

**Public Abstract**

First Name:Junlin

Middle Name:

Last Name:Wang

Adviser's First Name:Yi

Adviser's Last Name:Shang

Co-Adviser's First Name:

Co-Adviser's Last Name:

Graduation Term:SS 2015

Department:Computer Science

Degree:MS

**Title:MACHINE LEARNING METHODS FOR EVALUATING THE QUALITY OF A SINGLE PROTEIN MODEL USING ENERGY AND STRUCTURAL PROPERTIES****ABSTRACT**

Computational protein structure prediction is one of the most important problems in bioinformatics. In the process of protein three-dimensional structure prediction, assessing the quality of generated models accurately is crucial. Although many model quality assessment (QA) methods have been developed in the past years, the accuracy of the state-of-the-art single-model QA methods is still not high enough for practical applications. Although consensus QA methods performed significantly better than single-model QA methods in the CASP (Critical Assessment of protein Structure Prediction) competitions, they require a pool of models with diverse quality to perform well.

In this thesis, new machine learning based methods are developed for single-model QA and top-model selection from a pool of candidates. These methods are based on a comprehensive set of model structure features, such as matching of secondary structure and solvent accessibility, as well as existing potential or energy function scores. For each model, using these features as inputs, machine learning methods are able to predict a quality score in the range of. Five state-of-the-art machine learning algorithms are implemented, trained, and tested using CASP datasets on various QA and selection tasks. Among the five algorithms, boosting and random forest achieved the best results overall. They outperform existing single-model QA methods, including DFIRE, RW and Proq2, significantly, by up to 10% in QA scores.