BIOLOGICAL GENE SEQUENCE STUCTURE ANALYSIS USING HIDDEN MARKOV MODEL

Karuppusamy T[2], Dr.M.Sivasubramanian[2]

[1]Research Scholar, Department of Mathematics, Government Arts Udumalpet 642 126, Bharathiar University
karuppusamy26@gmail.com

[2]Assistant Professor, Department of Mathematics, Government Arts College, Udumalpet 642 126,Bharathiar University
msivasubramanian0@gmail.com

_____

Abstract

　　　　Identification or prediction of coding sequences from within genomic DNA has been a major part of the search of the gene. In this work real hidden Markov models (HMMs) to denote the consensus and deliver a beneficial tool in determining the splicing junction sites Markov models which has a recurring nature in computational biology leads to statistical models, in every sequential analysis it plays a role of putting up a right label on each residue. In sequential alignment and as well as in gene identification namely exons, introns or intergenic sequences which make in a sequence with homologous residue with the target database. Under the gene identification methodology Condon bias, exons, introns have length preference which leads to a combination of splice site consensus. Parameters are fixed on the onset while weight of the different information are polled together leading to the interception of result probability, which could lead to identifying the best score based on score mean and how confident are the best scoring answers are perfect. This leads to the concept of extendibility, to perfect and ad hoc gene finder, which is a modeled transitional methodology leading to the consensus, alternate splicing and offers polyadenylation signal. This leads to piling of authenticity against a delicate ad hoc program which could make to breakdown under its individual weightiness.

_____

1. Introduction

　　　　A proper substance for creation of probabilistic model towards a direct sequence and classification difficulties can provide the Hidden Markov Models (HMMs).To site a constructive interpretation a toy (HMM'S) and 5' splice site acknowledgement. Distortion of a 5's splice site recognition problem where a given DNA sequence which begins with an exons which contains one 5' splice site and end with introns. It is a issue where we have to classify, everywhere the switched since exons to introns happened and where the 5's splice site (5'ss) Sequence of exons and introns have a dissimilar statistical property in order for an intelligent activity just thing for some simple difference where exons have a even base arrangement on Middling, assume 25% each base, which introns on A/T rich (say has 40% A/T and 10% C/G) and the 5'ss consensus nucleotide is practically permanently a G (95% G and 5% A)

　　　　HMM bring in three state's which means, might assigned to a nucleotide E (exons) 5 (5'SS) and I (introns) each of it has emission probabilities. This could perfect up the base composition of exons, introns and the consensus G at 5 SS. This attempts to a transition probability which could make the probabilities to shift from existing state to a new state, under transition probability the linear order under which we expect the state's to occur may be one or more E's and one 5, one or more to interpret that HMM is generating a sequence under which the researcher visit a state, which emits a residue from the emission probability distribution under such a situation we have to clearly predict the transition probability distribution

　　　　In such a situation a model generates to setup the information where in one set under stressing the state path, which means labels as the transition from one point to another point. In the next category the observed sequence the DNA of each residue being emitted from one point to the next path point in markov chain methodology the point here means that it is based on the point which we put up. In this two observed sequence DNA, where each residue which is emitted from one point to the point of movement. The Markov Chain which means what we state , will go on and depend on point which comes in because only in the observed sequence the point path is hidden which are the residue labels that infers the point path in markov chain which is normally hidden

The overall paper has been describes as follows section 1 introduction about the HMM , section 2 describes the finding the best state path using HMM, section 3 represents the probability analysis of paths , section 4 deals with the main sequence of problem and result with hidden state Finally paper concluded on the section 5.

