Visual analysis of relevant features in Customer Loyalty Improvement Recommendation

Katarzyna A. Tarnowska, Zbigniew W. Ras, Lynn Daniel, Doug Fowler

Abstract This chapter describes a practical area of application of decision reducts to a real-life business problem. It presents a feature selection (attribute reduction) methodology based on the decision reducts theory, which is supported by a designed and developed visualization system. The chapter overviews an application area - Customer Loyalty Improvement Recommendation, which has become a very popular and important topic area in today's business decision problems. The chapter describes a real-world dataset, which consists of about 400,000 surveys on customer satisfaction collected in years 2011-2016. Major machine learning techniques used to develop knowledge-based recommender system, such as decision reducts, classification, clustering, action rules, are described. Next, visualization techniques used for the implemented interactive system are presented. The experimental results on the customer dataset illustrate the correlation between classification features and the decision feature called the promoter score and how these help to understand changes in customer sentiment.

Key words: NPS, recommender system, feature selection, action rules, meta actions, semantic similarity, sentiment analysis, visualization

Katarzyna A. Tarnowska
University of North Carolina at Charlotte, e-mail: ktarnows@uncc.edu

Zbigniew W. Ras
University of North Carolina at Charlotte, Warsaw University of Technology, Polish-Japanese Academy of IT, e-mail: ras@uncc.edu

Lynn Daniel
The Daniel Group, e-mail: LynnDaniel@theDanielGroup.com

Doug Fowler
The Daniel Group, e-mail: DougFowler@theDanielGroup.com

1 Introduction

Nowadays most businesses, whether small-, medium-sized or enterprise-level organizations with hundreds or thousands of locations collect their customers feedback on products or services. A popular industry standard for measuring customer satisfaction is so called "Net Promoter Score"[1] [sat12] based on the percentage of customers classified as "detractors", "passives" and "promoters". Promoters are loyal enthusiasts who are buying from a company and urge their friends to do so. Passives are satisfied but unenthusiastic customers who can be easily taken by competitors, while detractors are the least loyal customers who may urge their friends to avoid that company. The total Net Promoter Score is computed as %Promoters -%Detractors. The goal here is to maximize NPS, which in practice, as it turns out, is a difficult task to achieve, especially when the company has already quite high NPS.

Most executives would like to know not only the changes of that score, but also why the score moved up or down. More helpful and insightful would be to look beyond the surface level and dive into the entire anatomy of feedback.

The main problem we tried to solve is to understand difference in data patterns of customer sentiment on a single company personalization level, in years 2011-2015. The same we should be able to explain changes, as well as predict sentiment changes in the future.

The second section presents work related to customer satisfaction software tools. The third section describes dataset and application area - decision problem. The fourth presents approach proposed to solve the problem, including machine learning techniques and visualization techniques. Lastly, the fifth section provides evaluation results and sixth - final conclusions.

2 Customer Satisfaction Software Tools

Horst Schulz, former president of the Ritz-Carlton Hotel Company, was famously quoted as saying: "Unless you have 100% customer satisfaction… you must improve". Customer satisfaction software helps to measure customers' satisfaction as well as gain insight into ways to achieve higher satisfaction. *SurveyMonkey* [sura] is the industry leading online survey tool, used by millions of businesses across the world. It helps to create any type of survey, but it also lacks features with regard to measuring satisfaction and getting actionable feedback. *Client Heartbeat* [cli] is another tool built specifically to measure customer satisfaction, track changes in satisfaction levels and identify customers 'at risk'. *SurveyGizmo* [surb] is another professional tool for gathering customer feedback. It offers customizable customer satisfaction surveys, but it also lacks features that would help to intelligently an-

[1] NPS®, Net Promoter®and Net Promoter®Score are registered trademarks of Satmetrix Systems, Inc., Bain and Company and Fred Reichheld

alyze the data. *Customer Sure* [cus] is a tool that focuses on customer feedback: facilitates distribution of customer surveys, gathering the results. It allows to act intelligently on the feedback by tracing customer satisfaction scores over time and observe trends. *Floqapp* [flo] is a tool that offers customer satisfaction survey templates, collects the data and puts it into reports. *Temper* [tem] is better at gauging satisfaction as opposed to just being a survey tool. Similar to *Client Heartbeat*, Temper measures and tracks customer satisfaction over a period of time.

These types of tools mostly facilitate design of surveys, however, offer very limited analytics and insight into customer feedback. It is mostly confined to simple trend analysis (tracing if the score has increased or decreased over time). The *Qualtrics* Insight Platform [qua] is the leading platform for actionable customer, market and employee insights. Besides customers' feedback collection, analysis and sharing it offers extensive insight capabilities, including tools for end-to-end customer experience management programs, customer and market research and employee engagement.

