BOISE STATE UNIVERSITY

CS-334
Algorithms of Machine Learning

Topic: Sequential Modeling

Arthur Putnam

Overview

- Sequential data
- What we can do with sequential modeling
- What techniques are out there
- Introduction into Markov Chains

What is Sequential Data?

What comes to mind?

What is Sequential Data?

Sequential Data is any kind of **data** where the order "matters"

What makes **Sequential Data** different from Timeseries data?

Why is it important to know if our data is sequential ?

When should we consider using a sequential approach?

When “critical” information is lost if the order is lost or not represented

Does the order of this data matter? Why or why not?

Hint: this may be trick question

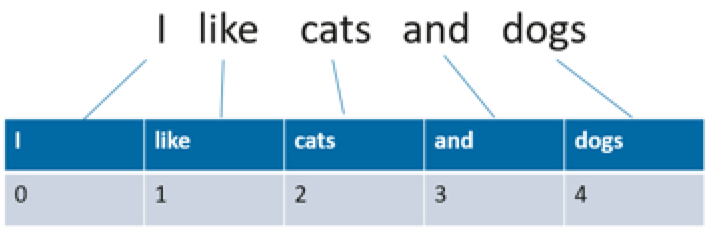
User Navigation

user	sequence
123	Homepage > Product1 > Privacy > Contact
456	Product 1 > Product 2 > Product 1 > Return Policy
789	Product 2

Events & System Logs

```
16:15:51,765 INFO   - Initializing database
16:15:55,218 INFO   - Database Service started
16:15:55,328 INFO   - Object-Relational Mapping Service
16:15:55,328 INFO   - Opening disk safes
16:16:03,468 INFO   - Disk Safe Service started
16:16:03,578 INFO   - Task Scheduler Service started
16:16:07,984 INFO   - Initializing Spring root WebApplic
16:16:13,312 INFO   - Web Server Service started
16:16:13,312 INFO   - CDP Server 3.3.0 build 8004 Starte
16:16:13,328 INFO   - Creating default agent
16:16:13,921 INFO   - Product STANDARD_SERVER_VER3_TRIAL
16:16:13,921 INFO   - License validity(true/false):  tru
16:16:13,921 INFO   - Valid until:  1/21/10 4:00 PM
16:16:13,921 INFO   - Allowed number of agents:  1
16:17:38,218 INFO   - Shutting down server
16:17:38,250 INFO   - Closing Spring root WebApplication
16:17:38,296 INFO   - Web Server Service shut down
16:17:38,296 INFO   - Task Scheduler Service shut down
16:17:38,296 INFO   - Closing disk safes
```

Words in a sentence



Key takeaway: just because the data has a sequential element doesn't mean we should model it that way

