Public Abstract

First Name:Meshari

Middle Name:Saud

Last Name:Alazmi

Adviser's First Name:Dong

Adviser's Last Name:Xu

Co-Adviser's First Name:Yi

Co-Adviser's Last Name:Shang

Graduation Term:SP 2012

Department:Computer Science

Degree:MS

Title:PROTEIN STRUCTURAL MODELS SELECTION USING 4-mer SEQUENCE AND COMBINED SINGLE AND CONSENSUS SCORES

Quality assessment for protein structure models is an important issue in protein structure prediction. Consensus methods assess each model based on its structural similarity to all the other models in a model set, while single scoring methods, such as Opus-ca and RW, evaluate each model based on its structural properties. In this work, a novel method proposed and developed to effectively combine consensus methods and single scoring methods for better quality assessment. At first, a new method called Single Position Specific Probability (SPSP) Score is proposed based on consensus method using 4-mer sequence. Specifically, every letter in the 4-mer sequence represents a state for a local region consisting of four amino acids. A machine learning method (Neural Network) helped to combine several single scoring methods, RW, DDFire, and OPusCa with consensus methods, SPSP and Consensus Global Distance Test-Total Score (CGDT-TS) to achieve a good combination of all the terms. The method was tested on two benchmark datasets and achieved improvements over the state-of-the-art methods. The first benchmark was on Yang Zhang's data containing 56 targets. The second benchmark was from Rosetta data containing 35 targets. For Zhang's data, the CGDT score is 0.6058, while combined method achieved 0.6105. For Rosetta data, the CGDT score achieved 0.4255, while combined method achieved 0.4529.