ESTIMATING THE IMPORTANCE OF CONSUMER PURCHASING CRITERIA IN DIGITAL ECOSYSTEMS

Jose Ignacio Peláez
Department of Languages and Computer Sciences
University of Malaga, Spain
jipelaez@uma.es

Francisco E. Cabrera
Department of Languages and Computer Sciences
University of Malaga, Spain
fecabrera@uma.es

Luis G. Vargas
Business Analytics and Operations
The Joseph M. Katz Graduate School of Business
University of Pittsburgh, USA
lgvargas@pitt.edu

ABSTRACT

The purchasing process is not an isolated or unique fact, but a process that contains distinct phases. One of these phases is the evaluation of alternatives, in which the consumer compares the benefits that will be obtained from each brand, product or service, depending on a set of characteristics or criteria. Knowing these criteria and their importance is essential to companies for the supply and development of their products. Commonly this process is developed through multiple criteria decision-making methods, where experts, using surveys and their own experience, ascertain the consumers' purchasing criteria, as well as the importance of them. But this process to determine the importance of the criteria is not simple, as the thought, attitude and learning that direct purchases contain relations that consumers are not able to express explicitly; and secondly because the current information society in which we live causes consumers to receive large volumes of information at all times, causing perception changes regularly. This causes growing discrepancies between the rankings offered by experts and the actual purchase ranking of products and services. In this paper we propose a model to estimate the importance of the criteria and alternatives driving purchases making use of the information expressed by consumers in digital ecosystems. This model takes a set of comparable alternatives defined by the client and valuation criteria defined by experts, and by making use of the assessments that consumers make in digital ecosystems, determines the synergies implicit between the criteria and calculates their weight. Lastly, making use of the Choquet integral, the model aggregates the information and determines

a purchase ranking. The model has been tested in real examples, obtaining satisfactory results.

Keywords: Marketing, Sentiment Analysis, Aggregation Operators, Synergy Criteria. Choquet Integral

1. Introduction

According to Philip Kotler (2016) the purchasing decision process is defined, commonly, in five stages:

- Stage 1. Recognition of a Need. The consumer recognizes a need, and this poses as a problem that must be solved.
- Stage 2. Search of Information. To solve the problem, the consumer looks for information. This search can be of two forms, passive, where consumers limit themselves to just being receptive to the information being provided by different means; or active, where the decision maker actively seeks information and asks various sources.
- Stage 3. Alternatives Valuation. From the information obtained, the consumer assesses the alternatives.
- Stage 4. Purchase Decision. The consumer carries out the purchase according to the assessment made in the previous stage, deciding aspects such as brand, quantity, moment and retailer of the purchased goods. But before the choice is made, it may happen that there are other factors affecting the decision that the consumer is not able to explicitly state in the assessment, despite being considered in the final purchase.
- Stage 5. Post-purchase Behavior. Depends on the satisfaction or dissatisfaction from the product perceived by the consumer once purchased and used.

In this paper we present a model to estimate the importance given by the final consumer to the purchase criteria using information available in digital ecosystems. The model asks the client which alternatives wants to compare, and the experts which criteria should be used to perform these comparisons. It then analyzes the users' communications in social networks and through a voting method for aggregation of social preferences determines the initial importance of the criteria and alternatives. Finally, the model determines the synergies between criteria making use of the initial importance of the alternatives and recalculates the weight of the criteria and the final rank of alternatives

2. Literature Review

Stages 3 (Alternatives Valuation) and 4 (Purchase Decision), are defined within decision making (DM) problems related to products purchases or services contracting and are linked to user preferences (Bernal, et al. 2016; Chang et al. 2016). One of the key activities of companies has been trying to detect the reasons that motivate people to buy a

product or use a specific service to adapt to their customers (Hocevar et al. 2014). To do this, experts analyze user preferences, obtained through personal experiences and surveys, to determine what aspects should be taken into account when improving their products or services. In the decision-making area, these aspects are called criteria and the importance values associated to them reflect the users' preferences (Bana e Cost et al. 1990; Batrinca and Treleaven 2015), 3]. User preferences are used to evaluate alternatives with the purpose of choosing the most appropriate one by using multiple-criteria decision analysis methods (MCDA).

The process of determining criteria and their importance is not simple for two reasons: First, because the thoughts, attitude and learning directing purchases contain relationships that the consumer is not able to explicitly express (Anstead and O'Loughlin 2015; Berger and Iyengar 2103; Verrochi-Coleman et al. 2017); and second, because the current information society constantly provides the consumer with large amounts of information, which causes quick changes in their perception (Villarroel-Ordenes et al. 2017).

3. Hypotheses/Objectives

The main problem these proposals face when seeking information from social media is that they require experts or consumers to make active comparisons between different alternatives. But making comparisons between all alternatives or criteria is not always possible, mainly because individuals hardly ever express their thoughts in intensity values, they merely state the order in which they prefer them (Hu and Liu 2004; Kotler 2016; Vargas 2016). To solve this problem, Vargas (2016) proposed a method that determines the importance of alternatives/criteria making use of a ranking.

