A Regression Model using K-Means Algorithm to Screen 32 Compound Dataset as COX-2 Inhibitors

Abstract

COX-2 provided a new class of anti inflammatory, analgesic and antipyretic drugs with significantly reduced side effects. It has been reported that inhibiting COX-2 could also be an important strategy for preventing or treating a number of cancers. A report with modified k-means clustering algorithm to cluster groups of compounds obtained from regression analysis along with few compounds which were non-tested against COX-2 and screened them using regression model. The regression model due to its high predictive ability can be utilized as an alternative aid to the costly and time consuming experiments for recognizing and determining compounds with high COX-2 binding affinity. Hence, a group of new derivatives from literature are subjected to screening utilizing the produced model. A set of 32 compounds with pyrazole ring as main nucleus was selected from a published review paper. In this work, a modification of k-means algorithm that efficiently searches data to cluster points by computing sum of squares within each cluster which makes the program to select the most promising subset of classes for clustering. From a set of 32 compounds, only the top 5 compounds are combined with 58 molecule data set to perform cluster analysis. From the analysis it is evidenced that k-means
clustering algorithm is able to group data objects of all molecules based on the 3 centroids provided and all top 5 compounds appear to be centred on one spade whereas Celecoxib appeared in another cluster.

**References**


Index Terms

Computer Science

Algorithms

Keywords
COX-2, Cluster analysis, k-means, phylogenetic tree