Discovering inappropriate billings with local density based outlier detection method

Yin Shan, D. Wayne Murray, Alison Sutinen
Program Review Division, Medicare Australia
134 Reed St. North, Tuggeranong
ACT 2900, Australia
{Yin.Shan, Wayne.Murray, Alison.Sutinen}@medicareaustralia.gov.au

Abstract
This paper presents an application of a local density based outlier detection method in compliance in the context of public health service management. Public health systems have consumed a significant portion of many governments’ expenditure. Thus, it is important to ensure the money is spent appropriately. In this research, we studied the potentials of applying an outlier detection method to medical specialist groups to discover inappropriate billings. The results were validated by specialist compliance history and direct domain expert evaluation. It shows that the local density based outlier detection method significantly outperforms basic benchmarking method and is at least comparable, in term of performance, to a domain knowledge based method. The results suggest that the density based outlier detection method is an effective method of identifying inappropriate billing patterns and therefore is a valuable tool in monitoring medical practitioner billing compliance in the provision of health services.

Keywords: local outlier factor, LOF, health data mining, fraud detection, open source data mining.

1 Introduction

In many countries, public health systems have consumed a more significant portion of governments’ expenditure than ever and this trend is likely to continue as the aging population will unavoidably require more medical resources. Thus, there is an urgent need to make sure that the scarce funding available for the public health service is spent appropriately. This usually involves analysing a large amount of data collected by the public health agencies. Data mining can provide one of the major instruments for exploring health service data. (Becker, Kessler and McClellan, 2005, Lin et. al., 2008, Yang and Hwang, 2006).

One of main concerns of public health agencies is the occurrence of inappropriate practice and fraud. Given the large amount of spending on public health, a small percentage of non-compliance would result in a huge loss of the public money. The US Department of Health and Human Services, estimated that improper Medicare benefit payments made during 2002 financial year totalled $13.3 billion, or about 6.3 percent of the $212.7 billion in processed fee-for-service payments reported by the Centers for Medicare and Medicaid Services (CMS) (US DHHS, 2003). Therefore it is critical to detect these non-compliant activities to facilitate the proper and efficient use of the resources.

In Australia, a government agency, Medicare Australia, administers Medicare, a fee for service national health funding system for Australians. There are over 400 million transactions processed by Medicare Australia and approximately 30 billion dollars benefit paid in 2007-2008 (MA 2008). Medicare Australia is also responsible for undertaking reviews to ensure the integrity of associated health programs it administers. There have been a series of studies that have applied a range of data mining techniques to the Medicare Australia data for various compliance purposes (Pearson, Murray and Mettenmeyer 2005, He, Graco and Yao 1999, Shan et. al, 2008). In this research, we focus on detecting inappropriate billing of one medical speciality group.

In this compliance domain, it often occurs that there is a lack of labelled cases, which makes it difficult to employ supervised machine learning techniques, while outlier detection becomes attractive. This work demonstrates that a local density based outlier detection method (Breunig et.al., 2000) significantly outperforms a few other methods and can be effectively used in the detection of inappropriate practice.

The remaining sections of this paper are organised as follows. The problem domain is briefly introduced in Section 2. Section 3 provides a brief background on outlier detection and density based method. The data set used is described in Section 4, whilst Section 5 outlines the experimental studies applied to the data set. The results and its evaluations numerically and by subject
matter experts are covered in Section 6. Section 7 presents discussions and the last section is the conclusions and future research.

2 Optometrists Billing Compliance

Optometrists claim a small yet significant portion of Medicare benefits. There are approximately 3,000 Optometrist practicing in Australia. In contrast to the group of General Practitioners (GPs) which has over 25,000 practitioners nation-wide, this is a relatively small group in size.

This group is quite unique in terms of the number of Medicare items it can bill. There are only 26 items in the Medicare Benefit Schedule (DHA, 2007) which optometrist can bill whilst there are commonly hundreds of items available to other groups of medical professionals.

Usually, the proportion of these items should be relative stable and we understand there are a small number of factors which may contribute to the variations, such as optometrist’s speciality (e.g. some optometrists specialised in contact lens fitting) and patient age. These items bear different level of compliance risk. There is a possibility that some items are overused resulting in over servicing or some items are inappropriately used to obtain more financial benefit to the optometrist. This research hypothesizes that by detecting outliers in terms of item usage and a few other factors, the individuals with unusual billing practice may be discovered. These individuals may bear higher risk of inappropriate practice as they are significantly different from their peers.

