# M5 Model Tree applied to modelling town centre area activities for the City of Nottingham

T. K. T. Le [1], R.J. Abrahart[2], N. J. Mount[3]
School of Geography, University of Nottingham, Nottingham NG7 2RD
Tel: +44 (0) 115 9515383, Fax: +44 (0) 115 951 5249
Email: [1]lgxktl@nottingham.ac.uk, [2]bob.abrahart@nottingham.ac.uk,
[3]nick.mount@nottingham.ac.uk

## 1. Abstract

For town centre planning and decision making purposes, it is important to acknowledge the related changes in social and economic structure, and to understand the various shifts and trends in the 'life and soul' of a modern urban system. The vitality of town centres can be evaluated through the combination of a set of town centre performance indicators (TCPIs). However, the evolving nature of town centres, coupled with the high variation in characteristics between different urban environments, makes the definition of an appropriate set of TCPIs and the subsequent evaluation of their relative importance a non-trivial process.

The innovation of machine learning methods, in particular, M5 Model Tree (M5MT) algorithms has paved the way for data mining. This machine learning technique offers the advantages of knowledge discovery through analysing the patterns of TCPI combinations. The predicted output map which is built on a set of rules can be used to represent the complex situation of urban dynamics. In other words, the M5MT can incorporate a great deal of statistical thinking, learn efficiently, and automatically produces multi-rule combinations over a set of TCPIs according to the overall quality of the target distribution.

The resulting M5MT technique is transparent and simple with the help of computing intelligence techniques. This machine learning technique is rather new to computer modellers and decision makers in the context of town and urban planning. This reported paper, therefore, introduces and highlights the important choice of the machine learning M5MT technique used in the compilation of TCPIs to foster an improved understanding of the changing distributions in urban modern land use and to support the development of better or more informed planning policies.

In order to examine the advantages of knowledge discovery in TCPI compilations using the machine learning M5MT technique, a single overall composite public mental town centre map was used as the target output of the model. This output resulted from 644 individual attempts at defining and mapping the natural characteristics of Nottingham's town centre using the Tagger toolkit[1] in a web-based GIS online survey (Figure 1). This survey also asked individual participant to identify the contribution of each indicator over a set of TCPIs. The collective responses were used to develop a set of weights that describes the percentage of each individual indicator. This information was compared with the M5MT result that produced multiple rule combinations over a set of TCPIs.

[1] Tagger toolkit is developed by Tim Water and Andy Evans, University of Leeds. This toolkit has been modified by author and uses three different types of spray can tool to shade the areas which users think they belong to the extent of town centre in the web-based GIS survey.

Figure 1: A single overall public mental town centre map
(web-based GIS survey)

In addition, a set of eight associated and lesser-associated TCPIs that agreed over the public responses in defining and delimiting the extent of the Nottingham's town centre was used as the input training set in the M5MT algorithms. This spatial information was transferred into density surfaces using the GIS computing techniques. Each of these surfaces represents commercial, car park, leisure, pedestrian flow, public building, industrial, residential and education indicators, of which the first five are considered associate TCPIs and the remaining are designated lesser-associated TCPIs (Figure 2).

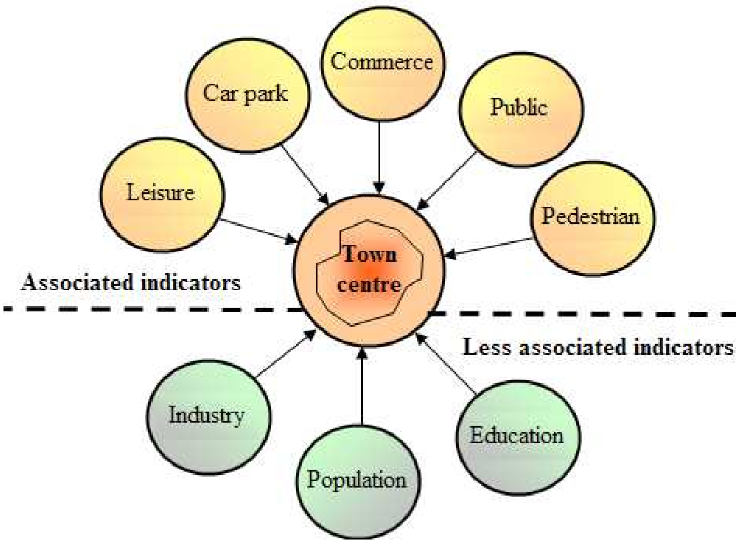


Figure 2: GIS input layers: eight considered TCPIs for Nottingham's town centre study

There is a strong requirement to produce a single "best" output map of town centre area activities. However, it is difficult to capture a universal description in defining and mapping this extent. As seen in Figure 1, the darker the red color is, the more confident those areas belong to town centre area activities. It is much easier to recognize where the town centre starts than to work out where it ends. The problem lies in the different perceptions of the significance of each TCPI and their relative importance in terms of an image[2] of town centre area activities.

