A Decision-Making Model for Environmental Behavior in Agent-Based Modeling*

Noelia Sánchez-Maroño[1], Amparo Alonso-Betanzos[1], Óscar Fontenla-Romero[1], Miguel Rodríguez-García[1], Gary Polhill[2], and Tony Craig[2]

[1] University of A Coruña, Department of Computer Science, 15071 A Coruña, Spain
{nsanchez,ofontenla,ciamparo}@udc.es, miguel.bhm@gmail.com
[2] The James Hutton Institute, Aberdeen, United Kingdom
{gary.polhill,tony.craig}@hutton.ac.uk

Abstract. Agent-based modeling (ABM) is an increasingly popular technique for modeling organizations or societies. In this paper, a new approach for modeling decision-making for the environmental decisions of agents in an organization modeled using ABM is devised. The decision-making model has been constructed using data obtained by responses of individuals of the organizations to a questionnaire. As the number of responses is small, while the number of variables measured is relatively high, and obtained decision rules should be explicit, decision trees were selected to generate the model after applying different techniques to properly preprocess the data set. The results obtained for an academic organization are presented.

1 Introduction

Nowadays, big companies and organizations require more and more precise models in order to monitor, inference or simulate their realities in a more detailed way. Agent-Based Modeling (ABM) has been proven as an effective tool for this purpose, allowing the direct modeling of those agents (workers, sections, departments...) participating on its daily life, instead of large and hard-to-understand equation models, which are also harder to develop (extra information has to be gathered in order to obtain the needed equations), justify, perform or even explain.

LOw Carbon At Work (LOCAW, http://www.locaw-fp7.com/) is a FP-7 European Union project, in which seven European research institutions participate with the aim of deepening the knowledge of barriers and drivers for healthy lifestyles concerning carbon, through an integrated investigation of daily practices and behaviors on different organizations, so they can achieve the European Union pollution agreements for the next years, and more specifically in 2050[1]. The project includes case studies of six organizations of different types

* This work has been funded in part by the European Commission through Framework Programm 7, grant agreement number 26515, LOCAW:LOw CArbon at Work.

I. Rojas, G. Joya, and J. Cabestany (Eds.): IWANN 2013, Part I, LNCS 7902, pp. 152–160, 2013.

and sizes, to be modeled using an ABM approach, that simulates everyday pro-environmental practices of the different kinds of workers, taking into consideration also these barriers and drivers. ABM has become more and more popular as a tool for modeling on social sciences, since it allows the construction of models where individual entities and their interactions are directly represented. Compared with Variable-Based Modeling (using structural equations) or approximations based on systems (using differential equations), ABM offers the possibility of modeling individual heterogeneity, representing explicitly the decision rules for the agents, and locating them on a geography or other kind of space. It allows the modelers to represent on a natural way multiple analysis scales, the emergency of structures at the macro or social level of the individual action and several kinds of adaptation and learning, which are not easy to achieve with other modeling approximations [2]. The potential of ABM is on the direct representation of each of the actors on a social system, and their behaviors, working on their natural environment. Thus, a model for the behavior of the agents is also needed. In this paper, the model for decision-making in environmental responses of the different types of agents involved in an organization is described. The structure of the model is derived based on two main restrictions: (a) the output of the model, that is, the environmental decision of the agent, needs to be explicit, (b) the model should be based on the reported actual behavior of the different individuals of the organization. This behavior is obtained through the responses to a questionnaire elaborated by the sociologists participating in the project.

2 The General Model

The LOCAW project uses ABM as a synthesis tool for representing everyday practices in the workplace pertaining to the use of energy and materials, management and generation of waste, and transport. Different types of organizations were selected as case studies, specifically, two public sector organizations, two private companies which belong to the energy sector and two private companies of the heavy industry sector. Each organization entails different degrees of autonomy for its workers; therefore, the possibilities for making a decision varies from one to another. For example, people involved on the daily activity of one of the public sector organizations (a university), enjoy considerably more autonomy than do factory workers in the private companies. Therefore, the model should be adjusted to these particularities of each organization, but maintaining a core model that facilitates comparative studies between them and to derive polices or guidelines to achieve a more pro-environmental behavior at the workplace. Bearing this in mind, a general ontology [3] and a general schema were developed. The idea is to simulate the behavior of every worker on the organization, according to the tasks they perform and the options available to implement these daily tasks. For instance, an agent has to move from home to the workplace, but there are choices available, such as going by car, bus, walking, etc. Thus, in order to reproduce the behavior of the agents, the ABM model will follow this schedule:

