PalStudent Journal
A Palestinian Scientific Journal for the Youth

ABSTRACTS: VOLUME 6, SPECIAL ISSUE

ABSTRACT

Design a Novel Method to Cluster Data Based on an Angular Randomization Test

Baneen Hussein Abd Zaid, Abeer Ali Naji, Kawthar Hataf Fazaa, Ahmed Jebur Ali.

Iraq, University of Kufa, College of Co-educational Education, Department of Mathematics.

Background: Cluster analysis of amino acids poses a difficult task in the field of predicting protein function and structure. This research proposes a novel method that focuses on using the p-value obtained from an angular randomization test that evaluates the similarity of dihedral-angle distributions as the main measure for clustering.

Objectives:(A)Explore innovative approaches for grouping amino acids, with a specific emphasis on their functional and structural consequences in protein forecasting. (B)Perform a simulation research to assess the operational features and performance of the suggested clustering method, with the goal of verifying its reliability and efficacy in real-world applications.

Methods: Angular Randomization Test, Permutation Test and Circular Data were applied.

The method can be summarized as follows:

1. Calculate the angular randomization test statistic E.

2. Shuffle group labels in the pooled dataset while keeping dihedral angles intact.

3. Calculate the energy statistic (E∗) for the shuffled data using the same equations.

Correspondence concerning this article should be addressed to the mentioned authors at the mentioned institutes.
Copyright © 2024 Al-Quds University, Deanship of Scientific Research. All rights reserved.
E-mail: research@admin.alquds.edu
Palestine, Abu Dis, Al-Quds University

4. Repeat steps 2 and 3 a specified number of times (nperm) to obtain a distribution of ($E*$) under the null hypothesis.

5.Calculate the P-value by counting the number of times ($E*$) is greater than or equal to E and dividing by the total number of permutations.

Results: Our results indicate that the amino acid clusters formed by glycine, proline, and asparagine are unique and easily distinguishable, providing consistent interpretations. In addition, a simulation study highlights the effectiveness of this technique, indicating positive operational features for amino acid clustering.

Conclusions: This work presents a novel approach using an angular randomization test-based distance metric that can be adjusted for use with any distance-dependent clustering algorithm to categories amino acids. This novel technique provides a different pathway for investigating clusters of amino acids using metrics that are directly linked to the structure and function of proteins, as well as those based on physicochemical and biochemical aspects. The provided two-sample test exhibits notable sensitivity and successfully manages type-I error.

Keywords: angular randomization test, squared Euclidean distance, permutation two-sample test, energy statistics

Correspondence concerning this article should be addressed to the mentioned authors at the mentioned institutes.
Copyright © 2024 Al-Quds University, Deanship of Scientific Research. All rights reserved.
E-mail: research@admin.alquds.edu
Palestine, Abu Dis, Al-Quds University