Discovering Marketing Rules for the Tourist Sector in Visitor Service Quality Surveys

George Stalidis
Alexander Technological Educational Institute of Thessaloniki

Abstract

Advances in information technology nowadays offer to marketers intelligent tools which are able to transform all sorts of data into useful knowledge. Well known methods for discovering knowledge in primary data and applying them to solve marketing problems can be roughly categorized in two large families, data mining and statistical data analysis.

In this paper, methods from both categories were applied on primary survey data in order to reveal the strengths and weaknesses of each approach for marketing decision support. The problem considered was to improve the positioning and the perceived quality of hotels, by relating tourist characteristics with their perceptions and satisfaction attributes.

The aim was to extract hidden knowledge from standard service quality surveys and to provide it in the form of rules to support the marketer in his decision making. Initially, data from a questionnaire-based survey on expectations and satisfaction of tourists from their hotels were analyzed using multidimensional factor analysis, followed by a knowledge modeling process, which resulted in a rule-set suitable for decisions in hotel positioning.

The next step was to apply association rule mining and decision tree methods to the same dataset, resulting in a new set of rules. It was found that the two approaches produced complementary knowledge, which revealed different aspects of the underlying trends that could be consolidated and increase the effectiveness of marketing decisions.

Keywords

Marketing Decision Support, Multidimensional Data Analysis, Data Mining, Knowledge Extraction, Rule-Based Systems, Hotel Service Quality Analysis.

Introduction

Questionnaire-based surveys are one of the most important sources of information for marketing and business planning. Advances in information technology nowadays offer to marketers analytic tools which enable them to dig into all sorts of market survey data and discover hidden information and useful knowledge, applicable to more effective marketing planning. The most typical route in order to draw conclusions from data is to interpret them with descriptive statistics, perform statistical tests to confirm hypotheses and often to apply more complex statistical methods such as factor analysis and Structural Equation Modeling.

In this work, we investigated alternative methods for explorative analysis, extraction of knowledge and modeling of the findings in the form of marketing rules. The goal was to drastically enlarge the potential of exploiting and reusing survey data by formulating computerized rules available to users who are not experts in analysis, through highly usable decision support tools.

Knowledge Discovery in Databases (KDD) is a well-known discipline that addresses exactly this goal and is defined as "the process of non-trivial extraction of implicit, previously unknown and potentially useful knowledge from data" (Adriaans and Zantige 1996). Not surprisingly, one of the major application fields of KDD is marketing, boosted by the continuously increasing abilities of information systems to gather and process large amounts of data. Popular applications in marketing are the shopping basket analysis, customer assessment and market segmentation.

While KDD is a multi-step process, the role of the discovery stage is usually undertaken by Data Mining, which is well suited to the analysis of large amounts of data, where traditional statistical analysis based on the "hypothesis and test" paradigm becomes an inefficient or simply inapplicable process. Data Mining is a combination of statistical and logical methods, together with computer science, that takes advantage of the high computational abilities of computer systems. Although statistical techniques are not replaced but rather extended, data mining is based on different philosophy which offers solutions to problems not easily handled by statistics (Han and Kamber 2001).

An alternative to data mining approach for knowledge extraction is Data Analysis, which has also proven its ability in typical marketing problems. By Data Analysis we refer to the family of multidimensional factor analysis methods, including as representative members the Multiple Correspondence Analysis (MCA) and multidimensional Hierarchical Clustering (CHA) (Benzecri 1992). These are explorative statistical methods, highly effective in the analysis of qualitative characteristics, i.e., in finding relations in a large number of nominal variables.

The typical motto for these methods is that they start from the data and show the way to the model and not the opposite (Greenacre 2007). Being highly flexible regarding data requirements, they are suitable for identifying population clusters and to highlight the properties which characterize more intensely each cluster (Stalidis and Karapistolis, 2012).

A challenging marketing problem, set as a goal in this paper in order to investigate the effectiveness of the above method families, was to extract from tourist satisfaction surveys solid and evidence-based guidelines and provide them to those involved in tourism marketing. Since more than two decades ago, it has been supported that successful marketing in tourism depends on the extent to which more specialized consumer demands or lifestyles can be identified, as opposed to massive generic approaches (Gunter and Furnham 1992; Weinstein 1994). As recent literature has shown (Kavoura and Bitsani 2013), a selective and adapted approach towards the implementation of communication activities is necessary.

The positioning strategy of a tourism product can thus be devised following the measurement of the customer's image of the tourism product (Etchner and Ritchie

1993) and his satisfaction from product attributes, in correlation with his needs and desires (Cho 1998). A two-way relationship has also been suggested (Ibrahim and Gill 2005) between the image that tourists have and the satisfaction they derive from their experience, a hypothesis which highlights the importance of placing more emphasis on the identification of customers' perception and satisfaction of a tourism product for positioning purposes.

