Recent Advances in Artificial Intelligence Research and Development
J. Vitrià et al. (Eds.)
IOS Press, 2004

A Qualitative Reasoning Approach for Measuring Citizens' Satisfaction

Marta DOMINGO[*], Núria AGELL [*+]
J. Carlos AGUADO[*†], Andreu CATALÀ[*†]

[*] *Grup de Recerca en Enginyeria del Coneixement*
[+] *Departament de Mètodes Quantitatius – ESADE, Universitat Ramon Llull*
[†] *ESAII-Universitat Politècnica de Catalunya*
domingo@iiia.csic.es, agell@esade.edu,
juan.carlos.aguado@upc.es, andreu.catala@upc.es

Abstract. This paper lies within the domain of learning algorithms based on fuzzy logic connectives. It is shown how the use of qualitative information in a learning process will increase the power of generalization. A software tool able to work simultaneously with qualitative and quantitative data is constructed over the discrete structure of absolute orders of magnitude spaces. The software developed will provide city hall councillors with useful information recovered from the data they have collected in a survey designed to evaluate sustainability. The use of LAMDA algorithm to measure the satisfaction of citizens in the patterns is presented.

Keywords. Data mining, learning algorithms, fuzzy logic, qualitative modelling, sustainability, hybrid operators

Introduction

A study of citizens' indicator was commissioned by the Council of Vilanova i la Geltrú, a town with 55.000 inhabitants in Catalonia, Spain. Initially, a descriptive study was performed, based on a frame questionnaire concerning different aspects of the city. The questionnaire provided a number of qualitative questions with five possible linguistic labels, expressing a degree of satisfaction. Furthermore, additional qualitative information related to the reasons of satisfaction or dissatisfaction of citizens was collected, in the form of open questions in order to collect more specific answers.

In the past decade, urban sustainability has emerged as a key concern of designers and urban managers, in order to improve the quality of life of citizens. Several aspects must be considered in studies dealing with urban sustainability such as environmental conditions, local economy and social or cultural opportunities. The usual approach to monitor sustainability at the local level is the definition of indicators to quantify particular parameters which are supposed to be related with the above aspects. In [5] a review of historical background of sustainability and the use of indicators can be found.

Usually, quantitative and qualitative indicators are defined. While the definition and calculation of the former is direct, the latter are difficult to manage with classic techniques. Thus, in many works quantitative indicators are taken into account as a central concern of urban sustainability while other subjective perceptions of people related with the global

welfare may not be properly treated. This is considered a weakness of such works because the local scale of the analysis of sustainability is seen as a crucial step to improve sustainability at the global scale.

In the European context, a definition of sustainability indicators was addressed in Hanover (February 2000) in the meeting: 'Towards a Local Sustainability Profile-European Common Indicators'. In this meeting, local authorities of many European cities took the compromise of integrating the European Common Indicators into the existing municipal management system.

The degree of satisfaction of citizens with their locality was chosen as the first Common Indicator. Other European Core Indicators were: CO_2 emissions, Passenger transport (km/capita), percentage of population with access to green spaces, and Number of days with good air quality. These indicators were designed to measure movement towards or away from sustainability, focusing on the extent of change over time and the identification of trends and directions, rather than on absolute measures.

Some of the authors of this paper were involved in the study of the first indicator "degree of satisfaction of citizens with their locality" commissioned by the Town Council of Vilanova i la Geltrú (Catalonia, Spain), a town with over 55.000 inhabitants. Initially, a descriptive study was performed on the basis of a representative collection of data. This information was based on a frame questionnaire concerning different aspects of the city. The questionnaire provided a number of qualitative questions with some linguistic labels as answer. Moreover, qualitative information related to the reasons of satisfaction or dissatisfaction of citizens was collected too.