3 Local density based outlier detection method

The concept of outlier detection originally came from the field of statistics (Hawkins, 1980). These methods are usually very well understood and have solid theoretical grounds. However, they often require the prior knowledge of probability distribution and more importantly they are usually univariate. In the domain of data mining we more often than not encounter huge datasets with at least dozens of variables, whose underlining distributions are unknown.

Recently there have been a number of algorithms proposed to address the above issues of the traditional statistical methods. Some of them are based on clustering methods (Ester, et al, 1996, Ng and Han 1994, Zhang et. al 1996). Outliers in those clustering methods are defined as those records, which cannot fit neatly into any clusters and thus need to be identified and treated as exceptions. Some other methods have been specifically designed to detect these outliers (Knorr and Ng 1998, Ramaswamy et al., 2000, Breunig et al 2000).

In this work, a local density based outlier detection method (Breunig et.al., 2000) is used. In this method, one single measure, the Local Outlier Factor (LOF), indicating the degree of outlier-ness is calculated for each record. The records having the largest LOF values are the most significant outliers. There are several reasons to choose this method. Firstly, this method is based on the local property, which is suitable for this problem. A series of K-means clustering analyses were performed on this data set with different number of clusters and there was no clear global cluster structure found, as evident by very small value of the silhouette information (Rousseeuw, 1987). The silhouette information measures how well the points are grouped into clusters. If the value is close to 1, it suggests the points in general clearly belong to certain clusters. The value is close to -1, otherwise. Small silhouette values obtained in our experiments undermines the basis of employing most of the clustering based outlier detection methods, as there is no obvious cluster discovered in this dataset. Secondly, this method has only one parameter to tune and requires minimum prior knowledge, such as probability distribution, which is unknown. Thirdly, this method provides one straightforward rating of the degree of outlier-ness - LOF.

The complete formal definition of LOF can be found in the original paper (Breunig et al., 2000). An accessible and simplified description is presented here for the completeness of the paper.

We assume that for any object p, there are no two objects that are the same distance from that object p. This greatly simplifies the discussion without compromising the basic idea. We need several simple notions before introducing the definition of Local Outlier Factor (LOF). For any positive integer k, the k-distance of object p, denoted as k-distance(p), is the distance of k-th object from p. The k-distance neighbourhood of p, denoted as \( N_k(p) \), contains every object whose distance from p is not greater than the k-distance. The local reachability density of p, denoted as \( lrd_k(p) \), is simply the inverse of k-distance(p):

\[
ldr_k(p) = \frac{1}{k - \text{distance}(p)}
\]

The local outlier factor is just the average of ratios of local reachability density of p to that of its neighbouring objects:

\[
\text{LOF}_{\text{MinPts}}(p) = \frac{\sum_{o \in N_{\text{MinPts}}(p)} lrd_{\text{MinPts}}(o)}{\text{MinPts}}
\]

where \( \text{MinPts} \) is a predefined constant, which can be loosely interpreted as a smoothing factor. It is easy to see that the lower \( p \)'s local reachability density is, and the higher the local reachability densities of \( p \)'s \( \text{MinPts} \)-nearest neighbours are, the higher is the \( \text{LOF} \) value of \( p \). From this definition, it is clear that LOF is basically a
measure comparing the density of region around the object we are interested to the densities of those regions of its surrounding objects. This definition implies that outlier-ness is a local property and it is very intuitively appealing when a global cluster pattern does not exist or is not the focus. For more justification and comparison, please refer to the original publication (Breunig et al., 2000).

There are many distance measures, such as Hamming distance and Euclidean distance. This local density based method is not limited to any particular measure. We choose Euclidean distance in this research.

4 Optometrist data set

The data set used in this study was drawn from Medicare Australia’s Enterprise Data Warehouse, covering billing records of optometrists for a rolling one year period, ending the third quarter of 2008 (1 Oct, 2008 – 30 Sept, 2008 inclusive).

