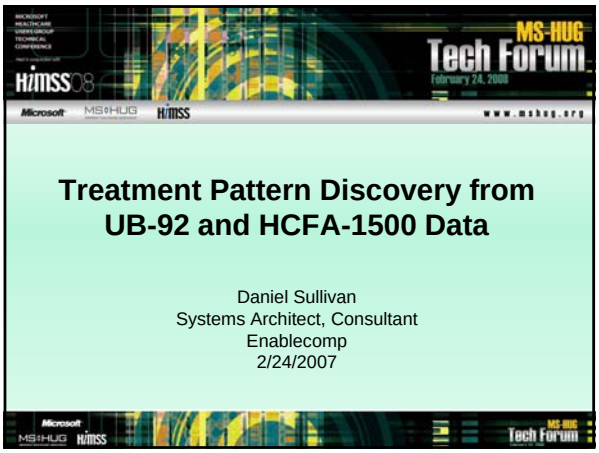


Treatment Pattern Discovery from UB-92 and HCFA-1500 Data

Daniel Sullivan
Systems Architect, Consultant
Enablecomp
2/24/2007

Agenda

- Introduction
  - My background
  - About Enablecomp
  - The work I do at Enablecomp
- What is Data Mining?
- What is our process?
- Why financial analysis of provider behavior?
- Recommendations before you begin a DM Project.
- Bayesian/Belief Networks and HMM (Hidden Markov Model Techniques) The practical example (if time permits) will only discuss the Markov model
- Practical Example "Workers Comp Provider Behavior"
- Questions/Issues/Concerns ***What about privacy??? HIPAA
- www.hhs.gov/ocr/hipaa/

Introduction

- Academic:
  - BA History University of WA (1993)
  - MA History Central WA University (1996)
  - BS in Informatics Indiana University (2003)
- Non-IT/Software Engineering experience as US Army Officer and Process Improvement Facilitator
  - US Army Air Defense Officer 1996-1999
  - Rolls-Royce PLC 2000 – 2001
  - Full Time Student and Consultant (2001-2003)
- Software engineer since 2003, working in Healthcare Finance since 2005.

Introduction (Continued)

- Enablecomp:
  - Primarily Specialized in Workers Compensation Collections
  - Revenue Cycle Management
  - Specialized Health Finance Informatics
  - 0 Balance or Small Bill Automated Payment Recovery
  - We have both Hospital and Physician Clients.
- What I do there:
  - Software Engineer and Architect.
  - Develop/Design Medicare based rules engine.
  - Develop/Design Claims Management System.
  - Datamining

Introduction (Continued)

- Enablecomp Protection:
  - I cannot discuss the following:
    - Any non-public domain intellectual property
    - Actual Projects at Enablecomp
    - Proposed Projects at Enablecomp
    - Healthcare Data which has not be obfuscated/protected
    - Our provider clients

Introduction (Continued)

- Find out more?
  - Visit our website: www.enablecomp.com
  - Phone Number: 615-791-4300
    - Questions about Hospital RCM: ext 19
    - Questions about Physician RCM: ext 21
    - General questions or proposals for Health Finance Informatics Projects: ext 17, or 615-500-7107

What is Data Mining?

- Applied AI?
- Applied statistics?
- ***Derivation of new (USEFUL) information from a large data store?***
- Scientific Method?
- The process we follow is fairly standard and abbreviated in proportion to the size of our organization and the data we manage.

What is our process?

- How is our Data Mining Process Structured
  1. Product Concept Development: we brainstorm about possible product or customer needs we could meet if we understood a particular problem better and determine if we have the data to meet this need or develop the product.
  2. Hypothesis Development: we brainstorm to come up with reasonable statements about the behavior of the system under analysis.
  3. Model Development: we build a model of what we think is happening and define the key business objects and properties as they apply.
  4. Test Development: We shape the data in order to make in amenable to analysis.

What is our process? (Continued)

- How is our Data Mining Process Structured
  5. Model Data and Test Data Selection: We select a portion of our data and set it aside for model generation and another set for experimentation.
  6. Test/Experiment: Upon generating the model we run the test data set through the model.
  7. Evaluate Results with Subject Matter Experts: We then apply a Delphi technique to determine if the system successfully predicted the outcomes we expected. If we are successful, we go to step 8, if not we re-start the process at the appropriate step.
  8. Feasibility analysis: we determine if there is a financial benefit to implementing the model, given the basic costs involved. If not, we go to step 1.
  9. Deployment and Model Management: We deploy the model, based upon the desired use. We manage the model and re-test periodically.

