World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:8, No:10, 2014

2D Fingerprint Performance for PubChem Chemical Database

Authors : Fatimah Zawani Abdullah, Shereena Mohd Arif, Nurul Malim

Abstract : The study of molecular similarity search in chemical database is increasingly widespread, especially in the area of drug discovery. Similarity search is an application in the field of Chemoinformatics to measure the similarity between the molecular structure which is known as the query and the structure of chemical compounds in the database. Similarity search is also one of the approaches in virtual screening which involves computational techniques and scoring the probabilities of activity. The main objective of this work is to determine the best fingerprint when compared to the other five fingerprints selected in this study using PubChem chemical dataset. This paper will discuss the similarity searching process conducted using 6 types of descriptors, which are ECFP4, ECFC4, FCFP4, FCFC4, SRECFC4 and SRFCFC4 on 15 activity classes of PubChem dataset using Tanimoto coefficient to calculate the similarity between the query structures and each of the database structure. The results suggest that ECFP4 performs the best to be used with Tanimoto coefficient in the PubChem dataset.

Keywords : 2D fingerprints, Tanimoto, PubChem, similarity searching, chemoinformatics
Conference Title : ICCST 2014 : International Conference on Computer Science and Technology
Conference Location : Bali, Indonesia
Conference Dates : October 09-10, 2014