Implementing TURF analysis through binary linear programming [1]

Daniel Serra[2]
Department of Economics and Business
UPF

Abstract

This paper introduces the approach of using Total Unduplicated Reach and Frequency analysis (TURF) to design a product line through a binary linear programming model. This improves the efficiency of the search for the solution to the problem compared to the algorithms that have been used to date. The results obtained through our exact algorithm are presented, and this method shows to be extremely efficient both in obtaining optimal solutions and in computing time for very large instances of the problem at hand. Furthermore, the proposed technique enables the model to be improved in order to overcome the main drawbacks presented by TURF analysis in practice.

Key words: Total Unduplicated Reach and Frequency; Market research; Competitive algorithm; Product optimization; Large datasets

---

[1] This Project has been financed in part by the Spanish Ministry of Education and Science, National Plan for the Promotion of Knowledge ECO2009-11307. The author wishes to recognize the work of Ana Micaelef and the comments by John P. Ennis that improved considerably this paper.

[2] Author for correspondence: daniel.serra@upf.edu. Department of Economics and Business, Universitat Pompeu Fabra.

1. Introduction

TURF (Total Unduplicated Reach and Frequency) analysis is a technique used in the world of marketing to optimize product lines. Specifically, it involves selecting the combination of product variants that will ensure that the overall product reaches its maximum penetration. Its applications are extensive, both in the world of mass commodities, and in durable goods and even in services (Conklin & Lipovetsky, 2000, Cohen 1993). The technique is applicable when we have a product that we want to launch on the market as a range in which only one attribute changes, for example, the choice of different flavours for an ice-cream, colours for an MP3 player or fragrances for an air-freshener.

In 1990 Miaoulis, Free and Parsons presented the technique. It involves an adaptation of tools from the world of advertising where the aim is to design a communication plan that will reach the highest number of potential customers. Transferred to the world of marketing, the technique concerns choosing out of all the possible combinations the product line that, in a similar way, will attract the highest number of potential consumers. The problem prioritizes new consumers over those who duplicate consumption, what it tries to maximize is the penetration of the product line as a whole, not of each of the variants that make it up.

An example will help to clarify the concepts. Suppose we have three variants of a product that are candidates for forming a product line. In our case, we will assume that we are limited to putting 2 varieties out of the possible 3 onto the market. We have to find the optimal combination that will achieve the greatest penetration of the product line. The following table shows the situation considered and the data used for the example.

	A	B	C
Interviewee 1	1	1	0
Interviewee 2	1	1	0
Interviewee 3	1	0	0
Interviewee 4	0	0	1
Total	**3**	**2**	**1**

If we were seeking to maximize the penetration of each of the variants separately, then we should opt for marketing the combination A+B, as each of them has 3 and 2 potential consumers, respectively. However, our goal is to maximize the penetration of the product line as a whole, and since the two consumers who choose B have already been reached by variant A, including B in the combination does not bring any increase in penetration, with the penetration for the complete range being 3 individuals. If, on the other hand, we choose the combination A+C we are increasing the overall product penetration by 1 individual, because C has been chosen by a consumer that did not choose A. Our overall penetration will be 4, the maximum that we could reach.

The calculation method proposed by the authors is that of exhaustive enumeration, calculating the overall penetration of all the possible combinations. This guarantees the optimal combination. However, as is acknowledged in the same article, as the alternatives being considered in the problem increase, the computation time required increases exponentially and therefore, as the authors propose, a more efficient search algorithm should be found using mathematical programming methods.

In 2000, Kreiger and Green proposed an alternative method to exhaustive enumeration: the greedy algorithm. This method consists of choosing, firstly, the alternative that attracts the highest number of individuals, including it in the final solution. Then the alternative that achieves the greatest penetration taking into account only the individuals who were not attracted by the first alternative is chosen, in other words, the unduplicated penetration is calculated, conditioned to the first alternative chosen. The process continues until all the individuals are covered or until none of the remaining alternatives manages to increase the overall penetration. As the authors themselves show, this process does not guarantee the optimal solution (Mullet, 2001).

Ennis et al. (2011) presented an exact algorithm to solve the problem with large datasets. The algorithm is based on the principle of non-synergy, which is a way of reducing the number of iterations when complete enumeration is used to solve TURF situations.

Markowitz (2004, 2005) extends the TURF model to different business marketing situations and compares the TURF model with existing others.

Adler et al. (2010) propose to run TURF on discrete choice data on which hierarchical Bayesian methods have been used to predict individual utilities on each of a large number of potential products.