(1) All (or some) agents make their choices; (2) Environmental impact of those choices is computed; (3)All agents who made a choice adjust their choice algorithm according to the inherent feedback from making that choice (i.e. their own personal enjoyment of it); (4)All (or some) agents forming an in connection to the agents who just made a choice reinforce or inhibit that choice; (5)All agents receiving at least one inhibition or reinforcement adjusts their choice selection algorithm accordingly; (6) Any adjustment to the choice set is made according to scenario conditions.

This paper is focused on the first point of this schema, i.e., the decision-making process of the agents. As mentioned before, this model has several restrictions. The first one is related to its output, as the decision of the agent needs to be explicit in order to check if it is theoretically consistent with the knowledge of the experts (psychologists and sociologists). Besides, a comprehensible output may help to its interpretation by the personnel of the organizations involved. These reasons determine the election of *if-then rules* to explain the decision-making process of the agents. The second limitation stipulates that the model should represent the actual behavior of workers, therefore actual data must be collected using a questionnaire. As there are different size organizations, and responding to the questionnaire will be voluntary, it is not expected to obtain a large number of samples, therefore it restricts the validity of the decision-making algorithms applicable. Giving these reduced data set and the need of deriving rules, decision-trees were selected to generate the decision-making process. A large amount of decision-trees are going to be derived, one for each decision with an environmental impact that the agent (worker) has to take under consideration, for example, going to work walking or using some transport, turning on/off the lights when going for lunch,etc. Thus, it is important to design an automatic procedure that help to derive those decision-trees from data. The different techniques applied are explained in the next section.

3 The Decision-Making Model

The LOCAW project is organized on seven work packages (WP) that pursue different objectives regarding the environmental behavior of individuals in organizations. The psychologists and sociologists in this project have discussed different theoretical models to explain human pro-environmental behavior, finally adopting the model presented in the upper section of Figure 1 where behavior is influenced by values, awareness of consequences, outcome efficacy and norms [4]. Values can be seen as abstract concepts or beliefs concerning a person's goals and serve as guiding standards in his or her life. Schwartz identifies 10 human value types [5], however only four different types were considered important for this project: egoistic, hedonic, altruist and biospheric.

Different quantitative and qualitative tools were used in LOCAW project to analyze the different organizations, for example, focus groups, interviews, life story's, etc. Among them, a questionnaire was designed to obtain data regarding individual factors that affects pro-environmental behavior at work. The questionnaire is based in the value-belief-model (VBM) shown in upper part of Figure 1, and therefore it includes three different blocks with questions about: a) values (some of them depicted at figure 2), b)motivations, i.e., efficacy, worldviews and norms, and c) behaviors. Regarding this last block, notice that not only the behavior at work is important for the aim of this project, but also the behavior at home in order to detect if there exists spillover between them. Therefore, 74 questions for behaviors regarding the use of energy and materials, the treatment of waste and the use of transport in both- work and home- were included.

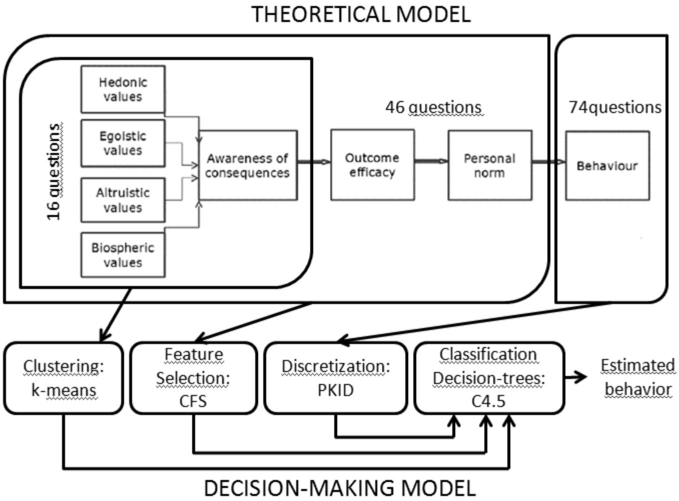

Fig. 1. The decision-making model for the agents in the LOCAW project

The information and data collected are being used to automatically obtain classification decision trees that could explain the agents' pro-environmental decisions when doing a daily task. However, before using decision-trees to determine the possible behaviors of the agents, some techniques have been applied to obtain a representative set of data that maximizes the generalization capability of the resulting decision trees (see Figure 1). The different algorithms are subsequently briefly described, all of them are available in the Weka tool environment [6]. This platform was chosen as it based on Java programming language and the whole project will be developed using this language.

