

# Strategic group identification using evolutionary computation

M.R. Martínez-Torres<sup>a,1</sup>, S.L. Toral-Marín<sup>b,\*</sup>

<sup>a</sup>Escuela Universitaria de Estudios Empresariales, Avda, San Francisco Javier, s/n, 41018 Sevilla, Spain

<sup>b</sup>Escuela Superior de Ingenieros, Avda, Camino de los Descubrimientos, s/n, 41092 Sevilla, Spain

## ARTICLE INFO

### Keywords:

Franchising  
Strategic groups  
Genetic Algorithms  
Evolutionary computation  
Performance

## ABSTRACT

This paper proposes to identify strategic groups among franchisors from a big set of franchisor variables. Genetic evolutionary computation was used to reduce a set of variables efficiently, and factor analysis was used to make up the strategic groups. Franchise 500 was used as database. The results suggest both that the general map of franchisor has changed since Carney and Gedajlovic's study, and that genetic evolutionary computation is a valid way to extract knowledge from a huge set of data. This paper proposes useful information for those retail firms considering internationalization via franchising. The originality of this paper is in the use of Genetic Algorithm to discriminate the final set of variables to be used for the identification of strategic groups instead of evaluating one by one the adequacy of each variable theoretically. The ability of evolutionary computation to create new knowledge is good to produce new insights into this topic.

## 1. Introduction

Franchise strategy refers to the outcome of the decision to operate and expand a business by franchising versus company-ownership (Chen & Ou, 2009; Falbe & Welsh, 1998).

Two major theoretical perspectives have been proposed to explain patterns of company-ownership versus franchisee-ownership: Resource Scarcity and Agency Theory (Alon, 2001; Carney & Gedajlovic, 1991; Combs & Castrogiovanni, 1994; Combs, Ketchen, & Hoover, 2004; Michael, 2003; Newkirk & Lederer, 2006; Paik & Choi, 2007). Both theories should be analyzed, as they determine the firm characteristics that are going to drive franchisors into strategic groups.

Advocates of the strategic group concept argue that industry members can be classified into groups along key characteristics, such as strategy and structure (e.g., Hatten & Schendel, 1977). In general, firms within an industry that follow a similar approach or strategy have been termed strategic groups (Porter, 1980), and strategic groups have been found to differ in performance (Ketchen, Thomas, & Snow, 1993). Opportunities are not evenly distributed across an industry; some strategies offer better profit potential than others. Firms may be tempted to change strategies to exploit opportunities as they arise, but shifting to a new strategic group can be risky because the necessary investment can be

substantial and the perceived opportunities may be short lived (Patterson & Smith, 2003).

All firms do not adopt franchising for similar reasons, but rather groups of firms share similar approaches (Carney & Gedajlovic, 1991). If strategic groups exist among firms that franchise and these groups differ in performance, examining all franchising firms as a set cannot capture the true picture of the franchising-performance relationship (Combs et al., 2004).

The typical approach to strategic group identification consists of collecting detailed industry data, and then identifying groups through clustering or other grouping algorithms (Shanley & Peteraf, 2005; Sohn & Kim, 2008). The variables used to group will have much influence on the identified groups. A new perspective is proposed in this paper. Instead of theoretically evaluating one by one the adequacy of each variable, a big set of them is used. Using an appropriate fitness function, an evolutionary algorithm will discriminate the final set of variables to be used for the identification, as well as the resulting strategic groups. We have used Genetic Algorithms (GA) for this purpose. One of the key advantages of evolutionary computation is its ability to discover new knowledge. The evolving nature of the computation can establish new relationships considered never before (Nanni & Lumini, 2009).

The next section is devoted to the theoretical perspectives to franchising. Then, strategic groups and performance are introduced. After that, the evolutionary technique used in this paper (Genetic Algorithm) is shown: first, the basis of the methodology is described, and then the particular application to the strategic group identification problem is detailed. The obtained results are illustrated later. Finally, some conclusions have been drawn.

\* Corresponding author. Tel.: +34 954 48 12 93; fax: +34 954 48 7373.  
E-mail addresses: rmtorres@us.es (M.R. Martínez-Torres), toral@esi.us.es (S.L. Toral-Marín).

<sup>1</sup> Tel.: +34 954 55 43 10; fax: +34 954 55 69 89.

## 2. Theoretical perspectives to franchising

Oxenfeldt and Kelly (1968) offered perhaps the first explanation of why the proportion of franchised outlets differs among franchisors. Under the Resource Scarcity view, franchisors use franchising as a way to overcome constraints to growth, including a lack of trained managers and a lack of financial capital (Doherty, 2007). Success requires financial, informational, and managerial resources, but these resources are hard to obtain in the short run (Dant, Kaufmann, & Paswan, 1992). The franchisee provides an infusion of capital through fees and royalties and offers the franchisor (relatively) inexpensive growth. However, subsequent research tended to focus on the fact that firms used franchising because they also needed human capital (Norton, 1968), managerial talent (Dant et al., 1992; Doherty, 2007; Falbe & Welsh, 1998), and local knowledge (Combs & Castrogiovanni, 1994).

