Application of Optimisation-based Data Mining Techniques to Medical Data Sets: A Comparative Analysis

Zari Dzalilov, Adil Bagirov, Musa Mammadov
The Centre for Informatics and Applied Optimisation
School of Science, Information Technology and Engineering,
University of Ballarat, Victoria, Australia
z.dzalilov@ballarat.edu.au, a.bagirov@ballarat.edu.au, m.mammadov@ballarat.edu.au

***Abstract* - Computational methods have become an important tool in the analysis of medical data sets. In this paper, we apply three optimisation-based data mining methods to the following data sets: (i) a cystic fibrosis data set and (ii) a tobacco control data set. Three algorithms used in the analysis of these data sets include: the modified linear least square fit, an optimization based heuristic algorithm for feature selection and an optimization based clustering algorithm. All these methods explore the relationship between features and classes, with the aim of determining contribution of specific features to the class outcome. However, the three algorithms are based on completely different approaches. We apply these methods to solve feature selection and classification problems. We also present comparative analysis of the algorithms using computational results. Results obtained confirm that these algorithms may be effectively applied to the analysis of other (bio)medical data sets.**

Keywords – data mining; optimisation; cystic fibrosis; tobacco control.

I. Introduction

Optimization plays a fundamental role in designing efficient data mining techniques. For example, the support vector machine algorithms are among the most efficient data classification techniques [7]. Such techniques have been applied to medical data sets over the last two decades to solve wide range of problems including feature selection, data classification and prediction problems. Despite of significant developments in this area there is still a lot of evaluation work to be performed due to the fact that medical data sets are diverse and it is difficult to formulate a unique criterion for all of them. Comparison of specific medical data sets was done in [5, 8,18,19, 23].

In this paper, we present the results of application of three optimization-based data mining algorithms to two different data sets: the CF (Cystic Fibrosis) data set [14] and the Tobacco Control data set [12,13]. These algorithms can be applied to solve three different problems of data mining: data regression, data classification and clustering. All three algorithms are based on nonlinear models, and therefore, can detect nonlinear relationships between both features and instances. Data sets used in evaluation are completely different which helps to have a clear picture about efficiency and accuracy of algorithms used in the comparison. Moreover, such data sets have not been thoroughly studied using data mining techniques.

The paper is organized as follows. Section II describes the two data sets used for the analysis, Section III briefly describes the algorithms and Section IV presents the results obtained by applying these algorithms to the data sets. Section V concludes the paper.

II. Data set description

The two data sets for analysis in this paper are the Cystic Fibrosis and Tobacco Control. These are described in more detail in this section.

A. Cystic Fibrosis data set

Cystic fibrosis is the most common fatal genetic disorder in the Caucasian population [11]. Clinical scoring systems for the assessment of Cystic fibrosis disease severity have been used for almost 50 years without being adapted to the milder phenotype of the disease in the 21st century [9,11,14, 20]. A fresh approach is needed for the development of comprehensive CF disease severity scales, which may be used as a disease predictor. The goal is to develop a scoring system to assess the longitudinal process of Cystic Fibrosis.

We propose to develop a new clinical scoring system by employing various statistical tools and optimisation methods. We previously identified an approach for developing a disease severity scale [14]. We now propose to refine this scale by using a hybrid model combining mathematical optimisation and data mining approach. We evaluate mathematical optimisation methods that can be used for the solution of feature selection problems. The advantage of these methods is that they allow one to consider datasets with an arbitrary number of classes.

The evaluation is based on the Cystic Fibrosis database from the cohort at the Royal Children's Hospital in Melbourne. The data base contains 212 subjects, with 69 features and 3 expert defined or 5 CAP defined classes. The methods applied to this data set are the *Linear Least Squares Fit (LLSF)* [21, 22] and the *Heuristic Algorithm for Feature Selection* [1,2,3]. Both of these methods explore relationships between features and classes. They allow to analyse data sets with an arbitrary number of classes. Our results show that the methods applied are helpful to determine the contribution of features to the effectiveness of disease severity classification, which is the main point for developing *a clinical scoring system.* The results obtained can be used in preparatory work for clinical trials. However, more data points are needed to finalize a clinical score, by re-running these methods in the larger data set.