- **Clustering:** Following the VBN model in Fig. 1, the behavior of individuals depends on four different types of values. Therefore, it is expected that workers (and so agents) behave in different ways according to these values, i.e., several profiles can be identified. To generate these profiles and so to represent variability in the model, a clustering technique was used. Since it is an adequate well-known technique, *K-means* [7] was employed in our model.

	Opposed to my values	Not important			Important				Very important
1. EQUALITY: equal opportunity for all	-1	0	1	2	3	4	5	6	7
2. RESPECTING THE EARTH: harmony with other species	-1	0	1	2	3	4	5	6	7
3. SOCIAL POWER: control over others, dominance	-1	0	1	2	3	4	5	6	7
4. PLEASURE: joy, gratification of desires	-1	0	1	2	3	4	5	6	7
5. UNITY WITH NATURE: fitting into nature	-1	0	1	2	3	4	5	6	7
6. A WORLD AT PEACE: free of war and conflict	-1	0	1	2	3	4	5	6	7
7. WEALTH: material possessions, money	-1	0	1	2	3	4	5	6	7
8. AUTHORITY: the right to lead or command	-1	0	1	2	3	4	5	6	7

Fig. 2. A part of the questionnaire related to values

– **Feature Selection:** Adequate identification of relevant features/variables is fundamental in real world scenarios because it may help to obtain simpler models and to focus experts' attention on the relevant data. In this problem, the ratio samples/features is low, because there are 68 questions (6 personal, 16 on values and 46 for motivations) while the number of responses is expected to be in the order of a few hundred (depending on the size of the organization), so the lack of samples prevents obtaining models that properly generalize in spite of the ability of decision trees to discriminate features. Therefore, feature selection (FS) was applied to determine the relevant features while eliminating the irrelevant or redundant ones [8]. From the different FS methods, a filter was chosen because of its independency of any learning algorithm, specifically, the Correlation-Feature-Selection algorithm (CFS) [9] has been applied to the whole set of data.
– **Discretization:** Most questions in the questionnaire use Likert scales, indicating the degree to which respondents agreed with a proposition, or the frequency with which they performed a behaviour (see Figure 2). Again, as the number of responses is not expected to be high, it could happen that not all the ranges could be equally represented in the final sample. To solve this problem a discretization step was considered as necessary, using the Proportional K-Interval Discretization (PKID) algorithm [10]. This algorithm automatically chooses a number of intervals to divide the sample, taking into account the number of samples obtained in each subinterval.
– **Classification:** Finally, once the data has been preprocessed by the previous steps, decision trees can be constructed to automatically derive rules that will lead to a specific behavior for the agents. For this, the C4.5 algorithm was employed [11] as it is one of the most successful methods for this purpose.

4 Experimenting Results

In this section, we will show the results obtained for one of the organizations to be modelled, the University of A Coruña (UDC). The UDC has a total of 2277 workers, between administration (790) and research/teaching (1487) personnel. The questionnaire has been passed down to workers (that could voluntarily answer it) automatically using the Qualtrics application (`https://www.qualtrics.com/`). The answers of the questionnaire have been preprocessed to clean highly-incomplete (more than 45% blanks) or ambiguous data which could contaminate the model. After that, a total amount of 237 different valid samples have been gathered. The different methods presented in the previous section were subsequently applied to this data set and the results obtained are detailed in the following subsections.