Viewing franchising primarily as a means to access resources, a firm's propensity to franchise varies over time. An implied tenet of Resource Scarcity Theory is the belief that the firm is more likely to increase company-ownership of sites as franchise systems mature and accumulate resources. This is precisely one of the criticisms that the Resource Scarcity thesis has received. Combs and Castrogiovanni (1994) observed that in contrast to the predictions of Resource Scarcity Theory, larger firms actually used more franchising in their development. They also found a weak negative relationship between the age of the franchisor and the use of franchising, and no relationship at all between the growth of the franchisor and the use of franchising. Some corporate giants as McDonald's and Budget Rent-A-Car endorse this asseveration.

The theory is also criticized by economists on the basis that capital can be raised more efficiently in the market. Although franchising may lower risk for the franchisor, it increases risk for the franchisee (Rubin, 1978).

Agency Theory explains the organization of relationships when one agent determines the work and another undertakes it (Mole, 2002; Shane, 1998). In franchising, the Agency Theory perspective discusses it as the relationship between one party (the franchisor) depending on another party (the franchisee) to undertake some action on the franchisor's behalf (Paik & Choi, 2007). Franchising encourages franchisees to maximize effort because, as owners, they must devote their own capital to build and operate outlets (Brickley & Dark, 1987). As a consequence, franchising lowers the cost of monitoring (Dant & Kaufmann, 2003; Pizanti & Lerner, 2003). Managerial talent and local knowledge are also eased by franchising because of the franchisees' risk to lose their upfront monetary investments if their outlets fail as a result of their own managerial inadequacies (Shane, 1998). Quite the opposite of the Resource Scarcity Theory, Lafontaine and Kaufman suggests that agency factors favour an increased use of franchising as a chain expands with maturity (Lafontaine & Kaufmann, 1994).

However, some agency problems are not solved by franchising. There are some situations in which the franchisee may be inclined to shirk by under investing and free riding at the expense of the chain's reputation (Michael, 2000). Although monitoring cost can be decreased, transferring specific knowledge to potential franchisees can also be costly (Jensen & Meckling, 1995).

Some authors have proposed a reconciliation of both theories. Martin and Justis (1993) found that short- and long-run incentives to franchise differ. Whereas resource scarcity reasons to franchise are most relevant for young franchisors seeking to expand, agency reasons become increasingly relevant with maturity (Castrogiovanni, Combs, & Justis, 2004).

## 3. Strategic groups and performance

Researchers have long suspected the presence within industries of subgroups of firms whose behaviours and results differ from those of the broader industry (Lee, Lee, & Rho, 2002; Porter, 1976, 1979). Currently, there are three theoretical perspectives regarding strategic group formation: the industrial organization, strategy, and cognitive/behavioural perspectives (Hoyt & Sherman, 2004).

The industrial organization perspective defines strategic groups as persistent features of the industry structure that result from entry and mobility barriers. Structural barriers impede new firms from entering new industries (Audretsch, Houweling, & Thurik, 2004). Firms may be tempted to change strategies to exploit opportunities as they arise, but shifting to a new strategic group can be risky because the necessary investment can be substantial, and the perceived opportunities may be short lived (Wheeler, Ibeh, & Dimitratos, 2008). Thus, firms generally choose not to change groups because it is unclear whether the performance enhancements gained will exceed the costs incurred (Mascarenhas & Aaker, 1989). This is a consequence of the empirically determined recognition that single groups are separated by barriers which restrict the strategic mobility of firms (Caves & Porter, 1977).

The strategy perspective is more internally focused and thus assumes that the firm's management makes decisions to configure internal resources, so as to establish a sustainable competitive advantage (Hirschsohn, 2008).

Finally, the cognitive perspective contends that strategic groups are formed by managers who partition their environment to reduce uncertainty and who possess bounded rationality (Peteraf & Shanley, 1997).

Based on strategic groups' theory as well as the evidence linking strategic groups and performance, we expect that some strategic groups among franchisors will have strategic profiles that are better suited to their environment than others. Nevertheless, identifying strategic groups involves more than just choosing initial groups and data sources. The variables used to group will have much influence on the identified groups. A detailed enumeration of nearly one hundred of such variables from prior studies has been detailed in Ketchen et al. (1993). Even if we were only concentrated in franchisor studies, the number of possible variables or indicators is huge. Insufficient consideration has been given to determine which variables are appropriate for the purpose of distinguishing strategic groups.

To avoid this, instead of theoretically evaluating one by one the adequacy of each variable, a big set of them will be used in this paper. Using an appropriate fitness function, an evolutionary algorithm will discriminate the final set of variables to be used for the identification, as well as the resulting strategic groups.

## 4. Methodology: Genetic Algorithms

To identify strategic groups, we must first identify the firm characteristics that are likely to drive franchisors into distinctive strategic groups. We have used Genetic Algorithms (GA) for this purpose. Genetic Algorithms (GA) have been used to solve a variety of optimization problems, such as natural gas pipeline control, structural optimization, image registration, job scheduling, path planning, product design, etc. (Fisco, 2003; Goldberg, 1989; Lee et al., 2002; Li, Deng, & Luo, 2009).