Conklin et al (Conklin et al. 2004) elaborate on the prediction of overall satisfaction from satisfaction attributes, suggesting that certain attributes, termed "Must-be" attributes by Kano (Kano et al. 1984), have a dramatic negative impact on satisfaction when they are not delivered, although these attributes have a minimal positive impact when they are improved from an acceptable level. According to their work, there is a non-linear nature of the relationship between "Must-be" attributes and overall satisfaction, which makes the identification of such attributes difficult with standard linear modeling and they propose an analytical design based on the Shapley Value (Shapley 1953) as a decision tool to identify key dissatisfiers.

The same problem of explaining overall satisfaction in terms of satisfaction attributes, as supporting evidence to decide on marketing priorities, is also discussed in a white paper by C. Colby (Colby 2013). It is argued that the typical methodology of joining importance levels and satisfaction levels per attribute does not always adequately predict overall satisfaction and that a more sophisticated, non-linear approach to combining satisfaction measures and importance can result in a model with greater ability to explain this phenomenon.

A quite popular tool for measuring service quality is servqual (Parasuraman et al. 1988), which is also attribute-based and may provide insights on how perceived quality is affected by individual factors/attributes (Tribe and Snaith 1998). The analysis of servqual surveys can thus be used to produce guidelines for enhancing performance (Zeithaml et al. 1990). This approach has been recently applied in the field of hotel industry by performing a primary survey based on a modified servqual questionnaire and analyzing quantitatively the factors which explained customer perceptions (Markovic and Raspor 2010).

The focus of this paper was to investigate intelligent computerized methods for knowledge extraction from primary survey data, as an alternative approach to classical statistical analysis. In particular, the presented work was aimed at the problem of positioning and enhancing the perceived quality of hotels by supporting the marketer in his decision making with engineered marketing rules, extracted from standard service quality surveys. In the next section, the adopted analysis and data mining methods are introduced and briefly presented. Then, the research application sought is presented and results are experimental results are provided. Finally, the outcomes of the two rule extraction approaches are discussed and conclusions are drawn.

Methods

Discovery of Marketing Rules Using Data Mining

Introduction to Data Mining Methods

In the context of marketing, data mining can be used to produce new knowledge, in order to take strategic decisions, specify targets, improve planning regarding the product, campaign or promotion (Cooper and Giufrida 2000). In order to achieve this, patterns and relationships are used to build models of the real world, suitable for estimations and predictions. As there are a number of methods for different purposes, successful application of data mining depends on clear definition of business goals, on the specification of the appropriate model and finally on the effectiveness of the applied algorithms (Shreiber 2008).

Among the possible problems are a classification (deciding or predicting in which category a case falls), a clustering (segmenting a population to homogeneous groups), a logical inference (predict an outcome given a number of facts), regression (estimate the value of a numerical variable given a set of parameters) or a time series prediction (learn from the past and extrapolate to the future). A model should then be selected to match the targeted problem. Decision trees can be used for classification, association rules for patterns, neural networks for regression or support vector machines for clustering (Roiger and Geatz 2003).

Many algorithms are available to build models, some of which had so wide acceptance that can be characterized as standard. In a survey paper by X. Wu et al (Wu 2008), the top 10 algorithms are presented among the most influential data mining algorithms in the research community. Among them are the C4.5 and CART algorithms for building decision trees and the apriori algorithm for solving the classical marketing problem of shopping basket analysis. Other standard algorithms are k-nearest neighbors, CHAID trees and backpropagation neural networks. In the present work, the focus was on the methods of Association Rules and Classification Trees, coupled with rule induction processes. These were selected as the most promising approaches to the problem dealt with in this paper and are briefly presented in the following.

Decision and Classification Trees

Classification trees are designed to examine the effects of a large number of predictor variables to a selected target variable, performing successive univariate splits. They are flexible regarding the variety of predictors that can be analyzed, including categorical (e.g. gender, opinion), continuous (e.g. price) or mix of the two types of predictors. This makes Classification Trees flexible and their interpretation quite simple and almost natural.

In the learning phase, a tree is built on a set of known examples from which it is expected to learn a "good" way of classifying these examples in the classes of a categorical dependant variable from their measurements on one or more predictor variable. The tree is then used to classify unseen examples in the same way (Han and Kamber 2001). There are mainly two methods of tree growing algorithms: CART and CHAID, differing in the type of splitting criterion used. The CHAID method uses the Chi2 statistic for contingency tables while the CART family is oriented to statistics using the concept of impurity (Breiman 1984).

A rule induction process can be used to transform a Tree into sets of logical rules by decomposing it to individual paths from the root to a leaf. Such paths are selected to formulate rules, which can be used in a subsequent knowledge engineering phase to build Knowledge Based Systems (Fayyad et al. 1996).

The format of such Rules is as follows (Ligeza 2006):

Condition 1 AND... AND *Condition n* → *Outcome* (confidence parameters)

Each condition of such a rule corresponds to a variable having a specific value or range of values (e.g. customer age < 30) and the strength of each rule is indicated by the confidence parameters "Support" (number of cases where all conditions are met) and "Confidence" (percentage of correct predictions in relation to the cases where the rule is applicable) (Ligeza 2006).

