ASSESSING THE QUALITY OF SERVICE USING BIG DATA ANALYTICS WITH APPLICATION TO HEALTHCARE

Authors: Feras A. Batarseh, Eyad Abdel Latif
Presenter: Saifuddin Hitawala
Outline

• Background of Big Data Analytics in Healthcare
• Motivation and Related Work
• Healthcare-Specific Software System: CHESS
• Analytical Studies
• Conclusion
• Discussions
Big Data Analytics in Healthcare

- Healthcare professionals have an easy access to abundant amounts of data
- Multiple data analytics software vendors provide tools specifically tailored to healthcare
- For e.g. IBM: North Dakota’s DHS, SAS: Centre for Health Analytics and Insights
Data Analytics Lifecycle

Background
Medical Expert Systems

- **Expert Systems**: Use heuristic rather than algorithmic approaches

- **MYCIN**: Landmark medical rule-based system (first expert system)

- These systems can act as repositories for human knowledge in the absence of human expertise

- With data analytics, knowledge is saved in form of data and transformed into knowledge on the fly using data mining methods
Motivations

• A study also showed that healthcare costs could be reduced by $300-$450 billion by applying lessons from success stories.

• This paper aims to locate such stories on state-level so that not-so-successful states can learn from these best practices.
Motivations

• Healthcare industry is also constantly changing in terms of data.

• In 2010, the Affordable Care Act (ACA) included a provision that let the Department of Health and Human Services (HHS) release their data to research institutions and the public.

• ACA changed the analytics in healthcare field and its impacts are discussed in the paper.
## Related Work

<table>
<thead>
<tr>
<th>Data Source</th>
<th>Data Generated</th>
</tr>
</thead>
<tbody>
<tr>
<td>EHRs</td>
<td>Clinical documentation, patient history, results reporting, and patient orders.</td>
</tr>
<tr>
<td>LIMS</td>
<td>Laboratory results. Typically interfaced with EHRs.</td>
</tr>
<tr>
<td>Diagnostic or monitoring instruments</td>
<td>Range from images (e.g., magnetic resonance imaging) to numbers (e.g., vital signs) to text report (result interpretation). May or may not be interfaced with EHRs.</td>
</tr>
<tr>
<td>Insurance claims/billing</td>
<td>Information on what was done to the patient during a visit, the cost of those services and the expected payment. The level of service is often determined from data in EHRs.</td>
</tr>
<tr>
<td>Pharmacy</td>
<td>Information on the fulfillment of medication orders. Not typically part of EHRs.</td>
</tr>
<tr>
<td>Human resources and supply chain</td>
<td>Lists of employees and their roles in the institution and the location and utilization of medical supplies. Not typically interfaced with EHRs.</td>
</tr>
<tr>
<td>Real-time locating systems</td>
<td>Positions and interactions of assets and people.</td>
</tr>
</tbody>
</table>

Most data sources used in literature
CHESS

- Comprehensive Healthcare Electronic Software System
- Data: Michigan Quality Improvement Consortium
- All the data files reside in Hadoop
- A state-level aggregation is then migrated to a SQL server for analysis
- That data is later used for advanced analytics using Tableau or Excel’s Pivot tables
1. Select a complete and verified two dimensional data set.
2. Enter the full name and address of your Hadoop cluster (You can get Hadoop for free from Apache’s website: hadoop.apache.org).
3. Put your data in excel, or if it exceeds excel’s limits, place it in a comma separated files (unless it’s an EHR file).
4. Use the CHESS tool interface to upload the data by browsing, choosing the file, and selecting “import my data into CHESS”.
5. Once uploading is done, choose your preferred tool for advanced analytics. Note: if the tool is not available in the selection, the user can connect the tool directly to the same Hadoop cluster, and use their tool of choice for statistical analysis.
Analytical Healthcare Experiments

• 3 analytical studies:
  • Descriptive: Describes historical health data from 2005 and after
  • Predictive: Looks at data for the end of 2015 and after
  • Prescriptive: Describes how recent health trends affect states

• Only complete and valid data was used
Health of the US states

- Each person reacts differently to his/her health requirements
- Hypochondriacs: Very anxious and paranoid of getting sick
- Emergency-driven: Only seek healthcare when severely sick
- Optimal: Engagement level referred to as preventive care
First, the **average number of visits per patient** at each state is measured by aggregation.

