Label Bias Problem

Training Data

X:Y

rib:123

rib:123

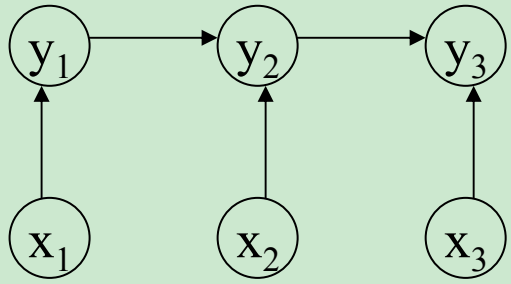
rib:123

rob:456

rob:456

Parameters

$$P(1 \mid r) = 0.6,\ P(4 \mid r) = 0.4,$$
$$P(2 \mid i,1) = P(2 \mid o,1) = 1,$$
$$P(5 \mid i,4) = P(5 \mid o,4) = 1,$$
$$P(3 \mid b,2) = P(6 \mid b,5) = 1$$



$$\sum_{\text{all states s reachable from 1}} p(s|o,1) = 1$$

In the training data, label value 2 is the only label value observed after label value 1
Therefore **P(2 | 1) = 1, so P(2 |x, 1) = 1 for all x**

$$P(s \mid s',o) = \frac{\exp(\sum_k \lambda_k f_k(o,s',s))}{\sum_s \exp(\sum_k \lambda_k f_k(o,s',s))} = \frac{\exp(\sum_k \lambda_k f_k(o,s',s))}{\exp(\sum_k \lambda_k f_k(o,s',s))} = 1$$

Single outgoing

Label Bias Problem

Training Data

X:Y

rib:123

rib:123

rib:123

rob:456

rob:456

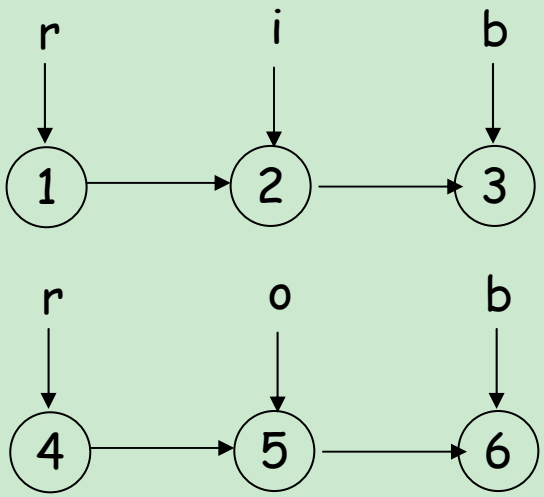
New input: rob

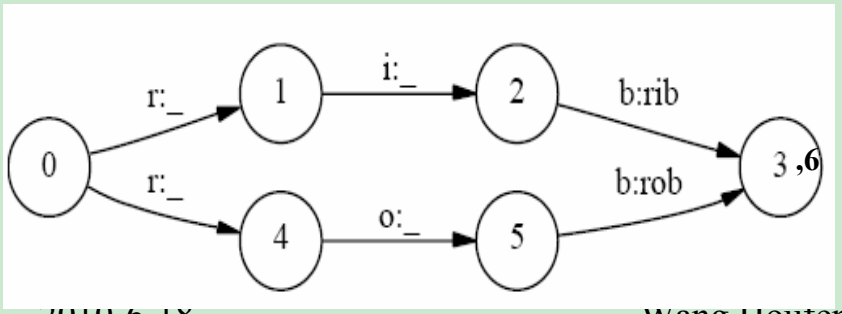
Parameters

$$P(1 \mid r) = 0.6,\ P(4 \mid r) = 0.4,$$
$$P(2 \mid i,1) = P(2 \mid o,1) = 1,$$
$$P(5 \mid i,4) = P(5 \mid o,4) = 1,$$
$$P(3 \mid b,2) = P(6 \mid b,5) = 1$$

$$P(123 \mid \text{rob}) = P(1 \mid r)P(2 \mid o,1)P(3 \mid b,2)$$
$$= 0.6 \times 1 \times 1 = 0.6$$
$$P(456 \mid \text{rob}) = P(4 \mid r)P(5 \mid o,4)P(6 \mid b,5)$$
$$= 0.4 \times 1 \times 1 = 0.4$$

Model





$$\sum_{\text{all states s reachable from 1}} p(s|o,1) = 1$$