Models and Methods for Quality Management Based on Artificial Intelligence Applications

Nafissa Yussupova[1], George Kovács[2], Maxim Boyko[1], Diana Bogdanova[1]

[1] Faculty of Informatics and Robotics, Ufa State Aviation Technical University, 12 K. Marx str., 45000 Ufa, Russian Federation, e-mail: yussupova@ugatu.ac.ru, maxim.boyko@ugatu.ru, diana.bogdanova@ugatu.ru

[2] Computer and Automation Research Institute, Hungarian Academy of Sences, Kende u. 13-17, 1111 Budapest, Hungary, e-mail: kovacs.gyorgy@sztaki.mta.hu

Abstract: This paper proposes a conceptual approach to the research into customer satisfaction based on a detailed analysis of consumer reviews written in natural languages using Artificial Intelligence (AI) techniques such as Text Mining, Aspect Sentiment Analysis, Data Mining and Machine Learning. Special Internet resources for accumulating customer reviews, such as yelp.com, tripadviser.com and tophotels.ru, are used as data sources. To evaluate the efficacy of the proposed approach, we have carried out an experiment on qualitative and quantitative research of hotel client satisfaction. Even "Big data" applications were taken into account as a possibility to evaluate quality of services. The obtained results prove the effectiveness of the proposed approach to decision support in product quality management and argues for using it instead of classical methods of qualitative and quantitative research into customer satisfaction.

Keywords: quality management services; analysis of customer feedback; CPM; sentiment analysis

1 Introduction

Quality assurance is currently realized by means of a process approach based on the model of a quality management system [1]. It describes the interaction of the company and the customer during the process of product production and consumption. To correct the parameters of product quality in order to improve it for the customer, the model includes feedback. For companies, one aspect of feedback during the process of quality management is information about the level of customer satisfaction, expressed in the form of customer reviews of the product quality. That is why customer satisfaction is the key information in quality management that influences decision-making.

To collect data and to evaluate customer satisfaction, the International Quality Standard ISO 10004 recommends using the following methods: personal interviews, phone interviews, discussion groups, mail surveys (postal questionnaires), online research and survey (questionnaire survey) [2]. However, these methods of collecting and analyzing customer opinions show a number of significant drawbacks.

A general drawback of the recommended methods is the need for a large amount of manual work: preparing questions, creating a respondent database, mailing questionnaires and collecting results, conducting personal interviews, preparing a report based on the results. All this increases the research costs. Due to their discreteness these methods do not allow for the continuous monitoring of customer satisfaction. For this reason, the data analysis is limited to one time period and does not give an insight into the trends and dynamics of customer satisfaction. This also has a negative influence on the speed of managerial decision making, which depends on the arrival rate of up-to-date information about customer opinions.

Existing scales of customer satisfaction and their subjectivity perception raise additional questions. Values of customer satisfaction expressed in the form of abstract satisfaction indices make it difficult to understand, compare and interpret the results. Methods of analysis of data collected through the recommended ISO 10004 procedures permit only the detection of linear dependencies.

In this paper, to increase the effectiveness of product quality management, we suggest approaching the research of customer satisfaction through the use of AI technologies.

1.1 Should We Use Big Data for Analysis?

Recently several works deal with the problems and applications of “Big data”, as for example [3, 4 and 5]. Most often, the problem is associated with the necessity of processing of structured and unstructured data of large volumes. The term “Big data” appeared lately, in 2008 [6]. However, already in 2001 specialists in artificial intelligence faced the Big data problem, when a program to create "Intelligent Image" ended in failure. That time "Big data" was not identified as a separate problem, and has been regarded as a temporary difficulty. In general, the problem and the fact that the emergence of big data can be correlated, became clear only, when it was investigated from the approach of analytical data processing.

Big data is a series of approaches, tools and methods for handling structured and unstructured data of huge volume and significant diversity for results perceived by humans, effective in the conditions of continuous growth. The distribution of information across multiple nodes of computing networks in the pst 15 years gave the alternative to traditional control systems databases and decisions as a class Business Intelligence.

According to one of the approaches to this issue, the concept of "Big data" refers to the operations that can be performed only on a large scale.

