Available online at www.sciencedirect.com
ScienceDirect
European Journal of Operational Research 177 (2007) 515–524
EUROPEAN JOURNAL OF OPERATIONAL RESEARCH
www.elsevier.com/locate/ejor

Decision Support

A classification model based on goal programming with non-standard preference functions with application to the prediction of cinema-going behaviour

D.F. Jones [a,*], A. Collins [b], C. Hand [c]

[a] Department of Mathematics, University of Portsmouth, Buckingham Building, Lion Terrace, Portsmouth, Hants PO1 3HE, UK
[b] Department of Economics, University of Portsmouth, Richmond Building, Portland Street, Portsmouth, Hants PO1 3DE, UK
[c] School of Marketing, Faculty of Business, Kingston University, Kingston Hill, Kingston upon Thames, Surrey KT2 7LB, UK

Received 5 May 2005; accepted 21 November 2005
Available online 3 March 2006

Abstract

This paper presents a goal programming model that allows for the flexible handling of the two group classification problem. The goal programming model is based around the concepts of non-standard preference functions and penalty function modelling. An extension to a generalised distance metric case is given. The inclusion of multiple levels of classification based upon different levels of certainty is incorporated into the model. The model is tested on a real-life data set pertaining to cinema-going attendance and conclusions are drawn both in the context of the methodology and of the application.

Keywords: Data mining; Pattern classification; Goal programming; Non-standard preference functions; Cinema going

1. Introduction

The need to classify objects into one of a number of distinct classes occurs in many situations in life. The most common situation and easily analyzable is the two-group classification problem where items have to be classified into one of two distinct groups by consideration of a number of attributes of the item. Two group classification models arise in fields such as finance (creditworthy or non-creditworthy), medicine (benign or malignant), and defence (friendly or hostile). Normally, a training set of observations whose class and attributes are known is available in order to train or form the

* Corresponding author.
E-mail addresses: dylan.jones@port.ac.uk (D.F. Jones), alan.collins@port.ac.uk (A. Collins), c.hand@kingston.ac.uk (C. Hand).

0377-2217/$ - see front matter
doi:10.1016/j.ejor.2005.11.014

model in its classification. Typical mathematical and computational techniques that have been used to form pattern classification models include genetic algorithms [3], simulated annealing [4], and neural networks [8,10]. The use of mathematical programming techniques for pattern classification is described by Baek and Ignizio [2]. Freed and Glover [5,6] describe and analyze the use of linear and goal programming methods for pattern classification and discriminant analysis. Nakayama and Kagaku [11] give a goal programming model that uses a piecewise linear discriminant line for the purposes of classification.

This paper concentrates on the use of distance-metric and preference modelling techniques to build a goal programming model that allows for the setting of parameters and weights so that the probability of correct classification and misclassification at a number of levels can be more accurately controlled by the modeller. The remainder of the paper is divided into three sections. Section 2 develops the model from the basic goal programme by progressively adding multiple classification levels, non-standard preference functions, and distance metrics other than the standard Manhattan (L_1) case. Section 3 presents a cinema-going example which is modelled and solved using the methodology developed in Section 2. Results are discussed from the viewpoints of the methodology and the application. Section 4 then draws conclusions.

2. Model development

Suppose in the training set there exist n_1 observations of type A and n_2 observations of type B. Each type 1 observation has a score of a_{ij} associated with the jth criteria where $j = 1, \ldots, m, m$ being the number of criteria under consideration. Similarly each type B observation has a score of b_{ij} associated with the jth criterion.

The most basic goal programming model for such a pattern classification problem can now be defined as

$$\mathrm{Min}\, z = \sum_{i=1}^{n_1} (n_i^{(a)}) + \sum_{i=1}^{n_2} (p_i^{(b)}) \tag{1}$$

Subject to

$$\sum_{j=1}^{m} a_{ij}x_j + n_i^{(a)} - p_i^{(a)} = x_0, \quad i = 1, \ldots, n_1,$$

$$\sum_{j=1}^{m} b_{ij}x_j + n_i^{(b)} - p_i^{(b)} = x_0, \quad i = 1, \ldots, n_2,$$

$$\sum_{j=1}^{m} x_j = 1,$$

$$-\alpha \leqslant x_j \leqslant \alpha, \quad j = 1, \ldots, m,$$

where the $m + 1$ decision variables $x_0, x_1, \ldots, x_m$ define the coefficients in the equation of the discriminant line given by

$$x_1y_1 + x_2y_2 + \cdots + x_my_m = x_0, \tag{2}$$

where $\mathbf{y}$ is an m-dimensional vector representing the co-ordinates of a generic observation in criterion space. α is a user defined parameter defined so as to be significantly larger than the largest absolute value of the $\mathbf{x}$ vector. This parameter is introduced for reasons of computational simplicity.

