Sequential pattern detection for identifying courses of treatment and anomalous claim behaviour in medical insurance

1st James Kemp
School of Computer Science and Engineering
University of New South Wales
Sydney, Australia
james.kemp@unsw.edu.au

2nd Christopher Barker
Compliance Analytics Branch
Australian Government Department of Health
Sydney, Australia
chris.barker2@health.gov.au

3rd Norm Good
Australian eHealth Research Centre
Commonwealth Scientific and Industrial Research Organisation
Brisbane, Australia
norm.good@csiro.au

4th Michael Bain
School of Computer Science and Engineering
University of New South Wales
Sydney, Australia
m.bain@unsw.edu.au

***Abstract*—Fraud and waste is a costly problem in medical insurance. Utilising sequence information for anomaly detection is under-explored in this domain. We present a multi-part method employing sequential pattern mining for identifying and grouping comparable courses of treatment, finding patterns within those courses, calculating the cost of possible additional or upcoded claims in unusual patterns, and ranking the providers based on potential recoverable costs. We applied this method to real-world radiation therapy data. Results were assessed by experts at the Australian Government Department of Health, and were found to be interpretable and informative. Previously unknown anomalous claim patterns were discovered, and confirmation of a previously suspected anomalous claim pattern was also obtained. Outlying providers each claimed up to $486,617.60 in potentially recoverable costs. Our method was able to identify anomalous claims as well as the patterns in which they were anomalous, making the results easily interpretable. The method is currently being implemented for another problem involving sequential data at the Department of Health.**

***Index Terms*—unsupervised learning, anomaly detection, medical information systems, data mining, decision support systems**

I. Introduction

Fraud and waste in medical insurance claims are costly problems, with fraudulent claims typically costing 3-8% of expenditure for OECD healthcare organisations [1], [2]. Challenges in detecting fraud and waste in healthcare are well documented, and include typical big data challenges such as: large volume of data; heterogeneity in both the sources and meaning of data; concept drift, whereby the meaning of the data changes over time; and class imbalance, where fraudulent or wasteful claims are expected to appear at a much lower rate than acceptable claims [2]–[6].

Detection of fraud can be viewed as an outlier detection problem [6]. The variation inherent in medicine contributes to the challenges, due to factors such as differing patient demographics, changes in provider training and techniques, and clinical uncertainty [7], [8]. This means that outliers can occur for a variety of reasons, and identifying instances of anomalies is not enough to indicate potentially fraudulent or wasteful behaviour. Rather, identifying repeated anomalous behaviour is preferable. Given the costs involved in identification and recovery of fraud and waste, calculating costs for the anomalous behaviour can also be a useful metric [9].

With these challenges in mind, new techniques are required to handle the problem [6]. While machine learning and data mining methods are becoming commonplace, there are challenges to typical machine learning approaches. These include the human expert time and effort required to label data for supervised approaches, and the need for human-interpretable results if they are to be used for audits or legal action (where many techniques are black-box methods) [4], [10]. Some potential solutions are under-researched, such as time-series or sequential approaches [4], [5]. We examine claims made from the Australian Medicare Benefits Schedule (MBS) [11].[1] We present a technique which identifies *courses of treatment* within patient data and groups comparable courses of treatment, then finds patterns within the courses of treatment and identifies unusual rates of patterns by providers compared with the body of peers, and finally ranks the providers based on potential recoverable costs.

This research is supported by an Industry PhD scholarship which includes funding from the Commonwealth Scientific and Industrial Research Organisation, the Department of Health, Australian Government, and an Australian Government Research Training Program (RTP) scholarship.

[1] Owing to privacy concerns it will not be possible to release this dataset. Source code is available: github.com/jpkemp/anomaly_detection_framework

978-1-6654-6819-0/22/$31.00 ©2022 IEEE

The remainder of the paper is structured as follows. In Section I-A we briefly review some of the relevant work in the literature. Section II develops our multi-part method, in which we construct an appropriate representation in terms of courses of treatment from the raw claims data, detect anomalous provider behaviour, and score it in terms of potentially recoverable costs. Since we do not have "ground truth" labelled data for this study, in Section III we give results centering on the anomalous behaviour identified by our approach and its validation by domain experts. Section IV discusses our approach and validated results, including limitations and topics for future research.

A. Related work

Many sequential pattern mining algorithms exist [12], however identifying sequences is only part of the challenge for this research, and useful representations of the underlying entities in the data (such as patients, providers, etc.) and methods to compare similar entities need further development. Few papers have been published which incorporate sequence information into fraud and waste detection in healthcare. Two recent papers have made some headway into the problem, and highlight some useful approaches as well as challenges in the area.

Farbmacher et. al. [13] trained a supervised neural network on data from a German private health insurer using an economic evaluation criterion as the loss function. The criterion was designed to capture the trade-off between gain from recovering fraudulent claims and cost of recovery of those claims; therefore, it incorporated factors such as costs of false positives and number of claims made by a provider. They used an attention mechanism to reduce claim sequences to fixed-length vectors suitable for machine learning, while capturing dependencies between the claims. This does, however, mean that the sequences themselves (and potential uncommon relationships between the items) were not modelled directly. They ranked the providers in the test set using the prediction score from the output layer. Their model outperformed a boosted tree model, particularly when total recovery and cost of recovery were considered. However, unsupervised approaches are preferred due to the high human time and effort costs of labelling data, and the rapid concept drift as medical practice evolves and deliberate fraudsters respond to compliance activities means that models trained on labelled data can quickly become irrelevant [4].