This work focus on the idea that qualitative data are usually undervalued in sustainability studies due to the lack of efficient tools to manage them. In other words, the potential information of qualitative indicators is not exploited because the existent data analysis tools are mainly based in classical statistics, and they are not effective tools to deal with this kind of information. Thus, a considerable amount of (qualitative) information is lost. In this article, an automatic classifier algorithm called LAMDA (Logical Association for Multivariate Data Analysis) [1,2] is applied to the analysis of a database dealing with the first indicator "degree of satisfaction of citizens with their locality". The objective is to identify and distinguish citizen profiles according to the satisfaction with the city they are living in. Thus, urban management policies could be directed to improve the aspects that are more important to satisfy a given profile of citizen.

In Section 2 the problem being treated in this study is presented. Section 3 gives the basic concepts of LAMDA supervised learning algorithm and its performance when working with quantitative or qualitative data. In Section 4, the experiments being performed and their results are described. Finally, the paper draws some conclusions and offers proposals for future research.

1. Problem definition

1.1 Description of the framework and involved variables

In the European meeting: 'Towards a Local Sustainability Profile-European Common Indicators' (Hanover, 2000), the degree of satisfaction of citizens with their locality was chosen as the first Common Indicator [3], [4], [6]. The methodology proposed to study this indicator was based on a frame questionnaire that was composed of 11 questions about several aspects of the urban environment. There were 10 variables involved with the basic urban services (natural resources protection, carefulness of public spaces, employment, cultural and leisure events, health services, education services, public transport, local governance (planning and decision making processes), public safety services, housing). An 11th variable was a more synthetic and subjective issue dealing with the convenience of the city for everyone as a good place to live and work in.

Table 1. Questionnaire

	QUESTIONS
1)	Are you satisfied about the natural resources protection (urban environment, beach, mountains)?
2)	Are you satisfied about the streets, public spaces, façades, pedestrian-only zones?
3)	Are you satisfied about the employment opportunities in Vilanova i la Geltrú?
4)	Are you satisfied about the level of cultural, sportive, and leisure services in Vilanova i la Geltrú?
5)	Are you satisfied about the level of public health and social services in Vilanova i la Geltrú?
6)	Are you satisfied about the education services in Vilanova i la Geltrú?
7)	Are you satisfied about the public transport services in Vilanova i la Geltrú?
8)	Are you satisfied about the opportunities of participation in the decision making processes of the city (municipal elections, forums, attention to the citizen problems...)?
9)	Are you satisfied about the level of security in Vilanova i la Geltrú?
10)	Are you satisfied about the accommodation opportunities in Vilanova i la Geltrú (easiness, quality, prices...)?
11)	Are you satisfied about Vilanova i la Geltrú as a globally good place to live and work in?

The questionnaire provided each question with five possible linguistic labels as answer, expressing a given degree of satisfaction.

Table 2. Possible answers

ANSWERS
Don't know
Very dissatisfied
Rather dissatisfied
Quite satisfied
Very satisfied

Additionally, an extra question was added to the frame questionnaire, asking for the best considered advantages of the city, and reasons of satisfaction or dissatisfaction. It was an open question, that is, a direct qualitative answer was allowed in order to collect more detailed answers.

1.2 Motivation and design of the qualitative reasoning approach

The questionnaire was administered to a sample of 1000 citizens from Vilanova i la Geltrú, and the qualitative reasoning approach followed up a previous classical statistic study. There were two reasons that motivated us to go further. First, the strong qualitative nature of the indicator "*degree of satisfaction of citizens with their locality*". Second, the fact that the percentage of citizens that were quite satisfied or very satisfied with the city as a good place to live and work in (variable 11) was perceptibly higher than the percentage of quite satisfied or very satisfied persons in the partial questions (variables 1 to 10). This led us to investigate whether the qualitative information collected in the additional question could reveal the reason for this difference.

The experiments presented in this paper consist in a supervised learning process using the LAMDA algorithm. They were designed under the hypothesis that, taking advantage of the whole amount of pure qualitative information in the form of the "free answer", it was possible to improve the learning process. Nevertheless, the collection and treatment of this kind of answers require considerable effort.