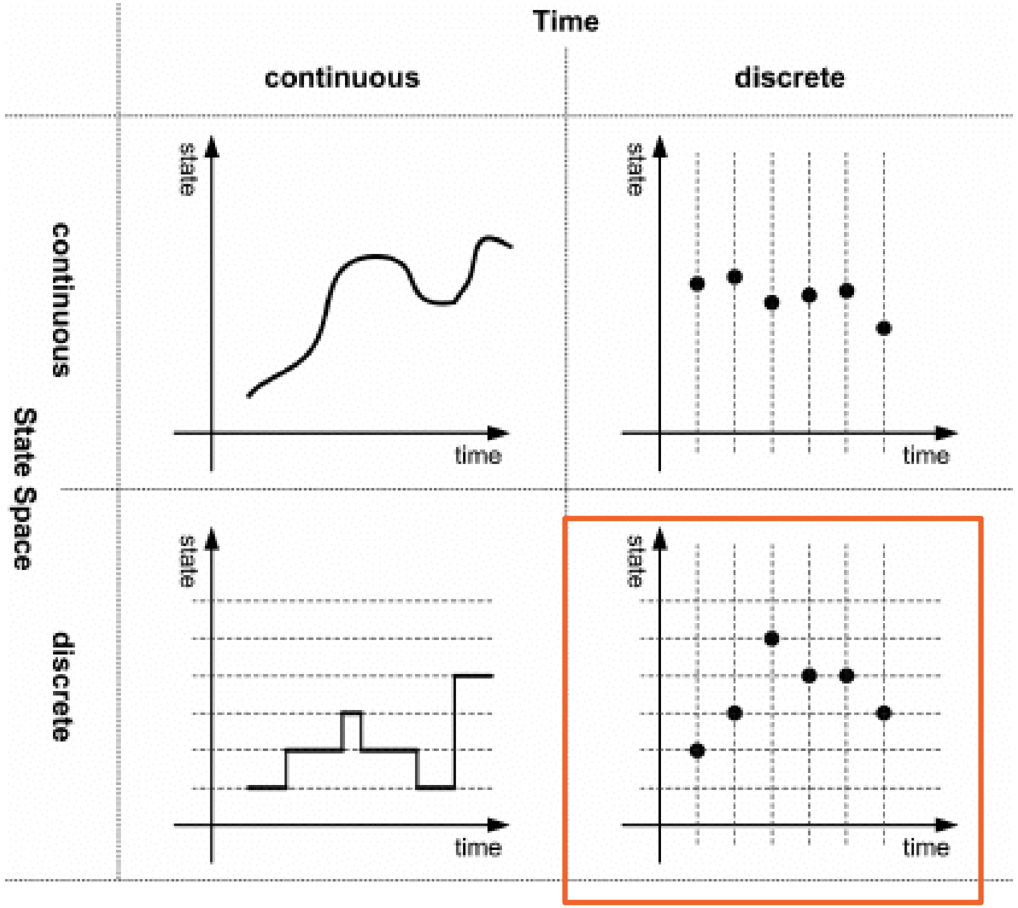
What we can do with sequential modeling?

- Classification
- Predict the next item in the sequence
- Determine if a given sequence is normal or abnormal
- Use it to create generative AI

Common Ways to Model Sequential Data

- Conditional Random Fields
- Recurrent Neural Networks (RNN)
- Markov Chains
- HMMs

Sequential State Spaces



For the rest of this lecture we are going to assume both time and states are discrete.

Sequential State Spaces

- Define a state space generically as:

$$State\ space = \mathbb{S} = \{s_0, s_1 \dots s_n\}$$

- s_i represents a discrete state

- $State\ space$ contains all possible states for seen and unseen sequences

Sequence

- Let's define a sequence as

$$X = (X_t) = (X_0, X_1, X_2, \ldots X_t)$$

$$\forall t \in \mathbb{N}, X_t \in \mathbb{S}$$

- t represents a discrete timestep
- X_t represents the discrete state at time t
- X is an ordered list where each element is in the state space

Sequence Example

- Let $\mathbb{S} = \{rock, paper, scissors\}$

Game 1

$X_0 = rock$

Game 2

$X_1 = rock$

Game 3

$X_2 = paper$

Game 4

$X_3 = scissors$

$$X = (X_0, X_1, X_2, X_3)$$
$$X = (rock, rock, paper, scissors)$$

Sequence Example 2

- Let $\mathbb{S} = \{rock, paper, scissors\}$

Game 1

$X_0 = rock$

Game 2

$X_1 = rock$

$$X = (X_0, X_1)$$
$$X = (rock, rock)$$

Let’s assume we are trying to predict the next move our opponent is going to make in a game of Rock, Paper, Scissors

Game 1

$X_0 = rock$

Game 2

$X_1 = rock$

Game 3

$X_2 = paper$

Modeling Sequential Data

- One way we might model sequential data is via a probabilistic approach where we predict future states based on the present and the past

$$P(future \mid present, past)$$

Example Assumptions

- Let’s assume we already know the probability distribution

$$P(future \mid present, past)$$

- This probability distribution is based on our opponent's previous moves

$$P(future \mid present, past)$$

Game 1

$X_0 = rock$

Game 2

$X_1 = rock$

Game 3

$X_2 = paper$

Game 4

$X_3 = scissors$

$$P(rock \mid X_2 = paper, X_1 = rock, X_0 = rock)$$
$$P(paper \mid X_2 = paper, X_1 = rock, X_0 = rock)$$
$$P(scissors \mid X_2 = paper, X_1 = rock, X_0 = rock)$$