Based on the realities of the world the use of "Big data" technology is becoming more and more important and commonly used. However, it is hard to define whether big data technology is really necessary and helpful for a given problem or not. The implementation of Big Data is guided by the concept of the four Vs: Volume, Variety, Velocity and Value [7]. To facilitate decision-making on implementing the technology "Big Data" in our case, we have estimated an “indicator of readiness” of transition to new technologies to work with a large volume of data - an indicator of readiness for Big Data is called Bigd. We applied the method, which is detailed in [8]. If the value of Bigd is more than 50%, the “t” technology of Big Data should be implemented. The Parameter «Volume» shows the size of the accumulated data, parameter «Velocity» is calculated from two values: the first describes the capture and processing of data in near-real-time (obtaining data by high-speed streaming); the second - is the rate of accumulation in the organization of data to be analyzed. If we estimate the increase of generated data to 60% or more per year, it results in unsolvable problems for companies with no appropriate IT infrastructure. The existing IT infrastructures would be soon exhousted and a necessary upgrade would cost much more than the benefits it would provide.

Parameter "Variety" is defined as follows: "the data is collected from one or more sources, and possibly in different formats." This parameter is determined by experts as the aggregate value of the quantitative sources. Parameter «Value» is determined by experts and is in the range from 0 to 1, and shows the value of the information from the source of data.

Our analysis showed that due to the features of our problems Big Data technologies cannot be effectively used to evaluate the quality of services. Therefore, to support decision making in the quality of services, management has proposed a method for evaluating the tone of reviews based on artificial intelligence technologies.

2 The Approach to Quality Management

Figure 1 represents the algorithm of the suggested approach to quality management based on research into customer satisfaction using AI applications. It consists of four main stages: 1. collection of reviews from Internet resources, data cleansing and loading data into the database. The second stage comprises the processing and analysis of the collected reviews. It includes marking reviews by their emotional response, i.e., sentiment (for example, negative and positive), identifying product aspects, and evaluating the sentiment of the comments on the separate aspects. Following the stage of data processing utilizing visualization tools, quantitative research is carried out. A qualitative research of customer satisfaction is undertaken by means of building models based on decision trees, where the review's sentiment serves as a dependent variable, and sentiment comments on separate product aspects as independent variables. Managerial decision development and making is carried out on the basis of this research.

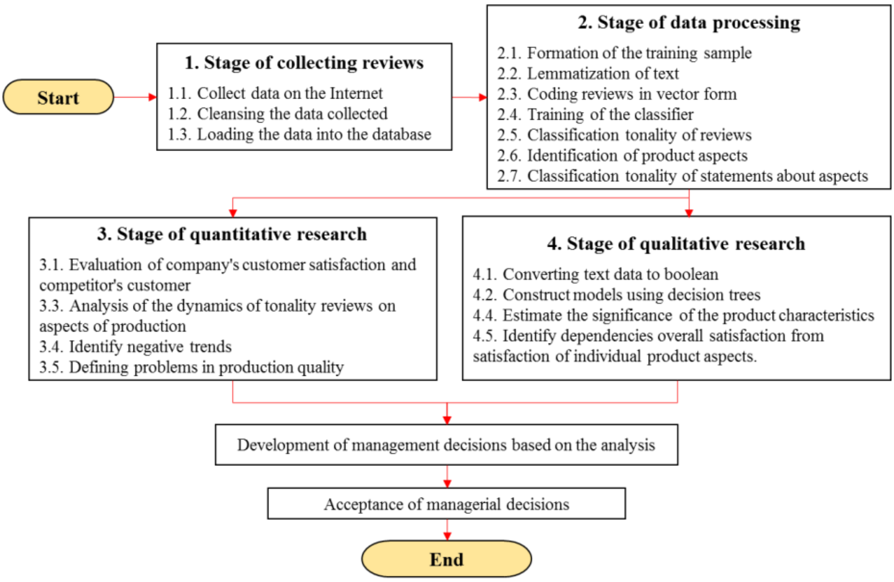

Figure 1

Quality management based on research of customer satisfaction using AI applications

3 Applied Artificial Intelligence Techniques

3.1 Data Collection

Nowadays there are a large number of Internet resources where users can leave their opinions about goods and services. The most popular examples are tophotels.ru (635,000 reviews), yelp.com (53 million reviews), tripadvisor.com (travels, 130 million reviews). Similar resources continue to gain popularity. Their advantage as a source of information for satisfaction evaluation lies in their purpose – the accumulation of customer reviews. As opposed to social network services, the web pages of review sources use XML that determines the structure typical for a review. Such a structure includes separate blocks with the name of a product or company and a review, and other blocks with additional information. Therefore, all reviews are clearly identified in relation to the review object. It significantly simplifies the process of data collection and excludes the problem of key word ambiguity. One further advantage is that many of such resources monitor the reviews and check the objectivity of the authors.

