



Journal of Applied Sciences

ISSN 1812-5654

science
alert

ANSI*net*
an open access publisher
<http://ansinet.com>

Gene Finding Using Hidden Markov Model

Srabanti Maji and Deepak Garg

Department of Computer Science and Engineering, Thapar University, Patiala-147004, India

Abstract: The objective of this study is to perform mini review on Hidden Markov Models (HMMs) which is recently important and popular among bioinformatics researchers and large no of software tools are based on this technique. The mathematical foundations of HMMs shall be considered first in brief manner and then the gene identification application. In the case of gene identification process, HMM basically resolve three basic problems: First is the evaluation problem, in this it computes the probability that a particular HMM will generate a given sequence of observations. Second is Decoding problem, in which it will uncover the most likely hidden state and Third is Learning problem, it is used to adjust the model parameter and train the HMM to find an optimal model. Evaluation problem can be solved by using Forward and Backward algorithm, Decoding problems are solved by using Viterbi algorithm and posterior decoding algorithm and then Learning problems are solved through Viterbi training algorithm and Baum-Welch algorithm. Finally, some limitations of the current approaches and future directions are also reported.

Key words: Markov chain, HMM, bioinformatics, gene finding

INTRODUCTION

The usually applied technique which is quite prevailing too, is Markov Processes and Markov Chains (De Fonzo *et al.*, 2007; Nasiri, 2011). A Markov process is a stochastic process satisfying a certain property, called the Markov Property. A process satisfies the Markov property if one can make predictions for the future of the process based exclusively on its present state just as well as one could express the process's full history (Parent, 2004; Cawley and Pachter, 2003).

A Markov chain is a first-order Markov process for which the probability distribution of a state at a given time is explicitly dependent only on the previous state and not on all the others (Garg, 2007a; Lifshits *et al.*, 2009). More specifically there is a finite set of possible states and the transitions among them are governed by a set of conditional probabilities of the next state given the present one, called transition probabilities (Garg, 2007b; Kumar and Raghava, 2009). The transition probabilities are implicitly (unless declared otherwise) independent of the time and then one speaks of homogeneous, or stationary, Markov chains. In case of DNA sequence; the "time" means the position along the sequence (Yoon and Vaidyanathan, 2004). Starting from a given initial state, the consecutive transitions from a state to the next one produce a time-evolution of the chain that is therefore completely represented by a sequence of states that a priori are to be considered random (Lunter, 2007; Dosay-Akbulut, 2006).

A Hidden Markov Model consists of two stochastic processes. The first stochastic process is a Markov chain that is characterized by states and transition probabilities (Ahmad, 2011; Tran *et al.*, 2009). The states of the chain are externally not visible, therefore "hidden". Another stochastic process will generate emissions which is observable at every instant. It is dependent on state probability distribution (Meyer and Durbin, 2002). In case of Hidden Markov Model the term "hidden" not indicates the parameter of the model, but it indicates the state of the Markov Chain (El-Sayed and Khedr, 2007).

A Hidden Markov Model is a generalization of a Markov chain, in which each ("internal") state is not directly observable (hence the term hidden) but produces ("emits") an observable random output ("external") state, also called "emission", according to a given stationary probability law (Oron *et al.*, 2008). In this case, the time evolution of the internal states can be induced only through the sequence of the observed output states.

If the number of internal states is N , the transition probability law is described by a matrix with N times N values; if the number of emissions is M , the emission probability law is described by a matrix with N times M values. A model is considered defined once given these two matrices and the initial distribution of the internal states. The study by Rabiner (Jing *et al.*, 2006; Frikha *et al.*, 2007; Rabiner, 1989) is widely well appreciated for clarity in explaining HMMs. It is a powerful class of model used in many fields including

gene finding, profile searches, error correction coding, multiple sequence alignment, speech recognition and regulatory site identification.

