KU KONKUK UNIVERSITY

Statistical Models

HMM, MEMM, and CRFs

자연어처리연구실

PART-I

HMM (Hidden Markov Model)

- What is the HMM?
  - Hidden + Markov Model
- Markov Model: Example
  - Training

		Tomorrow State (내일 상태)		
		Rain	Cloudy	Sunny
Today State (오늘 상태)	Rain	0.4	0.3	0.3
	Cloudy	0.2	0.6	0.2
	Sunny	0.1	0.1	0.8

  - Assumption: Tomorrow weather depends only on today one.
  - Problem: P(Rain, Rain, Sunny, Cloudy)?

마코프 모델 (Markov Model)

- Markov Model: Sequence Probability
  - 결합 확률(joint probability) 계산 모델
  - 연쇄 규칙과 마코프 가정을 이용하여 결합 확률을 단순화하여 결합 확률을 근사화시키는 모델

$$P(y_1, y_2, \dots, y_t)$$

$$= P(y_1)P(y_2|y_1)P(y_3|y_1,y_2) \dots P(y_t|y_1,y_2,\dots,y_{t-1})$$

Chain Rule

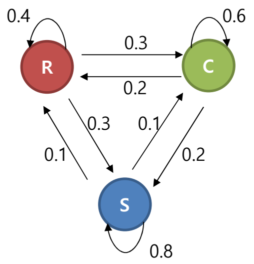
$$= P(y_1)P(y_2|y_1)P(y_3|y_2) \dots P(y_t|y_{t-1})$$

1'st-order Markov Assumption

마코프 모델 (Markov Model)

마코프 모델: 학습데이터로 부터 얻어진 전이 확률분포 (상태 간 이동 확률 분포)

		Tomorrow State (내일 상태)		
		Rain	Cloudy	Sunny
Today State (오늘 상태)	Rain	0.4	0.3	0.3
	Cloudy	0.2	0.6	0.2
	Sunny	0.1	0.1	0.8



State = $\{S_1: Rain, S_2: Cloudy, S_3: Sunny\}$

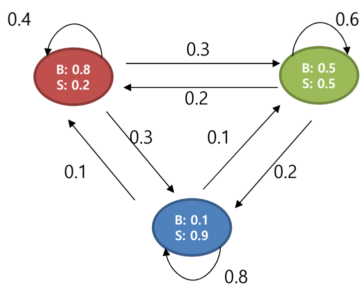
$P(S_1, S_1, S_3, S_2|model)$
$=P(S_1)P(S_1|S_1)P(S_3|S_1)\ P(S_2|S_3)$
=1 * 0.4 * 0.3 * 0.1
=0.012

What is Hidden?

- Hidden Markov Model
  - 상태(풀고자 하는 레이블)를 직접 관측할 수 없고 상태를 예측하는데 도움이 되는 특징(자질)만을 관측할 수 있음
  - 상태가 감춰져 있고(직접 관찰할 수 없고) 관측에 대한 확률로만 존재
  - 예제
    - 상태(State): Rain, Cloudy, Sunny
    - 관측(Observation): B (Rain Boots), S (Sports Shoes)
    - 문제(problem): P(B, B, S, S, Rain, Rain, Sunny, Cloudy)?

What is Hidden?

- 상태(State): Rain, Cloudy, Sunny
- 관측(Observation): B (Rain Boots), S (Sports Shoes)
- 문제(problem): “B, B, S, S”를 관측했을 때 날씨가 어떻게 예측하는게 최적일까?
  - P(Rain, Rain, Sunny, Cloudy) vs. P(Rain, Rain, Sunny, Sunny) vs. …



HMM (Hidden Markov Model)

Let $P(x_{1,t}) = P(x_1, x_2, \ldots, x_t)$

$\text{HMM} = \operatorname*{argmax}_{y_{1,t}} P(x_{1,t}, y_{1,t})$

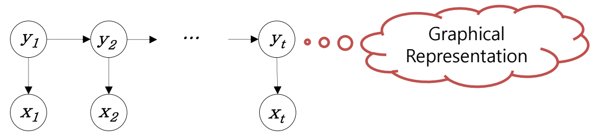
Chain Rule

$= \operatorname*{argmax}_{y_{1,t}} P(y_{1,t}) P(x_{1,t}|y_{1,t})$

$= \operatorname*{argmax}_{y_{1,t}} \prod_{i=1}^{t} \boxed{P(y_i|y_{i-1})} \boxed{P(x_i|y_i)}$

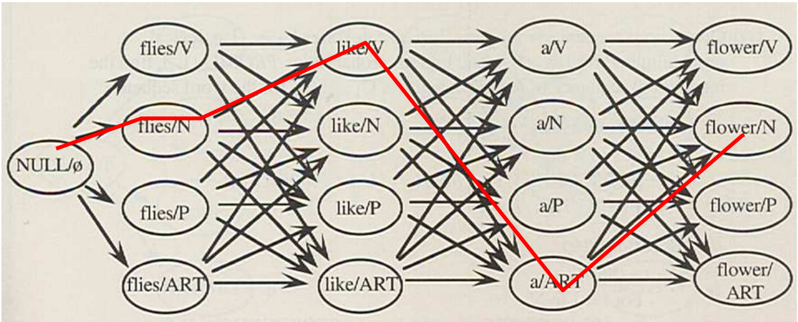
1'st-order Markov Assumption
→ 전이 확률 (Transition Probability)

Independent Assumption
→ 관측 확률 (Observation Probability)



Sequence Labeling Problem

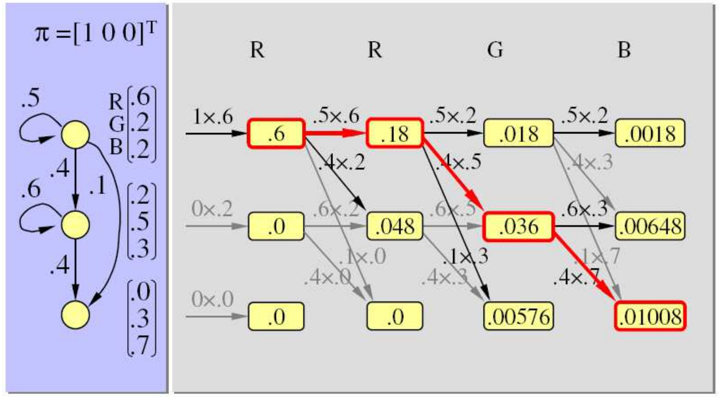
- Segmentation or path analysis problem
  - Application: Part-of-speech tagging



Viterbi Algorithm

- 모든 경로를 고려하지 않고도 빠른 시간 내에 최적의 경로를 찾는 알고리즘

Output: 1, 1, 2, 3



PART-II에 계속됩니다!