Conklin et al. (2004) and Conklin & Lipovetsky (2005) observe that it is often impossible to distinguish between subsets of different flavour combinations with practically the same level of coverage. They introduce in the model the Shapley Value (SV), also known as the fuzzy Choquet integral, tool borrowed from cooperative game theory, that permits the ordering of flavours by their strength in achieving maximum consumers' reach and provides more stable results than TURF. Lipovetsky (2008) adds to the TURF model the Lazarsfeld's Latent Structure Analysis (LSA), tool applied to problems in marketing involving the choice of products with maximum customer coverage. The LSA is combined with the TURF technique, and also with the SV. The SV is used for ordering the items by their strength in covering the maximum number of consumers, which provides more stable results than TURF. The blending of LSA with TURF and SV yields new abilities of the latent structured TURF and SV. The marketing strategy based on using these techniques permits the identification of the preferred combinations in media or product mix for different population segments.

In the following section we present the TURF model as an integer linear programming problem, the method which we establish as the framework for the analysis in order to apply an optimal search algorithm. Furthermore, the new method is applied to a real case, extended in such a way that some of the main drawbacks of TURF analysis are overcome. A series of adaptations of the model to multiple situations that may arise in real life are proposed. Finally, several randomly generated instances of the problem have been solved using Lingo, a commercial software used to solve linear, integer and binary linear problems using the Simplex algorithm, and branch and bound when needed. The results obtained through our exact algorithm are presented, and this method shows to be extremely efficient in obtaining optimal solutions and being extremely efficient in computing time for very large instances of the problem at hand.

2. Application of a binary linear programming model to TURF analysis

The problem we are going to study below comes under the group of binary linear programming models, linear programming because all the functions of the model (both the target function and all the constraints) are linear and binary functions because the variables we are going to introduce will be variables that can only have values of zero or one.

Once the model has been established, a series of mathematical algorithms can be applied that let us obtain the solution to the problem. In this regard, there are exact algorithms, which guarantee the optimal solution, and heuristic algorithms, which do not always give optimal solutions but do provide a good sub-optimal solution. In this case, we will use an exact algorithm, thus assuring that we will obtain the optimal solution.

As we have indicated above, the problem consists of finding the product line that, overall, attracts the highest possible number of customers. The person responsible for the product could present it as follows:

- *What is the minimum number of varieties I have to put on the market in order to attract the maximum possible number of buyers?*
- *What varieties should I market?*

To answer these questions, the first thing we have to do is compile the data. As required by TURF analysis, data observation is carried out through surveys in which the interviewee is asked to rate each of the proposals based on how attractive it is. This analysis does not need the use of a certain scale or volumetric measure (Miaoulis, Free, & Parsons, 1990), and it is common practice to use an "intention to buy" scale. The data we have to input in the model are binary data (buy / not buy) and therefore a criterion has to be established to distinguish between the two. Normally the highest score ("top box") or highest two scores ("top two boxes") of the scale are used as buy and the rest as not buy.

With the data matrix obtained, we will have for each individual consulted the alternatives he or she is willing to buy and those he/she is not willing to buy, in other words, we can calculate the penetration of each alternative separately and the penetration of each product line as a whole.[3]

The following phase consists of translating our real problem into a mathematical model. More precisely, the model will be cast as a binary linear program.

Let us suppose a problem in which we have n alternative flavours for an ice-cream. We carry out a survey of m individuals as to what flavours they would buy and what

[3] We should point out that in the original database it is necessary to screen out those individuals who do not show an interest in any of the varieties presented. We will not be able to attract them to any combination and therefore they should not be included in the problem. The model presented below works provided that this screening of the data is carried out. Otherwise, an easily adaptable alternative model has to be used.

flavours they would not. We have a data matrix in which for each individual we know the set of flavours he or she would be willing to buy. We can define the following set:

$N_i = \{j \,/\, \textit{consumer i chooses variety j}\}$

In other words, for every individual it is the set of flavours he or she indicates that he or she would be willing to buy. In this way we introduce the data in the model.

We also need to know what varieties we will put on the market, and what others we will not. These will be the binary variables of our model:

$$x_j = \begin{cases} 1, \textit{ we put } \text{variety } \textit{ j on the market} \\ 0, \textit{ we don't} \end{cases}$$

The problem, as we have presented it, consists of finding the minimum combination of varieties that means that all the individuals can buy at least one of the varieties they have chosen. Therefore, our target function will consist of keeping to a minimum the number of product varieties we will put on the market, in other words:

$$Min\ Z = \sum_{j=1}^{n} x_j$$

However, bearing in mind that all the interviewees have to be able to buy at least one of the varieties they have chosen, that is:

$$\sum_{j \in N_i} x_j \geq 1 \qquad i = 1, \ldots, m$$

We force at least one of the varieties chosen by each individual to be put on the market.[4]

Thus, the final formulation of the TURF model will be:

$$Min\ Z = \sum_{j=1}^{n} x_j$$

[4] If we work with a database that has not been screened, the constraint $\sum_{j \in N_i} x_j \geq 1$ will not let us solve the problem because not all the individuals will be buyers. In this case, the constraint has to be reformulated as follows:

$$\sum_{j \in N_i} x_j \geq c_j$$

With $c_j = 1$ if the individual shows interest in buying at least one variety, $c_j = 0$ if the individual does not show interest in any variety.
In any case, we reiterate the advisability of working with a screened database.

s.t.

$$\sum_{j \in N_i} x_j \geq 1 \qquad i = 1,\ldots,m$$

$$x_j = \{0,1\} \qquad j = 1,\ldots,n$$

The use of this model fulfils the purpose of the TURF analysis, because when an individual has already been “reached” by a variety, he or she is no longer considered for the choice of the rest of varieties. Thus the problem takes into account new buyers and does not consider at any time to what extent individuals who are already buyers duplicate the purchase of the product with more than one variety.