ISBN: 978-1-61208-227-1

B. Tobacco Control data set

Smoking is one of the leading causes of death around the world and as such, control of tobacco use is an important global public health issue [10]. The large detrimental impact, that smoking already has on a public health has the potential to become even greater as the population worldwide ages and dementia prevalence increases. Controlling tobacco smoking and determining effective policies is difficult because of the complexity of human nature. Nevertheless, there have been numerous attempts to describe and understand the effectiveness of tobacco control policies to smokers' quitting behaviour. Linear regression and logistic regression are currently very popular statistical techniques for modelling and analysing complex data in tobacco control systems [24]. However, in tobacco markets, numerous inter-related factors interact with tobacco control policies in non-trivial fashion, such that policies and control outcomes are non-linearly related. The use of linear and logistic regression is therefore fundamentally limited due to their inability to deal with these complex relationships. Compared with traditional statistical techniques, optimization-based methods have the potential to be more effective analysis tools of complex tobacco control systems. The Tobacco Control data set was collected in Australia for studying and evaluating the psychosocial and behavioural impact of diverse tobacco control policies to smokers' behaviour. This data set was collected in the frame of ITC project [6, 15, 17] and is shown in Figure 1.

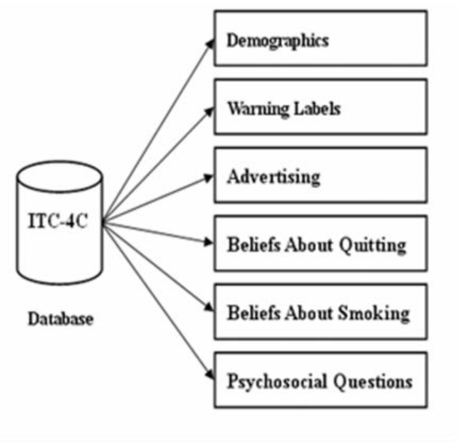

Figure 1. Description of Tobacco Control data set.

The data set was used to evaluate the optimization-based approaches in data mining [12, 13]. The aim of the exercise was to find *clusters* of smokers with similar beliefs about quitting for predicting the rate of quitting attempt. We apply the following optimization-based algorithms to the tobacco control data: the *Linear Least Squares Fit (LLSF),* the *Heuristic Algorithm for Feature Selection,* and the *Modified Global k-means Algorithm* [4]. We have obtained some promising preliminary results for covering potential solutions in tobacco control. A brief description of algorithms is given below

III. Algorithm description

The three algorithms used in this paper are now described in more detail.

A. The Linear Least Squares Fit (LLSF)

Let M be the number of all features and C be the number of classes. Data is given in the form of two matrices. Matrix $A=(a_{ij}),\ i=1,...,N,\ j=1,...,M,$ where N is the number of samples. Matrix $B=(b_{ik}),\ i=1,...,N,\ k=1,...,C,$ where vector $(b_{i1},\dots,b_{iC})$ describes class information for the row/sample i; $b_{ik}=1$ if sample i belongs to class k and $b_{ik}=0$ otherwise. Consider the matrix $X=(x_{jk}),\ j=1,...,M,\ k=1,...,C,$ that describes the relationships between features and classes. LLSF aims to find matrix X by minimizing the function $f(X)=\| AX-B\|^2$.

Take any feature p and eliminate it from the list of features. Denote

$$A(p)=(a_{ij}),\ i=1,...,\ N,\ j=1,...,\ p-1,\ p+1,\dots,M$$

and

$$X\,(p)=(x_{jk}),\ \ j=1,...,\ p-1,\ p+1,\dots,\ M,k=1,...,C.$$

Let

$$X^*(p)\ =\ \arg\min\{\ \| AX(p)-B\|^2:\ \ X(p)\in R^{M\times C}\ \}$$

Matrix $X^*(p)$ can be used to predict all samples $i=1,...,\ N$ using all features except p: $j = 1, \dots, p-1, p+1, \dots, M.$ Denote the average accuracy obtained in this way by $E(p)$. Clearly, the inequality $E(p_1) <\ E(p_2)$ for some features p_1 and p_2 means that the accuracy decreases more if we eliminate feature p_1 rather than p_2. Therefore, we can say that feature p_1 is more important than p_2; we write in this case $p_1 \square p_2$, arranging all features in a way that: $E(j_1) \leq\ E(j_2) \leq ..... \leq\ E(j_M)$ we obtain the order of features by their importance in ascending order $j_1 \succ\ j_2\ \succ .... \succ\ j_M$.

