

Evaluation Technique of the Spatial Development of Tourist Cluster of the Region

Natalya Stepanovna Martysenko and Angelina Georgievna Kim

Vladivostok state university economy and service,
Vladivostok, Russia, 690014, Vladivostok, street Gogol, 41

Abstract: This paper deals with technique for handling unstructured data, which significantly extends the capabilities of researchers in the study of socio-economic phenomena and processes. The concept of a composite attribute is defined. The technique allows producing the typologies of the values of attributes studied. For automation of the unstructured data processing an expert system based on the use of three types of computer dictionaries has been developed. Presentation of information in the form of composite attributes allows analyzing the interaction of different structural groups of consumers.

Key words: Unstructured data processing technique • Expert system • Data analysis • Tourism
• Questionnaire survey

INTRODUCTION

One of the key issues of fundamental and applied research in the field of tourism is the development of effective predicting and managing models of tourist movement. The main obstacle to their development and use is the lack of sufficient information on the processes of tourist product consumption in the region and the development trends of human needs evolving in a variety of types of recreational, leisure and other activities. Such information may be obtained from mass surveys of the consumers of tourist product [1].

More than ten years ago, the authors have started monitoring the behavior of the tourist product consumers of Primorsky Krai. A series of questionnaire surveys was carried out during this period. Nearly 1.5-2 thous. citizens of Primorsky Krai of Russia are polled annually.

While drawing up the questions for survey it was found that it turns to be impossible to determine a number of important characteristics of the process of consumption of tourist product when using the common structured questions. As numerous survey experiences have shown, a person answers more precisely (and more easily) the questions of qualitative or comparative nature than of quantitative one.

Therefore, our personal data is different in that they contain bulk non-numerical information, which is generated from a variety of measuring scales used in questionnaires. The extensive range of scales is caused by the desire of the researchers to obtain information that is more reliable from respondents rather than their whim. The researcher is always forced to seek a compromise between the desired information and the information he/she can get as a result of the questionnaire survey. The various forms of open questions were widely used when drawing up the questionnaires.

Open or unstructured questions are the most difficult in terms of computer processing of questionnaire data. Such questions, unlike the closed ones, do not contain any hints or "impose" any variant of the answer and are intended to receive a non-formalized opinion. Some of researchers do not apply the computer processing of open questions and use them to search for and obtain information for further researches. Meanwhile, the answers to these questions can be very informative.

The problem of computer analysis of qualitative data has attracted a large number of researchers around the world. The definition of problem of computer analysis of qualitative data is given in paper [2]. Among Russian scientists specializing in the processing of qualitative information are Davydov A.A. [3] and G.V. Kanygin [4].

Corresponding Author: Natalya Stepanovna Martysenko, Vladivostok state university economy and service,
Vladivostok, Russia, 690014, Vladivostok, street Gogol, 41.

The already-known means of computer processing of qualitative data, such as NVIO [5], are quite difficult for most of marketers. Therefore, an attempt was made to create a simple computational tool that could be straightforward for a wide range of researchers.

This paper suggests you first consider elements of the computer technology offered, which allows transiting from the unstructured form of data presentation to the structured one (typologies). Then we consider the examples of the construction of typologies and their use in analyzing the structures of spatial development of the tourist complex in the region.

Technique: It is customary to represent questionnaire data in the form of table “object-property”. This table is easily placed on a separate EXCEL sheet. Open question data presented in the form of text are placed into one column of the table. Moreover, we consider that the answer may be multiple. For example, when answering the question “What else do you like to do going to the seaside, except for sun bathing and swimming?” the respondent can give several variants, such as “play volleyball, admire landscape, fish” etc. An attribute in the table “object-property”, containing data on such issue, is called composite.

Answers in the composite attribute may consist of several more simple statements. Simple statements in the respondents' answers are separated by a character, such as “;” or “,”. In some cases, more complex statements may be in the form of complete sentences. The simplest answer may only consist of one simple statement. Admitting such answers to the open questions, we give free rein to a respondent.