- Let's assume

$$P(rock \mid X_2 = paper, X_1 = rock, X_0 = rock) = 0.3$$

$$P(paper \mid X_2 = paper, X_1 = rock, X_0 = rock) = 0.3$$

$$P(scissors \mid X_2 = paper, X_1 = rock, X_0 = rock) = 0.4$$

- What move should we go with?

$$P(scissors \mid X_2 = paper, X_1 = rock, X_0 = rock) = 0.4$$

Game 4

Opponent

Our Prediction

Our Move

First Approach Assessment

- What are some problems with this approach?

$$P(future \mid X_t, X_{t-1}, X_{t-2}, \ldots X_0)$$

- How would we calculate the probability distribution, if it wasn't given to us?

$$P(rock \mid X_2 = paper, X_1 = rock, X_0 = rock) = 0.3$$

$$P(paper \mid X_2 = paper, X_1 = rock, X_0 = rock) = 0.3$$

$$P(scissors \mid X_2 = paper, X_1 = rock, X_0 = rock) = 0.4$$

Markov Property

- **Markov Property** states that the conditional probability distribution of future states of the process **depends only** on the present state, not on the sequence of events that preceded it.

- **Markov assumption** is used to describe a model where the Markov property is assumed to hold

Using the Markov Property

$$P(future \mid present, past)$$

$$P(future \mid X_t, X_{t-1}, X_{t-2}, \dots X_0)$$

$$P(future \mid present)$$

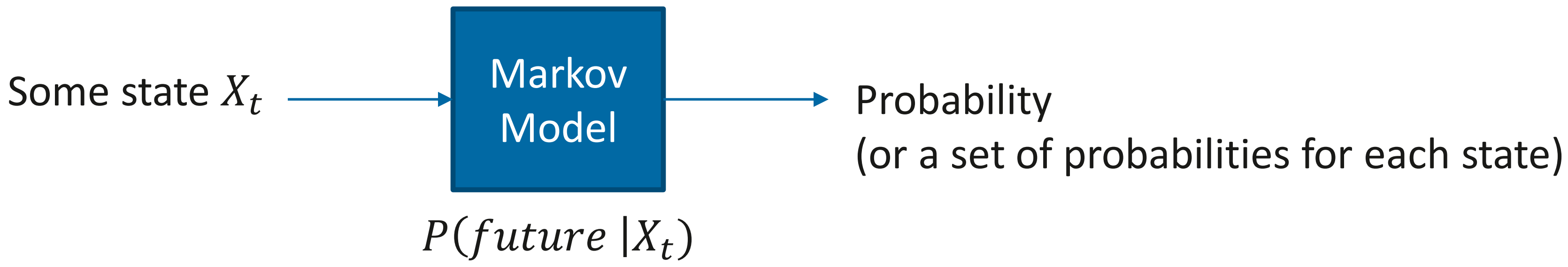
$$P(future \mid X_t)$$

This simplifies the probability function and is more robust at handling differently ordered sequences

Markov Chain

- A **Markov chain** is a stochastic model describing a sequence of possible events in which the probability of each event depends only on the state attained in the previous event.