Previously, we applied LLSF to the data set on Cystic Fibrosis [14]. Our preliminary results show that the methods applied are helpful in developing a clinical scoring system on Cystic Fibrosis.

B. Heuristic Algorithm for Feature Selection

This algorithm for the solution of the feature selection problem is based on techniques of convex programming [2] and allows one to consider data sets with an arbitrary number of classes. We consider feature selection in the context of the classification problem. The algorithm calculates a subset of

Copyright (c) IARIA, 2012.   ISBN: 978-1-61208-227-1

most informative features and a smallest subset of features. The first subset provides the best description of a dataset whereas the second one provides the description which is very close to the best one. A subset of informative features is defined by using certain thresholds. The values of these thresholds depend on the objective of the task.

The purpose of the feature selection procedure is to find the smallest set of informative features possible for the object under consideration, which describes this object from a certain point of view. The following issues are very important for understanding the problem:

- It is convenient to consider (and define) informative features in the framework of classification. In other words it is possible to understand whether a certain feature is informative for a given example if we compare this example with another one from a different class.
- Our goal is to find a sufficiently small set of informative features and to remove as many superfluous features as possible. Note that this problem can have many different solutions.
- It follows from the above that the set of informative features, which describe a given object, is a categorical attribute of this object. This is also a fuzzy attribute in a certain informal sense. It leads to the following heuristic conclusion: it is useless to apply very accurate methods in order to find this set. However, if we use heuristic (not necessarily very accurate) methods we need to have experimental confirmation of the results obtained.

This algorithm proceeds as follows. First, we find centres of each class and remove a feature which gives a smallest distance among all features. Using any classification method we compute classification accuracy with the rest of features. Then, we update the centres with one removed feature and remove again the closest feature and apply again the classification method. If the classification accuracy is reduced significantly, we stop and accept the rest of the features as the most informative. Otherwise, the algorithm continues. The "significant" reduction in accuracy is defined by the user. In our calculations this reduction is 1%.

C. Modified Global k-means Algorithm

Cluster analysis, also known as unsupervised data classification, is an important subject in data mining. Its aim is to partition a collection of patterns into clusters of similar data points. In cluster analysis we assume that we have been given a finite set of points A in the n-dimensional space R^n, that is $A = \{a^1,...,a^m\}$, where $a^i \in R^n, i = 1,...,m$. There can be different types of clustering. In this paper, we consider the hard unconstrained partition clustering problem, that is the distribution of the points of the set s A into a given number k of disjoint subsets $A^1,...,A^k$ with respect to predefined criteria such that:

1) $A^j \neq \emptyset, j = 1,...k;$

2) $A^j \cap A^l = \emptyset, j, l = 1,...k, j \neq l;$

3) $A = \bigcup_{j=1}^{k} A^j;$

4) no constraints are imposed on clusters $A^j, j = 1,...k.$

The sets $A^j, j = 1,...k;$ are called clusters. We assume that each cluster A^j can be identified by its centre (or centroid) $x^j \in R^n, j = 1,...k.$ Then the clustering problem can be reduced to the following optimisation problem (see [1, 2]):

find a collection $\overline{x} = (\overline{x}^1,...,\overline{x}^p)$ of n-dimensional vectors, which is a solution to the following problem:

$$\min f(x^1,...,x^p) = \sum_{i=1}^{m} \min_{j=1,...,p} \left\| x^j - a^i \right\|^2 \quad \text{subject to}$$

$$x^j \in R^n, j = 1,..., p. \qquad (1)$$

Here, ||.|| is an Euclidean norm. The problem (1) is also known as minimum sum of squares-clustering problem [16]. It is assumed that a given set of points contains p clusters and the solutions $\overline{x}^1,...,\overline{x}^p$ to the problem (1) are the centres of these clusters.

IV. Results and discussion

A. Cystic Fibrosis Data Set

Statistical approaches were tested on the Cystic Fibrosis database from the cohort at the Royal Children's Hospital in Melbourne. After data preparation which included expert-opinion of an individual's clinical severity on a 3 point-scale (mild, moderate and severe disease), two multivariate techniques were used to establish a method that would have a better success in feature selection and model derivation. The methods were *Canonical Analysis of Principal Coordinates (CAP'*) and *Linear Discriminant Analysis*. A 3-step procedure was performed which included selection of features, extracting 5 severity classes from the 3 original classes as defined by medical experts and establishment of calibration datasets.