Snorovikhina and Zaytsev [14] investigated unsupervised approaches to fraud detection using labelled data from Allianz insurance. They implemented a sequence-to-sequence autoencoder to detect anomalous sequences, based on the reconstruction error. They also used an unsupervised long short-term memory (LSTM) neural network for predicting anomalous individual treatments within sequences, by determining the probability of a treatment given the previous treatments. In both cases, a threshold was used to class patient sequences as anomalous or not. Both methods outperformed an isolation forest model, with area under the receiver-operating curve up to 0.771. Precision was low at a recall of 0.8 (up to 0.333 for the LSTM model). The LSTM model goes some way to allowing human interpretation, with its ability to determine which treatments are anomalous in which positions. However, it may not help with determining why the treatment appears anomalous, and as mentioned, individual instances of anomalous behaviour may not be indicative of fraud or waste and a provider-centred view is preferable. However, neither of these methods solves our problem, and cannot therefore be used for comparison in our empirical evaluation II-M.

II. Methods

While each of the papers covered in Section I-A partially address some of the challenges outlined, neither is a complete solution. In order to address these challenges, in this paper we develop an unsupervised approach based on sequential pattern mining and apply it to data from patients undergoing radiation therapy. Radiation oncology was selected as patients receive consecutive treatments, and the claims are therefore inherently sequential. The Australian Government Department of Health and Aged Care (DoH) is currently interested in radiation oncology claims, which facilitated validation of the results[2].

A. Data

The Medicare Program in Australia is a public health insurance scheme, providing reimbursement for medical costs to Australian residents and some categories of visitors. While technically the reimbursement is for the patient, typically the claims are made and the funds are received by the provider [11]. The MBS data set is comprised of claim rows, each containing information about a claim for a single item code representing a service. Multiple item codes can be claimed for an episode of care. Information in the claim rows includes details relevant to the claim such as the date and location of service, service provider, patient, and referrer identifiers, the item claimed, and reimbursement costs and modifiers for the item. For this study, no identifying information was included in the data except unique numerical identifiers for each patient and provider. Conceptually, each row in the dataset can be considered a four-tuple $\langle T, I, A, V\rangle$, for $T \in \mathit{Dates}$, $I \in \mathit{Item\ Codes}$, $A \in \mathit{Patients}$ and $V \in \mathit{Providers}$.

B. Data extraction

Radiation therapy treatment occurs regularly over a period of approximately 3-8 weeks, depending on the severity of cancer and condition of the patient [15]. The episodic nature of the treatment makes it a good candidate for sequence pattern mining. In the MBS, radiation therapy for cancer is expected to begin with a pair of items representing a computerised simulation and treatment plan, to be followed by a series of items representing treatment and verification of the treatment, with differences between the item codes primarily reflecting different treatment equipment [11]. For a typical patient only

[2]Ethical approval for this study was granted by the University of New South Wales Human Research Ethics Committee Executive.

one of the simulation/planning item pairs would be present, even over several courses of treatment, since the only valid reason for doing a new simulation and plan is discovery of a new cancer during a treatment [11]. All claim rows for patients with a claim for an item from the MBS Category 3: Therapeutic Procedures, Group T2: Radiation Oncology, Subgroup 5: Computerised Planning were identified from the 2021 MBS claims data. From those claim rows, claim rows claimed by any provider with claims from Category 3: Therapeutic Procedures, Group T2: Radiation Oncology were extracted. The data we used was therefore the claims for patients undergoing radiation therapy, by their radiation therapists (but not restricted to claims for Radiation Oncology items). The extracted data comprised 10,268,717 claim rows from 404 providers. There were 72,239 patients who received treatment, with 3,954 different item codes claimed.

C. Method overview

The method we used to identify providers engaging in unusual claim behaviour follows these broad steps:

1) Identify courses of treatment for each patient and extract claims made within them
2) Group courses of treatment based on the planning and simulation items claimed within the course of treatment
3) Within each group, identify patterns of sequential claims across each provider's courses of treatment, as well as rarely claimed items
4) Flag patterns where the rate for a provider is an outlier compared with their peers
5) For the flagged patterns, identify potential substituted/extra items
6) Identify occurrences of the flagged patterns within the provider's courses of treatment, and sum the costs of the substituted/extra items
7) Rank the providers based on their extra costs

D. Identifying courses of treatment

We used a two-stage process to identify the start and end dates of a course of treatment. Firstly, local periodic patterns were identified within the data for each patient. Secondly, the dates for the planning/simulation item claims and dates of the local periodic patterns were merged to create a date range for each course of treatment for a patient.