There are two main types of collecting Internet data on customer reviews: 1) by using API (application programming interface) and 2) by web parsing. API is a set of ready-to-use tools – classes, procedures, functions – provided by the application (Internet resource) for use in an external software product. Unfortunately, only few resources that accumulate reviews have their own API. In this case, to collect reviews we can use the second method for data collection – web parsing. Web parsing is a process of automated analysis and content collection from xml-pages of any Internet resource using special programs or script.

3.2 Sentiment Analysis

After the data has been collected and cleaned, we can start their processing with the help of Text Mining tools. Sentiment Analysis is used to evaluate the author's product satisfaction. Sentiment stands for the emotional evaluation of an author's opinion in respect to the object that is referred to in the text.

We can distinguish three main approaches to Sentiment Analysis: 1) linguistic, 2) statistical, and 3) combined. The linguistic approach is based on using rules and sentiment vocabulary [9, 10 ,11]. It is quite time-consuming due to the need of compiling sentiment vocabularies, patterns and making rules for identifying sentiments. But the main drawback of the approach is the impossibility of obtaining a quantitative evaluation of the sentiment. The statistical approach is based on the methods of supervised and non-supervised machine learning. The combined approach refers to a combined use of the first two approaches.

The present work uses the methods of supervised machine learning: Bayesian classification and Support Vector Machines. Software implementation is simple, and does not require generating linguistic analyzers or sentiment vocabularies. Text sentiment evaluation can be expressed quantitatively. To apply these methods, a training sample was created. To describe an attribute space, vector representation of review texts was used with the help of the bag-of-words model. Bit vectors - presence or absence of the word in the review text, and frequency vectors – the number of times that a given word appears in the text of the review, served as attributes. Lemmatization, a procedure of reducing all the words of the review to their basic forms, was also used. More detailed information about machine learning methods used in this paper can be found in articles [15, 16].

3.3 Aspect Sentiment Analysis

Sentiment Analysis of reviews allows to evaluate general customer product or company satisfaction. However, it does not make clear what exactly the author of the review likes and what not. To answer this question, it is necessary to perform an Aspect Sentiment Analysis. An aspect means characteristics, attributes, qualities, properties that characterize the product, for example, a phone battery or delivery period, etc. However, one sentiment object can have a great number of aspects. Furthermore, aspects in the text can be expressed by synonyms (battery and accumulator). In such cases it is useful to combine aspects into aspect groups. An example of such aspect groups is represented in Figure 3.

An Aspect Sentiment Analysis of a review is a more difficult task and consists of two stages – identifying aspects and determining the sentiment of the comment on them. To complete the task of the Aspect Sentiment Analysis, a simple and effective algorithm has been developed:

First stage.

1. *Extract all nouns S from the set of reviews D.*
2. Count the frequency of words $\forall i = \overline{1,|S|} : f_i = N_i / N$ in the whole set of reviews D, where N is the number of appearances of all words, N_i the number of appearances of the i noun.
3. Count the difference $\forall i : \Delta_i = f_i - f_i^V$ between the counted frequencies f_i and *vocabulary* frequencies f_i^V.
4. Sort the set of nouns S in descending order Δ_i.
5. Divide the set of nouns *S* from $\Delta_i > 0$ into aspect groups.

Second stage.

1. Divide a set of reviews into sets of sentences.
2. Perform classification of sentiment for each sentence.
3. Check each sentence for the condition: if a sentence has a negative or positive sentiment and contains at least one noun from any aspect group, then the given sentence is labeled as an opinion about the given aspect.

A frequency vocabulary (based on the corpus) that helps to compare the obtained frequencies with word frequencies is used to identify aspects. The nouns with maximum frequency deviations are candidates for inclusion into aspect groups. Division of the noun set into aspect groups was carried out manually. We should note that if a sentence includes nouns from several aspect groups, then it will appear in each of them.

The results of Sentiment Analysis and Aspect Sentiment Analysis can be represented in the form of text variables $Obj = (\mathrm{Re}\,v_i, Sent_i, Neg_i^1, ..., Neg_i^j, Pos_i^1, ..., Pos_i^j)$, where Obj is a sentiment object or a product, $\mathrm{Re}\,v_i$ the text of the *i* review, $Date_i$ the date of *i* review publication, $Sent_i$ the sentiment of *i* review, Neg_i^j the negative sentences about the *j* aspect in the *i* review, Pos_i^j the positive sentences about the *j* aspect in the *i* review, *i* the review number, *j* the aspect group number.

