

OneHotEncoding and LSTM based Deep Learning Models for Protein Secondary Structure Prediction

Vamsidhar Enireddy¹ and C Karthikeyan ¹

¹Koneru Lakshmaiah Education Foundation

August 14, 2020

Abstract

Protein Secondary Structure (PSS) prediction is crucial for examining and studying the protein structure and its function. PSS helps to predict the tertiary structure and offers to understand about its structures, which in turn helps to design various drugs. The existing PSS prediction techniques are capable of achieving Q3 accuracy of nearly 80%, and there has not any improvement till now. In this paper, we propose a novel technique that uses amino acid sequences alone as an input feature, and the respected feature vector matrix is given through the deep learning model (DLM) for PSS prediction. Apart from all deep learning methods, we use OneHotEncoding and LSTM (Long short term memory) technique to forecast PSS that helps to give more accuracy. The one hot encoder is used to extract the local contexts of amino-acid sequences, and BLSTM (Bi-directional LSTM) captures the long-distance interdependencies among amino-acids. LSTM is one of the new deep learning models successfully applied in the field of bioinformatics to solve problems. LSTM is very efficient in mapping the long term dependencies of sequence information, which is more capable than the convolutional neural networks (CNN's). The performance of the proposed system is estimated on the openly available datasets such as CullPDB, CASP10, and CASP11. Results show that the performance of the proposed technique achieved superior outcomes than the existing approaches on the three similar datasets.

Hosted file

Manuscript (5).docx available at <https://authorea.com/users/350904/articles/475611-onehotencoding-and-lstm-based-deep-learning-models-for-protein-secondary-structure-prediction>