Why analyze the financial side of treatment, if you are a Provider?

- Workers comp: Evidence Based Guidelines.
- Medicare: Pay 4 Performance.
- Contracts and Negotiation with Payers.
- Understanding Payer Denial Behaviors.
- Managing small claims as “claim groups” and negotiating payment based on this.
- Providing general information and models to assist Physicians and other healthcare professionals.

Some Recommendations when starting a DM project:

1. Don’t black-box it, subject matter experts are critical throughout the whole lifecycle.
2. Be Patient, it’s a lot like fishing and you don’t always catch fish your first time out.
3. See it as a Scientific Process, see yourselves as scientists and make sure you try to follow as much of the standard scientific process as possible.
4. Make sure you have enough data, there is no perfect answer as to how much data is enough. But data mining is really only useful when the numbers get large enough to be statistically relevant.
5. Start out simple, begin with algorithms and techniques that are tested and reliable and avoid techniques which have not been tested in other business domains.
6. You need someone who understands the math, you don’t have to be a PhD in order to use data mining tools, but it is extremely helpful to have someone with an understanding of how the tools work, how the algorithms work, what their complexity is and how they have been applied to other business contexts.

Bayesian/Belief Networks and Markov Models

- Algorithmically low complexity (with few exceptions)
- Fit nicely into supervised learning schemes
- Fit nicely into hybrid systems which utilize heuristics and business logic servers (PROLOG)
- Can be evaluated graphically by subject matter experts (like Doctors and Accountants) with very little prep.
- Can generate reasonably stable yet changing models.
- ***A good first step before trying other techniques…***
  - ***ANN, Decision Trees, ART, Genetic Algorithms, etc.***

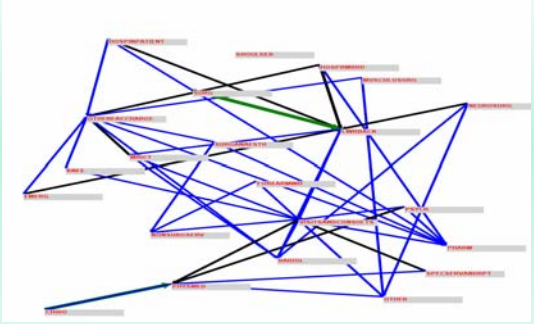
Bayesian/Belief Networks

- Bayesian Networks model conditional relationships between variables in a complex system.
- Bayesian networks are Weighted DAG's (Directed Acyclic Graphs). The weight represents a probability relationship between 2 different variables in the model and their possible states.
- MS Analysis services has a solid implementation of this data mining technique built in.
- http://research.microsoft.com/adapt/MSBNx/msbnx/Basics_of_Bayesian_Inference.htm

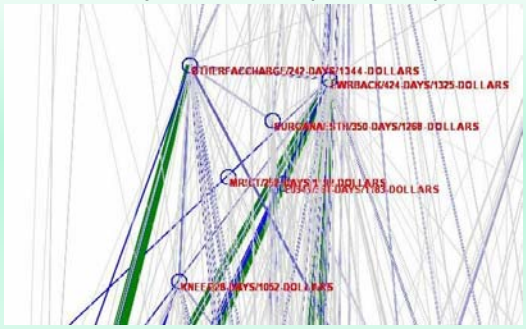
Markov Models

- Similar to Bayesian models, but with some differences:
  - GREAT for modeling hidden business process behaviors.
  - Graphs can have cycles (state machines).
  - Time is a factor - Markov Assumption: Events which occur within the model are conditionally dependent only upon the previous event.
  - Data is shaped into sequences or Markov Chains.
  - So lets look at this as it applies to Healthcare Billing?

