**Abstract**

Surveys are an easy, important and reliable method to measure the “pulse” of the organization’s stake-holders. A survey helps in identifying improvements to current products, services and business processes. With the advent of the Web, it is now easy to conduct large-scale on-line surveys. However, it is a challenge to analyze the responses to derive novel, interesting, actionable insights and to design effective improvement plans. In this paper, we describe a tool called QUEST for analyzing survey responses. QUEST’s pre-packaged knowledge containers provide frequently needed analysis of survey responses. Built-in analysis in QUEST varies from summaries, reports and charts to detailed statistical and data mining analysis and optimization. The analytics is designed to answer specific business questions, detect specific types of patterns, extract specific kind of useful and actionable knowledge and automatically suggest optimal improvement plans. We present a real-life case-study where QUEST was used to analyze responses from a real-life employee satisfaction survey in an IT company.

**Keywords:** Survey analysis, Employee satisfaction survey, data mining, text mining, Domain-driven data mining, Root cause analysis.

1 Introduction

1.1 Surveys: Motivations and Types

Knowing specific issues, problems and the likes and dislikes of employees, customers, suppliers etc. is critical for any organization. Surveys are an easy, important and reliable method to measure the “pulse” of the organization’s stake-holders. A survey helps in identifying improvements to current products, services and business processes. These improvements are directly driven from the stake-holders’ needs as reflected in the survey responses and hence are more effective in increasing satisfaction, reducing costs, improving products etc. Survey results can be used to make business decisions that are supported by findings of real issues, needs, feelings of the stake-holders.

Surveys are often designed to understand and extract specific needs of a well-defined class of stake-holders of an organization. For example, employee satisfaction survey (ESS), product evaluation survey, customer satisfaction surveys etc.

Different media can be used to conduct surveys. In telephonic surveys, an executive contacts the target person over a telephone, asks questions and notes down the answers in a survey form. Similar activities would take place in Gallup-style person-to-person surveys. Telephonic and manual surveys impose various constraints on the survey process, limiting the effectiveness of the survey. Number of questions cannot be too many, cannot be too specific (Eg., requiring respondent to quote specific dates or figures) and so on. The number of people covered is also limited and the accuracy of responses is often doubtful, due to manual data entry. Last, but not the least, the customer is not directly responding to the questions – there is an intermediary.

With the advent of computers and the Web, these limitations are easy to remove. Many organizations are moving towards online surveys, where respondents fill up the survey questionnaire over the web or the Intranet. The responses are stored in relational databases (see Figure 1). Online surveys offer more flexibility and convenience to the respondents; they can complete the surveys at any suitable time, can save partially filled up responses etc. Finally, online surveys have a much larger reach, much lower data collection costs (Eg., avoids manual interviews or data entry) and more reliable data collection (Eg., online validations can detect simple errors in responses and prompt the respondent accordingly).

Surveys typically contain several types of questions: value, structured and unstructured. **Value questions** ask the respondent to provide specific values (Eg., dates, numbers etc.). For example, “Enter the number of years you own product X: ___”. **Structured questions** offer a few fixed options (called domain of values) to the respondent, who chooses one from them. For example, “Type of food you prefer: Vegetarian, Non-vegetarian”, “Rate the food quality: 1 2 3 4”. **Unstructured questions** ask the respondent to provide a free-form answer without any restrictions. For example, “Please suggest ways in which we could improve the service”.

In this paper, we treat value questions as structured questions. Questions are often grouped into categories; questions within a category gather responses about a specific aspect of the product, service or organization. For example, an airline customer survey might group the questions into categories like in-flight services, airport services, check-in processes, food, tickets etc. Questions in in-flight services category gather responses about reading material, entertainment programmes, cabin temperature, seats, cleanliness, cabin crew, announcements etc.
Some surveys ask the respondents to provide an importance to each question (or to each category of questions). Possible values for importance are usually fixed; Eg., unimportant, moderate, high. Answers to structured questions are often mapped (internally) to an ordered (or unordered) set of numeric values; Eg., integers 0 to \(N\) for some \(N\). In some surveys, the respondents are allowed to not answer some questions. Lastly, questions may be classified as negative or positive and responses to these may need to be treated separately.

Some respondent data (i.e., information about the respondents themselves) may be collected during the survey; Eg., age, income, gender, education, location, products owned etc. Finally, some business data (Eg., details of products, prices, history of interactions with customers, marketing campaigns etc.) may also be available.

Once the survey responses are collected and the survey is closed (no more responses are coming in), the next task is to analyze the responses and decide the future course of action. There are basically several goals to this analysis (discussed later).

