

SMI-BLAST

In order to convert the ISH errors into True-PSSM in the iteration process of PSI-BLAST, in this study, we propose a framework (SMI-BLAST) to correct those ISH errors by embedding sequence-similarity-matrix-based iterative framework (SMI-based framework) in PSI-BLAST. SMI-BLAST cannot only correct the PSSMs with ISH errors, but also can improve the ranking quality for protein remote homology detection.

1. Usage of key codes

<p>run.sh</p>	<p>Describe: the script file for running SMI-BLAST</p> <p>Parameters:</p> <p>usr_dir The path of test sequence; test sequence should be saved on test.fasta</p> <p>cutoff The number of iteration. 2 represents iteration number = 2; 3 represents iteration number = 5; 4 represents iteration number = 10;</p> <p>CON_FOLD The path of source code</p> <p>Dependency: [RankLib-2.7] [Trained learning to rank model]</p>
<p>init.sh</p>	<p>Describe: Generating the features of query sequences</p> <p>Parameters:</p> <p>usr_dir The path of test sequence; test sequence should be saved on test.fasta</p> <p>Dependency: [Pse-in-One2.0]</p>
<p>formate_ vec_mer ge.py</p>	<p>Describe: Generating the similarity scores of query sequence and feedback sequences</p> <p>Parameters:</p> <p>-psi The alignment feature of feedback list</p> <p>-query_TOP1 Top-1-gram feature of query sequence</p> <p>-query_TOP2 Top-2-gram feature of query sequence</p> <p>-query_accpssm ACC-PSSM feature of query sequence</p> <p>-query_pdtppsm PDT-Profile feature of query sequence</p> <p>-query_dt DT feature of query sequence</p> <p>-query_kmer2 Kmer feature of query sequence</p> <p>-query_dr DR feature of query sequence</p> <p>-query_distancepair DP feature of query sequence</p> <p>-query_ac AC feature of query sequence</p> <p>-query_cc CC feature of query sequence</p> <p>-query_acc ACC feature of query sequence</p> <p>-query_pdt PDT feature of query sequence</p> <p>-query_pcpseaac PC-PseAAC feature of query sequence</p> <p>-query_scpsseaac SC-PseAAC feature of query sequence</p> <p>-aggre Result path</p>

	Dependency: <u>84 feature datasets of SCOPe2.06</u>
ltr-iter.py	Describe: Using the new selected sequence to run the iteration process of PSI-BLAST
	Parameters: usr_dir The path of test sequence selected Path The path of selected sequence for constructing PSSM DB_path The path of SCOPe dataset sealed by PSI-BLAST ori_dir The path of original search result of PSI-BLAST iteration_number the number of iteration
	Dependency: [PSI-BLAST 2.7.1]