2. Finding the best state path

In an examination issue, we are given a grouping and we need to construe the secret state way. There are possibly many state ways that could produce the succession. We need to locate the one with the most elevated likelihood.For example, if we were given the HMM and the 26-nucleotide sequence as

C T T C A T G T G A A A G C A G A C G T A A G T C A

And 14 possible paths have non-zero probability since the 5'SS must fall on one of 14 internal A'S or G's. Fig 1 enumerates the representation of a HMM for the given sequence.
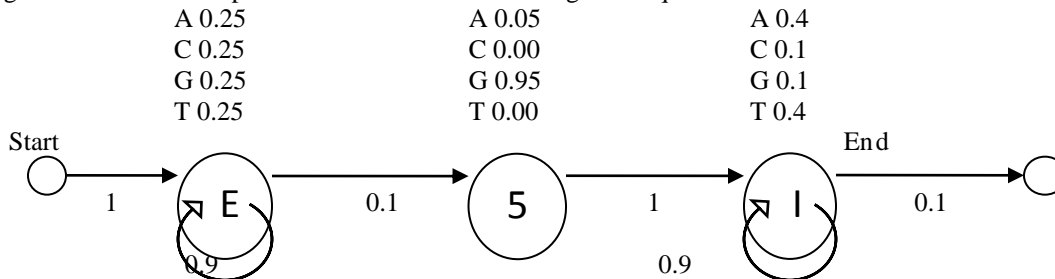
Six most noteworthy scoring ways (those with G at the 5'SS). The best one has a log likelihood of - 41.22 which induces that the most probable 5'SS position is at the fifth G.

3. Proposed Work

3.1 Probability Analysis

The probabilities related with different state changes are called progress probabilities (TP). The interaction is portrayed by a state space, a change network depicting the probabilities of specific advances, and an underlying state (or introductory appropriation) across the state space. Gee is a Markov interaction that at each time step creates an image from some letters in order, Σ, as indicated by outflow likelihood (EP) that relies upon the state. is the arrangement of states visited? of item: the likelihood of the way in the model occasions the likelihood of creating a given succession expecting a given way in the model.) The likelihood of way (PoP) is the result of progress likelihood and emanation likelihood and the log likelihood of way (log-PoP) is the amount of log esteems.

Path 1

C T T C A T G T G A A A G C A G A C G T A A G T C A

E E E E E E 5 I I I I I I I I I I I I I I I I I I I I

TP	$(1.0)(0.9)^5(0.1)(1.0)(0.9)^{18}(0.1)$
EP	$(0.25)^6(0.95)(0.4)^{11}(0.1)^8$
PoP	$(0.9)^{23}(0.1)^{10}(0.25)^6(0.95)(0.4)^{11}$
log -PoP	$\log\Pr(path1) = -2.42329 - 23.025 - 8.31777 - 0.05129 - 10.07920 = -43.89655 \approx -43.90$

Path 2

C	T	T	C	A	T	G	T	G	A	A	A	G	C	A	G	A	C	G	T	A	A	G	T	C	A
E	E	E	E	E	E	E	5	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I

TP	$(1.0)(0.9)^{7}(0.1)(1.0)(0.9)^{16}(0.1)$
EP	$(0.25)^{8}(0.95)(0.4)^{10}(0.1)^{7}$
PoP	$(0.9)^{23}(0.1)^{9}(0.25)^{8}(0.95)(0.4)^{10}$
Log-PoP	$\log\Pr(path2) = -2.42329 - 20.72327 - 11.09035 - 0.05129 - 9.16291 = -43.45111 \approx -43.45$

Path 3

C	T	T	C	A	T	G	T	G	A	A	A	G	C	A	G	A	C	G	T	A	A	G	T	C	A
E	E	E	E	E	E	E	E	E	E	5	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I

TP	$(1.0)(0.9)^{11}(0.1)(1.0)(0.9)^{12}(0.1)$
EP	$(0.25)^{12}(0.95)(0.4)^{7}(0.1)^{6}$
PoP	$(0.9)^{23}(0.1)^{8}(0.25)^{12}(0.95)(0.4)^{7}$
Log-PoP	$\log\Pr(path3) = -2.42329 - 18.42068 - 16.63553 - 0.05129 - 6.41404 \approx -43.94$