***Conceptual Model**



*We will define this model more formally later

Visualization

https://setosa.io/ev/markov-chains/

Markov Chain

- Often considered to be “memory-less” thanks to the Markov Property
- Markov Model can take a sequence as input and produce
  - the probability of that sequence
  - or the probability of the next states in the sequence
- Is trained from empirical data
  - (multiset of sequences)
- Pros: easy to train, simple to understand

Let’s Build a Markov Chain from Scratch

- We need to:
  - Define the state space
  - Find the initial probabilities
  - Find the transition probabilities
- Data Set
  - Sequences:
    - RRPSSRPSRP
    - PPPSPSPSRR
    - RPSSPSRPSP
  - R = Rock, P = Paper, S = Scissors

Define the state space

Empirical data:

- RRPSSRPSRP
- PPPSPSPSRR
- RPSSPSRPSP

$$\mathbb{S} = \{R, P, S\}$$

Find the Initial Probabilities

Empirical data:

- RRPSSRPSRP
- PPPSPSPSRR
- RPSSPSRPSP

R starting Probability	P starting Probability	S starting Probability
2/3 = 0.666	1/3 = 0.333	0/3 = 0

Find the Transition Probabilities

Empirical data:

- RRPSSRPSRP
- PPPSPSPSRR
- RPSSPSRPSP

$$P(X_{t+1} = R \,|X_t = R)$$

$$\frac{Number\ of\ pairs\ where\ X_t, X_{t+1}\ is\ seen}{Number\ of\ X_t}$$

State Transitions	$P(X_{t+1}\|X_t)$
RR	2/7
RP	
RS	
PR	
PP	
PS	
SR	
SP	
SS	

Find the Transition Probabilities

Empirical data:

- RRPSSRPSRP
- PPPSPSPSRR
- RPSSPSRPSP

State Transitions	$P(X_{t+1} \| X_t)$
RR	2/7
RP	5/7
RS	
PR	
PP	
PS	
SR	
SP	
SS	

$$P(X_{t+1} = R \mid X_t = P)$$

$$\frac{Number\ of\ pairs\ where\ X_t, X_{t+1}\ is\ seen}{Number\ of\ X_t}$$

Find the Transition Probabilities

Empirical data:

- RRPSSRPSRP
- PPPSPSPSRR
- RPSSPSRPSP

$$P(X_{t+1} = R \mid X_t = S)$$

$$\frac{Number\ of\ pairs\ where\ X_t, X_{t+1}\ is\ seen}{Number\ of\ X_t}$$

State Transitions	$P(X_{t+1} \mid X_t)$
RR	2/7
RP	5/7
RS	0/7
PR	
PP	
PS	
SR	
SP	
SS	

Find the Transition Probabilities

Empirical data:

- RRPSSRPSRP
- PPPSPSPSRR
- RPSSPSRPSP

$$P(X_{t+1} \,|X_t)$$

$$\frac{Number\ of\ pairs\ where\ X_t, X_{t+1}\ is\ seen}{Number\ of\ X_t}$$

State Transitions	$P(X_{t+1}\|X_t)$
RR	2/7 = 0.285
RP	5/7 = 0.714
RS	0/7 = 0
PR	0/10 = 0
PP	2/10 = 0.2
PS	8/10 = 0.8
SR	4/10 = 0.4
SP	4/10 = 0.4
SS	2/10 = 0.2

So now we have…

Initial Probabilities

R starting Probability	P starting Probability	S starting Probability
2/3 = 0.666	1/3 = 0.333	0/3 = 0

Transition Probabilities

State Transitions	$P(X_{t+1} \mid X_t)$
RR	2/7 = 0.285
RP	5/7 = 0.714
RS	0/7 = 0
PR	0/10 = 0
PP	2/10 = 0.2
PS	8/10 = 0.8
SR	4/10 = 0.4
SP	4/10 = 0.4
SS	2/10 = 0.2

Now we should be able to calculate the probability of the sequence of: RPS

So now we have…

Initial Probabilities

R starting Probability	P starting Probability	S starting Probability
2/3 = 0.666	1/3 = 0.333	0/3 = 0

