Abstract

Advances in DNA sequencing technology have caused generation of the vast amount of new sequence data. It is essential to understand the functions, features, and structures of every newly sequenced data. Analyzing sequence data by different methods could provide important information about the sequence data. One of the essential tasks for genome annotation is gene prediction that can help to understand the features and determine functions of the genes. One of the key steps towards correct gene structure prediction is accurate splice site detection. There are vast numbers of splice site prediction methods, however, a few of them can be incorporated in gene prediction modules because of their complexity. In this paper, a novel model is presented to recognize unknown splice sites in a new genome without using any prior knowledge. Our model is defined based on integrating Jensen-Shannon divergence and a polynomial equation of order 2. Finally, the proposed model is evaluated on Yeast’s genome to predict splice sites. The experimental results suggest that the proposed method is an effective approach for splice site prediction.
References


17. Saccharomyces Genome Database, Available online: http://www.yeastgenome.org/


Index Terms

Computer Science

Information Sciences

Keywords

Splice site, Position weight matrix, Entropy.