Predictive Knowledge Discovery by Multiobjective Genetic Fuzzy Systems for Estimating Consumer Behavior Models*

Jorge Casillas, Oscar Delgado
Dept. Computer Science and Artif. Intell.
Univ. Granada, 18071 Granada, Spain
casillas@decsai.ugr.es

Francisco J. Martínez-López
Marketing Department
Univ. Granada, 18071 Granada, Spain
fjmlopez@ugr.es

Abstract

The paper introduces a novel problem based on causal modeling in marketing where knowledge discovery is able to provide useful results (as shown in a real-world application). The problem features (with uncertain data and available expert knowledge) and the proposed multiobjective optimization approach makes genetic fuzzy systems to be a good framework to deal with that.

Keywords: knowledge discovery, genetic fuzzy systems, multiobjective optimization, causal modeling in marketing.

1 Introduction

In knowledge discovery in databases (KDD) we can distinguish between two different approaches [6]: *predictive* induction and *descriptive* induction. The difference lies in the main objective pursued and, therefore, the learning method used to attain that. On the one hand, predictive induction looks for generating legible models that describe with the highest reliability the data set that represent the analyzed system. In that case, the goal is to use the obtained model to simulate the system, thus getting an explanation of its complex behavior. On the other hand, descriptive induction looks for particular (interesting) patterns of the data set. In that case, we do not get a global view of the relationships among variables but we discover a set of rules (different enough among them) statistically significant.

This paper focus on predictive induction to extract useful knowledge from causal models used in marketing. Association fuzzy rules (with input and output variables previously fixed) are used. The extraction is performed by means of genetic fuzzy systems. Two questions arise at this stage: *why fuzzy rules?* and *why genetic algorithms (GAs)?*

The use of fuzzy rules (instead interval rules, decision trees, etc.) is justified mainly due to the kind of data set we are dealing (see Section 3.1). In our case, each variable is composed of a set of parameters (items) that add uncertainty to the data, since each of them provide partial information to describe the variable. Moreover, we are able to easily transform the available expert knowledge to linguistic semantics. Finally, the obtained fuzzy models can be linguistically interpreted, an important issue in KDD.

Regarding the use of GAs to derive these fuzzy models instead other well-known machine learning techniques, its application is justified by the following points. Firstly, since there are contradictory objectives to be optimized (such as accuracy and interpretability), we perform multiobjective optimization. It is one of the most promising issue and one of the main distinguishing marks of GAs compared to other techniques. Furthermore, we consider a flexible representation of fuzzy rules that can be properly developed by GAs.

The paper is organized as follows. Section 2 briefly describes the dealt problem based on consumer behavior models. Section 3 introduces the different KDD steps of the proposed methodology. Section 4 shows some obtained experimental results. Finally, Section 5 concludes.

2 Causal Modeling in Marketing

Marketing academics and practitioners have emphasized the need for knowing and explaining the consumer's behavior patters in a manner increas-

*Work supported in part by the Spanish Ministry of Science and Technology under Project TIC2002-04036-C05-01 and by the European Regional Development Fund

ingly efficient. This is mainly due to firms focused on final markets are immersed in highly competitive systems in which it is needed that their decision processes to be as correct as possible. In this sense, models of consumer behavior are considered as a specific case of *marketing management support system*, and throughout the time have demonstrated to be a source of transcendental relevance for the development of marketing science [8].

Notwithstanding, current models of consumer behavior do not seem to cover all the necessities that it should supposedly satisfy a model which aims to aid on the marketing decision making. Thus, as the main problem that actually face firms oriented to consumer markets is not the availability of information (data), but the possession of appropriate levels of knowledge to take the right decisions, the use of avant-garde knowledge discovery techniques able to exploit it may represent an essential source of competitive advantage.

In this respect, we focus our paper on the modeling estimation techniques and its improvement, presenting the potentials that methods of estimation based on fuzzy rules present in order to obtain estimated models of behavior more exhaustive, complex, flexible, interactive and which offer much more quantity of qualitative information than preceding estimation techniques used in this field [3, 8].

3 A Knowledge Discovery Method for Consumer Behavior Modeling with Fuzzy Rules

This section introduces the proposed KDD method to estimate consume behavior. Basically, it consists on preparing the data and on fixing the scheme we follow to represent the knowledge existing in the data. Once defined these aspects, a machine learning method based on GAs is proposed to automatically extract fuzzy models.

