

Projektarbeit:

Bewertung von de-novo Motifsuche-Algorithmen

Viele Abläufe auf genomischer Ebene werden durch kurze Signalsequenzen gesteuert. Dies können beispielsweise Bindestellen von Transkriptionsfaktoren sein. Transkriptionsfaktoren sind Proteine, die an die DNA binden und mit dieser Bindung regulatorische Prozesse in Gang setzen. So kann beispielsweise die Bindung des Transkriptions-Initiationskomplexes erleichtert oder aus sterischen Gründen verhindert werden. Häufig weiß man allerdings nur, dass eine bestimmte Menge von Genen koexprimiert und somit potentiell koreguliert ist, aber nicht, welcher Transkriptionsfaktor für dieses Verhalten verantwortlich ist. Ziel der do-novo Motifsuche ist es nun, aus einer Menge von Sequenzen, in denen ein gemeinsames Bindemotif vermutet wird, eben jenes zu bestimmen. Bekannte Verfahren sind der EM-Algorithmus [1], Gibbs-Sampling [5] oder k-mer basierte Algorithmen wie Weeder [3]. In letzter Zeit wurden mit FIRE [2] und DEME [4] außerdem diskriminative Ansätze publiziert, die in erster Linie auf differentielle Analysen zugeschnitten sind.

Ziel der Arbeit ist es, eine Menge von Benchmark-Datensätzen unterschiedlicher Schwierigkeit zu entwerfen. Darunter sollen auch Datensätze sein, die speziell für die Bewertung diskriminativer Algorithmen geeignet sind. Die erstellten Datensätze sollen dann genutzt werden, um einige der häufig genutzten Algorithmen zu bewerten und zu vergleichen.

Ansprechpartner:

Jan Grau, Raum 4.09, 55-24768, grau@informatik.uni-halle.de

Literatur

- [1] Timothy L. Bailey and Charles Elkan. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. In *Proceedings of the Second International conference on intelligent systems for molecular biology*, 1994.
- [2] Olivier Elemento, Noam Slonim, and Saeed Tavazoie. A universal framework for regulatory element discovery across all genomes and data types;. *Molecular Cell*, 28(2):337–350, 2007.

- [3] Giulio Pavese, Paolo Mereghetti, Giancarlo Mauri, and Graziano Pesole. Weeder Web: discovery of transcription factor binding sites in a set of sequences from co-regulated genes. *Nucl. Acids Res.*, 32(suppl_2):W199–203, 2004.
- [4] Emma Redhead and Timothy Bailey. Discriminative motif discovery in dna and protein sequences using the deme algorithm. *BMC Bioinformatics*, 8(1):385, 2007.
- [5] William Thompson, Eric C. Rouchka, and Charles E. Lawrence. Gibbs Recursive Sampler: finding transcription factor binding sites. *Nucl. Acids Res.*, 31(13):3580–3585, 2003.