1.2 QUEST Tool for Response Analysis
In this paper, we discuss QUEST: a tool for analyzing the responses of a survey. QUEST is aimed at non-specialist users who are not experts in statistics or data mining. Hence the central idea in QUEST is to pre-package a standard knowledge container, containing frequently needed analysis of survey responses. Thus QUEST follows a domain-driven data mining approach and specializes the data-mining and analysis algorithms to answer specific business questions in the survey domain. The built-in analysis in QUEST varies from summaries, reports and visual charts to detailed statistical analysis and data mining analysis designed to answer specific business questions, detect specific types of patterns and to extract specific kind of useful and actionable knowledge. We have also built a companion tool (not discussed here) called SEEK that can be used to design a survey questionnaire, deploy it on the Web, collect responses etc.

One unique feature of QUEST is the integration of statistical, data mining, and text mining functionality for response analysis. Another important feature of QUEST is its focus on providing answers to commonly-encountered business questions. QUEST does not cover specialized types of surveys such as Gap Analysis Surveys, Price Sensitivity Surveys etc. These specialized surveys are done to evaluate specific business hypotheses. QUEST is oriented towards satisfaction surveys which collect feedback from employees/customers. Further, statistical and data mining analysis in QUEST currently focuses on structured responses (Eg., satisfaction levels). Next version of QUEST will provide more facilities for analysis of numeric responses (Eg., preferred price).

QUEST provides facilities for creating many different reports and charts for quickly getting high-level summary of the responses. Statistical and data mining facilities can then be used to (a) do in-depth exploration of the data (b) get specific insights and (c) analyze the responses to answer specific business questions. Text-mining facilities can be used in a similar way to summarize, group, extract and analyze textual responses. Results from analysis of textual responses can be linked to the results from analysis done using statistical and data mining facilities. QUEST includes some optimization functionality as well.

This paper is organized as follows. Section 2 presents various types of analysis that can be performed on survey responses. Section 3 discussed the architecture and design of QUEST. Section 4 presents a real-life case-study where QUEST was used to analyze responses from a real-life ESS in an IT company. Section 5 provides some related work. We discuss conclusions and further work in Section 6.

2 Analysis of Survey Responses

2.1 Aims of Analysis
The survey responses are usually analyzed in an interactive and exploratory manner. The aims of this analysis are two-fold:

a) get a detailed understanding of the current status of the stakeholders needs, concerns and behaviour; and

b) design a future course of action for achieving specific improvements based on the findings in (a).

In practice, the users resort to different types of analysis to get a specific understanding and to design a specific course of action. Typical “business goals” of the analysis of an ESS are as follows (other types of surveys need similar analysis):

1. What are the major areas of concerns (or unhappiness)?
2. How does the “unhappiness” vary over employee attributes or their combinations (Eg., across age, branches, designations etc.)? Are there any unusual variations; Eg., are there any subsets of employees in specific branch, having a specific designation, who are more unhappy as compared to similar combinations?
3. Are there any inter-relationships between areas of concerns?
4. Can unhappy employees be partitioned into subgroups (or subsets), where employees within each subset share lot of common characteristics?
5. What are good predictors of employee unhappiness?
6. Identify “interesting” subsets of unhappy employees.
7. What are the root-causes of employee unhappiness?
8. Perform what-if (or impact) analysis to judge how specific changes in responses will affect overall employee satisfaction. For example, how will a 10% satisfaction increase in company transportation category among employees with designation = ITA affect the overall satisfaction level?
9. Identify optimal ways to achieve specified increase (Eg., 8%) in employee satisfaction.

Similar questions can be asked about “happy” employees. Depending on the kind of additional data available for employees much further analysis can be done. For example, if employee performance ratings and
work history are available then their relationships with satisfaction levels can be explored (Eg., do satisfied employees get better ratings? Do employees who travel abroad frequently have higher satisfaction levels?).

There are several challenges in the analysis of survey responses: large volumes, complex data and need to define specific goals for analysis. Another challenge is to design and apply appropriate techniques for specific analysis of textual responses. Combining results of analysis of structured and textual data can be difficult.

A number of simple reports and charts can be designed to summarize the basic facts regarding the survey responses. Some of these are shown in Section 4. In this section, we illustrate how statistical and data mining analysis can be performed to answer the business questions listed above for ESS. Analysis for other types of surveys is similar.

