Discovering Recursive Patterns in Biological Networks

Jacek Kukluk, Chang Hun You, Lawrence Holder, and Diane Cook, Department of Computer Science and Engineering, University of Texas at Arlington, Arlington, TX 76019-0015.

We investigate a machine learning algorithm for discovering recursive patterns from data. The data represents biological structures as a graph. Entities are represented as nodes and the relationships between them as edges. We show the mining process which infers hierarchical structure in the form of graph grammars by compressing portions of a graph described by a production and then inferring new productions on the compressed graph. We show how this algorithm performs in experiments with biological networks. One of the questions we address is if there is a common recursive pattern in several networks of the same species, e.g., glycolysis / gluconeogenesis, citrate cycle, and pentose phosphate pathways of Escherichia coli. We also show the results of experiments with chemical structures including macrocyclic gallium carboxylate, water-soluble tin-based metallodendrimer, and dendronized polymer. We found repetitive connected components which capture the underlying motifs, the basic building blocks of the structures.