For a perfect classification, every type A observation should lie on the positive side of the discriminant line and every type B observation should lie on the negative side of the discriminant line. Every misclassified observation will have either a positive $n_i^{(a)}$ or $p_i^{(b)}$ value. The goal programming achievement function therefore seeks to define the discriminant line so as to minimize the sum of the L_1 distances from the line of the set of misclassified observations. A perfect classification gives a z value of zero.

The basic model can however, lead to a large number of observations that lie exactly on the discriminant line and therefore are not perfectly classified in practice although they have associated positive and negative deviational values of zero. The worst case is for all of the observations to lie on the discriminant line. The probability of this type of situation occurring is greatest with high dimensionality and/or linear dependence of the factors.

To counteract this possibility, and to introduce new levels of certainty of classification a separation zone of size 2β is proposed. This leads to the following amended version of model (1):

$$\text{Min}\, z = \sum_{i=1}^{n_1} \left(n_i^{(a)}\right) + \sum_{i=1}^{n_2} \left(p_i^{(b)}\right) \tag{3}$$

$$\sum_{j=1}^{m} a_{ij}x_j + n_i^{(a)} - p_i^{(a)} = x_0 + \beta, \quad i = 1, \ldots, n_1,$$

$$\sum_{j=1}^{m} b_{ij}x_j + n_i^{(b)} - p_i^{(b)} = x_0 - \beta, \quad i = 1, \ldots, n_2,$$

$$\sum_{j=1}^{m} x_j = 1,$$

$$-\alpha \leqslant x_j \leqslant \alpha, \quad j = 1, \ldots, m.$$

The significant levels of the right hand value representing the discriminant line or lines parallel to it are the mid-section of the zone x_0 and the two constant values of the two lines defining the edge of the zone: $x_0 - \beta$ and $x_0 + \beta$. This leads to the new certainty classes which are denoted 'definite type A', 'probable type A', 'Unclassified', 'probable type B', and 'definite type B'. The parameter β reflects the size of the zone of the probable classification. That is, the larger the value of β, the more likely an observation is to be classed as 'probable' and the less likely as 'definite'. This is an advantage in the terms of avoiding complete misclassifications but a disadvantage in terms of ensuring definite correct classifications. Hence increasing the value of β can be thought of as more conservative behaviour. Initial estimates for β could be found by looking at the size of the deviations for a typical set of decision variables values and setting β to be a fraction of this range. This could be followed by sensitivity analysis to see the effect of smaller and larger β values. The relationship between certainty classes and the right hand side level is given in Table 1.

It is now possible to use a training set of observations to determine the probabilities of various types of errors occurring. The most serious misclassification is a type A observation being classified as definite type B or vice versa. A milder form of misclassification is that of a type A observation being classified as probable type B or vice versa. Two measures of the power of the classification are the percentage of observations correctly classified in the 'definite' classes, and the percentage of observations correctly classified in the 'definite' or 'probable' classes.

Table 1
Certainty classes

Achieved value (f_x)	Classification
$f_x < x_0 - \beta$	Definite B
$x_0 - \beta \leqslant f_x < x_0$	Probable B
$f_x = x_0$	Unclassified
$x_0 < f_x \leqslant x_0 + \beta$	Probable A
$x_0 + \beta < f_x$	Definite A