Two different methods, based on optimisation techniques have also been used for the solution of the feature selection problems. These methods are the *Linear Least Squares Fit (LLSF)* and the *Heuristic Algorithm for Feature Selection*. Since the data was already broken up into a number of classes the *Modified Global k-means Algorithm* was not used on this data set. We apply all methods to the data set containing 212 subjects, with 69 features and 3 expert defined classes or 5 CAP defined classes. The results are shown in Table 1. The methods are enumerated as follows: *LLSF*-1, *Feature Selection*-2, *Statistical*-3.

All three methods (1, 2, 3) indicate that the following features are the most significant:

TABLE I. CYSTIC FIBROSIS SIGNIFICANT FEATURES

Method	**Significant Features**
1, 2, 3	19,21,29,30, 31, 35, 47
1, 2	18,22,23, 26, 32,40, 41,48, 49, 50, 51, 55, 57

Information from Table 1 can be used in clinical practice. The highest preference should be given to the features in the first line, since they have been confirmed by all three methods. The features can be identified in the data set through the codes shown in Table 2.

TABLE II. CYSTIC FIBROSIS FEATURE CODES

Feature	**Code**	**Feature**	**Code**
19	NORESP	26	NTSURG4
21	NODAYS	32	CULTandBMI
29	FEV1P	40	ANTIBO
30	FVCP	41	ANTIBO
31	FEFP	48	THERAP2
35	BMIPCT	49	THERAP3
47	HERAP1	50	THERAP4
18	NOVIS	51	THERAP5
22	HTCOUR	55	NSUP1
23	HTDAYS	57	LTOXYYNU

Table 2 provides all the significant features that have been identified. The meaning of the features can be found from the data base of the Royal Children Hospital and is not included here.

Our preliminary results show that the optimization methods applied are helpful in developing a clinical scoring system. However, more data points are needed to finalize a clinical score, by re-running methods in the larger dataset.

B. *Tobacco Control Dataset*

As a preliminary work, we applied the *Linear Least Squares Fit*, the *Heuristic Algorithm for Feature Selection* and the *Modified Global k-means* algorithms to the four data sets containing:

- Data set 1 – 1458 subjects, with 71 features
- Data set 2 – 1477 subjects, with 69 features
- Data set 3 – 1260 subjects, with 60 features
- Data set 4 – 1350 subjects, with 60 features

1) *Linear Least Squares Fit*

Results obtained by LLSF algorithm for Data Set 1 are illustrated in Figure 2. This figure shows dependence of the classification accuracy on the number of features. One can see that this algorithm does not allow one to find the subset of most informative features.

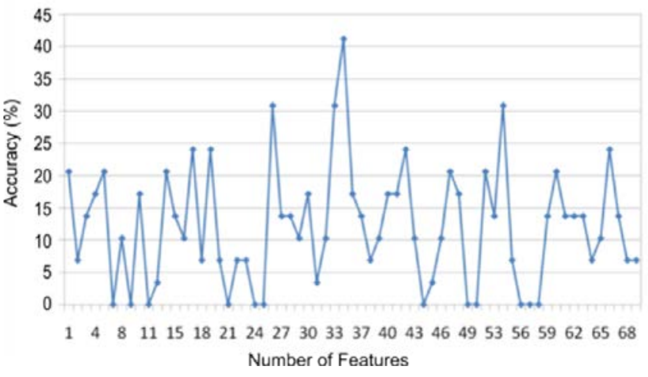

Figure 2. Results for Data Set 1 using LLSF algorithm.

2) *Heuristic Algorithm for Feature Selection*

Illustrations of numerical results of applications of the *Heuristic Algorithm for Feature Selection* to the Data Sets 1, 2 and 3 are shown on Figures 3-5 below. These figures show dependence of the classification accuracy on the number of features.