3 Dataset and application

The following section describes the dataset we worked on and corresponding application area. The problem we deal with includes attribute analysis and relevant feature selection (reduction) and ultimately an analysis of customer satisfaction and providing recommendations to improve it.

3.1 Dataset

The data was provided by a consulting company based in Charlotte within a research project conducted in the KDD Lab at UNC-Charlotte. The company collects data from telephone surveys on customer's satisfaction from repair service done by heavy equipment repair companies (called clients). There are different types of surveys, depending on which area of customer satisfaction they focus on: Service, Parts, Rentals, etc. The consulting company provides advisory for improving their clients' Net Promoter Score rating and growth performance in general. Advised companies are scattered among all the states in the US (as well as south Canada) and can have many subsidiaries. There are above 400,000 records in the dataset, and the data is kept being continuously collected. The dataset consists of features related to:

1. Companies' details (repair company's name, division, etc.);
2. Type of service done, repair costs, location and time;
3. Customer's details (name, contact, address);
4. Survey details (timestamp, localization) and customers' answers to the questions in survey;

5. Each answer is scored with 1-10 (optionally textual comment) and based on total average score (*PromoterScore*) a customer is labeled as either Promoter, Passive or Detractor of a given company.

Client attributes			Customer attributes			Service attributes		Survey attributes and questions (customer experience on client's service)				NP S	Stat us
ID	Name	Adress, …	Name	Location	…	Time	Cost ,…	Q1 (score)	Q2 (score)	Q… (score)	QN (score)		Promoter
1													Passive
2													Detractor

Fig. 1 *Illustration of NPS dataset structure - features and decision attribute.*

The data is high-dimensional with many features related to particular assessment (survey questions') areas, their scores and textual comments. The consulted companies as well as surveyed customers are spread geographically across United States. Records are described with some temporal features, such as *DateInterviewed*, *InvoiceDate* and *WorkOrderCloseDate*.

3.2 Decision problem

The goal is to find characteristics (features) which most strongly correlate with *Promoter/Detractor* label (*PromoterScore*), so that we can identify areas, where improvement can lead to changing a customer's status from Detractor to Promoter (improvement of customer's satisfaction and company's performance). Identifying these variables (areas) helps in removing redundancy.

Analysis should not only consider global statistics, as global statistics can hide potentially important differentiating local variation (on a company level). The problem is similar as in [GDST16] as correlations vary across space, with scale and over time. The goal is to explore the geography of the issue and use interactive visualization to identify interdependencies in multivariate dataset. It should support geographically informed multidimensional analysis and discover local patterns in customers' experience and service assessment. Finally, classification on NPS should be performed on semantically similar customers (similarity can be also based on geography). A subset of most relevant features should be chosen to build a classification model.

3.3 Attribute analysis

The first problem we need to solve is to find out which benchmarks are the most relevant for Promoter Status. There is also a need to analyze how the importance of benchmarks changed over years for different companies (locally) and in general (globally), and additionally how these changes affected changes in Net Promoter Score, especially if this score is deteriorated (which means customer satisfaction worsened). We need to identify what triggered the highest NPS decreases and the highest NPS growths.

The business questions to answer here are:

- What (which aspects) triggered changes in our Net Promoter Score?
- Where did we go wrong? What could be improved?
- What are the trends in customer sentiment towards our services? Did more of them become Promoters? Passives? Detractors? Did Promoters become Passives? Promoters become Detractors? If yes, why?

The problem with the data is that the set of benchmarks asked is not consistent and varies for customers, companies and years. Customer expectations change as well. Therefore, we have to deal with a highly incomplete and multidimensional data problem.

3.4 Attribute reduction

The consulting company developed over 80 such benchmarks in total, but taking into considerations time constraints for conducting a survey on one customer it is impossible to ask all of them. Usually only some small subsets of them are asked. There is a need for benchmarks (survey) reduction, but it is not obvious which of them should be asked to obtain the most insightful knowledge. For example, consulting aims to reduce the number of questions to the three most important, such as "Overall Satisfaction", "Referral Behavior" and "Promoter Score", but it has to be checked if this will not lead to significant knowledge loss in customer satisfaction problem. We need to know which benchmarks can/cannot be dropped in order to control/minimize the knowledge loss. For example, in years 2014-2015 some questions were asked less frequently since questionnaire structure changed and survey shortened for some companies. There is a need for analysis regarding how these changes in the dataset affect the previously built classification and model in the system.