Observe that the mathematical formulation of the problem at hand corresponds to the Location Set Covering Problem, formulated by Toregas & ReVelle (1972, 1973). This problem identifies the minimal number and the location of facilities, which ensures that no demand point will be farther than the maximal service distance from a facility. An updated description of the problem can be found in Daskin et al. (1999). This formulation has been extensively studied and a myriad of exact, heuristic and metaheuristic algorithms have been developed to solve it. These algorithms tend to be very efficient both in finding optimal solutions and in computing time for very large problems. See for example Almiñana & Pastor (1997) and Caprara et al. (1999).

3. Improving TURF

3.1. Main criticisms of classic TURF analysis

Although in theory TURF analysis appears to be a very useful tool in decision-making in this type of problems, it has received important criticisms, especially due to the fact that in practice the results are not as clear as would be expected. Below we detail some of the main drawbacks of the analysis:

- The most important problem of this tool is that it usually provides very similar (if not equal) results for different product lines, making it difficult to decide on the choice of a single combination.
- Furthermore, it usually includes in the solution varieties that only attract a small number of consumers and which are not very attractive to a broader public. This leads, in practice, to the analysis recommending that “strange” varieties be marketed that

contribute very little added penetration and which obtain results that are not very favourable in other indicators.

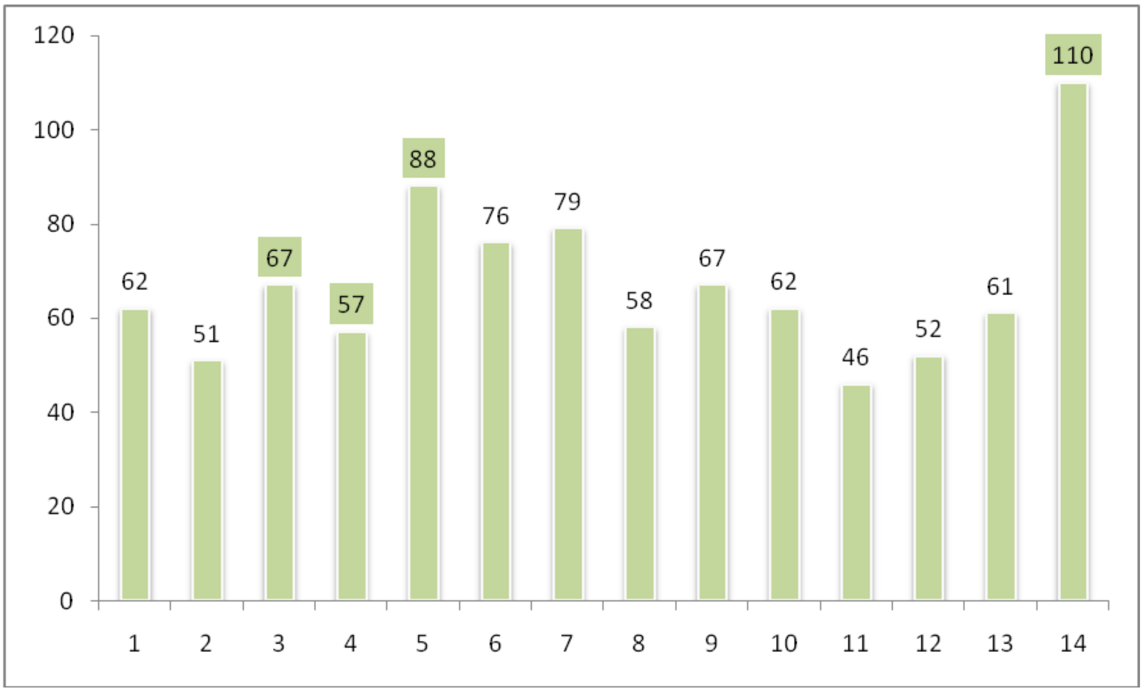
– On the other hand, marketing personnel do not have total control over the product line marketed. It may occur that, due to lack of shelf space or lack of stocks, the product line is not fully at the point of sale at the time the consumer makes his or her choice. These cases are not contemplated in the analysis because it only calculates the result assuming that all the varieties are available.

3.2. Possible improvements of the analysis

Bearing in mind the criticisms of TURF analysis, we present a procedure that in light of the practical results presented below would appear to be an adequate system for overcoming the main problems of this analysis, always taking into account the information that TURF provides us with in order to gain a better understanding of the behaviour of the product options we handle.

