Journal of Biomedical Informatics 52 (2014) 373–385

Contents lists available at ScienceDirect

Journal of Biomedical Informatics

journal homepage: www.elsevier.com/locate/yjbin

Combining data and meta-analysis to build Bayesian networks for clinical decision support

Barbaros Yet [a,*], Zane B. Perkins [b], Todd E. Rasmussen [c], Nigel R.M. Tai [d], D. William R. Marsh [a]

[a] School of Electronic Engineering and Computer Science, Queen Mary University of London, UK
[b] Centre for Trauma Science, Queen Mary University of London, UK
[c] The United States Army Medical Research and Materiel Command, MD, USA
[d] The Royal London Hospital, London, UK

ARTICLE INFO

Article history:
Received 17 June 2014
Accepted 31 July 2014
Available online 9 August 2014

Keywords:
Clinical decision support
Bayesian networks
Evidence-based medicine
Evidence synthesis
Meta-analysis

ABSTRACT

Complex clinical decisions require the decision maker to evaluate multiple factors that may interact with each other. Many clinical studies, however, report 'univariate' relations between a single factor and outcome. Such univariate statistics are often insufficient to provide useful support for complex clinical decisions even when they are pooled using meta-analysis. More useful decision support could be provided by evidence-based models that take the interaction between factors into account. In this paper, we propose a method of integrating the univariate results of a meta-analysis with a clinical dataset and expert knowledge to construct multivariate Bayesian network (BN) models. The technique reduces the size of the dataset needed to learn the parameters of a model of a given complexity. Supplementing the data with the meta-analysis results avoids the need to either simplify the model – ignoring some complexities of the problem – or to gather more data. The method is illustrated by a clinical case study into the prediction of the viability of severely injured lower extremities. The case study illustrates the advantages of integrating combined evidence into BN development: the BN developed using our method outperformed four different data-driven structure learning methods, and a well-known scoring model (MESS) in this domain.

1. Introduction

It is a challenge to build effective decision-support models for complex clinical problems; such problems involve multiple interacting factors [1,2] and to account for both the factors and their interaction a 'multivariate' model is needed [3]; these can have many forms: our focus is on Bayesian networks. In general, a multivariate model can be built in a number of ways: (1) purely from data using statistical and machine learning techniques [4], (2) from a combination of clinical knowledge and data [5–7] or (3) from published literature using multivariate meta-analysis techniques [8]. Each of these techniques has been shown to be successful in certain conditions but in this paper, we focus on clinical problems where none of these techniques is sufficient, on its own, to build a useful decision support model. That is, our focus is on problems that are complex, important but also rare: their rarity makes it hard to collect very large datasets (so called 'big data'); their complexity demands a sophisticated multivariate model but their importance ensures that a large number of relevant research studies is available.

In these domains, the first method of building models – purely from data – results in simple models that cannot deal with the complexity of the problem [1] because there is not enough data to support a complex model. The third approach fails because clinical studies rarely publish information detailed enough for multivariate meta-analysis [9]. Instead, many medical studies report 'univariate' relations between a single factor and an outcome. Randomised controlled trials, for example, analyse the effect of a single treatment by using randomisation to decrease the confounding effect of other variables [10]. Similarly, many observational studies report the relation between individual risk factors and outcomes even when their dataset contains information about multiple factors. The second approach – combining knowledge and data – could work but it ignores the large body of published evidence; our challenge is therefore to exploit the results of a meta-analysis of studies reporting univariate relations to supplement a dataset that is otherwise inadequate to support a complex multivariate model.

* Corresponding author. Address: Risk and Information Management Research Group, Room CS332, School of Electronic Engineering and Computer Science, West Square, Queen Mary University of London, E1 4NS London, UK.
E-mail address: b.yet@qmul.ac.uk (B. Yet).

http://dx.doi.org/10.1016/j.jbi.2014.07.018
1532-0464/

Decision support directly from univariate relations is limited, as the effects of interactions between variables are not taken into account. For example, evidence about individual effects of a treatment and a comorbidity factor can be analysed in separate meta-analyses. However, if the treatment and comorbidity factor interact with each other, their joint effect may be completely different from their individual effects. As a result, decision support provided by the meta-analysis of individual effects may be invalid for a patient who is exposed to both the treatment and the comorbidity factor (see [2,10,11] for a more detailed discussion of generalising clinical evidence).