To solve the issue of the implicit relations not expressed explicitly by the consumers, Bernal et al (2016) propose a Criteria Coalition Model (CCM). This model calculates a fuzzy measure based on experts' assessment on criteria importance and criteria coalitions, and then aggregates the evaluation alternatives for each criterion using the Choquet Integral. This aggregation method does not assume the independence of criteria and allows the aggregation of non-additive measures; therefore, it is able to represent dependence among criteria in many situations, and thus, consider synergistic effects among criteria (Marichal 2001)

4. Research Design/Methodology

4.1 **Coalition Model Based on Fuzzy Measure and Choquet Integral**

In a Multiple Criteria Decision Making (MCDM) problem, criteria are normally evaluated regarding a target. In general, this evaluation includes all the relevant criteria needed for decision making, but individually; thus, the synergies in the criteria interaction are not considered at the time of the evaluation. To provide greater consistency to the resolution of these problems, Bernal et al. (2016) proposed an interaction of the criteria

using fuzzy measures and the Choquet integral. This approach considers the synergy of the criteria at the time of the information necessary to establish the final rating of alternatives. The formal definition the proposal of Bernal et al. is as follows.

Let us consider $A = \{a_1, \ldots, a_k\}$ a set of alternatives to be evaluated with respect to a set of n criteria $N = \{c_1, \ldots, c_n\}$. Each alternative $a_i \epsilon A$, has a profile $x^a = \left(x_1^a, x_2^a, \ldots, x_n^a\right) \epsilon R^n$, where x_i^a is a partial valuation of a with respect to the criterion c_i. From x^a it is possible to calculate an overall measure $M(\lefteqn{\lefteqn{}}\textit{¿¿}a)$ for each alternative by an aggregation operator $M : R^n \to R$. Consider $P(N)$ the power set of N and also consider $A, B, C, \ldots \epsilon P(N)$ its subsets and $\mu(A), \mu(B), \mu(C), \ldots$ their weights.

Let us also consider:

(1) A set of labels $V = \{v_1, \ldots, v_j\}$; each v_i is used to provide the relative importance of each criterion with respect to the objective being evaluated.

(2) A set of labels $W = \{w_1, \ldots, w_j\}$; each w_i is used to provide the degree of synergy among criteria including the null label.

(3) A set of ratings $EL = \{el_1, \ldots, el_k\}$; each $el_i \epsilon [1, 2, 3]$ that characterizes each expert according to their expertise level.

(4) A set of pairs (el, lq) indicating the number of different linguistic labels of V that an expert can distinguish; the values of lq will be set depending on the expertise level.

(5) a set EC of pairs (e_i, lq) with the characterization profile of each expert e_i.

With these considerations, the calculation of coalitions is determined, firstly, by the linguistic labels of each expert according to their level of expertise; secondly, by valuing the individual criteria (each expert selects the whole label V that is more appropriate for each individual criterion). Then, each expert determines what are the relevant coalitions and evaluates them using the tags in the group W (the rest, i.e., what does not form any coalition is considered to have an additive relationship). It is also necessary to build the fuzzy measure for each coalition. A λ-fuzzy measure is a measure μ such that for all sets $A, B \in P(N), A \cap B = \varnothing$ satisfies

$$\mu(A \cup B) = \mu(A) + \mu(B) + \lambda \mu(A) \mu(B) \tag{1}$$

where $\lambda \in (-1, \infty)$. The value of λ is determined by the information and preferences provided by the experts. Then we proceed to calculate the Choquet integral (*CHI*) for each alternative according to (2).

$$CHI_\mu(x) = \sum_i^n \left[x_{(i)} - x_{(i-1)} \right] \mu\left(A_{(i..n)}\right) \tag{2}$$

where $x_{(i)}$ indicates a permutation such that $x_{(1)} \leq x_{(2)} \leq \ldots \leq x_{(n)}$; $A_{(i..n)} \in P(A)$; and $x_{(0)} = 0$.

Completed the previous steps, it is possible to obtain a valuation for each alternative, which represents the importance of each one as a possible solution to the problem of decision.

4.2 Voting Method to Rank Alternatives

Vargas (2016) proposed a new method to rank alternatives based in the voting concept. Voting is a modified form of ranking in which instead of working with cardinalities we work with the order of preferences of a set of alternatives. Often, the ranking obtained from the preferences of the voters needs to be transformed into intensities or weights from which decisions can be made. The Eigenvector Method for Pairwise Voting (EMPV) allows us to obtain an ordered ranking of alternatives from votes, in the form of a vector, that conveys information on the weight of each alternative.

The formal definition of the proposal of the EMPV is as follows (Vargas 2016):
Let us consider the following:
- Let $A = \{a_1, a_2, \ldots, a_n\}$ be the set of alternatives we want to rank.
- Let $N = \{1, 2, \ldots\}$ the set of voters which will emit their preference orders.
- Let $L(A)$ be the set of n! preference orders.
- A profile on a group of voters $M \subset_{\neq} N$ is a mapping $\phi: M \rightarrow L(A)$.
- Let Φ be the set of all possible profiles. For $\sigma \in L(A)$ and $\phi \in \Phi$, $n_\sigma(\phi)$ is the number of voters in profile ϕ that have the preference order σ.
- Let $(\sigma_1(\phi), \sigma_2(\phi), \ldots, \sigma_H(\phi))$ be the different orderings in profile ϕ. Let $v_{ij}[\sigma_h(\phi)]$ be the number of voters who prefer i to j in the ordering $\sigma_h(\phi)$ of profile ϕ.

