2-tuple fuzzy linguistic perceptions and probabilistic awareness-based heuristics for modeling consumer purchase behaviors

Jesús Giráldez-Cru*, Manuel Chica*†, Oscar Cordón* and Francisco Herrera*
*Andalusian Research Institute DaSCI "Data Science and Computational Intelligence"
University of Granada (Spain)
{jgiraldez,manuelchica}@ugr.es {ocordon,herrera}@decsai.ugr.es
†School of Electrical Engineering and Computing
The University of Newcastle (Australia)

***Abstract*—Agent-based modeling (ABM) is a simulation paradigm to model complex systems by defining heterogeneous individual-level behaviors in a bottom-up approach. ABM is typically employed to simulate markets to study consumer decisions and to see how consumers make their purchase decisions. In this work, we present a marketing ABM where consumer perceptions are modeled using 2-tuple fuzzy linguistic variables. These variables represent the opinions the consumers have on the different features of every product, which drive their decisions (e.g., price or quality). In contrast to numerical or crisp values, fuzzy linguistic variables are a realistic representation of these qualitative aspects. In our ABM, agents use a decision-making heuristic to select a product, which is based on those perceptions and a probabilistic utility maximization rule. This process requires a fuzzy aggregation of the perceptions of every product, based on an ordered weighted average (OWA). In addition, consumers can be aware or unaware of each product in the market. In our ABM, we model this information by introducing a brand awareness filter when applying the decision-making heuristic. Thus, consumer agents can only select those products they are aware of. Our experimental results show that our realistic representation of the consumer preferences is more accurate than other existing approaches.**

***Index Terms*—agent-based modeling, marketing, fuzzy linguistic information, fuzzy linguistic decision-making, awareness-based purchase heuristics.**

I. Introduction

The main objective when modeling a market is to understand the rules that govern it in order to test what-if scenarios afterwards. Therefore, it becomes crucial to understand and predict consumer purchases. In classical approaches, these decisions are commonly inferred from global variables in a top-down scheme. The main drawback of this paradigm is the inability to represent heterogeneous consumer behaviors and emergent events. For this reason, the latter classical approaches usually result in inaccurate representations of the market reality [1], [2].

An alternative approach to those classical models consists of studying the complex behavior of the market emerged from a bottom-up aggregation of consumer decisions [2], [3]. To this end, agent-based modeling (ABM) [4], [5] provides a suitable framework. ABM is a descriptive modeling technique (an aggregation of many individual decisions of every agent) which makes the modeler and the marketer better understand the market and its behavior. In most of the cases, modeling individual behaviors is simpler and more accurate than modeling the behavior of the whole system by global top-down rules. ABM has been successfully applied in many other diverse areas such as economics [6], [7], politics [8], trust-based social systems [9], [10], and contract farming [11].

Most of the existing marketing ABM represent consumer opinions using numerical or crisp values [12], [13]. This is an unrealistic representation of this kind of qualitative information. Moreover, consumer opinions are usually defined from consumer tracking data in the form of questions/surveys available at the company, which are commonly answered with linguistic terms. Therefore, handling numerical values requires the additional pre-processing of transforming the linguistic answers of the surveys into crisp data. As a consequence, the design of the model becomes more complex, and it may result in a loss of information.

In this work, we deploy an ABM system for modeling a virtual market with a realistic representation of consumer preferences based on fuzzy linguistic variables [14]–[16]. This fuzzy information allows us to give fuzzy values (e.g., high, low, or medium) to each of the product drivers (e.g., price, quality, or taste) [17]. In particular, we represent these preferences using fuzzy linguistic 2-tuple variables [18], which consist of pairs of a linguistic label and a symbolic translation. As explained in [17], this representation overcomes the drawbacks of other ordinal fuzzy linguistic approaches [19] since they allow us to assess different values to two linguistic variables with the same linguistic label (by having two distinct values in their symbolic translations). Also notice that this representation is significantly more realistic than other existing and standard representations such as numerical or crisp values. Every agent has a perception about all the drivers of every product in the market. Also, we present a fuzzy decision-making heuristic that aggregates those fuzzy linguistic perceptions in order to compute an assessment for every product and select one of them, which simulates consumer purchases in the market.

978-1-7281-6932-3/20/$31.00 ©2020 IEEE

Our consumer decision-making strategy requires to aggregate all consumer preferences for every product. Each aggregation represents the consumer assessment of each product, and the heuristic selects a product by an utility maximization function with a probability proportional to its assessment. Since consumer preferences are represented with fuzzy linguistic variables, their aggregation can be done by the ordered weighted averaging (OWA) operator [20]. This operator gives different weights to every variable in the aggregation. The OWA operator has been used in other fuzzy linguistic models [21], [22].

The ABM system for marketing analysis with fuzzy linguistic modeling presented in [17] properly models several real consumer behavior characteristics but still do not consider complex mechanisms such as word-of-mouth in a social network and other complex interactions between agents. To advance on the design of a realistic marketing ABM, in the current contribution we introduce two important modifications with respect to that model. First, consumer decisions are unlikely fully deterministic. We model this kind of behavior using a probabilistic decision-making heuristic, in contrast to the fully deterministic strategy used in [17]. Second, consumers are always aware of all the existing products in the market in previous studies [17]. In our proposed ABM, we include this awareness information by defining the brands each agent is aware of. Notice that this kind of information is usually available in consumers' tracking data from marketing agencies. Our decision-making heuristic implements a filter of brand awareness, such that consumer agents can only select products they are aware of.

