Chapter XIV

Differential Association Rules: Understanding Annotations in Protein Interaction Networks

Christopher Besemann, North Dakota State University, USA
Anne Denton, North Dakota State University, USA
Ajay Yekkirela, North Dakota State University, USA
Ron Hutchison, The Richard Stockton College of New Jersey, USA
Marc Anderson, North Dakota State University, USA

Abstract

In this chapter, we discuss the use of differential association rules to study the annotations of proteins in one or more interaction networks. Using this technique, we find the differences in the annotations of interacting proteins in a network. We extend the concept to compare annotations of interacting proteins across different definitions of interaction networks. Both cases reveal instances of rules that explain known and unknown characteristics of the network(s). By taking advantage of such data mining techniques, a large number of interesting patterns can be effectively explored that otherwise would not be.
Introduction

Many high-throughput techniques have been developed to identify protein interaction networks in organisms. One conventional definition of interactions within an organism considers whether the protein molecules interact physically. High-throughput techniques to test for such interactions include the popular yeast-two-hybrid method (Ito et al., 2001; Uetz et al., 2000). Another common definition of interactions relies on the concept of genetic interactions in which the lethality of two gene deletions is examined as in (Tong et al., 2001; Tong et al., 2004). Finally, there also exist a number of computer-driven procedures like the identification of domain-fusion interactions. Two genes have a domain fusion interaction in one species if their homologs occur as exons of a single gene in another species (Marcotte et al., 1999; Truong & Ikura, 2003). With so much interaction data available, there is a clear need to develop techniques that are targeted at understanding the role of networks in bioinformatics.

Interactions are not the only type of information available for proteins. Many genes and proteins are studied extensively in model species, such as yeast. A rich supply of experimental data, in the form of annotations, is available for these species and can be studied in much the same way as data associated with objects in other disciplines. These annotations are the “items” of information that belong to particular objects. A popular technique for the study of such items is association rule mining (ARM). ARM has the goal of discovering associations between frequently occurring sets of items. In the prototypical case, ARM is used in the analysis of shopping transactions (Agrawal, Imielinski, & Swami, 1993; Han & Fu, 1995); the technique has also been applied in relational database settings (Dehaspe & Raedt, 1997; Oyama, Kitano, Satou, & Ito, 2002). The combination of network data and protein annotation data is a special case of relational data. In a relational database, one table is used to describe protein annotations and one or more separate tables define the network(s). A relational framework allows analysis beyond statistical comparison of single annotations of interacting proteins (Schwikowski, Uetz, & Fields, 2000). Relational frameworks also support the study of associations involving more complex attribute combinations. Within the same framework, it also becomes possible to contrast different interaction definitions.

This chapter explores a novel ARM algorithm called differential ARM that enables the discovery of differences between interacting proteins and between multiple network types. In the Background section, we describe representative work on protein interactions and introduce key definitions in association rule mining. The next section, Mining Differential Association Rules, defines our goals and the problems to overcome, ending with a description of the differential ARM method. Finally, in the section on Differential Association Rule Experiments, an experiment on real-world yeast interaction data is described.
Prioritizing Disease Genes and Understanding Disease Pathways
[www.igi-global.com/article/prioritizing-disease-genes-and-understanding-disease-pathways/101241?camid=4v1a](www.igi-global.com/article/prioritizing-disease-genes-and-understanding-disease-pathways/101241?camid=4v1a)