Chapter 2.16
Differential Association Rules:
Understanding Annotations in Protein Interaction Networks

Christopher Besemann
North Dakota State University, USA

Anne Denton
North Dakota State University, USA

Ajay Yekkirala
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.

BACKGROUND

We begin by discussing related work in the area of protein interaction network analysis. Basic notation for association rule mining is covered and notes on related references in the area of ARM are given.

Protein Interaction Network Analysis

Several papers have analyzed yeast protein interaction networks. The majority of these works have studied isolated networks (Wagner, 2001) or used comparisons of the statistics of multiple networks (Maslov & Sneppen, 2002). In Schwikowski et al. (2000) the authors did some analysis of the yeast physical protein interaction network using "counting" methods to determine percentages of interactions that have certain functional behavior. Most examples of interest are techniques that require some manual intervention. In larger datasets, in which patterns of interest are harder to perceive, automated techniques would greatly assist analysis. Tong et al. (2004) performed statistical analysis on part of the yeast genetic interaction network. In this work, the authors begin with observations of interesting statistics observed within interaction networks, and conclude with clustering analysis and population genetics. Some of the conclusions drawn relate to the density of the network and comparison of how the genetic interaction network predicts parts of the physical interaction network. Again, the emphasis is on a single network and simple statistical comparisons. In Ozier, Amin, and

Related Content

Towards an Ethics Framework for Learning Analytics
André Pretorius (2023). *Investigating the Impact of AI on Ethics and Spirituality (pp. 107-136).*
www.irma-international.org/chapter/towards-an-ethics-framework-for-learning-analytics/331961

Assessing Smart-Home Platforms for Ambient Assisted Living (AAL)
Michiel Brink, Ignacio González Alonsoand Johanna E.M.H. van Bronswijk (2013). *International Journal of Ambient Computing and Intelligence (pp. 25-44).*
www.irma-international.org/article/assessing-smart-home-platforms-for-ambient-assisted-living-aal/104159

Content Based Search Engine for Historical Calligraphy Images
Xiafen Zhangand Vijayan Sugumaran (2014). *International Journal of Intelligent Information Technologies (pp. 1-18).*
www.irma-international.org/article/content-based-search-engine-for-historical-calligraphy-images/116740

Design, Measurements, and Analysis of Enhanced Bandwidth UWB: On-Body Antenna for Ambient Intelligence Environment
Raghvendra Singh, Kanad Ray, Preecha Yupapinand Jalil Ali (2021). *International Journal of Ambient Computing and Intelligence (pp. 140-158).*
www.irma-international.org/article/design-measurements-and-analysis-of-enhanced-bandwidth-uwb/272042

Towards Ontological Structures Extraction from Folksonomies: An Efficient Fuzzy Clustering Approach
Marouf Zahiaand Benslimane Sidi Mohamed (2014). *International Journal of Intelligent Information Technologies (pp. 40-50).*
www.irma-international.org/article/towards-ontological-structures-extraction-from-folksonomies/123943