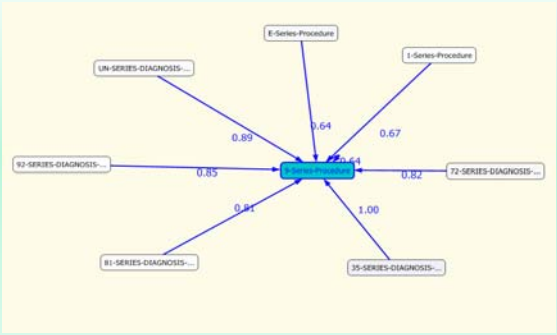
Example Visualization Without Analysis Services



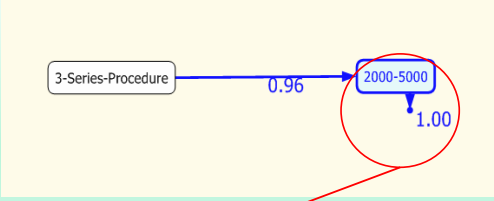
Example Visualization Without Analysis Services (Continued)



Example Visualization W/ Analysis Services

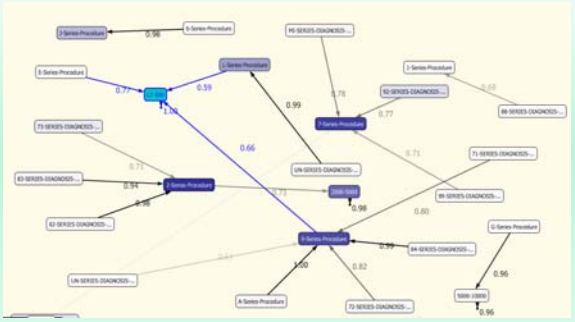


Example Visualization W/ Analysis Services (Continued)



This terminal node is a dollar range for total dollars spent on the given patient for a given treatment sequence.

Example Visualization W/ Analysis Services (Continued)



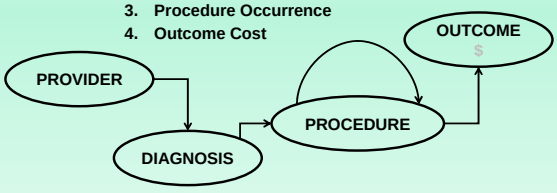
Treatment Patterns

- Data Preparation of UB-92/04 and HCFA-1500 Billing Data
  - Critical Variables and Events
    - Variables
      - ICD9/10 Diagnosis Code Categorization
      - Categorizing Provider Practice Specialty
      - CPT/HCPC Code categorization
      - Discretizing Line Charge Amount
      - Discretizing Total Bill Amount
      - Highest Valued Procedure on the Bill
      - Total Amount spent on Patient (for chain)

Treatment Patterns (Continued)

Critical Variables and Events (Continued)

- MARKOV Event Chain
  1. **Provider Choice**
  2. **First Admit Diagnosis**
  3. **Procedure Occurrence**
  4. **Outcome Cost**



Treatment Patterns (Continued)

Markov Chain in Raw Data:

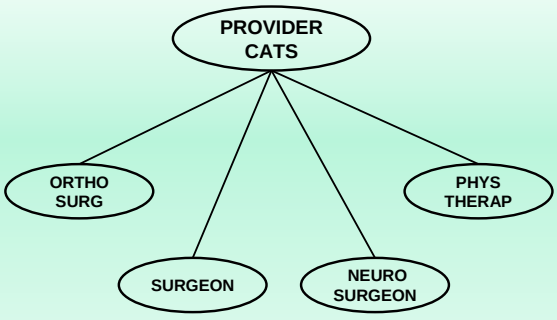
```
7, 29-SERIES-PROCEDURE, 20051201
8, CHARGES-OF-7554, 99999
9, 88-SERIES-DIAGNOSIS-diagnosed-by-Orthopaedic Surgeon, -1
10, 26-SERIES-PROCEDURE, 20070719
11, 26-SERIES-PROCEDURE, 20070816
12, 97-SERIES-PROCEDURE, 20070906
13, 97-SERIES-PROCEDURE, 20070906
14, 97-SERIES-PROCEDURE, 20070906
15, 97-SERIES-PROCEDURE, 20070906
16, 97-SERIES-PROCEDURE, 20070913
17, 97-SERIES-PROCEDURE, 20070913
18, 97-SERIES-PROCEDURE, 20070913
19, 97-SERIES-PROCEDURE, 20070913
20, 97-SERIES-PROCEDURE, 20070918
21, 97-SERIES-PROCEDURE, 20070918
22, 97-SERIES-PROCEDURE, 20070918
23, 97-SERIES-PROCEDURE, 20070920
24, 97-SERIES-PROCEDURE, 20070920
25, 97-SERIES-PROCEDURE, 20070920
26, 97-SERIES-PROCEDURE, 20070920
27, 97-SERIES-PROCEDURE, 20070920
28, CHARGES-OF-6784, 99999
29, 71-SERIES-DIAGNOSIS-diagnosed-by-Orthopaedic Surgeon, -1
```

