

Determining pattern element contribution in medical datasets

Anna Shillabeer and Darius Pfitzner

School of Informatics and Engineering
Flinders University of South Australia
PO Box 2100, Adelaide 5001, South Australia

[shil0014, pfit0022]@infoeng.flinders.edu.au

Abstract

Presented are two novel solutions addressing issues in the application of automated data analysis techniques in the medical domain. The primary aim of our work is to provide medical practitioners with patterns which can inform, and facilitate the development of subjective judgements regarding the content of those patterns. This is achieved by changing the focus of information evaluation and presentation from the broad pattern level to the finer pattern element level. We believe that our solutions provide a more informative pattern description than that of current data mining applications and also provide an opportunity for medical practitioners to increase the quality of care provided and create savings both in human and financial terms.

Keywords: medical data mining, automated data analysis, pattern evaluation, pattern element, information theoretic, TFIDF, element weighting.

1 Introduction

The acceptance of automated data analysis technology in medicine is relatively low compared to other domains (Imberman, Domanski et al. 2002; Hagland 2004). Many reasons have been documented for this and they can be generalised as follows:

1. The perceived low level of flexibility in data mining systems and the need for medical analytical processes to adapt to data mining methodologies rather than data mining adapting to the needs of medicine.
2. The lack of opportunity for incorporating subjectivity when mining medical data.
3. The lack of a mining protocol to guide the application of data mining technologies to produce trustworthy outcomes and eliminate the potential for data dredging.
4. The inability of data mining systems to reflect the decision making and diagnostic processes applied in medicine.
5. The lack of available data.

6. The production of too many irrelevant results, requiring a high level of user interpretation to discriminate those that are truly useful.

Points 1, 5 and 6 have been addressed in previously published work by the first author (Shillabeer and Roddick 2006; Shillabeer, Roddick et al. 2006) and are the subject of ongoing investigation. This paper presents work in progress and will address the fourth point listed above and highlights its relationship to the need expressed in the second point. This will be done by providing a discussion of the process of diagnostic decision making in the medical domain and how novel data pattern evaluation processes may augment the process. A discussion of a collaborative solution developed through a merging of theories from data mining, medicine and information retrieval theory will also be presented, together with experimental results to demonstrate the application of these solutions to the issues in focus.

2 Medical diagnostic decision making

It is well understood that the decision making processes of medical practitioners usually, if not always, occurs at a subjective level following an objective fact gathering process. There are essentially two methods of diagnosing a patient's condition during a consultation or series of consultations; by exclusion or by pattern categorisation (Frenster 1989, Merck 2003, Elstein & Schwartz 2002):

2.1 By Exclusion

This method is the safer but more expensive of the two options. Here a doctor will reach a decision on the most likely diagnosis and treatment pattern for a set of stated symptoms exclusively through a series of objective and often invasive tests. These tests will allow the doctor to reach a decision based upon an ability to exclude potential diagnoses based upon the results of the tests. A final diagnosis is usually not made until the results of all or any highly conclusive tests are known. The potential disadvantages of this option are that it results in a lengthy and costly process which can yield an unacceptable level of false negatives or positives and cause psychological stress for the patient during the wait; especially if the tests are for a condition for which the outcome is potentially life threatening, for example cancer, meningococcal infection or HIV.

2.2 Pattern Categorisation

This is the preferred option in the medical community as it is more cost and time effective than diagnosis by exclusion, however it is generally only applied by more

experienced doctors and specialists. In this scenario a doctor will reach a diagnosis by comparing the presenting patient's disease pattern and test results to known patterns of disease and treatment. It is often the case that rather than making a single diagnosis the practitioner will develop a cluster of potential diagnoses based upon the similarity of the potential diagnoses to the pattern of the patient's condition that initiated the medical consultation as shown in Figure 1. Probabilistic pattern completion is often required and the chosen diagnosis and treatment patterns are derived from a comparison to all others in the practitioner's body of knowledge and those nearest to the patients' pattern are selected. The diagnosis is therefore more holistically targeted to the patient and case rather than the cause or effect exclusively. A drawback of this is that a rare case or combination may mean the most suitable pattern is missed or incorrectly identified and errors can be made. With experience, doctors become efficient at recognising what's expected in a pattern and which patient or condition characteristics are most influential or critical when forming a diagnosis. The speed of pattern recognition may be increased by fuzzy matching which takes account only matches between critical attributes or attribute values which fall within a range. As attributes or values which occur frequently cannot be applied to discriminate between patterns, knowing which attributes to focus on and which to eliminate from the comparison is a skill developed through years of experience.