The data we are going to use to illustrate the procedure come from a real sample obtained for a study in which a total of 14 product options were presented to a sample of 150 individuals. The individual penetration of each of the different varieties is shown in graph 1.

Graph 1: individual penetration of each of the different varieties



Application of the basic TURF model shows us that with four varieties we reach the maximum possible penetration, which is 127 individuals, in other words, with the optimal combination of varieties we can reach 85% of the potential consumers. The analysis also shows us what varieties we have to choose for our product line (varieties shown as shaded areas in the above graph). As can be seen, there are varieties included in the final solution with an individual penetration notably lower than others that are not included.

It seems advisable to check whether there are other possible optimal combinations for our product line. A quick way of exploring possible alternative optimal solutions is to force each of the varieties not included in the first combination found to have a value of one, and in each case we recalculate the TURF. If the problem does not find an optimal solution or increases the minimum number of varieties necessary to obtain maximum penetration, the variety in question will never become part of an optimal combination. On the contrary, in other words, if by forcing a variety to be put on the market we obtain a combination of four varieties that reaches maximum penetration, this variety forms part of an optimal solution.

In our case, on carrying out the process described above, we reached the conclusion that there are indeed several optimal solutions. So, what is the best solution out of all the optimal solutions?

The answer to this question depends fundamentally on the specific goals of each problem. However, a good generic solution can be obtained by weighting each of the varieties according to its individual penetration. In other words, if we have two varieties that could form part of an optimum combination, we will opt for the variety which individually obtains greater penetration. As we have already discussed in the introductory section, the analysis prioritizes new buyers over buyers who duplicate varieties, however, under equal conditions we will always opt for the varieties that are contributing more penetration. In this way we also allow the consumer to have the greatest possible range of products letting him or her vary consumption in the case of goods with a high buying frequency or simply to have more options when making his or her choice.

The adaptation of the model is very simple, all we have to do is weigh the binary variables of the target function on the basis of the individual penetration of each of them so that if a variety attracts a higher number of buyers, it has priority for forming part of the optimal solution over other less attractive varieties.

Since in the basic problem we seek to minimize our target function, we have to weigh using a lower coefficient the greater the penetration of the variety it is associated to, for example the inverse of the individual penetration of the variety:

$$Min\,Z = \sum_{j=1}^{n} \frac{1}{p_j} x_j$$

s.t.

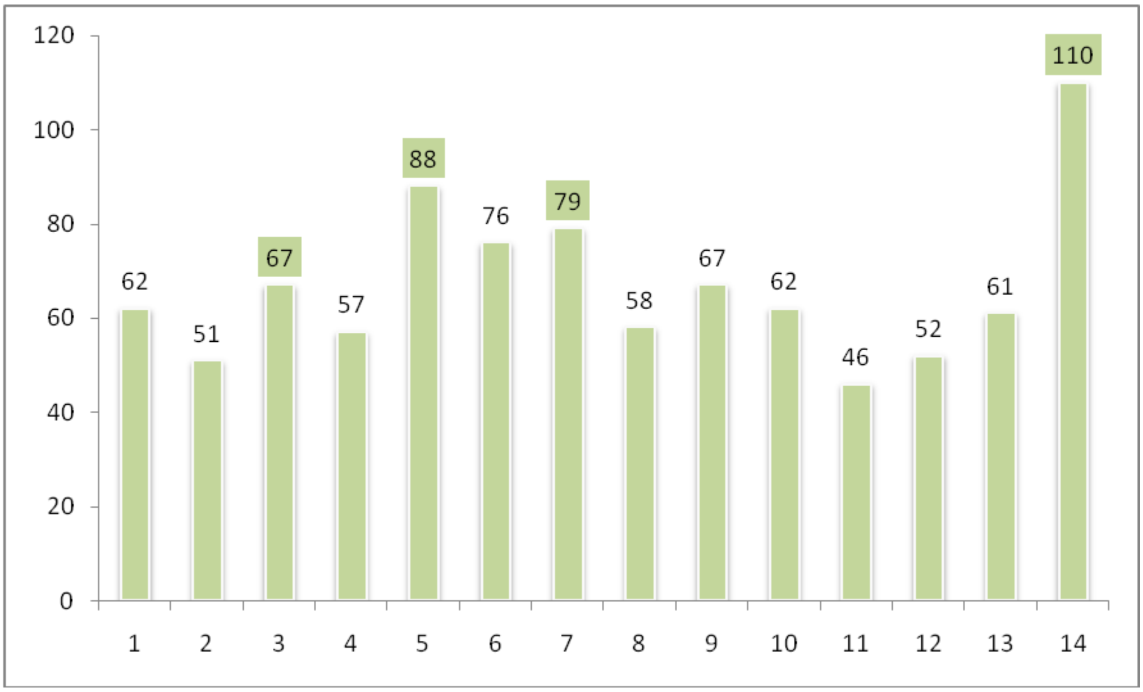
$$\sum_{j \in N_i} x_j \geq 1 \qquad i = 1,\ldots,m$$

$$x_j = \{0,1\} \qquad j = 1,\ldots,n$$

Where p_j is the individual penetration of each variety.