We evaluate our system in a real marketing case study and compare its performance to other traditional representations of consumer perceptions, analyzing their accuracy with respect to real data representing actual sales in this market. Our experimental analysis shows that our model is indeed the most accurate one.

The remaining of this work is organized as follows. In Section II we describe the usual procedure to model consumer perceptions as well as some preliminary concepts on fuzzy linguistic approaches. The structure and components of the marketing ABM considered are presented in Section III. An empirical evaluation to show the benefits of our approach is presented in Section IV. Finally, Section V presents concluding remarks and future works.

II. Background

In this section we review some preliminary concepts for modeling fuzzy linguistic information in the ABM model to handle consumer opinions and preferences.

In the literature, there are some previous works on the study of dynamics in ABM systems and the integration of fuzzy representations to their analysis. [23] analyzed the dynamics of opinions in an ABM system. A fuzzy representation of BDI agent perceptions is introduced in [24] to simulate decision-making processes in environments with imperfect information. The effects of product attributes on consumer decision-making strategies is analyzed in [25]. [26] presented an approach based on fuzzy rules to analyze human behaviors. None of the latter studies either uses a fuzzy linguistic representation or is focused on ABM systems for marketing analysis.

A. Numerical representation of consumer preferences from tracking data

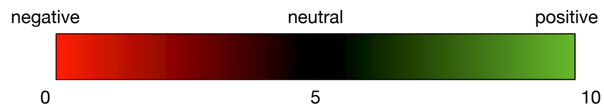

Fig. 1. Representation of consumer perceptions in numerical and linguistic scales.

A marketing ABM usually requires the definition of the consumer perceptions for every brand of the market to simulate a realistic virtual market. These perceptions of the consumers about the existing brands of the market are obtained from real consumers tracking data and brand health studies from well-established marketing consultants such as Kantar Millward-Brown [27]. These data are structured in the form of surveys including sets of answers to a series of questions about the brands [28].

In some cases, the survey answers are directly provided in the same scale required by the marketing ABM, a real value in [0,10] with 0 representing the most negative perception, 5 a neutral perception, and 10 the most positive perception for the consumer (see Figure 1). In those cases, the perceptions definition procedure only requires some grouping/selection of the answers and some statistic computation to obtain the average perception value of each brand for each consumer segment. A segment is a group of consumers who have a similar behavior. For instance, *heavy* consumers of a certain product purchase it regularly (e.g., daily), whereas *light* consumers purchase it occasionally (e.g., monthly). All agents of the same segment are characterized by having similar perceptions, defined from the consumers tracking data. The specific perception values for each agent of a segment in the ABM are randomly generated following a normal distribution with mean equal to the mean of the segment perceptions and small standard deviation. Therefore, using those segments is equivalent to have groups of very similar agents, whose behavior is expected to be resembling.

Nevertheless, the most usual situation is that those answers show a linguistic nature, being either linguistic labels or brand choices. In those cases, a manual pre-processing is required in order to translate those answers to the [0,10] scale, with the consequent *loss of information*. Our view is that the problem would be better tackled by directly working with the linguistic assessments following a fuzzy linguistic approach instead of transforming them into numerical values. Computing with words definitively provides a more natural representation when dealing with human perceptions, represented as words in natural language, as in our case.

B. Linguistic variables

Linguistic variables [14]–[16] are variables whose values are words or sentences in the natural language. They are used in *fuzzy linguistic approaches*, where the problem requires to deal with qualitative aspects [19]. This is a typical requirement in many contexts, where the most realistic and direct representation of the information is indeed the natural language. For instance, the *price* of two products can be easily compared in quantitative terms, but these two *numbers* do not provide any information to assess whether these products are *expensive* or *cheap*, according to a certain consumer. Notice that this is even more relevant when the variables cannot be represented in quantitative terms, e.g., comfort, quality or design.

In the *ordinal fuzzy linguistic approach*, a special case of fuzzy linguistic approaches, linguistic variables take values from a predefined totally ordered set of *linguistic labels* $S = \{s_0, \ldots, s_T\}$ of finite size $|S| = T+1$. We consider the usual definition of ordered set where $\forall s_i, s_j \in S \,.\, s_i \leq s_j \Leftrightarrow i \leq j$.

In this work, we consider triangular membership functions for linguistic variables [29]. In Figure 2, we represent an example of this fuzzy membership. In particular, we represent a triangular membership function for linguistic variables in the interval $[0, 10]$, for the set of linguistic labels $\{veryLow, low, medium, high, veryHigh\}$.