That is then contrasted with the **health of the states** derived from weights, BMI, and number of patients with diabetes and hypertension.

These two factors help define **Quality of Service**: A combination of states' health info and measures of patients engagement with their own health.
Health of the US states

Analytical Study: Descriptive
Trends in Patients’ Engagement

- Types of factors: Health system factors (HSF) and patient related factors (PRF)
- Previous study was based on PRF. Next two studies are based on HSF
- CHESS is used to create predictions for the number of visits for the US states based on the historical data
- Regression forecasting is done and linear forecasts are shown
Trends in Patients’ Engagement

Analytical Study: Predictive
Post-ACA Era

- Previously, Utah and Alabama were in the optimal area
- After introducing ACA, both states moved in very different directions in terms of health and patients’ engagement
- Also, Connecticut is doing much better than what was predicted
- Besides these 4 states, no obvious shifts were noticed in the data

<table>
<thead>
<tr>
<th>State</th>
<th>Smoking Rank</th>
<th>Obesity Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>Utah</td>
<td>1 (best)</td>
<td>4</td>
</tr>
<tr>
<td>Connecticut</td>
<td>4</td>
<td>8</td>
</tr>
<tr>
<td>Kentucky</td>
<td>49</td>
<td>46</td>
</tr>
<tr>
<td>Alabama</td>
<td>38</td>
<td>43</td>
</tr>
</tbody>
</table>
## Analysis and Corrections

### Analytical Study: Prescriptive

<table>
<thead>
<tr>
<th>State</th>
<th>Healthcare QoS Drivers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Utah</td>
<td>Low prevalence of smoking</td>
</tr>
<tr>
<td>Connecticut</td>
<td>Low prevalence of smoking, low obesity rates</td>
</tr>
<tr>
<td>Alabama</td>
<td>High prevalence of diabetes and abnormal child weights</td>
</tr>
<tr>
<td>Kentucky</td>
<td>High prevalence of smoking, high rate of ER-driven practices</td>
</tr>
</tbody>
</table>

- Utah: Low prevalence of smoking
- Connecticut: Low prevalence of smoking, low obesity rates
- Alabama: High prevalence of diabetes and abnormal child weights
- Kentucky: High prevalence of smoking, high rate of ER-driven practices
Summary

The Goal

<table>
<thead>
<tr>
<th>The Goal</th>
<th>Increase Healthcare “Value”</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Improve Quality and Efficiency of Care Delivery</td>
</tr>
<tr>
<td></td>
<td>Increase Healthcare QoS</td>
</tr>
<tr>
<td></td>
<td>State-level insights into the QoS and the Value of healthcare</td>
</tr>
</tbody>
</table>

Tactics

- Value-Based Purchasing
- Reduce Preventable Rermissions
- Reduce Hospital Acquired Conditions
- Bundled Payments
- Accountable Care Organizations

Prerequisite

Electronic Health Records

Electronic Health Records

Conclusion
References


- The US Department of Health and Human Services: http://www.hhs.gov/

References


• The American Health Rankings – by the United Health Foundation: http://www.americashealthrankings.org/
Strengths and Weaknesses

**Strengths**

- Novel contribution in healthcare analytics at state-level
- Well-designed studies for analysing the data
- A good background and clear motivations behind the studies

**Weaknesses**

- Reasons for the shifts from predictions to actual results not verified
- No solutions for enacting or implementing the analysis results
- Various health and socioeconomic factors not considered

Discussion
Related Papers

  • Uses predictive modeling to capture early life socioeconomic factors

  • Identifies variations in data access policies across federal and state agencies as well as hurdles in obtaining access to publicly funded DBs

Discussion
Future Work

• Perform a similar study on lower level of aggregation such as cities and counties and identify health habits in these areas

• Build advanced clustering models within CHESS such as k-means grouping states in different ways and thus identifying significance of certain healthcare factors

• Tell stories of more states in more detail like it was done for some states in paper like Alabama and Utah.
Questions

• Can this study be applied by gathering relevant data of other countries?

• Is there a way to personalize the CHESS system in order to monitor personal health as well as compare analysis results with the average human?

• Is it possible to have one big repository of health-related data that is both public and private (using access levels) thus furthering advancements in healthcare analytics?

• How much of an impact will big data analytics have on healthcare?