COMP 424 A4 TUTORIAL

Jade Yu
lei.yu@mail.mcgill.ca

Hidden Markov Model

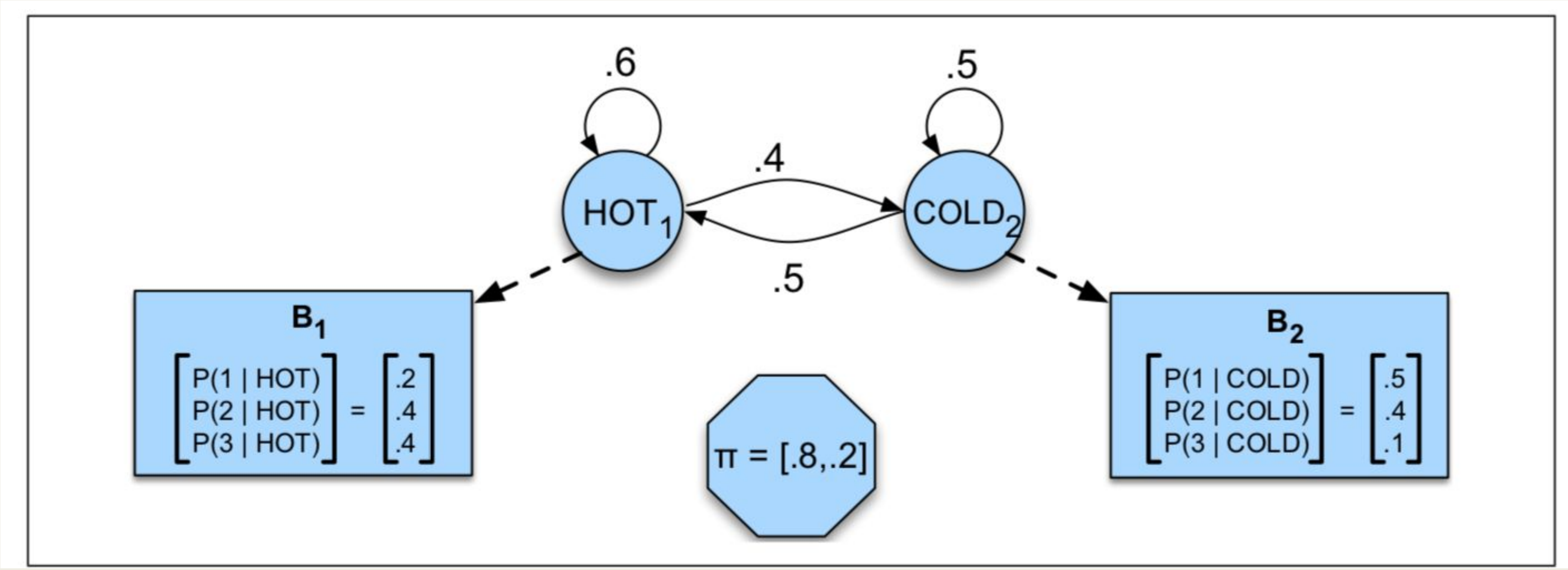
- Three fundamental problems for HMMs:
  - Likelihood: Given an HMM $\boldsymbol{\lambda} = (\boldsymbol{\pi}, \boldsymbol{A}, \boldsymbol{B})$ and an observed sequence O, find the probability $\boldsymbol{Pr}(\boldsymbol{O}|\boldsymbol{\lambda})$.
  - Decoding: Given an observation sequence O and HMM $\boldsymbol{\lambda} = (\boldsymbol{\pi}, \boldsymbol{A}, \boldsymbol{B})$, find the most probable sequence of hidden states: $\boldsymbol{Q}^* = \underset{\boldsymbol{Q}}{\operatorname{argmax}} \boldsymbol{Pr}(\boldsymbol{O}|\ \boldsymbol{\lambda}, \boldsymbol{Q})$
  - Learning: Given an observation sequence O, and sets of all possible hidden/observed states, find a best HMM model:

$$\boldsymbol{\lambda}^* = \boldsymbol{argmax}_{\lambda} \sum_{\boldsymbol{all\ possible\ hidden\ state\ seqs\ Q}} \boldsymbol{Pr}(\boldsymbol{O}|\ \boldsymbol{\lambda}, \boldsymbol{Q}) * \boldsymbol{Pr}(\boldsymbol{Q}|\boldsymbol{\lambda})$$

