Australasian Medical Journal [AMJ 2012, 5, 12, 609-618]

Respiratory knowledge discovery utilising expertise

Tristan Ling

School of Pharmacy, University of Tasmania, Australia

RESEARCH

Please cite this paper as: Ling T. Respiratory knowledge discovery utilising expertise. AMJ 2012, 5, 12, 609-618. http://doi.org/10.21767/AMJ.2012.1354

Corresponding Author:
Dr Tristan Ling
School of Pharmacy, University of Tasmania
Tasmania
Australia
Email: Tristan.Ling@utas.edu.au

Abstract

Background

Significant amounts of medical data are being archived, in the hope that they can be analysed and provide insight. A critical problem with analysing such data is the amount of existing knowledge required to produce effective results.

Aims

This study tests a method that seeks to overcome these problems with analysis, by testing it over a large set of archived lung function test results. A knowledge base of lung function interpretation expertise has been compiled and serves as a base for analysis.

Method

A user examines the dataset with the assistance of the knowledge discovery tool. Two pertinent respiratory research questions are analysed (the relative correlation between diffusing capacity and FEV_1 or FVC bronchodilator response, and the effects of BMI on various parameters of lung function), and the results compared and contrasted with relevant literature.

Results

The method finds interesting results from the lung function data supporting and questioning other published studies, while also finding correlations that suggest further areas of research.

Conclusion

While the analysis does not necessarily reveal groundbreaking information, it shows that the method can successfully discover new knowledge and is useful in a research context.

Key Words

Lung function, knowledge discovery, knowledge acquisition, data mining, MCRDR

What this study adds:

1. Knowledge discovery in medical areas needs extensive knowledge acquisition
2. Results of lung function data analysis for relationship between reversibility and D_LCO, and effects of high BMI on lung function
3. Knowledge discovery and knowledge acquisition can be combined

Background

The medical field presents unique challenges and benefits for discovering new knowledge.[1] To this end, archives of medical data are continually being added to in the hope that analysis of this data may provide valuable insights.[1,2] However, this data is difficult to analyse: it includes a large number of measurements of a variety of types, and requires extensive existing knowledge to be meaningfully interpreted.[1,3,4]

The study and treatment of the respiratory system is a typically complex medical field, dealing with one of the most critical components of the human body. There are a wide range of tests used to gather data on functioning lungs, generally divided into categories based on which aspect of lung function is being measured, and the inherent cost and difficulty of performing the tests.[5] The most common series of tests are spirometry tests, which concern the lungs' ability to move gas over time. These are often measured twice, with a bronchodilator administered between measurements which relaxes the airways into the lungs.[5] The second category of tests is static spirometry, or lung volumes, which attempt to measure the capacity of the lungs.[6,7] The final category of tests are the diffusing capacity tests, which measure gas transfer between the lungs and the blood.[8] Interpreting these test results is a complicated process that requires training and experience: specialists are employed to analyse patient test results and reach conclusions about the state of the patient's lungs. These interpretations typically then become one component in determining a diagnosis and treatment.

For all tests, the importance of a given value can vary significantly depending on the physiology, ethnicity, and living conditions of the patient being tested.[9,10] This is ameliorated by using data such as age, height, weight, and sex, with pre-determined reference equations, to calculate expected or predicted values for each of the tests. The discrepancies between measured values and expected values then becomes a primary source of information.[9] The reference equations are intermittently derived from large-scale studies into different healthy populations. [10-13]

Knowledge discovery

The term knowledge discovery describes any process for discovering new information about a subject using some combination of recorded data on the subject and existing knowledge about that data.[14] Effective knowledge discovery is a complex, multi-stage process of analysis and processing.[15,16] The data analysis or data mining stage has been heavily researched, with many methods available for finding patterns in data.[17,18] These methods encounter difficulty in analysing data when there is a large or complex existing body of knowledge about the meaning of that data.[19-21] A frequently stated problem is that the knowledge acquisition and engineering required makes useful analysis impractical.[19,22,23] This problem is addressed by the identification and incorporation of knowledge in the initial stages.[15,16] However, few methods have been developed to achieve this,[21] particularly for domains or applications of a realistic complexity.[23,24] A further concern is that the knowledge base needs to be updatable and extendable, and receive input from multiple sources, in order to ensure useful accuracy and longevity.[19,20] Acquiring and maintaining this knowledge is a difficult process.