A Genetic Algorithm is a computational abstraction of biological evolution which can be used to solve some optimization problems. The technique was first introduced by Holland (1975) for use in adaptative systems. It is an iterative process applying a series of genetic operators such as selection, crossover and mutation to a

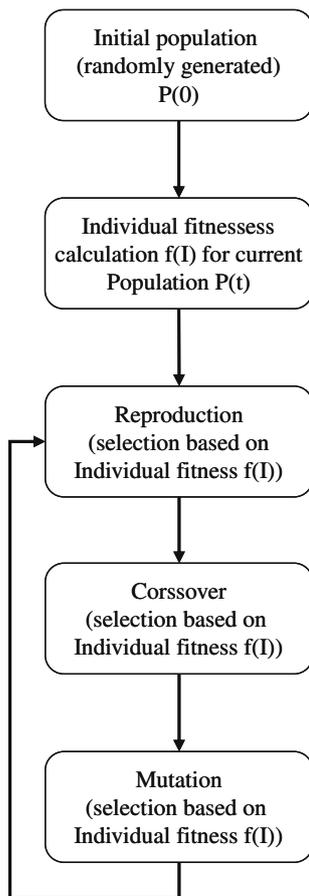


Fig. 1. Iterative process of Genetic Algorithm.

population of elements (Fig. 1). These elements, called chromosomes or individuals, represent possible solutions to the problem. The initial population is randomly selected from the solution space.

Genetic operators combine the genetic information of the elements to form new generations of populations. Each chromosome has an associated fitness value which quantifies its value as a solution to the problem. The chromosomes compete to reproduce based on their fitness values, thus the chromosomes representing better solutions have a higher chance of survival. The crossover involves two chromosomes whose portions are swapped. Selection according to fitness combined with crossover gives the GA its evolutionary power. The GA uses an elitist strategy meaning that the best individual is carried over to the next generation so that we can only improve the solution over the course of the genetic optimization. GA follows the iterative scheme shown in Fig. 1. The algorithm stops when some stopping criterions are satisfied.

To successfully apply GA to solve a problem, it must be considered the following items:

- Chromosomal encoding, how to represent possible solutions.
- Fitness function selection. It must accurately represent the value of the solution.
- Parameter values selection (population size, number of iterations, probabilities, etc.)

#### 4.1. Genetic Algorithm application to strategic group identification

Prior studies about strategic group identification among franchisor have only considered a few numbers of variables. For instance, Carney and Gedajlovic (1991) considered thirteen

Table 1

Candidate variables for factor analysis.

Var.	Description
Var01	No. of outlets 2005
Var02	Increment of outlets from 2004 to 2005
Var03	% of outlets located in US, 2005
Var04	% of outlets located in US, 2004
Var05	Average number of outlets opening per year
Var06	Average number of outlets opening per year since first franchise
Var07	Average investment
Var08	Standard deviation of investment
Var09	Average franchise fee
Var10	Standard deviation of franchise fee
Var11	Ongoing royalty fee
Var12	Term of agreement
Var13	Net worth requirement
Var14	Cash liquidity requirement
Var15	Number of employees per unit
Var16	Franchises more than 1
Var17	% of outlets franchised 2005
Var18	% of foreign outlets franchised 2005
Var19	Years since inception
Var20	Years between inception and first franchise
Var21	Ranking Franchise 500, 2005
Var22	Ranking Franchise 500, 2004
Var23	Average investment/Term of agreement
Var24	Average investment/Net worth requirement
Var25	Average investment/Cash liquidity requirement
Var26	% of outlets franchised in Canada, 2005
Var27	Average investment/Average franchise fee
Var28	Franchise department
Var29	No of outlets franchised/Franchise department
Var30	Foreign franchises/Franchise department
Var31	(Franchise fee + ongoing royalty fee) * term of agreement

variables for Canadian franchisor strategic group identification, while (Castrogiovanni, Bennett, & Combs, 1995) just used twelve variables for a similar study among US franchisors. Some other studies have replicated the former study in different countries (López & Ventura, 2001).

All of them have utilized factor analysis as the multivariate statistical tool for strategic group identification (Hsia, Hsu, & Jen, 2009; Sohn & Kim, 2008). Conventional techniques such as multiple regressions suffer the inconvenience of some restrictions like the independence of predictor variables or assumptions like linear relationships between variables.

Factor Analysis is a way to fit a model to multivariate data to estimate just their interdependence. In the factor analysis model, the measured variables depend on a smaller number of unobserved (latent) factors. Because each factor may affect several variables in common, they are known as "common factors". Each variable is assumed to be dependent on a linear combination of the common factors, and the coefficients are known as loadings. Factor analysis is useful in identifying inter-relationships between variables which are not directly observable, segmenting a sample into relatively homogeneous segments.

The estimated loadings from an unrotated factor analysis fit can usually have a complicated structure. The goal of orthogonal factor rotation is to find a parameterization in which each variable has only a small number of large loadings, i.e., is affected by a small number of factors. The rotated factor analysis fit ensures factors represent unidimensional constructs (Hsu, 2009).