Applying this system to our example gives the following combination of varieties:

Graph 2: individual penetration of each of the different varieties



In other words, we change variety number *4* of our initial solution (the fourth-worst individual penetration) and replace it with number *7* (third in the ranking of attractiveness to buying), without affecting our overall penetration.

The use of this adaptation of the model appears, in light of the results, to be fairly recommendable. In addition, it also lets us overcome the main drawbacks of the analysis we mentioned above:

- It lets us highlight one out of the possible optimal combinations, and therefore facilitates decision-making considerably.
- Out of all the optimal combinations, we choose those with the best possible results in individual penetration therefore reducing the possibility of finding varieties that only attract very minority publics.
- Furthermore, if the product line is not fully available at the point of sale, we know that at least individually each of the varieties available to the public at that time will be a sufficiently attractive variety, and therefore problems of lack of stocks or lack of shelf space will be less critical.

3.3. Comparison of solutions

On the other hand, we must not forget that all the information we handle is not drawn directly from the population that is the object of the study but from a sample. This sample will be more or less representative but in no case will it reflect perfectly the reality of the population. If we handle different product lines that are candidates for going on the market, the comparison between them should be carried out through comparing hypotheses in order to check whether the differences observed are significant at statistical level.

One of the checks that should be carried out by default is the comparison of the overall penetration of the best varieties according to TURF with the combination of varieties with the greatest individual penetration. If the overall penetration of the two combinations is statistically equal, we may be interested in opting for the second, even though the TURF suggests a different combination of varieties.

In our example, the combination of the four varieties with the greatest penetration (14, 5, 6 and 7) obtains an overall penetration of 125 individuals out of 150, two individuals fewer than the optimal combination. By applying a significance test for dependent proportions[5] we find that the two penetrations we are handling are statistically equal and

[5] McNemar's test was used to compare proportions in dependent samples. The level of signification obtained for the Chi-squared statistic was 0.5 and therefore the null hypothesis of equality between proportions is clearly accepted.

therefore it may be of interest to sacrifice two individuals of overall penetration to market the four varieties that obtain the greatest individual penetration. Likewise, we can compare any combination that is of special interest for the goals of the personnel responsible for the product.

In the event that the two proportions were significantly different, we would have to opt for one strategy or the other, always bearing in mind that the TURF analysis aims at maximizing the penetration of the line as a whole, sacrificing buyers who duplicate in favour of new buyers.

4. Adaptation of the basic model

As we have already seen, the fact of having found a theoretical model in which to place our problem makes our tool much more flexible when it comes to adapting to the different situations that may arise in real life. In this section, we present some of the most common problems that can be treated with TURF analysis.

4.1. Limited number of varieties

It is very common to have a limit to the number of varieties we wish to put on the market. In this case, the formulation of the problem would be:

What is the maximum penetration we can reach if we want to put a certain number of varieties on the market? Which varieties should I market?

In this case we have two groups of variables, not only the binary variables that determine what varieties will be marketed, but also variables that will let us know the number of buyers we will reach with the final combination. In other words, for each individual considered the solution will tell us whether he or she buys or not. We add these variables to the problem:

$$y_i = \begin{cases} 1, \; customer\ i\ will\ purchase \\ 0, \; customer\ i\ will\ not\ purchase \end{cases}$$

Also in this case the number of varieties is pre-determined so we have to introduce a new constraint to the problem:

$$\sum_{j=1}^{n} x_j = V$$

Where V is the number of varieties we want to put on the market.

In this case, not necessarily all the individuals considered will end up consuming, so we have to reformulate the constraint as follows:

$$\sum_{j \in N_i} x_j \geq y_i \qquad i = 1,\ldots,m$$

Finally, the problem no longer tries to minimize the number of varieties but to maximize the number of consumers, bearing in mind our limitation of varieties. Therefore, our new target function and our new model will be:

$$\max Z = \sum_{i=1}^{m} y_i$$

s.t.

$$\sum_{j \in N_i} x_j \geq y_i \qquad i = 1,\ldots,m$$

$$\sum_{j=1}^{n} x_j = V$$

$$\begin{aligned} y_i &= \{0,1\} \qquad i = 1,\ldots,m \\ x_j &= \{0,1\} \qquad j = 1,\ldots,n \end{aligned}$$

In other words, the target maximizes the number of buyers subject to a limitation on the number of varieties. When the problem decides that a variety goes on the market (x_j = 1) all the individuals who include it in their preferences will be considered buyers (their y_i will become 1) because the problem seeks to maximise the number of buyers. On the contrary, when the problem decides that none of the varieties chosen by an individual goes to the market (all the x_j chosen by the individual are zero) then that individual will not be considered a buyer (we force his y_i = 0)

Therefore, we consider that individuals are equal. However, it may be interest to differentiate between individuals on the basis of certain criteria. An extension of this version of TURF is presented in section 4.4, taking into account the frequency with which each individual buys the product.