BIOLOGICAL BACKGROUND

The central dogma of molecular biology pertains to DNA, RNA and Proteins. DNA can be converted to RNA by a process known as Transcription and RNA to Proteins by a process known as Translation. DNA can be represented by 4 characters, A, C, T and G. These are called which are called nucleotides. The same characters except for the nucleotide “T” which is replaced by “U” can represent RNA (Cheng and Zou, 2007; Kim, 2006; Pop and Salzberg, 2008). In case of proteins, the representation consists of 20 characters, corresponding to 20 amino acids of which they are composed (Sur, 2008). A one-to-one letter mapping occurs between a DNA molecule and its associated RNA molecule and a three-to-one letter mapping occurs between the RNA molecule and its associated protein molecule (Nath and Bhattacharjee, 2011; Tran *et al.*, 2008). The coordinates are known in rare cases where a protein sequence folds in a defined three dimensional structure. The defined structure is what actually provides the molecular function of the protein sequence.

Determining the DNA and RNA sequences are relatively cheaper than determining protein sequences. One of the most successful class of techniques for analyzing biological sequences has been the use of HMM (Larik and Soomro, 2003; Yang and Yang, 2008). HMM are mainly used for predicting protein sequences since it provides a good probabilistic framework for the same. Considering the work done, there are mainly two everyday tasks for bioinformatics: Deducing the protein DNA

sequence and Comparing protein sequences to an existing database of protein sequences, both of which utilize Hidden Markov Models.

Mathematical basics and Elements of HMM: A Hidden Markov Model is a finite learnable stochastic automate. It can be summarized as a kind of double stochastic process with the two following aspects: (a) the firstly stochastic process is a finite set of states, where each of them is generally associated with a multidimensional probability distribution (Wang *et al.*, 2008). The transitions between the different states are statistically organized by a set of probabilities called transition probabilities. (b) Secondly, stochastic process, in any state an event can be observed. Since we will just analyze what we observe without seeing at which states it occurred, the states are "hidden" to the observer, therefore the name "Hidden Markov Model".

Each Hidden Markov Model is defined by states, state probabilities, transition probabilities, emission probabilities and initial probabilities (Bhardwaj, 2007; Lee *et al.*, 2008). For the sake of simplicity, in the following notations we consider only one sequence of internal states and one sequence of associated emissions. Table 1 shows the notations.

BASICS OF HMM IN STOCHASTIC MODELING

There are important aspects of modeling Hidden Markov Models in order to solve real problems. The following two steps are included in the stochastic modeling of an HMM automate-first, to define the model structure; and second, to define the learning and operating algorithm.

Definition of HMM architecture: The generalized computational architecture of an operating HMM λ_i with two integrated stochastic processes is shown in Fig. 1.

Table 1: HMM Notations

U	The set of all the N possible internal states
X	The set of all the M possible external states
L	The length of the sequence
k	A time instant, where $k \in [1, \dots, L]$
s_k	Internal state at time k, where $s_k \in U$
$S = (s_1, s_2, s_3, \dots, s_L)$	A sequence of L internal states
e_k	Emission at time k, where $e_k \in X$
$E = (e_1, e_2, e_3, \dots, e_L)$	A sequence of L external states
$a_{uv} = P(s_k = v s_{k-1} = u)$	The probabilities of a transition from the state u to the state v
A	The NXN matrix of elements $a_{u,v}$
$b_u(x) = P(e_k = x s_k = u)$	The probabilities of the emission x from the state u
B	The NXM matrix of elements $b_u(x)$
$\pi_u = P(s_1 = u)$	The probability of the initial state u
Π	The N vector of elements π_u
$\lambda = (A, B, \Pi)$	The definition of the HMM model

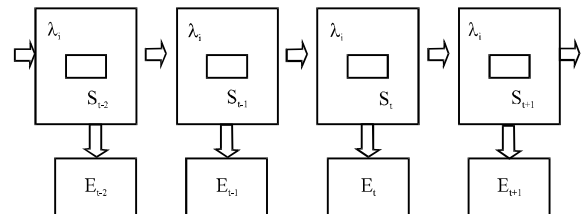


Fig. 1: Generalized Architecture of an operating Hidden Markov Model. In the above model, λ is HMM variable, there are 2 different state variable in the model, hidden state variable S_i and observation variable E_i . In this above HMM, 4 different times like t-2, t-1, t and t+1 are considered

Each shape represents a random variable that can adopt any of a number of values. The random variable S_t is the hidden state at time t . The random variable E_t is the observation at the time t . The law of conditional probability of the Hidden Markov variable S_t at the time t , knowing the values of the hidden variables at all times depends only on the value of the hidden variable S_{t-1} at the time $t-1$. Every values before are not necessary anymore, so that the Markov property as defined before is satisfied. By the second stochastic process, the value of the observed variable E_t depends on the value of the hidden variable S_t also at the time t .