2.2 Satisfaction Index

Computing satisfaction index (SI) is important for the analysis of survey responses. Suppose a survey consists of M structured questions Q1, Q2, ..., QM. Each question Qj has a fixed domain Dj of possible answers; |Dj| denotes the number of possible answers for Qj. Without loss of generality, we assume that each Dj is a set consisting of numbers 0, 1, ..., |Dj|−1. This ordered representation of answers to a question is sometimes inappropriate for categorical questions, whose answers are unordered. For example, possible answers to the question What is your current marital status? might be unmarried, married, divorced and widowed, which cannot be easily mapped to numbers (say) 0, 1, 2, 3. Assume that there are N respondents, each of whom has answered each of the M questions. For simplicity, we ignore the possibility that some respondents may not have answered some of the questions. Let Rij denote the rating (or the possibility that some respondents may not have answered each of the M questions. For simplicity, we ignore the importance of the categories.

algorithm BF_AOC_subsets

Let C be the set of k questions having the lowest SI; //global AOC
for i = 1 to P do
  for each i-subset B = {A[i], A[j], ..., A[P]} of A do
    // B is a subset of A and contains i attributes
    Let P_B = product of the domains of attributes in B
    for each i-tuple X in P_B do
      Let D_X be the subset of responses satisfying X
      Let C_X be the areas of concerns for D_X
      if C and C_X differ significantly then print X;
    endif
  endfor
endfor
endfor

For example, we might discover using the above algorithm that the subset of respondents described by age $\geq 30..35 \land$ designation $\in \{ITA, AST\}$ has substantially different areas of concerns than the rest of the employees. We can define similar algorithms for deriving insights (b) and (c). Alternatively, to choose most common areas of

2.3 Areas of Concern

An area of concern is either a category (i.e., a group of related questions) or a question, having very low SI. In the simplest case, k questions (or categories) having the lowest SI can be identified as top k overall areas of concern. Areas of concerns can be also computed for specific groups of respondents (Eg., age groups, locations, designations etc.) and then compared. Following interesting insights can then be derived: (a) subsets of respondents whose areas of concerns differ from those of the entire set of respondents (b) most common areas of concerns (c) Least common areas of concerns (d) subsets of respondents that include one (or more) of the least common areas of concerns.

Let $A = \{A_1, A_2, ..., A_P\}$ denote the set of discrete attributes of the respondents (Eg., age, gender, designation, location, experience etc.). Let $V_i$ denote the finite non-empty set of values of attribute $A_i$, $1 \leq i \leq P$. Following brute force subgroup discovery algorithm for (a) systematically examines subsets of respondents and identifies those whose areas of concerns differ from those of the entire set of respondents. A randomized version of this algorithm randomly picks the subsets $B$ of $A$ and their descriptors $P_B$. Another version of this algorithm adopts the beam search strategy to reduce the search space. Well-known Jaccard coefficient can be used to decide how dissimilar given two sets of categories are; the standard definition is adapted to take into account importance of the categories.

Clearly, 0 ≤ S(Qj) ≤ 100.0 for all questions Qj. If all employees answer 0 to a question Qj, then S(Qj) = 0. If all employees answer [Dj]−1 to a question Qj, then S(Qj) = 100.0. SI for a category (i.e., a group of related questions) can be computed similarly. The overall SI is the average of the SI for each question:

$$S = \frac{\sum_{j=1}^{M} S(Q_j)}{M}$$

We can analogously define SI S(i) for each respondent. Overall SI can be computed in several equivalent ways.
concern, we could select questions having largest support (number of respondents) in the lowest satisfaction level.

Inter-relationships (Eg., independence) between questions (i.e., areas of concerns) are often interesting. For example, sample correlation coefficient \( r_{ij} \) between questions \( Q_i \) and \( Q_j \) gives a good idea of the dependence between them. Instead of \( r_{ij} \), we could use non-parametric statistic such as \( \chi^2 \)-coefficient. Thus we can identify \( k \) pairs of questions with highest dependence between them. For example, we might find that compensation and immediate supervisor are the two most highly correlated questions (see Figure 6). This means that whenever an employee gives a low score to compensation, he/she is very likely to give low score to immediate supervisor also and vice versa. If the questionnaire contains many questions, then using \( r_{ij} \) as question similarity measure, clustering (Jain A.K., Murty M.N., and Flynn P.J. 1999) techniques (Eg., average linkage or Ward’s algorithm) can be used to identify groups (clusters) of most correlated questions. Thus we might find that the questions \{compensation, immediate supervisor, appraisal\} are most highly correlated. As another type of analysis, QUEST can identify \( k \) questions most highly correlated with the SI. Such questions are predictors for the respondent’s satisfaction level. For example, suppose that compensation and appraisal are 2 questions most highly correlated with the employee SI. Then knowing the responses of an employee to only these 2 questions, we could predict his/her final SI with good accuracy. Thus top \( k \) predictor questions are good candidates for overall areas of concerns for the set of respondents as a whole. We can also use association rule mining to find more complex dependencies among questions.