4.2. Extension of a product line already being marketed

Frequently, we are not dealing with a totally new product line, but with one that already exists; in other words, we already have some varieties marketed that we do not want to stop selling. If this is the case, the adaptation of the model is very simple, all we have to do is to force the corresponding variables to have a value of one, that is, for each of the varieties we want to include in the final solution we introduce a new constraint:

$$x_j = 1$$

Where j is the variety we want to continue marketing.

4.3. Maximizing the volume

Furthermore, just as we have previously weighed on the basis of the individual penetration of each of the varieties, it may be of interest to us to consider the volume we can obtain with each of the varieties. In goods with a high buying frequency it may be worthwhile to market a variety even though it is not as attractive in comparison with another if it obtains a more frequent statement of intention to purchase.

In this case, we weight the varieties by assigning greater weight to the varieties with more frequent purchases and less weight to the most sporadic. To do so, we weight by the inverse of the buying frequency obtained, so that the varieties with a more frequent consumption will have more possibilities of being put on the market.

The procedure is as follows. Each of the data in the frequency matrix is multiplied by the intention to buy the variety for each individual, so that if the individual has no intention to buy the specific variety, the declared frequency does not count for this variety. Once this has been done, we add up the columns of frequencies for each variety (f_j). All that remains for us to do is obtain the inverse of each of the f_j. These data will be the weightings used to weight the binary variables of our problem so that the varieties showing the greatest frequency will be weighted by a smaller factor that will have more weight on making our target function minimal.

$$Min\ Z = \sum_{j=1}^{n} \frac{1}{f_j} x_j$$

s.t.

$$\sum_{j \in N_i} x_j \geq 1 \qquad i = 1,\ldots,m$$

$$x_j = \{0,1\} \qquad j = 1,\ldots,n$$

4.4. Taking into account the different characteristics of the buyers

If our number of varieties is limited and is lower than the number of varieties offered by the optimal solution to the basic problem, in other words, if we do not manage to attract all the individuals considered, we can discriminate between buyers based on the criterion we are most interested in, for example, the frequency with which they consumer usually buys the product as a whole (if we are designing a new line of ice-creams, we could consider the frequency with which each individual usually buys ice-cream).

We use this information to weigh the interviewees' preferences so that if an individual buys very often, his or her preferences should have more weight than individuals who purchase more sporadically.

Since our number of varieties is limited, this version of TURF is an extension of the adaptation presented in section 4.1., in which we had a variable for each individual considered in the problem, a variable which we will weigh on the basis of the frequency of buying declared by each individual:

$$\max Z = \sum_{i=1}^{m} f_i y_i$$

s.t.

$$\sum_{j \in N_i} x_j \geq y_i \qquad i = 1,\ldots,m$$

$$\sum_{j=1}^{n} x_j = V$$

$$y_i = \{0,1\} \qquad i = 1,\ldots,m$$
$$x_j = \{0,1\} \qquad j = 1,\ldots,n$$

Where f_i is the number of times individual i buys the product in a given period[6].

This formulation for the TURF problem has its equivalent in the area of location analysis. More precisely, the Maximal Covering Location Problem (Church and ReVelle 1974).

4.5. More than one variety per consumer

On other occasions, it can be considered that attracting the consumer with just one variety limits the consumer's capacity to alternate between varieties, very common behaviour especially in food products. Accordingly, what interests us is that each consumer considered feel attracted by at least 2, 3,… varieties. The problem is also usually applied to products that include different varieties in the same products, for example, an air-freshener that combines three different fragrances, a box with a selection of biscuits, etc. In these cases, we can consider that the consumer will only buy the product provided he or she feels attracted by at least a determined number of varieties (greater than one).

The adaptation in this case is also very simple, and consists of changing the constraints relating to consumption as follows:

$$Min \;\; Z = \sum_{j=1}^{n} x_j$$

s.t.

$$\sum_{j \in N_i} x_j \geq P \qquad i = 1,\ldots,m$$

$$x_j = \{0,1\} \quad j = 1,\ldots,n$$

We consider that each individual has to feel attracted by at least a number of varieties greater than 1 (P)[7].

[6] Since we want to maximize the target function, we can multiply directly by the frequency with which each individual declares that he/she buys the product.

4.6. Combined with other variables

The TURF analysis calculated in this way permits the introduction in the model of other variables not included in the study and that have more to do with the production/organization of the business, such as the production cost of each of the varieties, their net profit, or the time necessary to produce them.

5. Solving Turf: Binary programming v. other algorithms

As it is well known, the use of the greedy algorithm does not guarantee that the optimal solution to the problem will be obtained. In order to quantify to what extent this deviation of results occurs, we have compared this method with the method proposed in this study, both in randomly generated samples and in real samples.