This paper stresses the potential use of M5MT technique in discovering the intelligent compilation of TCPIs based on the significance of each individual TCPI and their relative importance in terms of the natural characteristics of town centre area activities. The technique uses the idea of a decision tree by splitting the input space of the training set (*e.g.* town centre area activities) into sub-spaces (sub-areas) and building for each sub-space a linear regression model (at the leaves) that can predict continuous numeric attributes instead of the traditional method of providing a class label for each cluster (Figure 3).

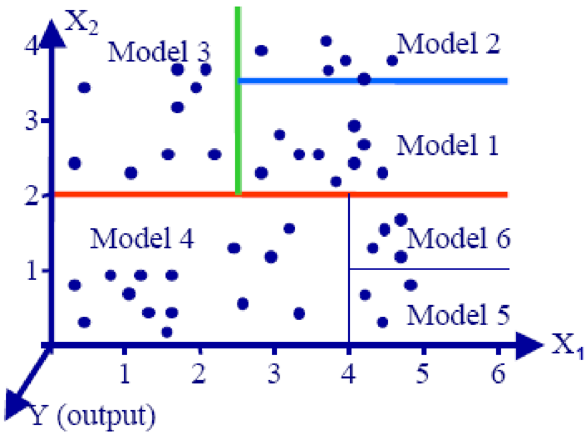


Figure 3: Splitting the input space of the training set [$X_1$, $X_2$] using M5MT algorithm. Each model is a linear regression model $y = a_0 + a_1x_1 + a_2x_2$

By doing this, each linear model at the leaf produces a certain rule for combining eight indicators using different weights. It also tells you about which indicators are predominant and important as the major representation in each leaf. The combination of a set of TCPIs in each leaf was found to be very much similar to public thoughts in defining the extent of town centre area activities as expressed in the questionnaire. However, the model produces multi-rules with a complexity of natural characteristics of urban landscape that the human mind is not able to handle since many complicated things are occurring in different places.

The challenge of this technique is how to choose the sample that is representative. No one can tell whether the sample is representative or not, but one simple thing should follow is that each TCPI is represented in about the right proportion in the training set. A number of randomly selected instances over a set of TCPIs were tested through an iterative (trial and error) testing procedure to find out the appropriate training-set size and the best representation for each TCPI. Figure 4 shows the influence of training-set size on the behaviour of functional trees over a set of TCPIs in terms of the variance of tree models and their correlation coefficient.

[2] The term "image" used in this paper describes the significant activities taken place within a town centre that everyone can obviously recognise in the city.