2.4 Interesting Subsets

An important goal of analysis is to find interesting subsets (subgroups) of “unhappy” respondents, such that respondents in each subgroup can be succinctly described by common (shared) characteristics. A subgroup of respondents is interesting if its statistical characteristics are very different from the rest of the respondents. For example, suppose a subgroup \( F_i \) of respondents is described by age \( \geq 30.35 \) and designation \( \in \{ITA, AST\} \) and suppose also that \( F \) contains 83% unhappy respondents whereas only 34% respondents are unhappy in the set of remaining respondents. Then clearly \( F \) is interesting (see Figure 10). If such an interesting subgroup is large and coherent enough, then one can try to reduce their unhappiness by means of specially designed programmes. We have designed and used subgroup discovery algorithms (Natu M., and Palshikar G.K. 2008) for discovering interesting subgroups of respondents. We have also adapted classification techniques, such as association based classification (CBA) (Li W., Han J., and Pei J. 2001) and decision tree, for this purpose.

2.5 Predictive Models

In supervised learning, we are given a training dataset of records (Eg., employees having attributes like age, gender, designations, location, experience etc.) along with a class label for each record (Eg., happy or unhappy). The well-known statistical classification problem consists of discovering classification rules which generalize the given labeled examples. These rules can then be used to predict the class label for unseen examples. Decision trees (Quinlan J.R. 1993), support vector machines (Vapnik V. 1995) and association rule based classification (Li W., Han J., and Pei J. 2001) are some of the techniques designed to discover classification rules from a labeled training dataset (Tan P.-N., Steinbach M., and Vipin Kumar 2005, Han J., and Kamber M. 2006). We discuss several ways in which statistical classification techniques can be applied to predict the satisfaction levels of respondents. First, we discretize the overall SI of each respondent to make it a class label (Eg., unhappy, ok, happy).

a) Using only responses to predict SI. In the simplest case, we use only the response data (responses to questions) to build a predictive model for the respondent’s SI. We do not use any respondent data (Eg., age, gender etc.). We might discover classification rules such as \( \text{IF } Q_{\text{compensation}} \in \{0, 1\} \text{ AND } Q_{\text{immediate supervisor}} = 0 \text{ THEN SI = unhappy}. \) Such predictive rules give a better idea of dependence between questions and overall SI.

b) Using respondent data and responses to predict SI. Next, we build a predictive model for SI using both the response data (responses to questions) and respondent data (Eg., age, gender etc.). We might discover classification rules such as \( \text{IF } Q_{\text{compensation}} \in \{0, 1\} \text{ AND } Q_{\text{immediate supervisor}} = 0 \text{ AND designation } \in \{ITA, AST\} \text{ THEN SI = unhappy}. \) Such predictive rules, if they have large support and confidence, give a better idea of the concerns of various subgroups of respondents (see Figure 9).

c) Using only respondent data predict SI. Lastly, we use only the respondent data (Eg., age, gender etc.) to build a predictive model for the respondent’s SI. We do not use any response data (responses to questions). We might discover classification rules such as \( \text{IF } \text{designation } \in \{ITA, AST\} \text{ AND location } = \text{AHMD} \text{ THEN SI = unhappy}. \) In effect, such rules are predictive models for identifying unhappy respondents.

2.6 Root Cause Analysis

An important analysis of survey responses is concerned with identifying subsets of unhappy respondents and then identifying root causes for their unhappiness. Analysis (b) in Section 2.5 discovers the surface (or apparent) causes of unhappiness for various subsets of unhappy respondents. Textual responses are likely to contain more information about the reasons for unhappiness (see Figure 10). For example, suppose a subset of employees is unhappy about cafeteria services, which is a surface cause. Can analysis of textual answers to cafeteria related questions shed more light on why this subset of employees is unhappy about cafeteria? Consider textual
answers to the question *Suggest how cafeteria services can be improved.* Text clustering techniques can be used to automatically partition (group) answers to this question into clusters, where each cluster represents a set of related, coherent aspects. For example, the clusters may represent aspects related to cafeteria such as *more variety, cheaper prices, more cleanliness,* extend timings. We can now consider as if the questionnaire contains 4 more questions (one for each cluster) having answers \{0, 1\} i.e., *Do you want more variety in cafeteria? Do you want cheaper prices in cafeteria?* etc. For example, if a respondent had wanted *more variety* and *cheaper prices,* then his responses to the corresponding two questions are 1 and 0 to the remaining two questions. Now we can repeat the predictive model of Section (2.4)(b) and get a more detailed understanding of the reasons for the respondents’ unhappiness. Another possible approach is to apply (Gorsuch, R. L. 1983) to response data and identify a set of factors that explains the observed responses. Each factor would be a combination of some of the questions, which helps in postulating a specific cause for unhappiness.

### 2.7 More Statistical Analysis

More detailed statistical analysis can be made on the responses to draw specific inferences. Student’s t-test can be used compare two specific groups (male vs. female; employees with experience less or more than 5 years etc.) and to decide whether SI level in one group is statistically different from that in the other group.

As another example, suppose an organization has 12 branches, with \(n_1, n_2, \ldots, n_{12}\) number of employees in them. An important question is: is there any (statistically significant) difference between the satisfaction levels of these branches. That is, are the satisfaction levels in all branches similar? 1-factor ANOVA analysis can be used to compare two specific groups (male vs. female; employees with experience less or more than 5 years etc.) and to decide whether SI level in one group is statistically different from that in the other group.