Instead of using the same variables than Carney and Gedajlovic (1991) and Castrogiovanni et al., 1995 to replicate both studies, we propose to use a big set of variables as possible candidates for factor analysis. A total of 31 candidate variables have been considered, and they will be applied for strategic group identification of US franchisors included in the Entrepreneur's Franchise 500 survey, January 2005. Table 1 details the variables used in this study.

The objective of this work is to find the optimum subset of variables from Table 1 leading to optimum strategic group identification. The obtained results will show if the thirteen variables considered in previous studies are enough, or if it would be necessary to extend the number of considered variables to achieve a better representation of US franchisor groups. But the number of possibilities when working with such huge number of variables suggests this problem is unapproachable if trying to explore all of them. The space of possible solutions is formed by  $2^{31} = 2.1475e + 009$  possibilities. That means that we would have to perform  $2^{31}$  different factor analyses to completely explore the space of possible solutions. In this kind of problems, evolutionary computation can perform a guided search of the optimum solution with lower computation cost than exploring one by one all the possibilities.

In this task, Genetic Algorithm will be applied to perform the evolutionary search of the optimum solution. The first condition to apply GA properly is a good selection of the chromosomal encoding, which should be valid and complete. Our chromosomal encoding is constituted by a 31 binary sequence in which ones are the variables that are going to be used in factor analysis and zeros represents variables that are going to be excluded from this analysis. Clearly, the encoding representation is complete, as the  $2^{31}$  possibilities are able to be represented, and valid, as all of them can be computed.

The next step is the fitness function selection. The fitness function quantifies the suitability of each chromosome as a solution. Chromosomes with high fitness have more chance of being selected, passing their genetic material (via reproduction or crossover) to the next generation. The fitness function provides the pressure for evolution towards a new generation with chromosomes of higher fitness than the previous ones. The chromosome representing the optimal solution should have the maximum fitness value for the solution space, and near optimal solutions should have high fitness values. In the context of strategic group identification by means of factor analysis, it is not possible to build a simple fitness function. Fitness function should be multi-objective fitness function considering several parameters.

$$F = c_1 Var + c_2 \frac{1}{n} \sum_{i=1}^k r_i^2 + c_3 Dist$$

This proposed fitness function consists of three parts.

- Explained variance (*Var*). Factor analysis results show the explained variance by the considered factors (usually, the number of factors is given by the number of eigenvalues of the correlation data matrix bigger than 1). The explained variance by the selected number of variables should be maximized. But it is not the unique parameter to be taken into account. A fitness function equal to the explained variance will tend to the trivial solution of just considering one variable. This is due to the fact that it is easier to explain the variance of a data set when it is formed by a small number of data.
- Correlations between variables ( $1/n \sum_{i=1}^k r_i^2$ ). The average of the sum of the squared correlation coefficients between variables is used as the second part of the fitness function. This part of the fitness function will tend by itself to the trivial solution of considering the whole data set. It is the reverse strength to the previous part of the fitness function.
- Number of factors. The third part of the fitness function measures the distance (*Dist*) to the typical number of factor of previous studies about strategic group identification. A distant number from 5 groups is penalized, as this is a typical value obtained en prior studies (Carney & Gedajlovic, 1991; Castrogiovanni et al., 1995).

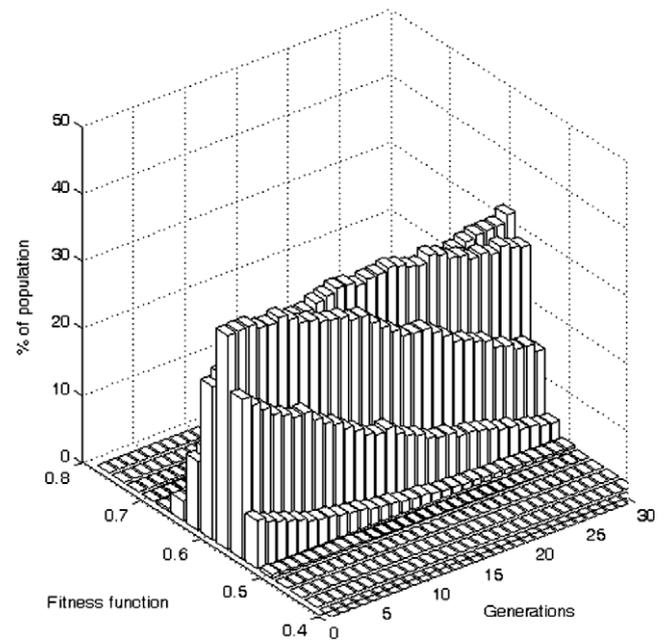


Fig. 2. Fitness distribution over 30 generations of the genetic algorithm.

$C_1$ ,  $C_2$ , and  $C_3$  are coefficients used to adjust the relative importance of the three parts of the fitness function. The range of them is  $[0, 1]$ , with the restriction of  $C_1 + C_2 + C_3 = 1$ .