Association Rules

Another approach to rule extraction from data is Mining of Association Rules (Agrawal and Srikant 1994). Association rules are if/then statements that help uncover relationships between seemingly unrelated data. Unlike classification trees which are forms of supervised learning (in which there is a specific target variable), in this case the learning method searches for any association between features, without any specific target variable. These associations are detected as combinations of properties or facts found frequently in large datasets. An example is the recommendation systems used in many online shopping systems: If you bought X, then you may also like Y.

Association rules of this kind are extracted in two phases. Firstly, a large number of cases are analyzed to detect frequent itemsets, that is combinations of attributes which are found together frequently. Standard methods used for this task are the Apriori algorithm and the FP-Growth tree (Piatetsky-Shapiro and Frawley 1991). The next step is to generate rules that predict the occurrence of an item if a set of other items are present, using the criteria of support and confidence to identify the most important relationships.

Support is an indication of how frequently the items appear in the database and is thus a measure of how important this rule is. Confidence indicates the number of times the rule has been found to be true, measuring the prediction accuracy of the rule. Other measures are also used to assess a rule, such as the lift, which shows how far from independence is the predicted association and therefore the interestingness of the rule (Quinlan 1986).

Rule Formulation Using Data Analysis

The problem of extracting marketing rules from survey data is by its nature qualitative and explorative and therefore calls for methods that perform well with nominal or logical variables, independently from quantitative scales and a-priori hypotheses. In previous work (Stalidis 2012), methods from the branch of multidimensional data analysis, in particular a combination of Multiple Correspondence Analysis (MCA) (Benzecri 1992; Greenacre 2007) and Hierarchical Cluster Analysis (CHA) based on Benzecri's chi-square distance and Ward's linkage criterion were used in order to extract knowledge from survey data. The results of the statistical analysis were of qualitative nature and were suitable for coding as computerized rules for knowledge-based decision support.

In order to uncover the underlying factors that explained the respondents' behavior, MCA was applied, including all the variables involved in a particular exploration. The MCA method handled all variables (including items of multi-item constructs) as categorical variables and all possible answers (i.e. categories) were seen as properties – equivalent to the definition of attribute values in data mining. The aim was to find significant associations and disjunctions among properties and to use the resulting patterns to construct behavior classes.

For example, a finding that a group of associated properties includes the answers age>65 and requirements for security and parking, as opposed to another group of properties including expectations for bar, swimming pool and age <35 would reveal the existence of two classes corresponding to different ages and priorities. The CHA method was applied in parallel to cluster individuals on the basis of the full set of their responses. The result was to segment the sample to homogeneous groups of measurable size.

The next step in the process was to apply MCA, including both the questionnaire-related variables and the group membership variable. In this way, groups of individuals were associated with classes and properties, and thus interpretations were obtained about who (or how many individuals) share similar views and which were their dominating properties. It is worth mentioning that the discovered associations were defined among properties (i.e. categories) and not among variables, as would be the case in quantitative analysis. In this way, it was possible to reveal and express heavily non-linear dependencies, while the result was not related to any a-priori hypothesis or model.

Although the Data Analysis concluded to the graphical interpretation of factorial planes, a methodology has been proposed for expressing the findings as rules, in a formulation similar with the one mentioned in the data mining section. More specifically, in (Stalidis 2012) the development of an ontology (Prantner et al 2007) for the tourist sector was presented, which was used as the vocabulary to formulate machine understandable marketing rules (e.g. John **is_a** *good_customer,* where John is an object of class *customer* and *good_customer* is a subclass of class *customer)*.

The resulting rule model was more sophisticated than the one produced by data mining processes, since conditions and outcomes were not restricted to expressions of variable values but could also express higher level facts, such as an object having a particular property, on the basis of a special vocabulary, However, it is also clarified that the rule formulation needed to be performed manually by the analyst, as opposed to the case of data mining, where rules were created automatically.

Applications and Experimental Results

Application Scenario and Input Data

The application was based on a service quality survey addressed to tourists who were asked to evaluate their hotel. The survey was carried out during the summer of 2010 at seaside destinations in Northern Greece. The questionnaire was in the category of servqual, designed to collect information about the expectations and satisfaction of the visitors regarding their hotel. The questionnaire also captured information about the demographic characteristics of the visitors and the purpose/type of their trip.