3.4 Decision Trees

The following paragraph focuses on an algorithm of the processing of data obtained with help of Sentiment Analysis and Aspect Sentiment Analysis. The task of the developed algorithm is the mining of data that can be used for decision support in product quality management. To realize this algorithm, we use the intelligent data analysis tool, i.e. the decision tree since this tool can be easily understood and its results can be clearly interpreted; it also can explain situations by means of Boolean logic.

The algorithm of processing of data obtained by means of Sentiment Analysis consists of the following procedures:

1. Convert a set of text data $Obj = (\mathrm{Re}\,v_i, Sent_i, Neg_i^1, ..., Neg_i^j, Pos_i^1, ..., Pos_i^j)$ into a Boolean type by the following rules:

1.1. If $Sent_i = negative$, *then* $blSent_i = 1$ *else* $blSent_i = 0$.

1.2. If $Neg_i^j \neq null$, *then* $blNeg_i^j = 1$ *else* $blNeg_i^j = 0$.

1.3. If $Pos_i^j \neq null$, *then* $blPos_i^j = 1$ *else* $blPos_i^j = 0$.

2. Creating a decision tree where the variable $blSent_i$ is a dependent variable from $(Neg_i^1, ..., Neg_i^j, Pos_i^1, ..., Pos_i^j)$.

3. Estimation of the significance of aspect groups and interpretation of results.

The algorithm we have described allows us to understand which sentiment comments on product aspects influence the whole text sentiment or, in other words, what product aspects influence customer satisfaction and in what way. Our decision tree model allows us to consider the influence not only of separate sentiment comments on aspects but also of their mutual presence (or absence) in the text on customer satisfaction. The decision tree model also enables us to detect the most significant product aspects that are essential for the customer. The logical constructions (called rules) that we have obtained can be expressed both in the form of Boolean functions in a disjunctive normal form and in natural language.

A decision tree model can help to predict sentiment in dependence on various inputting aspect comments of different sentiments. In fact, it makes it possible to evaluate experimentally customer satisfaction in dependence on satisfaction with different product attributes. As the final result, prediction and analysis of the influence of different inputting variants on customer satisfaction allows us to distribute the company's budget effectively to maintain a high product quality.

The significance of aspects group shows how much the sentiment of a review depends on the sentiment of the aspect group. If the number of aspect groups is $g/2$, then the number of independent variables is g (positive and negative statements in each group of aspect). The formula for calculating the significance of m variable is:

$$Sign_m = \frac{\sum_{j=1}^{k_m}\left(E_{m,j} - \sum_{i=1}^{n_{m,j}} E_{m,j,i}\cdot\frac{N_{m,j,i}}{N_{m,j}}\right)}{\sum_{l=1}^{g}\sum_{j=1}^{k_l}\left(E_{l,j} - \sum_{i=1}^{n_{l,j}} E_{l,j,i}\cdot\frac{N_{l,j,i}}{N_{l,j}}\right)}\cdot 100\%, \quad (1)$$

where k_l is the number of nodes that were split by attribute l, $E_{l,j}$ is the entropy of the parent node, split by attribute l, $E_{l,j,i}$ is the subsite node for j, which was split by attribute l, $N_{l,j}$, $N_{l,j,i}$ are the number of examples in the corresponding nodes, $n_{l,j}$ is the number of child nodes for j parent node.

The score of customer satisfaction S with products is calculated by the formula:

$$S = \frac{N^{pos}}{N^{pos} + N^{neg}}\cdot 100\%, \quad (2)$$

where N^{pos} is the number of positive reviews, N^{neg} the number of negative reviews.

The score of customer satisfaction S_j with j aspect group of products is calculated by the formula:

$$S_j = \frac{n_j^{pos}}{n_j^{pos} + n_j^{neg}}\cdot 100\%, \quad (3)$$

where n_j^{pos} is the number of positive comments containing mention of the j aspect group, n_j^{neg} the number of negative comments containing mention of the j aspect group.

4 Experiments

Effectiveness evaluation of the developed approach was performed on the data obtained from 635,824 reviews of hotels and resorts in Russian. The reviews were collected from a popular Internet resource for the period of 2003-2013. The initial structure of the collected data consisted of the following fields: hotel name; country name; resort name; date of visit; opinion of the hotel; author evaluation of food; author evaluation of service; review number. The data was preliminarily processed and loaded into the database SQL Server 2012.