When voters state a preference of alternative a_i over alternative a_j they are basically ranking them as $a_i \succ a_j$. So, the ratio a_i/a_j measures how much strongly a_i is preferred over a_j . For a given profile ϕ, $v_{ij}(\phi)$ is the number of voters that prefer alternative i to alternative j so the voting ratio between criteria i and j, $a_{ij}(\phi)$ is defined as:

$$a_{ij}(\phi) \equiv \frac{v_{ij}(\phi)}{v_{ji}(\phi)} \quad \text{with} \quad v_{ji}(\phi) > 0$$

The voting matrix $A(\phi)$ is the matrix would be given by:

$$A(\phi) = \left\{ a_{ij}(\phi) \equiv \frac{v_{ij}(\phi)}{v_{ji}(\phi)} \right\} \tag{3}$$

Since there are m alternatives, the matrix of voting ratios $A(\phi)$ is given by:

$$A(\phi) = \begin{bmatrix} 1 & \dfrac{v_{12}(\phi)}{v_{21}(\phi)} & \cdots & \dfrac{v_{1n}(\phi)}{v_{n1}(\phi)} \\ \dfrac{v_{21}(\phi)}{v_{12}(\phi)} & 1 & \cdots & \dfrac{v_{2n}(\phi)}{v_{n2}(\phi)} \\ & & \ddots & \vdots \\ \dfrac{v_{n1}(\phi)}{v_{1n}(\phi)} & \dfrac{v_{n2}(\phi)}{v_{2n}(\phi)} & \cdots & 1 \end{bmatrix} \qquad (4)$$

Special considerations must be made in this matrix if the voters do not produce all possible orderings of the alternatives. In these situations, there might exist some *i* and *j* where $v_{ji}(\phi) = 0$ when $v_{ij}(\phi) > 0$, which would make impossible to calculate the ratio $a_{ij}(\phi)$. To avoid this situation a fictitious voter could be introduced in the profile which prefers *j* to *i* and has no other preferences. For a sufficiently large number of voters, the addition of an extra voter will make no difference on the priorities of the alternatives.

As shown by Vargas (2016), the principal right eigenvector of $A(\phi)$ is the most accurate representation of the voter's preferences, which is the ordered ranking of alternatives that we are seeking.

5. Data/Model Analysis

The model used to estimate the importance of the criteria from the information generated by consumers in digital ecosystems is presented using The Open Group Architecture Framework (TOGAF) (The Open Group 2011) that provides, from a business approach, business, technology, data and application models.

Figure 1 shows the business model. It consists of three large blocks: objectives and motivation, data sources and interfaces, and lastly the proposed process model.
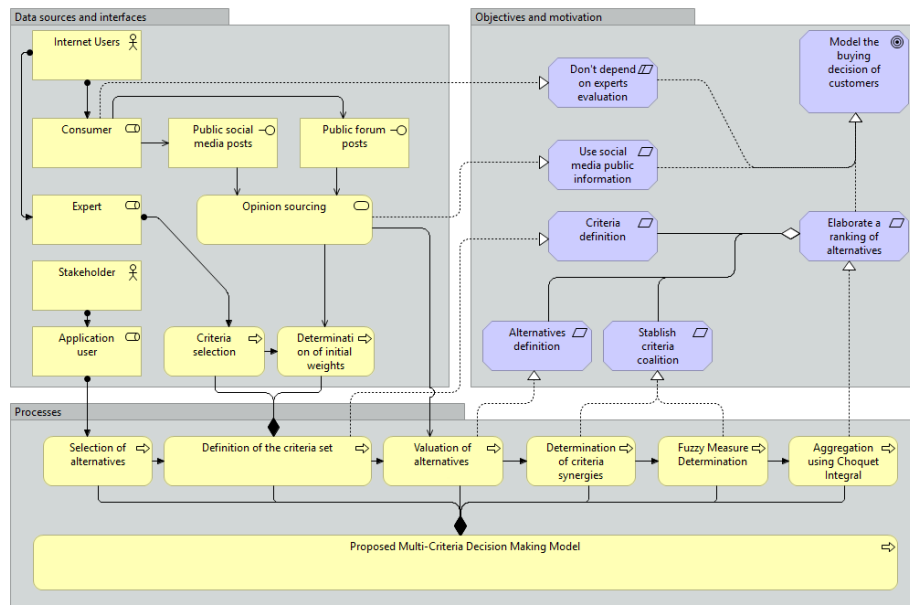
Figure 1. Proposed Multicriteria Decision Making Model.

5.1. Initial Data

The initial data required by the system are: A set $A = \{a_1, a_2, \ldots, a_n\}$ of alternatives that the customer wants to compare to determine the purchase decision ranking of their consumers; a set $C = \{c_1, c_2, \ldots, c_m\}$ of appropriate criteria defined by experts that allow comparison of the alternatives; and a set $CSO = \{cso_1, cso_2, \ldots, cso_k\}$ of opinions/communications published by consumers in digital ecosystems and related to the alternatives.