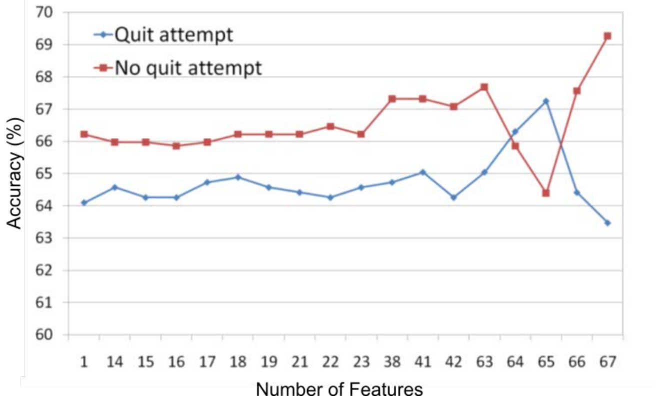

Figure 3. Results for data set 1 using Heuristic Feature Selection algorithm.

Figure 3 shows features obtained from the first survey. The red line shows the smokers with no intention to quit, with features 67 and 65 having maximal and minimal accuracies respectively. The blue line shows features associated with smokers with the intention to quit. This time the behaviour is reversed with the minimal accuracy for feature 67, and the maximal accuracy for feature 65. In general, the two feature sets exhibit complementary behaviour.

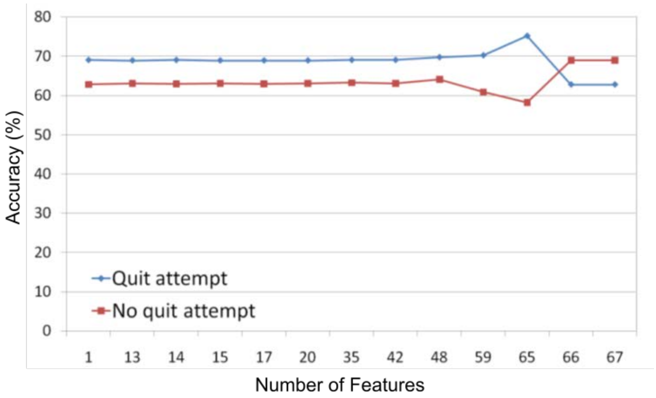

Figure 4. Results for data set 2 using Heuristic Feature Selection algorithm.

Figure 4 shows features obtained on the data from the second survey. The same maximal and minimal features are obtained for both datasets as in the previous case, however, in contrast to the data of the previous survey, the observed fluctuations in feature accuracy is much less pronounced.

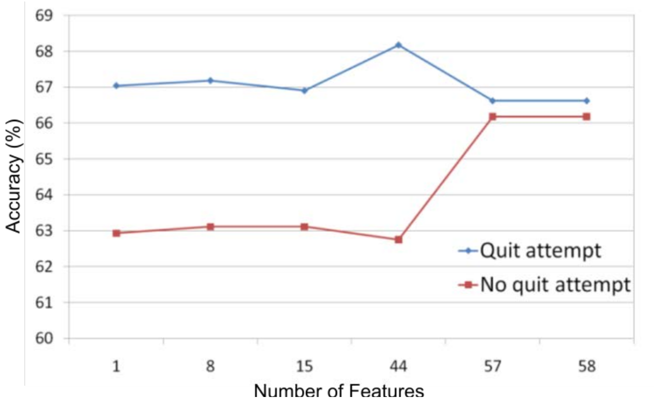

Figure 5. Results for data set 3 using Heuristic Feature Selection algorithm.

Figure 5 shows features obtained on the data from the third survey. This time the maximal feature for smokers with no intention to quit (red line) is 57 and the minimal feature is 44. For smokers with the intention to quit (blue line) the maximal feature is 44 and the minimal feature is 57. The figure once again shows complementary behaviour of these feature sets. At the same time however, the increased difference between the two feature sets, compared to previous sets, is apparent.

Some examples of the meanings of the significant features for quit attempt:

- 5 - Thoughts about danger of smoking
- 7 - Thoughts about harm to self
- 35 - Confidence to quit smoking
- 37 - Perception of quitting difficulty
- 66 - Disapproval of smoking in society
- 36 – How easy or hard to stop smoking permanently

The results obtained allow us to draw the following conclusions:

- *Heuristic Algorithm for Feature Selection* revealed complementary pattern of features for quitters and non-quitters
- Overlapping features are likely to be important in tobacco control programs

As a result of the feature selection algorithm we found a number of significant features, associated with quit attempts. The features in the neighbourhood of the maximal point can also be considered to be associated with quit attempts, with a reasonable level of accuracy. The maximum accuracy attained is 67.24. The list of smokers' response to the most significant questions in predicting the rate of quitting attempt is shown in Table 3.