Problem	Algorithm
Likelihood	Forward Algorithm / Backward Algorithm
Decoding	Viterbi Algorithm (Dynamic Programming)
Learning	Baum-Welch Algorithm (EM)

A toy example by Eisner et al (2002):

- Imagine that you are a climatologist in the year 2799 studying the history of global warming. You cannot find any records of the weather in Baltimore, Maryland, for the summer of 2020, but you do find Jason Eisner's diary, which lists how many ice creams Jason ate every day that summer.
- Construct an HMM for this problem:



Let's address three fundamental problems through this example:

- **Likelihood:** What is the probability of the observed sequence of ice cream amounts '1 3 3' , given the parameters in the previous slide?
- **Decoding:** Given an observed sequence of ice cream amounts '1 3 3', what is the most likely sequence of weathers in 3 days, given the parameters in the previous slide?
- **Learning:** If now the parameters ($\lambda = \{\pi, A, B\}$) are unknown, and we observed a sequence of ice cream amounts '1 3 3', what is the best set of parameters for this model?

Utility Theory

What you should know after Jackie’s lecture:

- MEU principle
- Expected utility and maximum expected utility
- Value of information and value of perfect information

Example: St. Petersburg Paradox

- Consider a repeated lottery in which a fair coin (equal probability for heads and tails) is tossed repeatedly. For each round, the player has to pay α dollars at first, and will receive (2^k) dollars if the first head, say, occurs after k tosses of the coin.
  - What's the expected utility from playing this lottery if we assume that the utility always equals to the expected payoff?
  - Is our assumption reasonable?

Example: St. Petersburg Paradox

- Bernoulli’s proposition to address the paradox: people are risk aversive: our utility function is sub-linear (he postulated that if the returned payoff is $N, then the utility would be $log(N))
- The new expected payoff would then be:

$$P_s = \frac{1}{2}\log_{10} 2 + \frac{1}{4}\log_{10} 4 + \ldots + \left(\frac{1}{2}\right)^k \log_{10} 2^k + \ldots,$$

,which is finite (0.60206).

- To design a similar “St. Petersburg Gamble” on a real scenario, we shall revise the rules so that players will not get infinite expected return (otherwise no bookies would pay off bets).
- We’ll prove that, in our new gambles, rational players will always win (have positive gains).
- The bookmakers are then motivated to earn profit by conspiracy: they’ll covertly use a coin with heavy tail to reduce players’ chances to win.
- What if the conspiracy is exposed? If you’re an “information broker” selling this secret to some players, what is the fair price for this information?

- **A modified version of SPP:** For each round, the player still pays α dollars at first, and will receive $\sqrt{\alpha * 2^k}$ dollars if the first head occurs after k tosses of the coin.
- To simplify our analysis, assume players are risk-neutral (that is, their utility function is equal to their expected net gain), and can always decide an optimal paying strategy given provided information.
- Let β denote the probability that the coin shows its head for each toss. (In our previous slides $\beta = 0.5$)

- Given the settings above, for $\beta = 0.5$, find an optimal α^* to maximize the expected utility, and the corresponding maximal utility.
- For any value of β, decide the optimal value $\alpha^*(\beta)$ to maximize the expected utility, and the corresponding maximal utility.
- If the bookmakers conspire to use a coin with heavier tail $(0 < \beta < 0.5)$ without telling the players. For what values of β could the bookies have positive expected gain?

- Assume the playmakers are using a coin with $\beta = 0.3$ (though he claims that the coin is fair).
- If you're a "information broker" who happens to know this conspiracy: **the playmaker is using an heavy-tailed coin with $\boldsymbol{\beta = 0.3}$.**
- Now if you want to sell this information to the players, can you set a fair price for it?