

{tag} International Journal of Computer Applications
Foundation of Computer Science (FCS), NY, USA

[Volume 151](#)

-
[Number 5](#)

Year of Publication: 2016

Authors:

Yousef Seyfari, Farzad Didehvar, Hadi Banaee, Fatemeh Zare-Mirakabad

10.5120/ijca2016911800

{bibtex}2016911800.bib{/bibtex}

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

1. Burset M., Seledtsov I. A., and Solovyev V. V. 2000. Analysis of canonical and non-canonical splice sites in mammalian genomes. *Nucleic acids research* 28, no. 21: 4364-4375.
2. Staden R. 1984. Computer methods to locate signals in nucleic acid sequences. *Nucleic acids research* 12, no. 1: 505-519.
3. Li J. L., Wang L. F., Wang H. Y., Bai L. Y., and Yuan Z. M. 2012. High-accuracy splice site prediction based on sequence component and position features. *Genetics and Molecular Research* 11, no. 3: 3432-3451.
4. Tavares L. G., Lopes H. S., and Lima C. R. E. 2009. Evaluation of weight matrix models in the splice junction recognition problem. *IEEE International Conference on Bioinformatics and Biomedicine Workshop*, 2009: 14-19.
5. Zhang M. Q., and Marr T. G. 1993. A weight array method for splicing signal analysis. *Computer applications in the biosciences: CABIOS* 9, no. 5: 499-509.
6. Burge C., and Karlin S. 1997. Prediction of complete gene structures in human genomic DNA. *Journal of molecular biology* 268, no. 1: 78-94.
7. Nassa T., Singh S., and Goel N. 2013. "Splice Site Detection in DNA Sequences using Probabilistic Neural Network." *International Journal of Computer Applications* 76, no. 4: 1-4.
8. Baten A. K. M. A., Halgamuge S. K., and Chang B. CH. 2008. Fast splice site detection using information content and feature reduction. *BMC bioinformatics* 9, no. Suppl 12: S8.
9. Wei D., Zhang H., Wei Y., and Jiang Q. 2013. A novel splice site prediction method using support vector machine. *Journal of Computational Information Systems* 9, no. 20: 8053-8060.
10. Wei D., Zhuang W., Jiang Q., and Wei Y. 2012. A new classification method for human gene splice site prediction. In *Health Information Science*. Springer Berlin Heidelberg, HIS 2012: 121-130.
11. Goel N., Singh S., and Aseri T. C. 2015. An Improved Method for Splice Site Prediction in DNA Sequences Using Support Vector Machines. *Procedia Computer Science*, 57: 358-367.
12. Bari A. G., Reaz M. R. and Jeong B. S. 2014. Effective DNA Encoding for Splice Site Prediction Using SVM. *Math-Communications in Mathematical and in Computer Chemistry*, 71(1): 241-258.
13. Salekdeh A. Y. and Wiese K. C. 2011. Improving splice-junctions classification employing a novel encoding schema and decision-tree. In *Congress on Evolutionary Computation (CEC)*: 1302-1307.
14. Wei D., Zhuang W., Jiang Q., and Wei Y. 2012. A new classification method for human gene splice site prediction. In *Health Information Science*. Springer Berlin Heidelberg: 121-130.
15. Huang J., Li T., Chen K., and Wu J. 2006. An approach of encoding for prediction of splice sites using SVM. *Biochimie*, 88(7): 923-929.
16. Meher P. K., Sahu T. K., Rao A. R., and Wahi S. D. 2014. A statistical approach for 5' splice site prediction using short sequence motifs and without encoding sequence data. *BMC bioinformatics*, 15(1): 362-376.
17. *Saccharomyces Genome Database*, Available online: <http://www.yeastgenome.org/>
18. Román-Roldán R., Bernaola-Galván P., and Oliver J. L. 1998. Sequence compositional complexity of DNA through an entropic segmentation method. *Physical Review Letters* 80, no. 6: 1344-1347.
19. Bernaola-Galván P., Grosse I., Carpena P., Oliver J. L., Román-Roldán R., and Stanley

H. E. 2000. Finding borders between coding and noncoding DNA regions by an entropic segmentation method. Physical Review Letters 85, no. 6: 1342-1345.

Index Terms

Computer Science

Information Sciences

Keywords

Splice site, Position weight matrix, Entropy.