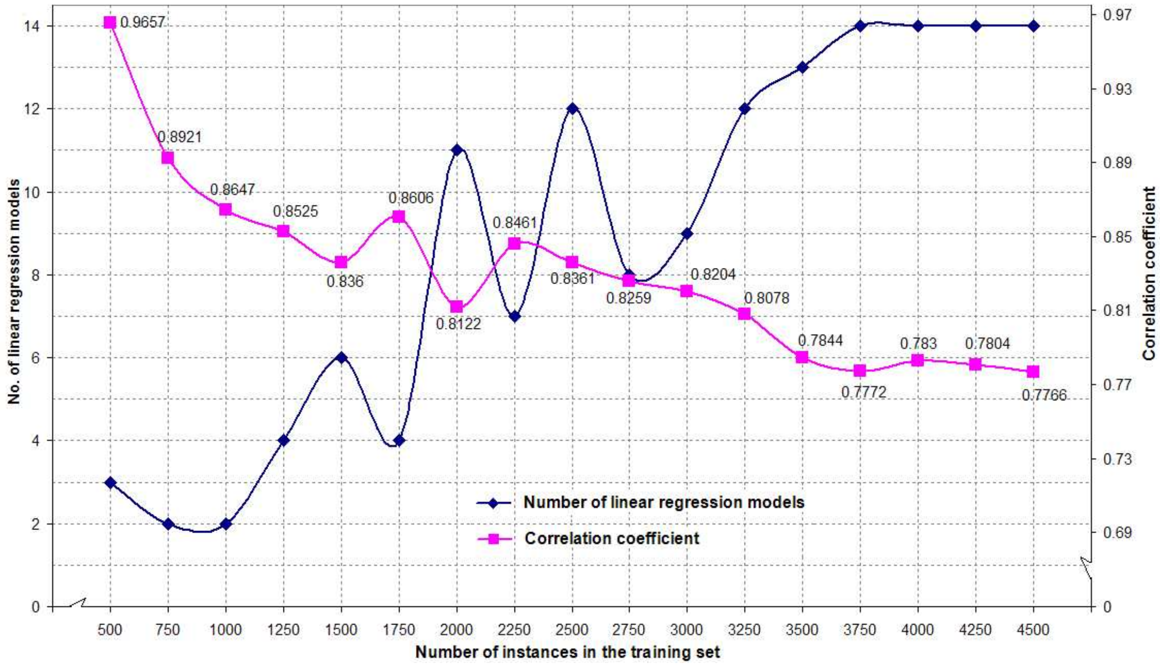

Figure 4: The influence of training-set size on the behaviour of functional trees

The study fixed the training-set size with 4250 instances over a set of TCPIs through the testing procedure. The result of fourteen Linear Models is produced with the correlation coefficient between the actual input dataset with the predicted output town centre being 78.04. Figure 5 shows the tree model with all its leaves.