Path 4

C	T	T	C	A	T	G	T	G	A	A	A	G	C	A	G	A	C	G	T	A	A	G	T	C	A
E	E	E	E	E	E	E	E	E	E	E	E	E	5	I	I	I	I	I	I	I	I	I	I	I

TP	$(1.0)(0.9)^{14}(0.1)(1.0)(0.9)^{9}(0.1)$
EP	$(0.25)^{15}(0.95)(0.4)^{6}(0.1)^{4}$
PoP	$(0.9)^{23}(0.1)^{6}(0.25)^{15}(0.95)(0.4)^{6}$
Log-PoP	$\log\Pr(path4) = -2.42329 - 13.81551 - 20.79442 - 0.05129 - 5.49774 \approx -42.58$

Path 5

C	T	T	C	A	T	G	T	G	A	A	A	G	C	A	G	A	C	G	T	A	A	G	T	C	A
E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	5	I	I	I	I	I	I	I	I

TP	$(1.0)(0.9)^{17}(0.1)(1.0)(0.9)^{6}(0.1)$
EP	$(0.25)^{18}(0.95)(0.4)^{5}(0.1)^{2}$
PoP	$(0.9)^{23}(0.1)^{4}(0.25)^{18}(0.95)(0.4)^{5}$
Log-PoP	$\log\Pr(path5) = -2.42329 - 9.21034 - 24.95330 - 0.05129 - 4.58145 \approx -41.22$

Path 6

C	T	T	C	A	T	G	T	G	A	A	A	G	C	A	G	A	C	G	T	A	A	G	T	C	A
E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	5	I	I	I	

Tp	$(1.0)(0.9)^{21}(0.1)(1.0)(0.9)^{2}(0.1)$
EP	$(0.25)^{22}(0.95)(0.4)^{2}(0.1)$
PoP	$(0.9)^{23}(0.1)^{3}(0.25)^{22}(0.95)(0.4)^{2}$
Log-PoP	$\log\Pr(path6) = -2.42329 - 6.90776 - 30.49848 - 0.05129 - 1.83258 \approx -41.71$

Thus the best is Path 5 which one has a log probability of - 41.22, which infers that the most likely 5'SS position is at the fifth G.

4. Result and Analysis:

Now let us find the probability of the 8 possible paths when 5'SS fall on internal A'S are,

Path 1

C	T	T	C	A	T	G	T	G	A	A	A	G	C	A	G	A	C	G	T	A	A	G	T	C	A
E	E	E	E	5	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I

Transition Probability: $(1.0)(0.9)^{3}(0.1)(1.0)(0.9)^{20}(0.1)$

Emission Probability: $(0.25)^{4}(0.05)(0.4)^{12}(0.1)^{9}$

Probability of path 1 is product of transition probability and emission probability equal to $(0.9)^{23}(0.1)^{11}(0.25)^4(0.05)(0.4)^{12}$

The calculation is as follows

$23\log(0.9) = -2.42329$
$11\log(0.1) = -25.32844$
$4\log(0.25) = -5.5451$
$\log(0.05) = -2.99573$
$12\log(0.4) = -10.99549$

The log probability of path 1 is sum of log values

$\log\Pr(path1) = -2.42329 - 25.32844 - 5.5451 - 2.99573 - 10.99549 = -47.28805 \approx -47.29$

Path 2

C	T	T	C	A	T	G	T	G	A	A	A	G	C	A	G	A	C	G	T	A	A	G	T	C	A
E	E	E	E	E	E	E	E	5	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I

Transition Probability: $(1.0)(0.9)^8(0.1)(1.0)(0.9)^{15}(0.1)$

Emission Probability: $(0.25)^9(0.05)(0.4)^9(0.1)^7$

Probability of path 2 is product of transition probability and emission probability equal to $(0.9)^{23}(0.1)^9(0.25)^9(0.05)(0.4)^9$

