Available online at www.sciencedirect.com
ScienceDirect
Tourism Management 28 (2007) 290–297
TOURISM MANAGEMENT
www.elsevier.com/locate/tourman

Forecasting tourists' characteristics by a genetic algorithm with a transition matrix

Montserrat Hernández-López*, José Juan Cáceres-Hernández

Faculty of Economics, University of La Laguna, Campus de Guajara, 38071 Spain

Received 14 April 2005; accepted 17 November 2005

Abstract

The number of tourists that visit a mature tourism destination becomes stable. If recommendations that could adjust the supply structure towards demand needs in this scenario are established as administration goals, then forecasting the total number of tourists is important, but of particular interest is the internal composition of the demand in terms of specific characteristics of the future groups. This view type of forecasts can be achieved by a genetic algorithm (GA). In this paper, a GA with a transition probability matrix is developed and, as an illustration of the methodology, one run of such an algorithm has achieved better forecasting performance than a simple GA.

Keywords: Tourists characteristics; Forecasts; Genetic algorithms; Transition matrix

1. Introduction

Tenerife (Canary Islands, Spain) is a consolidated tourism area whose tourism activity has a huge social and economic impact. Tourism regulation is one of the topics that concerns its political leaders. Local authorities are now aware that unplanned development of this area can eventually eliminate its attractiveness as a tourism destination. Consequently, the aim is to limit and customize tourism supply so that, firstly, the attraction of new clients does not imply a global reduction in demand in the long term and, secondly, mass tourism with clients characterized by lower purchasing power does not replace tourism of a higher status, as has been observed in traditional destinations.

These previously mentioned reasons show why it is important to forecast demand, not just in aggregate terms but in a way that can anticipate the composition of the tourist groups. That is, the types of tourists who visit a specific tourism site are of interest. Specifically, consider the case of a consolidated area. It would be rather unusual that the total number of tourists experienced significant variations. Nevertheless, individual groups within this aggregate could be subject to reductions when compared to others in the same aggregate.

Standard econometric and statistical techniques do not seem to be the most appropriate method when a productive activity of this nature requires this type of forecast. The usefulness of genetic algorithms (GAs) in order to predict changes in tourism demand has been mentioned by Cho (2003). However, a GA is also a tool capable of explaining changes in the composition of demand based on a set of assumptions. The first assumption is that tourists attempt to maximize utility from their holiday. It is also logical to assume that a tourist who is satisfied with his tourism destination is more likely to return to this destination than one who obtained a lower utility. A second assumption is that a satisfied tourist can influence other individuals with similar characteristics to choose the same destination. A third assumption is that an unsatisfied tourist can influence other potential tourists to avoid a destination that did not satisfy his needs and therefore search for a new destination. Thus, the composition of the tourist group would change as a result of greater participation of those types of tourists

*Corresponding author. Tel.: +34922317032; fax: +34922317042.
E-mail addresses: mhdezl@ull.es (M. Hernández-López), jcaceres@ull.es (J.J. Cáceres-Hernández).

0261-5177/$ - see front matter
doi:10.1016/j.tourman.2005.11.016

that adapt to the tourism supply and, consequently, obtain a higher utility from their holiday.[1]

Hernández-López (2004) applied a simple genetic algorithm (SGA) based on Goldberg's (1989) approach in order to explain changes in the tourism population that visited the south of Tenerife. However, the purpose of the present paper is to show how changes in a tourism population that visits a specific area can be forecasted using a different GA. The new algorithm depends on a transition probability matrix which is specifically adapted to this study. Incorporating such a transition operator is the main contribution of this paper to the literature. The transition matrix is useful to explain changes in a population of economic agents as a function of individual behavioural assumptions. The likelihood of specific changes depends, explicitly, on our knowledge about the observed behaviours.

