

AI METHODS IN BIOINFORMATICS

Iosif Vaisman

1998

Scientific Models

Physical models -- Mathematical models

Mechanistic models

Mechanism

Predictive power
Elegance
Consistency

Stochastic models

Black box

Predictive power

Hidden Markov models

Stochastic mechanism

Artificial Intelligence in Biosciences

- Neural Networks (NN)
- Genetic Algorithms (GA)
- Formal Grammars (FG)

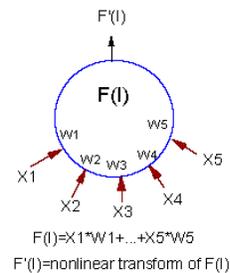
Artificial Intelligence in Biosciences

- Neural Networks (NN)
- Genetic Algorithms (GA)
- Formal Grammars (FG)

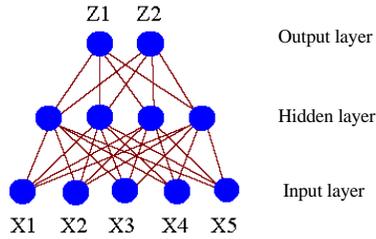
Neural Networks

- interconnected assembly of simple processing elements (units or nodes)
- nodes functionality is similar to that of the animal neuron
- processing ability is stored in the inter-unit connection strengths (weights)
- weights are obtained by a process of adaptation to, or *learning* from, a set of training patterns

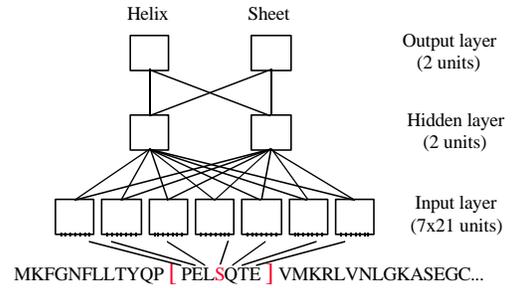
Neural Networks



Neural Networks



Neural Networks



Artificial Intelligence in Biosciences

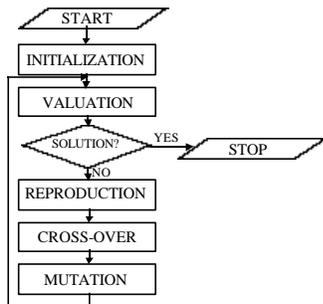
- Neural Networks (NN)
- Genetic Algorithms (GA)
- Formal Grammars (FG)

Genetic Algorithms

Search or optimization methods using simulated evolution.
 Population of potential solutions is subjected to natural selection, crossover, and mutation

choose initial population
 evaluate each individual's fitness
 repeat
 select individuals to reproduce
 mate pairs at random
 apply crossover operator
 apply mutation operator
 evaluate each individual's fitness
 until terminating condition

Genetic Algorithms



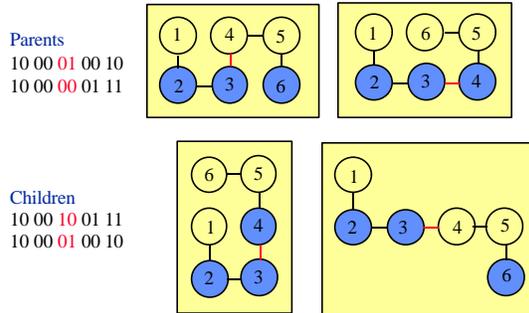
Crossover



Mutation



Genetic Algorithms Applications



Artificial Intelligence in Biosciences

Neural Networks (NN)
Genetic Algorithms (GA)
Formal Grammars (FG)

Grammars and Language

gram•mar *n.*

1. the study of the way the sentences of a language are constructed
- ...
4. *Generative Gram.* a device, as a body of rules, whose output is all of the sentences that are permissible in a given language, while excluding all those that are not permissible.