If we therefore accept that the ability to recognise which pattern elements or attributes are important in the clinical diagnostic process it is a logical step to move towards an automated process of identifying the important elements of patterns derived from the application of data mining technologies. It is important to note here that what is considered important may be a highly subjective decision in itself and hence the need for flexibility is required by both the system and the user. Whilst there are systems which aim to teach diagnostic methodology to medical trainees, there is little evidence to suggest that automated processes are able to accurately reflect the process and this may be one of the more important reasons for the slow uptake of medical data mining systems.

3 Terminology

This paper introduces new terminology and incorporates technical terminology which is defined for clarification here. This paper relies upon an understanding of the data mining process and traditional support metric as applied in data mining systems.

Data mining is an automated data analysis process that aims to reveal previously unknown patterns in large and complex collections of data. There are almost as many definitions of data mining as there are data miners. One of the more commonly quoted definitions is that it is a "non-trivial process of identifying valid, novel, potentially useful and ultimately understandable patterns in data" (Fayyad 1996). Data mining is essentially part of a larger knowledge discovery (KD) process (see Table 1.) whereby data is transformed into information and knowledge is then extracted from that information with

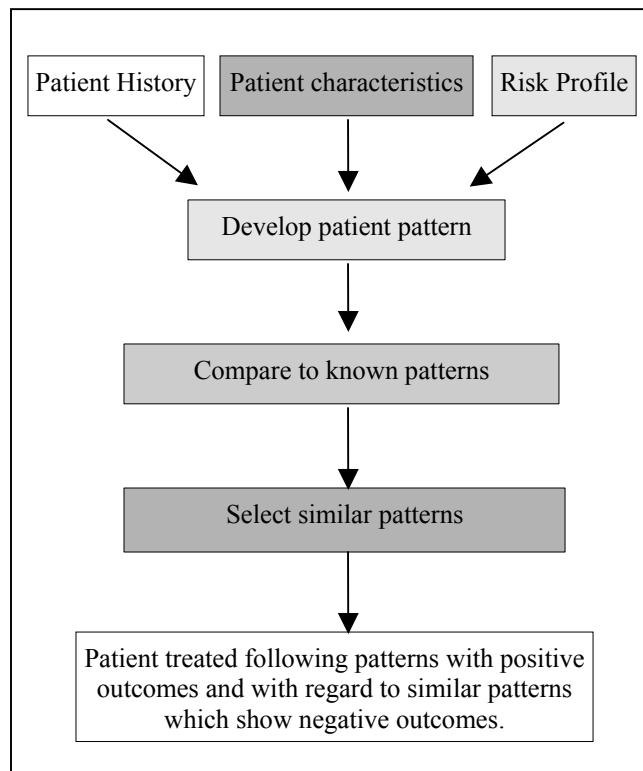


Figure1. Model of an effective diagnostic method.
 Note: Darker shading denotes higher level of subjective input.

the focus being on the identification of useful patterns in data. These patterns are often described through the standard heuristics of support and confidence.

One concept presented in this paper is an adaptation of the support heuristic. Support essentially quantifies the frequency with which a pattern occurs within the original

KD tasks	Description
Data cleaning	Management of missing irrelevant or inconsistent data irrelevant or inconsistent data
Data integration	Merging multiple data sources.
Data selection	Data relevant to the analysis being done is selected from the source.
Data transformation	Data is transformed into forms appropriate for the mining algorithm utilised.
Pattern extraction / data mining	An iterative process for revealing patterns in data.
Pattern evaluation	Determining which patterns are 'interesting' or useful – provision of information
Knowledge presentation	Transformation of the information into a suitable format – provision of knowledge.

Table 1. The data mining/knowledge discovery process.

data source, expressed as a percentage. The minimum support threshold is generally set by the user at runtime.

A pattern is defined as a set of attribute values which occur equal to, or greater than, the minimum support required by the user. Patterns take the general format as shown in the following example

Condition 'A' + Treatment 'B' lead to an outcome of 'recovery' with support of 20%.