The MBS has a five-layer tree structure representing the relationships between services which may be claimed, comprising $Category \rightarrow Group \rightarrow Subgroup \rightarrow Subheading \rightarrow Item$ (with subgroup and subheading being optional) [11]. To identify the local periodic patterns, the item codes were converted to their subgroup-level *ontology location* within the MBS. That is, all items within the Therapeutic Procedures → Radiation Oncology → Computerised Planning subgroup would be identified as "3_T2_5". Ontology structures are prominent in biomedical domains, and are typically built in a way that both machines and humans can interpret [16], [17], making them a useful way of incorporating domain knowledge into algorithms without explicitly designing new features or rules. The timestamps of the dates for these claims were converted to integer values from the first day of a patient's claims. That is, if a patient's first claim for 2021 was the 3rd of January, claims from that day would have a timestamp of '1', claims from the 4th of January would have a timestamp of '2', and so on. For each patient, the converted codes and timestamps were passed into the SPMF implementation of the LPP-Growth algorithm [18]. LPP-Growth can return multiple patterns, which may overlap in time. The returned timestamps were merged if the timestamps overlapped or were within 31 days of each other, in order to give the overlapping pattern timestamps. The timestamps for claims of planning/simulation items were identified and the timestamps merged with pattern timestamps if the claims were within 31 days of each other, in order to get the start and end timestamps of the course of treatment. These timestamps were then used to extract the claims for separate courses. Multiple courses of treatment may be identified for each patient. The ith course of treatment for a patient A is defined as $C_{A,i} = \{\langle T, I, A, V\rangle \mid T_{(m,i)} \leq T \leq T_{(n,i)}\}$, where $T_{(m,i)}$, $T_{(n,i)}$ are the start and end timestamps, respectively, identified as above.

E. Grouping courses of treatment

Courses were labelled according to the unique set of ontology simulation and planning items occurring within the course, creating a *context* with which to group similar courses for comparison. Comparing providers of a broad specialty type, such as radiation oncologists, can be a poor comparison as subspecialties and patient demographics may be not be available in the data. Instead, we attempt to compare like with like - i.e., what does a provider do when in a given situation; in this case, the situation is defined by the technology used to treat the cancer, as defined by the simulation and planning items chosen. The set of all subsets of item codes from ontology location O, in this case "3_T2_5", defines the set of "context labels" $\mathcal{L}$. The subset of items from O appearing in a given course C will correspond to exactly one of these context labels $L \in \mathcal{L}$. A context $N_L = \{C_{(A,i)} \mid \forall I \in L,\ \exists\langle T, I, A, V\rangle \in C_{(A,i)}$ and $\forall I' \notin L, I' \in O, \nexists\langle T, I', A, V\rangle \in C_{(A,i)}\}$. The remaining steps in the process (described below) are applied to each context individually.

F. Identifying sub-sequence patterns

To identify unusual claim patterns, we returned to sequence mining. We were informed by an expert at the DoH that in this area it is common for multiple providers to be involved with one patient, particularly in large private practices. As more than one oncologist may therefore be involved in any given course of treatment, any provider involved in a course was treated as if all the claims belonged to them. While this means some courses of treatment were assessed more than once, there were two advantages to this approach. Firstly, it kept the courses of treatment comparable by keeping the full sequence, rather than separating sub-sequences by provider and losing surrounding information. Secondly, it allowed for the detection of possible collusion among the practitioners, e.g., where two providers are making the same claims for a patient.

Each course has a set of involved providers. By inverting that view we can obtain the set of courses in which a provider is involved. A sequence S is the set of items in a course $C_{A,i}$, ordered by time. We obtain a set of sequences Q_V by converting the courses in which a provider is involved to sequences. For each Q_V, we used the CM-SPAM algorithm [19] to detect sequential patterns. Support of a pattern P for a provider is defined over Q_V. Specifically:

$$support(P) = \frac{|\{S \in Q_V \mid P \text{ is a sub-sequence of } S\}|}{|Q_V|}$$

For clarity, we will also refer interchangeably to support of a pattern as the *claim rate* of a pattern for a provider:

$$support(P) = rate_{P,V}.$$

CM-SPAM finds patterns up to a set maximum length across a collection of sequences, with minimum claim rate set as a percentage of courses in which the pattern occurred for the provider. The ability to set the maximum length improved tractability. Courses for each provider were passed to the SPMF implementation of the CM-SPAM frequent pattern mining algorithm [19], with maximum sequence length of 3 (i.e., no more than 3 items in the sequence, whether consecutively, concurrently, or a mix), maximum gap of 1 (i.e., no gap), and claim rate of 3% (or a percentage equating to a minimum of 3 sequences, if that was higher).

G. Identifying unusual patterns

For each sequential pattern P, there is a distribution of $rate_{P_V}$ across the providers (assuming 0 for providers with a claim rate $< 3\%$). The extreme outlier cutoff $thresh_P = Q3 + 3 \times IQR$, where Q3 is the third quartile and IQR is the inter-quartile range), was chosen to identify providers claiming patterns at an unusual rate. Providers with an extreme-valued claim rate for a pattern compared with their peers ($rate_{P,V} > thresh_P$) were noted to be an unusual claimant of that pattern. Note that it is not the pattern itself that is flagged as unusual — it is the combination of the pattern *and* the provider's claim rate that marks it as anomalous. Providers with fewer than 20 courses of treatment within the context were excluded completely.