There are only 26 items in the Medicare Benefit Schedule (Australian Government, 2007) which optometrists can bill Medicare Australia. Given the advice from the domain expert, some of these 26 items are combined thus we obtain a list of 12 unique variable combinations from the original 26. Each variable represents the number of services of a particular item or combined items. The patient age is obviously related to some optometry services and thus average patient age is included as an extra variable. The total number service is also included, resulting in 14 variables for each record. The data set contains 2893 optometrists, one record for each optometrist. The density based outlier detection and some other comparing methods are performed on this dataset.

5 Experimental study

We applied two major analyses on this data set, the density based LOF outlier detection method and a domain knowledge based univariate method, to discover inappropriate practice. In addition to these two methods, a random sampling was also conducted to serve as a baseline measure, as any more complicated method has to outperform this baseline.

1) LOF method. There is no prior knowledge of relative importance among the variables and thus we assume that each of these 14 variables has equal weight. The data is normalised before feeding into the LOF algorithm. We used one of common normalisation methods, i.e., each variable is normalised to mean 0 and standard deviation 1 by subtracting its mean and then dividing by its standard deviation (Sarle, 2002). There is only one tuning parameter MinPts in the LOF algorithm as explained in Section 3. We choose a lower bound of MinPts 30 and an upper bound of 50. We followed the heuristics offered in (Breunig et al., 2000) to determine the LOF, i.e., the data is fed into the LOF algorithm with two MinPts values, 30 and 50, the resulting larger LOF value is taken as the final value for that record. The LOF calculation was undertaken using the dprep package in the open source statistical software R.

2) Domain knowledge based univariate method. This is a crude univariate method. Based on the experience of the domain expert, amongst the 14 variables in the dataset, there are three variables which may be particularly related to high risk behaviour. If any of these three variables significantly deviates from its mean, it is often an indication of higher risk. So in this method, we pick those optometrists who have values for any of these three variables that are 4 or more standard deviations away from its mean. This gives us a list of 32 optometrists.

3) Baseline random sampling. A sample of 25 optometrists was drawn randomly as the most basic benchmark.

We propose two approaches to compare the results and assess the effectiveness of the methods in identifying potential non-compliant individuals in this research. The first validation method is indirect. The compliance history of identified optometrist is analysed. The intuition is that there should be some correlation between high risk optometrists and past records of non-compliance activities. The second one is a direct method. The identified optometrists were presented in a de-identified form to the subject matter expert for evaluation.

Admittedly, none of these two validation methods are perfect as the only relatively reliable method of validating non-compliant individuals is comprehensive desk and field audit, which is often prohibitively costly if it needs to be done on a large scale. The first method - validation using historical compliance records - has the implicit assumption that the past is an indication of the future which is obviously not always true. The second method requires the involvement of the subject matter expert, which takes advantages of prior human knowledge and at the same time presents the opportunity for human error. Furthermore, it is possible that a particular subject matter expert might tend to focus on a particular set of compliance risks. However, we speculate that the combination of these two complementary methods would give us an indication of the performance of the outlier detection methods.

6 Results

The results from three methods – density based LOF, domain knowledge based univariate method and baseline random sampling are compared with two approaches – history checking and domain expert manual validation. These two comparisons are presented in the following two subsections respectively, which give us similar conclusions.

6.1 Validation using historical compliance records

There is a compliance database available to us, containing records of medical practitioners and allied health
professionals who have been approached in relation to
previous compliance activities. The first validation is to
match specialists identified by three methods against their
compliance history in this compliance database. This
provides an estimate of the effectiveness of the various
outlier detection methods studied in this work in detecting
non-compliant practice.

<table>
<thead>
<tr>
<th>Method</th>
<th>Have compliance record</th>
</tr>
</thead>
<tbody>
<tr>
<td>LOF - High</td>
<td>36.00 %</td>
</tr>
<tr>
<td>LOF - Low</td>
<td>0.00 %</td>
</tr>
<tr>
<td>Univariate</td>
<td>31.25 %</td>
</tr>
<tr>
<td>Random</td>
<td>25.00 %</td>
</tr>
</tbody>
</table>

Table 1: Comparison of three methods. LOF method
is presented in first two rows.