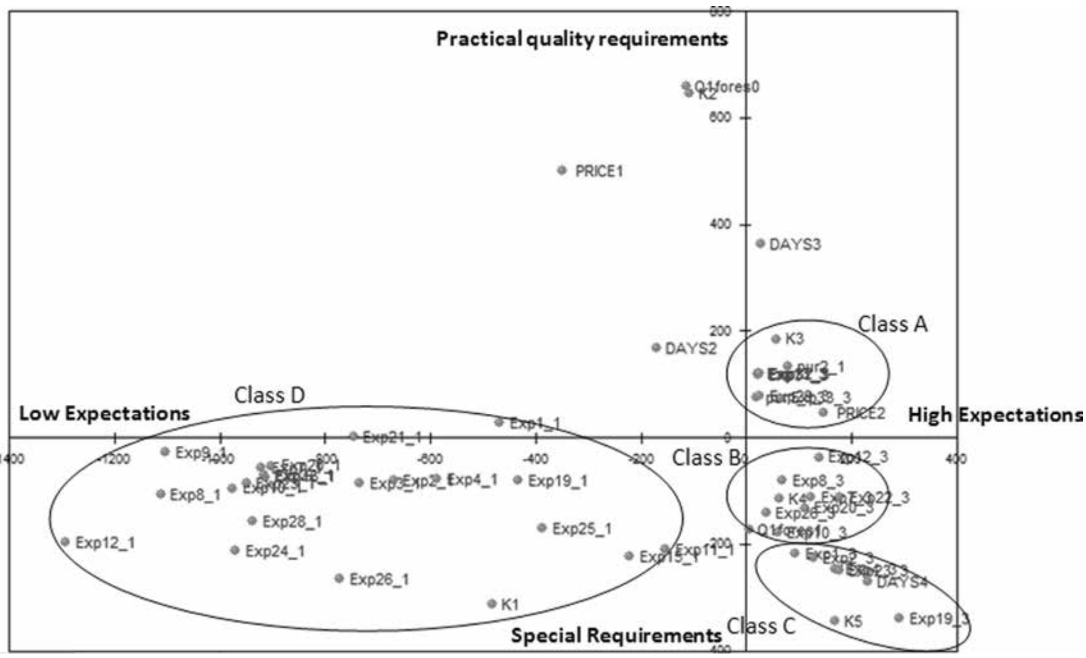
In addition to the main purpose of this survey – which was to evaluate the perceived quality of the hotels in the area and to study the factors explaining satisfaction – the aims of our analysis were to:

- Find some (possibly non linear) structure in the preferences and expectation attributes of the hotel customers, in order to identify representative visitor classes and to associate these classes with priority expectation attributes.
- Associate expectations from the hotel to demographic characteristics of the visitors and to the purpose of their visit.
- Come up with predictors of overall satisfaction, in the form of attribute combinations specialized to individual customer groups, instead of generic overall factors.

The goal was to discover conditional rules that predict visitor satisfaction or indicate patterns matching visitors with options and requirements, so that the marketer would be able to decide which hotel characteristics deserve more emphasis, given their target customers or, vice-versa, which customers are worth aiming at, given the characteristics of the hotel.

The questionnaire consisted of 21 closed-type questions. In two of them, the respondents indicated on a five-point scale the degree of their expectation/satisfaction for each one of 33 attributes. Additionally, they were asked about the purpose and duration of their visit, the amount they were willing to pay for their room as well as a few demographic characteristics. The dataset included a sample of 400 visitors. From those, 41% were men and 59% women, 33% were aged from 26 to 35 years old, 55% were from 36 to 55 and 19% were older than 55, while most of the respondents (67%) had completed higher education. Most of the visitors stated that the purpose of their visit was vacation and just 8% mentioned professional reasons.

Figure 1: The factorial plane 1X2 (explaining 41,5% of intertia), where four visitor classes were identified in terms of their expectation attributes. The horizontal axis differentiates properties expressing low expectations from those expressing high ones. The vertical axis expresses the contrast between practical requirements and special requirements



Rule Extraction Using Multidimensional Data Analysis

The analysis process was initially applied on the variables related to the characteristics of the visit (e.g. reason for visit, duration, cost category) together with the expectation attributes from the hotel, in order to explore possible patterns in visitor needs. The MCA method showed that the first four factorial axes explained just 55,9% of total inertia, indicating that the phenomenon was quite complex. The 1st factorial axis (29.7% of inertia), revealed a strong contrast between properties showing low expectations - low cost and those related to high expectations for specialized items. This axis thus represented the positioning from low budget to high quality vacation. The 2nd factorial axis (11.8% of inertia) differentiated basic practical requirements (e.g., cleanness, consumables) from more specialized demands (e.g. sports facilities and spa). The 3rd and 4th factorial axes (7.3% and 6.9% of inertia, respectively) revealed additional, less apparent associations among requirements, corresponding to different styles of visitors. Figure 1 illustrates the factorial plane 1X2, where properties are projected with respect to the factors: low cost vs high quality and practical requirements vs special requirements.

From the formations of properties on this plane, four classes were identified, corresponding to expectations for:

- Basic practical items and medium cost
- Entertainment and comfort
- Wellness, sports activities and special diet
- Lowest price

Additional trends, less prevailing but also interesting, were found on the factorial plane 3X4 (14,2% of inertia), for example the association among the expectations for beach nearby, satisfactory room size and the hotel to belong to a group of hotels in contrast to requirements for swimming pool, local traditional style and family-type hotel. In the next steps, visitors were segmented to five homogeneous groups using CHA. The associations among the centers of these groups and the previously identified classes, as well as with demographic properties were then investigated. The results are summarized in Table 1.