Knowledge acquisition

Ripple Down Rules (RDR) are an approach to acquiring and encoding expertise developed by Compton and Jansen.[25] In their work developing and maintaining expert systems, knowledge was encoded as a set of rules, each describing a conclusion to apply if the current case (scenario) met a series of conditions (a simplistic example of such a rule: If [BMI > 30] Then [Obese]; although rules are typically more complex, such as: If [$FEV_1/FVC < 0.7$] And [FEV_1 % of predicted ≥ 80] And [TLC % of predicted ≥ 80] Then [Mild Obstruction].). Based on their experiences with such systems, they made the observation that no rule can be guaranteed to be correct, with almost every rule subject to revision given enough time in use.[26] They therefore based their knowledge acquisition process on the idea that knowledge was context-dependent and would inevitably require correcting over time. The process functions on a case-by-case basis: the system is presented with a case and attempts to interpret it, based on its knowledge base, and provide its reasoning. If an expert disagrees with the system's conclusions, the system is updated with a new rule, based on the expert's justifications for why the system was wrong. This new rule is added within the context of the incorrect rule, overriding that rule when conditions are met, and stored with the evidence (case) which raised the issue. To maintain integrity and ensure that knowledge is never lost, rules are only ever added to the knowledge base, never modified or removed.[26] This technique was later adapted to a more generally applicable Multiple Classification RDR (MCRDR) structure, which allowed multiple conclusions (interpretations) to be reached for each given case with minimal loss of efficiency.[27] Under both RDR models, cases are considered sequentially and individually, and as much knowledge as possible is extracted from a case before moving onto the next. In 2006 Vazey presented a predictive model for the case-driven acquisition of knowledge,[28] concluding that to be most effective, knowledge acquisition should be Case And Rule-Driven (CARD). The CARD methodology allows knowledge to be added via a top-down, rule-driven approach, or via a bottom-up case-driven approach. Vazey found benefits to rule-based development and concluded that a combination of the two is likely to provide the best results.[29] The fundamental benefits of MCRDR were maintained: that the knowledge base can be easily corrected or added to without risking validity, and can be developed by a domain expert without needing assistance.

Method

The MCRDR method provides incremental knowledge acquisition that has been shown to help resolve problems which have excessive knowledge acquisition and engineering requirements.[30,31] The MCRDR-encoded form of the knowledge base is also eminently extensible and accessible to programmatic manipulation, and so this was chosen as a logical choice to build the knowledge base. This study used a modified MCRDR method, using the combined case and rule based approach developed by Vazey. The resultant knowledge base was used for decision support in the interpretation of lung function data, and to test a new method for the comparison ad consolidation of different experts' knowledge; but its relevance to this study was its use as a framework for knowledge discovery.

The MCRDR knowledge base was compiled from input by a leading respiratory scientist in Australia, another highly regarded clinical specialist, and another respiratory researcher. The knowledge was expressed by the expert defining their interpretations and conclusions for a series of cases, while providing justifications for which features of the case led to their decisions. The first expert began by defining rules independently of cases, and the knowledge

base was subsequently added to by other experts in the typical MCRDR style. Once completed, inconsistencies were resolved where necessary and consolidated into a final knowledge base.

The lung function data used in this study was an amalgamation of lung function case reports from Austin Health in Melbourne, Australia; the 2004 round of the Tasmanian Longitudinal Health Study (TAHS); and reports from the Royal Hobart Hospital in Hobart, Australia. All cases had any personally identifying data removed. Each constituted a single set of test results from a single patient, independent of history, resultant diagnosis or any other information. Test results were numeric values (excluding values such as sex and current smoking status), and were without any graphs (such as Flow Volume Loops) or notes.

Knowledge discovery system

The aim of the knowledge discovery system was to allow a user to explore a dataset and test hypotheses, while being notified of any relevant patterns within the data. This consisted of four components: a database of cases; a knowledge base of rules containing additional information about those cases; a composite of functions for calculating statistics and mining interesting relationships in the data; and an online interface for a user to explore these elements and direct the functions. As there are currently no generic software packages for this purpose, the software was custom-written for the research project (although the core of the knowledge acquisition and discovery code is generalised and applicable to other tasks). To use the system, the user described the type of patient case they were interested in examining, either by directly specifying ranges for particular attributes, or by specifying case interpretations or conclusions that the system was aware of. This functionality was offered in the online system through a simple web form. The system displayed those cases with summary statistics, and highlighted any correlated groups of cases, attribute ranges, or defined conclusions, using the data mining and statistical components. The user could further investigate by changing the selection criteria, and could at any time categorise a group of cases with a conclusion/interpretation. The system would then know how to reach such a conclusion for a case in the future, by using the case selection criteria as a definition. This feeds back into the knowledge base, expanding the scope of the system and its effectiveness at interpreting cases, giving it more information for pattern analysis.

The data mining techniques used were primarily taken from association rule mining, information theory, and probability measurements, given their applicability to exploratory data analysis and data mining in general.[32-34] These measures were chosen for their ease of calculation, such that the system was sufficiently responsive with a large database in an online environment, and for their ease of interpretation. The measures were categorised into two areas. The first, class statistics, displayed information on the knowledge-base-derived interpretations present in the current set; showing the number of cases in the current set with each conclusion, and the number of cases expected to have that conclusion, based on the ratio within a chosen base dataset (defaulting to the complete dataset). Additionally traditional data mining measures were also calculated: the confidence, gain, and Piatetsky-Shapiro gain (hereafter PS-gain).[33,35-37] In addition to showing the values, the user could specify thresholds of interestingness for each; if a threshold was reached for a given conclusion, the conclusion was marked as "interesting" and highlighted to the user, with the strength of highlight corresponding to the strength of correlation. Links to further details for any given set of cases were also provided.

