Artificial Intelligence ••• (••••) •••–•••

Contents lists available at ScienceDirect

Artificial Intelligence

www.elsevier.com/locate/artint

Learning Bayesian network parameters under equivalence constraints

Tiansheng Yao *, Arthur Choi, Adnan Darwiche

Computer Science Department, University of California, Los Angeles, Los Angeles, CA 90095, USA

ARTICLE INFO

Article history:
Received in revised form 15 May 2015
Accepted 23 May 2015
Available online xxxx

Keywords:
Bayesian networks
Parameter learning
Learning with constraints
Semi-supervised learning

ABSTRACT

We propose a principled approach for learning parameters in Bayesian networks from incomplete datasets, where the examples of a dataset are subject to equivalence constraints. These equivalence constraints arise from datasets where examples are tied together, in that we may not know the value of a particular variable, but whatever that value is, we know it must be the same across different examples. We formalize the problem by defining the notion of a constrained dataset and a corresponding constrained likelihood that we seek to optimize. We further propose a new learning algorithm that can effectively learn more accurate Bayesian networks using equivalence constraints, which we demonstrate empirically. Moreover, we highlight how our general approach can be brought to bear on more specialized learning tasks, such as those in semi-supervised clustering and topic modeling, where more domain-specific approaches were previously developed.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

In machine learning tasks, the examples of a dataset are generally assumed to be independent and identically distributed (i.i.d.). There are numerous situations, however, where this assumption does not hold, and there may be additional information available that ties together the examples of a dataset. We can then, in turn, exploit this background knowledge to learn more accurate models.

Consider, as a motivating example, the following scenarios that arise in medical diagnosis, where we would like to learn a model that could be used to diagnose diseases from symptoms. Typically, we would have data consisting of patient records, which we assume to be independent. However, we may obtain further information that ties some of these records together. For example, we may learn that two patients are identical twins, and hence may both be subject to increased risk of certain genetic diseases, i.e., they share the same genetic variants that may cause certain genetic disorders. We may also, for example, learn that two patients were both exposed to a third patient, who was diagnosed with a contagious disease. When learning a model from data, we would like to be able to take advantage of this type of additional information, when it is available.

We can view this type of additional information more generally as *equivalence constraints* that bear on an incomplete dataset, where we may not know the particular value of a variable, but whatever that value is, we know that it must be the same across different examples in our dataset. In this paper, we introduce a simple but principled way to deal with such additional information. In particular, we introduce and formalize the problem of *learning under equivalence constraints.* We

* Correspondence to: T. Yao, University of California, Los Angeles, 4801 Boelter Hall, Los Angeles, CA 90095, USA. Tel.: +1 310 794 4343; fax: +1 310 794 5057.

E-mail addresses: tsyao@cs.ucla.edu (T. Yao), aychoi@cs.ucla.edu (A. Choi), darwiche@cs.ucla.edu (A. Darwiche).

http://dx.doi.org/10.1016/j.artint.2015.05.007
0004-3702/© 2015 Elsevier B.V. All rights reserved.

Please cite this article in press as: T. Yao et al., Learning Bayesian network parameters under equivalence constraints, Artificial Intelligence (2015), http://dx.doi.org/10.1016/j.artint.2015.05.007

first introduce the notion of a *constrained dataset*, which implies a corresponding *constrained log likelihood*. We then define the problem of learning the parameters of a Bayesian network from a constrained dataset, by maximizing the constrained log likelihood.

There are a variety of applications, across a variety of different domains, that can be viewed as learning from a constrained dataset. For example, in the information extraction task of named-entity recognition, we seek to label the elements of a text by the type of entity that they refer to (e.g., in an abstract for a talk, we would want to identify those elements that refer to the speaker). Hence, if we see a name that appears multiple times in the same text, we may presume that they all refer to an entity of the same type [1] (an equivalence constraint). As another example, in the task of (vision-based) activity recognition [2], our goal is to annotate each frame of a video by the activity that a human subject is involved in. In this case, a video could be partially annotated by a human labeler, specifying that different frames of a video that depict the same activity (again, an equivalence constraint).