The calculation is as follows

$23\log(0.9) = -2.42329$
$9\log(0.1) = -20.72327$
$9\log(0.25) = -12.47665$
$\log(0.05) = -2.99573$
$9\log(0.4) = -8.24662$

The log probability of path 2 is sum of log values

$\log\Pr(path2) = -2.42329 - 20.72327 - 12.47665 - 2.99573 - 8.24662 = -46.86556 \approx -46.87$

Path 3

C	T	T	C	A	T	G	T	G	A	A	A	G	C	A	G	A	C	G	T	A	A	G	T	C	A
E	E	E	E	E	E	E	E	E	5	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I

Transition Probability: $(1.0)(0.9)^9(0.1)(1.0)(0.9)^{14}(0.1)$

Emission Probability: $(0.25)^{10}(0.05)(0.4)^8(0.1)^7$

Probability of path 3 is product of transition probability and emission probability equal to $(0.9)^{23}(0.1)^9(0.25)^{10}(0.05)(0.4)^8$

The calculation is as follows

$23\log(0.9) = -2.42329$
$9\log(0.1) = -20.72327$
$10\log(0.25) = -13.86294$
$\log(0.05) = -2.99573$
$8\log(0.4) = -7.33033$

The log probability of path 3 is sum of log values

$\log \Pr(path3) = -2.42329 - 20.72327 - 13.86294 - 2.99573 - 7.33033 = -47.33556 \approx -47.33$

Path 4

$C \quad T \quad T \quad C \quad A \quad T \quad G \quad T \quad G \quad A \quad A \quad A \quad G \quad C \quad A \quad G \quad A \quad C \quad G \quad T \quad A \quad A \quad G \quad T \quad C \quad A$
$E \quad E \quad E \quad E \quad E \quad E \quad E \quad E \quad E \quad E \quad 5 \quad I \quad I \quad I \quad I \quad I \quad I \quad I \quad I \quad I \quad I \quad I \quad I \quad I \quad I \quad I$

Transition Probability: $(1.0)(0.9)^{10}(0.1)(1.0)(0.9)^{13}(0.1)$

Emission Probability: $(0.25)^{11}(0.05)(0.4)^7(0.1)^7$

Probability of path 4 is product of transition probability and emission probability equal to $(0.9)^{23}(0.1)^9(0.25)^{11}(0.05)(0.4)^7$

The calculation is as follows

$23\log(0.9) = -2.42329$
$9\log(0.1) = -20.72327$
$11\log(0.25) = -15.24924$
$\log(0.05) = -2.99573$
$7\log(0.4) = -6.41404$

The log probability of path 4 is sum of log values

$\log \Pr(path4) = -2.42329 - 20.72327 - 15.24924 - 2.99573 - 6.41404 = -47.80557 \approx -47.80$

Path 5

C	*T*	*T*	*C*	*A*	*T*	*G*	*T*	*G*	*A*	*A*	*A*	*G*	*C*	*A*	*G*	*A*	*C*	*G*	*T*	*A*	*A*	*G*	*T*	*C*	*A*
E	*E*	*E*	*E*	*E*	*E*	*E*	*E*	*E*	*E*	*E*	*E*	*E*	*E*	5	*I*	*I*	*I*	*I*	*I*	*I*	*I*	*I*	*I*	*I*	*I*

Transition Probability: $(1.0)(0.9)^{13}(0.1)(1.0)(0.9)^{10}(0.1)$

Emission Probability: $(0.25)^{14}(0.05)(0.4)^{6}(0.1)^{5}$

Probability of path 5 is product of transition probability and emission probability equal to $(0.9)^{23}(0.1)^{7}(0.25)^{14}(0.05)(0.4)^{6}$

The calculation is as follows

$23\log(0.9) = -2.42329$
$7\log(0.1) = -16.11810$
$14\log(0.25) = -19.40812$
$\log(0.05) = -2.99573$
$6\log(0.4) = -5.49774$

The log probability of path 5 is sum of log values

$\log \Pr(path5) = -2.42329 - 16.11810 - 19.40812 - 2.99573 - 5.49774 = -46.44298 \approx -46.44$