The second area of measures concerned attribute statistics, grouped by Patient Details, containing attributes such as sex, age, and height; Common Spirometry containing those spirometric measurements most commonly used by experts, such as FEV_1, FVC and FEF_{25-75}; Other Spirometry containing 27 less commonly used spirometric measurements; Lung Volumes; Gas Transfer; and the source of the case. For each numeric attribute the minimum, maximum, mean, and the standard deviation were listed. For nominal attributes the frequency of each allowable value was listed, and for comparison, how frequently it occurred in the base dataset. Attributes were also tested for interestingness, tested through various measures. The conditions chosen by the user to select a set of cases were combined with the range of values for the current attribute to define an association rule[38] in the form $(current\ search\ conditions) \rightarrow (attr_{min} \leq x \leq attr_{max})$. Attributes were also rated with confidence, gain, and PS-gain. For numeric attributes, the z-score of the mean for the current set was also calculated and compared with a user-defined threshold. Interesting attributes were highlighted and the magnitude presented in simple terms, for example "Cases in this range have your conditions 50% more often than expected". Information gain[39-41] was also calculated for nominal attributes to determine if the user's conditions were a good predictor for each value. For numeric attributes, the system used information gain to find the optimal range for predicting the current rule conditions.

Results

To test that the method could successfully discover new knowledge from complex data, the system was used to resolve and expand on research questions that were raised

in respiratory literature, and to answer questions that were suggested by the experts involved as interesting topics to consider and answer through data analysis. The results of two of these questions are presented here, although other tests were performed, including: examining the general distribution and form of lung function for subjects who met the ATS/ERS reversibility guidelines (explained in the following section); and whether, and to what extent, Alveolar Volume can be used to estimate Total Lung Capacity in patients with airflow obstruction. Respiratory specialists were consulted throughout in order to ensure accuracy and a correct understanding, and to interpret results where necessary. Each data analysis study was performed in a single session, including analysing any relevant previous studies and interpreting the results; the system logs showed a mean time of slightly less than three hours of use per study.

In subjects who met the ATS/ERS criteria for reversibility is the correlation with D_LCO (Diffusing Capacity of Carbon Monoxide) stronger in the FVC (Forced Vital Capacity) responders than FEV_1 (Forced Expiratory Volume over 1 Second) responders?

The American Thoracic Society (ATS) and the European Respiratory Society (ERS) define a positive bronchodilator response as an increase post-bronchodilator of at least 12% and 0.2L for either FEV_1 or FVC,[42] conclusions which had already been defined by the experts through MCRDR knowledge acquisition. The system indicated a significant relationship between cases showing either FEV_1 Reversibility or FVC Reversibility and those showing Low D_LCO (another expert-defined conclusion), and as another recent study found, the relationship with FEV_1 responders is stronger.[43] Examining further, the absolute change of FEV_1 after bronchodilators (FEV_1 Δ) showed a strong Pearson correlation with all D_LCO measurements (R=0.527 for uncorrected D_LCO, p<0.0001), but with no significant correlation between the percentage change of FEV_1 (FEV_1 % Δ) and any D_LCO measurement. A weaker but still significant correlation was shown between the percentage improvement of FVC and D_LCO measurements, (R=0.295 for uncorrected D_LCO, p<0.0001, with stronger correlations between other D_LCO measurements), although there was no significant correlation shown between absolute change of FVC and D_LCO. The stronger relationship between FEV_1 Reversibility and D_LCO, than between FVC Reversibility and D_LCO, also matched the findings of Agahi.

The automated analysis also showed that FEV_1 responders with Obstruction had a statistically significant relationship to the conclusions Gas Trapping, Hyperinflation and Low D_LCO, as shown in Table 1.

Table 1: Classes showing the strongest association to cases with FEV_1 Reversibility, for the 150 cases with Obstruction

Class	Cases	*PS-gain*	*p*
Gas Trapping	42 (28%)	24 (58%)	< 0.0001
Hyperinflation	34 (22.7%)	16 (47.6%)	< 0.0001
Low D_LCO	88 (58.7%)	41 (46.8%)	< 0.0001

Table 2: Some of the attributes indicated as most related to the FEV_1 Reversibility class, for the 150 cases with Obstruction

Attribute	Expected Mean	Actual Mean	Std Deviations
D_LCO (Hb corrected)	14.78	18.16	0.5
D_LCO % of predicted (Hb corrected)	55.81%	65.37	0.4
V_A % of predicted	89.98%	98.14%	0.4

