

What comes next?

Description

It's a banal truism: events follow one another in sequence, inexorably. You brush teeth, you wash face, you pour cereal, you eat cereal, you rinse bowl, you attend online meeting, you get dressed &!

Such sequences often follow patterns. In some cases, a researcher might want to detect those patterns: to predict what comes next, to show how to reinforce or break out of a pattern, or to detect the variables that influence those event sequences.

I've seen at least one article that explains the usefulness of detecting patterns as people navigate from one web page to another on line ([link](#)). Each web page is an event.

A chain of events

I don't have access to that kind of data. I can however inspect the last 100 blog posts I've published. Blog posts appear in a time sequence. I'll pretend here that someone is interested to try and predict the topic of the next post.

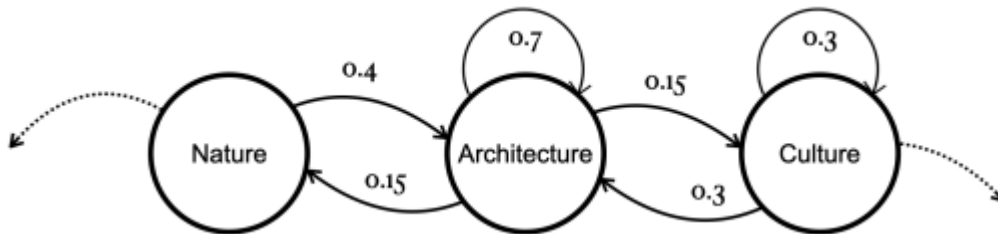
Blog platforms such as WordPress encourage the author to set up a series of categories. Here are 14 of my most common categories: Architecture, Artificial Intelligence (AI), Body, Culture, Economics, Ethics, Film, Media, Metaphor, Nature, Play, Research, Society, Voice.

An interested data analyst could say, "The blogger has just published a post about Architecture, what are the odds that the next post in the author's time sequence will also be about Architecture?" Let's assume the analyst has no access to previous posts, and that those buttons at the foot of a post that say "previous post" and "next post" are invisible.

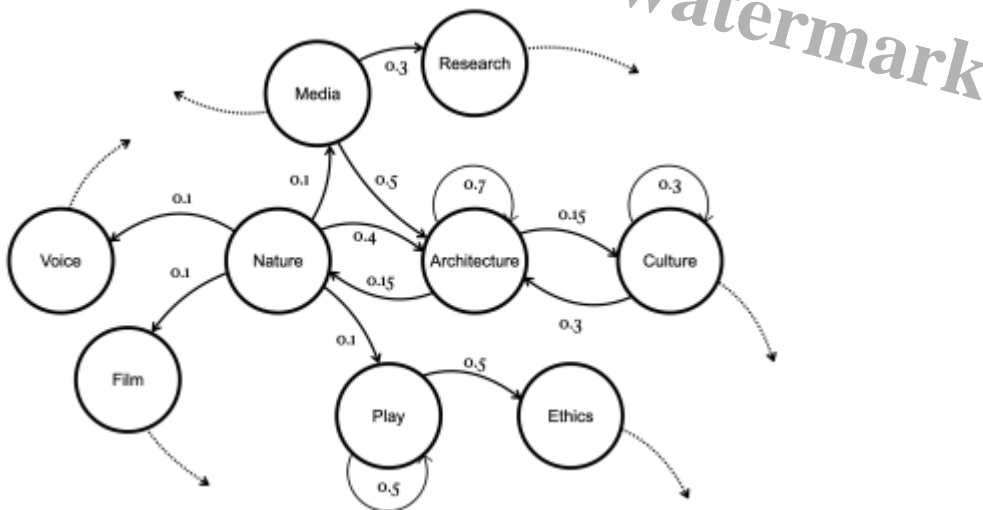
I've actually calculated those odds from the sequence of my past 100 posts. Of the 100 posts, 45 were in the Architecture category, and 32 of those Architecture posts were followed by another Architecture post. Judging by that 100 post history, the odds are therefore about 7:3, i.e. a 0.7 probability that if I'm looking at an Architecture post the next post will also be about Architecture. There's also a 0.15 probability it will be about Nature, and a 0.15 probability it will be about Culture.

The probabilities of what comes next should add up to 1.0.

If the next published post happens to be about Culture, the probability that the one after that in the author's time sequence will also be about Culture is 0.3 or it could be about Architecture, with a 0.3 probability. There's a 0.4 probability that a Culture post will be followed by something else, shown here with a dotted line. Here's a diagram showing those possibilities and their probabilities.



Think of a random walk through this network, where an agent makes a decision at each node (the circles) about which arrow to follow next. The decision is weighted according to the probabilities of the arrows exiting the node. I've added some more nodes and probabilities from my analysis of my last 100 posts here.



The network diagram soon gets out of hand. It's simpler as a matrix. I've also limited the values to one decimal place for simplicity. The grey cells indicate the probability that a post category in the sequence repeats, i.e. they are the loops in the network diagram above.

	Architecture	Artificial Intelligence	Body	Culture	Economics	Ethics	Film	Media	Metaphor	Nature	Play	Research	Society	Voice
Architecture	0.7			0.1						0.1				
Artificial Intelligence		0.3		0.3	0.3									
Body	1.0													
Culture	0.3	0.1		0.1	0.3	0.1			0.1					
Economics	0.3	0.2		0.2	0.3									
Ethics	0.7					0.2	0.2							
Film			0.3	0.3	0.3					0.3				
Media	0.5											0.3	0.3	
Metaphor														1.0
Nature	0.4						0.1	0.1			0.1			0.1
Play						0.5					0.5			
Research related	1.0													
Society				0.3			0.3	0.3						
Voice and text				0.1			0.1	0.1					0.1	0.4

Even though I derived these probabilities from my record of the post sequence, someone inspecting the network (or the matrix) doesn't know that history. A person (agent or algorithm) navigating through network doesn't need to consider the path just taken to work out what happens next.

Markov chains

Generalising this method, the nodes in the diagram could be any decision point, e.g. road junctions encountered while navigating on a bike through the city. The values on the arrows (arcs) could be the probability that a cyclist would take that route given statistical data about congestion or gradient. There's no planned destination in this scenario, and there's no account of where the cyclist has just come from. It's a predictive model that assumes no origin or aim.

What I have been describing is an event sequence as a *Markov process*. The OED describes that as any stochastic process for which the probabilities, at any one time, of the different future states depend only on the existing state and not on how that state was arrived at. The nodes in the Markov chain network above are known as states.

I'm interested in this kind of modelling and its application to the urban condition. That's as a precursor to examining the concept of Hidden Markov Models (HMM).

Also see posts tagged [maze](#).



Bibliography

- Singer, Philipp, Denis Helic, Behnam Taraghi, and Markus Strohmaier. 2014. Detecting Memory and Structure in Human Navigation Patterns Using Markov Chain Models of Varying Order. *Plos One*, (9) 12. [Link](#)

Note

- The image above shows a cyclist and a vehicle with Google Street View camera in Edinburgh 29 May 2015, 10:59 am.



Category

1. Architecture

Tags

1. Markov process
2. navigation
3. smart city

Date Created

September 26, 2020

Author

rcoyne99

default watermark