Pair wise Alignment of Pathways

GeorgiaState University

Qiong Cheng; Alexander Zelikovsky Department of Computer Science, Georgia State University, Atlanta, GA 30303

Outline

- Motivation
- Problem Formulation
- Solution & Results From the First Stages
- # Plans for Next Stage
- Helated Works
- Reference

Motivation

- The Homeomorphism problem of pair wise graphs has been a open problem in the theory of graph. There have been lots of applications which can be reducible to the problem. We aims to propose a feasible solution to calculate the minimal cost of homeomorphism alignment in biological pathways so that we can
- Study the approximate solution to the homeomorphism problem in the system biology;
- Study algorithms to calculate the significant similarity of pathways in biological networks;
- Study the evolution of pathways in different species and different functions in the same species;
- > Study the reconstruction problem of phylogenetic networks .

Problem Formulation

Given pattern tree T and text graph G, find a homeomorphic mapping that minimize the homeomorphic

- Def. A mapping from pattern $T = \langle V^T, E^T, L^T \rangle$ to text $G = \langle V^G, E^G, L^G \rangle (f : Pattern T \rightarrow$ [eztG) is called homeomorphic if
- 1. every vertex in V^T is mapped a vertex in V^G $(f_v : V^T \rightarrow V^G)$; \pm every edge $\epsilon = (u, v)$ in E^T is mapped to a path in $P^G(f_c : E^T \rightarrow P^G)$. The path is a ordered sequence of the seq $u_0=f(u), u_1, u_2, ..., u_k=f(v) \ of \ vertices \ in \ P^G;$

Def. Let $\Delta : L^T \times L^G \rightarrow \Re$ denote the 3-place relation on the set of all ordered pairs of a vertex in V^2 and a vertex in V^G to a real number. The relation represents a 2-dimension table of the vertex-to-vertex ularity cost scores. $\delta : L^G \rightarrow \Re$ denote the binary relation. Its domain is all vertices in V^G , and its unge is the real number set. § represents the penalty score for deleting a vertex from a path. The cost of a homeomorphic mapping $f: T \rightarrow G$ is

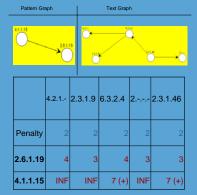
 $cost(f) = \sum \Delta(v, f(v)) + \delta \times \sum (|f(e)| - 1)$

Our Solution

- Employ dynamic program to get an exact result for the specified problem
- Build a two dimension table for dynamic program by the post order traversal; Recursive Function:



Table for Dynamic Program



Results from the First Stage

- We did experiments on alignment of pathways of bacterium Escherichia coli and the yeast Saccharomyces cerevisiae;
- It supports the alignment of directed cyclical and acyclical pattern graphs and directed cyclical and acyclical text graphs;
- It supports two types of calculation of significance based on random graph by reshuffling edges and nodes of the original text graph;
- Statistical Significant of alignments (An example-chorismate biosynthesis in E., Coli)

	1	2	3
ReshufflingEdge	0.07	0.07	0.07
ReshufflingNode	0.18	0.19	0.19
Dr.Pinter's solution	0.29	0.29	0.29

Plans for the Next Stage

- To further prove whether the homeomorphism problem on biological pathways is NP-completeness;
- To improve our approach to the node-disjoint path problem so that it has higher significant similarity of pathways;
- Try to employ linear programming to obtain the approximate solution;
- > Accomplish more experiments to study the evolution of > Try the clustering technology to study the evolution of
- pathways

Related Works

- ≻ Kelley et al (2003)
- Goal: identify conserved pathways (chains) - Idea: can be done efficiently by reducing it to sequence alignments
- ≻ Ron Pinter (2005)
- Goal: identify conserved pathways (directed acyclic graphs)
- Idea: Dynamic program for mapping on multi source trees

Reference

- 1. Ron Y.Pinter, Oleg Rokhlenko, Esti Yeger-Lotem and Michal Ziv-Ukelson (2005) Algnment of metabolic pathways Vol.21 no. 16 Bioinformatics
- Ming Chen and Ralf Hofest, An algorithm for linear metabolic pathway alignment, *In silico biology ISSN*, 1386-6338:111-128, (2005)
- Tohsata Y., Matsuda H. and Hashimot A. A Multiple Alignment Algorithm for Metabolic Pathway Analysis using Enzyme Hierarchy, Proc. 8th International Conference on Intelligent Systems for Molecular Biology, (ISMB 2000): 376-383
- M. Garey and D. Johnson. Computers and Intractability: A Guide to the Theory of NP-Completeness. W.H. Freeman, New York, 1979

For additional information, please contact:

Alexander Zelikovsky Giong Cheng Computer Science Department Georgia State University