Two different citizens' profiles are defined: satisfied profile (citizens that are very o quite satisfied with the city as a good place to live and work) and dissatisfied profile (citizens that are very o quite dissatisfied with the city as a good place to live and work). Each individual being classified is assigned to one of these classes according to his answer to the synthetic variable (question 11).

In the next step, the learning LAMDA algorithm is used to learn the citizen profiles from the data concerning the variables 1 to 11. Finally, the whole process is repeated but adding a new pure qualitative variable and the results are compared.

2. Automated Learning Algorithm

The classification methods based on hybrid connectives combine some of the most interesting trends from both the pure numeric and the pure symbolic classification algorithms. In order to achieve this, they take profit from the generalizing power of fuzzy logic and the interpolation capability of hybrid connectives.

As is usual with data mining tools, the input data for this algorithm are in the form of a set of observations or individuals recorded as rows of a text file, each of them divided into different values, one for every descriptor or column. Every descriptor can be either quantitative or qualitative and, in the latter case, each possible value is called a descriptor attribute or modality. Obviously, in order to compute an individual's degree of adequacy to a class both with regard to qualitative and quantitative descriptors, these must be coded as a function of the same descriptor set.

In the LAMDA (Logical Association for Multivariate Data Analysis) methodology, it is assumed that a Marginal Adequacy Degree *MAD(Xi/Pi)* can be evaluated for each descriptor, and consequently a Global Adequacy Degree *GAD(X/P)* is obtained by a logical combination, or connection operator *L*, such that:

$$GAD(X/P) = L\,[MAD(X_1/P_1),\ MAD(X_2/P_2),\ \ldots,\ MAD(Xn/Pn)]$$

The operator L is defined as a linearly compensated hybrid connective [7], i. e. the interpolation between a t-norm and a t-conorm:

$$L = (1-\beta)\,T + \beta\,T^*$$

by means of the β parameter such that $\beta = 0$ represents the t-norm, T, for example the Minimum and $\beta = 1$ means the t-conorm , T*, for example the Maximum. This parameter will -inversely- determine the exigency level of the classification, so it can be called *tolerance*. Some examples of t-norms currently used in the LAMDA algorithm are MinMax, Probabilistic, Frank and Lukasiewicz.

The implementation of these possibilities assumed the form of LAMDA algorithm, which was developed by Josep Aguilar [2] in collaboration with a series of authors (Ramon López de Mántaras, Núria Piera...) who from the eighties have been enhancing this original self learning classifying technique. The input data to this algorithm are in the form of a set of observations or individuals recorded as rows of a text file, each of them divided in different values, one for every descriptor or column. Every descriptor can be either quantitative or qualitative and, in this last case, each possible value is called attribute or modality of the descriptor. The basic classification operation always is (see figure 1):

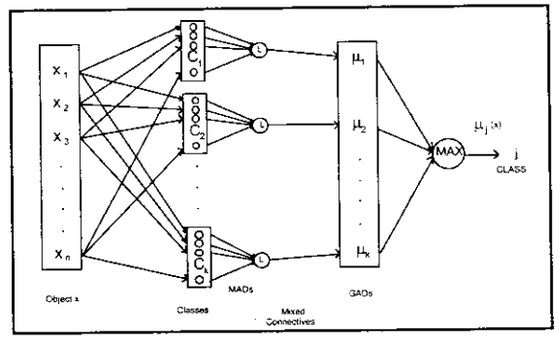

Figure 1. Hybrid connectives based classification

where MAD and GAD stand for Marginal and Global Adequacy Degree, respectively, of an individual to a given class. It should be noticed that this classifying structure remarkably resembles an artificial neural network. Thus, not amazingly, there are a training stage when classes can be modified and created, and a pure pattern recognition step when it is needed only to assign individuals to fixed classes. Nevertheless these stages can be compatible and learning could go on forever.