Definition of the learning and operating algorithms—three basic problems of HMMs: The main types of problems occurring in the use of Hidden Markov Models are:

Evaluation problem: Compute the probability that a given model generates a given sequence of observations. The most used algorithms are: (a) the forward algorithm: find the probability of emission distribution (given a model) starting from the beginning of the sequence. (b) the backward algorithm: find the probability of emission distribution (given a model) starting from the end of the sequence.

Decoding problem: Given a model and a sequence of observations, induce the most likely hidden states. In this stage we attempt to uncover the hidden part of the model, i.e., to find the “correct” state sequence. More specifically: (a) find the sequence of internal states that has, as a whole, the highest probability. The most used algorithm is the Viterbi algorithm. (b) find for each position the internal state that has the highest probability. The most used algorithm is the posterior decoding algorithm.

Learning problem: given a sequence of observations, find an optimal model. The observation sequence used to adjust the model parameters is called a training sequence, because it is used to “train” the HMMs. The training problem is very crucial for most applications of HMMs. The most used algorithms start from an initial guessed model and iteratively adjust the model parameters. More specifically: (a) find the optimal model based on the most probable sequences (as in problem B1). The most used algorithm is the Viterbi training (that uses recursively the Viterbi algorithm in B1). (b) find the optimal model based on

the sequences of most probable internal states (as in problem B2). The most used algorithm is the Baum-Welch algorithm (that uses recursively the posterior decoding algorithm in B2).

Gene finding: The term “gene finding” indicates the action of finding genes within a DNA sequence, but is often used with a more general meaning of labeling DNA tracts (Burge and Karlin, 1997), for example labeling them as coding, intergenic, introns, etc. In this last sense gene finding can be considered a special case (the most important in bioinformatics) of the more general action known as sequence labeling (also for non-DNA sequences). Determining the DNA and RNA sequences are relatively cheaper than determining protein sequences. One of the most successful class of techniques for analyzing biological sequences has been the use of HMM. HMM are mainly used for predicting protein sequences since it provides a good probabilistic framework for the same (Birney, 2001; Nath and Bhattacharjee, 2011). Considering the work done, there are mainly two everyday tasks for bioinformatics: Deducing the protein DNA sequence and Comparing protein sequences to an existing database of protein sequences, both of which utilize Hidden Markov Models. In the early 1990s, Krogh *et al.* (1994) introduced the use of HMMs for discriminating coding and intergenic regions in *E. coli* genome. The program GeneMark (Borodovsky and McIninch, 1993) finds genes in *E. coli* DNA using a Markov model for the coding region. It is a parser with a complex intergenic model. The more complex HMM (Fig. 2), intergenic model consists of several parts in addition to the start and stop codon models. After generating the stop codon, the model chooses either the transition to the long intergenic HMM or the short intergenic HMM, with appropriate probabilities. The short intergenic HMM tends to generate intergenic regions of lengths from 1 to 14 or so, with statistics determined from examples of such short intergenic regions in actual *E. coli* contigs. Similarly, the parameters of the long intergenic model are adjusted to capture the statistics of longer intergenic regions. The parameters of the two intergenic models were estimated from a set of known intergenic regions by a learning procedure known as the forward-backward algorithm. As a result of the training process, the long intergenic region develops patterns, without having to explicitly encode them. The complex parser has a better accuracy.