Now we will define what we have at the output of information technology of open question processing. First, simple statements will be considered, such as a special case of composite attribute (property). In case of open question it could be expected, that the respondents would not give the same answers. There are many identical or similar statements in practice, even without mentioning simple and orthographical errors. The list of answers, different in substance and not in form, to such questions is limited. When already having a sample of 700 questionnaires only 30-50 different variants of answers may be selected, which can be interpreted as the attribute value measured in nominal scale. An increase in sample volume does not practically change the list of variants.

For processing data of open question, the typing technique is used [6]. Typing technique is a substitution of original simple statement (in the form of text) with

similar or generalizing one (in the form of text). To perform typing the auxiliary table “List of attribute value” is formed. When calculating the table “List of attribute values” the composite attribute is divided into simple. One of the columns of this table includes all the unique values of the original attribute. In addition, it contains a column where the frequencies of values encountered are calculated. The operation of typing is applied to data of table “List of attribute values” rather than to original data of table “object-property”. First, the simple situations are processed. For example, different spelling of a word or different word order. The most appropriate (more accurate and competent) form of statement spelling is chosen from similar statements and then this statement is copied into the cells of table “list of attribute values” with similar statements. Replacing any of unique statements with one, presented in the list of values, we reduce the number of rows in the table “list of attribute values?”. After performing a series of substitutions it is advisable to perform the “compression”, which is based on the recalculation of the table “list of attribute values?”. The table “list of attribute values?” slowly decreases and becomes more illustrative.

After completing the processing of more simple situations, the more complex cases will be processed. A group of less common statements, still having the same subject, is found in the table “List of attribute values”. For this group of simple statements the researcher chooses some generalizing statement in the table and, if none is found, formulates new one, expressing the general meaning or subject of the group of simple statements.

For example, when answering the question “What else do you like to do going to the seaside, except for sun bathing and swimming?” the respondents have answered the following:

“Jet-skiing”, “hang-gliding”, “parasailing”, “cliff diving”, “rock-climbing” etc.

However, these statements have occurred quite rarely (less than 0.1%) and we decided to substitute them with a generalizing statement, such as “extreme”, which was found in the table “list of attribute values”. Generally, a meaning of the statements has remained the same.

In order to avoid the loss of information, especially when conducting the resurvey, we substitute the similar statements with the generalizing ones, being clarified. Clarification or nuance are given in brackets. For example, we have substituted the original values in the above cases with “extreme (jet-skiing)”, “extreme (hang-gliding)”, “extreme (parasailing)”, “extreme (cliff diving)” and “extreme (rock-climbing)”.

The nature of answer, which determines the type of personality of the respondent (consumer), is more important for us than the specific content of the response. If the original table “list of attribute values” may contain several thousand values, the same table after processing (typing) usually contains several hundred values, including the values clarified. Creation of this table completes the first stage of typing (the first level).

A new attribute obtained still has too many different values to be analyzed. Therefore this attribute is subjected to further processing (second level). At this stage, the clarifications given in brackets are simply excluded and another column is formed in the table “list of attribute values”, called subclass, which will contain from 30 up to 50 unique statements.

Availability of 30-50 different values is also a quite large number to analyze measurements in nominal scale. Therefore, the researcher has, after forming an acceptable list of really different answers, to group these answers, considering them as some of the characteristics of non-crossing classes, types or subjects, depending on the meaningful sense of the attribute and the definition of task, for which typing is made. Definition of personality type is more appropriate for our example. Combining the simple statements into classes is the third level of typing. The researcher himself formulates the name for each class due to the nature of statements combined.

In practice, the results of grouping, obtained by different researchers, are very similar. Differences may occur because certain statements may indeed occupy an intermediate state and can be attributed to several classes. Differences in the grouping of statements may be due to the differences between the criteria, which are used by various researches to group the statements. Each researcher may give absolutely different names of classes. The best thing is to use laconic names.

Therefore, the following result of processing of open question data will be obtained (at the output):

- Three new presentation of the attribute (property), which are included into original data table and may be subjected to further processing for obtaining meaningful conclusions;
- The table “list of attribute values”, which may be used in case of resurveying or in order to identify data typing of other questionnaires meant for investigation of this process.

