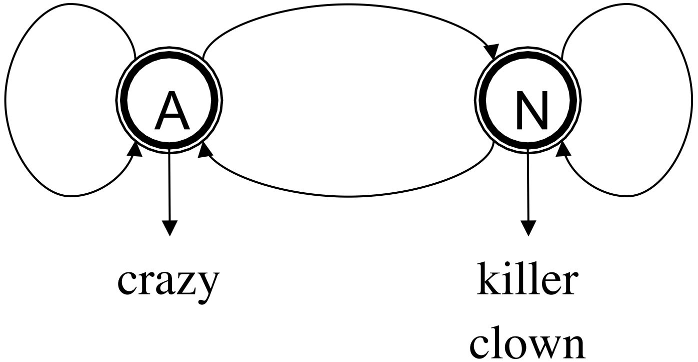
3 states: N, V, P
Observation sequence: (o_1, … o_6)
State sequence (6): (N, N, V, P, N, N)

Finite State Machines

Mealy Machine



Moore Machine

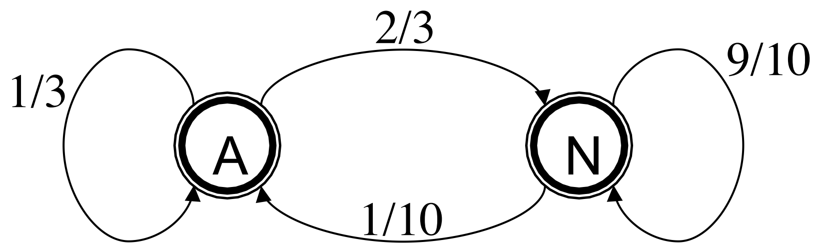


These two representations are equivalent

Probabilistic FSMs

- Start at a state *i* with a *start state probability*: π_i
- Transition from state *i* to state *j* is associated with a *transition probability*: $a_{ij} = P(j|i)$
- Emission of symbol *o* from state *i* is associated with an *emission probability*: $b_i(o) = P(o|i)$
- Two conditions:
  - All outgoing transition arcs from a state must sum to 1
  - All symbol emissions from a state must sum to 1

Probabilistic FSMs



0 killer	1/3 killer
1.0 crazy	0 crazy
0 clown	1/3 clown
0 problem	1/3 problem

Probabilistic FSMs

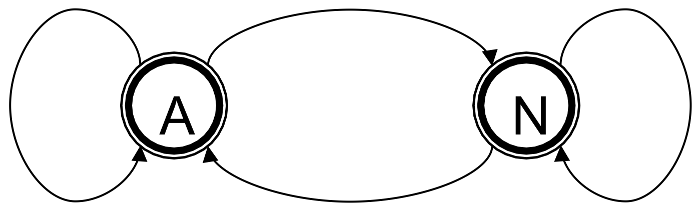
<u>Emission</u>

$b_A(\text{killer}) = 0$

$b_A(\text{crazy}) = 1$

$b_A(\text{clown}) = 0$

$b_A(\text{problem}) = 0$



killer	killer
crazy	crazy
clown	clown
problem	problem

<u>Emission</u>

$b_N(\text{killer}) = 1/3$

$b_N(\text{crazy}) = 0$

$b_N(\text{clown}) = 1/3$

$b_N(\text{problem}) = 1/3$

$$\sum_{o \in V} b_i(o) = 1$$

$$\sum_i \pi_i = 1$$

<u>Start state</u>

$\pi_A = 1/2$

$\pi_N = 1/2$

<u>Transition</u>

$a_{A,A} = 1/3$

$a_{A,N} = 2/3$

$a_{N,N} = 9/10$

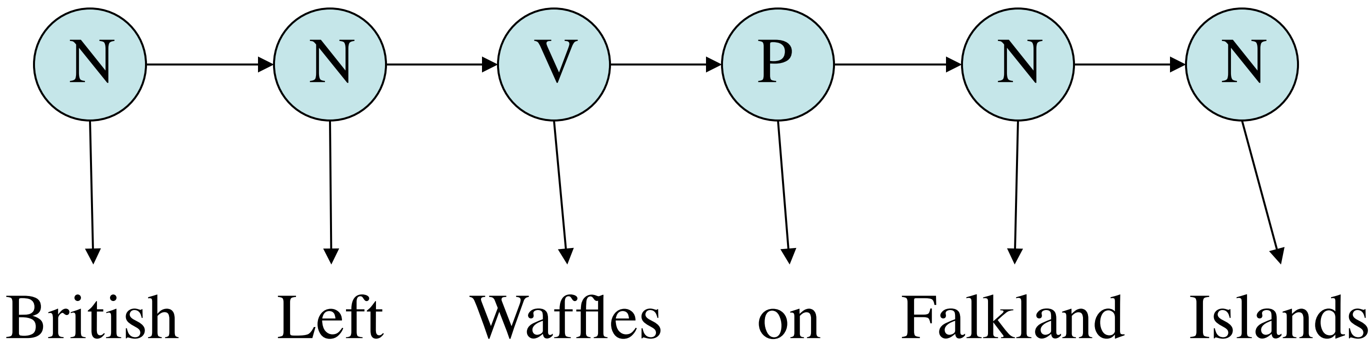
$a_{N,A} = 1/10$

$$\sum_j a_{i,j} = 1$$

Hidden Markov Models

- There are n states $s_1, \ldots, s_n$
- The emissions are observed (input data)
- Observation sequence $\mathbf{O}=(o_1, \ldots, o_T)$
- State sequence $\mathbf{X}=(X_1, \ldots, X_T)$
- Each X_i is one of the n states
- The states are not directly observed (hidden)

Sequence Learning



3 states: N, V, P
Observation sequence: $(o_1, \ldots o_6)$
State sequence (6): (N, N, V, P, N, N)

HMM Algorithms

- HMM as parser: *compute the best sequence of states for a given observation sequence*
- HMM as language model: *compute probability of given observation sequence*
- HMM as learner: *given a set of observation sequences, learn its distribution*, i.e. learn start, transition and emission probabilities

CMPT 413: Computational Linguistics

HMM1: Introducing Hidden Markov Models

Anoop Sarkar

http://www.cs.sfu.ca/~anoop