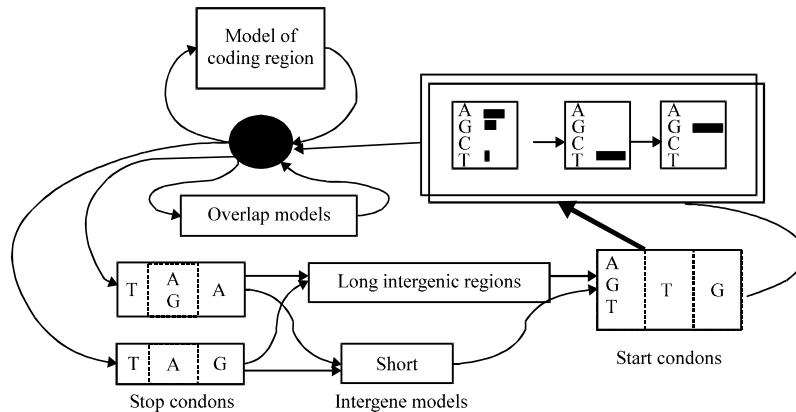


Fig. 2: HMM architecture for a parser for *E. coli* DNA with a complex intergenic model. The gene model above the central state that contains the 61 triplet. In the DNA sequence A, T, G, C is the four codon characters. This intergenic model consists of several parts in addition to the start and stop codon models. After generating the stop codon, the model chooses either the transition to the long intergenic HMM or the short intergenic HMM, with appropriate probabilities. The short intergenic HMM tends to generate intergenic regions of lengths from 1 to 14 or so, parameters of the long intergenic model are adjusted to capture the statistics of longer intergenic regions

The gene model above the central state that contains the 61 triplet. In the DNA sequence A, T, G, C is the four codon characters. This intergenic model consists of several parts in addition to the start and stop codon models. After generating the stop codon, the model chooses either the transition to the long intergenic HMM or the short intergenic HMM, with appropriate probabilities. The short intergenic HMM tends to generate intergenic regions of lengths from 1 to 14 or so, parameters of the long intergenic model are adjusted to capture the statistics of longer intergenic regions.

Many extensions to the original “pure” HMM have been developed for gene finding. For example, Henderson *et al.* (1997) designed separate HMM modules, each one appropriate for a specific region of DNA. Separate HMM modules were designed and trained for specific regions of DNA :exons, introns, intergenic regions and splice sites. In order to form a biologically viable topology, the models were coupled. Additionally, the integrated HMM was trained on a set of eukaryotic DNA sequences and then tested by using an unknown DNA sequences. Kulp *et al.* (1996) and Burge and Karlin (1998) used a Generalized HMM (GHMM or “hidden semi-Markov Model”) that allows more than one emission for each internal state.

A pair Hidden Markov Model (Durbin *et al.*, 1998) is having large no of similarity with standard HMM.

Basic difference is that it emits pairwise alignment, where as standard HMM emits a single sequence. This method provides parse only alignments between two sequences but, with suitable enhancements, it is sometimes applied to gene finding. For example, Meyer and Durbin (2002) resented a new method that predicts the gene structure starting from two homologous DNA sequences, identifying the conserved subsequences. A useful open-source implementation is described by Majoros *et al.* (2005). Lukashin and Borodovsky (1998) proposed a new algorithm (GeneMark.hmm) that improves the gene finding performance of the old GeneMark algorithm by means of a suitable coupling with an HMM model. Pedersen and Hein (2003) introduced an Evolutionary Hidden Markov Model (EHMM), based on a suitable coupling of an HMM and a set of evolutionary models based on a phylogenetic tree.

We propose an oversimplified biological example of an HMM, inspired by the toy example by Eddy (2004) with only two internal states but with exponential complexity. The model is detailed in Fig. 3. The set of internal states is $U \equiv \{ 'c', 'n' \}$ where 'c' and 'n' stand for the coding and non-coding internal states and the set of emissions is the set of the four DNA bases:

$$X \equiv \{ 'A', 'T', 'G', 'C' \}$$

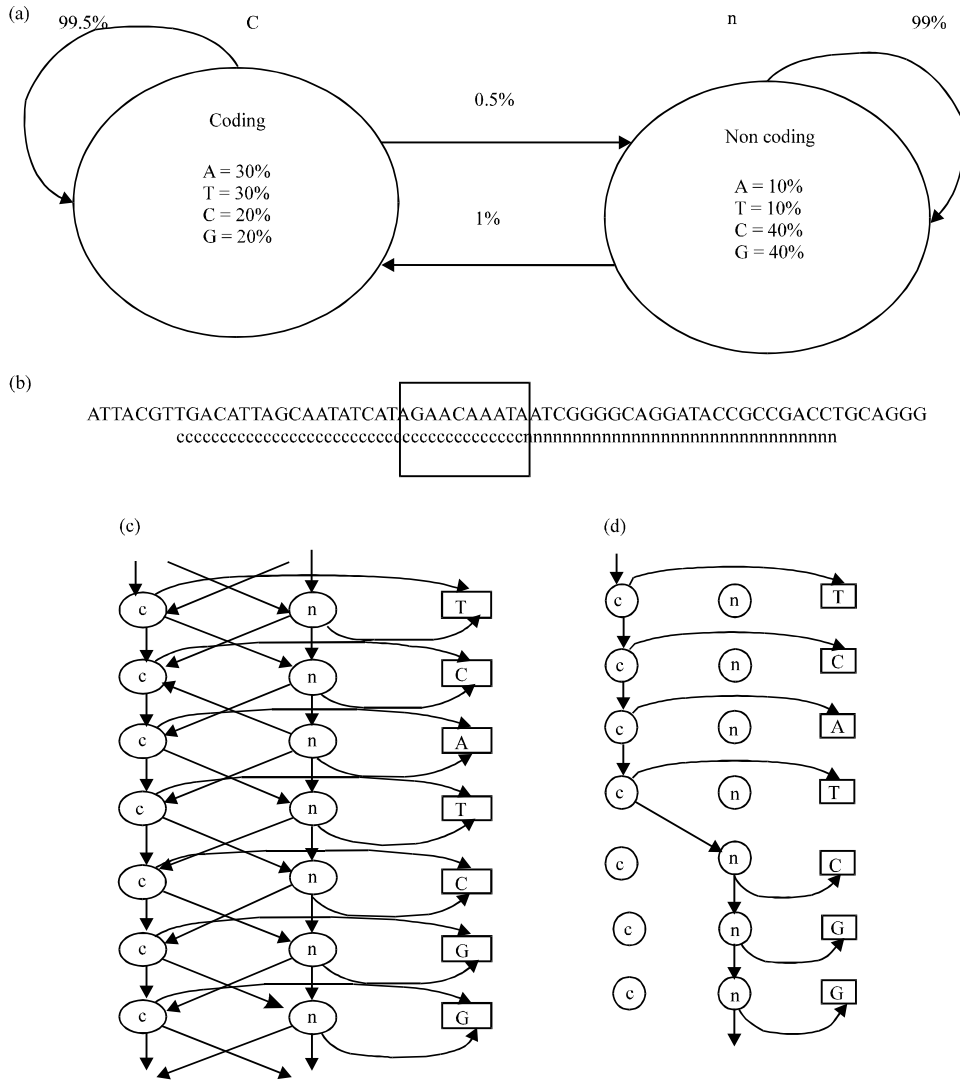


Fig. 3: An example of HMM, (a) The circular boxes represent the internal states 'c', indicating the coding and 'n' non coding or intron, inside the boxes there are the probabilities of each emission ('A', 'T', 'C' and 'G') for each state; outside the boxes four arrows are labelled with the corresponding transition probability, (b) The 66 observed emissions are representing a sample sequence and most likely, the sequences of internal states are given in the second row, the dotted part is depicted in (c) and (d), (c) The right-hand side column represents the boxed tract of bases in (b), the other columns represent the two possible alternatives (for each squared base) for the internal state ('c' or 'n') that emitted the base, each row refers to the same position along the sequence, the arrows represent all possible transitions and the emissions, (d) The figure shows a likely sequence of choices between the alternative internal states producing its sequence in (b), such a sequence of choices of internal state transitions amounts to choosing a path in (c)

We are given a DNA sequence that begins in an exon, contains one 5' splice site and ends in an intron. The problem is to identify-where the 5' splice site (5'SS) is located. Let's imagine that exons have a uniform base composition on average (25% each base), introns are A/T rich (say, 40% each for A/T, 10% each for C/G) and the

5'SS consensus nucleotide is almost always a G (say, 95 G and 5% A).

Figure 4 illustrates the action, on the same tract of the sequence in Fig. 3b, of the Viterbi algorithm used to decode the whole sequence by means of the model described in Fig. 3a.