Path 6

C	*T*	*T*	*C*	*A*	*T*	*G*	*T*	*G*	*A*	*A*	*A*	*G*	*C*	*A*	*G*	*A*	*C*	*G*	*T*	*A*	*A*	*G*	*T*	*C*	*A*
E	*E*	*E*	*E*	*E*	*E*	*E*	*E*	*E*	*E*	*E*	*E*	*E*	*E*	*E*	5	*I*	*I*	*I*	*I*	*I*	*I*	*I*	*I*	*I*	*I*

Transition Probability: $(1.0)(0.9)^{15}(0.1)(1.0)(0.9)^{8}(0.1)$

Emission Probability: $(0.25)^{16}(0.05)(0.4)^{5}(0.1)^{4}$

Probability of path 6 is product of transition probability and emission probability equal to $(0.9)^{23}(0.1)^{6}(0.25)^{16}(0.05)(0.4)^{5}$

The calculation is as follows

$23\log(0.9) = -2.42329$
$6\log(0.1) = -13.81551$
$16\log(0.25) = -22.18071$
$\log(0.05) = -2.99573$
$5\log(0.4) = -4.58145$

The log probability of path 6 is sum of log values

$\log\Pr(path6) = -2.42329 - 13.81551 - 22.18071 - 2.99573 - 4.58145 = -45.99669 \approx -46.00$

Path 7

C	*T*	*T*	*C*	*A*	*T*	*G*	*T*	*G*	*A*	*A*	*A*	*G*	*C*	*A*	*G*	*A*	*C*	*G*	*T*	*A*	*A*	*G*	*T*	*C*	*A*
E	*E*	*E*	*E*	*E*	*E*	*E*	*E*	*E*	*E*	*E*	*E*	*E*	*E*	*E*	*E*	*E*	*E*	*E*	*E*	5	*I*	*I*	*I*	*I*	*I*

Transition Probability: $(1.0)(0.9)^{19}(0.1)(1.0)(0.9)^4(0.1)$

Emission Probability: $(0.25)^{20}(0.05)(0.4)^3(0.1)^2$

Probability of path 7 is product of transition probability and emission probability equal to $(0.9)^{23}(0.1)^4(0.25)^{20}(0.05)(0.4)^3$

The calculation is as follows

$23\log(0.9) = -2.42329$
$4\log(0.1) = -9.21034$
$20\log(0.25) = -27.72589$
$\log(0.05) = -2.99573$
$3\log(0.4) = -2.74887$

The log probability of path 7 is sum of log values

$\log\Pr(path7) = -2.42329 - 9.21034 - 27.72589 - 2.99573 - 2.74887 = -45.10412 \approx -45.10$

Path 8

C	*T*	*T*	*C*	*A*	*T*	*G*	*T*	*G*	*A*	*A*	*A*	*G*	*C*	*A*	*G*	*A*	*C*	*G*	*T*	*A*	*A*	*G*	*T*	*C*	*A*
E	*E*	*E*	*E*	*E*	*E*	*E*	*E*	*E*	*E*	*E*	*E*	*E*	*E*	*E*	*E*	*E*	*E*	*E*	*E*	*E*	5	*I*	*I*	*I*	*I*

Transition Probability: $(1.0)(0.9)^{20}(0.1)(1.0)(0.9)^3(0.1)$

Emission Probability: $(0.25)^{21}(0.05)(0.4)^2(0.1)^2$

Probability of path 8 is product of transition probability and emission probability equal to $(0.9)^{23}(0.1)^4(0.25)^{21}(0.05)(0.4)^2$

The calculation is as follows

$23\log(0.9) = -2.42329$

$4\log(0.1) = -9.21034$

$21\log(0.25) = -29.11218$

$\log(0.05) = -2.99573$

$2\log(0.4) = -1.83258$

The log probability of path 8 is sum of log values

$\log \Pr(path\, 8) = -2.42329 - 9.21034 - 29.11218 - 2.99573 - 1.83258 = -45.57412 \approx -45.57$