The final decision for GA application refers to parameter values selection. GA performance may be sensitive to certain parameter values, particularly the population size, the frequency of operator selection and the termination criterion. All of them vary considerably, and there is little or no documented justification for the selection of them. Nevertheless, a high value for the population size may reduce this sensibility to GA parameters. In this paper, population size has been chosen equal to 10,000, with a 20% of reproduction rate. The value of 10,000 is considered a good value to obtain richness genetic content. These values are typical in the literature about GA (Cole, 1998; Goldberg, 1989).

## 5. Results

Beginning with an initial randomly generated population, GA has converged after 30 generations, with an explained variance of 82.8%, 20 variables and six factor (that means six eigenvalues higher than 1).

Time required by genetic algorithm execution is 750 s (12.5 min). This value is much smaller than the alternative option of exploring the whole solution space. Taking into account that each factor analysis requires 2.5 ms, the  $2^{31} = 2.1475e + 009$  possibilities of the solution space would require more than 62 days, 24 h a day.

The evolution of the genetic clustering algorithm is detailed in Fig. 2. The initial population (generation 0) has a low fitness value, which indicates that the individuals of the population are far from the optimum. As the number of generations increase, the fitness of individuals within the population also increases, as the genetic algorithm is biased towards the survival of genetic material contained within the individuals with high fitness function values.

The results from factor analysis using the set of variables selected by the genetic algorithm are detailed in Table 2.

Although the number of eigenvalues higher than 1 is six, due to the gap between the fifth and the sixth eigenvalue is preferable to choose just five components or factors. Factor loadings with

**Table 2**  
Explained variance of resulting factor analysis.

Component		Eigenvalues		
		Value	Proportion of total variance	Cumulative proportion
Var01	No. of outlets 2005	5127	25.636	25.636
Var02	Increment of outlets from 2004 to 2005	4438	22.191	47.826
Var03	% of outlets located in US, 2005	2391	11.953	59.780
Var04	Average number of outlets opening per year	1861	9305	69.084
Var05	Average number of outlets opening per year since first franchise	1561	7804	76.888
Var06	Average investment	1189	5943	82.831
Var07	Standard deviation of investment	.984	4921	87.752
Var08	Net worth requirement	.608	3041	90.793
Var09	Cash liquidity requirement	.493	2464	93.257
Var10	% of outlets franchised 2005	.367	1834	95.091
Var11	% of foreign outlets franchised 2005	.271	1354	96.445
Var12	Years since inception	.198	.992	97.437
Var13	Years between inception and first franchise	.174	.870	98.307
Var14	Ranking franchise 500, 2005	.107	.537	98.844
Var15	Average investment/term of agreement	.090	.451	99.295
Var16	Average investment/net worth requirement	.061	.307	99.602
Var17	Average investment/cash liquidity requirement	.035	.176	99.778
Var18	% of outlets franchised in Canada, 2005	.027	.136	99.914
Var19	No of outlets franchised/franchise department	.013	.065	99.978
Var20	Foreign franchises/franchise department	.004	.022	100.000

**Table 3**  
Factor loadings obtained by the principal component method with Varimax rotation.

		Component				
		1	2	3	4	5
Var01	No. of outlets 2005	.014	.794	.521	.054	.088
Var02	Increment of outlets from 2004 to 2005	.019	-.074	.723	-.119	.068
Var03	% of outlets located in US, 2005	.004	-.623	.187	-.084	-.583
Var04	Average number of outlets opening per year	-.027	.639	.666	-.022	-.156
Var05	Average number of outlets opening per year since first franchise	-.008	.567	.739	-.098	-.043
Var06	Average investment	.896	-.004	-.046	.344	.035
Var07	Standard deviation of investment	.936	-.012	-.031	.172	-.016
Var08	Net worth requirement	.177	8690E-05	-.088	.918	.074
Var09	Cash liquidity requirement	.148	.020	-.044	.907	.064
Var10	% of outlets franchised 2005	.126	.025	.146	-.435	-.615
Var11	% of foreign outlets franchised 2005	.098	.783	-.081	-.266	.229
Var12	Years since inception	.176	.224	.187	.062	.758
Var13	Years between inception and first franchise	.168	-.066	.158	-.107	.821
Var14	Ranking franchise 500, 2005	-.036	-.300	-.686	-.146	-.233
Var15	Average investment/term of agreement	.903	-.007	-.105	.306	.049
Var16	Average investment/net worth requirement	.839	-.013	.077	-.227	.214
Var17	Average investment/cash liquidity requirement	.867	.005	.106	-.195	.054
Var18	% of outlets franchised in Canada, 2005	.074	.201	-.281	-.362	.019
Var19	No of outlets franchised/franchise department	-.070	.826	.299	-.003	-.030
Var20	Foreign franchises/franchise department	-.047	.904	.028	.035	.012

varimax rotation are shown in Table 3. Each row represents the factor loadings of each variable. Moving horizontally from left to right across the five loadings in each row, the highest loading has to be identified. All the variables associated in this way with the same factor are hypothesized to share a common meaning that the analyst should discover.

The interpretation of the rotation loadings leads to the resulting strategic groups, represented by the different clusters identified (Table 4).

To facilitate the interpretation of these, Table 5 shows the mean value of the selected variables per each of the five factors.