2.1. Extended model using non-standard preference functions

Model (3) uses a standard linear preference function to penalise the non-achievement of deviational variables. Jones and Tamiz [7] propose a framework that allows any other preference structure to be used. This allows for changes in penalty weights and discontinuous penalties to be utilised at certain points in the objective scale. Model (3) has three points at which such penalty changes may be usefully employed. These are the points of classification class change: $x_0 - \beta$, x_0, and $x_0 + \beta$. Using the Jones and Tamiz methodology to allow weight changes at these points leads to the following goal programming model:

$$\begin{aligned}\text{Min}\, z = {} & W_a \sum_{i=1}^{n_1} \left(u_1 n_{i1}^{(a)} + u_2 n_{i2}^{(a)} + u_3 n_{i3}^{(a)}\right) \\ & + W_b \sum_{i=1}^{n_2} \left(v_1 p_{i1}^{(b)} + v_2 p_{i2}^{(b)} + v_3 p_{i3}^{(b)}\right)\end{aligned} \tag{4}$$

$$\sum_{j=1}^{m} a_{ij}x_j + n_{i1}^{(a)} - p_{i1}^{(a)} = x_0 - \beta, \quad i = 1, \ldots, n_1,$$

$$\sum_{j=1}^{m} a_{ij}x_j + n_{i2}^{(a)} - p_{i2}^{(a)} = x_0, \quad i = 1, \ldots, n_1,$$

$$\sum_{j=1}^{m} a_{ij}x_j + n_{i3}^{(a)} - p_{i3}^{(a)} = x_0 + \beta, \quad i = 1, \ldots, n_1,$$

$$\sum_{j=1}^{m} b_{ij}x_j + n_{i1}^{(b)} - p_{i1}^{(b)} = x_0 - \beta, \quad i = 1, \ldots, n_2,$$

$$\sum_{j=1}^{m} b_{ij}x_j + n_{i2}^{(b)} - p_{i2}^{(b)} = x_0, \quad i = 1, \ldots, n_2,$$

$$\sum_{j=1}^{m} b_{ij}x_j + n_{i3}^{(b)} - p_{i3}^{(b)} = x_0 + \beta, \quad i = 1,\ldots,n_2,$$
$$\sum_{j=1}^{m} x_j = 1,$$
$$-\alpha \leqslant x_j \leqslant \alpha, \quad j = 1,\ldots,m.$$

This model allows for any non-linear continuous preference structure to be used. The class weights W_a and W_b allow for differing importance to be given to the different classes. This can be useful when the significance of misclassification in type A is different to that in type B. It can also be used to allow equal weighting to be given to the two types in the case where the number of observations in each class is substantially different (i.e. $n_1 \gg n_2$ or vice versa).

The interior weights u and v allow for the penalisation of per unit misclassification beyond the associated boundary. The weights u_1 and v_3 are associated with the penalisation of complete 'A as definite B' and 'B as definite A' misclassifications. Increasing these weights will decrease the level of this type of misclassification but is also likely to reduce the power of the test as defined above. The weights u_2 and v_2 correspond to the increase in per unit penalisation beyond the discriminant line and hence are associated with the 'A as probable B' and 'B as probable A' types of misclassification. Increasing these weights will hence lower this type of misclassification but may increase either the power or the probability of complete 'A as definite B' type of misclassification. The weights u_3 and v_1 are associated with the per unit penalisation of deviations away from the 'A as definite A' and 'B as definite B' classes. Increasing these weights will increase the power of the test if measured as the number of observations successfully classified as 'definite' but not reduce the number misclassifications of both the 'A as definite B' and 'A as probable B' types.

The explanation of the weights in the preceding paragraph demonstrates that the modeller can impose the desired power and probabilities of misclassifications by the setting and adjustment of the weighting scheme. It may be necessary to use some interactive or genetic algorithm based approach that allows the decision maker to expressed the desired power and probability of misclassification and then iteratively adjusts the weights solving model (4) for the training set until as close as possible a match is achieved. This in itself is a multiobjective problem as the desired levels of power and misclassification may not be achievable and as close a solution as possible should be found.