### 2.8 What-If (Impact) Analysis

What-if (impact) analysis helps the users to judge how specific changes in responses will affect overall respondent satisfaction. For example, how will a 10% satisfaction increase in company transportation category among employees with *designation = ITA* affect the overall satisfaction level? Typical form of a what-if query specifies (a) a subset of respondents in terms of respondent attributes (b) a list of questions (c) change in the responses; and (d) target (currently, overall SI). An algorithm changes the responses of the given subsets of respondents for given questions as per the given change specification and computes the new overall SI.

### 2.9 Designing an Optimal Action Plan

An important outcome of an ESS is to design and implement an optimal action plan for improving SI of the respondent population as a whole. Clearly, a valid action plan must contain concrete actionable suggestions that can be implemented and that are likely to lead to a substantial increase in SI. In what sense is an action plan optimal? While different optimality criteria can be set for designing such a plan, we adopt the following:

The proposed plan should lead to the desired increase in the overall average SI at the least possible “cost”.

We assume that the expected increase \(L\) in the overall SI is specified by the user as input \(E\), if the user wishes to increase SI from 62.0 to 65.0 then \(L = 3.0\). The proposed action plan, if implemented, should achieve an increase of \(L\) in the overall SI. We assume that each category (or question) corresponds to a possible action.

<table>
<thead>
<tr>
<th>SI level</th>
<th>No. of respondents</th>
<th>No. of respondents for P1 m×10%</th>
</tr>
</thead>
<tbody>
<tr>
<td>≤ 25</td>
<td>166</td>
<td>149</td>
</tr>
<tr>
<td>25 - 50</td>
<td>1000</td>
<td>917</td>
</tr>
<tr>
<td>50 - 75</td>
<td>6948</td>
<td>6353</td>
</tr>
<tr>
<td>75-100</td>
<td>1452</td>
<td>2147</td>
</tr>
<tr>
<td>SI</td>
<td>62.8</td>
<td>64.94</td>
</tr>
</tbody>
</table>

**Table 1. SI values for a category C.**

We assume that SI values for \(C\) are discretized into 4 bins (intervals): \(B = \{B_1 = [0, 25], B_2 = (25, 50], B_3 = (50, 75], B_4 = (75,100]\} \). Let \(N\) denote the total number of respondents \((N = 9566\) in the example). Let \(b = \{b_1, b_2, b_3, b_4\}\) denote the set of values where each \(b_i\) denotes the no. of respondents whose SI falls into bins \(B_i (i = 1, 2, 3, 4)\); \(b_1 + b_2 + b_3 + b_4 = N\). For example, for \(C = \text{Career Opportunities}\) and the data in Table 1, we have \(b = \{b_1 = 166, b_2 = 1000, b_3 = 6948, b_4 = 1452\} \). The grouped average \(\psi_0(C) = 62.8\) is the average SI for \(C\).

Suppose we design and implement an improvement plan for some specific category \(C\). How do we predict the effect of such a plan on SI for \(C\)? This is a difficult question. As a simplification, we assume the following effect model: \(m\%\) respondents move from each lower SI level to its immediate higher SI level and no movement occurs in the “highest” SI bin. For \(C\), if plan \(P1\) is implemented, we expect 17 respondents (10% of 166) to move from satisfaction level 0–25 to 25–50, 100 employees move from 25 – 50 to 50 – 75 and 695 move from 50 – 75 to 75 – 100 (column 3 in Table 1). For given \(0 ≤ m ≤ 1\) called, *upward movement fraction*, the new values for the no. of respondents in each bin are:

\[
\begin{align*}
    b_1' &= b_1 - mb_1; \\
    b_2' &= b_2 - mb_2 + mb_1; \\
    b_3' &= b_3 - mb_3 + mb_2; \\
    b_4' &= b_4 + mb_4;
\end{align*}
\]

Clearly, \(b_1' + b_2' + b_3' + b_4' = N\). The grouped average of the new data is \(\psi_0(C) = 64.94\), which is the predicted average SI for \(C\) after implementing plan \(P1\). Thus plan \(P1\) leads to a benefit \(\psi(C, m) = 2.14\) (64.94 – 62.8) in overall SI for \(C\).