4.1 Step 1: Clustering for UDC

The clustering process was carried out using only those 16 questions pertaining to values included in the questionnaire (some of them shown in Figure 2), leading to a data set size of 237 instances $\times$ 16 dimensions. The *k-means* algorithm requires the number of clusters as a parameter, and since four different clusters have been theoretically identified by the experts as adequate for this application study, obtaining four clusters was our first attempt. In general, *k-means* is quite sensitive to how clusters are initially assigned, so different initializations were tested. However, none of the partitions obtained allowed for clearly distinguishing the profiles as indicated by the experts working in the project. Finally, in discussion with the experts, six clusters were identified that drive to adequate separation of the samples and contain hybrid groups. Specifically four "almost-pure" profiles can be identified on clusters zero through three (coinciding with the theoretical ones: egoistic, altruistic, biospheric and hedonic) and two more hybrid groups, that mixed similar profiles (biospheric-altruist and egoistic-hedonic). Columns 3-8 in Table 1 illustrate those clusters, and the parenthesis contain the value of the number of samples they represent. The table details the values of the centroid of each cluster for each value. It can be appreciated that each item is marked with a different symbol (square, triangle, etc.); these shapes are associated with a theoretical profile, so diamond represents biospheric questions, square is used for altruist ones, up-triangle is linked to egoistic items and, finally, down-triangle shows the hedonistic issues. As each dimension value in the centroid represents the mean value for that dimension in the cluster, high values of "up-triangle" dimensions (social power, wealth, authority, influential, ambitious) are expected for the "egoistic" profile, high values of "down-triangle" ones for the hedonic one and so on. Notice that the highest values for each row are in boldface letters. Another important aspect in the clustering with values, is that this section is the only one in the questionnaire that has a column entitled "Opposed to my values" , with a -1 value assigned, than can be checked by individuals answering it. The other sections of the questionnaire have a range between 0 and 7 for the answer. So, not all the ranges in the values part of the questionnaire have the same significance, as only the first column specifies opposing values, while the others specify a continuous range between 0 (Not important) and 7 (Very important). Then, the responses obtained in that column have been weighted with a factor that multiplies by 10 its importance regarding the responses obtained in the other 8 columns. That is the reason why some of the centroid values are negative.

4.2 Steps 2 and 3: Feature Selection and Discretization for UDC

Feature selection allows for determining the relevant features for a giving problem. Actually, this paper copes with 74 problems, one for each election the agent has to consider, i.e., one for each behavior to be modeled. Then, the CFS algorithm has been applied 74 times to determine the relevant inputs (values and motivations) for all the behaviors. Therefore, the final output of this step is a matrix relating behaviors and inputs that has been proven theoretically-consistent by our experts. This matrix shows similarities and differences between behaviors and an extract can be appreciated in Table 2. As explained before, the sample was discretized in order to obtain an adequate representation of the actual intervals obtained in the samples.

Table 1. Clusters obtained for the UDC case. Note that beside the four theoretical clusters initially devised, two more hybrid groups were added.

Attribute	**Full set**	**0**(2)	**1**(62)	**2**(85)	**3**(20)	**4**(56)	**5**(12)
■Equality	6.37	4.5	6.68	**6.71**	6.55	5.91	4.58
♦Respecting earth	5.71	2.50	**6.42**	6.22	6.20	4.73	2.83
▲Social Power	-3.20	1.00	-10.0	**1.35**	-8.50	-1.40	-0.75
▼Pleasure	4.87	2.5	**5.37**	5.27	4.10	4.07	4.83
♦Unity with nature	5.13	1.00	**6.02**	5.91	5.60	3.63	2.08
■A world at peace	6.41	2.50	**6.79**	**6.79**	6.75	6.02	3.67
▲Wealth	1.65	2.50	2.37	2.54	-7.00	2.30	**2.75**
▲Authority	1.24	**4.50**	-0.27	2.49	-1.15	1.46	2.58
■Social justice	6.37	3.50	6.55	**6.73**	6.45	5.95	5.25
▼Enjoying life	5.20	3.5	**5.84**	5.64	3.40	4.52	5.25
♦Protecting environment	5.75	4.00	**6.40**	6.33	6.15	4.75	2.58
▲Influential	2.25	**4.50**	1.73	3.13	1.05	1.84	2.33
■Helpful	5.39	2.00	5.68	**6.01**	5.85	4.30	4.33
♦Preventing pollution	5.55	3.50	6.15	**6.26**	6.05	4.41	2.33
▼Self-indulgent	4.06	2.00	**4.73**	4.59	2.00	3.36	3.92
▲Ambitious	3.31	**4.50**	4.03	3.84	-0.10	2.96	2.92
		Egoistic▲	Hedonic▼	Altruist■	Biospheric♦	Bio-Altruist	Ego-Hedonic