In Table 1, the results of the three methods are compared – density based LOF method, univariate method and random sampling. As discussed previously, Local Outlier Factor (LOF) is a measure of outlier-ness. Large LOF value indicates large deviation. However, is large deviation directly correlated to a high risk of potential compliant activities? In order to verify this, 25 records with highest LOF values and 25 records with the lowest LOF values were examined, as listed in the first row of Table 1. It is evident that the LOF value is a good risk indicator, as 36% of records with highest LOF have compliance record while none of the records with lowest LOF have a compliance record. It is clear from Table 1 LOF is significantly better than randomly sampling and at least as good as crude univariate method.

<table>
<thead>
<tr>
<th>Method</th>
<th>Average number of compliance records per optometrist who has at least one compliance record</th>
</tr>
</thead>
<tbody>
<tr>
<td>LOF - High</td>
<td>1.67±1.12</td>
</tr>
<tr>
<td>LOF - Low</td>
<td>N/A</td>
</tr>
<tr>
<td>Univariate</td>
<td>1.30±0.67</td>
</tr>
<tr>
<td>Random</td>
<td>1.00±0.00</td>
</tr>
</tbody>
</table>

Table 2: Comparison of three methods on the average number of compliance records per optometrist who has at least one compliance record.

We are aware that an optometrist may have different numbers of records in their compliance history, which suggests some of them have multiple incidents of non-compliant practice or have been engaged in multiple compliance activities. We speculate that the average number of compliance records per optometrist may be an indication of the severity or certainty of a possible non-compliant practice. We listed the average number of compliance records per optometrist who has at least one compliance record for each method in Table 2. For all those optometrists identified by the highest LOF values and who had compliance records, they had on average 1.67 records (with the standard deviation 1.12). While this is higher than 1.30 of the univariate and 1.00 of the random sampling, the difference is not statistically significant.

We also experimented larger sample size. We reduced the threshold of the univariate method to 3 standard deviations and that resulted 109 records and we drew same number of records with the highest LOF values. The average number of compliance records per optometrist from univariate and LOF methods was still different but not statistically significant.

Checking against an optometrist’s compliance history provides us with a measure to compare the density based LOF method with other methods. In this comparison, we find that LOF is at least as effective as the univariate method in identifying high risk individuals and significantly outperforms random sampling. However, if the certainty of a possible non-compliant practice can be measured by the number of compliance records one has, there is no significant correlation between LOF value and the number of compliance record per optometrist.

6.2 Validation by domain expert

Checking compliance history of an optometrist is an indirect way of validation. During this study we had access to a domain expert, a compliance optometrist, to provide us with a more direct evaluation. Although the best way to validate whether an individual optometrist was truly non-compliant, would be a review by a panel of peers consulting a domain expert was the best we had access to at the time.

We invited the domain expert to evaluate each group of de-identified optometrists and rate each individual with one of three levels of risks – low, medium and high. The evaluation is listed in the Table 3.

<table>
<thead>
<tr>
<th>Method</th>
<th>High</th>
<th>Medium</th>
<th>Low</th>
</tr>
</thead>
<tbody>
<tr>
<td>LOF - High</td>
<td>6</td>
<td>16</td>
<td>3</td>
</tr>
<tr>
<td>LOF - Low</td>
<td>0</td>
<td>0</td>
<td>25</td>
</tr>
<tr>
<td>Univariate</td>
<td>13</td>
<td>18</td>
<td>1</td>
</tr>
<tr>
<td>Random</td>
<td>0</td>
<td>8</td>
<td>17</td>
</tr>
</tbody>
</table>

Table 3: Rating of optometrists provided by subject matter expert.

As listed in Table 3, the majority of optometrists identified by LOF (with highest value) and univariate methods are rated as having a risk level of medium or above, whilst the randomly sampled optometrists were mostly rated as posing a low risk.

To gain a better idea of these results, we regroup these three groups of risk levels into two. Medium and high risks are grouped together, and classified as Unexplainable, as the behaviour of these individuals cannot be explained by the domain expert and thus may
be related inappropriate practice activities. Low risk optometrists were reclassified as *No Further Action* because they look normal and no further action needs to be taken.

The outcome of this regrouping is listed in Table 4. It clearly shows that density based LOF is a good indication of risk as 88% of individuals with highest LOF are risky and none of the individual with low LOF represent a risk. Compared to other methods, the LOF method significantly outperforms random sampling, which is consistent with the results from the previous history checking validation.