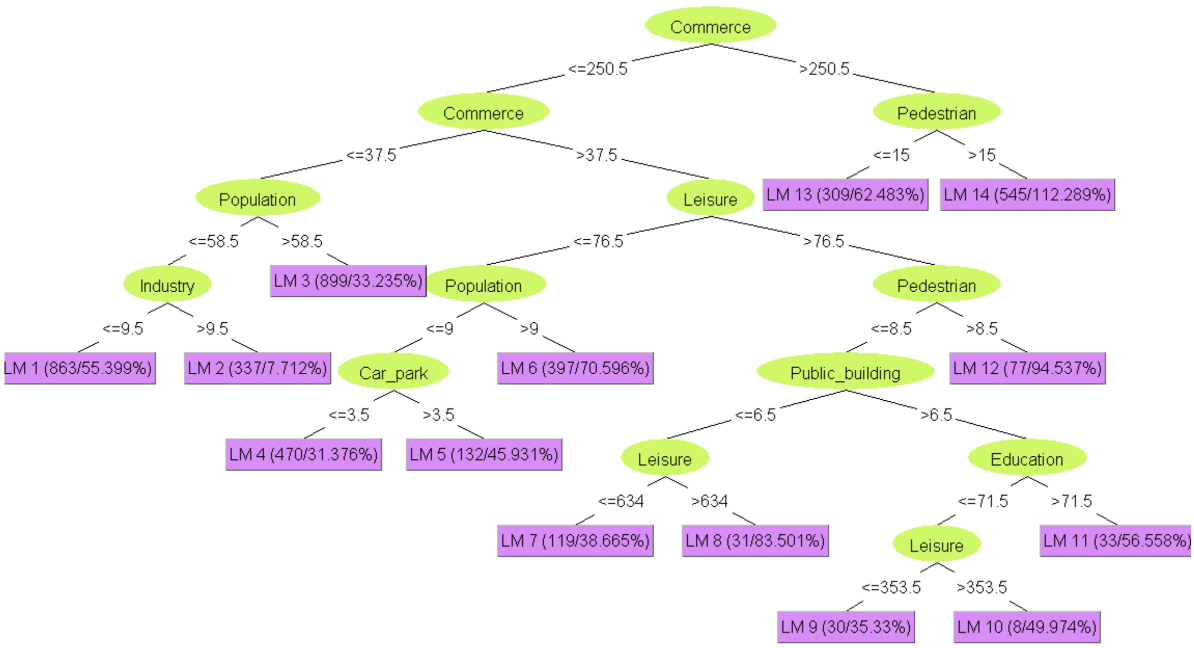


Figure 5: Tree model results from the 4250 instances for eight TCPIs

The above tree model obviously reveals that the commercial factor is considered an important indicator that represents the significant pattern of town centre activities. Following this root, the highly dense concentration of commercial use and pedestrian flow is used to develop Linear Model 14. These areas reflect the diversity of use in the core of town centre where many different activities take place (Figure 6a). Similarly, Linear Model 13 shows the town centre activities based on the high concentration of commercial use rather than other activities (Figure 6b). Linear Model 12 describes town centre area activities when commercial areas are less concentrated, and other areas, mainly leisure and pedestrian flow, are important considerations (Figure 6c). In contrast, town centre area activities will not be represented by those areas having sparse commercial and dense residential use (Linear Model 3; Figure 6d) or which have concentrated industrial activities predominate (Linear Model 2; Figure 6e). By following those leaves, the Model Tree will tell you the story of how significant each indicator is and to which degree these indicators explain the town centre area activities. Compared to multi-criteria evaluation based on public participant questionnaire surveys for determining the weight of each indicator, the result of M5 Model Tree is a reliable and logical for understanding the complexity of modern town centre activities.

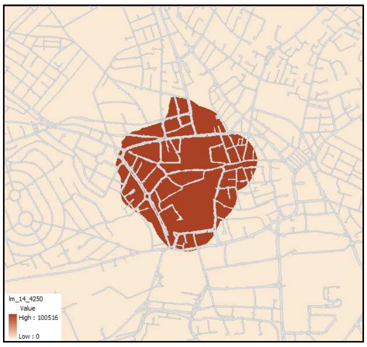

Figure 6a: Linear Model 14

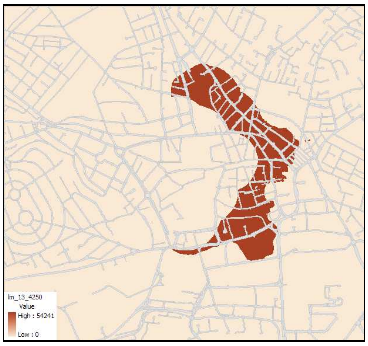

Figure 6b: Linear Model 13

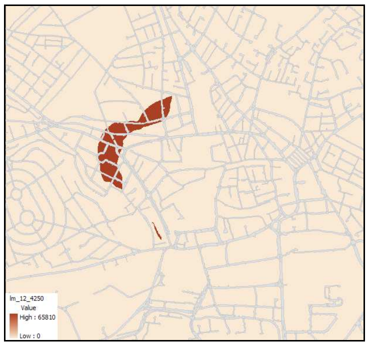

Figure 6c: Linear Model 12

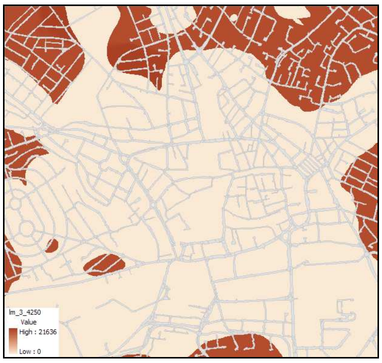

Figure 6d: Linear Model 3

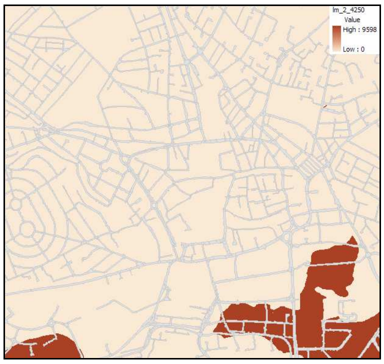

Figure 6e: Linear Model 2

It is believed that the use of machine learning M5MT for knowledge discovery offers a promising and effective technique for evaluating the vitality of towns and cities. This technique is particularly useful for monitoring and modelling the natural changing and complex characteristics of urban dynamics where most cultural, social and economic activities take place.

## 2. Bibliography


CASA- Centre for Advanced Spatial Analysis, 2002, *Producing boundaries and statistics for town centres: London pilot study technical report*, Office of the Deputy Prime Minister, Centre for Advanced Spatial Analysis, University College London.

Kingston R, Carver S, Evans A, and Turton I, 2000, Web-Based Public Participation Geographical Information Systems: An Aid To Local Environmental Decision-Making. *Computers, Environment and Urban Systems*, 24(2), 109-125.

Witten I H and Frank E, 2005, *Data mining: Practical machine learning tools and techniques*, Morgan Kaufmann, San Francisco, CA, 560 pages.