Table 1: Classification of Visitors in Terms of their Requirements

Group	Associated Class and Properties	Demographic properties
K1 N=174, 43,4%	**Class D – Low budget** Willing to spend up to 50€ Low expectations overall.	
K2 N=102, 25,5%	First time visit to this destination Willing to spend up to 50€	Age 26-35
K3 N=32, 8,1%	**Class A - Basic quality vacation** Purpose of visit is vacation or entertainment, Duration 2 weeks Willing to spend 50 to 100€, High expectations for security, cleanness, materials and consumables (i.e. shampoos, towels, etc.), hairdryer, restaurant in the hotel.	Age 56-65, Profession: retired
K4 N=16, 4%	**Class B – Recreational vacation** Has been in this destination once or twice before, The hotel to belong to a group of hotels, Expectations for swimming pool, entertainment activities, comfortable lobby.	Age 56-65, Profession: retired
K5 N=76, 18,9%	**Class C - Activities and wellness** Duration 3 weeks or more, Willing to spend 100-150 per night, Requirements for spa and wellness services, facilities for persons with special needs, sports facilities, special diet menu.	Age 46-55, Profession: freelancer, Medium income

The above findings were expressed as rules and stored in an electronic knowledge base, as presented in earlier work (Stalidis, 2012), utilizing the Protégé framework, the OWL-DL language for developing the ontology and the SWRL language for expressing the rules (Protégé 2012). The outcome of the rule formation process for the specific application was 64 rules.

As an example, the profile of Group "Basic quality vacation" drove the following two rules:

- **IF** *trip* hasTripFeature *Entertainement* **AND** *trip* hasDuration *TwoWeeks* **AND** *visitor* hasRequirement *Cost50to100* **THEN** *trip* hasTripFeature *BasicQualityVacationTrip*
- **IF** *trip* hasTripFeature *StandardQuality VacationTrip* **AND** *visitor* hasVisitorFeature *Age56-65* **THEN** *visitor* hasRequirement *Security*

Mining of Association Rules and Classification Trees

Following a parallel route with that of Data Analysis, the same dataset was explored by employing association rules and classification trees. The first analysis process was based on unsupervised learning, in order to discover patterns in the visitor expectations without any prior knowledge. The second one was supervised learning (classification) where the target variable was the general satisfaction of the visitor and the goal was to specify the factors that mostly influence this satisfaction and the elements that can be used to predict it. The data mining analysis was performed on RapidMiner 5.3 from Rapid-I GmbH, using the free version, licensed under AGPL (www.rapidminer.com). RapidMiner offered not only a complete set of algorithms for the most common data mining problems but also a convenient and effective graphical environment for building analysis and testing processes. In this environment, preprocessing and modeling components were offered as modules that could be connected to each other graphically, to compose complete and reusable processes.

Rules on Expectation Patterns

The problem of discovering patterns of associated visitor expectations was similar to the shopping basket problem. In the survey questionnaire, each visitor selects from a large list of possible expectations what he considers important or not, in the same sense that a supermarket customer buys a set of items. By identifying frequent item sets, i.e. expectations which are often selected together, it should be possible to predict what a visitor may also want, whenever he has a specific request.

