Are Doctors Overcharging you?

Huanye Liu
Uchicago MACSS

Introduction

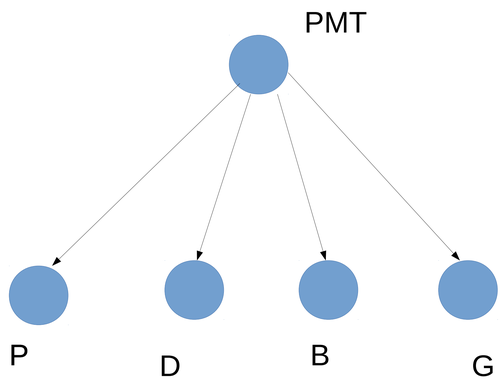
We use graphical model based methods for analyzing medical claims data to detect doctors' overcharging behavior

Data

Source: Medicare Claims Synthetic Public Use Files

Variable	Description
NPI	Doctor's Unique Identification
P	Claim Procedure Code
D	Claim Diagnostic Code
B	Patient's Year of Birth
G	Patient's Gender
PMT	Claim Average Payment

Bayesian Network Model



$$Pr(PMT|P,D,B,G) = \frac{Pr(PMT,P,D,B,G)}{\sum_{PMT} Pr(PMT,P,D,B,G)}$$

$$Pr(PMT) = Pr(PMT)Pr(P|PMT)Pr(D|PMT)Pr(B|PMT)Pr(G|PMT)$$

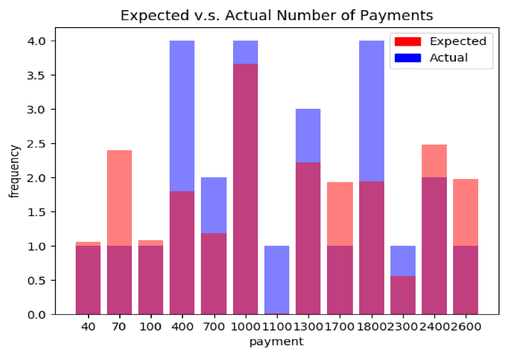
$$Pr(P|PMT) = \frac{N(P,PMT)}{\sum_P N(P,PMT)}$$

$$Pr(D|PMT) = \frac{N(D,PMT)}{\sum_D N(D,PMT)}$$

$$Pr(B|PMT) = \frac{N(B,PMT)}{\sum_P N(B,PMT)}$$

$$Pr(G|PMT) = \frac{N(G,PMT)}{\sum_P N(G,PMT)}$$

Result 2



Dynamic Programming

T × T 2D matrix for DP

The value of each cell is the posterior probability of k maximum payments among all i claim payments

i

k \ i	0	1	2	. . .	T
0	f_{00}	f_{01}	f_{02}	. . .	f_{0T}
1	f_{10}	f_{11}	f_{12}	. . .	f_{1T}
2	f_{20}	f_{21}	f_{22}	. . .	f_{2T}
.	.	.	.	. . .	.
.	.	.	.	. . .	.
T	f_{T0}	f_{T1}	f_{T2}	. . .	f_{TT}

$$Pr(N_{pmt} >= O_{pmt}) = \sum_{k=o_{pmt}}^{T} Pr(N_{pmt} = k)$$

$$f_{ki} = Pr(PMT_T = PMT|y_T) f_{k-1,i-1} + Pr(PMT_T \neq PMT_T|y_T) f_{k,i-1}$$

$$E_{PMT} = \sum_{k=1}^{T} Pr(N_{PMT} = k)k$$

Result 1

Rank of Doctors Who Overcharge Patient

NPI	f_{kT}	N	E[N]
8687459280	0.0000051545	1	0.0000051545
4257804925	0.0000107343	2	0.0238377507
1728991257	0.0000107579	4	0.2830734919
8234450508	0.0000141459	2	0.0078447315
9863658854	0.0000239282	2	0.0121832158
589759662	0.0000300942	1	0.0000300942
2371152333	0.0000324308	1	0.0000324308
5810836174	0.0000327865	3	0.2343714115
7811358993	0.0000344331	2	0.0318564656
6006804581	0.0000363304	1	0.0000363304

Doctors Who Charge Patient with Big N

NPI	f_{kT}	N	E[N]
3204358873	0.0480109606	14	8.9641158024
3733829210	0.7230098962	14	15.6702726184
8461982462	0.4153129262	14	12.9254407334
9943567290	0.3493275149	15	13.3559956746
9127235965	0.4230077793	16	14.9529958135
1308731628	0.4191260768	17	15.9035712276
2484358033	0.0703081096	19	13.5079929412
1586461282	0.865061931	24	28.805738982
2150853459	0.0349447214	24	16.6561038808
3177196045	0.1719397817	26	21.6349473007

Conclusions

The Bayesian Network Model provides a powerful tool

a) to capture the relationship among variables of concern and

b) to translate the target of anomaly detection into computing the aggregate posterior probability of the target variable.