Cluster 1 brings together variables related with investment. It is characterized by a high value of all the loadings and it can be resembled to the "expensive" group identified by Carney and Gedajlovic (1991). The identified group has anyway some differences with respect to the one they proposed, as they are not so conservative. They exhibit a moderate rate of growth, according to var04. Expensive franchisors are primarily concerned with resource scar-

city, demonstrated by the high value of all the variables in which investment is involved.

Cluster 2 refers to highly internationalized US franchisors. Var03, referring to the percentage of outlets located in US, shows a negative loading in Table 4, corresponding to a minimum value in Table 5 when compared with the other factors. They are characterized by a great confidence on franchising as a way of minimizing the agency cost. This conclusion is suggested by the high values of the ratios given by the number of outlets and the foreign outlets, and the size of franchise department (Var19 and Var20, respectively).

Cluster 3 is somewhat related with the previous one. They are also very big franchisor, with a high rate of growth, but, as a difference with the previous cluster, they are focused on domestic market. According to Table 5, var03 has the maximum value in cluster 3, just the opposite than in cluster 2. Var14 also exhibit a negative loading in Table 4. Var14 is the ranking 2005 in Franchise 500. The lower this value is, the more stability and reliability the franchisor

**Table 4**  
Clusters identified by factor loadings.

Cluster	Var.	Description	Loading
Cluster1	Var06	Average investment	0.896
	Var07	Standard deviation of investment	0.936
	Var15	Average investment/term of agreement	0.903
	Var16	Average investment/net worth requirement	0.839
	Var17	Average investment/cash liquidity requirement	0.867
Cluster2	Var01	No. of outlets 2005	0.794
	Var03	% of outlets located in US, 2005	-0.623
	Var11	% of foreign outlets franchised 2005	0.783
	Var19	No of outlets franchised/franchise department	0.826
	Var20	Foreign franchises/franchise department	0.904
Cluster3	Var02	No. of outlets 2004	0.723
	Var04	Outlets opening per year	0.666
	Var05	Outlets opening per year since first franchise	0.739
	Var14	Ranking Franchise 500, 2005	-0.686
Cluster4	Var08	Net worth requirement	0.918
	Var09	Cash liquidity requirement	0.907
Cluster5	Var10	% of outlets franchised 2005	-0.615
	Var12	Years since inception	0.758
	Var13	Years between inception and first franchise	0.821

is (the first franchisor of the ranking is the best one). So, it is natural a minimum value of var14 in Table 5 meaning that this cluster is formed by the best franchisor according to Franchise 500 criterions. Reasons for franchising are closer to resource scarcity than to the agency theory.

Cluster 4 is formed by conservative franchisor. They exhibit the lower rate of growth of all the groups (var04 and var05), with a lot of requirements about franchises, particularly net worth and cash

liquidity requirements. They are conservative, and moderate expensive compared with franchisors included in cluster 1, primarily concerned with resource scarcity.

Finally, cluster 5 can be assimilated to franchise converts group of Carney and Gedajlovic (1991) study. Converts typically begin operations as wholly-owned chains, and after many years decide to begin franchising.

According to the obtained results, it can be said that franchisors are generally concerned with both administrative efficiency and resource scarcity, though their concern over one relative to the other may vary somewhat with their particular circumstances.

In order to compare the means of variables from Table 5, a Kruskal-Wallis test has been performed. The Kruskal-Wallis test is a nonparametric version of one-way analysis of variance. The assumption behind this test is that the measurements come from a continuous distribution, but not necessarily a normal distribution. The test is based on an analysis of variance using the ranks of the data values, not the data values themselves. The low *p* value for each variable suggests that the mean of each one of them is significantly different than the other sample means. The null hypothesis can be rejected, so it can be concluded that the obtained strategic groups are discrete (see Table 6).

## 6. Conclusions and implications

The first conclusion that can be emphasized is that the general map of franchisor since Carney and Gedajlovic study has changed. Franchising as a way of growth has been adapted to the current circumstances through the years. Highly internationalized franchisors appears now as a clearly distinguished group. Probably, the old group of mature franchisor has split into the groups identified

**Table 5**  
Selected variables mean values per factor.

	Factor 1 (10.91%)	Factor 2 (19.09%)	Factor 3 (25.45%)	Factor 4 (27.27%)	Factor 5 (17.27%)
VAR01	1321.30	5336.60	4054.00	1393.30	2717.30
VAR02	67.25	94.86	342.89	45.35	87.05
VAR03	0.87	0.60	0.90	0.83	0.61
VAR04	44.11	128.16	161.89	43.66	46.64
VAR05	62.05	139.26	194.32	40.00	75.01
VAR06	2250.80	180.84	126.92	528.51	428.64
VAR07	2152.50	108.22	73.73	293.72	192.55
VAR08	364.50	189.55	181.73	550.77	354.79
VAR09	127.51	61.44	54.07	236.88	112.82
VAR10	1.00	0.98	0.99	0.87	0.74
VAR11	0.12	0.39	0.09	0.05	0.21
VAR12	27.92	29.62	25.11	28.23	54.26
VAR13	8.42	3.52	4.25	2.20	26.16
VAR14	67.42	62.38	38.00	65.83	47.74
VAR15	136.01	15.06	14.32	39.59	33.05
VAR16	5.59	0.92	1.30	1.01	1.88
VAR17	14.65	3.63	4.41	2.41	4.13
VAR18	0.04	0.09	0.03	0.01	0.03
VAR19	52.58	411.80	207.80	86.43	118.13
VAR20	4.87	221.68	17.57	5.61	28.57