5.2. Extraction and analysis of data from digital ecosystems

The extraction of information from digital ecosystems can be performed by automated procedures that make use of private or public APIs. The analysis of these communications requires semantic analysis algorithms, which allow the extraction of emotions and the determination of the topic of each communication in relationship to the alternatives (Batrinca and Treleaven 2015; Chan et al. 2016; Hu and Liu 2004). In our case we implemented a system which extracts the relevant communications and performs the analysis. Its architecture is described in Figure 2.

Before any analytic can be performed, the tracking terms must be defined. These are the words or sentence patterns that the watchtower service will consider when fetching communications. When a user of any social media whose API has been integrated in the system emits a statement containing any of the active tracking terms, their communication is stored in the system.
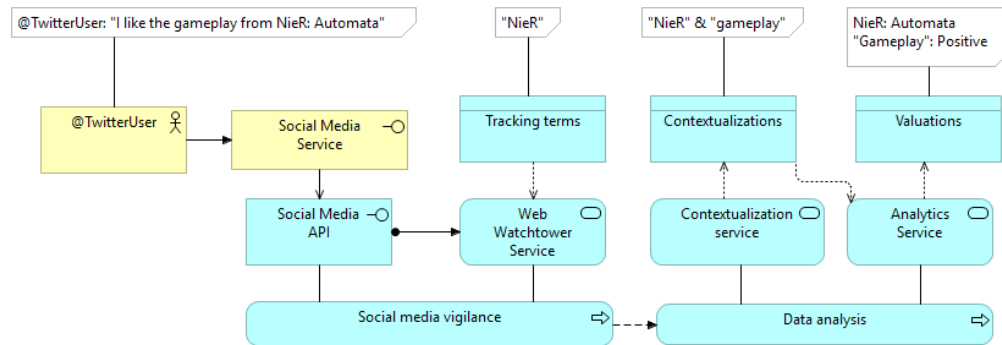
Figure 2. Social media vigilance and analytics system.

The stored communications are then analyzed by the analytics service using the NLTK library from python and a set of logical rules which define the actor to which an analytic must be performed. (i.e. In order to analyze the videogame "NieR: Automata" any occurrence of "NieR: Automata" or its known synonyms "NieR:A", "new NieR", etc. must be considered).

Since only the communications matching the contextualization pattern are analyzed, and they can be modified dynamically, when the obtained communications are not relevant enough more precision can be achieved by imposing tighter constraints in this step. Furthermore, by changing the contextualization rules, the same analytics can be performed in different alternatives or in a subset of their communications (i.e. "NieR: Automata" and "gameplay") which allows valuations in each of the defined criterions.

5.3. Application

The proposed model is illustrated with a real example for the selection of video games. The example illustrates the different processes of the proposed MCDM model. This example is based on sales of video games during the month of May 2017.

Step 1. Selection of Alternatives

After asking a series of clients, they decide that they want to determine the purchase decision of consumers on the following games: A = {Persona 5, Mass Effect: Andromeda, NieR: Automata, The Legend of Zelda: Breath of the Wild, Resident Evil 7, Horizon: Zero Dawn}, which is then the set of alternatives that the client wants to compare in order to determine the purchase rank of consumers.

Step 2. Definition of Criteria Set

The criteria set is compiled by experts after a thorough review of the most employed criteria in specialized game magazines, blogs, webs, forums and other media. The criteria set given by the experts is:

$$C = \{Graphics, Sound/Music, Length/Replay\ Value, Gameplay, Story\}$$

Once the set of criteria has been defined, a data mining process is performed to extract from social media, for each pair of criteria c_i and c_j a set of communications S_{ij} related to them in the context of video games. Then, each communication in S_{ij} is classified by its stated preference between c_i and c_j. Communications not clearly stating preference between criteria are discarded and then, two subsets $S_{i>j} \subseteq S_{ij}$ which contains the communications stating that $c_i > c_j$ and $S_{j>i} \subseteq S_{ij}$ which contains the set communications stating $c_j > c_i$. Once these two sets are assembled the ratio between their cardinalities $|S_{i>j}|/|S_{j>i}|$ is the voting ratio v_{ij} in the voting matrix V.