H. Identifying anomalous items

For each provider with outlying patterns, we identify similar or subsequence patterns for which the provider would not be considered an outlier, enabling us to identify potential substitute or additional items that have been claimed. Substitute items may indicate upcoding, where a more expensive item is claimed though a cheaper item should have been delivered [2]. This allowed us to better estimate the potential recoverable costs for the provider.

Firstly, we test whether the provider is substituting one item for another. Potential substitute patterns were identified by finding all patterns of the same length containing one different item, but sharing the same pattern of sub-group level ontology locations. That is, if an anomalous pattern had a treatment and verification item on one day, and a verification item on the next, similar patterns would be examined if they also had a treatment and verification item on one day and a verification item on the next, but for which one of the item codes was different. Using the ontology allowed us to incorporate domain knowledge without explicit consultation of experts — i.e., we could determine what items might reasonably be considered similar within the context by using the least-generalisation from the ontology.

If the provider would be considered a normal claimant for one of the potential substitute patterns (see Section II-I), the different item would be considered to be a substitute item, and the pattern would be given a cost $F_{P,V} = max(0, u - s)$, where u is the cost of the unusual item and s is the cost of the substitute item. If a substitute item is not identified, a potential extra item is looked for.

Potential subsequence patterns with the same items and pattern of days, minus one item (and potentially day), were identified. Normality was checked for the potential subsequence pattern with the highest outlying claim rate, as per Section II-I. If the minus-one-item subsequence pattern failed the normality check, potential subsequence patterns with two fewer items were identified. This procedure was repeated down to subsequence patterns of length 1. The unusual pattern is assigned a cost of the items which are additional to the subsequence pattern ($F_{P,V} = u$). If no potential substitute or subsequence patterns were found, all items in the original pattern were flagged at full cost for the occurrences of that pattern.

I. Normality check

At each stage of identifying potential replacement patterns (whether substitute or subsequence patterns), the provider's claim rate for the unusual pattern was compared with the threshold of the potential replacement pattern with the highest threshold for extreme outliers. That is, from the potential replacement patterns, the one with the highest value for $thresh_P = Q3 + 3 \times IQR$ was selected for comparison. If $rate_{P,V} < max(thresh_P), P \in R$ (the provider would not be considered an outlier if they were claiming the potential replacement pattern instead of the unusual pattern), the item(s) in the unusual pattern but not in the replacement pattern are considered to be the unusual items.

J. Example

As an example of the process in Sections II-H and II-I, say we have two sets of items in different ontology locations, $X = \{X_1, X_2\}, X_j \in O_X$ and $Y = \{Y_1, Y_2\}, Y_j \in O_Y$. In the following we use the notation $X \to Y$ to denote the sequential pattern, or simply pattern, where item X is followed by item Y in a course, with A and Y occurring on subsequent days.

Suppose a pattern $X_1 \to Y_2$ is flagged as an unusual pattern for a particular provider V_1, who claims that pattern in 70% of their courses, i.e., ($rate_{X_1 \to Y_2, V_1} = 70\%$). The outlier threshold for that pattern is, say, 10% ($thresh_{X_1 \to Y_2} = 10\%$). Potential replacement patterns would be as follows:

Pattern ID	Pattern	Replacement type	$thresh_P$
1	$X_1 \to Y_1$	Substitute	85%
2	$X_2 \to Y_2$	Substitute	15%
3	X_1	Subsequence	95%
4	Y_2	Subsequence	6%

First, the threshold of the substitute pattern with the highest outlier threshold (in this case pattern 1) would be compared with the provider claim rate for the unusual pattern. As 85% is higher than 70%, the provider would be considered a normal claimant if the unusual pattern was replaced with pattern 1. Therefore, Y_2 can be considered the unusual item, and the unusual pattern would be assigned a cost for this provider $F_{X_1 \to Y_2, V_1} = max(0, cost(Y_2) - cost(Y_1))$, as Y_2 may be upcoded from Y_1.

If no substitute pattern was found (say pattern 1 only had a threshold of 60%, i.e., less than 70%), the set of subsequence patterns would be examined. In the example pattern 3 has the highest outlier threshold from the subsequence patterns. As 95% is higher than 70%, the provider would be considered a normal claimant if the unusual pattern was replaced with pattern 3, and again Y_2 would be considered the unusual item. In this case, since we are dealing with subsequence patterns, the full cost of Y_2 would be assigned as a cost for this provider (i.e. $F_{P,V} = cost(Y_2)$) as Y_2 would be considered an additional, rather than a substituted, item.

K. Identifying rare items

In order to reduce the number of unusual patterns which required processing, patterns containing rare items were removed prior to step II-G. Rare items were defined as items occurring in $< 10\%$ of courses of treatment within a context. The nature of sequential pattern mining is to produce all possible combinations of sequences. Many patterns can be found which essentially describe the same behaviour.

