Viewing Predicted TFBS

Andrew A Carbonetto
project for UBC CPSC 533C

Contents

- Goal of Project

- Completed Features of Project

- Demo

- ...Unfinished Business

- What are Biologists looking for?

  - Conserved patterns of motifs

- How can they find it?

  - Comparison of multiple sequences

- Goal is to provide an interactive presentation of motifs over several sequences of DNA.

- A user should be able to easily see patterns of motifs that are common to *both* sequences.

  - Hopefully that means both patterns are important and the results can be published.

- Used Prefuse (Java toolkit for Infovis Graphics)

- Main object is the VisualItem (represents all the objects in a graph: Edges, Nodes, ...)

- VisualItems are organized via the ItemRegistry

- Have Prefuse read in a graph representing a set of predicted TFBS's (Motifs), and draw a comparison with other already read graphs.

  - Comparison: represented as Parallel Coordinates

- VisualItems:

  - NodeItems: represent the individual predicted Motifs, clustered by file read

  - ...they should also preserve spatial location (location in cluster should represent location on sequence of DNA)

- VisualItems:

    - EdgeItems: Represent common motifs found in each file

    - Because this is a parallel coordinates system, only those edges between adjacent clusters should be shown.

- Filters:

    - Color Filters: Motifs are initially colored by family (motif attribute)

    - Location Filters: Emphasis on nearby motifs

    - Mouse Pointed Emphasis: Even more emphasis on motifs around the mouse pointer

- Interactivity

  - Smooth animation and manipulation of sequences

  - Easy ordering of sequences

  - Ability to add and remove sequences

  - Ability to copy sequences

- Fast feedback of information

  - Show in the Demo

- Filtering:

  - Show in the Demo

- Demo

  - Demo1: Skinny dataset, with many sequences

  - Demo2: Large dataset example

- More filters!! More feedback!!

- Get around MAX_ITEM bound of prefuse

- Node occlusion

- (User Study) I'd like to get the opinion of some biologists that actually search these datasets.

Thanks

Questions? Comments?