To classify segments, we used a binary scale (negative and positive) on the hypothesis that the absence of negative is positive. A training sample of positive and negative opinions was created using the collected data on the author's evaluation of accommodation, food and service. The Internet resource tophotels.ru uses a five-point grading scale. A review can have a maximum total of 15 points, a minimum of 3 points. The training sample included 15,790 negative reviews that had awarded 3 and 4 points, and 15,790 positive reviews that had awarded 15 points. We did not use author evaluation for further data processing. The marking of the remaining 604,244 reviews was carried out using a trained classifier.

For the purpose of effectively creating a sentiment classifier, we evaluated the accuracy of the classification of machine learning algorithms and some peculiarities of their structure (Table 1). The criterion Accuracy (ratio of the number of correctly classified examples to their total number) was used to assess classification accuracy. Accuracy evaluation was performed on two sets of data. The first set (Test No. 1) represented a training sample consisting of strong positive and strong negative opinions. It was tested by cross validation by dividing the data into 10 parts. The second set (Test No. 2) included reviews covering different points and was marked manually (497 positive and 126 negative reviews). It was used only for the accuracy control of the classifier that had been trained on the first data set.

To assess the influence of the negative particles "not" and "no", we used tagging; for example, the phrase "not good" was marked as "not_good", and was regarded by the classifier as one word. This technique allowed for the increase of sentiment classification accuracy.

Table 1

Comparison of methods for sentiment classification

Machine learning methods	**Vector**	**Test No. 1**	**Test No. 2**
SVM (linear kernel)	Frequency	94.2%	83.1%
SVM (linear kernel)	Binary	95.7%	84.1%
NB	Binary	96.1%	83.7%
NB	Frequency	97.6%	92.6%
NB (exceptional words)	Frequency	97.7%	92.7%

Bagging NB	Frequency	97.6%	92.8%
NB (tagging “not” and “no”)	Frequency	98.1%	93.6%

For the marking of reviews and the Sentiment Analysis, we created a classifier on the basis of the NB method, with frequency vectors as attribute space, and with the use of lemmatization and tagging of the negative particles “not” and ‘no”.

Using the algorithm we had developed we extracted from all reviews the key words that were divided into seven basic aspect groups (a part is represented in Figure 2): *beach/swimming pool, food, entertainment, place, room, service, transport.* The following step was extracting and marking sentences with words from aspect groups by sentiment. However, not all sentences with aspects have a clearly expressed sentiment; therefore, the sentences which do not show a clearly expressed sentiment were filtered out.

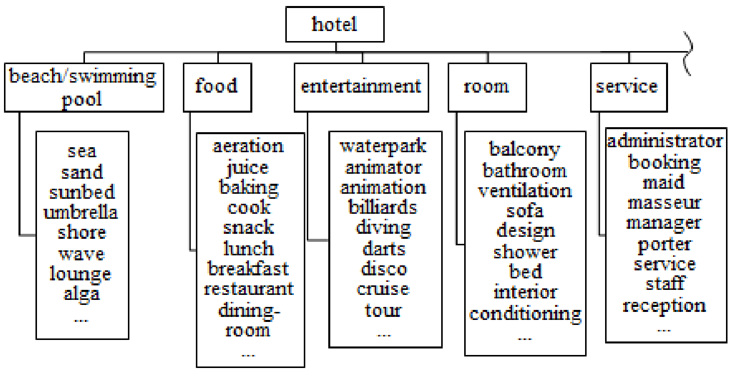

Figure 2
A part of the aspect groups of the object “hotel”

We will give an example of our qualitative and quantitative research for two 5 star hotels “A” (1,692 reviews) and “B” (1,300 reviews) located in the resort Sharm el-Sheikh (63,472 reviews) in Egypt. First, we will describe our quantitative research of consumer satisfaction dynamics, then we will compare this with the average satisfaction in the whole resort, detect negative trends in the different hotel aspects and identify problems in the quality of hotel services.

The dynamics of customer satisfaction is represented in Figure 3. Concerning Hotel “A”, there is a positive upward satisfaction trend beginning in 2009; it reaches the average resort level in 2013. Concerning Hotel “B”, in 2012 there was a sharp satisfaction decline and a similarly sharp increase in 2013. We can also notice this trend in a monthly schedule (Figure 4). Satisfaction decrease for Hotel “B” started in June 2012 and stopped in October 2012. Then, customer satisfaction with Hotel “B” grew to a level that was higher than the average resort level, being ahead of its competitor – Hotel “A”.