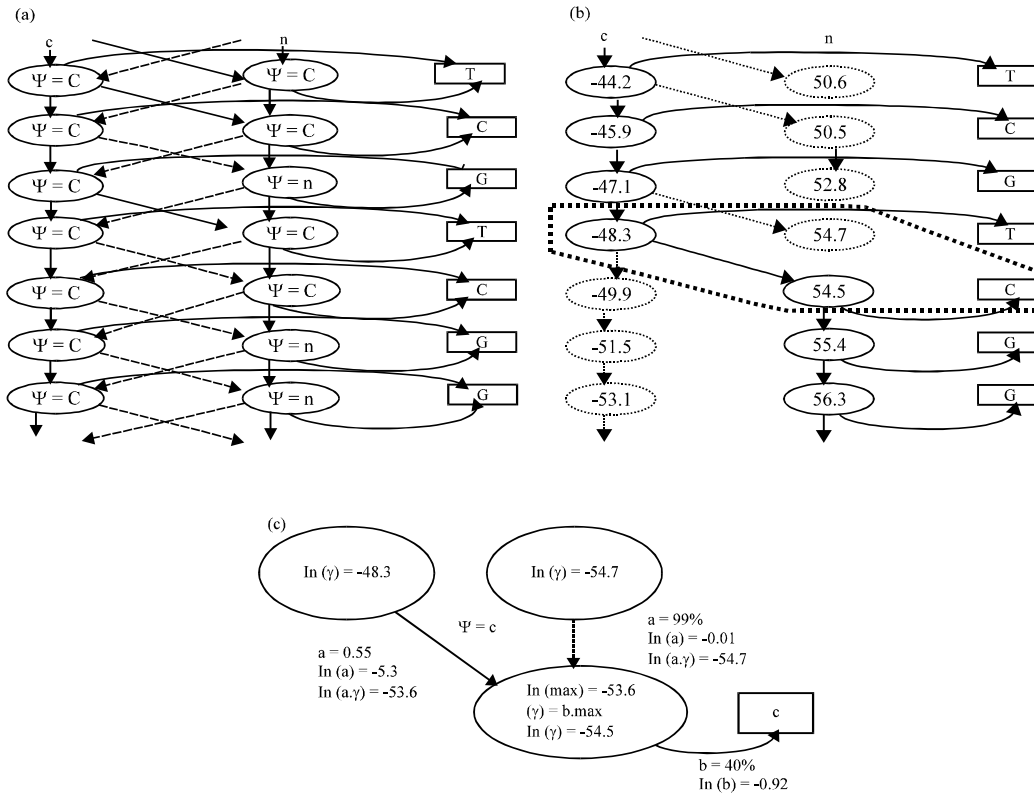


Fig. 4: Illustration of the action (in the same sequence) of the Viterbi algorithm used to decode the whole sequence by means of the model described in Fig. 3 (a), more specifically, 4 (a) and 4 (b) illustrates the transition from Fig. 3 (c), (d), (a) In each circular box, there is the value of the Ψ pointer, computed in the first phase, as illustrated in 4 (c), specifically, “ $\Psi = C$ ” means that we discard the hypothesis of the transition from the previous state 'n' (also indicated by dashing the corresponding incoming arrow), (b) In each circular box, the logarithmic value of the probability γ is shown, which is calculated and used in the first phase, dashed lines represent the transitions discarded in the second phase, note that logarithms of the probabilities are used in order to avoid troubles because of very small numbers, necessary for practical applications, (c) A zoom of the marked zone in 4 (b), where the computation of a recursion step of the Viterbi algorithm is done, is illustrated

CONCLUSIONS

Considering the present scenario of the HMM in bioinformatics, from the time of its introduction and from the wealth of available applications, it may be concluded that the concept has reached a developed state. In addition, since the beginning of the field, novel applications have been fostered by so many different extensions and modifications in respective techniques thus producing models that can be considered even today are still known as Hidden Markov Models.