For example, given the pair of criteria C_1, "graphics" and C_4, "gameplay" the comments shown in Figure 3 are obtained. There are 3 communications stating that $C_1 > C_4$ and 20 stating $C_4 > C_1$ so the ratio in the voting matrix for C_1 and C_4 is $\dfrac{3}{20}$.
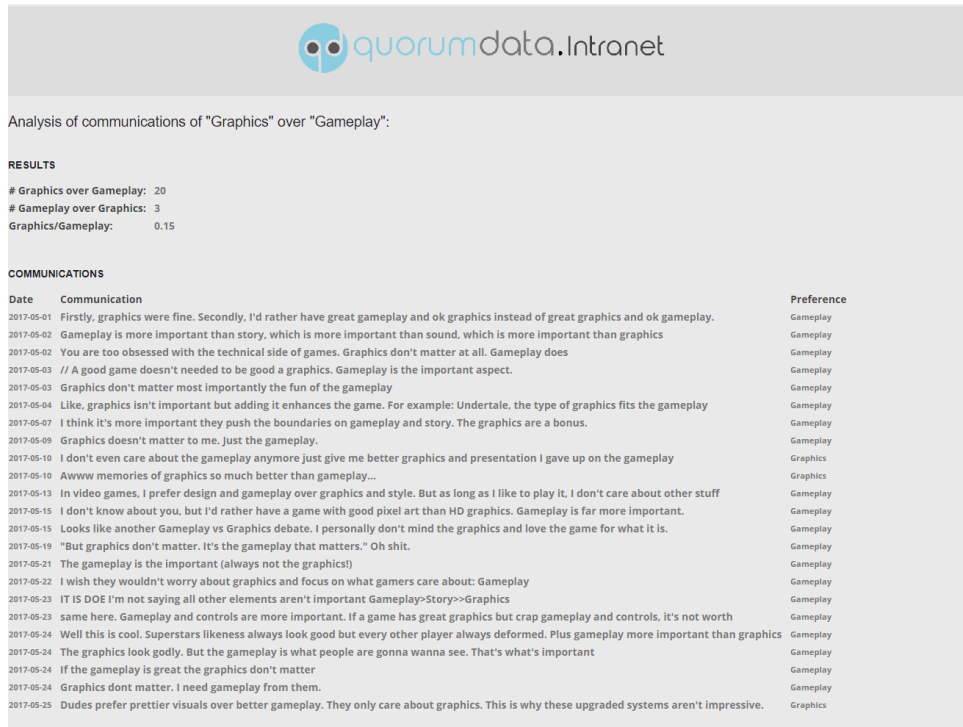

Figure 3. Comparisons found in social media between C_1 and C_4.

The same process is performed with the remaining criteria pairs, to obtain the voting matrix shown in Table 1. The initial weights of the criteria are calculated as the normalized principal right eigenvector of this voting matrix as shown in Table 2.

Table 1. Voting Matrix of Experts for Comparison Criteria.

Voting Matrix	C_1	C_2	C_3	C_4	C_5
C_1	1	0.35	0.36	0.15	0.25
C_2	2.80	1	1.43	0.20	0.51
C_3	2.72	0.69	1	0.17	0.57
C_4	6.50	4.85	5.83	1	2
C_5	3.88	1.93	1.73	0.50	1

Table 2. Initial Weights of Criteria. Normalized Eigenvector of the voting matrix.

C_1 (Graphics)	C_2 (Sound/ Music)	C_3 (Length/ Replay Value)	C_4 (Game play)	C_5 (Story)
0.05	0.12	0.10	0.49	0.21

Step 3. Valuation of Alternatives by Voting Procedure

In this step, we obtain the valuations performed by users in social media. This process is accomplished by means of an algorithm which captures comments in digital ecosystems, a semantic analysis algorithm and a valuation algorithm. The implemented system captures information from social media in which any of the keywords associated with the alternatives and the criteria are present. Then the communications are grouped by user and a semantic analysis algorithm is applied to obtain a valuation in each of the criteria explicitly mentioned in each communication. In Figure 4 we show how the valuation from the first user on the first alternative is obtained. In Table 3 we show some examples of the valuations obtained for the different alternatives.
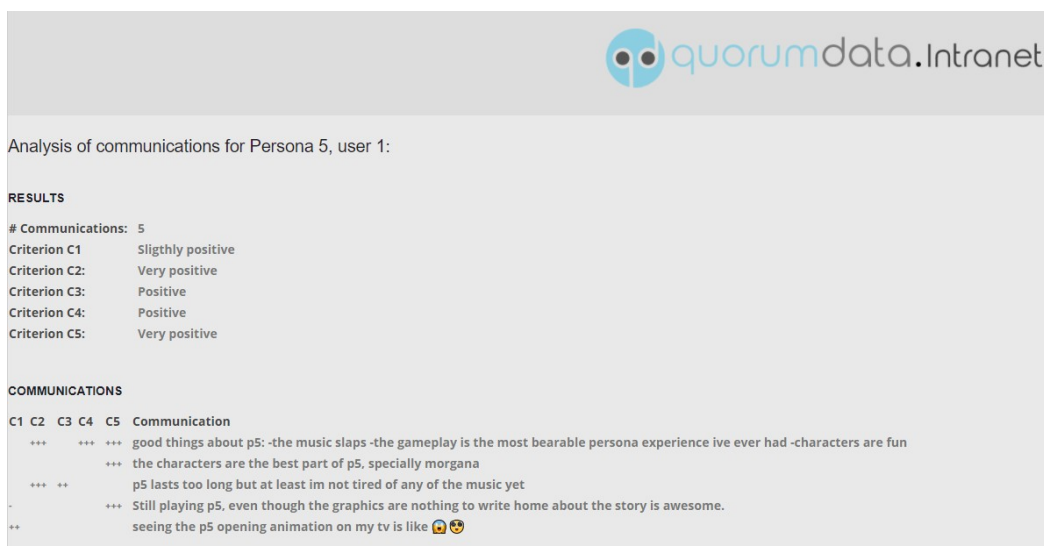

Figure 4. Valuation of Alternative 1 by user 1.

