Journal of Biomedical Informatics 48 (2014) 84–93

Contents lists available at ScienceDirect

Journal of Biomedical Informatics

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

Decision support from local data: Creating adaptive order menus from past clinician behavior

Jeffrey G. Klann [a,b,e,*,1], Peter Szolovits [c], Stephen M. Downs [d,e], Gunther Schadow [e,2]

[a] Laboratory of Computer Science, Massachusetts General Hospital, One Constitution Center, Suite 200, Boston, MA 02129, United States
[b] Harvard Medical School, 25 Shattuck St, Boston, MA 02115, United States
[c] Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Stata Center, 32 Vassar St, 32-254, Cambridge, MA 02139, United States
[d] Children's Health Services Research, Indiana University School of Medicine, 410 W. 10th St, Suite 1000, Indianapolis, IN 46202, United States
[e] The Regenstrief Institute for Health Care, 410 W. 10th St, Suite 2000, Indianapolis, IN 46202, United States

ARTICLE INFO

Article history:
Received 7 October 2013
Accepted 7 December 2013
Available online 16 December 2013

Keywords:
Clinical Decision Support
Data mining
Bayesian analysis

ABSTRACT

Objective: Reducing care variability through guidelines has significantly benefited patients. Nonetheless, guideline-based Clinical Decision Support (CDS) systems are not widely implemented or used, are frequently out-of-date, and cannot address complex care for which guidelines do not exist. Here, we develop and evaluate a complementary approach – using Bayesian Network (BN) learning to generate adaptive, context-specific treatment menus based on local order-entry data. These menus can be used as a draft for expert review, in order to minimize development time for local decision support content. This is in keeping with the vision outlined in the US Health Information Technology Strategic Plan, which describes a healthcare system that learns from itself.

Materials and methods: We used the Greedy Equivalence Search algorithm to learn four 50-node domain-specific BNs from 11,344 encounters: abdominal pain in the emergency department, inpatient pregnancy, hypertension in the Urgent Visit Clinic, and altered mental state in the intensive care unit. We developed a system to produce situation-specific, rank-ordered treatment menus from these networks. We evaluated this system with a hospital-simulation methodology and computed Area Under the Receiver–Operator Curve (AUC) and average menu position at time of selection. We also compared this system with a similar association-rule-mining approach.

Results: A short order menu on average contained the next order (weighted average length 3.91–5.83 items). Overall predictive ability was good: average AUC above 0.9 for 25% of order types and overall average AUC .714–.844 (depending on domain). However, AUC had high variance (.50–.99). Higher AUC correlated with tighter clusters and more connections in the graphs, indicating importance of appropriate contextual data. Comparison with an Association Rule Mining approach showed similar performance for only the most common orders with dramatic divergence as orders are less frequent.

Discussion and conclusion: This study demonstrates that local clinical knowledge can be extracted from treatment data for decision support. This approach is appealing because: it reflects local standards; it uses data already being captured; and it produces human-readable treatment-diagnosis networks that could be curated by a human expert to reduce workload in developing localized CDS content. The BN methodology captured transitive associations and co-varying relationships, which existing approaches do not. It also performs better as orders become less frequent and require more context. This system is a step forward in harnessing local, empirical data to enhance decision support.

Abbreviations: BN, Bayesian Network; ARM, Association Rule Mining; CPT, Conditional Probability Table; GES, Greedy Equivalence Search; ITS, Iterative Treatment Suggestion (the methodology defined in this manuscript); UVC, Urgent Visit Clinic.

* Corresponding author at: Laboratory of Computer Science, Massachusetts General Hospital, One Constitution Center, Suite 200, Boston, MA 02129, United States. Tel.: +1 617 643 5879; fax: +1 617 643 5280.

E-mail address: Jeff.Klann@mgh.harvard.edu (J.G. Klann).

[1] Dr. Klann is no longer affiliated with 'The Regenstrief Institute for Health Care'.

[2] Present address: Pragmatic Data LLC, 8839 Rexford Rd., Indianapolis, IN 46260, United States.

1532-0464/$ - see front matter
http://dx.doi.org/10.1016/j.jbi.2013.12.005

1. Introduction

A currently popular approach to improving the quality of health care is to make sure that similar cases are handled in similar ways, i.e., to reduce the variability of care [1]. Frequently this is accomplished through propagation of external protocols into practice, through mechanisms such as Clinical Decision Support (CDS) [2].

Unfortunately, computable CDS content is extremely expensive and time-consuming to create [3], maintain [4], and localize [5]. Consequently CDS has been much more slowly adopted than other components of Health Information Technology (HIT) [6]. Even when CDS available, the content is frequently inappropriate or incorrect [7]. Various projects are being undertaken to standardize computable CDS content in order to reduce the local implementer's work (e.g., [8]).