3.1 Data Gathering

First step is to collect the data related to the variables defining the theoretical model of consumer behavior proposed. In this sense, as it has been traditionally done in marketing, data are obtained by means of a questionnaire in a similar way to the models estimated by structural equation modeling. Thus, at first, attention should be paid to how consumer behavior modelers face and develop the measurement process of variables which complex behavioral models contain.

It can be said that measuring streams for these latent variables in consumer modeling can be classified into two groups depending on if they defend that these constructs can or cannot be perfectly measured by means of observed variables (indicators)—i.e., the existence or not of a one-to-one correspondence between a construct and its measurement. Certainly, though consumer behavior modelers tended to make use in the beginning of what was known as the *operational definition philosophy*, a more convenient and reasonable position is that ulterior based on the *partial interpretation philosophy* which distinguished between unobserved (constructs) and observed (indicators) variables. This later approach of measurement, being currently predominant in the marketing modeling discipline, poses to jointly consider multiple indicators—imperfect when considered individually, though reliable when considered altogether—of the subjacent construct to obtain valid measures.

For instance, we can consider a simple measurement model depicted in Figure 1, compounded by three construct or latent variables, two exogenous and one endogenous, where: (1) *interaction speed*: the consumer's perception about the Internet's capacity in general, and, more particularly, of different web-sites, to give a response when required; (2) *invasion of privacy*: the consumer's opinion regarding the invasion of his intimacy by the various agents with which he interacts in Internet applications; and (3) *attitude towards the Internet*: the consumer's overall about this communications medium. Likewise, with respect to the measurement scales, imagine, in one hand, that the first and second construct have been measured by means of several nine-points Likert scales ranging from 1: strongly disagree to 9: strongly agree. On the other, differential semantic scales with 9 points have been used for the third. Specifically, in Table 1 we show a hypothetical example of the set of items that could have been used for measuring each one, while Table 2 shows an example of data available for this problem.

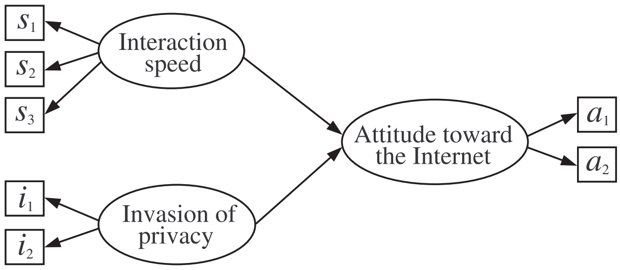

Figure 1: Example of a simple measurement (structural) model

Table 1: Example of a questionnaire associated to the measurement model shown in Figure 1

Interaction speed	
s_1:	Interaction with web pages is fast and stimulating
s_2:	The internet is quick
s_3:	Web pages that I usually visit download quickly enough
Invasion of privacy	
i_1:	When I surf the Internet, I feel my privacy has been invaded
i_2:	Online firms do not respect the visitor's intimacy
Attitude toward the Internet	
a_1:	*Negative* 1 2 3 4 5 6 7 8 9 *Positive*
a_2:	*Unfavourable* 1 2 3 4 5 6 7 8 9 *Favourable*

3.2 Data Processing

Secondly, it is necessary to adapt the collected data to a scheme easily tractable by fuzzy rule learning methods. Therefore, our methodological approach should be aware of the special features of the available data (with several items or indicators to describe a specific variable) when adapting the observed variables to a fuzzy rule learning method. An intuitive approach could directly reduce the items of a specific variable to a single value (e.g., by arithmetic mean). Another possibility would be to expand any multi-item example (the result of a questionnaire filled by a consumer) to several single-item examples and subsequently

Table 2: Example of four consumers' responses about the items shown in Table 1

Speed of interaction			*Invasion of privacy*		*Attitude Internet*	
s_1	s_2	s_3	i_1	i_2	a_1	a_2
2	3	2	6	7	2	2
6	6	7	3	2	8	7
8	8	9	2	3	9	9
5	5	5	4	4	4	4

reduce the data size with some instance selection process.