It should be noted, that the composite attributes would be also formed as a result of composite attributes typing. Special processing technique have been developed to analyze them [7].

The technology of open question processing has another one important result, which allows reducing significantly the time for data typing in resurveying (the process monitoring). When updating the original data table the typing procedure shall be repeated subject to new data input. To speed the work, the researcher can use two types of dictionaries, such as “Dictionary of substitutions” and “Dictionary of key words”, which are created for each attribute containing data on the open question. Such dictionaries are formed for each qualitative attribute. In addition, another one dictionary is used in data processing, which deals with various qualitative attributes and even with various questionnaires. This is a “Dictionary of redundant information”. It is used during the first stage of processing of qualitative textual data. This dictionary helps to remove or correct statements containing various redundant and insignificant information.

All dictionaries are stored in one file Access. Dictionaries store experience, gained by the researcher when working on typing of statements and represent the knowledge base.

Main Part: One of the simplest example of using the technology is typing of names of recreation areas visited by the citizens of the region in the summer, for instance, of Primorsky Krai. A questionnaire “Swimming season research” has included two open questions as follows:

- Most visited beaches: (name of the bay, the islands or the nearest town)
- Most visited recreation area: (name of the nearest locality, bay or island)

The first question relates to a one-day rest and the second-to an overnight rest. When answering questions, respondents can specify several beaches or recreation areas they like to visit. The list of possible places visited during the rest is very large and a complete list can not be made by the researcher in general. Therefore, only an open question can be used to determine structures of spatial distribution of recreation areas. Despite the illusion of simplicity of the task, it is quite difficult in

practice. The practice shows that there may be lot of different ways of writing names with the errors frequently repeated.

As a criterion for grouping the recreation areas, the affiliation to any administrative region may be used, for instance. After processing poorly structured data the information takes the structured form. In this case no loss of information occurs. The original information becomes arranged i.e. structured.

We shall consider a more complex situation. For example, our main interest in questionnaire “Swimming season research” was to find out the consumer preferences in organizing their pastime at the beach zone of recreation area. People, having a rest at the seaside, should do anything else apart from sunbathing and swimming (especially when staying at the seaside for several days). To study the pastime of vacationers a questionnaire has included the question "What else do you like to do going to the seaside, except for sun bathing and swimming:”.

As it turned out, the spectrum of interests of vacationers is not too wide. After typing, these interests have been combined into eight groups:

- Sportsmen – 26%
- Inactive persons – 19%
- Enthusiasts – 16%
- Gourmands – 13%
- Lyric persons – 12%
- Extroverts – 8%
- Sleepyheads – 5%
- Moms – 1%

The same questionnaire studied the negative statements of the respondents concerning their pastime at the seaside. To analyze the negative opinions a questionnaire has included the question “What did cloud your pastime at the beach zone: “. As a result of typing a quite stable structure of the distribution of consumers’ negative reactions was determined. After grouping all negative statements nine groups of recreants were identified:

- Green – 53%
- Choosy – 12%
- Optimists – 10%
- Unsociable – 8%
- Urbanists – 6%
- Intellectuals – 5%
- Intolerant – 2%

- Motorists – 2%
- Students – 1%

Computer typing technique was also used in analyzing a number of characteristics of pastime during holidays and vacations of students.

A questionnaire survey “Study of the tourist potential of Primorsky Krai and the prospects for its development” used the most complex form of open questions. This is a specialized questionnaire, based on questions involving the answers in the form of a few sentences. An example of such question is “What kind of requirements shall a municipal beach meet to make you visit it more often?”. Even using such a complex form of question the typing technique helped to identify quite certain groups of different opinions of residents of Primorsky Krai. Frequency series of distributing the opinions due to data grouped is shown in Fig. 1. The chart in Fig. 1 shows that most of respondents speak about the need to improve sanitary conditions of the beaches (32%) as the basic condition for their visiting.