Expanded this translates to a person having condition 'A' recovering after receiving treatment 'B' in 20% of all records held in the data source.

Further to the concept of a pattern is the finer notion of a pattern element. An element is a part of a pattern. In the example above, condition 'A', treatment 'B' and outcome 'recovery' are the three elements of the pattern. Treatment 'B' leads to an outcome of 'recovery' is also an element, as it represents only a part of the pattern.

4 The Argument for Elemental Support

This paper arose out of an understanding of the need to determine which pattern elements are important in medical decision making as raised in the previous section, and the realisation that not all elements of a pattern necessarily carry the same importance. Previous work has identified the need to calculate the importance or strength of each pattern derived from data mining activities in order to reduce the number of patterns presented and to increase the value of the set of patterns presented based upon the individual needs of each user (Shillabeer and Roddick 2006). However an issue was realised when attempting to describe the information held in the patterns to aid the subjective decision making of the medical practitioner. It was realised that it is not currently possible to identify which pattern elements directly affect the medical outcome or which have no effect and are therefore little more than confounders, although this is necessary if patterns derived from medical data are to assist in processes such as diagnosis as explained earlier. Two examples are provided to provide clarity.

1. Taking dymadon and lemsip together are unnecessary to control the symptoms of a cold, as they both contain the same active ingredient, but some patients may choose to do so from habit or ignorance. In a mining sense the pattern of dymadon + lemsip = reduction in symptoms, would be viable however we can see that in reality the removal of any one of the treatments would probably not affect the outcome but this information would not be available in traditional reporting of the presented pattern.
2. In the treatment of AIDS many medications were prescribed singularly and in combination before the specific combinations were found that made a genuine contribution to the well being and longevity of the patient. The individual medications may have demonstrated sufficient mining support through an analysis of their ability to associate with a positive outcome which would

have therefore been suggestive of an ability to facilitate a positive outcome alone, as demonstrated in figure 2. However, if in fact this were true only when combined with specific other medications this should be evident through the information or heuristics provided to describe the pattern. It should in fact be that the single medication does not reach sufficient support and is therefore not reported until combined with the other medications which result in a pattern that can be substantiated by medical fact.

Through considering these issues it was identified that data mining may not be able to ensure completeness, soundness and medical accuracy of results in the medical domain and a method for determining the importance of pattern elements rather than whole patterns is deemed necessary.

As shown in figure 2, traditional support does not provide sufficient information and may in fact be misleading. In this example, treatment A and procedure B are both quantified as strong and this could suggest that their individual implementation would lead to a positive outcome for the patient. However this may only be clinically true when in the presence of other treatments or interventions as the value for the pattern element includes instances of that element in all other patterns also and does not isolate to the incidence of only that specific pattern and no other. It is necessary to be able to discriminate on the most suitable pattern to apply by knowing the support for treatment A or procedure B in isolation rather than only as an element in a longer pattern as shown in figure 3. This more clearly denotes that applying either element alone will not necessarily lead to a positive outcome but when applied together there is a far greater chance of a positive outcome for the patient.

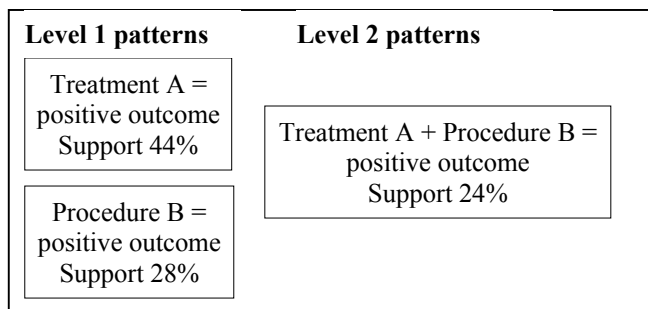


Figure 2. Traditional support values for a sample data pattern.