2.2. Extension to multiple levels of classification

The models (3) and (4) developed so far allow for four major classification type plus a 'no classification' if the observation lies exactly on the discriminant line. This can be extended to the general case of $2K$ major classes by the following formulation:

$$\mathrm{Min}\, z = W_a \sum_{i=1}^{n_1} \sum_{k=1}^{K} \left(u_k n_{ik}^{(a)}\right) + W_b \sum_{i=1}^{n_2} \sum_{k=1}^{K} \left(v_k p_{ik}^{(b)}\right) \tag{5}$$

Subject to

$$\sum_{j=1}^{m} a_{ij}x_j + n_{ik}^{(a)} - p_{ik}^{(a)} = x_0 + \beta_k,$$
$$i = 1,\ldots,n_1, \ k = 1,\ldots,K,$$
$$\sum_{j=1}^{m} b_{ij}x_j + n_{ik}^{(b)} - p_{ik}^{(b)} = x_0 - \beta_k,$$
$$i = 1,\ldots,n_2, \ k = 1,\ldots,K,$$
$$\sum_{j=1}^{m} x_j = 1,$$
$$-\alpha \leqslant x_j \leqslant \alpha \ j = 1,\ldots,m,$$

where β_k represents the adjustment of the right hand side from x_0 of the kth objective. Normally it is envisaged that symmetry about x_0 would be required. In this case K assumes an odd value, the middle objective corresponds to the discriminant line $\left(\beta_{\frac{K+1}{2}} = 0\right)$, and the other β values have the property that $\beta_k = -\beta_{K+1-k}$. There now exist K levels of misclassification and measures of power of classification.

2.3. Extension to distance metrics other than L_1

The achievement function of the generalised distance-metric formulation of model (5) is given as

$$\operatorname{Min} z = \left[W_a^\rho \sum_{i=1}^{n_1} \left[\sum_{k=1}^{K} \left(u_k n_{ik}^{(a)} \right) \right]^\rho + W_b^\rho \sum_{i=1}^{n_2} \left[\sum_{k=1}^{K} \left(v_k p_{ik}^{(b)} \right) \right]^\rho \right]^{\frac{1}{\rho}}. \tag{6}$$

The achievement function of model (5) is a specific case of (6) with the distance metric parameter ρ set equal to 1. This will give an optimisation that shows relatively little sensitivity to outliers in the data. Yu [14] shows that there exist a range of distance-metric model solutions corresponding to a continuous scale of ρ between 1 and ∞. These solutions will vary between the lesser outlier sensitivity approach at $\rho = 1$ and the inclusive, balanced approach at $\rho = \infty$. Solving the model for distance metrics other than $\rho = 1$ may yield results that better model the preferences of the decision maker. Values of less than $\rho = 1$ are possible too, but these are non-convex and hence harder to solve, especially in the case of large training sets. The metric $\rho = 0$ is established in the literature as the consideration of solely the number of misclassified observations. This requires the use of binary variables and is analogous to the 'discontinuity in preference' case in the Jones and Tamiz preference modelling system. There is an allowance in this system to form hybrids of discontinuities and penalty changes, equivalent to a combination of an $\rho = 0$ and $\rho \geqslant 1$ hybrid distance metric in pattern classification terminology. The algebraic form of such a hybrid model is given as

$$\operatorname{Min} z = \left[W_a^\rho \sum_{i=1}^{n_1} \left[\sum_{k=1}^{K} \left(u_k n_{ik}^{(a)} \right) + \mu_k s_{ik} \right]^\rho + W_b^\rho \left[\sum_{i=1}^{n_2} \sum_{k=1}^{K} \left(v_k p_{ik}^{(b)} \right) + v_k t_{ik} \right]^\rho \right]^{\frac{1}{\rho}} \tag{7}$$

Subject to

$$\sum_{j=1}^{m} a_{ij} x_j + n_{ik}^{(a)} - p_{ik}^{(a)} = x_0 + \beta_k, \quad i = 1, \ldots, n_1, \ k = 1, \ldots, K,$$

$$\sum_{j=1}^{m} a_{ij} x_j + M s_{ik} \geqslant x_0 + \beta_k, \quad i = 1, \ldots, n_1, \ k = 1, \ldots, K,$$

$$\sum_{j=1}^{m} b_{ij} x_j + n_{ik}^{(b)} - p_{ik}^{(b)} = x_0 - \beta_k, \quad i = 1, \ldots, n_2, \ k = 1, \ldots, K,$$

$$\sum_{j=1}^{m} b_{ij} x_j - M t_{ik} \leqslant x_0 - \beta_k, \quad i = 1, \ldots, n_2, \ k = 1, \ldots, K,$$

$$\sum_{j=1}^{m} x_j = 1,$$

$$-\alpha \leqslant x_j \leqslant \alpha, \quad j = 1, \ldots, m,$$

where μ_k and v_k are weights assigned to misclassifications of class A and class B, respectively, and M is an arbitrarily large positive constant. Setting $u_k = v_k = 0 \ \forall k$ gives a pure $\rho = 0$ model.