Dealing with fuzzy linguistic variables usually requires to aggregate their information, i.e., the variable values. A common approach is to transform these linguistic values into numbers, aggregate them by common methods, and finally transform the numerical aggregation into a linguistic label. To this purpose, we define two operations to transform linguistic variables into numbers and *vice versa*, as follows:

Definition 1 (Linguistic-Numerical Transformation)**.** *Given an interval $I = [a, b]$, a number $c \in [a, b]$, and a linguistic label s_k of a linguistic variable v, with k being the position of the label s_k in the ordered set of linguistic labels $S = \{s_0 \ldots s_T\}$ from which such a linguistic variable v takes values, we define the following linguistic-numerical transformation functions $\Delta : I \rightarrow S$ and $\Delta' : S \rightarrow I$ as:*

$$\begin{aligned} \Delta'(s_k) &= a + k \cdot (b-a)/T \\ \Delta(c) &= s_k \text{ s.t. } \forall_{s_i \in S \wedge i \neq k} \; |\Delta'(s_k) - c| < |\Delta'(s_i) - c| \vee \\ &\qquad \exists_{s_i \in S \wedge i = k+1} \; |\Delta'(s_k) - c| = |\Delta'(s_i) - c| \end{aligned}$$

Without loss of generality, we use the interval $I = [0, 10]$ in this work.

C. The OWA aggregation operator

The OWA is an aggregation operator that allows us to aggregate linguistic variables, considering they may have a distinct weight in the aggregation. For linguistic variables, it is defined as follows:

Definition 2 (OWA [20])**.** *Given a set of linguistic labels S, let $A = \{a_1, \ldots, a_m\}$ be a set of linguistic variables to be aggregated with $a_1 \ldots a_m \in S$. The OWA operator ϕ on these linguistic labels is defined as:*

$$\phi(A, W) = \Delta(W \cdot \Delta'(A^T))$$

where $W = [w_1, \ldots, w_m]$ is a weights vector such that $w_i \in [0,1]$ and $\sum_i w_i = 1$, and the functions Δ and Δ' are the linguistic-numerical transformation functions defined before.[1]

Without loss of generality, we consider A is an ordered set following the predefined order of assessments used in the weights vector.

D. 2-tuple fuzzy linguistic representation model

A drawback of the previous approach is the loss of information caused by the aggregation of linguistic labels. In particular, the aggregation of two distinct sets of linguistic labels may lead to the same value. As a result, it may be hard to assess whether one of these two sets is preferred to the other.

In order to solve the previous problem, the *2-tuple fuzzy linguistic representation* was proposed in [18]. In this approach, linguistic variables are represented by a linguistic label and a symbolic translation.

Definition 3 (2-tuple fuzzy linguistic variable [18])**.** *A 2-tuple fuzzy linguistic variable is a pair $\langle s, \alpha \rangle$, where $s \in S = \{s_0, \ldots, s_T\}$ is a linguistic label, and $\alpha \in [-t, t]$ is a symbolic translation.*

The semantics of the 2-tuples can be directly derived from their order in the ordered set of linguistic labels S. In particular, the following operators are: (i) equality: $\langle s_i, t \rangle = \langle s_{i+1}, -t \rangle$; (ii) negation: $neg(\langle s_i, \alpha \rangle) = \langle s_{T-i}, -\alpha \rangle$; (iii) maximization: $max(\langle s_i, \alpha_i \rangle, \langle s_j, \alpha_j \rangle) = \langle s_i, \alpha_i \rangle \Leftrightarrow s_i \geq s_j \vee (s_i = s_j \wedge \alpha_i \geq \alpha_j)$; and (iv) minimization: $min(\langle s_i, \alpha_i \rangle, \langle s_j, \alpha_j \rangle) = \langle s_i, \alpha_i \rangle \Leftrightarrow s_i \leq s_j \vee (s_i = s_j \wedge \alpha_i \leq \alpha_j)$.

The OWA operator can be directly applied to 2-tuple fuzzy linguistic variables by just redefining the linguistic-numerical transformation functions Δ and Δ' as follows. Recall that these functions are defined for an interval $[a, b]$ and for a set of linguistic labels $S = \{s_0 \ldots s_T\}$.

$$\begin{aligned} \Delta'(\langle s_k, \alpha \rangle) &= a + k \cdot (b-a)/T + \alpha \\ \Delta(c) &= \langle s_k, \alpha \rangle \text{ s.t. } \Delta'(\langle s_k, 0 \rangle) + \alpha = c \end{aligned}$$

Notice that, the value of t for the interval of symbolic translations is $t = (b-a)/2T$ when considering triangular membership functions (the ones that we use). An example of these transformations can be found in Figure 2, where it can be seen that $\Delta(6.4) = \langle high, -1.1 \rangle$.

III. Agent-based model with fuzzy linguistic information and awareness

In this section we define the ABM that simulates the behavior of a market with fuzzy consumer perceptions and linguistic

[1]For simplicity, we overload these functions Δ and Δ' for a vector of linguistic labels $A = \{a_1, a_2, \ldots, a_n\}$ and a vector of real numbers $B = \{b_1, b_2, \ldots, b_n\}$ as follows: $\Delta(B) = [\Delta(b_i)]_{1 \leq i \leq n}$ and $\Delta'(A) = [\Delta'(a_i)]_{1 \leq i \leq n}$.