Table 3. Examples of valuations obtained for the alternatives

Game/Users	C1	C2	C3	C4	C5
Criteria					
Persona 5					
1	0.67	1	0.83	0.83	1
2	1	1	0.67	1	1
3	0.67	0.67	0.67	0.5	0.83
4	1	0.67	1	1	0.83
…	…	…	…	…	…
Average	0.57	0.66	0.69	0.67	0.71
Mass Effect: Andromeda					
1	0.83	0.83	0.67	0.83	0.83
2	0.67	0.67	0.83	0.83	1
3	0.67	0.67	0.5	0.33	0.5
4	0.67	0	0.33	0	0
…	…	…	…	…	…
Average	0.63	0.59	0.58	0.48	0.5
NieR: Automata					
1	0.67	0.83	0.83	0.83	1
2	0.5	0.67	0.5	0.5	0.33
3	0.83	0.67	0.83	0.83	0.83
4	0.5	0.83	0.83	1	0.83
…	…	…	…	…	…
Average	0.63	0.7	0.7	0.72	0.63
Zelda: Breath Of The Wild					
1	0.83	0.67	1	1	0.67
2	0.67	0.83	1	1	0.83
3	1	1	0.83	1	0.83
4	0.5	0.83	0.83	0.83	0.5
…	…	…	…	…	…
Average	0.75	0.74	0.84	0.81	0.6
Resident Evil 7					
1	1	1	0.5	0.67	0.67
2	1	0.83	0.67	0.5	0.33
3	0.67	0.5	0.33	0.5	0.33
4	0.67	0.67	0.67	0.67	0.67
…	…	…	…	…	…
Average	0.73	0.67	0.44	0.56	0.52
Horizon: Zero Dawn					
1	1	1	0.67	0.83	0.83
2	0.83	0.67	0.83	0.83	0.83
3	0.67	0.5	0.33	0.5	0.5
4	0.67	0.67	0.67	0.67	0.67
…	…	…	…	…	…
Average	0.85	0.7	0.67	0.69	0.67

Step 4. Determination of Criteria Synergies

To penalize direct monotonic relationships and reinforce inverse monotonic ones, we calculate the Spearman's rank correlation coefficient $\rho_{\alpha,\beta}$ for every pair of criteria $\alpha,\beta \in C$ using the data in Table 3. Let $R(C)$ be the matrix of correlations:

$$R(C)=\begin{bmatrix} 1 & \rho_{1,2} & \cdots & \rho_{1,m} \\ \rho_{2,1} & 1 & \cdots & \rho_{2,m} \\ & & \ddots & \\ \rho_{m,1} & \rho_{m,2} & \cdots & 1 \end{bmatrix} \tag{9}$$

For the application the matrix of sample correlations is given by

Correlation Matrix	C_1	C_2	C_3	C_4	C_5
C_1	1	0.57	0.43	0.52	0.41
C_2	0.57	1	0.59	0.63	0.67
C_3	0.43	0.59	1	0.77	0.64
C_4	0.52	0.63	0.77	1	0.68
C_5	0.41	0.67	0.64	0.68	1

Given a threshold t, for each $|\rho_{\alpha,\beta}|>t$, some degree of redundancy is present, and thus, their synergy must be considered. The synergy sign of this interaction for every pair of criteria α, β is given by:

$$d_{\alpha,\beta}=\frac{-sgn(R(C)_{\alpha\beta}+t)+sgn(R(C)_{\alpha\beta}-t)}{2} \tag{10}$$

The degree of interaction $g_{\alpha,\beta}$ represents the intensity of the synergy and is dependent on the importance given to its underlying criteria. The importance of a criterion α has been previously determined as the αth component of $X_o(C), x_\alpha$, and as such, the degree of interaction between two criteria is defined by the following exponential function:

$$g_{\alpha,\beta}=e^{-4(x_\alpha+x_\beta-1)^2} \tag{11}$$

Considering the threshold t = 0.6, the criteria coalitions to be considered are: C_2 and C_4, C_2 and C_5, C_3 and C_4, C_3 and C_5, C_4 and C_5, all with negative sign given by (10).

For each pair of criteria, c_α, c_β, their coalition factor $S_{\alpha\beta}$ is calculated as the product of the synergy sign and the degree of interaction of the pair given by Spearman's rank correlation. We have

$$S_{\alpha,\beta}=d_{\alpha,\beta} \cdot g_{\alpha,\beta} \tag{12}$$

Step 5. Determination of the Fuzzy Measures

Finally, to capture the synergy between criteria, we use a λ-fuzzy measure μ such as for all $c_\alpha, c_\beta \in C$, $c_\alpha \cap c_\beta = \varnothing$ satisfies:

$$\mu(c_\alpha \cup c_\beta) = \mu(c_\alpha) + \mu(c_\beta) + \lambda \mu(c_\alpha) \mu(c_\beta) \tag{13}$$