3. An example

In this application consideration is given to the issue of identifying whether an individual ever goes to a cinema (movie theatre) or does not. In essence, we are considering the characteristics of the "cinema-prone" subset of the population. While it is possible to examine cinema going over a more tightly defined shorter time period, doing so may give misleading results. That a given individual did not go to the cinema during a certain short period may reflect either that none of the available films appealed during the period to that individual, or, that cinema going per se does not appeal. In this analysis by considering the question of whether an individual ever goes to a cinema focuses on the latter reason.

This analysis is facilitated by examination of the CAVIAR (Cinema And Video Industry Audience Research) survey conducted annually by a leading UK market research company—BMRB International for the Cinema Advertising Association. The primary purpose of the survey is to identify the characteristics of the audience for the cinema to help in targeting on-screen advertising. As such there is a slight bias in the sample in that younger respondents are over-represented, however, these younger age groups form the core of the cinema audience, as is also found in the US market [1].

3.1. Data

The CAVIAR survey is conducted as a face-to-face interview with respondents. The survey covers such areas as frequency of cinema going, frequency of video purchases and rentals, and amount of television watched. As such it offers a rich source for academic research. The Caviar 2001 survey is used as a training set. This consists of 1951 respondents, 1688 of whom were regular cinema-goers, and 263 of whom were not. The Caviar 2002 is used as a validation set. This consists of 1932 respondents, 1689 of whom were regular cinema-goers, and 243 of whom were not.

Turning to the variables used in our cinema application it may be noted that while price data is not included in the CAVIAR data set this poses no problem for the subsequent modelling phase since we are focusing on lifetime ever visitation which is unlikely to be inhibited by a single ticket price. The socio-economic group (seg) of the chief income earner in the respondent's household is used as the proxy for income as it is in many other studies. Socio-economic group is defined using a 6-point scale (A being the highest and E the lowest). The seg categories are presented as a series of 4 dummy variables: SEGAB, SEGC1, SEGC2, SEGDE, with socio-economic group E being omitted as the base category.

Two potential substitutes for the cinema visits are video and television. One may argue that video is better described as a complement to cinema going, particularly rented videos, which tend to be feature films. Frequent renting of videos or cinema going may increase awareness of what's available in the other medium. Alternatively, video may be regarded as a substitute for the cinema, especially if the convenience of watching a film at home and at a convenient time is regarded as more important than the wait for the film to be released on video. The frequency of viewing rented and bought videos is captured by two categorical variables: VIDRENT and VIDBUY. These variables take on values from 0 to 9 where 0 is never watch, 1 is less than once a year, 2 is once a year, 3 is 2–3 times a year, 4 is every 2–3 months, 5 is once a month, 6 is 2–3 days a month, 7 is once a week, 8 is twice a week and 9 is watch three times a week or more. This is the format of these variables in the CAVIAR dataset. The variable TVWEEK captures intensity of television viewing. The CAVIAR dataset contains data on number of days per week on which respondents watched television and the number of hours, on average, terrestrial, cable and satellite channels were watched. Respondents are asked to state the number of days per week they watch television and, on days they do watch television, how many hours on average each channel is watched. Television viewing patterns are not likely to be constant; individuals are not likely to watch the same proportions of terrestrial and satellite/cable channels every day. Rather, one channel may be predominantly watched at one time, a different channel may be watched at others. Arguably, the question has been misinterpreted by some respondents: in some instances, the implied total hours of television watched per day exceeded 24 hours, in other cases figures of more than 20 hours were obtained. An alternative proxy for television usage was developed. The highest value reported by each individual was used and multiplied by the number of days television was watched to give an estimate of television watched per week. For example if a respondent reported that he/she watched Channel 1 for 6 hours, Channel 2 for 2 hours, Channel 3 for 7 hours, Channel 4 for 1 hour on average, 7 hours would be used as an indicator of daily television use. This is the best proxy that can be obtained from the data. 'Non-economic' variables are also included in the model. The age of the individual (REALAGE) is expected to be a predictor of attendance; market research has found the core cinema audience to be aged 15–34, with the majority of this subset being 15–24.