Treatment Patterns (Continued)

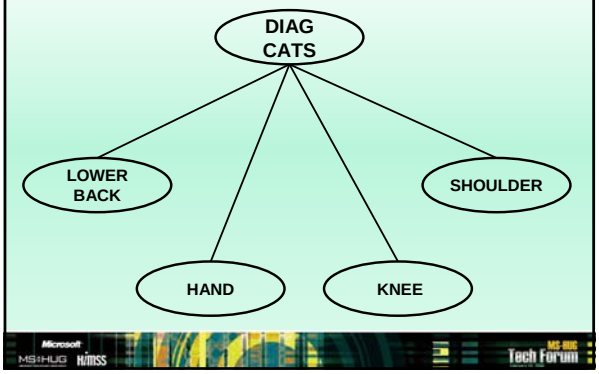
What other kinds of data are available from the bill?

- Place of Service
- Type of Service
- Other demographic (DOB, Gender, Patient Address)
- Employer Info
- Other diagnoses than the primary.

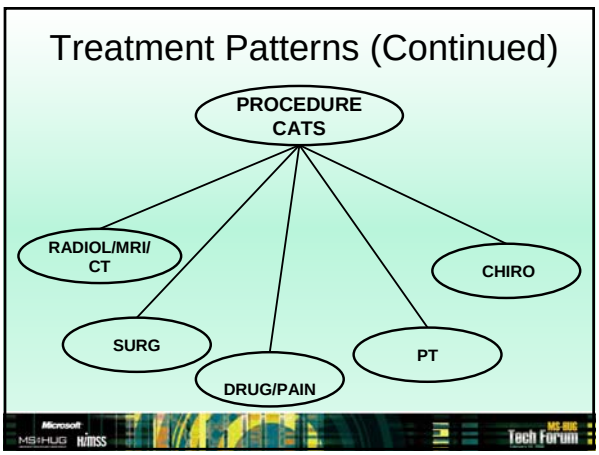
Treatment Patterns (Continued)



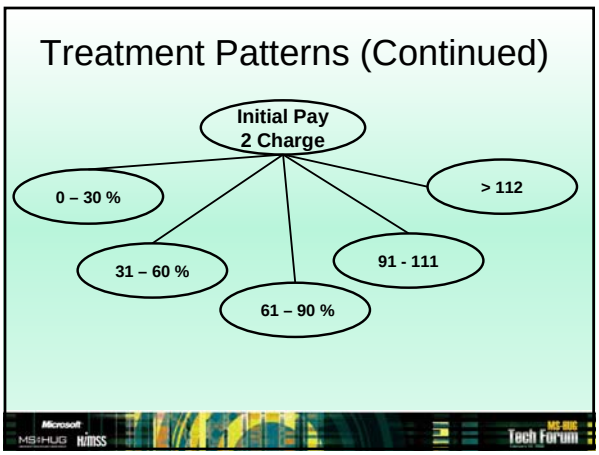
Treatment Patterns (Continued)



Treatment Patterns (Continued)



Treatment Patterns (Continued)



Markov File

Examples:

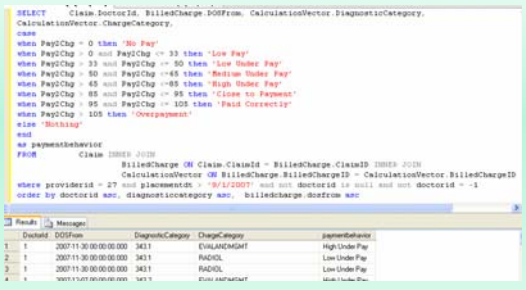
- DATA\HMM_20071203.txt
- results_20080210.txt

Example - Walkthrough

- Walk group through basic example of using analysis services.