**Table 6**  
Kruskal-Wallis test.

	VAR01	VAR02	VAR03	VAR04	VAR05	VAR06	VAR07	VAR08	VAR09	VAR10
Chi-2	14.944	21.454	40.163	17.565	23.990	40.960	33.109	26.825	38.161	31.975
gl	4	4	4	4	4	4	4	4	4	4
Sig. asintót.	.005	.000	.000	.002	.000	.000	.000	.000	.000	.000
	VAR11	VAR12	VAR13	VAR14	VAR15	VAR16	VAR17	VAR18	VAR19	VAR20
Chi-2o	40.835	27.346	31.750	14.993	36.867	22.947	24.957	19.485	11.943	20.370
gl	4	4	4	4	4	4	4	4	4	4
Sig. asintót.	.000	.000	.000	.005	.000	.000	.000	.001	.018	.000

by clusters 2 and 3. The former has adopted franchising as a way of internationalization, while the latter is still confident in franchising as a way to growth, primarily in US market.

The group “expensive and conservative” has been also split into two groups. Finally, convert franchisors remain the same.

The second conclusion is the application of evolutionary computation, particularly Genetic Algorithm, as a valid way to extract knowledge from a huge set of data. Replication of a previous study is useful to confirm obtained results, but the inconvenience is that replication does not allow new insights in a particular topic. The ability of Genetic Algorithm to explore new relationships within a big amount of data allows exploring new solutions or new perspectives in a particular theme, like the ones mentioned above.