There were 108 cases in the dataset with both FVC Reversibility and Obstruction. The class comparison between those cases and cases with Obstruction are shown in Table 3. Two of the same classes were identified as with FEV_1 Reversibility, although each to a lesser extent. Both were supported when looking at the attribute correlations (summarised in Table 4): cases with FVC Reversibility showed an association with RV, which increased from an expected mean of 130.51% of predicted to 150.57%. Cases with FEV_1 Reversibility showed no appreciable change in expected RV. In examining diffusion, cases with FVC Reversibility showed a small reduction in mean uncorrected D_LCO, dropping from 72.31% of predicted to 63.15%, although with a much smaller drop in corrected D_LCO (55.81% to 51.18%). This is the reverse of subjects with FEV_1 Reversibility which showed an increase in both those measurements. The differences are not conclusive, being no larger than half a standard deviation in either case. Further analysis showed that for obstructed cases, the percentage of FEV_1 change bears no significant correlation to the D_LCO (expressed as a percentage of the predicted value); whereas the percentage change of FVC showed some association, with a stronger correlation for cases that have FVC Reversibility without FEV_1 Reversibility (correlation = -0.32474, $p < 0.05$).

Table 3: Classes showing the strongest association to cases with FVC Reversibility, for the 108 cases with Obstruction

Class	Cases	*PS-gain*
Evidence of Gas Trapping	27 (25%)	14 (52.9%)
Low D_LCO	53 (49.1%)	19 (36.4%)

Table 4: Some of the attributes indicated as most related to the FVC Reversibility class, for the 108 cases with Obstruction

Attribute	Expected Mean	Actual Mean	Std Deviations
D_LCO (Hb corrected)	14.78	18.16	0.5
D_LCO % of predicted (Hb corrected)	55.81%	65.37	0.4
V_A % of predicted	89.98%	98.14%	0.4

A summary of several mean attribute comparisons between cases in the classes FEV_1 Reversibility and FVC Reversibility are presented in Table 5. The results support previous indications that cases with FVC Reversibility have generally lower values for spirometry tests than cases with FEV_1 Reversibility. Interestingly the mean diffusing capacity (D_LCO) is worse in FVC Reversible patients than FEV_1 Reversible, even though the correlation between FEV_1 Reversibility and D_LCO was stronger than FVC Reversibility and D_LCO.

Effects of BMI on lung function

As obesity is currently such a major health issue in the world today,[44] there is an increasing rate of studies trying to identify the effects of overweight and obesity on all aspects of lung function. Jones and Nzekwu performed a study into the effects of BMI on lung volumes,[45] and Stritt and Garland studied the effects of obesity on volumes and spirometry.[46] These studies are by no means the only examples of such work. Given these investigations, this section will focus on using this system to reproduce the results of those studies mentioned, and on examining what other information the data may provide.

Previous study results

Jones and Nzekwu's study collected results for 373 patients both male and female with a range of BMIs, but with a number of other fixed criteria, including: over 18 years of age; an FEV_1/FVC ratio over 90% of predicted; a RV less than the upper limit of normal; D_LCO above the lower limit of normal, when adjusted for VA. Results were analysed using linear or exponential regression, and analyses of variance.[45]

Table 5: Significant differences between attribute means for cases with FEV_1 Reversibility and cases with FVC Reversibility

Attribute	FEV_1 Reversible	FVC Reversible
FEV_1% Pred Pre-BD	63.35%	57.46%
FEV_1% Pred Post-BD	75.64%	65.59%
FVC% Pred Pre-BD	82.49%	72.11%
FEV_1 / FVC Pre-BD	0.62	0.59
FEV_1 / FVC Post-BD	0.68	0.57
FEF_{25-75}% Pred Post-BD	61.84%	46.87%
FEF_{25} Pre-BD	3.56	2.29
FEF_{25} Post-BD	4.59	2.66
FEF_{75} Pre-BD	0.47	0.31
FEF_{75} Post-BD	0.65	0.31
FEV_3 Post-BD	3.01	2.2
SVC Post-BD	1.3	0.22
V_A / TLC	0.84	0.78
D_LCO% Predicted	67.72%	53.54%

The study found linear relationships between BMI and VC, and between BMI and TLC, but without a significant change in either mean. FRC and ERV decreased exponentially as BMI increased, with the greatest rate of change in patients overweight or with mild obesity: at a BMI of 30, FRC was at 75% of the value of a person with a BMI of 20, and ERV at 47%.[45]

Stritt and Garland's study identified a lack of information regarding correlations between specific BMI levels and their effect on lung volumes, and even less information on correlations with spirometry. Patients were selected according to a series of criteria: a FEV_1/FVC ratio at least equal to predicted; a D_LCO at least 70% of predicted; and no evidence of respiratory muscle weakness. Patients were then grouped according to BMI.[46] Reported results were that the 13 patients with a BMI below 30 displayed a mean TLC of 93% of predicted, a mean FEV_1 at 73% of predicted, and a mean FVC at 77%; whereas the 10 patients with a BMI at 30 or above showed 81%, 67%, and 70% respectively.[46]

Testing procedures

Classes were defined for five BMI intervals: Underweight (BMI < 20), Normal Weight (20 ≤ BMI < 25), Overweight (25 ≤ BMI < 30), Obese I (30 ≤ BMI < 35), Obese II (35 ≤ BMI < 40), Obese III (40 ≤ BMI < 45), and Obese IV (BMI ≥ 45). The

following criteria were then added to remove extraneous factors: Age > 18; without Obstruction (FEV_1/FVC ratio ≥ 0.7); with an FEV_1/FVC ≥ 90% of predicted; normal D_LCO (above 80% of predicted); and RV < 120% of predicted.

