SELF-INDEPENDENT THEORITICAL STUDY ON MARKOV MODELS FOR LANGUAGE UNDERSTANDING AT DISCRETE TIME AND SPACE.

Vivek Suresh Raj
vivek.sureshraj {@ucalgary.ca}

Statement:

The Markov models and Markov chains has created a way to describe how random variables evolve at each time-step in the chain during the transition of states. Such evolution of random variables in the process contribute to the stochastic processes or methods. The on-going research is on deciding the next possible state for a long-term goal in a process under dynamic environment. This independent study is a small contribution in understanding the language with Markov model.

Related-work:

Humans, learn and evolve under trail-and-error learning algorithm. There are numerous methods on applying reinforcement method of learning in robots. Markov chain studies the best next possible state among all possibilities for an ‘uncertain’ and ‘dynamic’ environment.

Markov method of prediction is based on ‘conditional probability’ theorem. For instance, under conditional probability theorem, let us consider a sequence of random variable $X_1 \ldots . X_t$, on assuming that it is on a discrete time and finite space. Now, the future state, s at time (t+1) is “not only” dependent of its present states but also, its history, that is, X_0, X_1, X_2, $X_3 \ldots X_{t-1}$, X_t. Now, the burden of carrying the history is reduced in Markov chain, thus making it less computational. However, Markov’s chain or Markov model doesn’t mean that the history is independent of the future but it is conditionally independent, given the present state.

Experiment on Language understanding tasks:

Language is considered as a sequence of tokens in which the next-state token is considered to be dependent on the token contained in present state. Let us consider a simple example of three sentences in a common English language as,

1. He likes apple.
2. He likes chicken.
3. John was called for a cup of coffee.

The lemmatized word operation upon the above sentences is performed for further transition probability matrix formation.

Verification of theorem on language understanding:

The sequence of words in the sentences at initial state is considered for probability state-transition matrix calculations.

Probability state-transition, $P_{ss} = \begin{bmatrix} 0 & 0.7 & 0.1 & 0.1 & 0.1 & 0 \\ 0 & 0 & 0.25 & 0.25 & 0.25 & 0.25 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.4 & 0 & 0 & 0 & 0.6 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.5 & 0 & 0 & 0 & 0.5 \end{bmatrix}$

The corresponding current state and next state probability transitions are dependent on the current state word vectors. For example, the likelihood of having a word, w(apple) after word vector of, w(likes) is 10% in a sentence. Let us assume that the initial state distributions within the set of word vectors starts with the token, 'He' in the distribution.

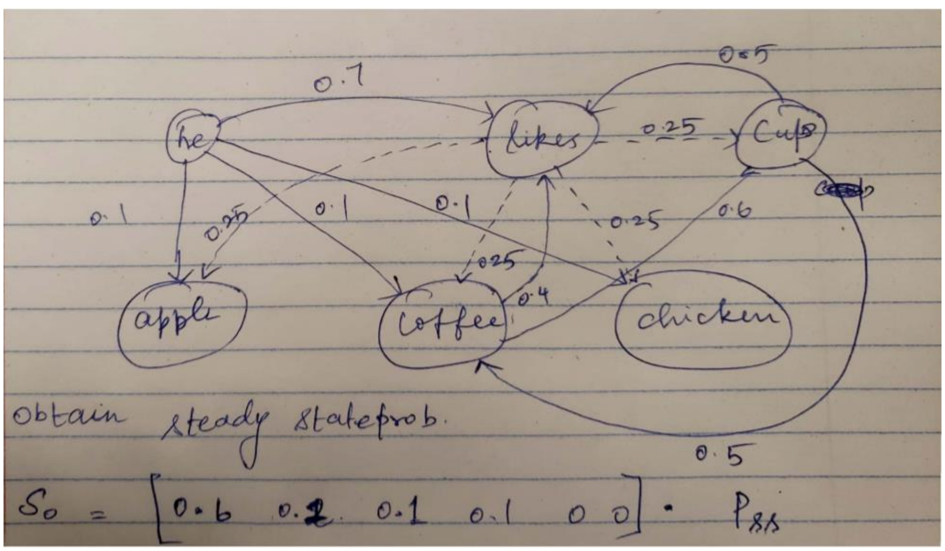
The initial state distribution at time, t_0 is given by, $S_0 = [0.6 \quad 0.2 \quad 0.1 \quad 0.1 \quad 0 \quad 0]$

The probability mass among the state distribution of the corresponding words is seen to be maximum for the token, w(He). Now, lets compute the next state possibility as,

Next state possibility, $S_t = \sum_{i=0}^{n} S_i \cdot P_{ss}$ where, i = {1…n}, until a stationary matrix is reached.

The stationary state matrices after a finite number of steps represents a steady-state system in Markov's chain.

Tree-Diagram:



APPENDIX-1:

Comparison study of the problem between Conditional Probability (CP) based maximum likelihood estimation and Markov chain process evaluation.

CP estimation of next possible outcome,

$P\ (X_{(t+1)} = i_{(t+1)}\ |\ X_t = i_t, X_{t-1} = i_{t-1}, \ldots, X_1 = i_1, X_0 = i_0\ )$

Whereas, by Markov Model it is possible to predict the best next possible state in a sequential decision-making problem on a long-run. The above statement holds only if it satisfies the following equation.

$P\ (X_{(t+1)} = i_{(t+1)}\ |\ X_t = i_t)$

APPENDIX – II:

Algorithm to Markov chain in the language understanding problem.

Step-1: Start
Step-2: Formulation of transition probability matrix,
$\sum_{i=1}^{t} X_i = x_t \epsilon X$, at time-step, t; $\sum_{i,j\,=\,0}^{t} \text{count}(x_{ij})\ /\ \text{count}\ (x_i)$
Step-3: Initialise the state-probability at time-step, t = 0.
Step-4: Markov chain process until the steady-state system is reached.
$S_{(t+1)} = \sum_{i=0}^{n} S_i\ .\ P_{ss}$ where, i = {1…n}
If $\sum_{i=0}^{n} S_{i+1} == S_i\ ,\ \forall\ S_i \in$ {Set of all possible states}
Step-5: Next possible state from Markov chain is given by, *argmax*($\sum_{i=0}^{n} S_{i+1}$)