If a provider was claiming a rare item at an unusual rate, multiple sequences would be found containing that item. The unusual item was therefore the rare item if the pattern contained only one unusual item. If the pattern contained multiple unusual items, other patterns (not containing the rare item) would also cover the claiming of the other unusual item(s). Removing patterns containing rare items therefore reduced processing time and complexity with little loss of information. Rare items were assigned a direct cost $F_I = u$, for rare item I.

L. Identifying unusual providers

Each provider was assigned a score, initially 0, which represented the potentially recoverable costs (PRC) from the provider in the context. To calculate the score, we iterated over each of the sequences in which the provider was involved. Each occurrence of an unusual pattern or rare item is marked, and the cost of the unusual items added to the score.

Continuing the example from Section II-J, say the provider V_1 had ten courses in total, and was not claiming any other patterns at an outlying rate, and was not claiming any rare items. Say that the unusual pattern was assigned a cost $F_{X_1 \to Y_2, V_1} = \$100$, and that the pattern occurred five times each in three of the courses, and three times each in four of the courses, giving a total of twenty-seven occurrences of the pattern. The provider would therefore have a PRC of $2,700.

Due to the overlap of unusual patterns - multiple patterns can be found that identify the same incorrect behaviour - any given item at a given sequence position was only scored once, regardless of how many unusual sequences it appeared in. One downside of this approach is that duplicate item claims on the same day would only be scored once, regardless of how many duplicate items were present. However, same-day duplicate item claiming is not a sequence-specific problem, and other techniques are available to address that issue. Providers were ranked based on their PRC.

We used ranking rather than attempting to identify a fixed threshold, e.g., in terms of dollar amount, for classifying suspicious/non-suspicious providers as defining such thresholds can be subjective and difficult to identify. Ranking allows investigators to prioritise cases and keep working until the results are no longer cost-effective to pursue [6]. Given the variation inherent in medicine, recoverable costs is a useful metric for prioritisation. Examining costs increases potential return on investment, rather than looking solely for anomalous claims which may be an artifact of medical heterogeneity rather than a representation of repeated, incorrect behaviour [9]. Cost of recovery can be a disincentive to action, so prioritising high-cost behaviours is beneficial for payers [2].

M. Validation of the algorithm

No labelled data was available to validate the results. The most common planning item represents dosimetry planning for intensity modulated radiotherapy (IMRT) [11]. Consequently, we selected contexts containing courses of treatment with that planning item for analysis, in order to validate the results. Those contexts were examined for the number of courses and providers, and the clinical relevance of simulation/planning item combinations was discussed with two senior compliance analysts (SCAs) at the DoH. For more in-depth examination we selected the context containing courses of treatment containing only the simulation/planning pair representing simulation and planning of IMRT, which is the most common combination, and the only one that is explicitly permitted by the MBS for IMRT [11]. The patterns associated with the top 10 high-scoring providers within that context were examined in detail, and results discussed with the SCAs.

III. Results

From discussion with the SCAs, two points of interest were highlighted: unexpected simulation/planning item combinations in the discovered contexts, and unusual patterns claimed by high-cost providers.

A. Contexts

We identified 72,473 courses of treatment with 344 combinations of simulation and planning items (contexts). Many of

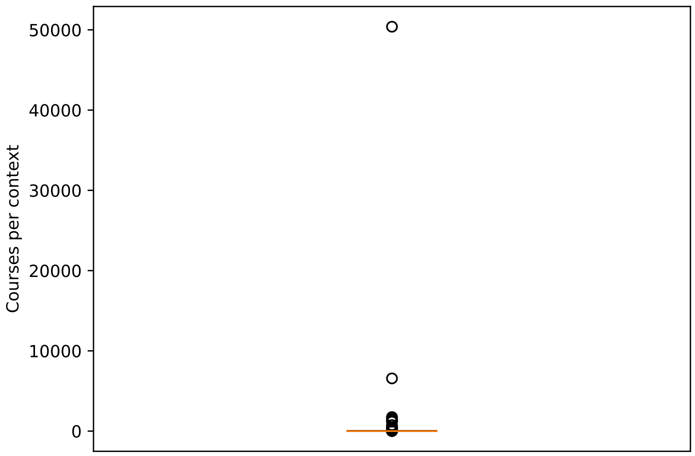

Fig. 1. Box plot of number of courses of treatment per context.

the contexts contained a single simulation or planning item (22) or three or more simulation or planning items (231), where only a pair of items is expected according to the MBS [11]. The number of courses of treatment assigned to these contexts was generally low, and is likely to be an artifact of either administration error by the provider, or the time period for connecting the context items to the periodic patterns being too low (thereby not connecting the simulation item to the course of treatment, for the single-item contexts) or too high (thereby joining two courses of treatment, for the three or more item contexts) (see section II-D). Course counts per context can be seen in Figure 1[3]. Most of the identified contexts contained very few courses, and may be the result of administrative errors; the combination of simulation and planning items they represent may not be expected according to the MBS. Only 10 contexts contained courses of treatment from more than 20 providers, and only 8 had more than 500 total courses of treatment. While the number of courses of treatment for most of the providers in the unexpected contexts was low, some providers had many courses of treatment in these contexts. An example of the courses of treatment per context is shown in Figure 2. The SCAs were very interested in the outlier providers (those involved in many courses of treatment) from contexts with unexpected simulation/planning item combinations, as the combinations are of questionable clinical relevance and repeated courses of treatment involving these combinations may be indicative of fraudulent or wasteful behaviour. This shows that the method we have outlined can be useful as part of a decision-support system.