The Hidden Markov Model can be used in various machine learning techniques. Main advantages of HMMs are-the ease of use, less training sets required and deeper understanding of the phenomenon due to the observation of the inner structure of the model. Various gene prediction tools are available which are based on HMM. The gene prediction tool Genscan can predict the gene

with accuracy of 80%. Several improvements have been reported in HMM based gene prediction tools like VEIL (Viterbi Exon-Intron Locator), having overall accuracy of 92 %. The correlation coefficient of VEIL, on which total bases are correctly labeled, is 0.73. To improve predictive accuracy, hybrid models are frequently designed with the combination of techniques like Support Vector Machine and Artificial Neural Network.

REFERENCES

- Ahmad, M., A. Abdullah and K. Buragga, 2011. A novel optimized approach for gene identification in DNA sequences. *J. Applied Sci.*, 11: 806-814.
- Bhardwaj, L.M., D. Garg and S.C. Saxena, 2007. NCM-2 for normalizing blast values for simple regions. *J. Bioinform. Trends*, 2: 29-36.

- Bimey, E., 2001. Hidden Markov models in biological sequence analysis. *IBM J. Res. Dev.*, 45: 449-454.
- Borodovsky, M. and J. McIninch, 1993. GENMARK: Parallel gene recognition for both DNA strands. *Comp. Chem.*, 17: 123-133.
- Burge, C. and S. Karlin, 1997. Prediction of complete gene structures in human Genomic DNA. *J. Mol. Biol.*, 268: 78-94.
- Burge, C.B. and S. Karlin, 1998. Finding the genes in genomic DNA. *Curr. Opin. Struct. Biol.*, 8: 346-354.
- Cawley, S.L. and L. Pachter, 2003. HMM sampling and applications to gene finding and alternative splicing. *Bioinformatics*, 19: ii36-ii41.
- Cheng, K. and C.H. Zou, 2007. Informatics models of recognitions of protein synthesis. *Asian J. Biochem.*, 2: 432-436.
- De Fonzo, V., F. Aluffi-Pentini and V. Parisi, 2007. Hidden markov models in bioinformatics. *Curr. Bioinformatics*, 2: 49-61.
- Dosay-Akbulut, M., 2006. Group I introns and splicing mechanism and their present possibilities in elasmobranchs. *J. Biol. Sci.*, 6: 921-925.
- Durbin, R., S. Eddy, A. Krogh and G. Mitchison, 1998. *Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids*. Cambridge University Press. Cambridge.
- Eddy, S.R., 2004. What is a hidden markov model? *Nat. Biotechnol.*, 22: 1315-1316.
- El-Sayed, M.H. and A.M. Khedr, 2007. Improving the performance of an HMM for protein family modelling. *J. Applied Sci.*, 7: 1626-1632.
- Frikha, M., Z.B. Messaoud and A.B. Hamida, 2007. Towards discriminative training estimators for HMM speech recognition system. *J. Applied Sci.*, 7: 3891-3899.
- Garg, D., 2007a. Efficient algorithm for pattern discovery in bioinformatic sequences. Ph.D. Thesis, Thapar University, India.
- Garg, D., 2007b. Multi hit, dropoff percentage and NCM-2: Three improvements in BLAST. *J. Clin. Diagn. Res.*, 1: 339-346.
- Henderson, J., S. Salzberg and K.H. Fasman, 1997. Finding genes in DNA with a hidden markov model. *J. Comput. Biol.*, 4: 127-141.
- Jing, L., H. Hua and L. Sufang, 2006. Voice identification based on HMMs. *Trends Applied Sci. Res.*, 1: 79-82.
- Kim, J.S., 2006. Characterization of a gene encoding DNA repair protein XRCC3 from *Arabidopsis thaliana*. *Int. J. Bot.*, 2: 23-28.
- Krogh, A., I.S. Mian and D. Haussler, 1994. A hidden Markov model that finds genes in *Escherichia coli* DNA. *Nucleic Acids Res.*, 22: 4768-4778.
- Kulp, D., D. Haussler, M.G. Reese and F.H. Eeckman, 1996. A generalized hidden Markov model for the recognition of human genes in DNA. *Proc. Int. Conf. Intell. Syst. Mol. Biol.*, 4: 134-142.
- Kumar, M. and G.P.S. Raghava, 2009. Prediction of nuclear proteins using SVM and HMM models. *BMC Bioinformatics*, Vol. 10.
- Larik, A.S. and Z.A. Soomro, 2003. Elucidation of gene structure and function of prokaryotes and eukaryotes through DNA information technology. *Asian J. Plant Sci.*, 2: 551-560.
- Lee, S.A., C.H. Chan, C.H. Tsai, J.M. Lai, F.S. Wang, C.Y. Kao and C.Y.F. Huang, 2008. Ortholog-based protein-protein interaction prediction and its application to inter-species interactions. *BMC Bioinformatics*, Vol. 9.
- Lifshits, Y., S. Mozes, O. Weimann and M. Ziv-Ukelson, 2009. Speeding up HMM decoding and training by exploiting sequence repetitions. *Algorithmica*, 54: 379-399.
- Lukashin, A.V. and M. Borodovsky, 1998. Genemarkhmm: New solutions for gene finding. *Nucleic Acids Res.*, 26: 1107-1115.
- Lunter, G., 2007. HMMoC: A compiler for hidden Markov models. *Bioinformatics*, 23: 2485-2487.
- Majoros, W.H., M. Pertea and S.L. Salzberg, 2005. Efficient implementation of a generalized pair hidden Markov model for comparative gene finding. *Bioinformatics*, 21: 1782-1788.
- Meyer, I.M. and R. Durbin, 2002. Comparative *ab initio* prediction of gene structures using pair HMMs. *Bioinformatics*, 18: 1309-1318.
- Nasiri, J., A. Haghazari and M. Alavi, 2011. Evaluation of prediction accuracy of genefinders using mouse genomic DNA. *Trends Bioinform.*, 4: 10-22.
- Nath, D.C. and A. Bhattacharjee, 2011. A Bayesian approach for autoregressive models in longitudinal data analysis: An application to type 2 diabetes drug comparison. *Asian J. Applied Sci.*, 4: 640-648.
- Oron, A.P., Z. Jiang and R. Gentleman, 2008. Gene set enrichment analysis using linear models and diagnostics. *Bioinformatics*, 24: 2586-2591.
- Parent, A., I. Benzaghoul, I. Bougie and M. Bisailon, 2004. Transcription and mRNA processing events: The importance of coordination. *J. Biol. Sci.*, 4: 624-627.
- Pedersen, J.S. and J. Hein, 2003. Gene finding with a hidden Markov model of genome structure and evolution. *Bioinformatics*, 19: 219-227.
- Pop, M. and S.L. Salzberg, 2008. Bioinformatics challenges of new sequencing technology. *Trends Genet.*, 24: 142-149.

