Data Mining Methodologies to Evaluate the Efficacy of Evidence-Based Practice Guidelines

Andrew Hangsleben1, Katherine Hauwiller1, Pranjul Yadav1, Sanjoy Dey1, Lisiane Pruinelli2, Connie Delaney2,3, Vipin Kumar1, Gyorgy Simon3, Bonnie Westra2,3, Michael Steinbach1

Motivation

The use of multidisciplinary scientific evidence based practice (EBP) guidelines during hospitalization can assist low income and minority populations to regain and maintain health, thus reducing complications from diseases and unnecessary services such as emergency department (ED) visits or re-hospitalization. However, EBP guidelines may not be equally effective across all populations. The study aims to analyze the effectiveness of multidisciplinary EBP interventions during hospitalization and follow-up and identify new patient derived evidence.

Analysis Plan

The analysis plan was created and is being tested using the diabetes data. It will subsequently be applied to sepsis data.

How effective are the guidelines?

<table>
<thead>
<tr>
<th>Patient 1:</th>
<th>Patient 2:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hypertension</td>
<td>Hypertension</td>
</tr>
<tr>
<td>High LDL</td>
<td>High LDL</td>
</tr>
<tr>
<td>Hospitalized with</td>
<td>Hospitalized with</td>
</tr>
<tr>
<td>Cardiac Problem</td>
<td>Cardiac Problem</td>
</tr>
</tbody>
</table>

Test Results (HDL, LDL, etc.)

Time since Hospitalization

Goal: Analyze trajectory differences to evaluate effectiveness of guidelines

We are first evaluating the guidelines for patients with similar backgrounds and conditions. Then we are applying the same analysis to compare populations with and without health disparities to discover any changes in guideline effectiveness.

Identifying New Evidence-Based Guideline Elements

- **Electronic Health Record (EHR) Challenges**
  - Length of observation period
  - Determine methods to manage missing data
    - Encounters update only a small fraction of the variables
  - Missing information has information
  - Seemingly irrelevant information may be relevant
  - New and changing diagnoses
  - Fragmentation

**Accomplishments**

Mining Interpretable and Predictive Diagnosis Codes from Multi-source Electronic Health Records

We developed a sparse, hierarchical canonical correlation analysis (SHHCCA) technique to find groups of ICD-9 diagnosis codes from EHRs that can predict the improvement of urinary incontinence of home health care (HHC) patients and are also interpretable to domain experts:

1. Incorporates prior information available from clinical domain knowledge using the clinical classification system (CCS)
2. Incorporates prior clinical information for the same patients, such as demographic, behavioral, physiological, and psycho-social variables
3. Hybrid framework that can combine both prior information and data-driven clinical information in a predictive model framework

Mining Patterns Associated with Mobility Outcomes in Home Health Care

Analysis was conducted within each subgroup to

1. Identify patient and support system characteristics associated with improvement or no improvement in mobility
2. Evaluate the consistency of these variables across subgroups
3. Find groups of variables that provide a higher level understanding of factors related to improvement of no improvement in mobility

Discriminative pattern mining study for knowledge discovery was conducted with electronic health record (EHR) data documented by home care clinicians.

Fused Sequential Logistic Models for Censored Data

We developed a temporally fused sequence of logistic regression models, which can handle the censored nature of the data. A sequence of logistic regression models allows for the intercept and the coefficients to be represented as functions (of time), and the fused lasso penalty we impose on each coefficient function promotes smoothness over time. Moreover, we will show that the coefficient functions are smooth and the temporal information they encode is relevant and meaningful.

Progression and Risk Assessment of Comorbid Conditions in Type 2 Diabetes Mellitus

Analyses extending previous association rule mining to longitudinal data can

1. Extract progression patterns of diabetes-related complications from a large longitudinal EHR data set, quantifying the risk of adverse outcomes, and taking the interactions between risk factors and comorbid conditions into account
2. Help to monitor patients’ transitions (along with the risk) from one set of comorbid conditions to another over time, and holds the promise for informing interventions and treatment strategies

Analysis

- **Motivation**
  - Because of the variations of predictors among subgroups, it was critical to perform subgroups to better understand variables associated with outcomes
  - Within each subgroup, variables that contribute to improvement or lack of improvement were identified and clustered to related variables and provide a higher level of abstraction. See figure below.
  - These approaches have the potential to provide new insights for clinicians to tailor their interventions for outcome improvement.
  - This study confirms the high prevalence of mobility limitations in primarily older HHC patients, their low rate of improvement, and provides new information on the variables associated with improvement or no improvement in mobility.

- **Accomplishments**
  - Hybrid framework enhances clinical interpretability as compared to the baseline model obtained from ICD-9 codes only, with achieving almost the same predictive capability.

- **Analysis**
  - Our methodology outperforms commonly applied techniques and identifies factors predictive of improvement that are very similar to the factors identified by survival models, the gold standard method for analyzing censored data.
  - Unlike Cox survival models, this method developed has the ability to provide virtually unbiased estimates for the probability of improvement.
  - Example: The existence of a surgical wound tends to increase the odds of improvement in urinary incontinence up to 30 days, at which time it stabilizes. A surgical wound that has not yet fully healed, increases the odds of improvement even more than just the existence of a wound. A non-healing surgical wound, on the other hand, tends to substantially decrease the odds of improvement relative to just having a wound. This leads to the following hypothesis: if a surgical wound is the impediment to the improvement in urinary incontinence, then as the wound heals over time, urinary incontinence improves. When the wound does not heal, the odds of improvement in urinary incontinence decrease. Other models cannot provide enough information to capture this effect.

Acknowledgements and Affiliations

This study was supported by NSF grant: NSF IIS-1344135 and by Grant Number 1UL1RR033183 from the National Center for Research Resources (NCRR) of the National Institutes of Health (NIH) to the University of Minnesota Clinical and Translational Science Institute (CTSI).

1 Department of Computer Science and Engineering, 2 School of Nursing, 3 Institute for Health Informatics, University of Minnesota

Contact Information/Project Site: vk.cs.umn.edu/disparities/website