B. High-cost providers

The context selected for in-depth examination contained 50,390 courses of treatment. Across the top 10 providers in that context, there were 42 patterns flagged as being claimed at an outlying rate by at least one of those providers.

[3]Note that in Figures 1 to 3 the standard outlier threshold of $Q3 + 1.5 \times IQR$ is used, which is not the same as the extreme threshold defined for $thresh_P$ in the text.

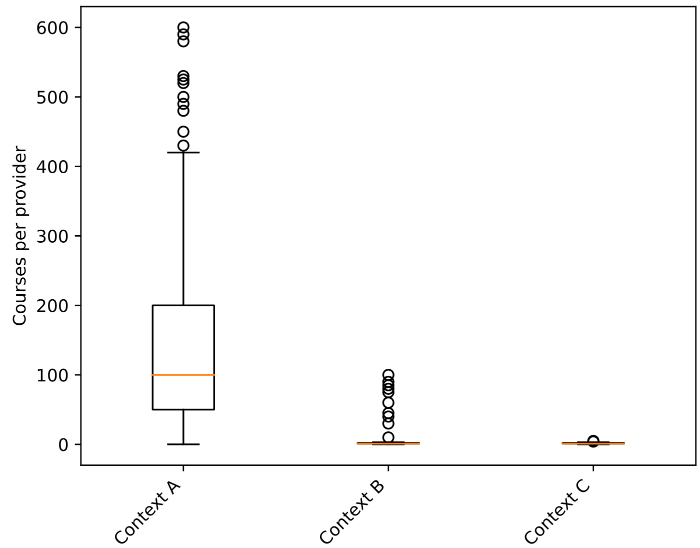

Fig. 2. Box plots illustrating the number of courses of treatment per provider in different contexts. Context A represents a context covering an expected simulation/planning item combination. Contexts B and C represent contexts with unexpected simulation/planning item combinations. Context B includes providers who are repeatedly claiming the unexpected combination. For privacy reasons these plots were created from mock data.

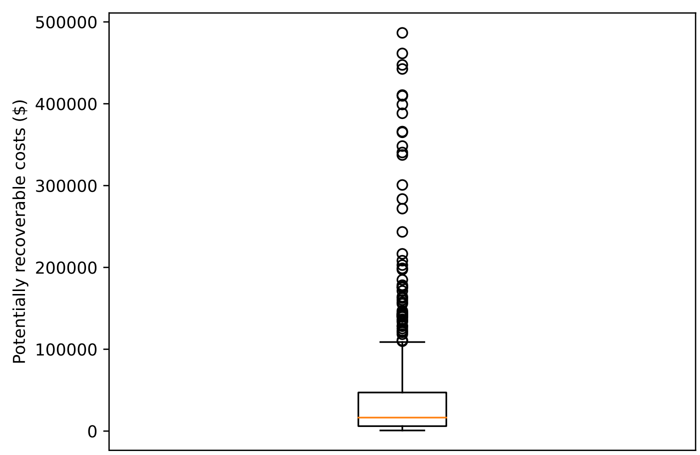

Fig. 3. Box plot of provider potentially recoverable costs, showing that most providers have a small PRC compared with the outliers.

On examination of those patterns, all could be classed as describing at least one of the following claiming behaviours:

1) Duplicate claiming of the planning item (indicated either as part of the pattern, or claimed subsequently to a treatment item - i.e., a previous planning item would have already been claimed)
2) Claiming treatment or verification items concurrently with the simulation or planning item claim; this could be early in the sequence, as part of the first claims of the simulation/planning pair, or later in the sequence, covering the same incorrect claims as the first behaviour
3) Unusual claiming of a phone attendance item

The SCAs found that the potential substitute/subsequence patterns for the unusual patterns were reasonable, i.e., the item identified as the likely unusual claim was appropriate in each case. PRCs ranged from $768.65 to $486,617.60, with a median of $16,495.60 and a mean of $51,140.19. The distribution was highly right-tailed, and PRCs from outlying providers were far in excess of most providers (see Figure 3).

The contribution of item costs to the PRCs was examined for the top 10 providers. The primary cost was from the planning item, which contributed 88.3-98.9% of the PRCs of the top 10 providers. This indicates that the first two types of claiming behaviours were the major drivers behind the ranking, rather than the third type of claiming behaviour or over-claiming of rare items. The first behaviour is a known incorrect claiming behaviour, which the DoH is interested in identifying. While some circumstances allow for duplicate claiming of a planning item within a course of treatment, the advice we have been given is that it is intended to be a rare occurrence. The way it is being claimed by some providers is almost certainly incorrect, given the rates of occurrence.

For the second claiming behaviour, the cost of which was also driven by the planning item, the advice we have been given is that it is possible that sloppy administration would account for some of the behaviour, however the items are not intended to be claimed in this way and the rates at which it is occurring in the high-scoring providers makes it unusual and warrants further investigation.