Figure 2: The Datamining Process Created with the Graphical Tool of RapidMiner

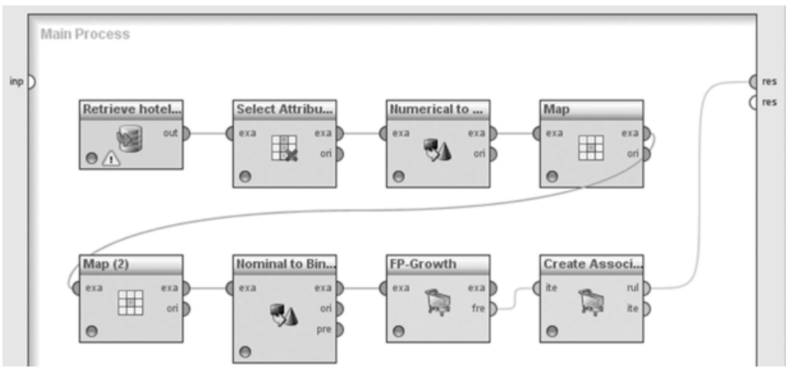


Figure 3: An Extract of the Association Rules on Expectation Patterns

No.	Premises	Conclusion	Support	Confidence	LaPlace	Gain	p-s	Lift	Conviction
52	pur12, DAYS = 4	Exp7	0.217	0.938	0.988	-0.246	0.112	2.069	8.836
51	DAYS = 4	Exp7	0.224	0.931	0.987	-0.258	0.115	2.052	7.886
50	Exp1	Exp4	0.210	0.926	0.986	-0.243	0.125	2.456	8.454
49	pur2, Exp15	Exp4	0.224	0.913	0.983	-0.267	0.132	2.420	7.129
48	pur12, Exp11	Exp4	0.222	0.912	0.983	-0.265	0.130	2.418	7.060
47	Exp23	Exp4	0.217	0.910	0.983	-0.260	0.127	2.413	6.921
46	pur12, pur2, Exp15	Exp4	0.210	0.907	0.983	-0.253	0.123	2.406	6.714
45	Exp11	Exp4	0.243	0.903	0.979	-0.296	0.142	2.394	6.399
44	DAYS = 4	pur12, Exp7	0.217	0.901	0.981	-0.265	0.114	2.109	5.785
43	pur2, Exp15	pur12, Exp4	0.210	0.854	0.971	-0.282	0.125	2.469	4.490
42	pur12, Exp21	Exp15	0.222	0.853	0.970	-0.298	0.130	2.416	4.406
41	Exp21	Exp15	0.239	0.847	0.966	-0.325	0.139	2.399	4.240
40	Exp11	Exp15	0.224	0.832	0.964	-0.315	0.129	2.355	3.847
39	Exp11	pur12, Exp4	0.222	0.823	0.962	-0.317	0.129	2.378	3.695
38	pur12, pur2, Exp7	Exp25	0.243	0.823	0.959	-0.348	0.137	2.283	3.605
37	pur2, Exp7	Exp25	0.251	0.814	0.956	-0.365	0.140	2.259	3.438
36	Exp7, Exp4	Exp25	0.208	0.798	0.958	-0.313	0.114	2.215	3.169

The mining process is shown graphically in Figure 2. The input data were constructed from the section on expectations (33 items for which the tourist stated the degree of importance), the purpose of visit and the duration of the visit. The values were transformed to binary by setting low importance and neutrality to 0 (negative) and high / very high importance to 1 (positive). Frequent itemsets were detected using the FP-Growth algorithm, while association rules were extracted using as criterion the lift of the rule. After some experimentation with the criteria of rule confidence, lift and conviction, as well as with the corresponding thresholds, the most interesting results were derived using the lift (min value = 2.0, gain theta = 2, Laplace-k = 1). This criterion prefers rules that deviate from independence, and are thus the most interesting, as opposed to the criterion of confidence which prefers rules with maximum prediction accuracy, which however are often trivial without much practical value.

The algorithm produced 52 rules, an extract of which are shown in Figure 3, sorted according to their confidence. For example, rule 52 which had the highest confidence, predicts that if the duration of visit is three weeks or more (Days=4) and the purpose of visit is contact with nature (pur12) then with probability 93,8% the visitor also expects swimming pool for children (exp7). Rule 50 predicts that if the visitor expects facilities for people with special needs (exp1) then with probability 92,6% he also requires lighting at the external spaces at night (exp4).

Rules Explaining Overall Satisfaction

The next mining process was applied on the variables regarding satisfaction (overall satisfaction and specifically from a large number of individual items) and the variable "purpose of visit". A classification tree was then grown, as shown in Figure 4, setting the general satisfaction as the target variable. The trained classification tree worked as a predictor of the overall visitor satisfaction from the individual satisfaction levels per attribute and was thus a data-driven model reflecting the role of each attribute to overall visitor's opinion.

The extracted rule model consisted of the decision paths which lead with high confidence to the most "pure" classes of low, medium or high contentment, so the combinations of attributes participating as antecedents in such rules were interpreted as the most decisive contributing factors explaining the outcome. Figure 4 illustrates the tree resulted from 400 examples, achieving a prediction accuracy of 370/400 (parameters: criterion Gini index, min gain = 0.15, confidence 0.2, min leaf size = 4, min size for split = 4, num of prepruning alternatives = 3). Among the results it can be highlighted that the root node was "security" (Sat29), meaning that this criterion gave the first best split for classifying the sample.

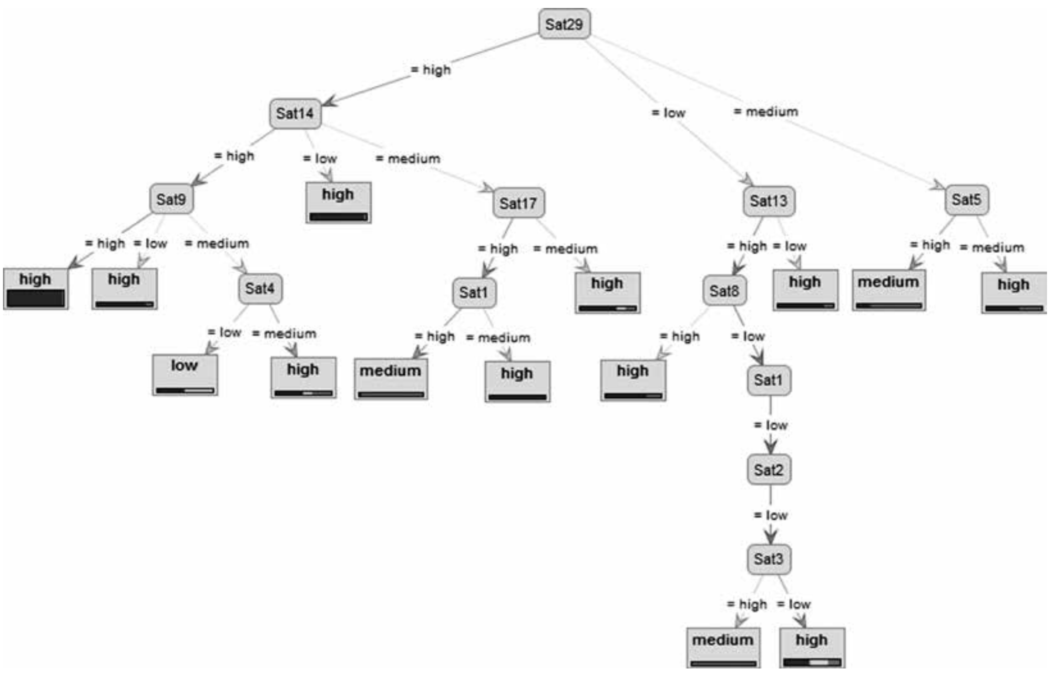
The rule with the highest discriminative power, classifying 212 from 217 cases correctly to the high satisfaction class was the following:

- **If** security(Sat29) = high **and** skilled and helpful personnel(Sat14) = high **and** quality and variety of food(Sat9) = high **then** high (212 / 0 / 5)

Another rule, predicting medium satisfaction was:

- **If** security(Sat29) = low **and** Laundry(Sat13) = high **and** hotel location (Sat8) = low **and** facilities for special needs(Sat1) = low **and** relaxation facilities &Spa(Sat2) = low **and** Sat3(lighting at the external spaces at night) = high **then** medium (0 / 0 / 17)

Figure 4: The classification tree which explains overall satisfaction using as criteria the satisfaction levels in individual attributes.



The resulting rule model consisted of 14 rules. More rules were initially produced but were rejected either due to their low reliability or because they were simply not meaningful to the marketer.

Discussion and Conclusions

A qualitative comparison between the outcomes of data analysis vs data mining, confirmed that the results reflected the different nature of the two method families. Multidimensional data analysis handled complex profiles of both individuals and properties, synthesized by all variables simultaneously. The produced associations or classifications reflected a holistic view of similarities/dissimilarities, where each variable contributed according to its significance.

Furthermore, the formation of factorial axes was based on divergence from independence, rather than absolute frequencies so that the factors revealed trends and contrasts in a qualitative way. The produced rules reflected the existence of profiles and qualitative trends, such as the class “basic quality vacation”, characterized by high expectations for security, cleanness, materials and consumables, duration 2 weeks, cost 50 to 100€, etc.

However, when counting the individuals who gave exactly this set of responses, the result was zero. The specific profile used to classify tourists was therefore not a real case but a center in a multidimensional space around which there was a cloud of properties and individuals. The classification of a tourist in such a class would depend on complex as well as relative decision borders instead of hard decisions.

Conversely, in the case of data mining, all association and classification rules were derived by counting and splitting sets of real cases. Variables were considered one at a time and the formulated rules seemed more clear, yet simplistic. An important advantage of data mining was the automatic creation of rules and the availability of confidence measures. This means that these methods were better candidates for fully automated processes to drive an inference engine, whereas data analysis was dependent on the human analyst to perform rule elicitation.

A formal comparison between the actual contents was not possible because of the different nature of the two kinds of rules, however it was evident that there was a small overlap and a large margin for complementary usage. A limitation of the current work was that the methods were applied on a single dataset, since the scope was a feasibility study for extracting marketing rules from quality assessment surveys and a qualitative comparison of the two promising approaches.

The presented work is also limited to the knowledge extraction phase and does not include the full knowledge engineering cycle. As future work, it is planned to expand the rule extraction processes to more datasets and to evaluate the problem solving abilities of the produced knowledge in realistic pilot cases.

In this paper, an unconventional approach in processing service quality surveys for marketing purposes was illustrated. Instead of classical statistical analysis to interpret customer perceptions and needs, experimentations were performed with knowledge extraction methods. It was set as a goal to formulate the survey results as marketing rules that can be used through computerized decision support tools by non-analysts to perform marketing planning.

The expectations of tourists from their hotel were explored, deriving rules on how to classify tourists in terms of their needs and how to predict their requirements according to the type of their visit. Rules were also extracted on how to predict/control overall satisfaction from individual attributes, without being limited to linear relations. As a conclusion, both multidimensional data analysis and data mining methods were found promising, each one revealing different aspects of the study and therefore offering its own value.