Let \(C = \{C_1, C_2, \ldots, C_n\}\) denote the set of \(n\) available actions (Eg., categories). Let \(0 ≤ M ≤ 100\) be a user-specified upper limit on the upward movement percentage; Eg., \(M = 50\%\) means that no more than 50% employees move from one bin to the next higher bin. Let
M = [0, M] denote the interval from 0 to M, inclusive. Then the set A of all possible action steps is defined as A = C x M; an action step is a tuple (C, m) where C \in C, m \in [0, M]. An action plan \( P = \{(C_1, m_1), (C_2, m_2), \ldots, (C_n, m_n)\} \) is a finite non-empty set of action steps, in which (i) an action appears at most once (C \neq C, for any two action steps in P) and (ii) \( m > 0 \) for every action step (C, m) in P; Eg., \( P = \{(\text{Career Opportunities}, 10\%), (\text{Company Image}, 20\%)\} \). Given an action plan \( P = \{(C_1, m_1), (C_2, m_2), \ldots, (C_n, m_n)\} \), the total benefit of P is \( \Phi(P) = \Phi(C_1, m_1) + \Phi(C_2, m_2) + \ldots + \Phi(C_n, m_n) \).

Many different improvement plans can be proposed for C, which differ in their effect (the value of m) and in their costs. Cost not only measures the monetary spending needed to implement a plan but also refers to time, efforts, resources etc. needed. Much domain knowledge is required to construct a suitable cost function. We assume that the cost function is given as input by the user. Let A denote the set of all possible action steps. A function \( f: A \rightarrow \mathbb{R}^+ \) associates a cost \( f(C, m) \) with any given action step \((C, m) \in A\). The function f should satisfy the following properties to be a valid cost function: (i) \( f(C, m) \geq 0 \) for any action steps \((C, m) \in A\); and (ii) f is a non-decreasing function in m for the same C i.e., \( f(C, m_1) \leq f(C, m_2) \) whenever \( m_1 \leq m_2 \). QUEST supports a simple built-in cost function:

\[
f(C, m) = m * g(C)
\]

Here \( g(C) \) is the user-specified relative cost for category \( C \), (Eg., improving canteen is cheaper than improving transport facilities). For any category, the cost increases linearly with m (larger upward movement will cost more). If \( g(C) = 1 \) for all categories, then we have a uniform cost function (all actions cost the same). Clearly, this cost function satisfies the conditions in the above definition of a valid cost function. More complex cost functions can be defined; but this cost function is easier for the user to specify and understand and also it simplifies the search for an optimal plan. Given an action plan \( P = \{(C_1, m_1), (C_2, m_2), \ldots, (C_n, m_n)\} \), the total cost of P is \( F(P) = f(C_1, m_1) + f(C_2, m_2) + \ldots + f(C_n, m_n) \).

Suppose the user’s goal is to design an action plan increase the overall SI to L%; Eg., if current SI is 64.82 then one possible goal would be to increase it to 70.0 (set \( L = 70.0 - 64.82 = 5.18 \)). The problem of finding an optimal plan of action is now defined as a linear program as follows. Find a subset P of action steps such that they satisfy the constraints in the definition of an action plan and total benefit \( \Phi(P) \geq L \) and total cost \( F(P) \) is minimized. QUEST solves this linear program using a standard solver.

\[
\text{minimize } f(C_1, m_1) + f(C_2, m_2) + \ldots + f(C_n, m_n)
\]

subject to

\[
m_1 \geq 0 \text{ and } m_i \leq M \quad // \text{ } 0 \leq m_i \leq M \text{, } M \text{ is a constant say 50}
\]

\[
\Phi(P) \geq L \quad // \text{ total benefit } \geq L \text{ (L is a constant say 5.18)}
\]

\[// b, \text{ values used to compute } \Phi \text{ are constants}
\]

QUEST also analyzes textual responses to identify specific actionable suggestions made by the respondents for each of the categories identified in the optimal plan. For example, if the optimal plan includes an action step of obtaining upward movement percentage of \( m=20% \) for category \( C=\text{Canteen Facilities} \), then QUEST mines the textual responses and identifies all actionable suggestions related to this category (see section 2.10).

2.10 Mining of Textual Responses

In many surveys, the respondents give free-form unrestricted textual responses to some questions. Several types of analysis can be done on the responses to a specific question.

Text clustering: The responses (or sentences in them) could be grouped using text clustering techniques (Zhao Y., and Karypis G. 2005) into clusters, such that each cluster indicates a coherent type of response. For example, responses to the question Why would you recommend our company to others? could be automatically grouped into say 3 clusters – each cluster described by keywords like helpful staff, well-known brand, excellent service.

Sentiment Analysis: Sentiment analysis techniques (Pang B., and Lee L. 2008) could be used to assign a sentiment level to each response; Eg., (+, 0, –). Assuming that there are multiple questions requiring textual responses, we could aggregate (for each respondent) the sentiments of responses to individual questions into an overall sentiment. One can then test for the correlation between the overall sentiment and overall SI for the respondents i.e., is overall sentiment a good predictor of the overall SI? We could also check whether the textual responses are consistent with related questions. For example, if a respondent has strongly positive sentiment in his answer to the above question, and his textual response to that question falls into the cluster labeled excellent service, then we may also expect a high rating from him for the question. Rate the quality of our services: 0 1 2 3. If that is not the case, then such an exception is interesting. We could analyze such exceptions: how many exceptions each question has and how many respondents are inconsistent in their responses to text and structured questions.

