International Journal of Medical Informatics 102 (2017) 71–79

Contents lists available at ScienceDirect

International Journal of Medical Informatics

journal homepage: www.ijmijournal.com

Decaying relevance of clinical data towards future decisions in data-driven inpatient clinical order sets

Jonathan H. Chen[a,*], Muthuraman Alagappan[d], Mary K. Goldstein[b,c], Steven M. Asch[a,f], Russ B. Altman[a,e]

[a] Department of Medicine, Stanford University, Stanford, CA, USA
[b] Geriatrics Research Education and Clinical Center, Veteran Affairs Palo Alto Health Care System, Palo Alto, CA, USA
[c] Primary Care and Outcomes Research (PCOR), Stanford University, Stanford, CA, USA
[d] Internal Medicine Residency Program, Beth Israel Deaconess Medical Center, Boston, MA USA
[e] Departments of Bioengineering and Genetics, Stanford University, Stanford, CA, USA
[f] Center for Innovation to Implementation (Ci2i), Veteran Affairs Palo Alto Health Care System, Palo Alto, CA, USA

ARTICLE INFO

Article history:
Received 29 September 2016
Received in revised form 1 January 2017
Accepted 12 March 2017

Keywords:
Electronic health records
Data mining
Collaborative filtering
Practice variability
Prediction models

ABSTRACT

Objective: Determine how varying longitudinal historical training data can impact prediction of future clinical decisions. Estimate the "decay rate" of clinical data source relevance.
Materials and methods: We trained a clinical order recommender system, analogous to Netflix or Amazon's "Customers who bought A also bought B..." product recommenders, based on a tertiary academic hospital's structured electronic health record data. We used this system to predict future (2013) admission orders based on different subsets of historical training data (2009 through 2012), relative to existing human-authored order sets.
Results: Predicting future (2013) inpatient orders is more accurate with models trained on just one month of recent (2012) data than with 12 months of older (2009) data (ROC AUC 0.91 vs. 0.88, precision 27% vs. 22%, recall 52% vs. 43%, all $P < 10^{-10}$). Algorithmically learned models from even the older (2009) data was still more effective than existing human-authored order sets (ROC AUC 0.81, precision 16% recall 35%). Training with more longitudinal data (2009–2012) was no better than using only the most recent (2012) data, unless applying a decaying weighting scheme with a "half-life" of data relevance about 4 months.
Discussion: Clinical practice patterns (automatically) learned from electronic health record data can vary substantially across years. Gold standards for clinical decision support are elusive moving targets, reinforcing the need for automated methods that can adapt to evolving information.
Conclusions and relevancm: Prioritizing small amounts of *recent* data is more effective than using larger amounts of older data towards future clinical predictions.

1. Introduction

1.1. Background and significance

Variability and uncertainty in medical practice compromise quality of care and cost efficiency, with overall compliance with evidence-based guidelines ranging from 20 to 80% [1]. Even after current reforms [2], evidence-based medicine from randomized controlled trials cannot keep pace with the perpetually expanding breadth of clinical questions, with only ~11% of guideline recommendations backed by high quality evidence [3]. Clinicians are left to synthesize vast streams of information for each individual patient in the context of a medical knowledge base that is both incomplete and yet progressively expanding beyond the cognitive capacity of any individual [4,5]. The practice of medicine is thus routinely driven by individual expert opinion and anecdotal experience.

Clinical decision support (CDS) seeks to reinforce best-practices by distributing knowledge-based content through order sets, alerts, templates, and prognosis scoring systems [6–10]. Here we pay special attention to clinical orders (e.g., labs, imaging, medications) as the concrete manifestation of point-of-care decision making. Computerized provider order entry (CPOE) [11] typically occurs on an "a la carte" basis where clinicians search for and enter orders to trigger subsequent clinical actions (e.g., pharmacy dispensing and nurse administration of a medication, or phlebotomy collection and

* Corresponding author at: 1265 Welch Road, Stanford, CA 94305, USA.
E-mail address: jonc101@stanford.edu (J.H. Chen).

http://dx.doi.org/10.1016/j.ijmedinf.2017.03.006

laboratory analysis of blood tests). Because clinician memory and intuition can be error-prone, health systems produce order set templates as a common mechanism to distribute standard practices and knowledge (in paper and electronic forms) as the current standard for executable clinical decision support. Clinicians can search by keyword for common scenarios (e.g.; "pneumonia"); and hope they find a preconstructed order set that includes relevant orders (e.g.; blood cultures; antibiotics; chest X-rays) [12–14].

While existing approaches to clinical decision support can already reinforce consistency with best-practices [6,7,15–18], production of this content is limited in scale by the human-expert, knowledge-based authoring necessary for each intervention [19]. If medical knowledge were static, such manual approaches might eventually converge towards a comprehensive set of effective clinical decision support content from the top-down. The reality is instead a perpetually evolving practice of medicine that responds to new evidence, technology, epidemiology, and culture that requires ongoing content maintenance to adapt to changing clinical practices [20–22].