In the test with random samples, a total of 100 samples were generated, considering 500 records (individuals) for 15 product options. On the other hand, we applied the same comparison in 68 real samples collected in several studies on different consumer goods. As we show below, the results of the analysis differ according to the type of sample analysed, although we can reach the same conclusion: the advisability of using the exact algorithm.

The test carried out on random samples shows alarming results, as only 6 of every 10 solutions obtained through the greedy algorithm provided the optimal solution.

In the case of the comparison of methods with real samples, the results are less shocking, due to the greater concentration of replies in the more attractive varieties. The analysis showed that of 68 samples analysed, in 9 of them the optimum result was not obtained when the greedy algorithm was used, in other words, 13% of the solutions recommended by that algorithm were erroneous. The error is much smaller compared to that obtained with random samples. However it is still significant because not reaching the optimum result means the erroneous recommendation to market at least one variety more than those that are strictly necessary.

In terms of computing times, the use of binary integer programming (BIP) is extremely efficient. For example, Ennis et al (2011) used the eTurf algorithm to find 12 out

[7] We have to exclude all those who do not feel attracted by at least P varieties because we assume that they will never consider buying my product and therefore we have to screen our database, retaining only those who have declared an intention to buy at least P product alternatives.

of 100 concepts. They report that the problem was solved in less than 24 hours. Several randomly generated Turf problems of similar size were solved using Lingo, a commercial software for linear and integer programming, with the formulation presented in section 4.1. Computing time never exceeded 1 second to find the solution. In the following table, the efficiency in using BIP in terms of computing time (seconds) is presented. M is the total number of interviews, N represents the total number of products, and P the desired number of products.

	N=100			N=500			N=1000		
	P=6	P=12	P=18	P=10	P=20	P=30	P=20	P=50	P=80
M=150	0.5s	0.6s	0.4s	0.9s	0.8s	0.8s	-	-	-
M=500	0.6s	0.6s	0.6s	2.3s	2.4s	2.1s	4.0s	4.1s	4.3s
M=1000	-	-	-	4.6s	4.9s	4.3s	8.7s	9.0s	9.8s

Presenting the analysis using BIP lets us benefit from a series of advantages compared to the algorithms used previously. The main advantages are:

- Compared to exhaustive enumeration: it represents a considerable saving in computation time for large problems. The algorithm that solves the problem no longer has to calculate each and every one of the possible combinations. This fact becomes relevant when the problems presented have a great quantity of variables and/or records. The main software on the market that enables TURF to be obtained does so using the exhaustive enumeration method and therefore presents limitations as to the number of varieties to be introduced in the problem. The implementation of TURF with the proposed model would let this limitation be extended. Likewise, it would also improve the efficiency of such software.
- Compared to the greedy algorithm: it guarantees the optimal solution. In section 4 we compared the two methods. The conclusions of this comparison show the advisability of using an exact algorithm.
- Compared to eTurf: using the binary formulation is much more efficient in terms of computing time. Very large problems can be solved in a matter of seconds.

- Furthermore, the presentation of the problem through BIP means that the adaptations of the model to real situations can be done easily since the model contributes great flexibility to the problem. In section no. 4, we presented some of the situations that habitually arise in real life and the adaptations of the model in order to obtain the goal.
- A large number of commercial LP solvers that are easy to use can be found, such as Lindo, Cplex or Gams. The most commonly used spreadsheets normally include add-ins that let mathematical programming problems to be solved, such as the Solver add--in that comes with Excel
- As mentioned above, the formulations presented in this paper belong to the class of binary linear programs. These correspond to the Location Set Covering Problem (LSCP) and to the Maximal Covering Location Problem (MCLP). These are well known NP-Hard problems that are applied here in the context of TURF analysis. Linear programs are solved using the Simplex algorithm together with the branch and bound algorithm, which is an exact algorithm that finds the optimal solution, regardless of the size of the problem. In other words, the iterative procedure of the Simplex algorithm is exactly the same for small, medium or large problems. The limit of the problem size that can be solved using the Simplex Algorithm is set by the software itself, or by the RAM of the computer. For example, the add-in Solver from Excel allows a very limited number of variables and constraints, while other professional packages such as Lingo may allow unlimited number of variables and constraints. A good description of these algorithms can be found in Winston (2003. As mentioned before, ReVelle (1993) proved that both the LSCP and the MCLP are "integer friendly" linear programs, and in general most of the times there is no need to use the branch and bound algorithm, since the Simplex algorithm finds optimal binary solutions. An excellent review of these problems and the way they are solved can be found in Snyder L. (2011)

6. CONCLUSIONS

This study represents a revision of the popular TURF analysis used in the world of marketing for the definition of product lines. The aim of this analysis is to maximize the penetration of a product line taken as a whole, sacrificing buyers attracted by different product alternatives in favour of new buyers.