The respondent's gender was also included in the model as a dummy variable (GENDER) where 1 indicates male and 0 female. The respondent's gender may capture both the difference (if any) in leisure time which men and women have and the difference in the appeal of the cinema. In the leisure literature it has been suggested that (married) women are likely to have less leisure time than men, depending on their involvement in the labour market and the allocation of household chores in the household. This has been termed "women's double day". Empirical investigations of the distribution of leisure time have reported conflicting results; some finding gender being significant in models of leisure time availability, others finding gender's effect is negligible [13].

3.2. Experimentation

In order to demonstrate the methodology developed in this paper, the following pattern classification models are developed and solved using version 5.0 of the LINGO linear, integer, and non-linear modelling system [9]. An L_0 model is not built due to complexity involved in solving integer programmes with the required number of variables. The largest distance metric used is $\rho = 3$ as higher values give too much consideration to outliers which in turn would not be beneficial in terms of the classification power. All models are solved on a 900 MHz PC.

- **1** A basic goal programming classification model (GPCM) as defined by model (3) with a classification gap of $\beta = 0.15$.
- **2** A GPCM as defined by model (4) with a classification gap of $\beta = 0.15$ and weighting scheme defined by $\mathbf{u} = \mathbf{v} = \mathbf{1}$.
- **3A** A GPCM with variable penalties as defined by model (4) with the following weighting scheme: $\beta = 0.15$, $u_1 = 1$, $u_2 = 2$, $u_3 = 3$, $v_1 = 3$, $v_2 = 2$, $v_1 = 1$.
- **3B** A GPCM with variable penalties as defined by model (4) with the following weighting scheme: $\beta = 0.15$, $u_1 = 3$, $u_2 = 2$, $u_3 = 1$, $v_1 = 1$, $v_2 = 2$, $v_1 = 3$.
- **4** A GPCM with seven levels as classification as defined by (5) with the following parameter set: $\beta_1 = 0.075$, $\beta_2 = 0.15$, $\mathbf{u} = \mathbf{v} = \mathbf{1}$.

Table 2
Model sizes and solution times

Model	Constraints	Non-linear	Variables	Solution time (seconds)
1	3902	n	7812	88
2	5852	n	11,713	196
3A	5852	n	11,713	222
3B	5852	n	11,713	202
4	9752	n	19,513	807
5A	5850	y	11,713	940
5B	5850	y	11,713	1036

- **5A** A Euclidean distance metric model as defined by model (7) with the following parameter set: $\beta_1 = 0$, $\beta_2 = 0.15$, $\mathbf{u} = \mathbf{v} = \mathbf{1}$, $\mu = \nu = \mathbf{0}$, $\rho = 2$.
- **5B** An L_3 distance metric model as defined by model (7) with the following parameter set: $\beta_1 = 0$, $\beta_2 = 0.15$, $\mathbf{u} = \mathbf{v} = \mathbf{1}$, $\mu = \nu = \mathbf{0}$, $\rho = 3$.

The model sizes and solution time using LINGO are given in Table 2.

Tables 3 and 4 records the percentage of classifications of the go and no-go (NG) observations in the validation set.

3.3. Analysis of the results in the context of the methodology

Considering the solution times given in Table 2 it can be seen that the addition of the extra constraints and variables needed to model multiple classification layers increases the solution time but not in an exponential manner, as all models registered acceptable solution times. The increase in solution time between the smallest model (model 1) and the largest of this set (model 5B)

Table 3
Classification percentages for the go observations in the validation set

Class	Model						
	1 (%)	2 (%)	3A (%)	3B (%)	4 (%)	5A (%)	5B (%)
Definite Go	46.2	26.0	32.6	13.0	17.4	26.1	26.1
Intermediate Go	–	–	–	–	33.5	–	–
Probable Go	34.6	56.0	47.8	68.4	30.7	55.9	56.0
Probable NG	13.6	15.6	16.0	17.6	12.7	15.6	15.6
Intermediate NG	–	–	–		4.0	–	–
Definite NG	5.6	2.4	3.6	0.9	1.7	2.4	2.4