Transition Probabilities

State Transitions	$P(X_{t+1}\|X_t)$
RR	2/7 = 0.285
RP	5/7 = 0.714
RS	0/7 = 0
PR	0/10 = 0
PP	2/10 = 0.2
PS	8/10 = 0.8
SR	4/10 = 0.4
SP	4/10 = 0.4
SS	2/10 = 0.2

calculate the probability of the sequence of: RPS

$$P(X_0 = R) * P(X_1 = P|X_0 = R) * P(X_2 = S \mid X_1 = P)$$

$$\frac{2}{3} * \frac{5}{7} * \frac{8}{10} = \frac{80}{210} = \frac{8}{21} \approx 0.38$$

Looking up each probability is still tedious

- Lucky we can redefine this process as a series of matrix multiplications
- We can also define 1xN (denoted by π) vector to represent our initial state probabilities
- We can define a NxN matrix (denoted by P) which represents the transition probabilities
- From there we can use

$$P(sequence) = \pi_{starting\_state} * P_{transition_1} * P_{transition_2} \dots * P_{transition\_n}$$

Initial State Probabilities

- We can also define 1xN (denoted by π) vector to represent our initial state probabilities

$$\pi = \left[p_{\mathbb{S}_{inital_1}}, p_{\mathbb{S}_{inital_2}}, \dots p_{\mathbb{S}_{inital_n}}\right]$$

Initial Probabilities

R starting Probability	P starting Probability	S starting Probability
2/3 = 0.666	1/3 = 0.333	0/3 = 0

$$\pi = [0.666, 0.333, 0]$$

$$\mathbb{S} = \{R, P, S\}$$

Transition Matrix

- Let p be an NxN matrix where N is the number of discrete states

$$P = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22} & \cdots & p_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ p_{n1} & p_{n2} & \cdots & p_{nn} \end{bmatrix}$$

$$P_{ij} = P(\mathbb{S}_j | \mathbb{S}_i)$$

*Assume $\mathbb{S}$ is ordered

$$P = \begin{array}{c} \\ R \\ P \\ S \end{array} \begin{array}{c} \begin{array}{ccc} R & P & S \end{array} \\ \begin{bmatrix} 0.285 & 0.714 & 0 \\ 0 & 0.2 & 0.8 \\ 0.4 & 0.4 & 0.2 \end{bmatrix} \end{array}$$

$$P_{ij} = P(\mathbb{S}_j | \mathbb{S}_i)$$

$$\mathbb{S} = \{R, P, S\}$$

Transition Probabilities

State Transitions	$P(X_{t+1} \mid X_t)$
RR	2/7 = 0.285
RP	5/7 = 0.714
RS	0/7 = 0
PR	0/10 = 0
PP	2/10 = 0.2
PS	8/10 = 0.8
SR	4/10 = 0.4
SP	4/10 = 0.4
SS	2/10 = 0.2

Markov Chain

- We can now calculate the probability of a sequence by "chaining" together the elements of the matrix

$$P(sequence) = \pi_{starting\_state} * P_{transition_1} * P_{transition_2} \dots * P_{transition\_n}$$

Let’s check

Previously we calculated the probability of the sequence of: RPS using the “old” method. Let’s try the matrix method.

$$P(X_0 = R) * P(X_0 = P | X_0 = R) * P(X_1 = S \mid X_1 = P)$$

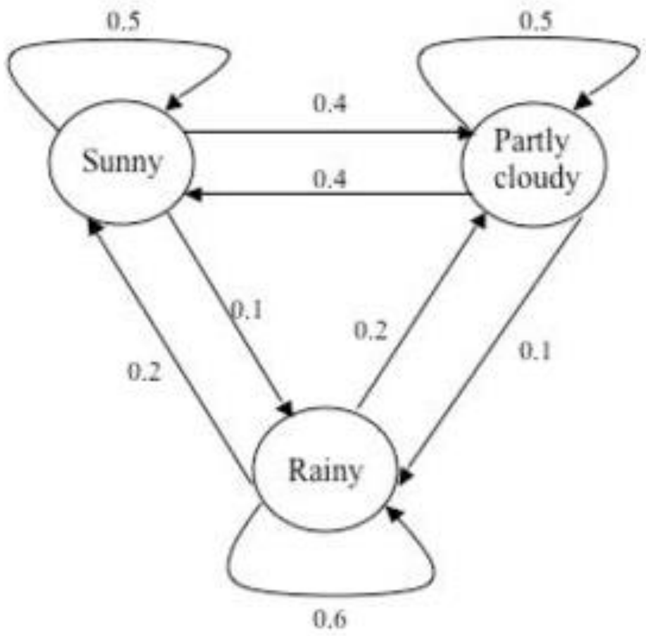
$$\frac{2}{3} * \frac{5}{7} * \frac{8}{10} = \frac{8}{21} \approx 0.38$$