Obviously, in order to compute the adequacy degree of an individual to a class it is necessary that both of them are coded as a function of the same descriptors set. That means that in fact a marginal adequacy degree (MAD) will be calculated for each individual, every class and each descriptor, and these partial results will be aggregated in order to get a global adequacy degree of an individual to a class (GAD). The simplest way to build this system would be by using probability distributions and functions, and aggregating them by the simple product, but that would force to impose a series of hypothesis on the data distribution and independence which are too arbitrary.

3. Experiments and Results

Our first objective was to try and classify the citizens in the four natural classes: very unsatisfied, rather unsatisfied, quite satisfied and very satisfied. Not surprisingly, the initial results were poor, taking into account the undefinition and subjectivity of the problem addressed. On the other hand, we were also interested in showing how the addition of more qualitative data could improve the results, as it certainly did. Three tests were conducted, the first with quantitative treatment of the first 10 variables, the second with a qualitative treatment and the last one with the addition of the qualitative variable obtained from the open answers. The best degrees of correctly classified individuals were only of 53.3 % for quantitative processing, 56.6 % for qualitative treatment and 60 % by incorporating the open answer qualitative information.

The second test tried to make classes easier to separate, and consequently their number was reduced to the minimum: satisfied and unsatisfied citizens. Again, the three tests were conducted and the results turn out to be really good this time. Now the global best degrees of correctly classified individuals were of 80 % for quantitative processing, 73.3 % for qualitative treatment and 83.3 % by incorporating the open answer qualitative information.

Table 3. Best results with two classes

MinMax, Frank	Classified as Satisfied	Classified as Unsatisfied
Really Satisfied	**73.3%**	26.7%
Really Unsatisfied	6.7 %	**93.3 %**

This means that, from the town council point of view, we can tell that an unsatisfied citizen really is dissatisfied with a 93.3% of accuracy; meanwhile we will success in a 73.3% of the satisfied citizens. Thus, if the Town Council addresses policies to improve satisfaction, only a 6.7 % of unsatisfied people will be unattended.

4. Conclusion and Future Work

The citizens' satisfaction indicator, as most sustainability indicators, is measured through questionnaires. Habitually, those questionnaires are rather rigid and do not allow to capture the real reasons for the citizens to feel satisfied or unsatisfied. In fact, it is very difficult to rightly compared answers from different individuals. Nevertheless, a significant coherency can be detected if, instead of simple quantitative treatment, qualitative and open answers are allowed.

This paper belongs to a bigger project aiming at employing Artificial Intelligence Techniques (Fuzzy Logic and Qualitative Reasoning) to assess sustainability indicators, and future research will further progress in that direction.

Acknowledgements

We want to specially thank the City Council of Vilanova i la Geltrú for their kind collaboration.

References

[1] J.C. Aguado et al., "Comparison of structure and capabilities between a nonstandard classification technique and the radial basis function neural networks". Proceedings of the 13th European Simulation Multiconference (ICQFN 99), Vol. II, pp. 442-448, Warsaw, Poland, 1999.

[2] T.Kempowsky et al.,"Classification tool based on interactivity between expertise and self-learning techniques" Proceedings of the 5th IFAC Symposium on Fault Detection, Supervision and Safety of Technical Processes (SAFEPROCESS'2003), Washington (USA), pp.723-728, 2003.

[3] F.M. Andrews, S.B.Withey, *Social indicators of well being*, PlenumPress, NewYork, 1976.

[4] N.M. Bradburn, *The structure of psychological well being*, Aldine, Chicago, 1969.

[5] J. M. Castro, *Indicadores de Desarrollo Sostenible Urbano. Una Aplicación para Andalucia*, Ph D Thesis, Universidad de Málaga, 2002.

[6] P. Ferrer, A. Català, C. Angulo, Estudi dels Indicadors de Sostenibilitat 1 I 6, ESAII-RR-02-10, 2002.

[7] J.G. Klir, B. Yuan, *Fuzzy Sets and Fuzzy Logics: Theory and Applications*, Prentice Hall PTR, USA, 1995.