The problem of these approaches is that the data are transformed, so relevant information may be lost or strained. We propose a more sophisticated process that allows working with the original format without any pre-processing stage: the *multi-item fuzzification*. Thus, a T-conorm operator (e.g., maximum), traditionally used in fuzzy logic to develop the union of fuzzy sets, is applied to aggregate the partial information given by each item during the inference process. Since it is not a preprocessing data but a component of the machine learning design, the details of that treatment of the items is described in Section 3.4.2.

3.3 Fuzzy Model Structure from Expert Knowledge

Several issues should be tackled at this step: the set of variables to be modeled, the transformation of marketing scales used for measuring such variables into fuzzy semantic, and the fuzzy rule structure (relations among constructs). As mentioned, the expert is able to provide its knowledge about the problem by a measurement structural model like shown in Figure 1 (of course, a real problem would work with a more complex model). From this information, we can deduce the variables and the direction (in terms of antecedents and consequents) of the relationships existing among them. Therefore, we can easily fix the input and output variable of the analyzed relationship. For example, from the measurement model of Figure 1, the fuzzy rule structure have the following form:

IF *InteractionSpeed* is A_1 and *InvasionPrivacy* is A_2 **THEN** *AttitudeInternet* is B.

With respect to the fuzzy semantic used for each variable, it is also possible to fix it according to expert knowledge. Indeed, when she/he build the questionnaire in order to collect data, she/he fixes the kind of scale and precision (number of points) used to measure each variable. From this information it is possible to define a fuzzy semantic. At this point, several marketing scale types can be used for its measurement. With the aim of simplifying the problem, in this paper we focus on interval scale (i.e., Likert differential semantic or rating scale), which is one of the most commonly used in marketing.

We suggest to transform these scales to Ruspini's strong fuzzy semantics with uniform density of the fuzzy membership functions to statistically unbias the significance of every linguistic term. Thus, we define the membership function shapes such as, given the set $S = \{min, \ldots, max\}$ defining an interval variable, they hold the following condition:

$$\sum_{k \in S} \mu_{A_i}(k) = \frac{max - min}{l}, \quad \forall A_i \in \mathcal{A},$$

with l being the number of linguistic terms and $\mathcal{A} = \{A_1, \ldots, A_l\}$ the set of them.

Figure 2 shows an example based on the transformation of a nine-point rating scale (a typical marketing scale used to measure the observed variables related to a construct) into a fuzzy semantic with the three linguistic terms *Low*, *Medium*, and *High*.

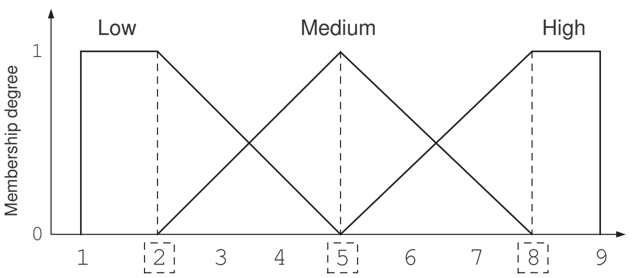

Figure 2: Transformation of a nine-point rating scale into a fuzzy semantic

At this stage, one could think about using some mechanism to automatically generate fuzzy partitions from data [5], or design a tuning method to adapt uniformly initialized fuzzy semantics [1], or both of them. However, in the analyzed problem we are not (more accurately, the expert is not) interested on generating fuzzy semantics that accurately cover data. It is because of the faithfully way to interpret the semantic considered by each consumer that filled the form is the uniform one. If we apply any automatic process to generate/tune fuzzy membership functions, we are adapting to the context, i.e., the answers of the consumer, but not to the meaning of the variables. Therefore, in this problem the KDD process is focused on the relationship among the variables (fuzzy rule surface structures).

3.4 Data Mining Process

Once fixed the linguistic variables that properly represent the tackled information, a machine learning process must be used to automatically extract the knowledge existing in the considered data. This process is, without any doubt, the most important issue from the KDD point of view. As mentioned in Section 1, in this paper we are interested on predictive induction. Of course, since we are performing knowledge discovery, the obtained model should not be only accurate enough but also be easily legible in order to be able to linguistically describe the real system. As it is known, accuracy and interpretability are two contradictory properties. To address that, we consider multiobjective genetic fuzzy system thanks to their good behavior to deal with multiple, contradictory objectives. Next subsections describe the main components of the proposed method.