New results

No class associations were indicated for any of the BMI categories defined. Table 7 summarises the mean values for volume measurements, FEV_1 and FVC over each category (SVC showed no substantial or consistent change). Mean TLC showed a relatively consistent linear downward trend culminating in a 7.07% (0.7 standard deviations) drop in mean percent of predicted between normal BMI and a BMI above 40. FRC showed a strong decrease as BMI increased, dropping consistently until the last group. This trend also seemed to continue in the opposite direction, with the group of patients with a BMI below 20 showing a higher FRC. Using all BMI groups, comparing BMI to FRC (expressed as a percentage of predicted) gave a correlation coefficient of -0.38; although FRC appeared to improve slightly as BMI became very high indicating it may not be a linear trend. ERV showed a similar relationship, including the Underweight group. Unfortunately a lack of ERV prediction equation in the data meant no percentage of predicted value was available. The ERV data also showed a slight improvement as BMI becomes very high, again possibly suggestive of a non-linear trend.

The numbers show a small but insignificant decrease in RV as BMI increased. Subjects with a BMI between 25 and 40 showed the largest decrease, although still only a drop from 93.05% to 90.22%. IC showed a consistent increase as BMI increased, although subjects with a BMI above 45 displayed a sufficiently smaller mean. Underweight subjects also continued the trend with a decreased value. Although not displayed in the table, V_A/TLC showed a trend similar to other attributes, with a very small decrease from 0.95 to 0.9 between Normal Weight and Obese III, with Obese IV subjects showing an increase to 0.97.

These results match those in the Jones study, identifying similar relationships between TLC, FVC, and ERV. Notably the results here also identified the slight improvement for very high values of BMI, although the Jones study identified this improvement at slightly lower BMI levels.

In comparing the results to those of the Stritt study, there was no equivalent relationship found between BMI and TLC: Stritt found a drop from 93% of predicted to 81% when comparing patients without obesity to those with. The results here did show a negative correlation, but not to the extent described by Stritt and Garland. Comparing Normal Weight subjects to those with a BMI above 45 found at most a drop from 104.3% to 97.3% (0.7 standard deviations). Comparing subjects with a BMI above 30 to those below, as in the Stritt study, again showed only a slight decrease as shown in Table 6. The general trend in this data appears to be a slight negative correlation between BMI and TLC, but nothing in the order of the data in Stritt's results.

Similarly, the system showed no significant change in the FEV_1. Closer analysis showed a trend of a slight decrease as BMI increased, with a small increase for very high BMI, as shown in Table 6. A direct comparison of BMI < 30 to BMI > 30 again showed a slight drop (0.36 standard deviations). Hence again, a trend seems to be evident, but not in the strength reported by Stritt. A very similar trend is apparent for FVC, although slightly more pronounced and with no increase as BMI becomes very large, although the rate of decrease slows significantly. The trend becomes smaller for post-bronchodilator FVC however (correlation coefficient -0.169). The percentage change of FVC also appeared to be correlated with BMI, increasing as BMI increases with a correlation coefficient of 0.206. FEF_{25-75}, FEV_1/FVC ratio, PEF, and FEV_1 post-bronchodilator change showed no correlation.

Table 6: A comparison of Stritt and Garland's results[46] to those found from this data

		TLC % pred.	FEV_1 % pred.	FVC % pred.
Stritt and Garland	**BMI < 30**	93%	73%	77%
	BMI ≥ 30	81%	67%	70%
This study	**BMI < 30**	103.7%	99.73%	100.24%
	BMI ≥ 30	100.67%	95.47%	93.7%

Discussion

The discovered results of the analyses presented are not necessarily groundbreaking, but this was expected as the studies were performed by a non-expert in the field, with very little experience or knowledge of available literature and domain knowledge. Importantly however, the results were new to the user, expanding the user's understanding of the data and the field. That the results were developed independently of similar conclusions, after only a few hours of analysis, indicate that the approach can be used to successfully, and efficiently, discover new knowledge.

Rather than discovering revolutionary knowledge for lung function, the primary goal of this study was to demonstrate the application and efficacy of the system. Of course this does not make the results irrelevant: the results provide useful evidence supporting or expanding on the results of other lung function studies, and provide direction for future

research. The results of the BMI analysis may prove useful given the scarcity of currently published results (relative to the interest in the area), and the small number of subjects used in many of the relevant studies. It should be noted however that there are issues to be considered in generalising results found by this approach.

