

Clustering Analysis of 2012 Medicare Part B

FFS Payments

By: Dave Miller, Xiang Li for MSiA 431

The Obama administration has made an effort to make the healthcare system more transparent, and released a data set of Medicare Part B Fee-For-Service (FFS) payments, which includes services provided to Medicare beneficiaries from (calendar year) 2012. All claims contained in the data set have been resolved. The data set contains provider information and demographics, beneficiary (patient) and service counts, provider type, provider charges, allowable amounts and payments, and provider participation in Medicare. Any aggregated information which is derived from 10 or fewer beneficiaries were excluded from the data set (to protect privacy).

The data set was clustered using an Euclidean distance algorithm in Hadoop. The data set was reduced to contain only the numeric variables (see Table 1). The data set was first normalized using Fig. A python script was then used to call mapreduce multiple times (10 iterations) in order to update the cluster centers (centroids). Four clustering iterations were run, with 3 to 6 clusters.

Table 1: Numeric Variables to Cluster

Variable	Description
line_srvc_cnt	Number of services provided.
bene_unique_cnt	Number of distinct beneficiaries receiving the service.
bene_day_srvc_cnt	Number of distinct beneficiaries/day service.
average_Medicare_allowed_amt	Average of the Medicare allowed amount.
stdev_Medicare_allowed_amt	Standard deviation of the allowed amounts.
average_submitted_chrg_amt	Average of the charges submitted by the provider.
stdev_submitted_chrg_amt	Standard deviation of the average submitted charges.
average_Medicare_payment_amt	Average amount that Medicare paid after coinsurance and deductible have been paid.
stdev_Medicare_payment_amt	Standard deviation of the Medicare paid amount.

An optimal cluster solution was found using 5 clusters. The iteration from 5 to 6 clusters showed no noticeable improvement from the 5 cluster solution. Figure 1 shows the 5 cluster solution.

Figure 1: Optimal 5 Cluster Solution

*Variables are in average amount



Cluster Descriptions

Cluster 1: Boss Balla - This group of providers sees the fewest patients per day and provides the fewest services, but gets paid (and charges) the most by far.

Cluster 2: Costco Doc - This group of providers provides the most services, and sees the 2nd most (to Cluster 5) number of patients, but has the cheapest services.

Cluster 3: Dr. Normal - This group provides a medium amount of services, but usually just one service per beneficiary, and it tends to have middle-of-the-road charges.

Cluster 4: Dr. Greed - This group of providers doesn't see very many patients, and usually it's limited to one service per beneficiary, but they are fairly expensive and charge Medicare a lot.

Cluster 5: Dr. Target - This group sees the most patients and provides a lot of services (usually about 2 per beneficiary), but the services tend to be affordable.

This 5 cluster solution of the Medicare Part B FFS payment data can be a useful tool for the government to check for fraud. The government can profile providers based on what cluster they belong to, and could use this information to determine which (if any) providers may be abusing the system and inflating their charges to Medicare. In addition, this could be a valuable resource to help identify and educate groups of providers in billing practices in the hopes of keeping costs down. Considering the rising healthcare costs and the increase in the Medicare population in the not too distant future (due to the Baby Boomers), this could be critical to Medicare's solvency.

It would be interesting to add the billing level (Medicare's 5 point scale which is supposed to reflect how complex - and thus costly - the visit was) to the data set to see if any new clusters form.