3.4.1 Fuzzy Rule Structure

In data mining is crucial to use a learning process with a high degree of interpretability. Therefore, we opt by a compact description based on the disjunctive normal form (DNF) [4]. This kind of fuzzy rule structure has the following form:

IF X_1 is $\widetilde{A_1}$ and ... and X_n is $\widetilde{A_n}$ **THEN** Y is B

where each input variable X_i takes as a value a set of linguistic terms $\widetilde{A_i} = \{A_{i1} \vee \ldots \vee A_{il_i}\}$, whose members are joined by a disjunctive (T-conorm) operator, whilst the output variable remains a usual linguistic variable with a single label associated. We use the *bounded sum* ($min\{1, a+b\}$) as T-conorm. The structure is a natural support to allow the absence of some input variables in each rule (simply making $\widetilde{A}_i$ to be the whole set of linguistic terms available).

3.4.2 Multi-item Fuzzification

In order to properly consider the set of items available for each input/output variable (as discussed in Section 3.2), we propose an extension of the membership degree computation, the called multi-item fuzzification. The process is based on a union of the partial information provided by each item. Given X_i and Y_j measured by the vectors of items $\vec{x}_i = (x_1^{(i)}, \ldots, x_{h_i}^{(i)}, \ldots, x_{p_i}^{(i)})$ and $\vec{y}_j = (y_1^{(j)}, \ldots, y_{t_j}^{(j)}, \ldots, y_{q_j}^{(j)})$, respectively, the fuzzy propositions "X_i is $\widetilde{A}_i$" and "Y_j is B_j" are respectively interpreted as follows:

$$\mu_{A_i}(\vec{x}_i) = \min\left\{1, \bigvee_{h_i=1}^{p_i} \sum_{A \in \widetilde{A}_i} \mu_A(x_{h_i}^{(i)})\right\},$$

$$\mu_{B_j}(\vec{y}_j) = \bigvee_{h_j=1}^{p_j} \mu_{B_j}(y_{h_j}^{(j)}),$$

with $\vee$ being a T-conorm (the *maximum* in this paper).

3.4.3 Coding scheme

Each individual of the population represents a set of fuzzy rules (i.e., Pittsburgh style). Each chromosome consists of the concatenation of a number of rules. The number of rules is not fixed *a priori* so, the chromosome size is variable-length. Each rule (part of the chromosome) is encoded by a binary string for the antecedent part and an integer coding scheme for the consequent part. The antecedent part has a size equal to the sum of the number of linguistic terms used in each input variable. The allele '1' means that the corresponding linguistic term is used in the corresponding variable. The consequent part has a size equal to the number of output variables. In that part, each gene contains the index of the linguistic term used for the corresponding output variable.

For example, assuming we have three linguistic terms (S, M, and L) for each input/output variable, the fuzzy rule [IF X_1 is S and X_2 is {M or L} THEN Y is M] is encoded as [100|011||2]. Therefore, a chromosome would be the concatenation of a number of these fuzzy rules, e.g., [100|011||2 010|111||1 001|101||3] for a set of three rules.

3.4.4 Objective Functions

We consider two objective functions to assess the quality of the generated fuzzy systems, the former (approximation error) to improve the accuracy and the latter (linguistic complexity) to improve the interpretability.

Approximation Error: Since the output variable is a composition of several items (see Table 2 and Section 3.2), we have adapted the root mean square error (RMSE) computation to consider that. As mentioned before, the aggregation of items is made by the union. So, let us suppose that the output variable is composed by two items and the prediction had a success degree of S_1 for the first item and S_2 for the second. The total success degree would be $S_1 \vee S_2$. From De Morgan's laws, $S_1 \vee S_2 = \overline{\overline{S_1} \wedge \overline{S_2}}$, i.e., $\overline{E_1 \wedge E_2}$—with E_1 and E_2 being the errors (complement of the success degrees) done over the corresponding items. If the objective of the algorithm is to maximize $S_1 \vee S_2$, the complement will be to minimize $E_1 \wedge E_2$. Therefore, the objective function (for minimization) in a MISO (multiple-input, single-output) system is as follows:

$$F_1(S) = \sqrt{\frac{1}{N}\sum_{e=1}^{N} \min_{t=1}^{q} \left(\mathcal{F}(\mathbf{x}^{(e)}) - y_t^{(e)}\right)}$$

with $(\mathbf{x}^{(e)}; \vec{y}^{(e)})$ being the eth example, $\mathbf{x}^{(e)} = (\vec{x}_1^{(e)}, \ldots, \vec{x}_n^{(e)})$ the input item vectors, and $\vec{y}^{(e)} = (y_1^{(e)}, \ldots, y_q^{(e)})$ the output item vector. Notice that $\mathcal{F}(\mathbf{x}^{(e)})$ performs the fuzzy inference using the multi-item fuzzification described in Section 3.4.2.