- *Opportunity to play sports*
- *Cultural behavior*
- *Accommodation*
- *Provision of family recreation*
- *Service cost reduction*
- *Entertainments*
- *Food and drinks*
- *Rare demands and wishes*
- *Safety*
- *Transport and accessibility*
- *Special demands*
- *Improvement of beach maintenance*
- *Requirements for development*
- *Requirements for clean environment*

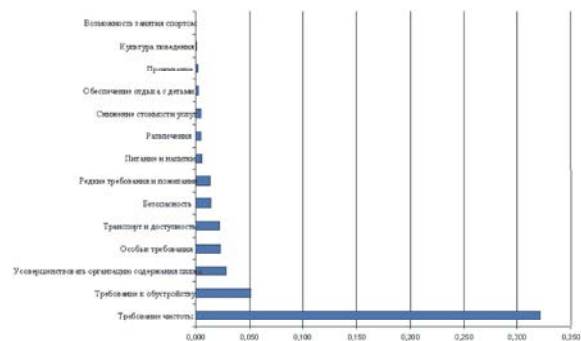


Fig 1: Grouping of consumers’ statements concerning the improvement of the condition of beach zones.

Table 1: Classes intersection matrix

Typing	Sportsmen	Enthusiasts	Sleepyheads	Gourmands	Lyric persons	Inactive persons	Extroverts	Moms
Sportsmen	0.708	0.089	0.013	0.087	0.088	0.003	0.069	0.003
Enthusiasts	0.141	0.714	0.014	0.070	0.082	0.000	0.039	0.005
Sleepyheads	0.067	0.044	0.698	0.149	0.102	0.003	0.048	0.010
Gourmands	0.177	0.090	0.060	0.608	0.069	0.005	0.110	0.005
Lyric persons	0.178	0.105	0.041	0.069	0.631	0.003	0.060	0.006
Inactive persons	0.004	0.000	0.001	0.003	0.002	0.993	0.001	0.000
Extroverts	0.210	0.075	0.029	0.165	0.090	0.002	0.563	0.012
Moms	0.079	0.079	0.048	0.063	0.079	0.000	0.095	0.667

These are quite reasonable numerical estimates, which shall be considered by both regional governments and commercial structures servicing the tourists and vacationers, in order to organize customer service due to evolving needs of tourists in different types of leisure, recreation, cognitive and other activities that are the basis and incentive for business push to the research and development of new regional products and types of tourism.

We have considered three types of open questions. The first case supposes the answer in the form of one or more words, the second case-one or more simple phrases and the third case supposes composite sentences.

Analysis of structures of spatial development of the tourist complex in the region on the basis of typologies. Any recreational resources assume their reference to a specific place or determination of their spatial location [8]. To predict and regulate the processes of consuming the tourist resources, such processes shall be described by data structured. A number of important structural characteristics are obtained in the process of typologies development.

Characteristics of consumption structure start interacting and become time-dependent. The aim of strategic management is to steer these changes in the right direction. In addition, the technique of structural change analysis shall be developed for this purpose.

It turns out that objects in their pure form cannot always be described by one structural feature. For example, some of respondents may be referred to several classes according to their statements. For instance, a respondent, being asked the question “What did cloud your pastime at the beach zone: “, may give the following answer: environmental conditions, crowding, unsettled beach etc. Single simple statements in typing an attribute have been referred to different classes. In this case, the original values have been substituted with names of classes such as "Green", “Unsociable”, “Urbanists”. Therefore, the classes may intersect along the whole set of statements. However, it

appears that some classes are more closer, some are less closer and some are totally isolated. We shall consider the principle of calculating the estimates of classes intersection.

Each composite answer, prescribed in the form of classes, has a combination of class pairs allocated. For example, if a single answer has a form of “S, I, I, S, M” (where letters denote classes), the respondent may be referred to several classes due to his/her statements. The following three pairs of classes may be formulated according to the answer given, such as “I, S”, “S, M”, “I, M”. Therefore, possible pairs of classes intersections can formulate an intersection matrix. Matrix size is $\kappa \times \kappa$, where κ is a number of classes allocated. Matrix is symmetric in the diagonal. Elements of the matrix represent the sum of pairs of classes encountered in the whole sample. To eliminate the influence of sample and classes size on the estimation of classes the matrix elements are normalized by dividing the rows into number of pairs of statements according to classes. Diagonal elements characterize the degree of isolation of the individual classes. For example, the matrix, shown in table 1, was calculated according to data grouped by preferences concerning pastime at the beach zone.

