World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:11, No:12, 2017

Structure Clustering for Milestoning Applications of Complex Conformational Transitions

Authors : Amani Tahat, Serdal Kirmizialtin

Abstract : Trajectory fragment methods such as Markov State Models (MSM), Milestoning (MS) and Transition Path sampling are the prime choice of extending the timescale of all atom Molecular Dynamics simulations. In these approaches, a set of structures that covers the accessible phase space has to be chosen a priori using cluster analysis. Structural clustering serves to partition the conformational state into natural subgroups based on their similarity, an essential statistical methodology that is used for analyzing numerous sets of empirical data produced by Molecular Dynamics (MD) simulations. Local transition kernel among these clusters later used to connect the metastable states using a Markovian kinetic model in MSM and a non-Markovian model in MS. The choice of clustering approach in constructing such kernel is crucial since the high dimensionality of the biomolecular structures might easily confuse the identification of clusters when using the traditional hierarchical clustering methodology. Of particular interest, in the case of MS where the milestones are very close to each other, accurate determination of the milestone identity of the trajectory becomes a challenging issue. Throughout this work we present two cluster analysis methods applied to the cis–trans isomerism of dinucleotide AA. The choice of nucleic acids to commonly used proteins to study the cluster analysis is two fold: i) the energy landscape is rugged; hence transitions are more complex, enabling a more realistic model to study conformational transitions, ii) Nucleic acids conformational space is high dimensional. A diverse set of internal coordinates is necessary to describe the metastable states in nucleic acids, posing a challenge in studying the conformational transitions. Herein, we need improved clustering methods that accurately identify the AA structure in its metastable states in a robust way for a wide range of confused data conditions. The single linkage approach of the hierarchical clustering available in GROMACS MD-package is the first clustering methodology applied to our data. Self Organizing Map (SOM) neural network, that also known as a Kohonen network, is the second data clustering methodology. The performance comparison of the neural network as well as hierarchical clustering method is studied by means of computing the mean first passage times for the cis-trans conformational rates. Our hope is that this study provides insight into the complexities and need in determining the appropriate clustering algorithm for kinetic analysis. Our results can improve the effectiveness of decisions based on clustering confused empirical data in studying conformational transitions in biomolecules.

Keywords : milestoning, self organizing map, single linkage, structure clustering
Conference Title : ICADMA 2017 : International Conference on Advanced Data Mining and Applications
Conference Location : Barcelona, Spain
Conference Dates : December 14-15, 2017