Internal G's Probability (when 5'SS falls on G)

Probability of 1st path $= 8.6218 \times 10^{-20}$

Probability of 2nd path $= 1.3472 \times 10^{-19}$

Probability of 3rd path $= 8.2224 \times 10^{-20}$

Probability of 4th path $= 3.2118 \times 10^{-19}$

Probability (best) of 5th path $= 1.2546 \times 10^{-18}$

Probability of 6th path $= 7.6577 \times 10^{-19}$

Total $= 2.6447 \times 10^{-18}$

Internal A's Probability (when 5'SS falls on A)

Probability (best) of 1st path $= 2.9042 \times 10^{-21}$

Probability of 2nd path $= 4.4314 \times 10^{-21}$

Probability of 3rd path $= 2.7696 \times 10^{-21}$

Probability of 4th path $= 1.7310 \times 10^{-21}$

Probability of 5th path $= 6.7618 \times 10^{-21}$

Probability of 6th path $= 1.0565 \times 10^{-20}$

Probability of 7th path $= 2.5794 \times 10^{-20}$

Probability of 8th path $= 1.6121 \times 10^{-20}$

Total $= 7.1078 \times 10^{-20}$

Grand Total = Sum of all probabilities of internal G's and internal A's

$$= 2.71577 \times 10^{-18}$$

The probability of all A path (8 paths) and G path (6 paths) has been calculated and finally measured Sum of all Probability (14 paths) $= 0.99785 \approx 1$

5. Real time assessment

A base pair is two engineered bases clung to one another outlining a "Rung of the DNA ladder." The DNA molecule includes two strands that breeze around each other like a bended ladder. Each strand has a spine made of turning sugar (deoxyribose) and phosphate social occasions. Joined to each sugar is one of four bases- - adenine (A), cytosine (C), guanine (G), or thymine (T). The two strands are held together by hydrogen associations between the bases, with adenine outlining a base pair with thymine, and cytosine forming a base pair with guanine.

Base pair depicts the association between the construction blocks on the strands of DNA. So every DNA particle is included two strands, and there are four nucleotides present in DNA: A, C, T, and G. Moreover, all of the nucleotides on one side of the strand sets with a specific nucleotide on the contrary side of the strand and this makes up the twofold helix. For instance, if there is a G on one side of the strand, there will reliably be a C (Cytosine) on the other strand. Assume by some coincidence if there is a T (Thymine) on one side of the strand, there will reliably be A (Adenine) on the other strand. Likewise, those nucleotides reliably pair. We furthermore count DNA and the proportion of DNA, or the length of DNA by using units of base sets. So we use base pair as a unit of assessment of DNA and RNA similarly as a term to depict the mixing relationship. In the United Kingdom ten STR loci are regularly utilized for recognizable proof purposes (Using the National DNA Database accepting D7S820 as standard for which there are mixes of allele lengths [Example 6 and 11 rehashes would appear as comparable to 11 and 6 rehashes on a gel $\dfrac{(14-6+1)[(14-6+1)+1]}{2} = 45$ combinations of allele lengths (recall that there are 6 to 14 repeats per allele for this STR, and know that the number of combinations of n objects taken 2 at a time $= \dfrac{n(n+1)}{2}$, when order is not important. [Example 6 and 11 repeats would show up as equivalent to 11 and 6 repeats on a gel

Hidden Markov models (HMM) have generally shown their handiness in the fields of statistics and pattern recognition to analyze hereditary qualities, similar standards of insights and likelihood. DNA fundamentally has four bases: adenine, guanine, thymine, and cytosine, which when combined together to form nucleotides. Nonetheless, the length of a nucleotide chain can be dubious. The DNA arrangement establishes the heritable hereditary data in crores that shapes the reason for the formative projects of every single living organic entity. Deciding the DNA succession is in this manner helpful in contemplating key natural cycles, just as in analytic or measurable units. In this investigation, we will use concealed Markov models (HMM) to decide DNA succession probabilities.