The links of classes are given in the form of graph (Fig. 2) for clarity of classes intersections.

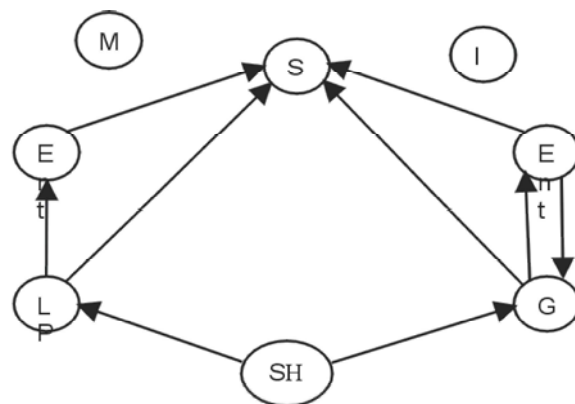


Fig 2: Classes intersection graph

A certain threshold value is set to intersection assessment when constructing a graph. In this case, the graph will show only the most essential links. The graph in Fig. 2. has a threshold value of 1.5. Classes at the graph vertices are marked with first letters of their names. The graph gives an indication of how to combine the range of services offered, i.e. to design a regional tourist product for different regional consumers.

CONCLUSION

Primorsky Krai of Russia is one of the most promising regions for tourism development [9]. In an increasingly competitive external and internal environment, the tourist complex development program shall be developed on the basis of modern scientific technique, which require new methods of collecting and analyzing data on consumers' preferences [10]. We have developed several methods of data analysis that showed high efficiency and allowed acquiring new knowledge quantified.

Summary:

- This paper deals with technique for handling unstructured data, which significantly extends the capabilities of researchers in the study of socio-economic phenomena and processes.
- This technique includes an expert system based on the use of three types of computer dictionaries. These dictionaries will help the researchers, involved in the same field, to share their experience in data systematization.
- Presentation of information in the form of composite attributes allows analyzing the interaction of different structural groups of consumers.
- The proposed data processing technique has passed evaluation test in processing large volumes of real data and can be recommended for use by a wide range of researchers dealing with questionnaire surveys in their practice.

REFERENCES

1. Martysenko, N.S., 2009. Data acquisition and analysis technique for evaluating the structure of consumers of tourist services of the region // *Practical marketing*, 11: 16-28.
2. Lewins, A., C. Taylor and G.R. Gibbs, 2014. What is Qualitative Data Analysis, 2005, [Electronic resource]. Access mode: http://onlineqda.hud.ac.uk/Intro_QDA/what_is_qda.php (accessed date: 25.01.2014).
3. Davydov, A.A., Qualitative researches: prospects for development. M.: IS RAS, 2008. [Electronic resource]. Access mode: http://www.isras.ru/index.php?page_id=922 (accessed date: 25.01.2014).
4. Kanygin, G.V., 2006. Constructing the constructivism // *Sociological research*, 11: 19-28.
5. Explore. Discover. Share http://www.qsrinternational.com/#tab_you (accessed date: 25.01.2014).
6. Martysenko, N.S. and E.A. Egorov, 2011. Information technology for increasing qualitative information processing efficiency // *Journal of Modern Applied Statistical Methods*, 10(1): 207-213.
7. Martysenko, N.S., 2010. Study of structure of tourist product consumption in Primorsky Krai // *Regional economics: theory and practice*, 26: 60-68.
8. Alexandrova, A.Yu., 2008. Spatial structure of the global tourism market // *Bulletin of the Tver State University. Series: Geography and Geoecology*, 4: 83-102.
9. Baklanov, P.Ya. and M.T. Romanov, 2013. Major factors and the ways of tourism development in the Far East region of Russia // *Bulletin of the National Academy of Tourism*, 1(25): 37-43.
10. Martysenko, N.S., 2008. Formation of tourism and recreation cluster as a factor of social and economic development of Primorsky Krai // *Regional issues*, 10: 95-101.