To improve this situation, we propose a method of combining the results of meta-analyses, clinical knowledge and data to provide decision support for complex decision problems where the data is scarce. Our method combines 'univariate' meta-analysis following a systematic review, with a small 'multivariate' dataset and expert knowledge. Bayesian networks (BN) offer a powerful framework to combine evidence from different sources [1,5,12,13]. Our methodology integrates the evidence from a meta-analysis into BN development by using it first to identify the BN structure and then to help determine the BN parameters; this second step uses auxiliary parameter learning models similar in some ways to techniques that can be used for meta-analysis. We illustrate the application and results of this method with a clinical case study into the prediction of the outcomes of severely injured lower extremities.

In the remainder of this paper, Section 2 recaps of a Bayesian meta-analysis technique to combine probabilities. Section 3 describes our methodology for developing a BN based on the results of a meta-analysis. Sections 4–6 present the case-study, results and conclusions respectively.

2. Meta-analysis of probabilities

The method we propose in Section 3 assumes a possibly small multivariate patient dataset is available together with univariate results of a meta-analysis of probabilities. In this section, we give a recap of the meta-analysis of probabilities by briefly presenting an existing Bayesian technique [14,15]. The results obtained from this meta-analysis technique can be used in the method of Section 3, though other techniques could also be used. The recap also serves to introduce hierarchical Bayesian models, which are also used in Section 3.

Meta-analysis is an important form of clinical evidence as it combines and summarises the relevant published evidence that is identified by a systematic literature review. Meta-analysis can be used to combine different types of statistics including odds ratios, risk ratios and probabilities [14]. We focus on the meta-analysis of probabilities as the parameters of a BN are composed of probabilities. Fig. 1 shows a random-effects Bayesian meta-analysis model that takes the variation between studies into account, and does not assume normality for the distribution of the individual studies.

The binomial distribution is the probability distribution of the number of positive outcomes in n independent experiments where the probability of a positive outcome is p for every experiment. In the meta-analysis model, the result of each individual study i is modelled with the binomial distribution shown below, where r_i is the number of positive outcomes observed in the study i, p_i is the true study probability of the study i, and n_i is the sample size of the study i.

$$r_i \sim Binomial(p_i, n_i)$$

The normal distribution is a convenient way of modelling the pooled estimate and the variation between studies. We use an inverse logit transformation to model the true study probability

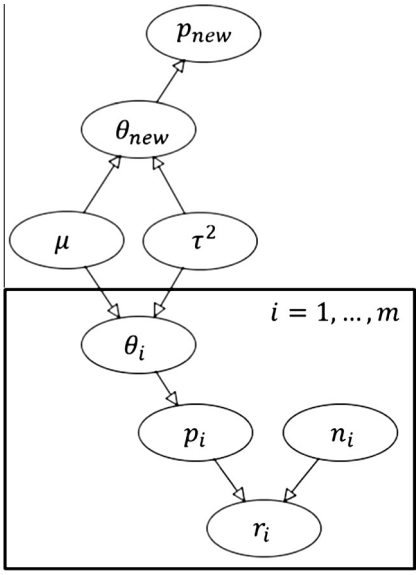

Fig. 1. Bayesian meta-analysis model for combining probabilities.

p_i with the normal distribution. The mean μ of this distribution represents the transformed pooled estimate, and the variance τ^2 represents the variation between studies.

$$logit(p_i) = \theta_i$$

$$\theta_i \sim Normal(\mu, \tau^2)$$

The predictive probability distribution can also be calculated by using an inverse logit transformation of this normal distribution. The predictive distribution is a recommended way of presenting the results of a meta-analysis as it represents the uncertainty from both the pooled estimate and the variation between studies (see [14] and chapter 8 of [16] for more detailed information on predictive distributions in meta-analysis).

$$\theta_{new} \sim Normal(\mu, \tau^2)$$

$$logit(p_{new}) = \theta_{new}$$

Finally, priors must be selected for the pooled estimate and between-study standard deviation. The non-informative priors shown below can be used if informative priors are not available.

$$\mu \sim Normal(0, 1000)$$

$$\tau \sim Uniform(0, 2)$$

In order to calculate the posteriors of μ, τ^2 and p_{new}, we enter the observed number of positive outcomes r_i and sample sizes n_i from each reviewed study. The posteriors can be calculated by using the dynamic discretisation algorithm [17] in AgenaRisk [18] or the Markov Chain Monte Carlo (MCMC) sampling technique in OpenBUGS [19].

3. Building BNs based on meta-analysis

The previous section described a Bayesian meta-analysis technique for pooling probabilities. In this section, we present a methodology that uses data, expert knowledge and the pooled probabilities from a meta-analysis to define the structure (Section 3.1) and parameters (Section 3.2) of a BN decision support model. Our methodology assumes that expert knowledge, a meta-analysis of univariate relations from a relevant systematic