Chapter 18
On Bias–Variance Analysis for Probabilistic Logic Models

Huma Lodhi
Imperial College London, UK

ABSTRACT
Deliberate exploitation of natural resources and excessive use of environmentally abhorrent materials have resulted in environmental disruptions threatening the life support systems. A human centric approach of development has already damaged nature to a large extent. This has attracted the attention of environmental specialists and policy makers. It has also led to discussions at various national and international conventions. The objective of protecting natural resources cannot be achieved without the involvement of professionals from multidisciplinary areas. This chapter recommends a model for the creation of knowledge-based systems for natural resources management. Further, it describes making use of unique capabilities of remote sensing satellites for conserving natural resources and managing natural disasters. It is exclusively for the people who are not familiar with the technology and who are given the task of framing policies.

INTRODUCTION
During the past few years, many methods have been proposed that are at the intersection of logic and probability. The distinguishing characteristic of these techniques is the provision of efficient representation for complex real world problem. For example, Stochastic Logic Programs (SLPs) (Muggleton, 1996) combine probabilistic models such as Stochastic Context Free Grammars (SCFG) and Hidden Markov Models (HMM) with first order logic. SLPs provide an efficient representation for complex biological systems such as metabolic pathways. Another competitive approach is termed Bayesian Logic Programs (BLPs) (Kersting & De Raedt) that are considered as generalization of Bayesian nets and logic programs. Newly proposed Markov logic networks (MLNs) (Richardson & Domingos, 2006) are also a useful example of combining probabilistic graphical models and logic.
More recently, there has been a growing interest in learning probabilistic logic representations. For instance, Failure Adjusted Maximization (FAM) (Cussens, 2001) is a useful tool to learn parameters in SLPs and Balios (Kersting & Dick, 2004) is a system that performs inference and learning in BLPs. While systems and techniques have been proposed for probabilistic logic learning, research has not been conducted in the important direction of analyzing the performance of probabilistic logic learners. Tools and methods to study the performance of probabilistic logical learning algorithms have not been investigated. In this article we focus on the unexplored research direction. We propose bias-variance (BV) decomposition to analyze and investigate the prediction (estimation) performance of the probabilistic logic learning algorithms for parameter estimation task.

In this article specifically focus on a particular approach, namely Stochastic Logic Programs that provide formalism for probabilistic knowledge representation. We employ FAM to learn parameters on the SLP. In order to study and analyze the prediction performance of probabilistic logic learning algorithms such as FAM we present definitions for bias and variance.

In order to empirically analyze the BV definitions we focus on a challenging and fundamental task in computational systems biology, namely quantitative modelling of metabolic pathways. Recent research has shown critical importance of quantitative aspects of biological information stored in complex networks and pathways for the system level understanding of biology (Kitano 2002a, 2002b). We have conducted experiments using metabolic pathways in Saccharomyces cerevisiae. We have applied the proposed bias-variance definitions for estimating the rates of reactions catalyzed by enzymes in pathways using FAM.

The article is organized as follows. Section 2 briefly explains metabolic pathways. In Section 3 we describe logic programming concepts, SLPs and FAM. We present BV decomposition for parameter estimation task in Section 4. Experimental results are described in Section 5 and Section 6 concludes the article.

METABOLIC PATHWAYS

Metabolic pathways, an important class of biological systems, represent chemical reactions within the confines of a cell. They comprise metabolites and enzymes and may be viewed as series of enzyme-catalyzed reactions in which product of one reaction becomes substrate for the next reaction. Dynamics of biological system and behaviour of enzymes in metabolic pathways can be studied by applying the Michaelis-Menten (MM) enzyme kinetic function, but the well-known method, namely Lineweaver-Burk or double reciprocal method (Lineweaver & Burk, 1934) is not free of problems. Dowd and Riggs (1945) analysis of the method gave discouraging and unsatisfactory results for estimating the kinetic parameters. Ritchie and Pravan (1996) have also observed that Lineweaver-Burk method can lead to unsatisfactory results. Furthermore it is computationally exhaustive to solve the MM equation using numerical methods (Duggleby 1994, 1995; Schnell & Mendoza, 2001).

In order to analyze and model metabolic pathways Petri nets have been used. Petri nets (Petri, 1962) are bipartite directed graphs. In a Petri net representation metabolites are represented by place nodes and transition nodes represent enzyme catalyzed reactions. Reddy, Mavrovouniotis, and Liebman (1993) and Hofestaedt (1994) pioneered the use of Petri nets for modelling metabolic pathways. In Angelopoulos and Muggleton (2002) SLPs have been applied to represent metabolic pathways and in order to learn metabolic pathways models FAM has been used. Lodhi and Muggleton (2004) have presented ensembles of FAM to induce quantitative models of metabolic pathways. In this article we analyze the performance of FAM for
Related Content

An Overview of Intra-Body Communication Transceivers for Biomedical Applications
www.igi-global.com/chapter/an-overview-of-intra-body-communication-transceivers-for-biomedical-applications/112359?camid=4v1a

Probabilistic Methods in Automatic Speech Recognition
www.igi-global.com/chapter/probabilistic-methods-in-automatic-speech-recognition/112333?camid=4v1a

Evaluating IS Quality: Exploration of the Role of Expectations on Stakeholders' Evaluation
Carla Wilkin, Rodney Carr and Bill Hewett (2001). Information Technology Evaluation Methods and Management (pp. 111-129).
www.igi-global.com/chapter/evaluating-quality-exploration-role-expectations/23671?camid=4v1a

Modified Distance Regularized Level Set Segmentation Based Analysis for Kidney Stone Detection
www.igi-global.com/article/modified-distance-regularized-level-set-segmentation-based-analysis-for-kidney-stone-detection/133531?camid=4v1a