Discovering important suggestions: Textual responses typically contain short and very generic sentences: staff is very helpful and friendly or he solved my problem very quickly. Such comments provide little insight into reasons for respondent’s satisfaction or dissatisfaction. On the other hand, some comments are important because they are very specific and provide actionable suggestions; Eg., Please provide the facility to leave message for support executive or The fact that the company Image is not good, could be grouped using text clustering techniques (Zhao Y., and Karypis G. 2005) into clusters, such that each cluster indicates a coherent type of response. For example, responses to the question Why would you recommend our company to others? could be automatically grouped into say 3 clusters – each cluster described by keywords like helpful staff, well-known brand, excellent service.
semantic depth indicates if the answer has something specific to say. Semantic depth of a word is the distance of the word in WordNet (an online dictionary) concept hierarchy from the root word (such as entity). Average semantic depth of a sentence is the average of the semantic depths of all words in the sentence. More the average semantic depth, more specific the answer is likely to be. (c) Unique words: Document frequency of a word (i.e., the number of responses in which that word appears) indicates the specificity of the response in which that word appears. A word is unique if its document frequency is low (i.e., the word appears in only 1-2 responses). More the number of unique words in a sentence, more are the chances that the sentence is an important suggestion. A simple empirical rule for classifying a sentence as important or not is as follows: IF length > 10 AND 6.5 ≤ average semantic depth ≤ 8.0 AND 3.27 ≤ count of unique words ≤ 6.56 THEN important suggestion = 1. The algorithm to discover important suggestion computes the features for each sentence and classifies it as important or not using such rules. Some examples of discovered important suggestions are:

(a) I was phoning on my mobile and she wanted me to do a customer service survey after the call and I said no as I was paying for the call she kept going on about how it would not take long and it just irritated me so I hung up. (b) The customer service on a whole was okay but I was surprised to know that they were not prepared to tell me why my policy was not performing which made me consider cashing it in.

3 QUEST: Functionality, Architecture

QUEST is a tool for processing and analyzing responses to various types of surveys. QUEST is developed at Tata Research Development and Design Centre (TRDDC), Pune, India, which is a part of Tata Consultancy Services (TCS), QUEST is aimed at non-specialist users who are not experts in statistics or data mining. Hence the central idea in QUEST is to pre-package a standard knowledge container, which includes frequently needed types of analysis of survey responses. The built-in analysis in QUEST includes summaries, profiles, reports, various charts, KPI reports, frequently used analytics (FUA) designed to answer specific business questions, detection of specific types of patterns and to extract specific kind of useful and actionable knowledge. We have also built a companion tool (not discussed here) called SEEK that can be used to design a survey questionnaire, deploy it on the Web, collect responses etc.

3.1 Features

QUEST is a user-friendly application system which provides structured and unstructured data analysis capabilities.

Structured data analytics: User can easily select and work with a subset of responses (Eg., all respondents having location = MUM) and can create various reports for structured responses. The types of reports available with QUEST are as follows:

Satisfaction index (ASI) report: SI is an aggregated measure indicating the extent of respondents’ satisfaction (100% completely satisfied, 25% - 0% - extremely dissatisfied). As explained earlier, SI is a weighted average of the rating (responses) and importance values given by a respondent (see Figure 5).

4X4 Reports (Important categories): The notion of importance (not to be confused with importance of a category as given by the respondent) in this report is defined based on the number of respondents giving very low rating (Eg., 1 and 2 out of 4) and very high importance (Eg., 3 and 4 out of 4) to a particular category. The more number of respondents in this quadrant indicates the overall ASI is greatly influenced by these categories and hence more important the category is. The report gives sorted (descending order) list of important categories for a given intersection of data based on the above quadrant count.

Average rating and importance reports: These reports give user the sorted list of categories/questions based on their average importance or rating (see Figure 7).

Text report (for further use in clustering): The output text file generated by this functionality is used as input for text data analysis under unstructured data analysis. It is a simple report of all text comments given by an employee/customer for a particular category of open questions.

Text data analytics: QUEST also includes a text clustering functionality. QUEST can group similar respondent comments into number of clusters (groups) specified by user. Each cluster then indicates reason for satisfaction or dissatisfaction. Number of respondents in the group indicates how important a particular reason is. Instead of clustering whole responses, QUEST can build clusters out of individual sentences in responses as well. This is important because a single response may include many different types of suggestions. Then each sentence can belong to a different cluster. QUEST can also automatically identify the optimum number of clusters for the responses (see Figure 8).

Figure 2. Architecture of the QUEST tool.