Table 4
Classification percentages for the no-go observations in the validation set

Class	Model						
	1 (%)	2 (%)	3A (%)	3B (%)	4 (%)	5A (%)	5B (%)
Definite NG	35.7	17.5	28.3	7.1	12.1	18.3	17.5
Intermediate NG	–	–	–	–	26.2	–	–
Probable NG	27.9	46.7	36.7	55.4	25.9	45.8	46.7
Probable Go	22.5	31.3	29.2	35.8	21.2	31.3	31.3
Intermediate Go	–	–	–	–	12.5	–	–
Definite Go	12.1	4.6	5.8	1.7	2.1	4.6	5.0

is around an order of magnitude. The major increase in computational time occurred when the distance metrics other than 1 (models 5A, and 5B) are used.

The classification probabilities in Tables 3 and 4 give a demonstration of the impact of altering the preference structure on the classification probabilities. The initial model 1 gives the highest probability of correct classification but also the highest probability of total misclassification. Model 3B, by contrast, with its heavier weighting beyond the total misclassification boundary gives the lowest probability of total misclassification but also the lowest probability of correct classification. Models 2 and 3A form intermediates between these two extremes. Thus it can be seen that the probabilities of classification/misclassification can be effectively managed by use of the preference structure. Model 4 gives extra information over and above models 2, 3A, and 3B, with the 'probable' classes separated into two levels to give added information to the decision maker. Models 5A and 5B provides little change in classification power from model 2 which seems poor return for the extra computing time needed.

With reference to the discriminant line given in Table 5, it can be seen that the lines produced by the different models give some variation in magnitude although not in sign. In part this is due to a linear dependence in the socio-economic grouping of variables, which is taken into account when analyzing the results in Section 3.4. However, certain global trends occur, such as the pattern over the socio-economic groups and strength and correlation of the factors. These trends have been used to extract the analysis undertaken in Section 3.4.

Table 5
Numerical coefficients of the discriminant line

Variable	Model						
	1	2	3A	3B	4	5A	5B
OWN	−0.118	−0.047	−0.076	−0.068	−0.039	−0.048	−0.049
KID06	0.058	0.034	0.050	0.049	0.031	0.035	0.035
KID714	0.097	0.070	0.085	0.086	0.057	0.069	0.069
REALAGE	−0.396	−0.264	−0.301	−0.301	−0.226	−0.264	−0.264
TVWEEK	−0.173	−0.151	−0.149	−0.147	−0.115	−0.151	−0.151
GENDER	0.019	0.008	0.003	0.004	0.007	0.008	0.008
SEGAB	0.415	0.377	0.392	0.389	0.354	0.377	0.377
SEGC1	0.342	0.313	0.319	0.317	0.305	0.313	0.313
SEGC2	0.276	0.259	0.266	0.263	0.260	0.259	0.260
SEGDE	0.152	0.204	0.181	0.179	0.201	0.204	0.204
VIDRENT	0.264	0.161	0.178	0.177	0.133	0.161	0.161
VIDBUY	0.064	0.036	0.051	0.052	0.031	0.036	0.037
RHS	0.025	0.035	0.033	0.031	0.043	0.035	0.036

3.4. Analysis of the results in the context of the application

The generated results present a range of interesting points of guidance regarding the relative importance and signs of a number of key variables. Turning first to the socio-economic group variables (the income proxies), one may note that across models 1, 2, 3A, 3B, 4 and 5B they are in ascending order of impact from segDE through to segAB. This indicates support for the view that cinema-going features the characteristics of a normal good by being positively related to income level. This evidence somewhat undermines the more anecdotal views of cinema going, articulated in some quarters, as a source of ‘escapism for the masses’. One frequently cited British study in this vein is that of Spraos [12]. That study sought explanations for the decline in cinema attendance over the period 1950–1959. Those interviewed in the film exhibition sector implied that the cinema was an inferior good, being regarded as entertainment for the ‘masses’ or ‘escapism’ for the poor. These results suggest otherwise. The amount of television viewing (TVWEEK) variable is slightly negative across all the model specifications suggesting this activity is as might be expected, a weak substitute for cinema going. Conversely, the video renting variable is consistently positive across all the models and is suggestive of this activity being weakly complementary with cinema going. In part, this effect may reflect a general positive disposition towards movies in an individual and/or some degree of successful impact arising from the on-screen or on-video movie trailers. Perhaps reflecting the core 16–25 year old cinema-going age group it may be noted that increasing age lowers the likelihood of an individual engaging in cinema going, as does (but more marginally) the prospect of lone cinema going (as indicated by the (OWN) variable). It may also be noted that as expected having children of age 7–14 seems to increase the likelihood of engaging in cinema going presumably as a shared family leisure experience. Having very young children (Kids 0–6) presents more marginal results across the models tested. In the same vein it may be noted that women are less likely to be cinema goers than men. In part these more marginal results may simply reflect the presence of search costs and other problematic issues related to finding suitably qualified childminders/babysitters.