4.3 Step 4: Classification for UDC

For each behavior, the relevant inputs selected by CFS together with the discretized output provided by the previous step form the data set to be fed to C4.5 algorith for training and testing. In all these cases, 66% of the data has

Table 2. An extract of the results of the feature selection process

Behavior	Sex	Studies Lv	Organiz. Lv	Exempl. role	Equality	Resp.earth	Peace
Total Flights	X	X	X	X			X
Turn lights	X	X	X	X	X	X	

been used for training while the remaining 34% is employed for testing. As 6 different clusters where obtained in Step 1 and 74 different behaviors must be modeled, $74 \times 5 = 444$ decision-trees were generated, as cluster 0 (column 3 in Table 1) has only 2 samples and thus it was not automatically treated. An example showing one of the trees derived can be seen on Fig. 3.

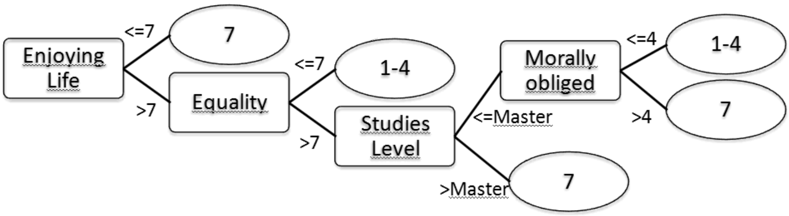

Fig. 3. One example of a tree derived for a behavior related to waste separation, specifically, separating glass from regular garbage at home

5 Conclusions and Future Work

LOCAW project focuses on everyday practices in the workplace and on the interplay of barriers and drivers of sustainable behavior. It will use ABM to study the possible large scale effects of introducing low carbon strategies in the workplace, in different organizations. ABM will include a decision-making algorithm to determine how agents choose between different environmental options in their daily tasks. This paper presents the decision-making algorithm designed based on decision-trees for practical restrictions. This algorithm takes, as input data, the workers' responses to a questionnaire designed by the psychologists in the project. Different methods were employed to make data tractable and, more important, to enhance decision-trees generalization capability. Between the different organizations involved in the project, UDC was selected as starting point because of proximity and familiarity. However, in future stages, this decision-making algorithm has to be adapted to the remaining organizations. Moreover, the decision-making algorithm has to be integrated in the ABM to reflect how the interaction between agents and environment may vary the possible options.

References

1. European Commission: What is EU doing about climate change? `http://ec.europa.eu/clima/policies/brief/eu/index_en.htm` (last visited on January 2013)

2. Gilbert, N.: Agent-based models. SAGE Publications, University of Surrey (2007)
3. Polhill, G., Gotts, N., Sánchez-Maroño, N., Pignotti, E., Fontenla-Romero, O., Rodríguez-García, M., Alonso-Betanzos, A., Edwards, P., Craig, T.: An ontology-based design for modelling case studies of everyday proenvironmental behaviour in the workplace. In: Proc. of International Congress on Environmental Modelling and Software Managing Resources of a Limited Planet, Leipzig, Germany (2012)
4. Steg, L., De Groot, J.I.: Environmental values. In: The Oxford Handbook of Environmental and Conservation Psychology, Oxford University Press (2012)
5. Schwartz, S.H.: Universals in the content and structures of values: Theoretical advances and empirical tests in 20 countries. Advances in Experimental Psychology 25, 1–65 (1992)
6. Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.H.: The WEKA data mining software: An update. SIGKDD Explorations 11(1) (2009), `http://www.cs.waikato.ac.nz/ml/weka/` (last visited on January 2013)
7. MacQueen, J.B.: Some methods for classification and analysis of multivariate observations. In: Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability, pp. 281–297. University of California Press (1967)
8. Guyon, I., Elisseeff, A.: An introduction to variable and feature selection. Journal of Machine Learning Research 3, 1157–1182 (2003)
9. Hall, M.A.: Correlation-based Feature Selection for Machine Learning. PhD thesis, University of Waikato, Hamilton, New Zealand (1999)
10. Yang, Y., Webb, G.I.: Proportional k-interval discretization for naive-bayes classifiers. In: Flach, P.A., De Raedt, L. (eds.) ECML 2001. LNCS (LNAI), vol. 2167, pp. 564–575. Springer, Heidelberg (2001)
11. Quinlan, J.R.: C4.5: Programs for Machine Learning. Morgan Kaufmann Publishers Inc. (1993)