Where $\lambda \in (-1, \infty)$. Since λ conveys the information regarding the degree of interaction of the criteria, we compute λ as follows. For a given criterion c_α, let $\mu(c_\alpha) = x_\alpha$ be the αth component of $X_o(C)$. We have

$$\lambda = S_{\alpha, \beta} \lambda_{max} \tag{14}$$

$$\lambda_{min} = \frac{max(\mu(c_\alpha), \mu(c_\beta)) - (\mu(c_\alpha) + \mu(c_\beta))}{\mu(c_\alpha) \mu(c_\beta)} \tag{15}$$

$$\lambda_{max} = -\lambda_{min} \tag{16}$$

Table 4 gives the λ-fuzzy measure μ for every combination of criteria

Table 4. λ-fuzzy measure μ

Combination	Fuzzy Measure
$\mu(c_1)$	0.05
$\mu(c_2)$	0.13
$\mu(c_3)$	0.11
$\mu(c_4)$	0.51
$\mu(c_5)$	0.22
$\mu(c_1, c_2)$	0.19
$\mu(c_1, c_3)$	0.17
$\mu(c_1, c_4)$	0.57
$\mu(c_1, c_5)$	0.28
$\mu(c_2, c_3)$	0.24
$\mu(c_2, c_4)$	0.57
$\mu(c_2, c_5)$	0.33
$\mu(c_3, c_4)$	0.56
$\mu(c_3, c_5)$	0.32
$\mu(c_4, c_5)$	0.58
$\mu(c_1, c_2, c_3)$	0.3
$\mu(c_1, c_2, c_4)$	0.67
$\mu(c_1, c_2, c_5)$	0.4
$\mu(c_1, c_3, c_4)$	0.66
$\mu(c_1, c_3, c_5)$	0.38
$\mu(c_1, c_4, c_5)$	0.74
$\mu(c_2, c_3, c_4)$	0.71
$\mu(c_2, c_3, c_5)$	0.45
$\mu(c_2, c_4, c_5)$	0.78
$\mu(c_3, c_4, c_5)$	0.77
$\mu(c_1, c_2, c_3, c_4)$	0.79
$\mu(c_1, c_2, c_3, c_5)$	0.51
$\mu(c_1, c_2, c_4, c_5)$	0.88
$\mu(c_1, c_3, c_4, c_5)$	0.86
$\mu(c_2, c_3, c_4, c_5)$	0.92

$\mu(c_1, c_2, c_3, c_4, c_5)$	1

For example, $\mu(c_2 \cup c_4) = \mu(c_2) + \mu(c_4) + \lambda\mu(c_2)\mu(c_4)$ where $\mu(c_2)$ and $\mu(c_4)$ are the priorities of the criteria C_2 and C_4, respectively, and λ is obtained using (14). We have $\mu(c_2 \cup c_4) = 0.13 + 0.51 - 1.14 \times 0.13 \times 0.51 = 0.57$. When considering interacting criteria, it is necessary to define these μ coefficients for every combination of criteria. Then, the Choquet Integral can be used to value each alternative [4] as follows:

For the kth alternative, let $(¿¿k1, x_{k2}, \ldots, x_{kn})$ be a permutation of $x_k = ¿$

$z_{kj} = (z_{k1}, z_{k2}, \ldots, z_{km})$ where $x_{k(1)} \leq x_{k(2)} \leq \ldots \leq x_{k(n)}$ and let $x_{k(0)} = 0$. Then, the final results for the kth alternative is given by

$$\left[x_{k(i)} - x_{k(i-1)}\right] \cdot \mu(¿¿k(i..n))$$

$$W(A_k) = \sum_{i=1}^{n} ¿$$

For the first alternative the average valuations for each criterion given in Table 3, are V = {0.5715, 0.6620, 0.6857, 0.6666, 0.7095} which for alternative 1 gives the following order $V(c_1) < V(c_2) < V(c_3) < V(c_4) < V(c_5)$ of criteria, so the ordered vector of weights is $x = \{c_1, c_2, c_4, c_3, c_5\}$. The Choquet integral operation (2) applied on the ordered set of valuations of the alternative is given by

$W(A_1)$ = c₁·μ(c₁,c₂,c₃,c₄,c₅) + (c₂-c₁)·μ(c₂,c₃,c₄,c₅) + (c₄-c₂)·μ(c₃,c₄,c₅) + (c₃-c₄)·μ(c₃,c₅) + (c₅-c₃)·μ(c₅) = (0.5715)×1 + (0.6620-0.5751)×0.9228 + (0.6666-0.6620)×0.7683 + (0.6857-0.6666)×0.3179 + (0.7095-0.6857)×0.2240 = **0.6700.**
Table 5 gives the final ranking of the alternatives.

Table 5. Final ranking

Alternative	Weight	Purchase rank
Persona 5	0.6700	4
Mass Effect: Andromeda	0.5158	6
NieR: Automata	0.6926	3
The Legend of Zelda: Breath of the Wild	0.7527	1
Resident Evil 7	0.5611	5
Horizon: Zero Dawn	0.6935	2

6. Limitations

One of the main problems that exists in the modeling of purchasing decision processes is that the purchasing rankings by the experts, in many cases, differ from the actual ranking. Tables 6 show the actual ranking of the actual purchases and Table 7 shows the one proposed by experts. These tables show that the actual order of purchase does not match the forecast of purchases made by the experts.