A known flaw in the method used is the use of fixed predicted values in interpretations, as the international standard is now to use statistically derived, individually calculated normal limits. This flaw was unfortunately identified too late in development to alter the outcome. Any further development of the method would certainly incorporate statistically derived limits of normal. However, this flaw does not detract from the demonstration of the applicability and efficacy of the approach. If incorporated the use of limits of normal would not cause any change in the computation time of the system or in its use, except perhaps to make the experts' task simpler. Other potential flaws in the study concern the dataset used. In order for results to be considered representative, this approach needs a large set of cases to work from. The generic dataset used in this study would not be suitable to test ideas about particularly specialised groups; for example, there are only two patients with a BMI above 30 and Restriction. Furthermore the generic database does not have all the data points that a specialised study might require; for example many cases lacked D_LCO values corrected for haemoglobin, or the time of day at which the tests were performed. Extra information can certainly be added, and is not necessarily a limitation of the approach, but is a potentially limiting factor that must be considered in evaluating the significance of these results.

It is also worth noting that the calculations used were chosen for their simplicity to ensure interface responsiveness, a necessity as the system was accessed through an online interface and with a simple, inexpensive web server. More complex interestingness measures could be adopted by calculating them on specific request for a section of data (as the current system did with the optimal range information gain calculations), by implementing the system offline, or by simply improving the hardware.

Given the limitations, it is suggested that the current system is best used as a source of preliminary analysis: testing a hypothesis against a store of data and existing knowledge to verify a suspected pattern. The BMI studies provide examples of this sort of use. Another useful application of the current system is to find supporting evidence for existing studies that require additional data. A good example of this is Stritt and Garland's study,[46] which presented interesting findings but based on small numbers of subjects (comparing a group of 13 patients to a group of 10). By testing their hypothesis with this system stronger supporting evidence was found, with discrepancies in the actual findings suggesting that a larger study is required.

Knowledge acquisition

A significant feature of this knowledge discovery approach is the integration of a knowledge acquisition method, meaning that the analysis results and any information generated from the process all feed into the existing knowledge base. This automatically adds to the store of data available for analysis, improving the effectiveness of both the knowledge discovery process and the detail of the expert classification system. This is shown by the results of the data analysis studies: the definition of classes such as FEV_1 reversibility and FVC reversibility add this information to the cases which would not have been identified in any other way, yet which can provide significant information on the relationships between attributes and their meaning – in this case, that a capacity to reverse FEV_1 has a stronger correlation to a reduced diffusing capacity than FVC reversibility, among other correlations. This additional information is also retained, such that if any future data analysis study defines any set which has a significant relationship to these classes, this will be displayed to the user, thus adding to the information discovered about the new study.

The necessity for incorporating domain knowledge is the biggest challenge to data mining for this data. Without being able to incorporate the level of knowledge used in performing the analysis here, the amount of time required to determine the usefulness of the results would be prohibitive.[19,47,48] Examining the relationships that were discovered during the testing of the method, it is apparent that interestingness measures alone do not indicate that a relationship is important to the user. For example, in order to conclude that relationship of increased BMI and decreased diffusion was significant depended not only on identifying a measurable correlation, but that that correlation was not expected. In examining the various BMI classes for significant trends, an average of 43 attributes per class were identified by the system; of these, approximately a quarter were considered interesting and examined further, based on a tacit understanding of expectations, and an understanding of the dependencies between attributes. Were the system to automatically perform detailed analysis of all of these attributes, the resultant list of "interesting" patterns would be far too large to be easily reviewed. The author is not aware of any knowledge discovery techniques that can sufficiently account for the issue of expectation.

Further application

In considering adopting this technique for other domains and datasets, the number of cases used is an important

consideration. While no tests have been performed to identify an optimal number, it is suggested that for a domain similar to this study, at least a few thousand cases are needed for analysis. This is a very rough estimate however: the number is of little importance except to try and ensure that the cases are representative of the domain. This is, however, of particular importance here as the data is not only used for analysis but also to acquire the knowledge base. The acquisition generally requires at least a few hundred cases, but the results will become more detailed and accurate as more cases are used.

Conclusion

The results of the studies performed show that this method can successfully perform knowledge discovery tasks in a complex field such as lung function. That the analysis of this data was performed by a relatively inexpert user, in a very short time frame, indicates that the approach has a reasonable level of efficiency and simplicity. The analysis for each of the topics was performed without any specialised preparation of subjects or a clinical study, and with very little individual preparation.

It is known that the necessity for complex domain knowledge in effective data mining is a major difficulty for data of the complexity seen here. This is relevant to the solution presented in this method in two ways. Firstly, the incorporation of a knowledge acquisition process, so that the expert can easily formalise their expectations: the system can then show the relationship between that classification, and the results of the data analysis. Secondly, this approach does not overwhelm the expert with results that may or may not be significant, and instead affords greater control over the analysis that is performed. This uses the tacit knowledge that the expert has, which is very difficult to otherwise access and encode into a knowledge base.

Further Information

For further details of the derived rules, or access to the system or code, please contact the author at Tristan.Ling@utas.edu.au.

