

Desenvolvimento de um configurador gráfico para o ToPS

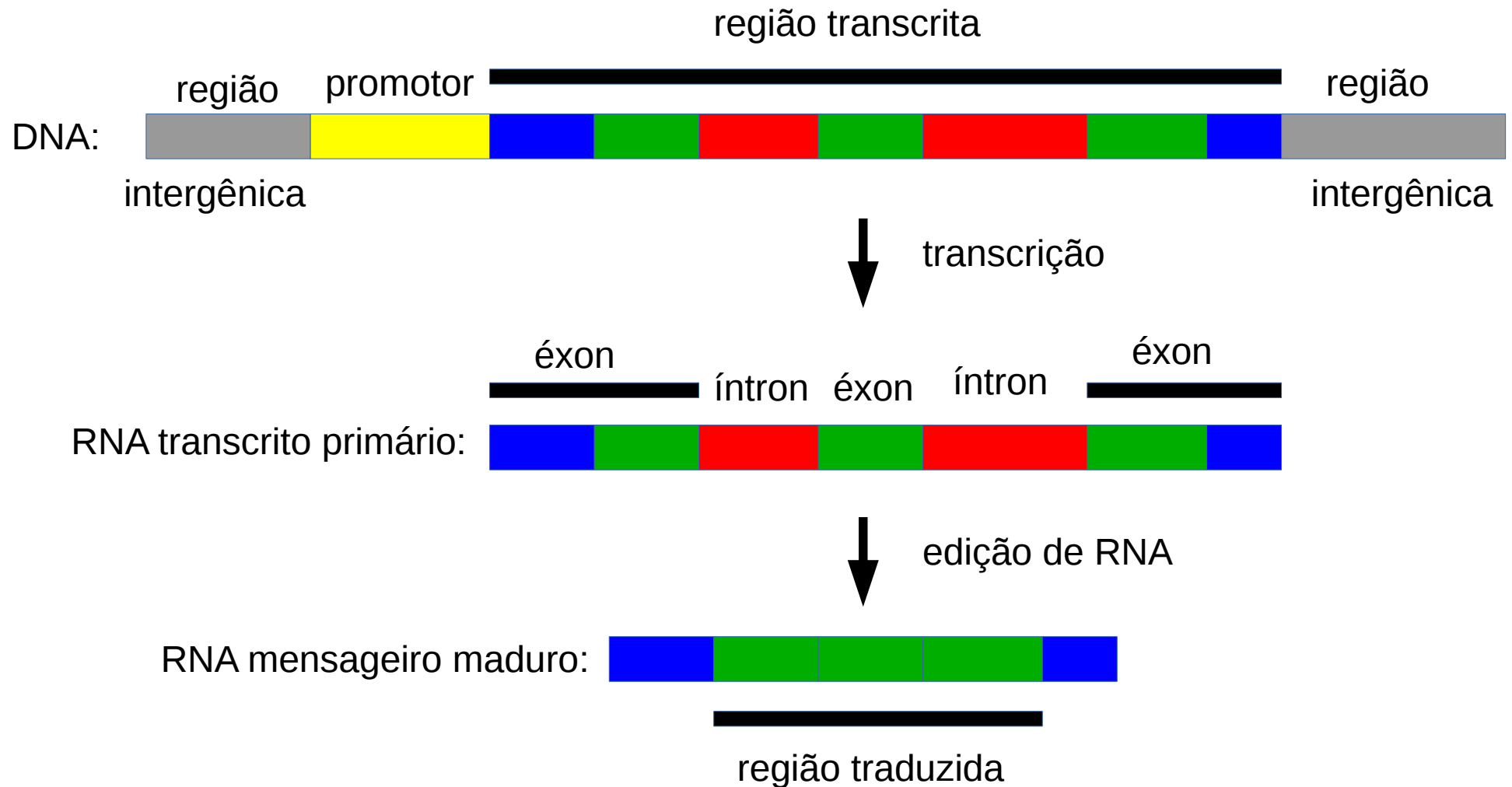
Trabalho de formatura supervisionado – MAC0499 – 2015

Aluno: Douglas Vasconcelos Cancherini

Supervisor: Alan Mitchell Durham

Introdução

Estrutura dos genes eucarióticos



Introdução

Predição de genes

- Conceito e importância
- Ab initio = a única entrada é a sequência genômica
- Treinamento prévio
- Predição por múltiplas informações