$$\pi = [0.666, 0.333, 0]$$

$$P = \begin{array}{c} \\ R \\ P \\ S \end{array} \begin{array}{c} \begin{array}{ccc} R & P & S \end{array} \\ \begin{bmatrix} 0.285 & 0.714 & 0 \\ 0 & 0.2 & 0.8 \\ 0.4 & 0.4 & 0.2 \end{bmatrix} \end{array}$$

$$\pi_1 * P_{12} * P_{23}$$
$$0.666 * 0.714 * 0.8 = 0.38$$

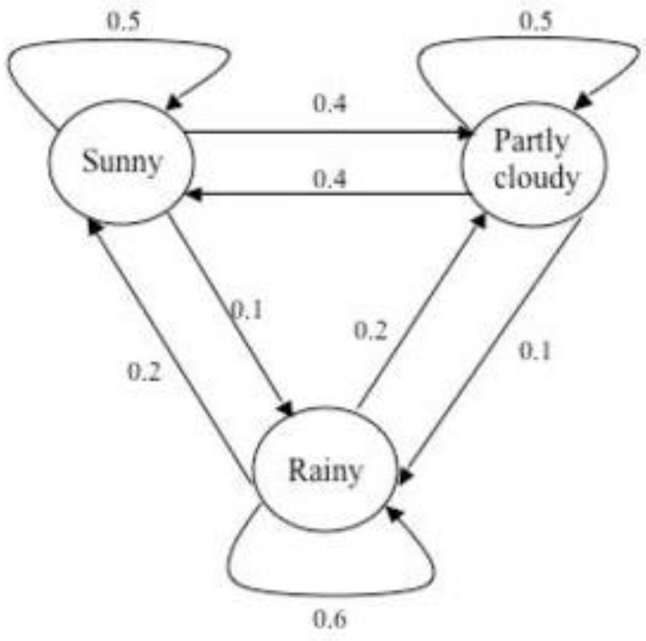
Your Turn! Class Challenge



- Assume $\mathbb{S} = \{S, P, R\}$
- Assume $\pi = [0.8, 0.1, 0.1]$
- Create the Transition matrix P

- Try to find the probability for:
  - SPR

Your Turn! Class Challenge



https://setosa.io/ev/markov-chains/

- Assume $\mathbb{S} = \{S, P, R\}$
- Assume $\pi = [0.8, 0.1, 0.1]$
- Try to find the probability for:
  - SPR

$$P = \begin{bmatrix} 0.5 & 0.4 & 0.1 \\ 0.4 & 0.5 & 0.1 \\ 0.2 & 0.2 & 0.6 \end{bmatrix}$$

$$\pi_1 * p_{12} * p_{23} = 0.8 * 0.4 * 0.1 = 0.032$$

So what can we do with our Markov chain

- Calculate the probability of a sequence
  - This is useful for comparing on sequence to another
  - We could also use this to classify by pick a probability cutoff point

- We can calculate the probability of ending in state S after some number of transitions T

Recommended Next Steps

- Hidden Markov Models (HMM)
  - We are making the assumptions that are likely untrue, HMMs help address or model hidden states

- Smoothing & Normalization
  - Some transitions might never happen in our data set, thus the probability will be zero which is probably not what we want

Arthur Putnam

Email: arthurputnam@boisestate.edu

THANK YOU