References

1. Cios K, Moore GW. Uniqueness of medical data mining. Artif Intell Med. 2002;26(1-2):1-24.

2. Roddick J, Fule P, Graco W. Exploratory medical knowledge discovery: Experiences and issues. ACM Special Interest Group on Knowledge Discovery and Data Mining (SIGKDD) Explorations. 2003;5(1):94-9.

3. Cios K, Kacprzyk J. Medical data mining and knowledge discovery. Denver: Physica-Verlag; 2001.

4. Prather J, Lobach D, Goodwin L, Hales J, Hage M, Hammond W, editors. Medical data mining: knowledge discovery in a clinical data warehouse. 1997 Annual Conference of the American Medical Informatics Association; 1997; Philadelphia: American Medical Informatics Association. 101.

5. Pellegrino R, Viegi G, Brusasco V, Crapo R, Burgos F, Casaburi R, et al. Interpretative strategies for lung function tests. Eur Respir J Suppl. 2005;26(5):948.

6. Laszlo G. Pulmonary Function: A Guide for Clinicians. Cambridge: Cambridge University Press; 1994. 245

7. Miller A. Pulmonary Function Tests. Orlando: Grune and Stratton; 1987. 275

8. Ruppel GL. Manual of Pulmonary Function Testing. 7 ed. St Louis: Mosby; 1994.

9. Collen J, Greenburg D, Holley A, King C, Hnatiuk O. Discordance in spirometric interpretations using three commonly used reference equations vs National Health and Nutrition Examination Study III. Chest. 2008;134(5):1009.

10. Subbarao P, Lebecque P, Corey M, Coates A. Comparison of spirometric reference values. Pediatr Pulmonol. 2004 Jun;37(6):515-22.

11. Crapo R, Morris A. Standardized single breath normal values for carbon monoxide diffusing capacity. Am Rev Respir Dis. 1981 Feb;123(2):185-9.

12. Goldman H, Becklake M. Respiratory function tests; normal values at median altitudes and the prediction of normal results. Am Rev Tuberc. 1959 Apr;79(4):457-67.

13. Hankinson J, Odencrantz J, Fedan K. Spirometric reference values from a sample of the general US population. Am J Respir Crit Care Med. 1999 Jan;159(1):179-87.

14. Goebel M, Gruenwald L. A survey of data mining and knowledge discovery software tools. SIGKDD Explorations. 1999;1(1):20-33.

15. Fayyad U, Piatetsky-Shapiro G, Smyth P. From data mining to knowledge discovery in databases. AI Magazine. 1996;17(3):37.

16. Kurgan LA, Musilek P. A survey of Knowledge Discovery and Data Mining process models. The Knowledge Engineering Review. 2006;21(01):1-24.

17. Brachman R, Anand T, editors. The Process of Knowledge Discovery in Databases: A Human-Centered Approach. Advances in Knowledge Discovery and Data Mining; 1996; California: Menlo Park. AAAI Press. Calif. 37-58.

18. Witten IH, Frank E. Data Mining: Practical machine learning tools and techniques. 2 ed. San Francisco: Morgan Kaufmann; 2005.

19. Liu B, Hsu W, Chen S, editors. Using general impressions to analyze discovered classification rules. 3rd International Conference on Knowledge Discovery and Data Mining; 1997; Newport Beach, California: AAAI. 31-6.

20. Piatetsky-Shapiro G, Matheus C, Smyth P, Uthurusamy R. Kdd-93: Progress and challenges in knowledge discovery in databases. AI Magazine. 1994;15(3):77.

21. Sinha AP, Zhao H. Incorporating domain knowledge into data mining classifiers: An application in indirect lending. Decision Support Systems. 2008;46(1):287-99.

22. Kotsifakos EE, Marketos G, Theodoridis Y. A framework for integrating ontologies and pattern-bases. In: Nigro HO, Cisaro SG, Xodo D, editors. Data Mining with Ontologies: Implementations, Findings, and Frameworks Information Science Reference: Idea Group Inc., Hershey; 2008.

23. Zhang C, Yu PS, Bell D. Domain-driven data mining. IEEE Transactions on Knowledge and Data Engineering. 2009;21(2):301.

24. Adejuwon A, Mosavi A. Domain Driven Data Mining–Application to Business. International Journal of Computer Science Issues. 2010;7(4).

25. Compton P, Jansen R, editors. A Philosophical Basis for Knowledge Acquisition. European Knowledge Acquisition for Knowledge-Based Systems; 1989; Paris. 75-89.

26. Compton P, Jansen R. Knowledge in context: A strategy for expert system maintenance. AI'88. 1990:292-306.

27. Kang B, Compton P, editors. Knowledge Acquisition in Context: the Multiple Classification Problem. Proceedings of the Pacific Rim International Conference on Artificial Intelligence; 1992; Seoul. 847-54.

28. Vazey M. Stochastic foundations for the case-driven acquisition of classification rules. Managing Knowledge in a World of Networks. 2006:43-50.