The rest of the paper is organized as follows. Section 2 describes the characteristics of the GA and the transition matrix used in the study. A transformation of the individuals in the original population is performed based on the assumption that those tourists whose utility level was the highest would represent the greatest percentage of the composition of future tourism populations. So, an intermediate population is obtained. Then, a transition matrix is defined that allows the individuals in the intermediate population to be transformed with fixed probabilities according to specific criteria. Section 3 defines specific elements of the GA applied to a tourism population and the data used as input for the analysis. The results obtained are used to illustrate the behaviour of the algorithm with a transition matrix. This analysis is performed in Section 4 and the results of one run of such an algorithm are compared to those generated from an SGA by Hernández-López (2004). Of course, inferences are not drawn, but it is shown the way in which the accuracy of the forecasts could be achieved. The paper finishes with some conclusions in Section 5.

2. GAs and the transition matrix

Suppose that a tourism population visits a specific destination and that its composition undergoes changes. An increase is expected in the percentage of individuals with characteristics similar to the tourists with highest utility level. GAs are then a useful tool to describe this dynamic process of population transformation. The changes are guided by maximization or simply improvement of a hypothetical fitness function. In this case this fitness function could be defined as the tourist's utility level after the stay. Each tourist could be categorized according to a series of characteristics, such as origin country, length of the stay or type of accommodation. So, given a tourism supply, the utility level could be expressed as a function of these characteristics. Once the fitness function is fitted, the GA is able to forecast changes in the internal composition of the population in terms of higher or lower presence of individuals with specific characteristics.

The first step in the application of the GA consists in formulating a fitness function which provides a value for the utility of the ith tourist, F_i, based on a set of k explanatory variables, $X_i : \{X_{i,1}, \ldots, X_{i,k}\}$, which represent specific characteristics of such a tourist. Once the fitness function is obtained, the performance of the GA depends on specific operators which specify change patterns in tourism population.

Individuals in the population are identified by structures or bit chains that indicate their characteristics. The algorithm proposed in this paper modifies the original population—generation t—in two phases. In the first phase, individuals from the original population are chosen using a selection operator and an intermediate population is obtained. The probability that an individual is selected is proportional to the value of the fitness function for such an individual. It is likely that the percentage of individuals with similar characteristics who showed a high utility level will increase in the next time period when compared to the previous time period. The expected number of times that an individual is replicated depends on the relationship between the sizes of the original and the final populations.

In order to obtain an intermediate population of size n, the proportional selection operator applies the following procedure. If $\Omega_1 : \{I_{1,1} \ldots, I_{1,r}\}$ denotes the set of individuals in the original population of size r and $W : \{1, \ldots, n\}$ is defined as the set of n positions where the individuals copied from the original population are located, then the intermediate population $\Omega_2 : \{I_{2,1}, \ldots, I_{2,n}\}$ is obtained through the selection operator $s(j) = I_{2,j}$, and is defined as $s : W \to \Omega_1$, such that

$$P(s(j) = I_{1,i}) = P(I_{2,j} = I_{1,i}) = p_i = \frac{F_i}{\Sigma_{i=1}^{r} F_i},$$
$$i = 1, \ldots, r, \quad \forall j = 1, \ldots, n,$$

where p_i, $i = 1, \ldots, r$, is the probability of copying individual i and this probability is defined as the quotient of the quality of individual i and the sum of all qualities of the r individuals in the original population.

The intermediate population is obtained by randomly generating the results of n multinomial tests of size r with probabilities $p_1, \ldots, p_r$. The copies of the individuals from the original population are completed according to a specified number and a new population is obtained with the desired size. If the selection operator is defined in this way it can be said that it modifies the original population by transforming it into a new population with the hope that this resultant population is characterized by a higher fitness level.

[1]The adaptive philosophy that inspires genetic algorithms was proposed by Holland (1975), although the methodological elements are attributable to Goldberg (1989). See also Mitchell (2001). Some applications in economics are given by Green and Smith (1987), Arifovic (1994), Arifovic (1995), Sierra and Bonsón (1995), Dawid (1996), Mahfoud and Mani (1996), Hurley, Moutinho, and Witt (1998), Bullard and Duffy (1999), Bee-Hua (2000), Nag and Mitra (2002), Bielli, Caramia, and Carotenuto (2002), Venkatesan and Kumar (2002), Álvarez and González (2003) and Casado-Yusta and Pacheco-Bonrostro (2003).