Table 6. Actual sales of the video games

Alternative	Units sold	Rank order
Persona 5	67.782	4
Mass Effect: Andromeda	52.611	6
NieR: Automata	86.671	3
The Legend of Zelda: Breath of the Wild	374.045	1
Resident Evil 7	66.076	5
Horizon: Zero Dawn	152.099	2

Table 7. Aggregated ranking of experts of video game reviews. Source: Metacritc

Alternative	Score (Out of 100)	Rank order
Persona 5	93	2
Mass Effect: Andromeda	72	6
NieR: Automata	88	4
The Legend of Zelda: Breath of the Wild	97	1
Resident Evil 7	86	5
Horizon: Zero Dawn	89	3

These differences are motivated mainly because of the implicit relationships between the customers' buying criteria that the experts cannot express when they are asked directly. The model proposed based on finding the weights of criteria and its synergies in buying processes yields the same ranking as the actual sales. The actual sales are can be written as a polynomial function of the proposed model scores given by the Choquet Integral:

Expected Sales = 5042.73-169.794*Proposed Model scores*100 + 1.42825*(Proposed Model scores*100)^2

Parameter	Estimate	Standard Error	T Statistic	P-Value
CONSTANT	5042.73	1509.53	3.3406	0.0444
Proposed Model scores*100	-169.794	48.7459	-3.48326	0.0400
Proposed Model scores*100^2	1.42825	0.387553	3.6853	0.0346

Analysis of Variance

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
Model	68898.8	2	34449.4	14.96	0.0275
Residual	6909.35	3	2303.12		
Total (Corr.)	75808.2	5			

R-squared = 90.8857 percent
R-squared (adjusted for d.f.) = 84.8096 percent
Standard Error of Est. = 47.9908
Mean absolute error = 29.8004
Durbin-Watson statistic = 1.13755 (P=0.2791)
Lag 1 residual autocorrelation = 0.406267
Residuals: Statistically independent Normal (0, Constant Variance).

7. Conclusions

The model proposed in this work estimates the importance of buying criteria and alternatives by making use of information expressed by consumers in digital ecosystems. The model takes a set of comparable alternatives defined by the client and valuation criteria provided by experts. By making use of the comments voiced by consumers and experts, a rank of alternatives is produced from which the initial weights of criteria and alternatives are obtained. With these initial alternative weights which have been calculated from the social media consumers/users, the synergies between criteria are determined and the criteria weights are recalculated. The Choquet integral aggregates the information and determines a purchase intention ranking. The model has been applied to different examples as the one shown above, obtaining similar results as those describe here.

In this work, we have shown how to obtain the criteria importance using data from social media, but the criteria were defined by the experts. Finding the criteria using data from social media, would allow the design of systems able to consider the actual criteria used by consumers in the decision process. This is the subject of our future research.

8. Key References

Anstead, N. and O'Loughlin, B. (2015). Social Media Analysis and Public Opinion: The 2010 UK General Election. *J'l of Computed-Mediated Communication*, 20, 204–220.

Bana e Costa, C. A., Brans, J.P. & Mareschal, B. (1990). Readings in Multiple Criteria Decision Aid. *Readings in Multiple Criteria Decision Aid,* (March 2004), 216–252.

Batrinca, B. & Treleaven, P.C. (2015). Social media analytics: a survey of techniques, tools and platforms. *AI & Society,* 30(1), 89-116.

Berger, J. and Iyengar, R. (2013). Communication channels and word of mouth: How the medium shapes the message. *Journal of Consumer Research,* 40(3), 567-579.

Bernal, R., Karanik, M., & Peláez, J.I. (2016). Fuzzy measure identification for criteria coalitions using linguistic information. *Soft Computing, 20*(4), 1315–

1327.

Chan, H. K., Wang, X., Lacka, E., & Zhang, M. (2016). A Mixed-Method Approach to Extracting the Value of Social Media Data. *Production and Operations Management*, *25*(3), 568–583.

Hocevar, K.P., Flanagin, A.J., Metzger, M.J. (2014). Social media self-efficacy and information evaluation online. *Computers in Human Behavior,* 39, 254-262.

Hu, M. and Liu, B. (2004). Mining and summarizing customer reviews. *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 168-177.

Kotler, P. (2016). *Marketing Management*. 15 ed. ISBN-13: 978-0133856460. Pearson.

Marichal, J. L. (2001). An axiomatic approach of the discrete Sugeno integral as a tool to aggregate interacting criteria in a qualitative framework. *IEEE Transactions on Fuzzy Systems*, *9*(1), 164–172.

Verrochi-Coleman, N., Williams, P., Morales, A.C., White, A.E. (2017). Attention, Attitude, and Action: The Effect of Incidental Fear on Choice Deferral. *Journal of Consumer Research,* 44 (2), 283-312.

The Open Group. (2011, December). The Open Group-TOGAF 9.1. ADM Overview Reference Card. Retrieved June 14th, 2013, from: http://www.opengroup.org/public/member/proceedings/q312/togaf_intro_weis man.pdf.

Vargas, L.G. (2016). Voting with intensity of preferences. *International Journal of Information Technology & Decision Making*, 15(4): 839–859.

Villarroel-Ordenes, F., Ludwig, S., de Ruyter, K., Grewal, D., Wetzels, M. (2017). Unveiling What Is Written in the Stars: Analyzing Explicit, Implicit, and Discourse Patterns of Sentiment in Social Media. *Journal of Consumer Research,* 43 (6), 875-894.