Mining Protein Family Specific Residue Packing Patterns From Protein Structure Graphs

von

Jun Huan, Wei Wang, Deepak Bandyopadhyay,
Jack Snoeyink, Jan Prins,
Alexander Tropsha

gehalten von Martin Jess

Problembeschreibung

Aus „Proteingraphen“ wiederkehrende 3-D Strukturen finden, die eine Proteinstrukturfamilie charakterisieren.

Biologische Sicht

Proteine

- Ketten von direkt/indirekt verknüpften Aminosäuren (AS)

wiederkehrende Strukturen

- aktives Zentrum, funktionsbestimmende Bestandteile

Protein als Graph

Verschiedene Möglichkeiten

- Graphen, Bäume, direkte, indirekte
- Atome als Knoten, Atombindungen als Kanten
- Struktureinheiten als Knoten, Tertiärstruktur-interaktionen als Kanten

Allgemeine Erstellung

verwendetes Modell

- Indirekter Graph mit AS als Knoten
- Koordinaten des C_α -Atomes
- Bezeichnung nach AS und Position in Primärstuktur
- „Bindungskanten“

Erweiterung durch proximity edges

- Nach drei unterschiedlichen Modellen

1. Abstandsmethode

Contact Distance Graph (CD)

Kanten mit allen C_α -Atomen im Abstand δ

2. Delaunay tessellation

Delaunay tessellation Graph (DT)

Kanten für alle 4 C_α -Atome die ein Tetraeder mit einer leeren Kugel darin ergeben

3. Almost Delaunay tessellation

Almost Delaunay Graph (AD)

DT erweitert mit weiteren Delaunay Kanten bei denen die Positionen der C_α -Atome um den Parameter c verändert sind

Frequent Subgraph Mining Algorithm

FFSM

1: $S \leftarrow$ { the CAMs of the frequent nodes }
2: $P \leftarrow$ { the CAMs of the frequent edges }
3: FFSM-Explore(P, S);

FFSM-Explore (P, S)

1: **for** $X \in P$ **do**
2: **if** $(X.isCAM)$ **then**
3: $S \leftarrow S \cup \{X\}$
4: $C \leftarrow$ { all matrices $M \mid X$ is submatrix of M }
5: remove CAM(s) from C that is either infrequent or not optimal
6: FFSM-Explore(C, S)
7: **end if**
8: **end for**

Classification Modeling

Support Vector Machine (SVM)

1. hoch-dimensionale Datenmengen
2. verschiedene Sets für kernel learing functions

Benutzung von `libsvm` mit radius kernel

http://www.csie.ntu.edu.tw/~cjlin/libsvm/