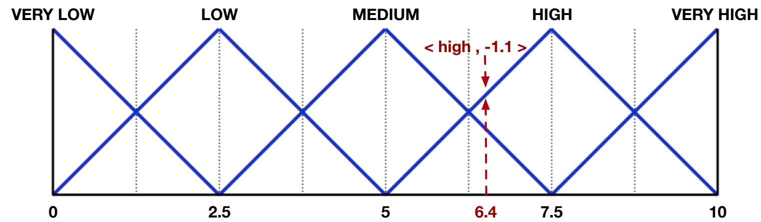

Fig. 2. Triangular membership functions for linguistic variables in the interval $[0, 10]$, for the set of linguistic labels $\{veryLow, low, medium, high, veryHigh\}$. It also includes the 2-tuple numerical-linguistic transformation $\Delta(6.4) = \langle high, -1.1\rangle$.

decision-making. As previously presented, our proposed ABM system uses fuzzy linguistic variables and fuzzy decision-making to characterize and handle brand perceptions [17]. This is the natural way of representing this kind of qualitative information. In particular, we use the fuzzy linguistic variables to represent the different aspects of each brand or product (e.g., *price*, *quality*, *comfort*, ...). These aspects are called *drivers* as they drive consumer choices. We model consumer perceptions on each brand using these drivers. For instance, a consumer can have a *low* perception about the price and a *high* perception about the quality of a certain product in the market.

We use 2-tuple fuzzy linguistic variables for storing the perceptions about the drivers that rule the market. This way, we can aggregate consumer perceptions on each brand without undergoing the problem of loss of information existing in the ordinal fuzzy linguistic approach.

In our model, agents represent consumers who carry out a decision-making process in order to select a product among a set of available brands. This process is performed according to their perceptions and their assessments on each brand. The agents population can be organized in segments, groups of very similar agents in terms of behavior. All of this allows us to simulate the behavior of a market and make predictions on it. In what follows, we precisely define the elements of our marketing ABM based on 2-tuple fuzzy linguistic representation.

A. Brands

In our ABM, consumer decision-making strategies will be only performed among a finite set of available of n brands $B = \{b_1, \ldots, b_n\}$. In order to model the attributes of each brand, we also consider a set of m drivers $D = \{d_1, \ldots, d_m\}$. These drivers are fixed for all the brands in the market as these drivers of importance depend on the product category rather than the brand. Additionally, all the respondents are asked their opinions on the same set of aspects of a product or brand in tracking surveys.

B. Consumer perceptions

Every consumer is represented by an agent of the system. Each agent has its own perceptions (positive, neutral, or negative) about each driver of each brand. In order to represent driver preferences, we define for each agent x a vector of weights $W^x = [w_1^x, \ldots, w_m^x]$, such that all weights must be in the interval $[0, 1]$ and their sum must be equal to 1. These weights represent the importance of each driver when a consumer agent x makes a decision. Notice that these driver weights are independent from each specific brand. It means agents will have the same driver weights for the whole market and its brands.

Consumer perceptions are modeled by defining, for each agent x in the ABM, a matrix of perceptions P^x of dimension $n \times m$, where each element $p_{i,j}^x \in P^x$ represents the perception of agent x on brand $b_i \in B$ about driver $d_j \in D$. In our model, these perceptions are represented using 2-tuple fuzzy linguistic variables, all of them taking values from a common ordered set of linguistic labels (see Definition 3). This allows us to represent the qualitative view of the consumer on each brand.

C. Consumer brand awareness

In a real market, a consumer can be aware or not about the available brands. To model this information, we define in our model the awareness of every agent about every brand. In particular, for every agent x it is defined a vector A^x of n Boolean variables, where $a_i^x \in A^x$ represents whether agent x is aware of brand $b_i \in B$.

D. Consumer agents

Based on the characterization of brands and consumers presented above, we now define consumer agents. Notice that this definition represents the mental state of the consumers, i.e., their knowledge about the market and their perceptions about its products.

Definition 4 (Consumer agent)**.** *A consumer agent x is defined as the tuple $\langle A^x, W^x, P^x\rangle$, where A^x is a vector of n Boolean variables representing the brand awareness of the agent x, W^x is a vector of m weights satisfying that $\forall w_i^x \in W^x \,.\, w_i^x \in [0, 1]$ and $\sum_{1 \leq i \leq m} w_i^x = 1$ and representing the preferences that agent x has on each driver, and P^x is a $n \times m$ matrix of fuzzy linguistic 2-tuples representing the perceptions that agent x has on each pair brand-driver.*

E. Decision-making heuristic

The decision-making process of each agent consists of selecting one of the available brands in the ABM, based on the agent perceptions on its purchase drivers. This decision simulates the purchase decision of a consumer. As a result, the ABM will describe the global behavior of the population of consumers, emerged from the individual decisions of agents.

For each agent, this fuzzy decision-making process can be divided into two steps: (i) aggregation of the assessment for each brand, and (ii) selection of a brand. In the first step, the agent needs to aggregate their perceptions on all the drivers for each brand. This aggregation is computed using the OWA operator for 2-tuples (see the previous section for more details) as follows:

Definition 5 (Brand Assessment). *For a consumer agent x and a given a brand b_i, we define the assessment $as(x, b_i)$ of this agent x on this brand b_i as the aggregation of its perceptions on this brand computed with the OWA operator:*

$$as(x, b_i) = \phi(P_i^x, W^x) = \Delta(W^x \cdot \Delta'((P_i^x)^T))$$

where $P_i^x = [p_{i,1}^x, \ldots, p_{i,m}^x]$ is the i-th row of matrix P^x, and Δ and Δ' are the linguistic-numerical transformation functions for 2-tuples defined in Section II.

