

Novel Methods and Tool for Clustering of Simple and Bipartite Graphs: Applications in Ecology and Computational Biomedical Research

Md. Altaf-Ul-Amin
Associate Professor
Nara Institute of Science and Technology, Japan

Abstract

Network analysis particularly graph clustering has become a useful and important technique in data mining applications. It provides a global view of data structure where highly concentrated data are grouped based on their common properties. Previously we developed graph clustering algorithms DPCLUS and DPCLUSO. Recently, we proposed a novel biclustering approach called BiCLUSO. We compared our biclustering algorithm with five different algorithms using biological and synthetic data and evaluated the performances. Our algorithm shows the best performance over the selected five biclustering algorithms. We also present new integrated software implementing the DPCLUSO and BiCLUSO algorithms to be utilized for simple and bipartite graph clustering. This tool provides the user with GUI based facilities for simple and bipartite graph clustering along with filtering and amalgamation facilities, hierarchical node analysis, node distribution among cluster set and visualization of all or partial portion of a big cluster set. We used this tool to analyze the bipartite relations between species and volatile organic compounds (VOCs). VOCs emitted by different species have huge environmental and ecological impacts. Biosynthesis of VOCs depends on different metabolic pathways based on which the species can be categorized. Our experiment shows that VOC based classification is consistent with taxonomy based classification of the species. Furthermore, we applied simple graph clustering algorithm DPCLUSO for finding inflammatory Bowel Disease (IBD) related genes. We also analyzed the mRNA and miRNA bipartite relations. In such relation a miRNA-regulatory module (MRM) is a subset of miRNA target interactions (MTI) where a group of miRNAs participate cooperatively by regulating a bunch of genes to control different biological processes. We mainly focused on MRMs detection from MTIs involving the IBD related genes. We evaluated the relevance of the miRNAs with IBD by counting their occurrences in different MRMs and their interactions with known IBD genes. Finally, we successfully identified some important IBD related miRNAs. Also, we will explain KnapSack database which has been developed in our lab.

Md. Altaf-Ul-Amin received B.Sc. degree in Electrical and Electronic Engineering from Bangladesh University of Engineering and Technology (BUET), Dhaka, M.Sc. degree in Electrical, Electronic and Systems Engineering from Universiti Kebangsaan Malaysia (UKM) and PhD degree from Nara Institute of Science and Technology (NAIST), Japan. He received the best student paper award in the IEEE 10th Asian Test Symposium. Also, he received two other best paper awards as a co-author of journal articles. He previously worked in several universities in Bangladesh, Malaysia and Japan. Currently he is working as an associate professor in Computational Systems Biology Lab of NAIST. He is conducting research on Network Biology, Systems Biology, Cheminformatics and Biological Databases. He published around 90 peer reviewed papers in international journals and conference proceedings. Current google scholar citation index of his publications is more than 7000.

TRADUÇÃO LIVRE

Novos métodos e ferramenta para agrupamento de gráficos simples e bipartidos:
Aplicações em Ecologia e Investigação Biomédica Computacional

Md. Altaf-UI-Amin
Professor Associado
Instituto de Ciência e Tecnologia de Nara, Japão

RESUMO

A análise de redes, particularmente a agregação de gráficos, tornou-se uma técnica útil e importante em aplicações de mineração de dados. Fornece uma visão global da estrutura de dados onde os dados altamente concentrados são agrupados com base nas suas propriedades comuns. Anteriormente desenvolvemos algoritmos de agregação de gráficos DPCLUS e DPCLUSO. Recentemente, propusemos uma nova abordagem de biclustering chamada BiCLUSO. Comparámos o nosso algoritmo de biclustering com cinco algoritmos diferentes utilizando dados biológicos e sintéticos e avaliamos os desempenhos. O nosso algoritmo mostra o melhor desempenho em relação aos cinco algoritmos de biclustering seleccionados. Apresentamos também novo software integrado que implementa os algoritmos DPCLUSO e BiCLUSO a serem utilizados para agrupamento simples e bipartido de gráficos. Esta ferramenta fornece ao utilizador facilidades baseadas em GUI para agrupamento de gráficos simples e bipartidos juntamente com facilidades de filtragem e amalgamação, análise hierárquica de nós, distribuição de nós entre conjuntos de agrupamento e visualização de todo ou parte de um grande conjunto de agrupamento. Utilizamos esta ferramenta para analisar as relações bipartidas entre espécies e compostos orgânicos voláteis (COVs). Os COVs emitidos por diferentes espécies têm enormes impactos ambientais e ecológicos. A Biossíntese de COVs depende de diferentes vias metabólicas com base nas quais as espécies podem ser categorizadas. A nossa experiência mostra que a classificação baseada em COV é consistente com a classificação das espécies baseada na taxonomia. Além disso, aplicamos o algoritmo simples de agrupamento de gráficos DPCLUSO para encontrar genes relacionados com a doença inflamatória intestinal (IBD). Analisamos também as relações bipartites mRNA e miRNA. Nesta relação, um módulo regulador do miRNA (MRM) é um subconjunto de interações alvo do miRNA (MTI) onde um grupo de miRNAs participa cooperativamente regulando um grupo de genes para controlar diferentes processos biológicos. Concentramo-nos principalmente na detecção de MRMs a partir de MTIs envolvendo os genes relacionados com o IBD. Avaliamos a relevância dos miRNAs com IBD contando as suas ocorrências em diferentes MRMs e as suas interações com genes IBD conhecidos. Finalmente, identificamos com sucesso alguns miRNAs importantes relacionados com a DII. Também explicaremos a base de dados KNApSACk que foi desenvolvida no nosso laboratório.

Md. Altaf-UI-Amin é licenciado em Engenharia Electrotécnica e Electrónica pela Universidade de Engenharia e Tecnologia do Bangladesh (BUET), Dhaka, M.Sc. em Engenharia Electrotécnica, Electrónica e de Sistemas pela Universiti Kebangsaan Malaysia (UKM) e doutoramento pelo Nara Institute of Science and Technology (NAIST), Japão. Recebeu o prémio de melhor trabalho de estudante no 10º Simpósio de Testes Asiáticos do IEEE. Recebeu ainda dois outros prémios de melhor artigo como co-autor de artigos de jornal. Trabalhou anteriormente em várias universidades no Bangladesh, Malásia e Japão. Actualmente, trabalha como professor associado no Laboratório de Biologia de Sistemas Computacionais do NAIST. Está a conduzir investigação sobre Biologia de Redes, Biologia de Sistemas, Cheminformática e Bases de Dados Biológicos. Publicou cerca de 90 artigos revistos por pares em revistas internacionais e anais de conferências. O índice actual de citação académica google das suas publicações é superior a 7000.