TABLE III. TOBACCO CONTROL DATA SET SIGNIFICANT FEATURES

Data Set	Significant Features
Data Set 1	5, 7, 35, 37, 66
Data Set 2	35, 37, 36
Data Set 3	35, 37, 36
Data Set 4	35, 37

These findings have demonstrated that the most significant features for pushing smokers to make a quit attempt are focused on the following aspects: knowledge about the harm of smoking, worry about health, confidence about quitting and addiction to smoking. Our findings are consistent with those that we have found before using other methods. Our methods have answered that smokers' motivations, knowledge about harm of smoking, beliefs about quitting and so on are key factors for making a quitting attempt. If we consider 64 to be sufficiently accurate, then the number of significant features increases.

3) *Modified Global k-means Algorithm*

Computational results to the Australian tobacco control data set can be summarized as follows:

- The modified global k-means algorithm allows one to find global or near global solutions to the clustering problems in the tobacco control data set.
- Results demonstrate that the modified global k-means algorithm detects correct number of clusters and the further reduction of the tolerance ε>0 do not lead to the increase of the number of clusters. This means that the modified global k-means algorithm is able to find stable cluster structure of the tobacco control data set.
- The clustering algorithm allows one to find stable clusters in the data set and these clusters do not change as the number of clusters increases. Moreover, we demonstrate that the cluster structure is not changing if one removes stable clusters one by one. This structure changes only when all stable clusters are removed from data set.

Our results show that the modified global k-means algorithm is efficient and robust for solving clustering problems in the tobacco control data sets. Future work in this area includes classification of the set of clusters associated with each data set. Our methods aim to answer a key question: “How can we predict the response of smokers within the clusters to tobacco control policies?” Compared with the traditional statistical techniques, the new methods have potential to become a good theoretical and methodological framework for modelling and analysing complex tobacco control systems. The results of analysis of the given data set are most likely to develop new models for a new survey, more accurate than the previous one.

V. CONCLUSION

We evaluated three optimization-based data mining methods on two distinct medical data sets. The Cystic Fibrosis is a medical data set built around measurements of disease severity. The results show that all three methods worked equally well and may consequently be used for analysis of similar medical data sets.

The Tobacco Control data set is a massive survey for studying and evaluating the psychosocial and behavioural impact of diverse tobacco control policies to smokers from many countries. This kind of data sets tend to be noisy and the design of the survey may not be optimally suited to evaluate the accuracy of the outcome. The results demonstrate that the LLSF is very sensitive to the noise whereas other two optimization-based methods (both clustering and classification) perform well in the analysis of these types of data sets. More informative data sets enriched by health parameters will help to find the links from smoking to the risk of diseases such as dementia, stroke, lung cancer, vascular dementia, oxidative stress and inflammation. The reference to the research outcome could greatly impact the health choices of smokers.

We conclude by noting the usefulness of optimisation-based methods (both clustering and classification) in the analysis of distinct types of medical data sets.

REFERENCES

[1] A. M. Bagirov, A. M. Rubinov, and J. Yearwood, “A global optimisation approach to classification,” Optimization and Engineering Journal, vol. 3, no. 2, 2002, pp. 129-155.

[2] A. M. Bagirov, A. M. Rubinov, and J. Yearwood, “A heuristic algorithm for feature selection based on optimisation techniques,” in Heuristic and Optimization for Knowledge Discovery, C. Newton, H. Abbas and R. Sarker, Eds. Idea Group Publishing, 2002, pp. 13-26.

[3] A. M. Bagirov, A. M. Rubinov, N. V. Soukhoroukova, and J. Yearwood, “Supervised and unsupervised data classification via nonsmooth and global optimisation,” TOP: Spanish Operations Research Journal, vol. 11, no. 1, 2003, pp. 1-93.

[4] A. M. Bagirov, “Modified global k-means algorithm for sum-of-squares clustering problems,” Pattern Recognition, vol. 41, no. 10, 2008, pp. 3192-3199.

[5] A. M. Bagirov, A. M. Rubinov and J. Yearwood, “Using global optimisation to improve classification for medical diagnosis and prognosis,” Topics in Health Information Management, vol. 22, no. 1, 2001, pp. 65-74.

