Food Quality and Preference 28 (2013) 382–388

Contents lists available at SciVerse ScienceDirect

Food Quality and Preference

journal homepage: www.elsevier.com/locate/foodqual

Implementing TURF analysis through binary linear programming ☆

Daniel Serra *

Department of Economics and Business, Universitat Pompeu Fabra, Spain

ARTICLE INFO

Article history:
Received 24 October 2011
Received in revised form 1 October 2012
Accepted 2 October 2012
Available online 26 October 2012

Keywords:
Total Unduplicated Reach and Frequency
Market research
Competitive algorithm
Product optimization
Large datasets

ABSTRACT

This paper introduces the approach of using Total Unduplicated Reach and Frequency (TURF) analysis 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.

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 (Cohen, 1993; Conklin & Lipovetsky, 2000). 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 flavors for an ice-cream, colors for an MP3 player or fragrances for an air-freshener.

Miaoulis, Free and Parsons (1990) 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

☆ 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.

* Tel.: +34 935421666; fax: +34 935421746.
E-mail address: daniel.serra@upf.edu

0950-3293/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.foodqual.2012.10.001

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.

Kreiger and Green (2000) 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 and 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), Markowitz (2005) extends the TURF model to different business marketing situations and compares the TURF model with existing others.

Adler and 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 and et al. (2004) and Conklin and Lipovetsky (2005) observe that it is often impossible to distinguish between subsets of different flavor 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 flavors 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 et al., 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.[1]

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 flavors for an ice-cream. We carry out a survey of m individuals as to what flavors they would buy and what flavors they would not. We have a data matrix in which for each individual we know the set of flavors he or she would be willing to buy. We can define the following set:

$$N_i = j/consumer\ i\ chooses\ variety\ j$$

In other words, for every individual it is the set of flavors 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:

[1] 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.