29. Vazey M, Richards D. Evaluation of the FastFIX prototype 5Cs CARD system. Advances in Knowledge Acquisition and Management. 2006:108-19.

30. Kang B. Validating Knowledge Acquisition: Multiple Classification Ripple Down Rules. Sydney: PhD Thesis, University of New South Wales; 1996.

31. Kang B, Compton P, Preston P, editors. Multiple Classification Ripple Down Rules: Evaluation and Possibilities. Proceedings 9th Banff Knowledge Acquisition for Knowledge Based Systems Workshop; 1995 Feb 26 - March 3; Banff. 17.1-.20.

32. Creighton C, Hanash S. Mining gene expression databases for association rules. Bioinformatics. 2003;19(1):79.

33. Lenca P, Vaillant B, Lallich S, editors. On the robustness of association rules. 2nd IEEE International Conference on Cybernetics and Intelligent Systems and Robotics, Automation and Mechatronics; 2006; Bangkok, Thailand: IEEE. 596-601.

34. Marinica C, Guillet F, editors. Improving Post-Mining of Association Rules with Ontologies. 13th International Conference on Applied Stochastic Models and Data Analysis; 2009; Vilnius, Lithuania. 76-80.

35. Bayardo RJ, Agrawal R, editors. Mining the Most Interesting Rules. 5th ACM Special Interest Group on Knowledge Discovery and Data Mining (SIGKDD) Conference; 1999; San Diego, California. 312219: ACM Press. 145-54.

36. Brin S, Motwani R, Ullman J, Tsur S, editors. Dynamic itemset counting and implication rules for market basket data. 1997 International Conference on Management of Data; 1997: ACM New York. 255-64.

37. Piatetsky-Shapiro G. Discovery, Analysis, and Presentation of Strong Rules. Knowledge Discovery in Databases. 1991:229-38.

38. Agrawal R, Imielinski T, Swami A. Mining association rules between sets of items in large databases. Proceedings of the 1993 ACM SIGMOD international conference on Management of data. Washington, D.C., United States: ACM Press; 1993. p. 207-16.

39. Freitag D. Machine Learning for Information Extraction in Informal Domains. Machine Learning. 2000;39(2):169-202.

40. Kent J. Information gain and a general measure of correlation. Biometrika. 1983;70(1):163.

41. Kullback S, Leibler R. On information and sufficiency. The Annals of Mathematical Statistics. 1951;22:79-86.

42. American Thoracic Society. Lung function testing: selection of reference values and interpretative strategies. Am Rev Respir Dis. 1991 Nov;144(5):1202-18.

43. Agahi A. Patterns of Lung Function in Health and Disease. Hobart: Honours Thesis, University of Tasmania; 2007.

44. Caballero B. The global epidemic of obesity: an overview. Epidemiologic Reviews. 2007;29(1):1.

45. Jones RL, Nzekwu MMU. The Effects of Body Mass Index on Lung Volumes. Chest. 2006;130(3):827-33.

46. Stritt M, Garland J. Effects of Obesity on Lung Volumes and Spirometry. Am J Respir Crit Care Med. 2009;179(1 MeetingAbstracts):A5526.

47. Piatetsky-Shapiro G, Matheus CJ, editors. The Interestingness of Deviations. AAAI-94 Workshop on Knowledge Discovery in Databases; 1994; Seattle, Washington. 25-36.

48. Silberschatz A, Tuzhilin A. What makes patterns interesting in knowledge discovery systems. IEEE Transactions on Knowledge and Data Engineering. 1996;8(6):970-4.

PEER REVIEW

Not commissioned. Externally peer reviewed

CONFLICTS OF INTEREST

The authors declare that they have no competing interests.

ETHICS COMMITTEE APPROVAL

Ethics approval obtained from the Tasmania Social Sciences Human Research Ethics Committee, approval number H0010834.

Table 7: Mean values for volume and spirometry, expressed as percentages of the predicted value (no ERV predicted data was available, so direct measure is shown)

	TLC	FRC	ERV	RV	IC	FEV_1	FVC
Underweight	103.25%	106.77%	1.86	93.36%	108.4%	95.9%	98.3%
Normal Weight	104.33%	98.95%	1.64	93.05%	113.2%	100.4%	101.3%
Overweight	103.32%	91.62%	1.22	90.22%	118.4%	99.5%	99.7%
Obese I	100.85%	82.18%	0.82	90.74%	123.7%	96.1%	94.3%
Obese II	100.35%	80.24%	0.84	90.39%	128.1%	93.8%	92.6%
Obese III	101.25%	76.64%	0.64	92.9%	135.7%	94.5%	91.9%
Obese IV	97.26%	87.46%	1.04	92.31%	110.2%	94.1%	91.8%
BMI correlation	-0.146	-0.38	-0.5	-0.04	0.19	-0.16	-0.252
Significance	<.0001	<.0001	<.0001	0.27	<.0001	<.0001	<.0001