

A Framework for Data Mining Visualization using Cluster Analysis: An Experimental Study on HIV-AIDS Datasets

S. N. Junaini¹

S. H. Huspi²

A. R. Mat³

N. A. Bolhassan⁴

Faculty of Computer Science & Information Technology

Universiti Malaysia Sarawak, 94300 Kota Samarahan, Sarawak, MALAYSIA

E-mail: ¹syahruln@fit.unimas.my, ²hshazlin@fit.unimas.my, ³marahman@fit.unimas.my, ⁴bnoralam@fit.unimas.my

Abstract

With a need for rational selection of a highly needed bio-informatics-related data, a technique to select, process and visualize the data is crucial. Through visualization, certain hidden form of knowledge can be achieved. Visual data exploration is really useful when little is known about the data and the exploration goals are vague. In this paper we present the method to be developed and implemented in a system prototype that will be able to produce an interactive visualization. The objective of this paper is twofold. First, we would like to investigate which visual data mining technique is the most effective for conveying and understanding bioactivity-related data visualization. Second, to explore new techniques to present and visualize the data as well as to validate the techniques by creating prototype of visualization tools to display and analysis the data. We develop a new framework using clustering technique for visualization and model the data. The dataset that will be used in our study is a subset of a particular AIDS dataset. For this study, 1000 molecules will be tested and analyzed. This paper explains an ongoing project.

activity visualization. Then we would develop a new framework using clustering techniques for visualization and modeling the data. Second, to explore new technique to present and visualize the data. We also would like to validate the technique by creating a prototype of a visualization tools to display and analyze the data.

This paper is organized as follows. First, the basic intuition and explanation behind visualization and data mining is introduced, followed by descriptions of its application in bio informatics. Then, a short review of existing visual data mining techniques is provided. In section 3, we discuss the details about our framework. Section 4 explains the proposed prototype design. The paper concludes with a sketch of future work to improve the framework as well as the prototype.

Keywords

Visualization, data mining, clustering.

1. Introduction

With a need for rational selection of a subset of highly needed bioinformatics data, a technique to select, process and visualize the data is crucial [4]. The aim of the selection process is to obtain maximum amount of bioactivity-related information just by synthesizing and testing minimum numbers of compounds. Also it is important to predict the bioactivity and its functions in related context [10]. A key to gaining an understanding of the bioactivity results is the ability to visually explore and interact with the data.

We present the method that to be developed and implemented in a prototype that will be able to produce an interactive visualization. The user will be able to 'see' what happened to the data tested using the visualization tool. We would also find a more efficient algorithm for solving the problems for data mining of the bioactivity-related data.

The objective of this paper is twofold. First, we would like to investigate which visual data mining technique is the most effective for conveying and understanding bio-related