Scaled Fuzzy Graph for Cluster Analysis in DNA Sequence of Olfactory Receptors

Satya Ranjan Dash, School of Computer Application, KIIT University, Bhubaneswar, India
Satchidananda Dehuri, Department of Systems Engineering, Ajou University, Suwon, Korea
Uma Kant Sahoo, School of Computer Application, KIIT University, Bhubaneswar, India

ABSTRACT

Olfactory receptors (ORs) are responsible for recognition of odor molecules. The deoxyribonucleic acid (DNA) sequences of these receptors are severely affected by local mutations. Therefore, to study the changes among affected and non-affected ORs, the authors attempted to use unsupervised learning (clustering) algorithm. In this paper, they have used a scaled fuzzy graph model for clustering to study the changes before and after the local mutation on DNA sequences of ORs. Their simulation study at the fractional dimensional level confirms its accuracy.

Keywords: Clustering, Deoxyribonucleic Acid (DNA), Fractional Dimension, Fuzzy Graph Model, Olfactory Receptors,

1. INTRODUCTION

The non-trivial method of identifying suitable, narrative, potentially helpful, and ultimately clear/understandable patterns in data is Knowledge Discovery in Databases (KDD) (Cao, 2012). Data mining algorithms find patterns in bulky amounts of data by appropriate models that are not essentially statistical models. In general, the data mining method is to extract information from a data set and convert it into an understandable composition for additional use, the key term is discovery, normally defined as “extracting something new”. Discovered patterns should be true on fresh data with some amount of confidence. Simplify to the future use in some other data. Some of the important tasks of data mining includes: association rule mining, classification, clustering, regression, etc.

At the other hand, Life science research has witnessed a paradigm shift over the last decade with a data rich environment (Frawley, Piatetsky-Shapiro, & Matheus, 1992). The ongoing incursion of these data, the inherent uncertainties in data collection processes, and the gap between data collection and knowledge curation have created ample of scopes for data mining researchers. In fact, most of recent
researches in various life-science related disciplines such as personalized genomics, functional genomics, proteomics and structural genomics, DNA sequence analysis are data driven, where knowledge discovery and data mining processes are playing increasingly important roles.

While tremendous progress has been made over the years, many of the fundamental problems in bioinformatics, such as protein structure prediction, gene–environment interaction, and regulatory pathway mapping are still open. Data mining will continue to play an essential role in understanding these fundamental problems and in the development of novel therapeutic/diagnostic solutions in post-genome medicine.

The olfactory system (Firestein, 2001) has the notable capability to discriminate a wide range of odor molecules. In humans, smell is rather considered to be an esthetic sense in contrast to most other species, which rely on olfaction to detect food, predators, and mates. Terrestrial animals, including humans, smell air-borne molecules, whereas aquatic animals smell water-soluble molecules with low volatility, such as amino acids. Humans are thought to have a poor olfactory ability compared with other animals such as dog or rodents, and yet they can perceive a vast number of volatile chemicals. Of the millions of volatile molecular species that have been catalogued by chemists, hundreds of thousands of distinct odors can be detected by the human nose.

Odorants, typically small organic molecules of less than 400 Da, can vary in size, shape, functional groups and charge (Sosinsky, Glusman, & Lancet, 2000). They include a set of various alcohols, aliphatic acids, aldehydes, ketones and esters; chemicals with aromatic, alicyclic, polycyclic or heterocyclic ring structures; and innumerable substituted chemicals of each of these types, as well as combinations of them.

However, subtle differences in the structure of an odorant, even between two enantiomers, can lead to pronounced modifications in odor quality. In this paper our special focus is to study the changes of OR before or after local mutations, through difference in clustering results on DNA sequences of ORs this work can detect the changes.

Clustering is a widely used knowledge discovery technique. It helps to uncover structures in data that were unknown in the past. The clustering of huge data sets has been the center of attraction in recent years; however, clustering is a difficult task since many existing algorithms fail to do well in scaling with the data size and the amount of dimensions that describe the points. In this paper, we present new clustering techniques, based on the fractal properties of the data sets, using fuzzy graphs and its cohesive matrix.

Clustering has been a fundamental problem in areas like bioinformatics (Jiang, Jang, & Zhang, 2004), data mining, pattern recognition (Jain, Murty, & Flynn, 1999), and image analysis. Clustering techniques used in many applications are either dominated by distance based clustering or connectivity based clustering. A few a-like algorithms have been used in Yeh and Bang (1975). However, the other popular category of clustering based on graph theory approach too igniting many researchers, for application in bioinformatics field.

Graph based clustering is the task of grouping the vertices of the graph into clusters, taking into consideration the edge structure of the graph in such a way that there should be many edges within each cluster and relatively few between the clusters.

Recent study Mathew and Sunitha (2010) shows that the fuzzy graph (Rosenfeld, 1975; Yeh & Bang, 1975) approach is more powerful in cluster analysis than the usual graph theoretic approach, due to its ability to handle the strengths of arcs effectively. Fuzzy graph models can be used to solve various practical issues in clustering analysis, network analysis, information theory, and database theory (Mordeson & Nair, 2000). Obviously this relies on the fact that, these problems should involve uncertainty to get better results. Interested reader can refer (Bhattacharya, 1987; Bhattacharya & Suraweera, 1991) to know more about fuzzy graph theory.
A Framework for Conceptualizing the Current Role and Future Trends of Information Systems in Medical Training
[www.igi-global.com/article/framework-conceptualizing-current-role-future/64351?camid=4v1a](www.igi-global.com/article/framework-conceptualizing-current-role-future/64351?camid=4v1a)

Diabetes-Related Cognitive Decline, a Global Health Issue, and New Treatments Approaches