Recall that the OWA aggregation of a set of 2-tuples (i.e., the perceptions of each driver) is also a 2-tuple. Therefore, the assessment of an agent on each brand is indeed a 2-tuple fuzzy linguistic variable.

The second and final step is the selection of a brand. This selection represents the brand preferred by each consumer and is based on the assessments of such a consumer on each brand. In this work we use the probabilistic utility maximization function $maxUtil^P$ as the brand selection function in the marketing ABM. This function assigns to each brand a probability proportional to its assessment, and randomly selects one brand using a probabilistic roulette. We use this non-deterministic function inspired by some previous works on marketing analysis and consumer behaviors [30]–[32]. This probabilistic version will help to show and understand the benefits of the proposed ABM with 2-tuple fuzzy linguistic perceptions, as we will see in Section IV.

Definition 6 (Probabilistic Utility Maximization). *For a consumer agent x, the function $maxUtil^P$, which stands for probabilistic utility maximization, is the probabilistic brand selection function that randomly selects a brand using a probabilistic roulette with the following probabilities p_x for each brand $b_i \in B$:*

$$p_x(b_i) = e^{\Delta'(as(x,b_i))} = e^{W^x \cdot \Delta'((P_i^x)^T)}$$

Without loss of generality, these non normalized probabilities can be simply normalized as: $p_x(b_i) = p_x(b_i)/\sum_{b_j \in B} p_x(b_j)$.

IV. Experiments and model evaluation

In this section, we present an empirical evaluation of the marketing ABM with a fuzzy linguistic modeling in a real case study. The motivation of this analysis is to show that our ABM offers a realistic representation of perceptions via 2-tuple fuzzy linguistic variables while it does not suffer any loss of information existing in other approaches. To this purpose, we present a comparison between our ABM and two other models by only differing in the representation of perceptions (the rest of the model remains unaltered).

First, we compare our ABM to a model with a numerical representation of perceptions in the interval $[0, 10]$ (this is the same interval used by 2-tuples in the numerical-linguistic transformation in our ABM with fuzzy linguistic modeling). This is the usual representation marketers must consider to process the available information.

Second, we compare our ABM to a model where perceptions are represented just by linguistic labels. In both cases, we use the same set of labels. Although this second model also uses a realistic representation of perceptions, it is less expressive. As a consequence, this models returns different results. This is because the assessments of each brand may differ even if the decision-making function is the same in both models.

We will also compare the results of these three models to real data representing actual sales and analyze their performance measuring their accuracy in terms of predicted sales of each model (aggregated for all the consumer agents in the ABM).

In the following sub-sections we first present a description of the experimental set-up of our analysis and next we present a large-scale marketing case study from a real marketing analysis provided by Zio Analytics, a Spanish marketing company, in order to show the different results produced by the three models. Recall that the only difference between these three models is the representation of the agent perceptions: (i) 2-tuple fuzzy linguistic variables, (ii) numerical variables, and (iii) fuzzy linguistic variables.

A. Description of the ABM simulation conditions

In order to initialize agent perceptions and driver weights, we use consumers segments. Recall that a segment is a group of consumers who have a *similar* behavior. In our ABM all agents of the same segment are characterized by the same driver weights W^x and similar perceptions P^x, randomly generated following a normal distribution with mean equal to the mean of the segment and small standard deviation.[2] We emphasize that segments do not introduce any change either in the structure and components of the ABM or in the decision-making process, but they only affect the consumer agents population structure and thus the initialization of the agent perceptions.

Since agent perceptions are randomly generated, the results of two simulations (where agents have been distinctly initialized) may differ. In order to reduce the influence of outliers and to obtain meaningful results, we perform a number of distinct Monte Carlo (MC) simulations, differing in the initialization of agents.[3] This is a common procedure when working with ABMs.

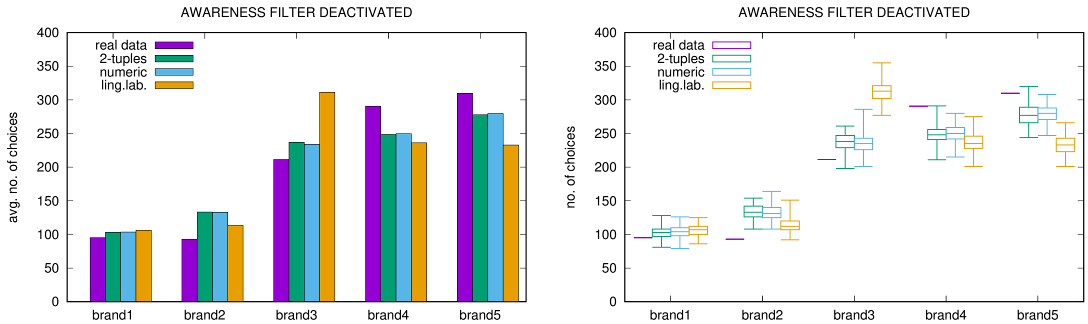

Fig. 3. Comparison of perceptions representations with histograms (left) and box-plots (right), without the brand awareness filter.