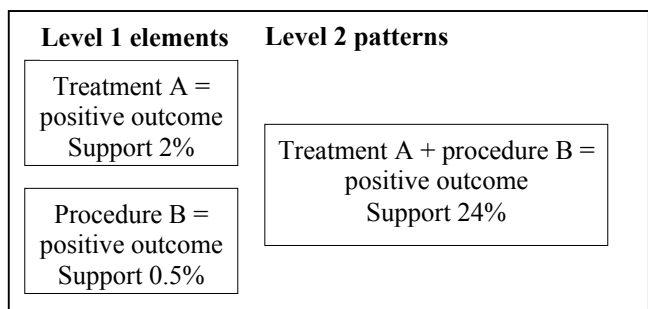


Figure 3. Elemental support values for the sample data pattern.

This understanding resulted in the following requirements to be addressed:

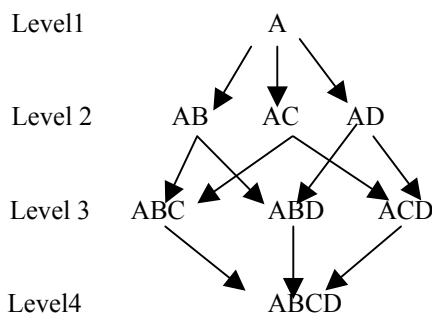
1. The need for data mining outcomes to be reported through a non-traditional application of the support measure.
2. The need to discriminate between the overall incidence of an element in any pattern and the incidence of that element in isolation.
3. The need to provide a clear description of the information held in a pattern to aid subjective judgement.

We propose two methods to address these needs. These methods have been termed support deconstruction and element weighting.

5 Support Deconstruction

Support deconstruction is a novel method which aims to determine the value of each element to a pattern. If 3 element patterns have the same support as 2 element patterns then we can logically deduce that the third element does not add any extra value, as the likelihood of achieving the same outcome is equal. In our earlier example of cold medications, the addition of lemsip to the pattern 'dymadon = relief of symptoms' would probably not increase the support for a relief in symptoms.

To determine the effect of each element we need to calculate the support for the pattern before and after adding each extra element to the pattern. In effect we need to take the incidence of BC from BCD to determine the effect of D. In a pattern ABCD = a positive outcome, each element appears as often as each other; therefore by applying traditional support metrics $suppA = suppB = suppC = suppD$ and each element could be considered equally strong. To determine the effect of element 'A' on pattern 'ABCD', its participation in all of the following relationships must be considered:



From being a single element (level 1) to being an element of the 4 element pattern (level 4), element 'A' participates in 7 other elements or relationships and in this example it would lend part of its traditional support to each of these. Support sharing on any level can be denoted as the sum of supports for all elements on lower levels occurring alone. Whilst this is logical there is further complexity, as traditionally the support for AB has included the instances of ABC, ABD and ACD and so on, and it is not clear how many times AB occurs in seclusion. To determine this there is a need to reverse engineer or deconstruct the support as demonstrated in Table 2.

To reveal the deconstructed elemental support for each element the participation of elements at all levels must be calculated. It is important to consider every element and pattern to get an accurate figure for each element in isolation not simply in those patterns with a direct linear relationship. This can be denoted as follows:

$Esupport(\text{focus element}) = Tsupport(\text{focus element}) - Esupport$ for each lower level element in which the focus element occurs. (E denotes elemental and T traditional).

Elements	A	AB	ABC	ABCD	TOTAL
Traditional support %	16	10	5	4	35
Elemental support %	6	5	1	4	16

Table2. Comparison of traditional and elemental support.

In the simplified example in table 2, we can see that the effect of adding element C to element AB is negative in that it has a lower elemental support than AB alone. In contrast adding element D to element ABC increases the elemental support of ABC, and hence denotes an increase of a positive outcome overall using our example. The obvious other comparisons needed here are the elemental supports for ABD, ACD, AC and AD to determine whether C actually adds value in the presence of D or if it is redundant and D is really the differentiator. This method has the potential to show how much each element affects the outcome described in the pattern. In a real world example this could show the most or least effective combinations of medications which would allow for more accurate targeting of medications and potentially a reduction in the number of medications taken.

Through developing a greater knowledge regarding the importance of an element to an overall pattern we can identify three types of element; positive, negative and inert.

- Positive elements are those which increase the elemental support of an element it joins with in a manner which would influence the subjective interest in the resultant pattern.
- Negative elements are those which decrease the elemental support of an element it joins with and again would influence the subjective interest in the resultant pattern.
- Inert elements are those which do not affect the elemental support of an element it joins with. The subjective interest in this element cannot be determined. It could be that the lack of effect would in itself be valuable knowledge and would therefore affect a decision based upon it.

