Using Sequence Transitions to Predict Future Costs

Health Equity Data Set

Introduction

Provided claim information for members.

Predict when a user should add money into their account.

Approach

- How is money taken out of the system?
  - As members receive services, claim contains *when* (ServiceEnd), *what* (CPTCode), and *how much* (PatientResponsibilityAmount).
  - Ordered in a sequence of time.

Approach

In order to determine how much money a person should have in their account, we want to determine what the next procedure could be and how much it might cost.

Procedures can often follow a pattern.

CPT to CCS

CPT codes can be too granular, need a way to group similarities.

CCS groups them from 52,000~ codes to 242.

Denormalized Data

NewMemberID, DependentID, CPTCode, CCSCode, BirthYear, Zip, State, ServiceEnd

Ordered by NewMemberID, DependentID, ServiceEnd in ascending order.

Transition Per Person

Every NewMemberID/DependentID represents a different person.

Transitions also represent features about the person:

- Age group (Under 30, Under 60, Over 60)
- Location (State)

Transition Probabilities

Under60_UT169 -> Under60_UT147 : 0.01357

Probabilities are calculated after counting all the instances.

Emissions

Emissions are the cost of the service.

Calculated in bigram transitions.

Under30_UT231_Under30_UT240 -> 1425.22 : 0.00008321

Training Set and Gold Set

Built a training set with 97% of the records.

3% were used for verification in the gold set, and tested.

Gold Set Prediction Results

Gold Set Prediction Results

Gold Average: 83% (Gold / Expected)

Gold to Expected Stdev: .63 (amount of variation between most of the results - most fell between 20% and 146%)

Expected Amount Query

Expected Amount Query Location

Input: BirthYear, Location, and Previous CPTCode

Output: Expected amount of next cost:

- Top 50% > transitions with their expected amount + standard deviation.

List of Predictions For Members

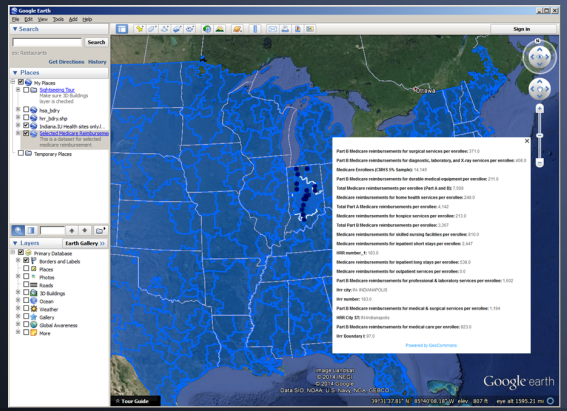
Member Predictions Location

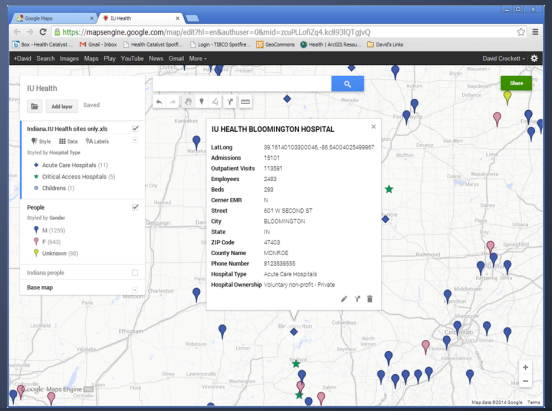
Recommended Balance ($ Amount)

Sufficient Amount (Yes/No)

Future Considerations

Claims data can be mapped per zip code and easily compared within and between two or more health service regions.





Future Considerations

Test against random subsets of transition sequences.

Better involvement of time - prediction is based on next CPTCode but isn’t specific on when that could be.

Future Considerations

Finish all member calculations (time constraint).

Continue experimenting with age/location and even gender groupings.

Test current predictions against future.