References

- Adriaans, Pieter & Dolf Zantige (1996). Data mining. Edingburgh: Addison Wesley.
- Agrawal, Rakesh & Ramakrishnan Srikant (1994). Fast algorithms for mining association rules. In: *Proceedings of the 20th VLDB conference*, pp. 487–499.
- Benzecri, Jean-Paul (1992). Correspondence Analysis Handbook. New-York: Dekker, P.
- Breiman, Leo (1984). Classification and regression trees. Belmont, Calif.:Wadsworth International Group.
- Cho, BaeHaeng (1998). Assessing tourist satisfaction, *Tourism Recreation Research*, 23(1), pp. 47-54.
- Colby, Charles (2013). Measuring the Importance of Customer Satisfaction Attributes, White paper by Rockresearch. Available at::http://www.rockresearch.com/measuring-theimportance-of-customer-satisfaction attributes.
- Conklin Michael., Ken Powaga. & Stan Lipovetsky. (2004). Customer Satisfaction Analysis: Identification of Key Drivers, *European Journal of Operational Research*, 154(3), pp.819-27.
- Cooper, Lee G. & Giovanni Giuffrida (2000). Turning Datamining into a Management Science Tool: New algorithms and Empirical Results, *Management Science*, 46(2), February 2000 pp. 249-264.
- Etchner, C. & Brent Ritchie (1993). The measurement of destination image: An empirical assessment, *Journal of Travel Research*, Spring 1993, pp. 3-13
- Fayyad, Usama., Gregory, Piatetsky-Shapiro & Padhraic Smyth ed. (1996). From data mining to knowledge discovery: an overview. In:2nd Int Conf on Knowledge Discovery and Data Mining.
- Greenacre, Michael (2007). Correspondence Analysis in Practice. Chapman & Hall.
- Gunter, Barrie & Adrian Furnham (1992). Consumer Profiles: An Introduction to Psychographics.
- London: Routledge.
- Han, Jiawei & Micheline Kamber (2001). Data Mining – Concepts and Technique. San Diego: Academic Press.
- Ibrahim, Essam E. & Jacqueline Gill (2005). A positioning strategy for a tourist destination, based on analysis of customers’ perceptions and satisfactions, *Marketing Intelligence & Planning,* 23(2), pp. 172-188.
- Kano, Noriaki., Nobuhiko Seraku,Fumio Takahashi & Shin Tsuji (1984). Attractive Quality and Must Be Quality, *Quality*, 14(2), pp. 39-48.
- Kavoura, Androniki & Evgenia, Bitsani (2013). Managing the World Heritage Site of the Acropolis, Greece, Advances in Culture*, Tourism and Hospitality Research*, 7, pp.58-67.
- Ligêza, Antoni (2006). Logical foundations for rule-based systems, Studies in Computational Intelligence, vol. 11, 2nd Ed., Berlin Heidelberg: Springer-Verlag.

- Markovic, Suzana & Sanja, Raspor (2010). Measuring Perceived Service Quality Using servqual: A Case Study of the Croatian Hotel Industry, *Management,* 5 (3), pp:195–209.
- Parasuraman, A.., Valarie, A.. Zeithaml. & Leonard, L. Berry. (1988). Servqual:A multiple-item scale for measuring consumer perceptions of service quality, *Journal of Retailing,* 64 (1), pp.14–40.
- Piatetsky-Shapiro, Gregory & William Frawley (1991). Knowledge discovery in databases. Menlo Park, Calif.: AAAI Press: MIT Press.
- Prantner Kathrin., Ying Ding., Luger, Michael., Zhixian Yan & Christoph Herzog (2007). Tourism Ontology and Semantic Management System: State-of-the-Arts Analysis. In: Proceedings of IADIS International Conference WWW/Internet 2007, pp.111-5.
- Protégé, (2012). http://protege.stanford.edu/
- Quinlan, J.R.. (1986). Introduction of Decision Trees, *Machine Learning,* 1, pp.81-106.
- Roiger, Richard J. & Michael Geatz (2003). Data Mining. A Tutorial-Based Primer*.* Addison Wesley,
- Shapley, Lloyd Stowell (1953). A Value for n-Person Games. In: Kuhn H.W., Tucker A.W. (ed), Contribution to the Theory of Games, II. Princeton, NJ: Princeton University Press, pp.307-17.
- Schreiber, Guus (2008). Knowledge Engineering.In: F. van Harmelen, V. Lifschitz, B. Porter (ed). Handbook of Knowledge Representation, Elsevier, pp. 929–46.
- Stalidis, George, (2012). Explorative Data Analysis and Knowledge Modeling methods for marketing decision support applied in the tourist sector, ICCMI 2012, Thessaloniki, pp. 601-607.
- Stalidis, George & Dimitrios Karapistolis (2012). Data Analysis to support business planning: application to a novel internet radio business, *Data Analysis Bulletin*, 13(12), pp. 59-72.
- Tribe, John & Tim, Snaith (1998). From SERVQUAL to HOLSAT: holiday satisfaction in Varadero, Cuba, *Tourism Management*, 19(1), pp. 25-34.
- Weinstein, Art (1994). Market Segmentation: Using Demographics, Psychographics, and other Niche Marketing Techniques to Predict and Model Customer Behaviour, Probus, Chicago, IL.
- Wu, Xindong et al (2008). Top 10 algorithms in data mining, Knowledge Information Systems, Vol.14, pp.1–37, DOI10.1007/s10115-007-0114-2
- Zeithaml, Valarie A., Parasuraman,A. & Leonard L. Berry (1990). Delivering Quality Service Balancing Customer Perceptions and Expectations**.** New York: The Free Press**.**