Elemental support can report on how important each element is to each pattern it participates in. A remaining issue is how to determine those pattern elements which have been recorded frequently enough to be eliminated due to the probability that they would already be known.

Essentially the more relationships an element is involved in the more likely it is to be uninteresting. By evaluating the overall frequency of the pattern we can determine which are more unique or important and this would also assist in reducing the overall numbers of patterns by facilitating the removal of patterns or pattern elements which contain knowledge encountered frequently. This issue can be addressed by a determination of element weighting.

6 Element Weighting

There is a frequent criticism of automated data analysis in medicine that too many rules are produced and they often represent knowledge which is commonly known. For example if an element 'A' participates in only one pattern whereas 'C' is involved in many patterns it could be assumed that C is more likely to be known, as it has been recorded more frequently. Solutions have been documented which involve the development and referencing of a knowledge base to hold 'known' patterns (Kononenko 1993; Lucas 1996; Perner 1996; Moser 1999). Whilst this has been demonstrated to be a valuable tool for reducing the numbers of known patterns presented, there are potential issues with this approach:

1. The human resource time required to build and maintain a sufficiently complete knowledge base for practical use;
2. The need to apply the knowledge subjectively to prevent exclusion of unusual or marginally different patterns and inclusion of different but non-informative patterns, and;
3. The need to be sure that the knowledge base was developed from a sufficiently similar data set so that a comparison can be confidently reported.

We present a solution that begins to address these issues through a method of evaluating patterns based on pattern element weightings. The proposed solution uses element weightings to compare the importance of elements within patterns, and combined element weights to compare patterns within levels. In this way element weights allow for subjective judgements to be made from original data rather than from external sources which may have been developed using different heuristics and/or from data with different characteristics, and/or origins. Pattern element weightings also facilitate subjective pattern evaluation through an understanding that the patterns and elements are representative of the data set from which they were derived and the rate of occurrence in the patterns can be compared with the occurrence in the data source to ensure that what seems frequent or rare is actually so. As suggested above, an element which participates in many relationships is more likely to be known and cannot effectively be used as a discriminator, but elements which participate in few relationships can be considered a discriminating participant and this method will also allow for accurate quantification of this quality.

As a first step toward automating the determination of element representation, we are currently testing a technique commonly used in the field of Information

Retrieval (IR) that finds its origins in Information Theory (IT). IR is concerned with the classification of text and text based documents and offers a valuable perspective on the evaluation of worth for individual text elements of a set in the context of document classification, indexing and searching. IT is rooted in applied mathematics and involves the quantification of data to enable a high level of data as possible to be stored and/or communicated over a channel, in our case the visual channel of a data sets analysis.

Modern applications have included the evaluation of internet search results to quantify the relevance of a document to the user supplied search terms. Results containing a higher frequency of terms or a greater number of terms would be classified as more relevant. Baeza-Yates & Ribeiro-Neto [bzRn99] suggest that rather than having a binary weighting, a non-binary measure should be used to determine the degree of similarity between sets of terms in a document or query. This would be applied to rank results even if they only partially matched the requirements. Fuzzy set theory was used to allow for partial membership of a set and gives a measure of closeness to ideal membership i.e. how similar each set of terms is to each other. Key analysis and development in this area was published in 1988 by Salton and Buckley [SaltBuck88] which describes the use of a formula named *tf.idf* (TFIDF) to quantify the similarity between the contents of a document and user requirements as defined through a set of query terms.

As an information-theoretic TFIDF describes the entropy of a piece of information. The concept of information entropy was introduced by Shannon [shan48] who defined entropy as a measure of the average information content associated with a random outcome. Simply put, information entropy relates to the amount of uncertainty about an event associated with a given probability distribution. TFIDF's ability to represent the amount of information conveyed by a word in a sequence or document suggests a similar approach may be useful in determining element and/or pattern relevance.

TFIDF is a weighting comprised of the two functions *Term Frequency (TF)* & *Inverse Document Frequency (IDF)*. *Term frequency* is the frequency of a specific term in a document which indicates the importance of a term

to that document. Formally described as;

$$TF = \frac{n_i}{\sum_k n_k} \text{ where } n_i \text{ is the number of occurrences of}$$

the specific term in a document and $\sum_k n_k$ the number of occurrences of all terms in the document.