Introdução

Modelos ocultos de Markov e extensões

- Cadeias de Markov cujos estados não podem ser diretamente observado, mas devem ser inferidos por símbolos emitidos
- Reunidos a uma série de extensões do conceito, mostraram-se bastante adequados para modelar elementos gênicos

Introdução

Modelos ocultos de Markov generalizados

- Cada estado passa a ser um submodelo probabilístico
- Permitem usar modelagens particulares para cada elemento gênico
- Permitem o uso de modelos de duração
- Boa parte dos melhores preditores gênicos da atualidade são baseados em GHMMs

Introdução

Dois problemas fundamentais

- Problema da decodificação: dados um modelo e uma sequência de símbolos, determinar qual o conjunto de estados sucessivos com máxima probabilidade de gerá-la
- Problema do treinamento: dados um modelo e uma sequência de símbolos, determinar quais parâmetros do modelo maximizam a probabilidade de emissão da sequência

Introdução

Ferramentas criadas pelo grupo

ToPS = *Toolkit for Probabilistic Models Of Sequence*

- Implementação eficiente de algoritmos de treinamento e decodificação para vários modelos probabilísticos
- Grande flexibilidade para combinar vários modelos num modelo maior baseado em GHMM

MYOP = *Make Your Own Predictor*

- Preditor gênico *ab initio* que faz uso do ToPS

Motivação

Arquivos de configuração do ToPS

- numerosos
- relacionam-se entre si
- dificuldades para um usuário explorar a flexibilidade potencialmente oferecida pelo programa

Objetivos

Usuários do ToPS que façam uso de um modelo baseado em GHMM

- arquitetura do GHMM
 - compreensão
 - visualização
 - modificação
- edição de parâmetros
 - submodelos
 - treinamento
 - modelos de duração

Implementação

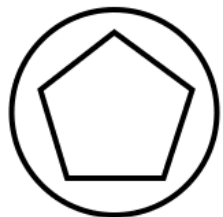
Características do sistema

- Cliente em HTML5 e JavaScript, com elementos SVG, compatível com navegadores recentes
- Restrições de segurança dos navegadores tornam necessário que haja um servidor, para que seja possível invocar o ToPS por linha de comando

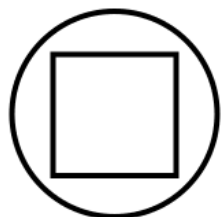
Implementação

Funcionalidades da interface I

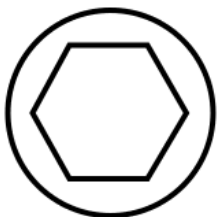
- Representação do GHMM como um grafo
 - vértices = submodelos
 - arestas = transições entre submodelos
- Um polígono diferente para cada tipo de modelo probabilístico
- Modelos podem ser arrastados na tela
- Modelos de duração como elipses/círculos ao redor dos submodelos



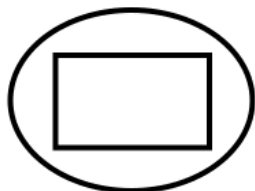
independent and identically
distributed model