Table 5 gives the numerical coefficients of the discriminant line (2) for each of the models.

4. Conclusions

It has been shown in this paper that goal programming is indeed a flexible tool suitable for forming pattern classification models. Advances in preference modelling techniques have allowed for goal programming pattern classification models that can alter the classification and misclassification probabilities by means of a weighting scheme. Of course the decision maker wishes to both maximise the level of correct classifications and minimize the level of misclassifications. The example demonstrates that these two objectives are in fact conflicting and hence brings out the multi-objective nature of pattern classification. The different weighting schemes in this model can thus be regarded as points on the trade-off curve between the two mentioned objectives. The relationship between different weighting schemes and the levels of classification and misclassification they produce is a topic for future research. The optional addition of multiple levels of classification may give the modeller added options that will be appropriate in certain types of classification that may contain legitimate ‘grey areas’. The use of distance metrics other than L_1 is by no means unknown in the literature. However, the final model (7) is the first overarching distance metric model definition of pattern classification that encompasses all possible distance metric solutions. Although some of its realisations may require a heavy computational burden (e.g. the L_0) and others may give large importance to outliers (e.g. the L_p as $p \to \infty$) it is a valid theoretical model. Many of the other realisations lead to appropriate solvable models. The application described herein demonstrates the promising scope and utility of the harnessing of distance metrics and non-standard preference functions in commercial arenas such as the detection of the more cinema prone segments of a population. Intuitively sensible results are produced with regard to the discerned impact on cinema going of complementary and substitute leisure activities, the impact of children, age and gender.

References

[1] B. Austin, Motivations for moviegoing, Communications Quarterly 34 (2) (1986) 115–126.

[2] W. Baek, J.P. Ignizio, Pattern-classification via linear-programming, Computers and Industrial Engineering 25 (1993) 393–396.

[3] S. Bandyopadhyay, C.A. Murthy, S.K. Pal, Pattern-classification with genetic algorithms, Pattern Recognition Letters 16 (1995) 801–808.

[4] S. Bandyopadhyay, S.K. Pal, C.A. Murthy, Simulated annealing based pattern classification, Information Sciences 109 (1998) 165–184.

[5] N. Freed, F. Glover, Resolving certain difficulties and improving the classification power of LP discriminant-analysis formulations, Decision Sciences 17 (1986) 589–595.

[6] N. Freed, F. Glover, Simple but powerful goal programming-models for discriminant problems, European Journal of Operational Research 7 (1981) 44–60.

[7] D.F Jones, M. Tamiz, Improving the flexibility of goal programming via preference modelling techniques, Omega 23 (1995) 41–48.

[8] D. Kim, Improving prediction performance of neural networks in pattern classification, International Journal of Computer Mathematics 82 (2005) 391–399.

[9] LINGO software, details available at <www.lindo.com>.

[10] R.P. Lippmann, Pattern-classification using neural networks, IEEE Communications Magazine 27 (1989) 47–64.

[11] H. Nakayama, N. Kagaku, Pattern classification by linear goal programming and its extensions, Journal of Global Optimization 12 (1998) 111–126.

[12] J. Spraos, The Decline of the Cinema: An Economist's Report, George Allen and Unwin, London, 1962.

[13] C. Thrane, Men, women and leisure: Scandinavian evidence of gender inequality, Leisure Sciences 22 (2000) 109–122.

[14] P.L. Yu, A class of solutions for group decision problems, Management Science 19 (1973) 936–946.