Linguistic Complexity: This second objective intends to assess the linguistic complexity of the generated fuzzy rule set. Firstly, it is clear that higher number of rules, higher complexity. Therefore, we measure the number of rules of the fuzzy system $\mathcal{F}$ as $C_1(\mathcal{F})$. However, since each DNF-type fuzzy rule has also a complexity degree itself, we should also consider this aspect. Then, let $C_2(\mathcal{F}) = \sum_{R_r \in \mathcal{F}} \prod_{i=1}^{n} l_{ri}$ be the complexity of the fuzzy system $\mathcal{F}$, with l_{ri} being the number of linguistic terms used in the ith input variable of the rth DNF-type fuzzy rule. The total number of available linguistic terms is computed when an input variable is not considered (i.e. "don't care").

Therefore, the joint objective (to be minimized) is the combination of both complexities as follows:

$$F_2(S) = C_1(\mathcal{F}) \cdot C_2(\mathcal{F})$$

We opt for this combined measurement instead considering two independent objectives because both are deeply related and assess the same concept (complexity of the system).

3.4.5 Evolutionary Scheme

A generational approach with the multiobjective NSGA-II replacement strategy [2] is considered. Binary tournament selection based on the crowding distance in the objective function space is used.

3.4.6 Genetic Operators

The *crossover* operator randomly chooses a cross point between two fuzzy rules at each chromosome and exchanges the right string of them. Therefore, the crossover only exchanges complete rules, but it does not create new ones since it respects rule boundaries on chromosomes representing the individual rule base. In the case that inconsistent rules appear after crossover, the ones whose antecedent is logically subsumed by the antecedent of a more general rule are removed. Redundant rules are also removed.

The *mutation* operator randomly selects an input or output variable of a specific rule. If an input variable is selected, one of the three following possibilities is applied: *expansion*, which flips to '1' a gene of the selected variable; *contraction*, which flips to '0' a gene of the selected variable; or *shift*, which flips to '0' a gene of the variable and flips to '1' the gene immediately before or after it. The selection of one of these mechanisms is made randomly among the available choices (e.g., contraction can not be applied if only a gene of the selected variable has the allele '1'). If an output variable is selected, the mutation operator simply increases or decreases the integer value. In the same way, specific rules appeared after mutation are subsumed by the most general ones and redundant rules are removed.

4 Experimental Results and Interpretation

The consumer behavior model we have used for the experimentation is based on analyzing the consumer's *flow state* in interactive computer-mediated environments. Data have been obtained from the survey used in [7] to test a conceptual model previously presented by the same authors. To illustrate the flow state concept, it is achieved when the consumer is so deeply involved in the process of navigation on the Web that "nothing else seems to matter" [7]. Training data is composed by 1,154 examples (consumers' responses). Because of paper space limitations, we focus the analysis on a specific relationship among the six relationships with a total of 12 variables available in the data set. The four constructs used as input variables of the system (*interaction speed*, *skill/control*, *chall/arousal*, and *telepress/time distortion*) are considered as primary antecedents of the consumer's flow state. In this sense, it is been hypothesized that they are positively related to this central construct.

We have run 10 times the proposed genetic fuzzy system. The fuzzy semantic shown in Figure 2 is considered for all variables. The resulting average convergence plot and joint Pareto-front are depicted on Figures 3 and 4, respectively.

Likewise, Table 3 shows the most accurate obtained fuzzy rule set from the Pareto optimal set. According to this example fuzzy model, we can observe that the three first constructs (input variables) seems to exert a poor influence over the consumer's flow state, even with no repercussions depending on the case. However, *Telepresence/TimeDistortion* is, with no doubt, the most relevant construct. As it can be easily seen, it plays a key role to specially determine low and high levels of flow state and shows a positive (direct) relationship with it.