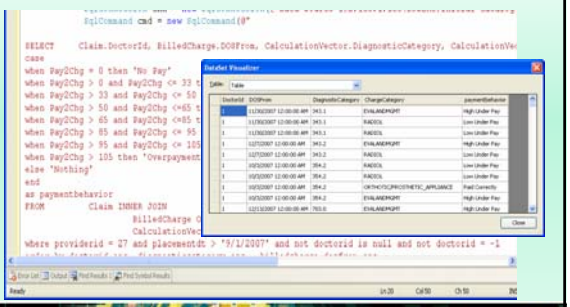
Example (Continued)

- Create Markov Chain Data – Select Data Set



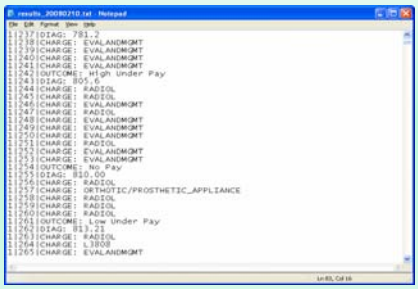
Example (Continued)

- Build Chain – write algorithm, create import file



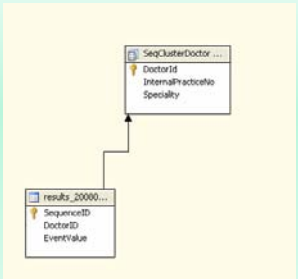
Example (Continued)

- Sequence/Chain File Complete



Example (Continued)

- Create Data Source View



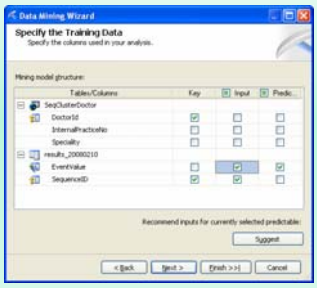
Example (Continued)

- Select and Generate MS Sequence Cluster Model
  - Right click to add new mining structure
  - Select “Sequence Clustering”
  - Step through wizard to generate model
  - For more detail on this, you should read:

  Data Mining with SQL Server 2005 by ZhaoHui Tang and Jaimie MacLennan (Wiley Press)

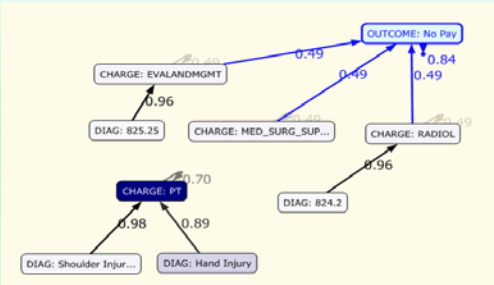
Example (Continued)

- Selection criteria for chain/sequence relation



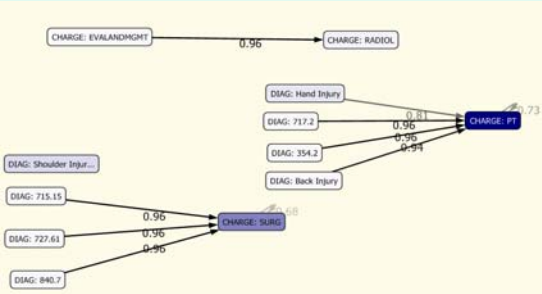
Example (Continued)

- Cluster 2



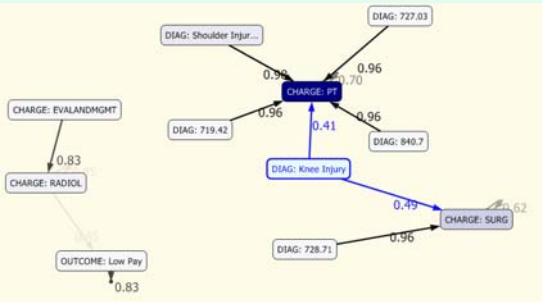
Example (Continued)

- Cluster 4



Example (Continued)

- Cluster 5



Recommended Material for Further Study:

- Data Mining with SQL Server 2005 by ZhaoHui Tang and Jaimie MacLennan (Wiley Press), © 2005
- Data Mining: Concepts and Techniques by Jiawei Han and Micheline Kamber (MK Publishers), © 2006
- Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations by Ian H Witten and Eibe Frank (MK Publishers), © 2000
- Bayesian Artificial Intelligence by Kevin B Korb and Ann E Nicholson (Chapman & Hall), © 2004
- AI Application Programming, 2nd Edition by M. Tim Jones (Charles River Media), © 2005
- Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference by Judea Pearl (MK Publishers), © 1988

Questions/Issues

- Performance and infrastructure requirements.
- HIPAA, Anonymity and Discretization.
- What else?