- Rabiner, L.R., 1989. A tutorial on hidden Markov models and selected applications in speech recognition. *Proc. IEEE*, 77: 257-286.
- Sur, S., M. Bhattacharya, A.K. Bothra, L.S. Tisa and A. Sen, 2008. Bioinformatic analysis of codon usage patterns in a free-living diazotroph, *Azotobacter vinelandii*. *Biotechnology*, 7: 242-249.
- Tran, D.H., K. Satou and T.B. Ho, 2008. Finding microRNA regulatory modules in human genome using rule induction. *BMC Bioinformatics*, Vol. 9. 10.1186/1471-2105-9-S12-S5
- Tran, T.T., F. Zhou, S. Marshburn, M. Stead, S.R. Kushner and Y. Xu, 2009. *De novo* computational prediction of non-coding RNA genes in prokaryotic genomes. *Bioinformatics*, 25: 2897-2905.
- Wang, X., D. Wu, S. Zheng, J. Sun, L. Tao, Y. Li and Z. Cao, 2008. Ab-origin: An enhanced tool to identify the sourcing gene segments in germline for rearranged antibodies. *BMC Bioinformatics*, Vol. 9.
- Yang, J.Y. and M.Q. Yang, 2008. Identification of intrinsically unstructured proteins using hierarchical classifier. *Int. J. Data Mining Bioinformatics*, 2: 121-133.
- Yoon, B.J. and P.P. Vaidynathan, 2004. HMM with auxiliary memory: A new tool for modeling RNA structures. *Asilomar Conf. Signals Syst. Comput.*, 2: 1651-1655.