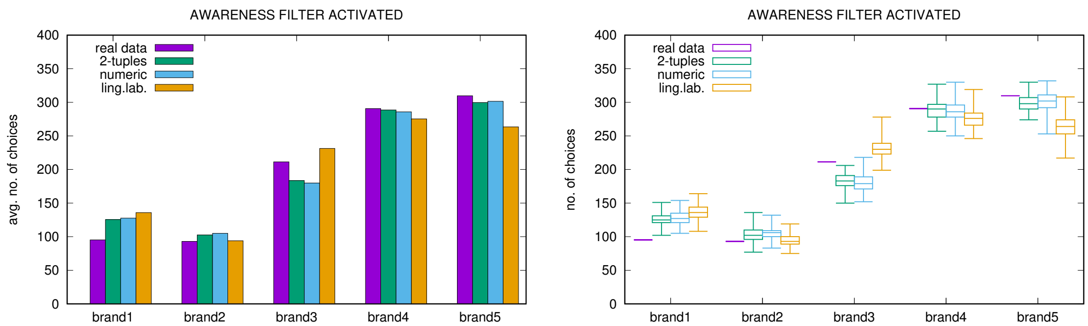

Fig. 4. Comparison of perceptions representations with histograms (left) and box-plots (right), with the brand awareness filter.

B. Validation in a real marketing case study

In this subsection, we present the results of the execution of the marketing ABM based on 2-tuple fuzzy linguistic representation in a real-world marketing case study. The results represent the number of choices of every brand, cumulative for every agent, considering that each agent only performs one decision-making process, and hence only chooses one brand.

The aim of this experiment is to show that our model achieves the main objective: a realistic representation of consumer perceptions without loss of information. In fact, we will later show that the representation of agent perceptions can dramatically affect the output of the ABM (i.e., the number of choices per brand), and therefore any prediction using those results may be inaccurate.

In this case study, we run our ABM with 1000 agents, whose driver weights and perceptions are initialized from existing marketing studies as described before. Recall that results represent the aggregation of 100 MC realizations of the ABM. In the two fuzzy linguistic approaches, we use as linguistic labels the set $S = \{low, medium, high\}$. This case study contains 5 brands and 6 drivers. For anonymity reasons, we have omitted the brand names in this work.

We carry out two distinct experiments, with and without the brand awareness filter in the decision-making heuristic. Recall that when such a filter is activated, a consumer agent is only able to choose a product among the set of available products it is aware of. In order to deactivate the awareness filter, we set the awareness of every brand to 100%, i.e., all agents are aware of every brand.

Figure 3 shows the comparison of these three models with the awareness filter deactivated. In the left column, we represent in histograms the average number of choices of the 100 MC simulations, whereas in the right column we provide box-plots representing the maximum, minimum, median and quartiles 1 and 3 of that number of choices in the same executions. In Figure 4 we represent the same experiment with the awareness filter activated.

We observe that the results in both experiments (with and without brand awareness) differ, for any representation of consumer perceptions. See, for instance, the average number of sales of brand2 using the 2-tuples fuzzy linguistic representation. When brand awareness is not considered (Fig. 3), this number is much greater than the number of actual sales than when this awareness filter is activated (Fig. 4). Therefore, we consider that modeling consumer awareness of brands is a crucial component to understand the market behavior.

Also, it can be seen that our ABM based on the 2-tuple fuzzy linguistic representation has approximately the same number of choices than the model with numerical perceptions. Both models cannot have exactly the same results because a probabilistic roulette is in play, hence a certain variability is expected. Nevertheless, both models return very similar results in general. On the contrary, we observe remarkable differences between the model with 2-tuples and the model with linguistic labels. See, for instance, the differences in the number of sales of brand5 (either with or without brand awareness), where the number of consumers preferring this brand is much smaller in

[2]In our experiments, we use a standard deviation of 1.5.

[3]In our experiments, an execution of the ABM is composed of 100 MC simulations.

the ABM modeling consumer perception with linguistic labels. Similar differences can be also observed for the other brands.

In order to measure the accuracy of each model, we compute their Mean Absolute Error (MAE) with respect to real data (i.e., actual sales). For two sets of observations X and Y of the same size, their mean absolute error is computed as:

$$MAE(X,Y) = \frac{1}{n}\sum_{i=1}^{n}|x_i - y_i|$$

where x_i (resp. y_i) stands for the i-th element of the set X (resp. Y) and n for its size. In our case, the observations X and Y respectively correspond to the number of choices of each brand for actual sales and for each of the three models.

Since the MAE weights the same all errors, we also compute the Root Mean Squared Error (RMSE), which penalizes variance by giving more weight to the errors with larger absolute values [33]. It is computed as:

$$RMSE(X,Y) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - y_i)^2}$$

TABLE I
MEAN ABSOLUTE ERROR (MAE) AND ROOT MEAN SQUARED ERROR (RMSE) OF THE THREE MODELS WITH RESPECT TO REAL DATA, WITH AND WITHOUT AWARENESS FILTER. THE MOST ACCURATE MODEL IS MARKED IN BOLD.

Error estimator	Awareness filter	Representation of consumer perceptions numeric	2-tuples	linguistic labels
MAE	No	2.83	2.95	5.25
	Yes	1.78	**1.60**	2.46
RMSE	No	3.08	3.20	6.23
	Yes	2.14	**1.95**	2.98

MAE and RMSE values for the three models are reported in Table I. It can be seen that the models with numerical and 2-tuples representation of consumer perceptions are much more accurate than the model with linguistic labels. This is a consequence of the less expressive representation of those perceptions used in the ordinal fuzzy linguistic model. We emphasize that this problem does not appear in our ABM based on 2-tuple fuzzy linguistic representation. We can also observe that the performance of every model is much worse if brand awareness is not considered. This suggests that brand awareness is another crucial component in order to accurately simulate a market. In fact, the most accurate model for both error estimators (MAE and RMSE) is indeed the one representing consumer perceptions with 2-tuple fuzzy linguistic variables and taking into account brand awareness, i.e., the model proposed in this work.