For the third claiming behaviour, while more patterns covered the behaviour than covered the first two behaviours, the rates of occurrence and cost contribution to the score were low. The SCAs noted that not all instances of the claiming of the item were being flagged by the algorithm, even for the high-scoring providers - the item itself was not the issue, but the order of occurrence and the rates at which it was occurring. While this was unusual compared with the peer body, the patterns could potentially be explained by combinations of patient demographics and provider locations, and were less interesting as a target to pursue, particularly when the low item cost was considered.

IV. Discussion

The multi-part algorithm we have presented here is a novel unsupervised learning algorithm for identifying anomalous behaviour in an interpretable way. It provided informative results when applied to radiation therapy providers, but is generalisable to other sequential problems. We were able to identify:

- comparable courses of treatment from patient information
- unusual claiming behaviours based on outlying rates of claim sequences
- costly providers based on the potentially recoverable costs of the aberrant claims

Validating results for studies on medical fraud and waste can be a time-consuming task for human experts, and exploration of the results by the experts at the DoH is at an early stage [4]. The output was readily interpretable by the SCAs, which is a high priority for the DoH. Other methods such neural networks are more opaque, which is not ideal from a compliance perspective [10]. While explainable frameworks such as Shapley Additive Explanations (SHAP) can identify features and values which contribute to anomalies in unsupervised autoencoders, we are not looking for anomalous samples but unusual, *repetitive* behaviour [20]. Viewing SHAP as a form of sensitivity analysis [21], the approach of Section II-H can also be seen as a form of sensitivity analysis, where a sequential pattern is altered by substitution or shortening, with the resulting change in PRC indicative of the effect due to changing an anomalous item. The ability to determine not only which item was unusual, but the circumstances in which it was unusual, was appreciated by the SCAs, and the patterns flagged as unusual were generally of interest as investigative targets. While one type of pattern was found to be less interesting, the contribution to the PRC was low and providers with unusual claim rates of those patterns would not be scored highly as a result. A previously unidentified unusual claiming behaviour was found, and is being investigated further. Further investigation of the results is also planned, and the algorithm has generated considerable interest in several sections of the DoH.

The method is generalisable to other episodic sequences. We have been requested to adapt the algorithm to a second project, regarding claims from the Pharmaceutical Benefits Scheme of fertility medication for assisted reproductive technology. The structure of the courses of treatment may be different for that situation, as dispensation of medication is not as regular as radiation therapy and treatment may be more tailored for each patient. This means that the first part of the algorithm (to identify the course dates) will need some alteration in order to identify comparable courses of treatment. However, early results [data not presented] have shown that the latter parts of the algorithm - once the courses of treatment have been identified - will be appropriate for this work, indicating that there is a class of problems (which may be unrelated medically) to which this algorithm will apply.

A. Limitations

Due to the volume of data processed (across a range of types of radiation therapy, each of which has different rules for claiming according to the MBS), and limited access to expert knowledge, validation to date has necessarily been limited in scope. The DoH is interested in this method, and discussions with experts are ongoing; more in-depth validation is expected to take place, but could take at least 12 months to complete.

Due to tractability issues, pattern lengths were limited to 3. A consequence of that decision is that some patterns will not be identified. Any unusual subsequence involving more than 3 items (whether occurring consecutively, concurrently, or a combination of the two) cannot be identified with patterns of length 3. While this did not seem to unduly impact our results, further investigation should be conducted on detecting longer sequences.

We have not directly compared our method with other common methods for sequence detection, such as Hidden Markov Models (see Section II-D) or recurrent neural networks. The lack of labelled data limits our ability to assess the performance of the model in relation to other methods. Although publicly available datasets could be used for comparison purposes, these are typically unlabelled in terms of the objectives of our approach. We note further that due to rapid concept drift in this type of domain, labels quickly become out of date. For example, within the DoH it is typically necessary to retrain models on an annual basis.

B. Future work

Future work could examine the timing of claims, rather than just the order. Identifying patterns where items are missing from the sequence, rather than added, could also be useful for identifying potential low-value care or non-compliance. Sequential association analysis could be helpful in that regard. Throughout the development and testing of the algorithm, particularly the pattern mining components, choices of parameters were made. These were chosen with a combination of experimentation and discussion with medical advisors, however more rigorous work should be done to determine the sensitivity of these parameters over a reasonable range of values. Further work could also be conducted on the identification of courses of treatment - not just on methods for finding and extracting the courses, but automatically detecting which method might apply or determining, from complete patient data, whether there might be courses of treatment within that data.

C. Conclusion

We presented a generalisable, interpretable algorithm for identifying anomalous patterns in medical claim sequences, with real-world applications as a decision-support system. This method has generated useful information for the Australian Government Department of Health, and exploration of the results is ongoing. Claim sequences and methods for identifying them such as sequential pattern mining are under-explored in the literature, and further work may improve the state-of-the-art with regard to detecting fraudulent and wasteful claims in health insurance.

Acknowledgment

Thanks to Roger Cleary and Val Lim at the Australian Government Department of Health for their contribution to validation of the results.