Indeed, the notion of an equivalence constraint, for the purposes of learning, has appeared before in a variety of different domains (either implicitly or explicitly), where a variety of domain-specific approaches have been developed for disparate and specialized tasks. One notable domain, is that of semi-supervised clustering [3]. Here, the notion of a *must-link constraint* was proposed for k-means clustering, to constrain those examples that are known to belong to the same cluster[1]; see, e.g., [4,5]. For example, when clustering different movies, a user may find that the clusters they learned assigned two different movies to two different clusters, when they should have been assigned to the same cluster (say, based on their personal preferences). In the topic modeling domain, a significantly different approach was proposed to accommodate must-link constraints (based on Dirichlet forest priors), to assert that different words should appear in the same topic (with high probability) [6].

In this paper, we show how the different tasks described above can be viewed uniformly as learning a Bayesian network from a dataset that is subject to equivalence constraints. We further propose a simple but principled way of learning a Bayesian network from such a dataset, which is competitive with, and sometimes outperforming, more specialized approaches that were developed in their own domains. Given the simplicity and generality of our approach, we further relieve the need to (a) derive new and tailored solutions for applications in new domains, or otherwise (b) adapt or generalize existing solutions from another domain (both non-trivial tasks).

Our paper is organized as follows. In Section 2, we review the task of learning Bayesian networks from incomplete datasets. In Section 3, we introduce the notion of a constrained dataset, and in Section 4 we introduce the corresponding notion of a constrained log likelihood. In Section 5, we consider the problem of evaluating the constrained log likelihood, and in Section 6, we discuss an iterative algorithm for optimizing it. In Section 7, we evaluate our approach for learning Bayesian networks from constrained datasets, further comparing it with more specialized approaches from two different domains: semi-supervised clustering and topic modeling. Finally, we review related work in Section 8, and conclude in Section 9.

2. Technical preliminaries

We use upper case letters (X) to denote variables and lower case letters (x) to denote their values. Sets of variables are denoted by bold-face upper case letters ($\mathbf{X}$), and their instantiations by bold-face lower case letters ($\mathbf{x}$). Generally, we will use X to denote a variable in a Bayesian network and $\mathbf{U}$ to denote its parents. A network parameter will further have the general form $\theta_{x|\mathbf{u}}$, representing the probability $Pr(X = x|\mathbf{U} = \mathbf{u})$. We will further use θ to denote the set of all network parameters.

Given a network structure G, our goal is to learn the parameters of the corresponding Bayesian network, from an incomplete dataset. We use $\mathcal{D}$ to denote a dataset, and $\mathbf{d}_i$ to denote an example. Typically, one seeks parameter estimates θ that maximize the log likelihood, defined as:

$$LL(\theta \mid \mathcal{D}) = \sum_{i=1}^{N} \log Pr_\theta(\mathbf{d}_i), \tag{1}$$

where Pr_θ is the distribution induced by network structure G and parameters θ. In the case of complete data, the maximum likelihood (ML) parameters are unique and easily obtainable. In the case of incomplete data, obtaining the ML parameter estimates is more difficult, and iterative algorithms, such as Expectation–Maximization (EM) [7,8], are typically employed.

In this paper, we are interested in estimating the parameters of a Bayesian network from a similar perspective, but subject to certain equivalence constraints, which we introduce in the next section. Our approach is largely motivated by the use of *meta-networks*, which are more commonly used for Bayesian parameter estimation [9,10]. In a meta-network,

[1] Similarly, *must-not-link constraints* were also considered, for examples that belong to different clusters.

Please cite this article in press as: T. Yao et al., Learning Bayesian network parameters under equivalence constraints, Artificial Intelligence (2015), http://dx.doi.org/10.1016/j.artint.2015.05.007

Download English Version:

https://daneshyari.com/en/article/4942069

Download Persian Version:

https://daneshyari.com/article/4942069

Daneshyari.com