Table 3: An obtained fuzzy rule set—*mdm* stands for *medium*. $F_1(\mathcal{F}) = 1.617885$, $F_2(\mathcal{F}) = 272$

InteractionSpeed			*Skill/Control*			*Chall/Arousal*			*Telepress/TimeDistortion*			*Flow*		
low	mdm	high	low	mdm	high	low	mdm	high	low	mdm	high	low	mdm	high
				×	×				×	×		×		
×		×		×		×	×		×		×	×		
	×			×	×					×	×		×	
×		×	×		×						×			×

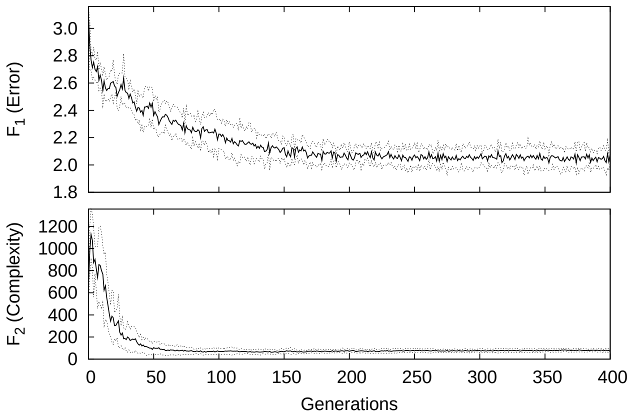

Figure 3: Convergence plot (solid lines) of mean error (top) and mean complexity (bottom) of the Pareto set. Dotted lines show $\bar{x} \pm \sigma$.

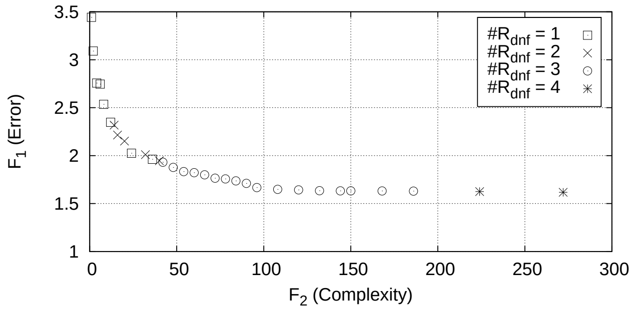

Figure 4: Joint Pareto-front. Each symbol type represents fuzzy models with a specific number of DNF-type fuzzy rules—$C_1(\mathcal{F})$

5 Concluding Remarks

The paper has introduced a novel problem—causal modeling in marketing—where KDD by genetic fuzzy systems can help to generate highly understandable fuzzy models for predictive induction. The problem provides a specific kind of uncertain data set that justifies the use of fuzzy rules. We develop multiobjective optimization to obtain accurate and legible fuzzy models. The proposed KDD methodology has been appropriately applied to a real-world causal modeling problem that analyzes interactive computer-mediated environments.

References

[1] J. Casillas, O. Cordón, M. del Jesus, and F. Herrera. Genetic tuning of fuzzy rule deep structures preserving interpretability and its interaction with fuzzy rule set reduction. *IEEE Trans. Fuzzy Systems*, 13(1):13–29, 2005.

[2] K. Deb, A. Pratap, S. Agarwal, T. Meyarevian, A fast and elitist multiobjective genetic algorithm: NSGA-II, *IEEE Trans. Evolutionary Computation*, 6(2):182–197, 2002.

[3] H. Gatignon, Commentary on P. Leeflang and D. Wittink's "Building models form marketing decisions: past, present and future.", *Int. J. Research in Marketing*, 17:209–214, 2000.

[4] A. González and R. Pérez. Completeness and consistency conditions for learning fuzzy rules. *Fuzzy Sets and Systems*, 96(1):37–51, 1998.

[5] S. Guillaume and B. Charnomordic. Generating an interpretable family of fuzzy partitions from data. *IEEE Trans. Fuzzy Systems*, 12(3):324–335, 2004.

[6] N. Lavrac, B. Cestnik, D. Gamberger, P. Flach, Decision support through subgroup discovery: three case studies and the lessons learned, *Machine Learning*, 57(1-2):115–143, 2004.

[7] Y. Novak, D. Hoffman, Y. Yung, Measuring the customer experience in online environments: a structural modelling approach, *Marketing Science*, 19(1):22–42, 2000.

[8] G. van Bruggen, B. Wierenga, Broadening the perspective on marketing decision models, *Int. J. Research in Marketing*, 17:159–168, 2000.