[6] R. Borland, H. H. Yong, N. W. Geoffrey, G. T. Fong, D. Hammond, K. M. Cummings, W. Hosking, and A. McNeill, “How reaction to cigarette packet health warnings influence quitting: findings from the ITC four country survey,” 2009, preprint.

[7] C. Burges, “A tutorial on support vector machines for pattern recognition,” Data Mining and Knowledge Discovery, vol. 2, 1998, pp. 121-167.

[8] W. P. Chang and D. M. Liou, “Comparison of three data mining techniques with genetic algorithm in the analysis of breast cancer data,” Journal of Telemedicine and Telecare, 2008.

[9] S. P. Conway and J. M. Littlewood, “Cystic fibrosis clinical scoring systems,” in Cystic fibrosis - Current Topics, vol. 3, J. H. Widdicombe Ed. New York: John Wiley & Sons Ltd, 1996, pp. 339-358.

[10] R. M. David, “50 years of reporting on tobacco and health,” British Medical Journal, 2000, vol. 320, no. 74.

[11] P. B. Davis, “Cystic fibrosis since 1938,” American Journal of Respiratory and Critical Care Medicine, vol. 173, no. 5, 2006, pp. 475-482.

[12] Z. Dzalilov and A. M. Bagirov, “Cluster analysis of a tobacco control data set,” International Journal of Lean Thinking, vol. 1, no. 2, pp. 40-45.

[13] Z. Dzalilov, J. Zhang, A. M. Bagirov, and M. A. Mammadov, “Application of optimisation–based data mining technique to tobacco control dataset,” International Journal of Lean Thinking, vol. 1, no. 1, pp. 27-41.

[14] G. Hafen, C. Hurst, J. Yearwood, M. A. Mammadov, J. Smith, Z. Dzalilov, and P. Robinson, “A new clinical scoring system in cystic fibrosis: statistical tools for database analysis – a preliminary report,” BMC Medical Informatics and Decision Making, 2008, 8: 44.

[15] D. Hammond, G. T. Fong, R. Borland, K. M. Cummings, A. McNeill, and P. Driezen, “Text and graphic warnings on cigarette packages: findings from the international tobacco control four country study,” American Journal of Preventive Medicine, vol. 32, 2007, pp. 202-209.

[16] P. Hansen, E. Ngai, B. K. Cheung, and N. Mladenovic, “Analysis of global k-means, an incremental heuristic for minimum sum-of-squares clustering,” Journal of Classification, vol. 22, no. 2, 2005, pp. 287-310.

[17] ITC International tobacco control policy evaluation survey. 2002-2010; http://www.itcproject.org, [retrieved: August, 2011]

[18] M. A. Mammadov, A. M. Rubinov, and J. Yearwood, “The study of drug-reaction relationships using global optimization techniques,” Optimization Methods and Software, vol. 22, no. 1, 2007, pp. 99-126.

[19] M. A. Mammadov, A. M. Rubinov and J. Yearwood, “An optimization approach to identify the relationship between features and output of a multi-label classifier,” Data Mining in Biomedicine, vol. 7, P. Pardalos, V. Boginski and A. Vazacopoulos, Eds. Series: Springer Optimization and its Applications, 2007, pp. 141-168.

[20] M. H. Schoeni, “Presentation and critical comparison of clinical scoring systems in patients with cystic fibrosis,” Klin Paediatr, vol. 205, 1993, pp. 3-8.

[21] Y. Yang, “An evaluation of statistical approaches to text categorization,” Information Retrieval, vol. 1, 1999, pp. 69-90.

[22] Y. Yang and X. Liu, “A re-examination of text categorization methods,” Proceedings of the 22nd annual international ACM SIGIR conference on research and development in information retrieval, Berkeley, California, 1999, pp. 42-49.

[23] D. D. Walker and G. Kadam, “Predicting breast cancer survivability: a comparison of three data mining methods,” Artificial Intelligence in Medicine, vol. 34, no. 2, 2005, pp. 113-127.

[24] J. Y. Zhang, D. Young, K. Coghill., S. Petrovic-Lazarevic, R. Borland, C. H. Yeh, and S. Bedingfield, “A new theoretical framework for modelling and analysing complex tobacco control systems,” Proceedings of the Ninth Global Business and Technology Association's Annual International Conference, 2007, pp. 811-817.