Still, standardized CDS does not address the following issues: the frequency of content change in medicine, physician attitudes toward guidelines, and terminology challenges. First, much content, both routine and complex, is not distilled into guidelines [9]. This might be quite common; in one study, the literature provided answers to primary care providers' routine clinical questions only 56% of the time [10]. Second, studies have shown that physicians value colleagues' advice at least as much as guidelines [11]. This might be because medicine is locally situated, and colleagues can provide a local frame of reference through which to decide if and how external guidelines relate to particular local cases [12]. Third, standardized content databases require translation of codes into standard terminologies, which is difficult and frequently causes failures in interoperability.

Electronic Medical Record (EMR) data is rapidly proliferating [13], in part due to the Meaningful Use incentive program [14]. These data offer the opportunity to harness local physician wisdom – how care is actually delivered – to augment and suggest protocols, vastly decreasing human effort in developing CDS content and making knowledge available in complex scenarios. It is possible to partially reconstruct physician decisions by aggregating the millions of treatment events in medical record systems. Such locally generated CDS content avoids the three issues discussed above. This fits into the Office of the National Coordinator for HIT's strategic plan, which centers on building a "learning healthcare system" that can perform dynamic analysis of existing healthcare data to glean various information, including best practices [15].

1.1. The wisdom of the crowd

Despite the incompleteness of guidelines and poor maintenance of expert-curated CDS, individual physician behavior is not reliable either. Studies show that care continues to be widely variable and that physicians' treatment does not align well with guidelines [16]. Therefore we suggest two important goals in the design of a CDS tool based on local wisdom.

First, the *average* behavior of many physicians is usually much better than any *individual* physician. Condorcet's jury theorem, upon which voting theory is grounded, proves that when each member in a group of independent decision makers is more than 50% likely to make the correct decision, averaging those decisions ultimately leads to the right answer [17]. If we believe that a physician is more likely than chance to make the correct decision, we can trust the averaged decision. The theorem does have two important caveats. First, it is only guaranteed to apply to binary choices (plus an unlimited number of irrelevant alternatives) [18]. Thankfully, many high-level medical decisions are of this type (e.g., "do I anticoagulate this patient or not?"). Second, crowd wisdom can become crowd madness when decision-makers are not truly independent but are influenced by some outside entity [19]. And of course, practitioners are influenced by colleagues, formularies, available equipment, local culture, etc. The Dartmouth Atlas project has found that the quality of care in a region is profoundly influenced by the 'ecology' of healthcare in that region, including resources and capacity, social norms, and the payment environment [20].

This leads to our second design requirement. Even when averaging decisions, it is impossible to guarantee that results are not influenced by these caveats. Therefore we do not seek to *replace* manual content development with automatically generated CDS content. Instead, our goal is to *complement* content development with knowledge distilled from EMR data. To this end, it was important to choose a data mining approach which produces output that a human expert could understand and update before inserting it into a clinical system.

1.2. Mining EMR data

A handful of studies have explored methods to abstract treatment decisions captured in EMR data into knowledge bases [21–25] or to find knowledge on-demand [26]. The majority of work in abstracting EMR data have used variations of Amazon.com's pairwise Association Rule Mining (ARM) algorithm [27], which has shown good results when capturing global linkages where little variability exists (e.g., drugs used for HIV treatment) [28]. However, researchers have struggled with both transitive associations and the long, static lists of associations that do not take context into account. In one case, the results of such an approach required a great deal of manual editing before incorporation into a decision support system [29]. Other studies have used this approach only as a rudimentary starting point for content developers. For example, the condition-treatment linkages in the National Drug File Reference Terminology (NDF-RT) were 'jumpstarted' by this approach [24].

Bayesian Networks (BNs) are an appealing alternative for mining wisdom from EMR data. BNs are a powerful multivariate, probabilistic reasoning paradigm that naturally model interactions among associations. BNs have a two-phase lifecycle. First, they are constructed, either by hand – which has been widespread in medical informatics research (see e.g., [30]) – or more recently from databases of observational data [31]. Such 'structure learning algorithms', as they are called, take into account transitive associations and co-varying relationships that pairwise rule mining cannot. Therefore, BN structure learning might be able to make sense out of the tangled correlations in clinical data that have hampered other approaches. The second phase of the BN lifecycle is its use – rather than being static networks or rules, BNs enable rapid, iterative exploration of decisions as context evolves.

In a previous study, we piloted a BN approach to produce static order menus for complications of inpatient pregnancy [32]. Our results were very promising, but our scenarios were fixed, they only explored one small domain of medicine, and they relied on the opinion of a single nurse practitioner to evaluate our results. In this study, we more fully flesh out our previous work to use BNs to learn the typical successions of orders made by clinicians for a variety of types of cases. Next, we build a recommendation system that responds adaptively to suggest the most common next orders based on what has been ordered and diagnosed previously. Third, we evaluate this system on hospitalization order-entry data in a multitude of scenarios across four domains. Finally, we undertake a brief comparison of this dynamic approach to a static ARM-like approach.

1.3. Objective

Our goal was to develop a methodology to produce adaptive, patient-tailored, situation-specific treatment advice from order-entry