Inverse document frequency indicates a terms importance with regard to the corpus. Based in part on information theory it is the logarithm of all the documents divided by the number of documents containing a specific term.

Formally it is; $IDF = \log\left(\frac{|D|}{|d_i \supset t_i|}\right)$ where $|D|$

is the total number of documents in the corpus and $| (d_i \supset t_i) |$ is the number of document within which term t_i occurs assuming $n \neq 0$.

This shows that TFIDF incorporates the word frequency in the document such that the more a word appears in a document (e.g., its TF, term frequency is high) the more it is estimated to be significant in that document while IDF measures how infrequent a word is in the collection and accordingly if a word is very frequent in the corpus, it is considered to be non-representative of this document (since it occurs in most documents; for instance, stop words). In contrast, if the word is infrequent in the text collection, it is suggested to be very relevant for the document. This characteristic results in common terms being filtered out due to lower ranking scores/weights.

We suggest this approach can be translated to the evaluation of medical patterns by simply re-stating the concepts and substituting comparable terms as follows:

- TF can be applied to determine the frequency with which a pattern element occurs within a particular pattern sub-set at the same level or cardinality for a given support e.g. all 4 element patterns with a minimum support of 40%. This is given the assumption that patterns of the same cardinality and support are contextually similar enough to be associated/chunked much like that of the sentences of a document and thus form a pseudo document.
- IDF can be applied to determine the frequency with which a pattern element occurs within the original data set.

When brought together in the TFIDF calculation the information represented by patterns of equal cardinality can be compared to the total set to derive an entropic measure at the element, pattern and sub-set levels. We suggest that this is useful for the subjective analysis of medical patterns as it allows for the recognition of the amount of intra and inter-pattern importance (information value) of individual elements.

In data-mining terms TF can be seen as comparable to the traditional support of the pattern as it could be applied to determine the frequency of the pattern within the total pattern set or data. The novelty of the solution becomes apparent when we discuss the application of IDF. In pattern analysis this could be defined as either the frequency of the pattern against all patterns in the present set or data source or the frequency of the element in all patterns in the set or data source. As we are focussing on the value of an element within a pattern, we chose to apply the second definition.

By combining both of these into a value for TFIDF we are actually calculating the frequency of element ABC within the pattern set multiplied by the log of inverse frequency of element ABC in the data source. This will show us if that element participates in many or few other patterns and whether the element frequency is representative of the data source.

If we achieve a low value this will show that the occurrence of the element in the pattern set is dissimilar to that in data source. If we have a high value then this is indicative of an inert element for example the gender element in the gender = female and condition = pregnancy pattern which would occur as frequently in each location.

Traditional metrics have not been documented in such an application, but for the reasons discussed above it is an important issue in medical pattern analysis. Our preliminary results of testing these theories are presented following.

7. Experimental results.

7.1 Methodology

We produced association rule patterns from breast cancer data provided by a Royal Australasian College of Surgeons National Breast Cancer Audit with a minimum support of 40%. This support was chosen to create a small result set for rapid processing from the initial 22,479 records. The patterns produced were developed from values for previous surgery, tumor size and type, menopausal status and four standard treatments.

The result set contained 45 patterns ranging from 1 element patterns to 4 element patterns. We then undertook a process of support deconstruction and element weighting determination on the longest and therefore most complex patterns.

The key to the patterns described in this section is:

- A No previous surgery
- B menopause status – post menopause
- C Ovarian ablation performed
- D Arom_inhib
- E Tumor type – ductal carcinoma NOS
- F Chemo therapy applied

7.2 Support deconstruction results.

The scope of testing the concept of support deconstruction was constrained to demonstrate that support for each element could be produced and applied to provide information that could assist in a subjective appraisal of elements and patterns.

The results showed that the method has promise as a tool for providing information about data patterns which could be used by a medical practitioner to make a subjective judgement about the worth of that pattern element.

