2006

A Tool for Molecular Similarity Networks

○Chunfei Huang(黄　春飛)[1, 2], Zsolt Lepp[1], Takashi Okada(岡田　孝)[1]
1. Science & Technology Department of Kwansei Gakuin University, Sanda, Hyogo, 651-1413 Japan
2. Computer Science & Technology College of Jilin University, Changchun, 130021 China

ABSTRACT: *In this paper, we introduced a tool for molecular similarity networks. With this tool user can create a molecular similarity network from an SDfile, and can view graphical molecule structures on the network and some of chemical properties about molecules. It is implemented in Java using Cytoscape and JChem libraries. It is a useful tool for chemical analyses of molecular similarity networks.*

1. Introduction

In daily research work we often use Cytoscape to analyze protein interaction networks and JChem to analyses chemical compound fragments separately. It is very inconvenient when we want to create a molecular similarity network and analyze it on molecular level. We searched software on internet for our purpose, but did not find any software to fit our needs. So we have an idea that integrates the above two software's together. According to the idea, we build a tool system to deal with our work using software libraries of Cytoscape[1] and JChem[2].

2. Functions of the tool

The main functions of the tool are creating molecular similarity network and viewing graphical structure of molecules on the network. The details are described as followings:

1) Read information of molecules form SDfile
   Read molecules form an SDfile and save each molecule as a node of the network. Transform each molecular data into Smiles, and save the Smiles string as an attribute of the nodes.
2) Create molecular similarity network
   First, calculate fingerprint for every node. Then, calculate similarities between nodes based on their fingerprints. Last, create edges between nodes according to their similarities.
3) View graphical structure of nodes on the network
   On the molecular similarity network, user can select nodes or edges and then can view graphical molecular structures of selected nodes or nodes linked by selected edges.
4) View similarity between nodes and some node properties on the network
   When edges are selected, user can view the similarity between nodes linked by the edges. When nodes are selected, user can view Smiles string, count of atoms, and other attributes of the nodes.

3. Software designs and implementation

We designed the tool as plugins of Cytoscape. Cytoscape is an open source software and let users extend plugin class to add user defined functions. To create molecular similarity network and display graphical molecular structures, we included JChem library in our classes. We divided the tool into two modules, the one is for creating molecular similarity networks named as "Molecule Network" and the other is for displaying graphical molecular structures named as "Molecule Viewer".

1) Molecule Network Module

In this module, the CreateMoleculeNetwork class fulfils main functions of the module by algorithm MolNetTask.

Algorithm MolNetTask: (This algorithm read molecules from SDfile, and produces smiles node attribute file and similarity network file.)

[1]. Read database connection information and similarity threshold from user defined property file;

[2]. Get SDfile path selected by user and read in molecule data from the file;

[3]. Transform molecule data to Smiles, calculate fingerprints and store this information into database;

[4]. Calculate similarities between molecules; create edges between nodes if their similarity is greater then the threshold given by user;

[5]. Save smiles strings to node attribute file; save similarity edges to network file. ‖

2) Molecule Viewer Module

In this module, the MoleculeViewerAction class monitors user actions by listener. The actions are describe as following:

Algorithm ViewerAction: (This algorithm listens to user requires and display molecular structures.)

[1]. If mouse is on a node then display the smiles string of the node;

[2]. If display command is required then do from [3] to [5];

[3]. dispNodes←{selected nodes}＋{nodes linked by selected edges};

[4]. If ｜dispNodes｜＞2 then display graphical molecular structures on an independent window;

[5]. Else if ｜dispNodes｜＞0 then display graphical molecular structures on the left tab panel. ‖

The tool is developed with Eclipse3.2.2[4] and Java 2SE5.0[3] run time library. It runs as a java application and requires that Cytoscape 2.4.1, JChem 3.2.7 and a database system (MySQL, Oracle, and Access etc) must be preinstalled.

4. Conclusions

We introduced a tool for molecular similarity network. With the help of the tool we can do similarity analyses easily and efficiently. But there are some further works to be done. Firstly, in the network, node denotes molecule, edge denotes similarity between nodes. If apply other chemical properties between molecules to edge, we can do more chemical analyses from the network. Secondly, the tool runs as a java application, we can improve the tool to run as web services. And finally, the complexity of similarity calculation is n*(n-1)/2, where n denotes the number of node. So the network can not be too big. We must work out a faster similarity calculation method for larger networks.

5. References

[1]. http://www.cytoscape.org/cgi-bin/moin.cgi/Developer_Homepage

[2]. http://www.chemaxon.com/dev_guide_land.html

[3]. http://java.sun.com/

[4]. http://www.eclipse.org/