

Identifying DNA-binding proteins based on multi-features fusion and LASSO feature selection

Shengli Zhang¹, Fu Zhu¹, Qianhao Yu¹, and xiaoyue Zhu¹

¹Xidian University

July 6, 2020

Abstract

DNA-binding proteins, performing an indispensable function in the maintenance of genetic information and holding significances for biomedical research, are inefficiently identified by traditional experimental methods due to their huge quantities. On the contrary, the machine learning method as an emerging technique demonstrates satisfactory speed and decent accuracy. Thus, this work focuses on extracting four different features from primary and secondary sequence features, i.e., RS, PseAACs, PSSM-ACCT and PSSM-DWT. With the LASSO dimension reduction method, we experiment on the combination of feature submodels to obtain the optimized number of top rank features. These features are input into the training Ensemble subspace discriminant descriptor to predict the DNA-binding proteins. Three different datasets are adopted to evaluate the performances of the as-proposed approach in this work. The PDB1075 and PDB594 datasets are adopted for the 5-fold Cross-Validation, and the PDB186 is used for the independent experiment. In the 5-fold Cross-Validation, the PDB1075 and PDB594 both show extremely high precision reaching 86.98% and 88.2%, respectively, while the accuracy of independent experiment is 75.8%, which suggests that the methodology proposed in this work is capable of predicting DNA-binding proteins effectively.

Hosted file

Identifying DNA_binding proteins.pdf available at <https://authorea.com/users/340288/articles/467362-identifying-dna-binding-proteins-based-on-multi-features-fusion-and-lasso-feature-selection>

Hosted file

Figure1.eps available at <https://authorea.com/users/340288/articles/467362-identifying-dna-binding-proteins-based-on-multi-features-fusion-and-lasso-feature-selection>

Hosted file

Figure2.eps available at <https://authorea.com/users/340288/articles/467362-identifying-dna-binding-proteins-based-on-multi-features-fusion-and-lasso-feature-selection>

figures/Figure3/Figure3-eps-converted-to.pdf

figures/Figure4/Figure4-eps-converted-to.pdf