References

[1] M. Button and J. Gee, "The financial cost of healthcare fraud," University of Portsmouth, 2015. [Online]. Available: https://pure.port.ac.uk/ws/portalfiles/portal/17778636/The_Financial_Cost_of_Healthcare_Fraud_Report_2015.pdf

[2] A. Couffinhal and A. Frankowski, *Wasting with intention: Fraud, abuse, corruption and other integrity violations in the health sector*. OECD Publishing, 2017, pp. 265–301.

[3] U. Sivarajah, M. M. Kamal, Z. Irani, and V. Weerakkody, "Critical analysis of big data challenges and analytical methods," *Journal of Business Research*, vol. 70, pp. 263–286, 2017. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S014829631630488X

[4] A. Abdallah, M. A. Maarof, and A. Zainal, "Fraud detection system: A survey," *Journal of Network and Computer Applications*, vol. 68, pp. 90–113, 2016. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1084804516300571

[5] J. West and M. Bhattacharya, "Intelligent financial fraud detection: A comprehensive review," *Computers & Security*, vol. 57, pp. 47–66, 2016. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0167404815001261

[6] T. Ekin, F. Ieva, F. Ruggeri, and R. Soyer, "Statistical medical fraud assessment: Exposition to an emerging field," *International Statistical Review*, vol. 86, no. 3, pp. 379–402, 2018.

[7] H. A. Buchan, A. Duggan, J. Hargreaves, I. A. Scott, and L. Slawomirski, "Health care variation: Time to act," *Medical Journal of Australia*, vol. 205, no. S10, pp. S30–S33, 2016. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.5694/mja15.01360

[8] T. Badgery-Parker, S.-A. Pearson, K. Chalmers, J. Brett, I. A. Scott, S. Dunn, N. Onley, and A. G. Elshaug, "Low-value care in Australian public hospitals: Prevalence and trends over time," *BMJ Quality and Safety*, vol. 28, no. 3, p. 205, 2019. [Online]. Available: http://qualitysafety.bmj.com/content/28/3/205.abstract

[9] S. M. Weiss, C. A. Kulikowski, R. S. Galen, P. A. Olsen, and R. Natarajan, "Managing healthcare costs by peer-group modeling," *Applied Intelligence*, vol. 43, no. 4, pp. 752–759, 2015. [Online]. Available: ⟨GotoISI⟩://WOS:000364385100004

[10] R. Zhao, R. Yan, Z. Chen, K. Mao, P. Wang, and R. X. Gao, "Deep learning and its applications to machine health monitoring," *Mechanical Systems and Signal Processing*, vol. 115, pp. 213–237, 2019. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0888327018303108

[11] Australian Government Department of Health, "Medicare benefits schedule," 2021.

[12] P. Fournier Viger, C.-W. Lin, U. Rage, Y. S. Koh, and R. Thomas, "A survey of sequential pattern mining," *Data Science and Pattern Recognition*, vol. 1, pp. 54–77, 02 2017.

[13] H. Farbmacher, L. Löw, and M. Spindler, "An explainable attention network for fraud detection in claims management," *Journal of Econometrics*, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0304407620302852

[14] V. Snorovikhina and A. Zaytsev, *Analysis of Images, Social Networks and Texts*. Springer International Publishing, 2020, ch. Unsupervised anomaly detection for discrete sequence healthcare data.

[15] Cancer Council NSW, "Understanding radiation therapy," Online, Level 14, 477 Pitt Street, Sydney NSW 2000, 2021. [Online]. Available: https://www.cancer.org.au/assets/pdf/understanding-radiation-therapy-booklet

[16] M. Ivanović and Z. Budimac, "An overview of ontologies and data resources in medical domains," *Expert Systems with Applications*, vol. 41, no. 11, pp. 5158–5166, 2014. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0957417414001213

[17] B. M. Konopka, "Biomedical ontologies — a review," *Biocybernetics and Biomedical Engineering*, vol. 35, no. 2, pp. 75–86, 2015. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0208521614000503

[18] P. Fournier-Viger, P. Yang, R. U. Kiran, S. Ventura, and J. M. Luna, "Mining local periodic patterns in a discrete sequence," *Information Sciences*, vol. 544, pp. 519–548, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0020025520309567

[19] P. Fournier-Viger, A. Gomariz, M. Campos, and R. Thomas, "Fast vertical mining of sequential patterns using co-occurrence information," in *Advances in Knowledge Discovery and Data Mining*, V. S. Tseng, T. B. Ho, Z.-H. Zhou, A. L. P. Chen, and H.-Y. Kao, Eds. Cham: Springer International Publishing, 2014, pp. 40–52.

[20] L. Antwarg, R. M. Miller, B. Shapira, and L. Rokach, "Explaining anomalies detected by autoencoders using shapley additive explanations," *Expert Systems with Applications*, vol. 186, p. 115736, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0957417421011155

[21] E. Plischke, G. Rabitti, and E. Borgonovo, "Computing Shapley effects for sensitivity analysis," *SIAM/ASA Journal on Uncertainty Quantification*, vol. 9, no. 4, pp. 1411–1437, 2021.