3.2 Architecture

QUEST has three main components (see Figure 2):

Data repository: The survey response data can be divided into following groups:

- Employee/ Customer/ Product attribute master data - These attributes comprise of respondent attributes (age, type of responder, experience etc.), organizational structure attributes (geography, country, center etc.).
Survey attributes – The categories of questions, questions themselves, question types, their ids etc.

Answer data/Survey response collected – The actual numeric and text responses given by the respondents taking the survey

Respondent attributes are available in master tables maintained by the organization and the survey responses are available in another table. QUEST combines these two views of the data in single table and then uses it for dimensional modeling. QUEST uses two separate tables for text and numeric data.

Model repository: QUEST uses a definition file to create dimensional model of the data and to process the aggregations. The model repository stores these model definitions and the aggregations of the numeric responses such as rating and importance given to a question. This component also provides important services for processing dimensional query and rendering the results.

Analytics function/interface: The QUEST interface is divided into two sub-menus – structured data analysis and unstructured data analysis. With structured analysis interface, user can map any dimension on rows or columns and set additional data filter and can create reports for metric like satisfaction index. The QUEST interface reads the dimension model and displays the attributes of the data for taking intersection. QUEST use MS-EXCEL component to connect to the dimensional model and aggregations. The dimensional query posed by the interface is processed and the charts are plotted by excel components.

QUEST provides text-clustering functionality through unstructured data analysis interface. QUEST uses the repeated bisection algorithm (Zhao Y., and Karypis G. 2005) to group the textual responses into various clusters. The collection of responses is converted to the standard vector space model based on TF/IDF. Repeated bisection algorithm then uses the vector space model to arrive at similar documents for forming groups.

4 QUEST: A Case Study

We present, in the following section, a case study where QUEST is been successfully used to provide insights for a real-life ESS in a large IT company (the client). The client, a large software organization, values contributions made by its associates and gives paramount importance to their satisfaction. It launches an ESS every year on its Intranet to collect the feedback for various organizational functions such as human resources, work force allocation, compensation and benefits etc. The survey contains a mixture of structured and textual questions. The goal is to analyze the responses and get insights into employee feedback which can be used to improve various organization functions. Figure 3 shows sample questions; Figure 4 shows a sample response. Figure 5 – 10 present sample results from various analysis techniques discussed earlier.
Figure 8. Clusters of textual responses to organizational changes.

**Prediction rule:**

**IF** EXPERIENCE RANGE = '1-3' AND GEOGRAPHY = INDIA AND GENDER = Male **THEN** RATING for C5 = 1

Figure 9. Model to predict rating.

Figure 10. Root-cause Analysis.

QUEST was used to apply the association-rules based classification algorithm (CBA) to survey responses. This algorithm discovered the following association rule (among many others) which describes a subset of 29 unhappy employees:

\[
\text{customer} = \text{X AND designation} = \text{ASC} \Rightarrow \text{ASI} < 65
\]

This rule states that employees having designation=ASC and who are working at Customer X are unhappy. The chart in Fig. 10 then tries to find the root causes of the unhappiness of that particular subset of employees. Basically, this chart identifies the categories which are rated very low by these employees. As seen, the employees in this subgroup are significantly unhappy about categories career opportunities and compensation and benefits. Further insights into the root causes can be obtained by analysis of their responses to relevant questions.

5 Conclusions and Further Work

With the advent of the Web, it is now easy to conduct large-scale on-line surveys. However, analyzing the responses to derive novel, interesting and actionable insights to design effective improvement plans remains a challenge. In this paper, we described a tool called QUEST for analyzing survey responses. QUEST’s pre-packaged knowledge containers provide frequently needed analysis of survey responses. Built-in analysis in QUEST varies from summaries, reports and charts to detailed statistical and data mining analysis designed to answer specific business questions, detect specific types of patterns and to extract specific kind of useful and actionable knowledge. We presented a real-life case-study where QUEST was used to analyze responses from a real-life ESS in a large IT company.

We are working on enhancing the built-in analytics in QUEST and on providing a better alignment of the analytics with the business goals of conducting and analyzing the survey. In particular, we are interested in adding facilities to allow the users to state different types of (statistical) hypotheses regarding the responses, which the tool can then verify or reject. We are also working on automatically building statistical models of subsets of respondents. Incorporating more data mining techniques (e.g., anomaly detection, clustering and sequence mining) is also of interest. Another significant area of further work is better integration of text and data analytics. We also wish to provide a framework to allow the user to easily build specialized “recipes” containing sequences of analytics. Finally, we wish to link the results obtained from analysis of survey responses with applications such as customer churn prediction.

6 References


Li W., Han J., Pei J. (2001): CMAR: Accurate and Efficient Classification Based on Multiple Class-Association Rules. in Proc. 2001 IEEE Int. Conf. on Data Mining, Nov. 29-Dec. 02, 369-376.