All 3 four-element patterns were deconstructed and elemental support was calculated for each element in each pattern e.g. the affect of element A on all relationships, then element B on all and so on. Each element participated in 7 other elements and 12 steps between the elements as shown in section 4. Types were then assigned for each element based upon changes in elemental

	Positive	Negative	Inert
Pattern 1			
Element A	10	1	1
Element B	11	1	0
Element C	11	1	0
Element D	9	3	0
Pattern 2			
Element A	10	0	2
Element E	10	0	2
Element C	11	0	1
Element D	10	1	1
Pattern 3			
Element A	9	1	2
Element F	10	0	2
Element C	11	0	1
Element D	9	3	0

Table 3. Elemental support results.

support as described in section 4. The results of the typing are shown in Table 3.

By reading the values in figure 4 for element A in pattern 3, we can see that the step from A to AC, or the effect of adding C to A is equal and F to A is negative. (Equality in elemental supports was deemed where the change was less than 1.) This suggests that the frequency of the outcome of this pattern is increased by adding C and decreased by adding F to A.

$$A=1.96(85.95)$$

$$AF=0.71(47.59) \quad AC=1.81(67.54) \quad AD=11.97(78.56)$$

$$AFC=2.55(43.47) \quad ACD=22.26(63.18) \quad AFD=3.41(44.33)$$

$$AFCD=40.92(40.92)$$

Figure 4. Matrix for element A in pattern 1 showing elemental supports with traditional supports in brackets.

From these results it is possible to see the difference in the effect of an element when combined with various other elements. Element A has a positive effect in 29 out of 36 relationships, a negative effect in 2 relationships and no significant effect in 5 relationships. From this information a medical practitioner may choose to further investigate where element A has a negative or insignificant effect and instead apply the more positive treatment patterns. This would however require a subjective appraisal but the information provided would serve to highlight patterns or elements of interest and show the effect of each.

7.3 Element representation results.

In initial experiments we have applied a basic TFIDF formula, as defined above, while acknowledging that this may be far from optimal given the many variants of this formula such as those proposed by Salton & Buckley (SaltBuck88), Johnson-Laird, Girotto & Legrenzi (JohnsGiroLeg98) and Treharne, Pfitzner & Powers (TrehPfitPow06).

TFIDFs' information theoretic basis suggests that it is translatable to the application of determining element and/or pattern relevance by allowing elements to be ranked against their information content derivable from available information and not external sources. This being the case we applied a modified form of the TFIDF approach to the frequent item sets generated from breast cancer data as a first step to testing the usefulness of this type of information for subjective analysis in the medical domain.

In modifying the TFIDF approach to fit the evaluation of medical patterns by simply re-stating the concepts and substituting comparable terms as follows:

When calculating our IDF equivalent, each row in the original data set is treated as a separate text block/document to generate block and element counts for the original data-set. D was treated as being equal to the number of rows in the original set and ti was considered equal to the number of rows the i^{th} element occurred in. Our reasoning for treating the rows individually is that no contextual information about each row has been realised at this point so they must be assumed to have no relationship.

In calculating TF the rows of patterns were grouped together by number of elements to create a pseudo document. As they were generated using the same heuristics they represent the same contexts. In other words our TF can be applied to determine the frequency with which a pattern element occurs within a particular pattern set at the same level. This is assuming that frequent patterns the same level for a given support are contextually similar enough to be associated/chunked much like that of the sentences of a document and thus as suggested form a pseudo document.

The TFIDF weighting generated in this fashion for each element now represents the expectation that it will occur in its relative pattern and allows the analyst to make judgements on an elements importance given its relative entropic measure. We suggest that this is useful for the subjective analysis of medical patterns as it allows for the recognition of the importance (information contained within) of individual elements within their pattern and subset, the importance of a pattern within its sub-set and each sub-set against each other sub-set. One method of subjective analysis of medical patterns is by looking for the unexpected or low element values, or by ordering the patterns of a relative item-set using the totalled weights

	Pattern Elements and Associated Weights	Pattern Weights
1	A = 0.031, F = 0.0501, C = 0.066, D = 0.0216	0.1687
2	A = 0.031, B = 0.0405, C = 0.066, D = 0.0216	0.1591
3	A = 0.031, E = 0.0381, C = 0.066, D = 0.0216	0.1567

Table 4. Four element patterns used for analysis.

for each pattern. For example, in Table 4 pattern number 3 has the lowest row weighting indicating this sequence of elements is least expected from that data source and might therefore be of interest. Also, when assessing all rows the analyst might note that element *D* has a significantly lower score than the others again flagging the fact it was least expected and possibly of interest.