To find reasons for the Hotel “B” satisfaction decrease, we will examine the diagrams in Figure 5. We can see that in 2012, Hotel “B” on average was second to Hotel “A” in such aspects as “Room” (Δ12%), “Place” (Δ8%), “Service” (Δ5%), “Beach/swimming pool” (Δ3%) and “Entertainment” (Δ3%). Besides, in 2012, Hotel “B” had more registered cases of food poisoning as well as cases of theft in August 2012. We should also note that one of the reasons of client dissatisfaction with Hotel “B” as a place was the beginning of the renovation of the hotel building and the rooms. These measures, however, were rewarded in 2013, when customer satisfaction with Hotel “A” aspects equaled the average resort level.

In 2013, customer satisfaction with Hotel “B” exceeded the average level in all aspects (Figure 6). Customer satisfaction with Hotel “A” dropped to lower than average values in such aspects as “Service” (Δ3%), “Food” (Δ3%), “Beach/swimming pool” (Δ3%) and “Transport” (Δ4%). The manager of Hotel “A” could be advised to direct efforts to increase the quality of all aspects, but would this be the most effective solution? Which aspects are the most significant for the customer and should consequently be improved in the first place? Is it possible to offset the dissatisfaction with the service, for example, by healthier food or an animated evening performance and achieve client satisfaction? A qualitative research of the Sentiment Analysis results can give answers to these questions.

Figure 3

Dynamics of customer satisfaction by year

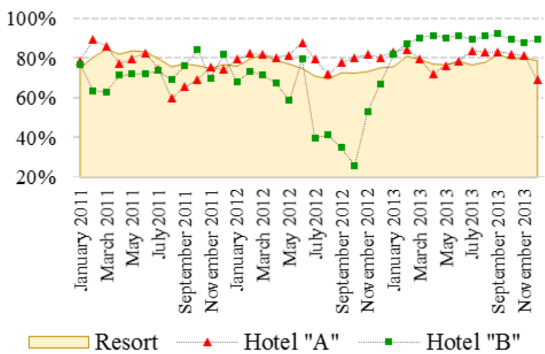

Figure 4

Dynamics of customer satisfaction by month

Decision trees were created using algorithm C4.5. The first step was creating a tree for the total sample of the reviews on the given resort to detect general principles. Extracted rules that have a reliability >80% are represented in Table 3. The second step is developing decision trees for the sample of Hotel "A" and Hotel "B" reviews to identify principles on the hotel level. Aspect significance is represented in Table 3.

Analyzing values of aspect significance (Table 6), we can say that the main factors of consumer dissatisfaction are a low service level, problems with food, and complaints about the hotel rooms. The most critical aspect for Hotel "B" is "Room". Without negative opinions on the aspect "Room", the reviews would be positive with a probability of 95.5% (Rule No. 10, Table 2). That is why the performed repair work facilitated a significant increase of consumer satisfaction. The most critical aspect for Hotel "A" is "Service", which corresponds with the findings for the resort as a whole.

The aspects which are significant both for the resort and for the two hotels and contributing to customer satisfaction are good food and amusing entertainment activities. The combination of these aspects can counterbalance negative emotions from the service or complaints concerning hotel rooms and leave the client with a favorable impression of the time spent in the hotel (Rules No. 5, 7, 11 Table 2). We should note that positive opinions about service, beach/swimming pool or place do not have a powerful influence on sentiment. That means the consumer priori focuses on a high-level of service, a well-kept place and the beach/swimming pool as a matter of course.

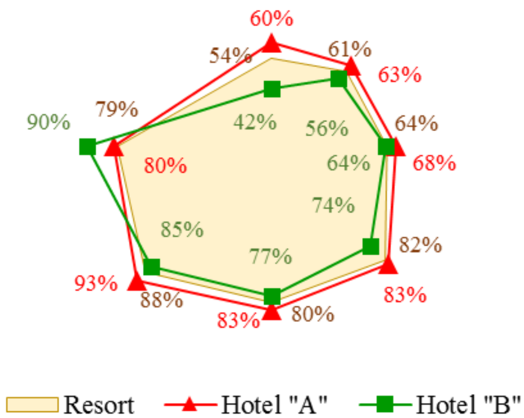

Figure 5
Comparison of customer satisfaction by aspects in 2012

The qualitative research we have undertaken enabled us to detect the main ways for Hotel "A" to increase customer satisfaction (Table 4). The problematic aspects identified in the course of our quantitative research correspond to the most significant aspects detected during the qualitative research stage. A search for alternative aspects that can lead to customer satisfaction in the presence of negative opinions about the significant aspects “Service” and “Food” was carried out. To accomplish this, the rules (Table 3) containing negative sentiment in problem aspects, but which eventually lead to a positive review, were filtered out by the decision tree. The rules obtained and examples of appropriate managerial decisions are represented in Table 4.

