Detecting Excessive Claim Behavior in Medical Insurance Claims

Rob M. Konijn konijn@liacs.nl, r.m.konijn@vu.nl
Arno Knobbe knobbe@liacs.nl
LIACS, Leiden University, Niels Bohrweg 1, 2333 CA Leiden, The Netherlands
VU University Amsterdam, De Boelelaan 1081a, 1081 HV Amsterdam, The Netherlands

Keywords: subgroup discovery, outlier detection, fraud detection

Health care spending is a heavy financial burden that many nations have to face. Rising health care costs are not only caused by aging populations and new medical technologies, but also by unnecessary services, inflated prices, or even fraud. In our research we are interested in finding patterns that correspond to the latter three causes.

When a patient visits a medical practitioner (a dentist, pharmacy, GP, hospital etc.), the practitioner charges an amount of money corresponding to the treatment the patient received. Because the patient generally does not know exactly what service is charged, there is a risk of erroneous claim behavior or even fraud, since the practitioner is the only one who knows the treatment actually performed and charged. Our data, made available by Achmea (the largest Dutch health insurance company), describes claims resulting from treatments provided by several classes of medical practitioners, including dentists, general practitioners (GPs) and pharmacies.

1. Subgroup Discovery

The problem of identifying interesting patterns in claim behavior is essentially an unsupervised learning problem. We have no claims that are labeled as fraudulent beforehand. The data we consider describes patients. A single record summarizes the care a patient received during a certain period (usually one year). The approach we take, is to single out a practitioner and compare its claim behavior against all other practitioners. We assume there is a single practitioner under investigation (the target practitioner). There will be a single target column t with domain $\{0, 1\}$ (or $\{true, false\}$), which identifies whether or not each patient visited this practitioner over a given period of time. Describing differences between target and non-target examples is called Subgroup Discovery (SD).

In (Konijn et al., 2013a) we describe how *local subgroups* of patients can be found. For a local subgroup, patients in the subgroup are much more present at the target practitioner (the target is true more often), while 'similar' patients outside the subgroup are much less frequently present. An example result is found in the case of dentistry, where the subgroup {*Consult, X-Ray picture*} is much more frequently true than the (roughly similar) subgroup {*Consult*}, meaning that the dentist under investigation charges an X-Ray picture next to a consult more often than other dentists.

Additional to counts of patients, also the costs spent on the (treatments of) patients are important. Subgroups with more money involved, are more interesting. We can view the data as having two target columns: one binary target and one continuous target describing costs. In (Konijn et al., 2013b) we investigate possible quality measures that take into account both the binary target as well as the costs-target. We aim to produce interpretable valuation of subgroups, such that data analysts can directly value the findings, and relate these to monetary gains or losses.

Current research is about including prior knowledge in the SD process. In our application, practitioners that mainly treat old patients will have a different claim distribution than practitioners that mainly treat young patients. To still be able to compare practitioners, we will incorporate this knowledge in the SD process.

References

Konijn, R. M., Duivesteijn, W., Kowalczyk, W., & Knobbe, A. (2013a). Discovering local subgroups, with an application to fraud detection. *PAKDD 2013*.

Konijn, R. M., Duivesteijn, W., Meeng, M., & Knobbe, A. (2013b). Cost-based quality measures in subgroup discovery. *New Frontiers in Applied Data Mining - PAKDD 2013 International Workshops.*