8 Conclusion

This paper has presented two novel applications for the determination of information value in medical data patterns. The theories are sound and early results show that they have potential to better describe the information held in medical data patterns. Their contribution to the domain is valuable not only in terms of making the results of automated data analysis more accessible to medical professionals but also in human and financial terms. Elemental support is able to describe patterns in terms of the effect of adding new elements to those patterns thus allowing a doctor to see which treatments will have the greatest chance of producing a favourable outcome for their patients. It can also be applied to highlight where treatments are not compatible or where the prescription of new medications will have little positive effect, hence saving the patient and pharmaceutical insurance companies valuable dollars.

There is more work to do on these theories, and an important next step will be to apply the theories to a wider data source and to present the patterns with their associated typing and information value weightings to medical professionals who can verify the value of the information provided.

9 Acknowledgements

The authors would like to acknowledge the medical knowledge and practical wisdom supplied by Associate Professor Robert Bryce, Clinical Director of Obstetrics and Gynaecology at the Flinders Medical Centre South Australia, in the development of the theories presented in this paper.

10 References

Baeza-Yates, R. & Ribeiro-Neto, B., (1999), "Modern Information Retrieval", Addison-Wesley, Sydney, p.27-30.

Salton, G. and Buckley, C., (1988), "Term-Weighting Approaches in Automatic Text Retrieval", Information Processing Management, Vol. 24, No. 5, pp.513-523.

Shannon, C.E., (1948), "A Mathematical Theory of Communication", Bell Systems Technical Journal, Vol.27, pp.379-423, 623-656, July, Oct.

Johnson-Laird, P. N., Girotto, V., Legrenzi, P. (1998). Mental models: a gentle guide for outsiders. Retrieved August 7, 2003 from The Interdisciplinary Committee on Organizational Studies, University of Michigan: <http://www.si.umich.edu/ICOS/gentleintro.html>.

Trehanne, K., Pfitzner, D. & Powers, D., (2006), "Towards Cognitive Optimisation of a Search Engine Interface", To be published in: Australian Language Technology Workshop, Sydney, Australia, 2006.

Fayyad, U. M. P.-S., G. Smyth, P. (1996). "From Data Mining to Knowledge Discovery in Databases." AI Magazine 17(3): 37-54.

Hagland, M. (2004). "Data Mining." Healthcare Informatics Online.

Imberman, S. P., B. Domanski, et al. (2002). "Using dependancy/Association rules to find indications for computerised tomography in a head trauma dataset." Artificial Intelligence in Medicine 26(1-2).

Kononenko, I. (1993). "Inductive and Bayesian Learning in Medical datasets." Applied Artificial Intelligence 7: 317-337.

Lucas, P. (1996). Enhancement of Learning by Declarative Expert-based Models. Intelligent data analysis in biomedicine and pharmacology, Budapest Hungary.

Moser, S. A. J., W.T. Brossette, S.E. (1999). "Application of Data Mining to Intensive Care Unit Microbiological Data." Emerging Infectious Diseases. 5(3): 454-457.

Perner, P. (1996). Mining Knowledge in X-Ray Images for Lung Cancer Diagnosis. Intelligent data analysis in biomedicine and pharmacology, Budapest Hungary.

Shillabeer, A. and J. F. Roddick (2006). "Towards Role Based Hypothesis Evaluation for Health Data Mining." Electronic Journal of Health Informatics 1(1): 1-8.

Shillabeer, A., J. F. Roddick, et al. (2006). On the Arguments Against the Application of Data Mining to Medical Data Analysis. Intelligent data Analysis in BioMedicine and Pharmacology, Verona, Italy.

Merck & Co. Medical Decision Tests. www.merck.com/mmhe/print/sec25/ch300/ch300.html Accessed 14/09/2006.

Frenster, J. H. Matrix Cognition in Medical Decision Making. Congress on Medical Informatics San Francisco 1989. Volume 7. pp 131-134.

Elstein, A. S. and Schwartz, A. Evidence Base of Clinical Diagnosis. BMJ March 2002. Volume 324, pp 729-732.