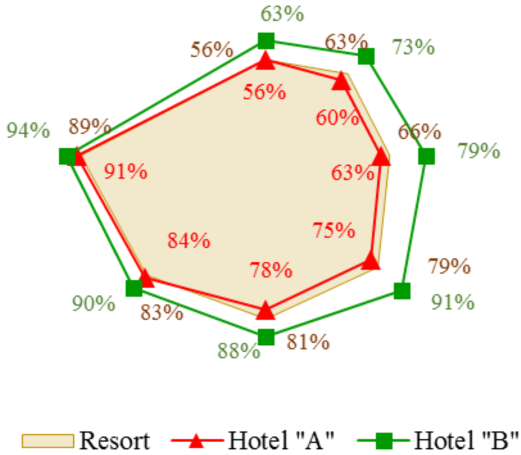

Figure 6
Comparison of customer satisfaction by aspects in 2013

Table 2
Significance of aspects

Aspect	Resort	Hotel “A”	Hotel “B”
Service⁻	34.8%	60.2%	-
Food⁺	30.3%	27.2%	30.3%
Food⁻	16%	-	-
Entertainment⁺	8.5%	12.7%	12.4%
Room⁻	4%	-	57.3%
Beach⁺	2.5%	-	-
Room⁺	2.1%	-	-
Territory⁺	1%	-	-
Service⁺	0.7%	-	-
Beach⁻	0.2%	-	-
Transport⁺	-	-	-
Theft	-	-	-
Food poisoning	-	-	-
Entertainment⁻	-	-	-
Territory⁻	-	-	-
Transport⁻	-	-	-

Table 3

Rules extracted by Decision Trees

№	Rules	Sentiment	Support	Reliability
Extracted rules on Resort reviews sample *				
1	$Food^{+} \cap \overline{Service^{-}} \cap \overline{Food^{-}}$	Positive	37-2%	97.4%
2	$Food^{+} \cap \overline{Service^{-}} \cap \overline{Food^{-}} \cap Beach^{+}$	Positive	11%	86.2%
3	$\overline{Food^{+}} \cap \overline{Service^{-}} \cap \overline{Service^{-}} \cap \overline{Room^{-}}$	Positive	10-6%	83.9%
4	$\overline{Food^{+}} \cap Service^{-} \cap \overline{Entertainment^{+}}$	Negative	6.9%	92.3%
5	$Food^{+} \cap Service^{-} \cap \overline{Food^{-}} \cap Entretainment^{+}$	Positive	5.8%	88.4%
6	$\overline{Service^{-}}$	Positive	62.9%	88.3%
7	$Food^{+} \cap Service^{-} \cap Entertainment^{+}$	Positive	20.5%	74.1%
8	$\overline{Food^{+}} \cap Service$	Negative	9.4%	86.2%
9	$Food^{+} \cap Service^{-} \cap \overline{Entertainment^{+}}$	Negative	7.2%	65.6%
* Accuracy of the created model by training sample 83.6% , by control sample 83.4%				
10	$\overline{Room^{-}}$	Positive	51.2%	95-5%
11	$Food^{+} \cap Room^{-} \cap Entertainment^{+}$	Positive	27.9%	81%
12	$\overline{Food^{+}} \cap Room^{-}$	Negative	11.1%	84%
13	$Food^{+} \cap Room^{-} \cap \overline{Entertainment^{+}}$	Negative	9.9%	55.8%

Table 4

Application of the results for the development of management solutions for Hotel "A"

Problematic Aspect Group	Rule level	Rules with result "Positive"	Support	Relia-bility	Examples of recommended managerial decisions
1. $Service^{-}$	Hotel	No.6: $\overline{Service^{-}}$	62.9%	88.3%	Educate and motivate service staff; check food service quality; organize entertainment activities.
		No.7: $Food^{+} \cap Service^{-} \cap Entertainment^{+}$	20.5%	74.1%	
	Resort	No.5: $Food^{+} \cap Service^{-} \cap \overline{Food^{-}} \cap Entertainment^{+}$	5.8%	88.4%	
2. $Food^{-}$	Hotel	-	-	-	Diversify menu; organize garbage collection on the beach.
	Resort	No.2: $Food^{+} \cap \overline{Service^{-}} \cap Food^{-} \cap Beach^{+}$	11%	86.2%	
3. $Beach^{-}$	Hotel	-	-	-	See above
	Resort	-	-	-	
4. $Transport^{-}$	Hotel	-	-	-	Not significant or outside of competence.
	Resort	-	-	-	