6. Conclusion

Under the tested above biological sequence the inference which has been derived very prominently identifies the hidden point path which put up many such point path. Thus it could be leading to generate a sequence or sequences.

The identification of a particular point with the highest probability under the testing sequence has prominently made in putting up the right label on each resistance

References

(1) Baum, L.E., and Eagon, J., "An Inequality with Application to Statistical Prediction for Function of Markov Processes and to a Model for Ecology, Bull.Amer.Math.Soc" 73(1963), 360-363.

(2) Baum, L.E., Petrie, T., Soules, G., and Weiss, N.,"A Maximization Techniques Occurring in the Statistical Analysis of Probabilistic Functions of Markov Chains," Ann. Math Statistic 41 (1970), PP. 164-171.

(3) Durbin, R., Eddy, S.R., Krogh, A. & Mitchison, G.J.Biological Sequence Analysis;

Probabilistic Models of Proteins and Nucleic Acids (Cambridge University Press-Cambridge UK, 1998)

(4) Elliott, R.J; Aggoun, L. & Moore, J.B. (1994).Hidden Markov Models; Estimation and Control. Springer-Verlag, New York

(5) Harte, D.S.(2005).Package, "Hidden Markov"; Discrete Time Hidden Markov Models. Statistics Research Associates, Wellington. URL WWW.Statsresearch.Co nz/Software.html.

(6) Juang, B.H., and Rabiner, L.R.,"Mixture Auto-Regressive Hidden Markov Models for Speech Signal Processing, vol.Assp-33,No.6, PP.1404-1413, Dec.1985.

(7) Juang, B.H.,"On the Hidden Markov Model and Dynamic Time Warping for Speech Recognition- A Unified View," AT&T B.L.T.J., Vol.63, No.7, PP 1213-1243, September 1984.

(8) Markel, J.D., and Gray, Jr., A.H., Linear Prediction of Speech, Springer- Verlag, New York, 1976.

(9) MacDonald, I.L. & Zucchini, W. (1997). Hidden Markov and other Models for Discrete valued Time Series. Chapman and Hall/CRC, Boca Raton. ISBN: 0-412-55850-5

(10) Poritz, A.B., "Linear predictive Hidden Markov Models and the Speech Signal," Proc. ICASSP '81, PP. 1291-1294, Paris, France, May 1982.

(11) Rabiner, L.R., Levinson, S.E., and Sindhi, M.M., "On the Application of Vector Quantization and Hidden Markov Models to Speaker-Independent, Isolated Word Recognition," Bell System Tech. J., Vol.62, No.4, PP. 1075-1105, April 1983.

(12) Rabiner, L.R., Juang, B.H., Levinson, S.E., and Sondhi, M.M., "Recognition of Isolated Digits Using Hidden Markov Models with Continuous Mixture Densities," AT&T B.L.T.J., Vol. 64, No, 3 PP. 1211-1234, July-August 1985.

(13) Rabiner, L.R. A Tutorial on Hidden Markov Models and selected applications in speech recognition. Proc. IEEE 77.257-286(1989).

(14) Rao, A.R. & Hamed, K.H. (2000). Flood Frequency Analysis. CRC, Boca Raton. ISBN; 0-412-55280-9

(15) Zucchini, W. (2005). Hidden Markov Models Short Course, 3-4 April 2005. Macquarie University, Sydney.