Random House Unabridged Dictionary

Language Components

Semantics (meaning)
Syntax (structure, form)

Language Syntax

Alphabet

Primitive elements
Letters, phonemes

Vocabulary

Elements composed from the alphabet
Words, phrases, sentences,...

Grammar

Legal composition of vocabulary
Rules, operators

Semantics

Derived from syntax

Semantic content derived from vocabulary within a context

Vocabulary element has its own meanings
dictionary lookup
meanings depending on context

Time **flies like an arrow**
Fruit **flies like a banana**

Formal Grammars

formal grammar

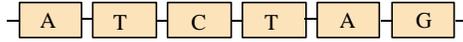
a means for specifying the syntactic structure of natural language by a set of transformation functions

Chomsky hierarchy (for string grammars)

type 0: phrase structure
type 1: context sensitive
type 2: context free (SCFG)
type 3: regular (Hidden Markov models)

Chomsky, *Syntactic Structures* (1957)

Markov Model (or Markov Chain)



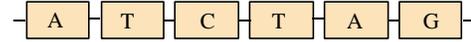
Probability for each character based only on several preceding characters in the sequence

of preceding characters = **order** of the Markov Model

Probability of a sequence

$$P(s) = P[A] P[A,T] P[A,T,C] P[T,C,T] P[C,T,A] P[T,A,G]$$

Hidden Markov Models



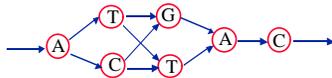
Observed frequencies	A 0.7	A 0.1	C 0.8	A 0.4	A 0.8	C 0.3
	T 0.3	T 0.9	G 0.2	T 0.6	T 0.2	G 0.7

Probabilistic model - true state is unknown

Hidden Markov Models

States -- well defined conditions

Edges -- transitions between the states



ATGAC
ATTAC
ACGAC
ACTAC

Each transition assigned a probability.

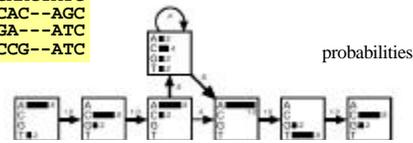
Probability of the sequence:

single path with the highest probability --- *Viterbi* path

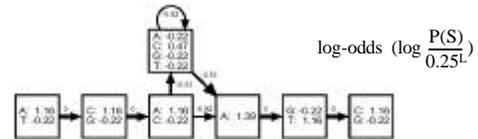
sum of the probabilities over all paths -- *Baum-Welch* method

Hidden Markov Models

ACA---ATG
TCAACTATC
ACAC---AGC
AGA---ATC
ACCG---ATC



probabilities



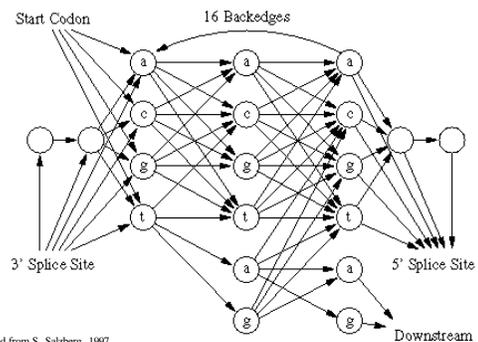
log-odds $(\log \frac{P(S)}{0.25^L})$

Adopted from Anders Krogh, 1998

Hidden Markov Model of Biased Coin Tosses

- States (S_i): Two Biased Coins {C1, C2}
- Outputs (O_j): Two Possible Outputs {H, T}
- $p(\text{Outputs } O_{ij})$: $p(C1, H), p(C1, T), p(C2, H), p(C2, T)$
- Transitions: From State X to Y {A11, A22, A12, A21}
- $p(\text{Initial } S_1)$: $p(I, C1), p(I, C2)$
- $p(\text{End } S_i)$: $p(C1, E), p(C2, E)$

Hidden Markov Model for Exon and Stop Codon (VEIL Algorithm)



Adopted from S. Salzberg, 1997