Our contribution to the analysis involves considering the problem as a Binary Linear Programming problem. This has enabled us to develop a model that fits the objective of the TURF analysis perfectly and allows us, through mathematical programming and the search algorithms used by it, to reach the solution to the problem in a more efficient and exact manner than with the previously proposed calculation algorithms.

Furthermore, the theoretical framework found lets the problem be adapted simply to multiple situations that may arise in practice. These extensions of the model have shown themselves to be very useful in overcoming the main drawbacks of the analysis.

The proposed working method shows that we must not forget the individual penetration of each of the proposed varieties, despite the fact that the objective of the TURF analysis does not consider this information at any time. As we have seen, it has been shown to be a very useful criterion when it comes to discriminating between different optimal solutions, making the objective of the analysis compatible with other objectives the person responsible for the product may have, such as choosing the varieties that are most attractive individually.

We should never make the final decision considering only the results obtained through the TURF analysis, particularly if we take into account that these are decisions based on information provided by market studies that do not reflect perfectly the reality of the population. However, and although ultimately the decision does not coincide with the recommendation of the analysis, the fact of carrying out the entire process provides us with very valuable information on the behaviour of the different alternatives we are handling, the optimum size of our product line and the number of individuals we manage to attract with each of the combinations being considered.

Finally, the formulations presented are extremely efficient in term of computing times. These formulations are "integer friendly", as defined by ReVelle (1993). This means that in general no branch and bound is needed to find the optimal solution when using the simplex algorithm to solve the problem.

REFERENCES

Adler,T. et al. (2010), Optimizing product portfolios using discrete choice modelling and turf, in S. Hess, A. Daly (Editors): *Choice modelling: The state-of-the-art and the state-of-practice. Proceedings from the inaugural international choice modelling conference*, Emerald Publishing Group Limited (2010), pp. 485–497.

Agresti, A. (2002). *Categorical Data Analysis.* Florida: Wiley Interscience.

Alminana M. & J. Pastor (1997), "An Adaptation of SH heuristic to the Location Set Covering Problem", *European Journal of Operational Research* 100(3), 583-596

Caprara A., Firchetti M. & P. Toth (1999), "A Heuristic Method for the Set Covering Problem", *Operations Research* 47(5), 730-743.

Church R. & C. ReVelle (1974), "The maximal covering location problem". *Papers of the Regional Science Association*, 32, 101–118.

Cohen, E. (1993). TURF analysis. *Quirk's Marketing Research Review*. July/August, 10-13.

Conklin, M., & Lipovetsky, S. (2000). A Winning Tool for CPG. *Marketing Research* Winter 2000, 30-36.

Conklin, M., & Lipovetsky, S. (2005). Marketing decision analysis by TURF and Shapley Value. *International Journal of Information Technology & Decision Making* 4(1), 5-19.

Conklin, M. et al. (2004). Customer satisfaction analysis: Identification of key drivers. *European Journal of Operational Research* 154(3), 819-827.

Current J., Daskin M. & D. Shilling (2001), "Discrete Network Location Models", in Z. Drezner & H. Hamacher (eds.): *Facility Location: Applications and Theory.* Springer-Verlag, 83-118.

Daskin, M. S. and S. H. Owen (1999), "Two New Location Covering Problems: The Partial Covering P-Center Problem and the Partial Set Covering Problem," Geographical Analysis, **31** , pp. 217-235.

Ennis J. et al. (2011). eTurf, a competitive Turf algorithm for large data sets, *Food Quality and Preference* 23(1), 44-48.

Kreiger, A. M., & Green, P. E. (2000). Turf Revisited: Enhancements to Total Unduplicated Reach and Frequency Analysis. *Marketing Research* 12, 30-36.

Lipovetsky S., (2008). SURF - Structural unduplicated reach and frequency: Latent class Turf and Shapley value Analyses, *International Journal of Information Technology & Decision Making* 7(2), 203-216

Markowitz, L. (2004). Appropiate Applications of TURF and Shapley Value for Product Line Optimization. *Ipsos Insight*.

Markowitz, L. (2005). Going Beyond Turf to Complement and Extend Existing Product Lines. *Ipsos Insight*, november.

Miaoulis G. et al. (1990). Turf: A New Planning Approach for Product Line Extensions. *Marketing Research*, march.

Mullet, G. M. (2001). TURF redux. *Quirk's Marketing Research Review*, march.

ReVelle C. (1993), "Facility Siting and Integer-Friendly Programming", *European Journal of Operational Research* 65(2), 147-158

Snyder L. (2011), "Covering Problems". In HA. Eiselt and V. Marianov (eds.): *Foundations of Location Analysis, International Series in Operations & Management Science*, Springer, 109-136

Toregas C. and C. ReVelle (1973), "Binary Logic Solutions to a Class of Location Problems", *Geographical Analysis*, V, No. 2, 145-155

Toregas C. and C. ReVelle (1972), "Optimal Location under Time or Distance Constraints", *Papers of the Regional Science Association*, XXVIII, 101-118

Winston W. (2003), *Operations Research: Applications and algorithms*. Duxbury.