<table>
<thead>
<tr>
<th>Method</th>
<th>Unexplainable</th>
<th>No Further Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>LOF - High</td>
<td>88.00%</td>
<td>12.00%</td>
</tr>
<tr>
<td>LOF - Low</td>
<td>0.00%</td>
<td>100.00%</td>
</tr>
<tr>
<td>Univariate</td>
<td>96.887%</td>
<td>3.13%</td>
</tr>
<tr>
<td>Random</td>
<td>32.00%</td>
<td>68.00%</td>
</tr>
</tbody>
</table>

Table 4: Risks of optometrists provided by domain expert.

From the Table 4, we can see over 97% of optometrists identified by univariate method cannot be explained while only 88% by LOF in this evaluation. However, this needs to be interpreted with care. As mentioned before, the univariate method is derived from the experience of the domain expert and is based upon unusual values for one or more of the three variables believed to be related to high risk behaviour. When the domain expert evaluates the risk, these are the most obvious indicators to look at. Not surprisingly, the univariate method has an accuracy approaching 100% in this evaluation. On the other hand, the density based LOF method considers multiple variables at the same time and thus the results are not as straightforward for the domain expert to interpret. With these considerations in mind, we argue that LOF might have comparable performance as univariate methods in the real world if the bias is removed in this analysis. Furthermore, LOF method looks at multiple factors simultaneously. It is possible that it might help identify individuals with multiple compliance issues or subtle issues that univariate method cannot identify.

7 Discussion

Combining the results from two evaluation methods – compliance history checking and domain expert validation, we can see that LOF is significantly better than the baseline random sampling and has the comparable accuracy of the univariate method. However a number of questions remain, including, does similar performance suggest that the LOF and the univariate method identify similar groups of people? Is LOF just an alternative way to employ a domain knowledge based univariate method of discovering individuals with extreme values of single variable? There are many different types of non-compliant practice and thus we do not expect one single method would be able identify all of these different types. Therefore, in this research, what may be useful is to investigate whether there is a significant overlap between high risk individual identified by LOF method and the domain knowledge based univariate method. If there is a significant overlap, that would render the LOF redundant.

We matched the list of 25 individuals identified by LOF with highest LOF values against 32 individuals identified by the univariate method with extreme values. There is only one individual appearing in both and thus the overlap is minimum, which indicates LOF is not a simple replication of univariate analysis on this problem.

To further demonstrate that the LOF and univariate analysis actually identify different types of high risk individuals. The individuals identified by LOF and the univariate method are plotted against three variables which are used by univariate method in Figure 1. It is clear that those individuals identified by different methods cover different regions of the problem space, which shows that these two methods are complementary to each other.

8 Conclusions and future research

This paper presents a novel application of density-based local outliers and demonstrates how it can be used with the aims of detecting inappropriate practice in the health service management domain, in this case, using optometrist billing compliance.

The results suggest the density based LOF outlier detection method is effective. We validated the results in two major ways. The first validation method is matching against compliance history and the second is the domain expert confirmation. Although there are different emphasis and bias for both of validation methods, they provide us similar conclusions.

These validation methods shows that LOF is a reliable indication of high risk individuals i.e. where low LOF values clearly related to individuals behaving consistent
to their peers and high LOF values clearly related to higher risk. The LOF method significantly outperformed the baseline method, which was randomly sampling of optometrists. It is also at least as effective as a univariate method derived from prior domain knowledge. Furthermore, LOF method looks at multiple factors at the same time. It is possible that it might help identify individuals with multiple compliance issues or subtle issues that univariate method cannot identify.

In practice, as one of the outlier detection methods, the density based method can be integrated into a bigger system, which employs multiple methods to detect inappropriate practice, for better accuracy. From our experience, the hybrid approach involving both automatic data mining techniques, such as LOF, and domain experts’ theory driven methods, often facilitate us constructing very effective systems for compliance purposes. Thus, the future work includes further analysis on the effectives of LOF methods and, if appropriate, how to best integrate it with other methods in Medicare’s operational systems.

9 Acknowledgements

The authors would like to acknowledge the input of Dr. David Jeacocke for his insightful comments during this research, and Dr Steve Zantos for his advice on the interpretation of optometric data.

10 References