V. CONCLUSIONS AND FUTURE WORK

In this work we have presented a fuzzy linguistic representation of consumer preferences and a fuzzy decision-making heuristic that handles them in order to simulate consumer purchases. They are integrated into an ABM for marketing analysis, where agents represent the consumers of the market. Consumer perceptions about the products are modeled using drivers, i.e., the different aspects of every product that drive consumer decisions (e.g., price, comfort, or quality).

Consumer perceptions are usually qualitative aspects. In contrast to numerical or crisp values, fuzzy linguistic variables provide a realistic representation of that kind of information. In our model, we use fuzzy linguistic 2-tuple variables, which do not suffer any loss of information, even when the information has to be aggregated. In reality, consumers can have limited knowledge about the existing brands (i.e., brand awareness). In order to model this mechanism, we define a consumer brand awareness variable for each consumer, meaning whether a consumer knows each brand of the market or not. Consumer agents compute assessments for every product they are aware of based on an aggregation of their perceptions. Those assessments drive their decision-making process, i.e., the process of selecting one of those products. We use a non-deterministic decision-making function based on a probabilistic roulette, where the probability of selecting a certain product is proportional to the assessment of the consumer on that product.

We have presented an empirical evaluation showing the importance of modeling realistically consumer perceptions. We observe than an incorrect representation of consumer perceptions may lead to dramatic differences in the results of the ABM, and as a consequence this may result in inaccurate predictions. In particular, we have analyzed the results of our ABM in a large-scale real marketing case study, showing the remarkable differences of different scenarios only differing in the representation of consumer perceptions (i.e., without altering their values). Additionally, we have studied the importance of including the consumer brand awareness information into the mdoel. Our experimental results show that omitting brand awareness in the model may produce significant differences in the simulation results.

As future work, we plan to extend our marketing ABM based on 2-tuple fuzzy linguistic variables in two directions. On the one hand, we plan to investigate and incorporate other decision-making heuristics in our system [30]–[32], adapting them to handle fuzzy linguistic information. On the other hand, we plan to extend our marketing ABM by incorporating temporal behavior to build a discrete-events simulation [3]–[5]. By doing so, the perceptions of each consumer agent can change overtime and thus, we can analyze how this affects the decision-making output. The extended simulation model could also incorporate media advertising and/or word-of-mouth processes among the consumers of the market through a social network.

ACKNOWLEDGMENTS

This work was supported by the Spanish Ministry of Science, Innovation and Universities, the Andalusian Government, the University of Granada, the Spanish National Agency of Research (AEI), and European Regional Development Funds (ERDF) under grants EXASOCO (PGC2018-101216-B-I00), AIMAR (A-TIC-284-UGR18), and SIMARK (PY18-

4475). J. Giráldez-Cru is also supported through the Juan de la Cierva program (FJCI-2017-32420).

We thank the company ZIO Analytics (`https://www.zio-analytics.com/`) for providing us with the real data used in this work.

REFERENCES

[1] G. Overgoor, M. Chica, W. Rand, and A. Weishampel, "Letting the computers take over: Using AI to solve marketing problems," *California Management Review*, p. 0008125619859318, 2019.

[2] W. Rand and R. T. Rust, "Agent-based modeling in marketing: Guidelines for rigor," *International Journal of Research in Marketing*, vol. 28, no. 3, pp. 181–193, 2011.

[3] M. Chica and W. Rand, "Building agent-based decision support systems for word-of-mouth programs. a freemium application," *Journal of Marketing Research*, vol. 54, pp. 752–767, 2017.

[4] C. M. Macal and M. J. North, "Tutorial on agent-based modeling and simulation," in *Proceedings of the 37th Conference on Winter Simulation*. ACM, 2005, pp. 2–15.

[5] J. M. Epstein, *Generative social science: Studies in agent-based computational modeling*. Princeton University Press, 2006.

[6] L. Hamill and N. Gilbert, *Agent-Based Modelling in Economics*, 1st ed. Wiley Publishing, 2016.

[7] M. Gallegati, *Complex Agent-Based Models*, ser. New Economic Windows. Springer, 2018.

[8] I. Moya, M. Chica, J. L. Sáez-Lozano, and O. Cordón, "An agent-based model for understanding the influence of the 11-M terrorist attacks on the 2004 spanish elections," *Knowledge-Based Systems*, vol. 123, pp. 200–216, 2017.

[9] M. Chica, R. Chiong, M. Kirley, and H. Ishibuchi, "A networked N-player trust game and its evolutionary dynamics," *IEEE Transactions on Evolutionary Computation*, vol. 22, no. 6, pp. 866–878, 2018.

[10] M. Chica, R. Chiong, M. Adam, and T. Teubner, "An evolutionary game model with punishment and protection to promote trust in the sharing economy," *Scientific Reports*, vol. 9, p. 19789, 2019.