Appendix 1

The calculation of path 1 is as follows

$23\log(0.9) = -2.42329$
$10\log(0.1) = -23.025$
$6\log(0.25) = -8.31777$
$\log(0.95) = -0.05129$
$11\log(0.4) = -10.07920$

The calculation of path2 is as follows

$23\log(0.9) = -2.42329$
$9\log(0.1) = -20.72327$
$8\log(0.25) = -11.09035$
$\log(0.95) = -0.05129$
$10\log(0.4) = -9.16291$

The calculation path 3 is as follows

$23\log(0.9) = -2.42329$
$8\log(0.1) = -18.42068$
$12\log(0.25) = -16.63553$
$\log(0.95) = -0.05129$
$7\log(0.4) = -6.41404$

The calculation path 4 is as follows

$23\log(0.9) = -2.42329$
$6\log(0.1) = -13.81551$
$15\log(0.25) = -20.79442$
$\log(0.95) = -0.05129$
$6\log(0.4) = -5.49774$

The calculation path 5 is as follows

$23\log(0.9) = -2.42329$
$4\log(0.1) = -9.21034$
$18\log(0.25) = -24.95330$
$\log(0.95) = -0.05129$
$5\log(0.4) = -4.58145$

The calculation path 6 is as follows

$23\log(0.9) = -2.42329$

$3\log(0.1) = -6.90776$

$22\log(0.25) = -30.49848$

$\log(0.95) = -0.05129$

$2\log(0.4) = -1.83258$

Appendix 2

Probability of G's paths (6 paths)

$$\text{Probability (G's path 1)} = \frac{(8.6218 \times 10^{-20})}{(2.71577 \times 10^{-18})}$$

$$= 0.0317$$
$$= 3\%$$

$$\text{Probability (G's path 2)} = \frac{(1.3472 \times 10^{-19})}{(2.71577 \times 10^{-18})}$$

$$= 0.0496$$
$$= 4.9\%$$

$$\text{Probability (G's path 3)} = \frac{(8.2224 \times 10^{-20})}{(2.71577 \times 10^{-18})}$$

$$= 0.030$$
$$= 3\%$$

$$\text{Probability (G's path 4)} = \frac{(3.2118 \times 10^{-19})}{(2.71577 \times 10^{-18})}$$

$$= 0.1182$$
$$= 11.8\%$$

$$\text{Probability (G's path 5)} = \frac{(1.2546 \times 10^{-18})}{(2.71577 \times 10^{-18})}$$

$$= 0.4619$$
$$= 46.19\%$$

$$\text{Probability (G's path 6)} = \frac{(7.6577 \times 10^{-19})}{(2.71577 \times 10^{-18})}$$

$$= 0.2819$$
$$= 28.19\%$$

Probability of A's paths (8 paths)

$$\text{Probability (A's path 1)} = \frac{(2.9042 \times 10^{-21})}{(2.71577 \times 10^{-18})}$$

$$= 0.001$$
$$= 0.106\%$$

$$\text{Probability (A's path 2)} = \frac{(4.4314 \times 10^{-21})}{(2.71577 \times 10^{-18})}$$

$$= 0.001632$$
$$= 0.1631\%$$

$$\text{Probability (A's path 3)} = \frac{(2.7696 \times 10^{-21})}{(2.71577 \times 10^{-18})}$$

$$= 0.001019$$
$$= 0.1019\%$$

$$\text{Probability (A's path 4)} = \frac{(1.7310 \times 10^{-21})}{(2.71577 \times 10^{-18})}$$

$$= 0.000637$$
$$= 0.0637\%$$

$$\text{Probability (A's path 5)} = \frac{(6.7618 \times 10^{-21})}{(2.71577 \times 10^{-18})}$$

$$= 0.002490$$
$$= 0.2489\%$$

$$\text{Probability (A's path 6)} = \frac{(1.0565 \times 10^{-20})}{(2.71577 \times 10^{-18})}$$

$$= 0.003890$$
$$= 0.3890\%$$

Probability (A's path 7) $= \dfrac{(2.5794 \times 10^{-20})}{(2.71577 \times 10^{-18})}$

$$= 0.009498$$
$$= 0.9498\%$$

Probability (A's path 8) $= \dfrac{(1.61215 \times 10^{-20})}{(2.71577 \times 10^{-18})}$

$$= 0.005936$$
$$= 0.5936\%$$