

Um modelo desenvolvido buscando responder as seguintes questões: Análise de sequências biológicas, identificação de genes, 'rotular' éxon/intron/sequência intergênica e em alinhamentos de sequências.

O modelo oculto de Markov (HMM) é uma fundamentação formal para modelos probabilísticos de sequências lineares rotulando problemas.

HMM é o “coração” de muitos programas, incluindo busca de genes por perfis (profiles) e identificação de sítios regulatórios.

Figura 1) Mostra como estimar uma sequência, construindo no exemplo um HMM com sítio de reconhecimento de splice. Para sugestões inteligentes, considera-se as diferentes propriedades estatísticas, vendo as diferentes porcentagens de composição de bases. Partindo dessas probabilidades é possível construir uma HMM.

Cada estado tem uma probabilidade de emissão e uma probabilidade de transição (probabilidade do estado atual para um novo estado) que vai descrever um modelo linear.

Porque é dito oculto? (hidden)

- Partindo de um estado (state path), o próximo só depende da posição imediatamente anterior, do que ele está. Uma vez que só vemos a sequência observada, esse estado imediatamente anterior (state path) é oculto.

O que significa dizer: HMM é um modelo probabilístico completo?

- Os parâmetros e os scores são todos probabilidades.

*Ainda assim, a teoria de probabilidades bayesiana pode manipular esse números em poderosos padrões, incluindo otimização de parâmetros e interpretação da significância de scores.

Encontrando o melhor *State path*

- Algoritmo de Viterbi é um algoritmo de programação dinâmica usado para alinhamentos de sequências padrões.
- Algoritmo de Viterbi garante encontrar a melhor probabilidade do estado dado uma sequência e um HMM.

Além do melhor score de alinhamento

- HMM confirma a confiabilidade da escolha,
- Usa uma técnica de somar todas possibilidades de caminhos antes de escolher a melhor.
- Não possui a propriedade 'remembering' para estados distantes, o que implica ser inapropriado para análise de estrutura secundária de RNA.

Ainda que existam outras ferramentas disponíveis baseadas em outros modelos, HMM é ainda a ferramenta mais poderosa segundo o autor.