[11] H. K. Nguyen, R. Chiong, M. Chica, R. H. Middleton, and D. K. Pham, "Contract farming in the mekong delta's rice supply chain: Insights from an agent-based modeling study." *Journal of Artificial Societies and Social Simulation*, vol. 22, no. 3, 2019.

[12] J. F. Robles, M. Chica, and O. Cordon, "Evolutionary multiobjective optimization to target social network influentials in viral marketing," *Expert Systems with Applications*, vol. 9, no. 113183, 2020.

[13] M. Janssen and W. Jager, "An integrated approach to simulating behavioural processes: A case study of the lock-in of consumption patterns," *Journal of Artificial Societies and Social Simulation*, vol. 2, no. 2, 1999.

[14] L. A. Zadeh, "The concept of a linguistic variable and its application to approximate reasoning - I," *Information Sciences*, vol. 8, no. 3, pp. 199–249, 1975.

[15] ——, "The concept of a linguistic variable and its application to approximate reasoning - II," *Information Sciences*, vol. 8, no. 4, pp. 301–357, 1975.

[16] ——, "The concept of a linguistic variable and its application to approximate reasoning - III," *Information Sciences*, vol. 9, no. 1, pp. 43–80, 1975.

[17] J. Giráldez-Cru, M. Chica, O. Cordón, and F. Herrera, "Modeling agent-based consumers decision-making with 2-tuple fuzzy linguistic perceptions," *International Journal of Intelligent Systems*, vol. 35, pp. 283–299, 2020.

[18] F. Herrera and L. Martínez-López, "A 2-tuple fuzzy linguistic representation model for computing with words," *IEEE Transactions on Fuzzy Systems*, vol. 8, no. 6, pp. 746–752, 2000.

[19] E. Herrera-Viedma, "Modeling the retrieval process for an information retrieval system using an ordinal fuzzy linguistic approach," *Journal of the American Society for Information Science and Technology*, vol. 52, no. 6, pp. 460–475, 2001.

[20] R. R. Yager, "On ordered weighted averaging aggregation operators in multicriteria decisionmaking," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 18, no. 1, pp. 183–190, 1988.

[21] E. Herrera-Viedma, F. J. Cabrerizo, I. J. Pérez, M. J. Cobo, S. Alonso, and F. Herrera, "Applying linguistic OWA operators in consensus models under unbalanced linguistic information," in *Recent Developments in the Ordered Weighted Averaging Operators: Theory and Practice*, ser. Studies in Fuzziness and Soft Computing, R. R. Yager, J. Kacprzyk, and G. Beliakov, Eds. Springer, 2011, vol. 265, pp. 167–186.

[22] A. Valdivia, M. V. Luzón, E. Cambria, and F. Herrera, "Consensus vote models for detecting and filtering neutrality in sentiment analysis," *Information Fusion*, vol. 44, pp. 126 – 135, 2018.

[23] P. Dykstra, W. Jager, C. Elsenbroich, R. Verbrugge, and G. Renardel de Lavalette, "An agent-based dialogical model with fuzzy attitudes," *Journal of Artificial Societies and Social Simulation*, vol. 18, no. 3, p. 3, 2015.

[24] G. P. Farias, G. P. Dimuro, and A. C. da Rocha Costa, "BDI agents with fuzzy perception for simulating decision making in environments with imperfect information," in *Proceedings of The Multi-Agent Logics, Languages, and Organisations Federated Workshops (MALLOW 2010)*, 2010.

[25] S. T. Kim, S. Hong, and C. O. Kim, "Product attribute design using an agent-based simulation of an artificial market," *International Journal of Simulation Modelling*, vol. 13, no. 3, pp. 288–299, 2014.

[26] P. Shi, F. Liu, M. Yang, and Z. Wang, "A fuzzy rules-based approach to analyzing human behavior models," in *Proceedings of the International Conference on Computer Modelling and Simulation (UKSim'11)*, 2009, pp. 346–351.

[27] M. Brown. (2019) Brand tracking studies. [Online]. Available: http://www.millwardbrown.com/mb-global/what-we-do/brand/brand-tracking

[28] A. Charan, *Marketing analytics: a practitioner's guide to marketing analytics and research methods*. World Scientific Publishing Company, 2015.

[29] E. Herrera-Viedma, "An information retrieval model with ordinal linguistic weighted queries based on two weighting elements," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 9, no. Supplement, pp. 77–88, 2001.

[30] J. R. Bettman, M. F. Luce, and J. W. Payne, "Constructive Consumer Choice Processes," *Journal of Consumer Research*, vol. 25, no. 3, pp. 187–217, 1998.

[31] G. Gigerenzer and W. Gaissmaier, "Heuristic decision making," *Annual Review of Psychology*, vol. 62, no. 1, pp. 451–482, 2011.

[32] C. del Campo, S. Pauser, E. Steiner, and R. Vetschera, "Decision making styles and the use of heuristics in decision making," *Journal of Business Economics*, vol. 86, no. 4, pp. 389–412, 2016.

[33] T. Chai and R. R. Draxler, "Root mean square error (RMSE) or mean absolute error (MAE)? - Arguments against avoiding RMSE in the literature," *Geoscientific Model Development*, vol. 7, no. 3, pp. 1247–1250, 2014.