In order of preference, the manager of Hotel "A" should first of all make decisions on increasing the service quality, and then on increasing the quality of food and beach/swimming pool maintenance. Transport problems – concerning flights, early check-in, and baggage storage – are not significant and can be solved within the frames of service improvement. The process of service quality increase can take much time; that is why organizing entertainment and animated programs together with solving problems in connection with restaurant service and beach/swimming pool maintenance can serve as immediate measures to increase client satisfaction. Specification of managerial decisions can be performed on the basis of the information on existing problems contained in negative reviews. The extracted opinions on aspects can be used by hotel managers to improve specific service areas.

Conclusion

1) The suggested conception, based on the approach of text data processing and analysis that we have developed, allows us to undertake quantitative and qualitative research of customer satisfaction using computer-aided procedures and thus enabling the making of effective managerial decisions about product quality management. The present conception allows for the effective reduction of labor intensity of customer satisfaction research that makes it available for use by a wide range of companies.

2) The experiment performed has proved its effectiveness for solving real problems of quality management, a satisfactory accuracy of Text Mining algorithms, and consistency of the results obtained.

Future work in this research area can be devoted to the automatic annotation of text data, to the representation of the huge amount of text found in of the reviews in the form of a summary, and to extracting useful and unique information.

References

[1] ISO 9000:2008. The Quality Management System. Fundamentals and vocabulary

[2] ISO10004:2010. Quality Management. Customer satisfaction. Guidelines for monitoring and measuring

[3] Viktor Mayer-Schönberger, Kenneth Cukier Big Data: A Revolution That Will Transform How We Live, Work, and Think

[4] Chernyak L. Big Data - a New Theory and Practice // Open Systems. Database 2011, № 10, pp. 18-25

[5] Jacobs A. The Pathologies of Big Data // Communications of the ACM (2009) T. 52, № 8, pp. 36-44

[6] Lynch C. Big Data: How do Your Data Grow? // Nature (2008) T. 455, №. 7209, pp. 28-29

[7] Michael Minelli, Michele Chambers, Ambiga Dhiraj // Big Data, Big Analytics: Emerging Business Intelligence and Analytic Trends for Today's Businesses ISBN: 978-1-118-14760-3, p. 224

[8] Babaev E. O., Basha N. V., Tomsha P. P., St. Petersburg State University of Economics «Big Data» Defenition. Company Readyness Indicator to Implement New Techologies to Work with Big Data Ammount

[9] Yi J., Nasukawa T., Niblack W., Bunescu R. "Sentiment Analyzer: Extracting Sentiments about a Given Topic Using Natural Language Processing Techniques", In Proceedings of the 3rd IEEE international conference on data mining, ICDM 2003, pp. 427-434

[10] Pazelskaya A. G., Soloviev A. N. "Method of the Determination Emotions in the Lyrics in Russian", Computer Program Linguistics and Intellectual Technologies, Issue 10 (17), 201, pp. 510-522

[11] Ermakov A. E., Kiselev S. L. "Linguistic Model for the Computer Analysis of Key Media of Publications", Computational Linguistics and the Intellectual Technology: proceedings of the International Conference Dialog' 2005, 2005, pp. 172-177

[12] Pang B., Lee L. "Opinion Mining and Sentiment Analysis. Foundations and Trends in Information Retrieval", Vol. 2, 2008, pp. 1-135

[13] Pang B., Lee L., "Thumbs up? Sentiment Classification using Machine Learning Techniques", Proceedings of the Conference on Empirical Methods in Natural. Language Processing (EMNLP), Philadelphia, July 2002, pp. 79-86

[14] Manning C., Raghavan P., Schuetze H. "An Introduction to Information Retrieval", Cambridge University Press. Cambridge, England, 2009, pp. 1-544

[15] Yussupova N., Bogdanova D., Boyko M. "Algorithms and Software for Sqentiment Analysis of Text Messages using Machine Learning", Vestnik USATU, T. 16-6(51), 2012, pp. 91-99

[16] Yussupova N., Bogdanova D., Boyko M. "Applying of Sentiment Analysis for Texts in Russian Based on Machine Learning Approach", IMMM 2012, Venice, Italy, 2012, pp. 8-14