Natural Language Processing for Knowledge Representation and Reasoning

Michaël Thomazo

April 14th, 2014
Dresden

Markov processes

Markov processes are widely used in NLP. We start today with the simplest one, Markov chains.

- presentation of the POST problem
- definition of a Markov chain
- examples of modelling
- intuitive presentation of HMM
- intuitive use of HMM for POST
- formalization comes at the next course

Part of speech tagging

Example taken from http://www.cs.umd.edu/ nau/cmsc421/part-of-speech-tagging.pdf

Part of speech: linguistic category of words Part of speech tagging (POST): associate a POS with each word in a sentence.

Part of speech tagging

Example taken from http://www.cs.umd.edu/ nau/cmsc421/part-of-speech-tagging.pdf

Part of speech: linguistic category of words Part of speech tagging (POST): associate a POS with each word in a sentence.

Example: Heat (V) water (N) in (P) a (DET) large (ADJ) vessel (N).

Why is it hard?

More explanation on Wikipedia about this sentence

Buffalo buffalo Buffalo buffalo buffalo buffalo Buffalo buffalo.

An explanation

Text excerpted from the Wikipedia articles *Buffalo buffalo Buffalo buffalo buffalo buffalo Buffalo buffalo.*, *Homonym* and *Homophone*. 26 March 2007

Simple approach work already quite good

- Assigning to each word its most common POS results in already 91% of success for English
- this is still troublesome: POST is often a pre-processing step; any error there has a big influence on the following

Several approaches for POS-tagging

- rule-based
- transformation-based
- stochastic

More importantly, the degree of “supervision” (labeled corpus available?)

Markov chains through examples

The Land of Oz

Weather in the Land of Oz: they never have two nice days in a row.

- if they have a nice day, nice and snow are equally likely the next day
- snow or rain, even chance of having the same next day
- if it still changes, half a chance it becomes nice weather

Transition Matrix

The transition matrix summarizes the probabilities to go from one state to another.

Transition Matrix

The transition matrix summarizes the probabilities to go from one state to another.

$$M = \begin{array}{c} \\ R \\ N \\ S \end{array} \begin{array}{c} \begin{array}{ccc} R & N & S \end{array} \\ \begin{pmatrix} 1/2 & 1/4 & 1/4 \\ 1/2 & 0 & 1/2 \\ 1/4 & 1/4 & 1/2 \end{pmatrix} \end{array}$$

(Homogeneous) Markov chains in a nutshell

- there is a set of states $S = \{s_1, s_2, \ldots, s_n\}$
- the process starts from one of these steps and moves form one state to another
- if the chain is in state s_i ($X_k = s_i$), there is a probability $p_{i,j}$ that it moves to s_j ($X_{k+1} = s_j$)

Very important: the *transition probability* depends *only* on the last state.

Campaign

The President tells A whether he will run for his succession. A tells to B, B to C, and so on, alway to somebody new. There is a probability a that a person change the answer from yes to nom and a probability b from no to yes.

Ehrenfest Model

Two urns contains d balls. At teach step, one of the d balls is chosen at random and moved from the urn that it is in into the other urn.

Ehrenfest Model

Two urns contains d balls. At teach step, one of the d balls is chosen at random and moved from the urn that it is in into the other urn.

$$M = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{array}{c} \begin{array}{ccccc} 0 & 1 & 2 & 3 & 4 \end{array} \\ \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 1/4 & 0 & 3/4 & 0 & 0 \\ 0 & 1/2 & 0 & 1/2 & 0 \\ 0 & 0 & 3/4 & 0 & 1/4 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix} \end{array}$$

Gene Model

We consider a population of animals, and a pair of genes, each of which may be of two types, G or g.

- GG and Gg are often indistinguishable
- dominant (GG), recessive (gg), hybrid (Gg)

Inheritance of one of the two genes by the offspring. It is assumed to be arbitrary.

- represent the result of mating an animal with an hybrid
- with a recessive
- with a dominant

Many other modeling targets

- biology (interacting molecules, predator-prey models)
- medicine (epidemiology, gene transmission, population dynamocs)
- computer science (internet traffic...)
- and more!

The Markov Property

The Markov Property

A process $X_0, X_1, \ldots$ satisfies the Markov property if:

$$\mathbb{P}\{X_{n+1} = i_{n+1} \mid X_n = i_n, \ldots, X_0 = i_0\} = \mathbb{P}\{X_{n+1} = i_{n+1} \mid X_n = i_n\}$$

for all n and all $i_j \in S$.

A first exercise

Taken from http://www.isn.ucsd.edu/classes/beng260/complab/week2/Chang_1999.pdf

Let $X_0, X_1, \ldots$ be a Markov chain. Let A and B be subsets of the state space.

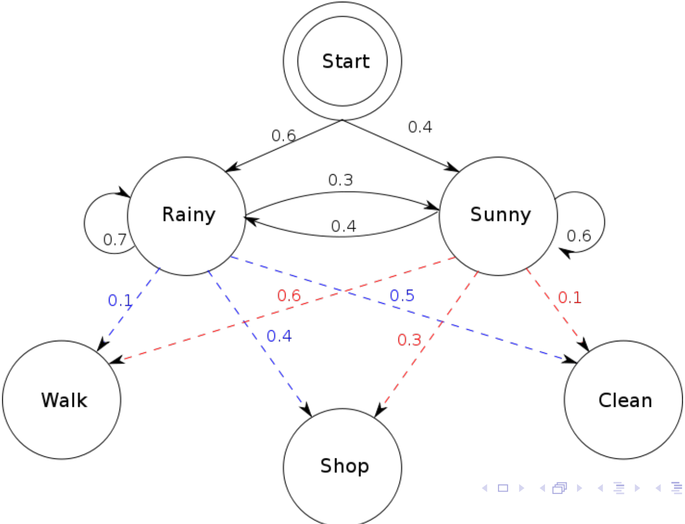
1. Is it true that:
   $\mathbb{P}\{X_2 \in B \mid X_1 = x_1, X_0 \in A\} = \mathbb{P}\{X_2 \in B \mid X_1 = x_i\}$?
2. Is it true that:
   $\mathbb{P}\{X_2 \in B \mid X_1 \in A, X_0 = x_0\} = \mathbb{P}\{X_2 \in B \mid X_1 \in A\}$?

Hidden Markov Models

Markov property still holds... but one does not have direct access to the state the process is in.
One only have access to an output, which depends on the state.

An Example

Source: Wikipedia



Two important problems

- given a model, can we find the parameters of the model that best explain a set of observations? (training part)
- the parameters of the model being fixed and given an observation, can we find the sequence that best explain this observation?