## References

- Alon, I. (2001). The use of franchising by US – based retailers. *Journal of Small Business Management*, 39(2), 1–12.
- Audretsch, D. B., Houweling, P., & Thurik, A. R. (2004). Industry evolution. Diversity, selection and the role of learning. *International Small Business Journal*, 22(4), 331–348.
- Brickley, J. A., & Dark, F. H. (1987). The choice of organizational form: The case of franchising. *Journal of Financial Economics*, 18, 401–420.
- Carney, M., & Gedajlovic, E. (1991). Vertical integration in franchise systems: Agency theory and resource explanations. *Strategic Management Journal*, 12(8), 607–629.
- Castrogiovanni, G. J., Bennett, N., & Combs, J. G. (1995). Franchisor types: Reexamination and clarification. *Journal of Small Business Management*, 33(1), 45–54.
- Castrogiovanni, G. J., Combs, J. G., & Justis, R. T. (2004). Toward a reconciliation of resource and agency views on franchising. In *Academy of management meeting*, New Orleans, LA (pp. 1–7).
- Caves, R. E., & Porter, M. E. (1977). From entry barriers to mobility barriers: Conjunctural decisions and contrived deterrence to new competition. *Quarterly Journal of Economics*, 91, 241–261.
- Chen, F. L., & Ou, T. Y. (2009). Gray relation analysis and multilayer functional link network sales forecasting model for perishable food in convenience store. *Expert Systems with Applications*, 36(3), 7054–7063.
- Cole, R. M. (1998). Clustering with genetic algorithm. Ph.D. Dissertation, University of Western Australia.
- Combs, J. G., & Castrogiovanni, G. J. (1994). Franchisor strategy: A proposed model and empirical test of franchise versus company ownership. *Journal of Small Business Management*, 32(2), 37–48.
- Combs, J. G., Ketchen, D. J., Jr., & Hoover, V. L. (2004). A strategic groups approach to the franchising–performance relationship. *Journal of Business Venturing*, 19, 877–897.
- Dant, R. P., Kaufmann, P. J., & Paswan, A. K. (1992). Ownership redirection in franchised channels. *Journal of Public Policy and Marketing*, 11(1), 33–44.
- Dant, R. P., & Kaufmann, P. J. (2003). Structural and strategic dynamics in franchising. *Journal of Retailing*, 79, 63–75.
- Doherty, A. M. (2007). The internationalization of retailing. Factors influencing the choice of franchising as a market entry strategy. *International Journal of Service Industry Management*, 18(2), 184–206.
- Falbe, C. M., & Welsh, D. H. B. (1998). NAFTA and franchising: A comparison of franchisor perceptions of characteristics associated with franchisee success and failure in Canada, Mexico, and the United States. *Journal of Business Venturing*, 13, 151–171.
- Fisco, J. H. (2003). Optimal dispersion of R&D activities in multinational corporations with a genetic algorithm. *Research Policy*, 32, 1381–1396.
- Goldberg, D. A. (1989). *Genetic Algorithm – in Search, Optimization and Machine Learning*. Addison-Wesley Publishing Company, Inc.
- Hatten, K. J., & Schendel, D. E. (1977). Heterogeneity within an industry: Firm conduct in the US brewing industry, 1952–1971. *Journal of Industrial Economics*, 26, 97–113.
- Hirschsohn, P. (2008). Regulating the ‘Animal Spirits’ of entrepreneurs? Skills development in South African small and medium enterprises. *International Small Business Journal*, 26(2), 181–206.
- Holland, J. (1975). *Adaptation in natural and artificial systems*. Ann Arbor, MI: University of Michigan Press.
- Hoyt, J., & Sherman, H. (2004). Strategic groups, exit barriers and strategy decision constraints in high-tech companies. *Management Research*, 15, 237–247.
- Hsia, T.-C., Hsu, Y.-L., & Jen, H.-L. (2009). A factor analysis based selection process for predicting successful university color guard club members. *Expert Systems with Applications*, 36(2), 2719–2726.
- Hsu, L.-C. (2009). Forecasting the output of integrated circuit industry using genetic algorithm based multivariable grey optimization models. *Expert Systems with Applications*, 36(4), 7898–7903.
- Jensen, M. C., & Meckling, W. (1995). Specific and general knowledge and organizational structure. *Journal of Applied Corporate Finance*, 8, 4–18.
- Ketchen, D., Thomas, J., & Snow, C. (1993). Organizational configurations and performance: A comparison of theoretical approaches. *Academy of Management Journal*, 36, 1278–1313.
- Lafontaine, F., & Kaufmann, P. (1994). The evolution of ownership patterns in franchise systems. *Journal of Retailing*, 70(2), 97–113.
- Lee, J., Lee, K., & Rho, S. (2002). An evolutionary perspective on strategic group emergence: A Genetic algorithm-based model. *Strategic Management Journal*, 23, 727–746.
- Li, X., Deng, Z., & Luo, J. (2009). Trading strategy design in financial investment through a turning points prediction scheme. *Expert Systems with Applications*, 36(4), 7818–7826.
- López, M. B., & Ventura, J. (2001). Grupos estratégicos en las franquicias españolas: Características básicas. *Economía Industrial*, 340, 163–176.
- Martin, R., & Justis, R. (1993). Franchising, liquidity constraints and entry. *Applied Economics*, 25, 1269–1277.
- Mascarenhas, B., & Aaker, D. A. (1989). Mobility barriers and strategic groups. *Strategic Management Journal*, 10, 475–485.
- Michael, S. C. (2000). The effect of organizational form on quality: The case of franchising. *Journal of Economic Behavior and Organization*, 43, 295–318.
- Michael, S. C. (2003). First mover advantage through franchising. *Journal of Business Venturing*, 18, 61–80.
- Mole, K. (2002). Business advisers' impact on SMEs: An agency theory approach. *International Small Business Journal*, 20, 139–162.
- Nanni, L., & Lumini, A. (2009). A genetic encoding approach for learning methods for combining classifiers. *Expert Systems with Applications*, 36(4), 7510–7514.
- Newkirk, H. E., & Lederer, A. L. (2006). Incremental and comprehensive strategic information systems planning in an uncertain environment. *IEEE Transactions on Engineering Management*, 53(3), 380–394.
- Norton, S. (1968). An empirical look at franchising as an organizational form. *Journal of Business*, 61, 197–217.
- Oxenfeldt, A. R., & Kelly, A. O. (1968). Will successful franchise systems ultimately become wholly-owned chains? *Journal of Retailing*, 44(4), 69–83.
- Paik, Y., & Choi, D. Y. (2007). Control, autonomy and collaboration in the fast food industry. A comparative study between domestic and international franchising. *International Small Business Journal*, 25(5), 539–562.
- Patterson, P. G., & Smith, T. (2003). A cross-cultural study of switching barriers and propensity to stay with service providers. *Journal of Retailing*, 79, 107–120.
- Peteraf, M., & Shanley, M. (1997). Getting to know you: A theory of strategic group identity. *Strategic Management Journal*, 18, 165–186.
- Pizanti, L., & Lerner, M. (2003). Examining control and autonomy in the franchisor–franchisee relationship. *International Small Business Journal*, 21(2), 131–159.
- Porter, M. (1976). *Interbrand choice, strategy and bilateral market power*. Cambridge, MA: Harvard University Press.
- Porter, M. (1979). The structure within industries and companies' performance. *Review of Economics and Statistics*, 61, 214–227.
- Porter, M. (1980). *Competitive strategy*. New York: Free Press.
- Rubin, P. H. (1978). The expansion of firms. *Journal of Political Economy*, 81(4), 936–949.
- Shane, S. (1998). Explaining the distribution of franchised and company-owned outlets in franchise systems. *Journal of Management*, 24, 717–739.
- Shanley, M., & Peteraf, M. (2005). Balancing theory and technique: Methodological issues in strategic groups research. *Research Methodology in Strategy and Management*, 2, 65–92.
- Sohn, S. Y., & Kim, Y. (2008). Searching customer patterns of mobile service using clustering and quantitative association rule. *Expert Systems with Applications*, 34(2), 1070–1077.
- Wheeler, C., Ibeh, K., & Dimitratos, P. (2008). UK export performance research. *International Small Business Journal*, 26(2), 207–239.