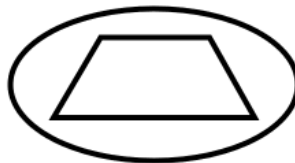
hidden Markov model



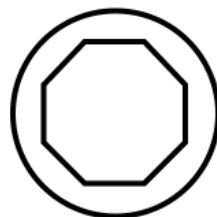
inhomogeneous
Markov chain



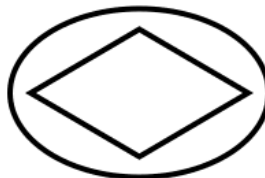
similarity-based
sequence weighting



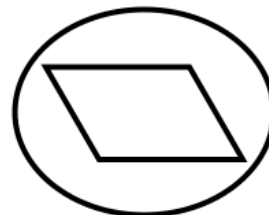
variable-length
Markov chain



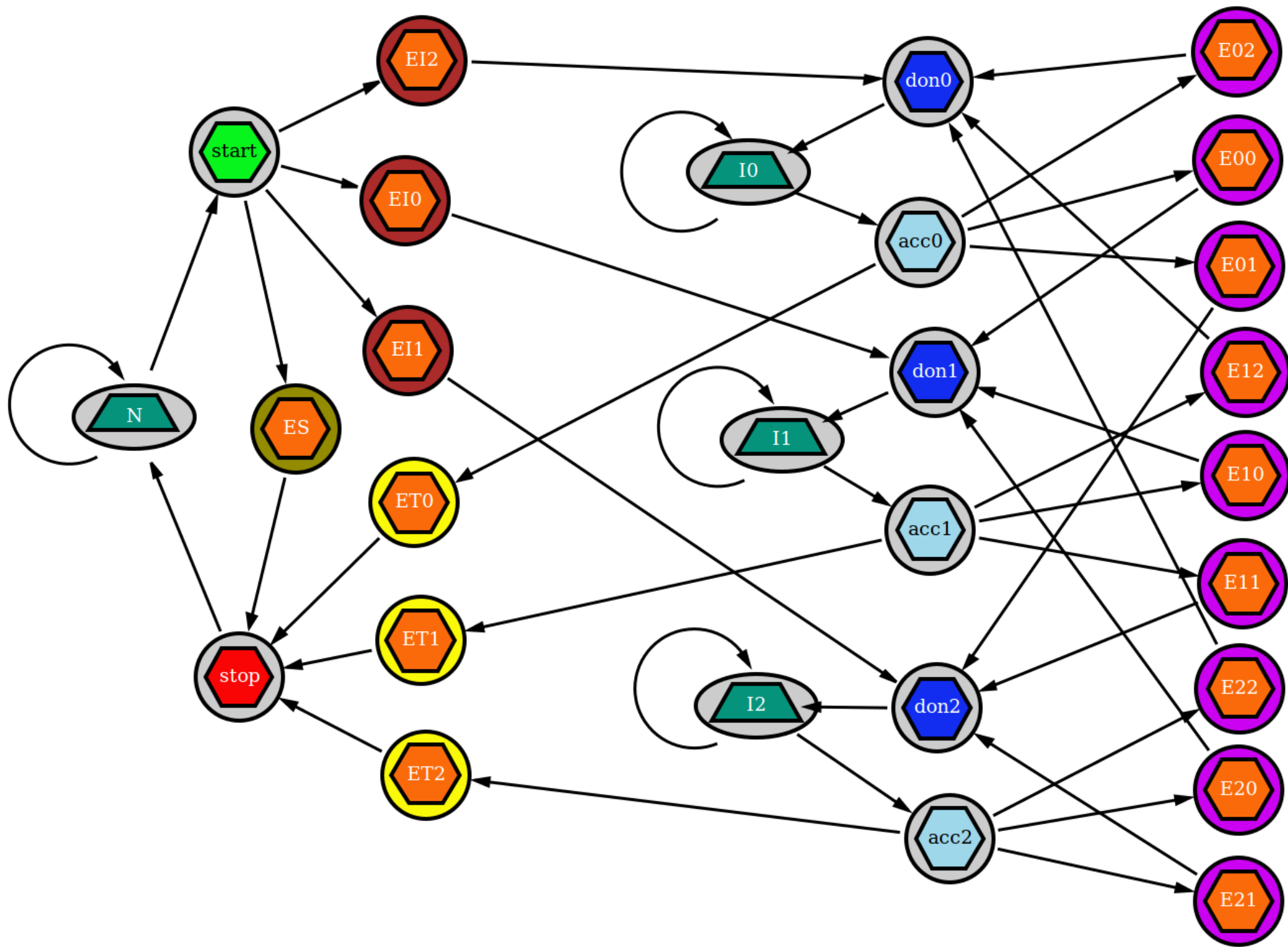
paired hidden Markov
chain



profile hidden Markov
chain



reserved for
future model



Funcionalidades da interface II

- Duplo clique sobre aresta, polígono ou elipse/círculo abre uma janela para a edição dos parâmetros correspondentes
- Compartilhamento de parâmetros entre modelos
 - facilidade para compartilhar
 - facilidade para reconhecer, por meio de cores

Model edition

Model type: pairHMM

Model name

Model specification

Enter model specification below

or just select an specification from models of the same type:

Model color

Choose a color for the model representation:

or just click for an automatically generated color:

Implementação

Detalhes da solução para o cliente

- Volume de dados
 - grande para um humano
 - pequeno para um computador
- Opção por armazenar os dados no próprio DOM
- Persistência por exportação de XML
- Funções acionadas por eventos na página agem diretamente sobre o DOM

Implementação

Perspectivas futuras

- Servidor que disponibilize, para as configurações escolhidas no cliente
 - treinamento
 - decodificação
 - simulação
- Possibilidade de escolher os parâmetros de treinamento no cliente

Considerações finais

- Temos um sistema que já apresenta uma interface bastante amigável para a edição das configurações do ToPS
- Aliada a um servidor com as características propostas, ele deverá em breve ser capaz de facilitar de fato o uso do ToPS, em toda a sua flexibilidade de configurações