The meaningful use era of electronic health records (EHR) [23] creates an opportunity for data-driven clinical decision support (CDS) to reduce detrimental practice variability with the collective expertise of many practitioners in a learning health system [24–28]. Specifically, one of the "grand challenges" in clinical decision support is data-mining content from the bottom-up in clinical data sources [29]. Such algorithmic approaches to clinical information retrieval could greatly expand the scope of medicine addressed with effective decision support, and automatically adapt to an ongoing stream of evolving practice data. This would fulfill the vision of a health system that continuously learns from real-world practices and translates them into usable information for implementation back at the point-of-care. Prior research into data-mining for decision support content includes association rules, Bayesian networks, and unsupervised clustering of clinical orders and diagnoses [30–37]. In our own prior work, inspired by analogous information retrieval problems in collaborative filtering and market basket analysis, we produced a clinical order recommender system [38,39] analogous to Netflix or Amazon.com's "Customer's who bought A also bought B" system [40].

Accumulating data in EHRs makes these concepts possible, but the dynamic nature of clinical practices over time challenges the presumption that learning from historical clinical data will inform current and future clinical practices. Prior work already demonstrates the importance of temporal patterns between clinical events towards outcome predictions [39,41–43]. Another important relationship is the separation between when data is generated relative to the time learned prediction models are applied and evaluated. Prior clinical prediction modules from mortality risk scores like APACHE and SAPS [44] to hospital readmissions models that risk adjust quality indicators [45] to modern systems based on electronic medical record data [10,46,47] all tend to evaluate their utility by assessing prediction accuracy on a (randomly) separated validation subset of the same data source. This is not representative of a realistic applied scenario where we must make recommendations and predictions towards *future* events that have not yet occurred [48].

1.2. Objective

To determine how varying longitudinal historical training data usage can impact prediction of future clinical decisions. Determine which inpatient admission diagnoses exhibit the most stability vs. variability of clinical practice patterns over time. Estimate the "decay rate" of the relevance of clinical data sources for informing future predictions.

Table 1
Example non-zero counts per ICD9 admission diagnosis from 2008 to 2014. Noise (extra counts) have been added to avoid potentially identifying data bins with counts <10. Detailed five digit ICD9 codes often lead to sparse elements, such as only a handful of admissions coded as 787.24. To compress the hierarchy, instances of five digit codes (e.g., 787.24) were also counted as the respective four digit code (e.g., 787.2), which were in turn also counted as the three digit code (e.g., 787). Thus, the aggregated 787.2 admission diagnosis code accounts for direct codes for 787.2, as well as all instances of 787.2x sub-codes. Likewise, the 787 admission diagnosis code accounts for all 787.x and 787.xx sub-codes.

Raw Count	Aggregate Count	ICD9	Description
0	1934	787	Symptoms involving digestive system
0	1111	787.0	Nausea and vomiting
872	872	787.01	Nausea with vomiting
100	100	787.02	Nausea alone
125	125	787.03	Vomiting alone
14	14	787.04	Bilious emesis
0	259	787.2	Dysphagia
215	215	787.20	Dysphagia, unspecified
13	13	787.22	Dysphagia, oropharyngeal phase
11	11	787.24	Dysphagia, pharyngoesophageal phase
20	20	787.29	Other dysphagia
83	83	787.3	Flatulence, eructation, and gas pain
3	17	787.6	Incontinence of feces
14	14	787.60	Full incontinence of feces
0	464	787.9	Other symptoms involving digestive system
450	450	787.91	Diarrhea
14	14	787.99	Other symptoms involving digestive system

2. Materials and methods

2.1. Collaborative filtering for clinical order decision making

We extracted deidentified patient data from the (Epic) electronic medical record for all inpatient hospitalizations at Stanford University Hospital via the STRIDE clinical data warehouse [49]. The structured data covers patient encounters from their initial (emergency room) presentation until hospital discharge. With five years of data spanning 2008–2014, the dataset includes >74 K patients with >55 M instances of >45 K distinct clinical items. The clinical item elements include >10,000 medication, >1600 laboratory, >1200 imaging, and >1000 nursing orders. Non-order items include >7000 lab results, >7800 problem list entries, >5300 admission diagnosis ICD9 codes, and patient demographics. Medication data was normalized with RxNorm mappings[50] down to active ingredients and routes of administration. Numerical lab results were binned into categories based on "abnormal" flags established by the clinical laboratory, or being outside two standard deviations from the population mean. We aggregated ICD9 codes up to the three digit hierarchy as in Table 1. This helps compress the sparsity of diagnosis categories, while retaining the original detailed codes if they are sufficiently prevalent to be useful. The above preprocessing models each patient as a timeline of clinical item event instances, with each instance mapping a clinical item to a patient at a discrete time point.

With the clinical item instances following the "80/20 rule" of a power law distribution [51], most item types may be ignored with minimal information loss. In this case, ignoring rare clinical items with <256 instances reduces the effective item count from >45 K to ~4.6 K (10%), while still capturing 54.5 M (98%) of the 55.4 M item instances. After excluding common process orders (e.g., vital signs, notify MD, regular diet, transport patient, as well as most nursing and all PRN medications), 2030 clinical orders remain.

Using our previously described method [38,39,52], we algorithmically mined temporal association rules for clinical item pairs from past clinician behavior. Based on Amazon's product recommender [40], we collected patient counts for all clinical item