3.5 Customer satisfaction analysis and recognition

The second application area is tracking the quality of the data/knowledge being collected year by year, especially in terms of its ability to discern between different types of customers defined as *Promoters*, *Passives* and *Detractors*. The main questions business would like to know the answers to are:

- What (which aspect of service provided) makes their customers being promoters or detractors?
- Which area of the service needs improvement so that we can maximize customer satisfaction?

For every client company we should identify the minimal set of features (benchmarks) needed to classify correctly if a customer is a promoter, passive or detractor. We need to determine the strength of these features and how important a feature is in the recognition process. However, answering these questions is not an easy task, as the problem is multidimensional, varies in space and time and is highly dependent on the data structure used to model the problem. Sufficient number of customer feedback on various aspects must be collected and analyzed in order to answer these questions. Often human abilities are not sufficient to analyze such huge volume of data in terms of so many aspects. There is a need for some kind of automation of the task or visual analytics support.

3.6 Providing recommendations

The ultimate goal of our research project is to support business with recommendations (recommendable sets of actions) to companies, so that we can improve their NPS. The items must be evaluated in terms of some objective metrics.

Besides, we need to make the recommendation process more transparent, valid and trustworthy. Therefore, we need to visualize the process that leads to generating a recommendation output. The end user must be able to understand how recommendation model works in order to be able to explain and defend the model validity. Visual techniques should facilitate this process.

4 Proposed approach

Addressing the problem of feature analysis, there are two approaches to feature evaluation and its importance towards classification problem.

The first one is based on the discrimination power of a set of features and how the classification problem (in terms of accuracy) is affected if we discard one or more of them in a dataset. It always starts with the set of all features used in classification. This approach is logic-based and it can be called top-down approach.

The second one is a statistic-based and it can be called bottom-up approach. It talks about the discrimination power of a single feature or a small set of features. It does not make any reference to discrimination power of combined effect of features together. To compare them, we can say that the first approach is focused more on minimizing knowledge loss, the second one more on maximizing knowledge gain.

4.1 Machine learning techniques

In this subsection we present major machine learning and data mining techniques we used to solve the problem stated in the previous section. These include: decision reducts and action rules (stemming from Pawlak theory on Information Systems, Decision Tables and Reducts), classification and clustering.

4.1.1 Decision reducts

To solve decision problems as stated in the previous sections we propose applying attribute reduction techniques. The one we propose is based on decision reducts and stems from rough set theory (logic-based).

Rough set theory is a mathematical tool for dealing with ambiguous and imprecise knowledge, which was introduced by Polish mathematician Professor Pawlak in 1982 [PM81]. The rough set theory handles data analysis organized in the form of tables. The data may come from experts, measurements or tests. The main goals of the data analysis is a retrieval of interesting and novel patterns, associations, precise problem analysis, as well as designing a tool for automatic data classification.

Attribute reduction is an important concept of rough set for data analysis. The main idea is to obtain decisions or classifications of problems on the conditions of maintaining the classification ability of the knowledge base. We introduce basic concepts of the theory below.

Information Systems

A concept of *Information System* stems from the theory of rough sets.

Definition 1. An *Information System* is defined as a pair $S = (U,A)$, where U is a nonempty, finite set, called *the universe*, and A is a nonempty, finite set of attributes i.e. $a : U \to V_a$ for $a \in A$, where V_a is called the domain of a [RW00].

Elements of U are called *objects*. A special case of *Information Systems* is called a *Decision Table* [Paw85].

Decision Tables

In a decision table, some attributes are called *conditions* and the others are called *decisions*. In many practical applications, decision is a singleton set. For example, in table in Figure 1 decision is an attribute specifying *Promoter Status*. The conditions would be all the attributes that determine *Promoter Status*, that is, question benchmarks and also other attributes (geographical, temporal, etc.).

Based on knowledge represented in a form of a decision table, it is possible to model and simulate decision-making processes. The knowledge in a decision table is represented by associating or identifying decision values with some values of conditional attributes.

For extracting action rules, it is also relevant to differentiate between so-called *flexible* attributes, which can be changed, and *stable* attributes [RW00], which cannot be changed. $A = A_{St} \cup A_{Fl}$, where A_{St} and A_{Fl} denote *stable* attributes and *flexible* attributes respectively. Example of stable attributes in customer data would be company's and survey's characteristics, while flexible would be assessment areas (benchmarks), which can be changed by undertaking certain actions (for example, staff training).

Reducts

In decision systems often not every attribute is necessary for the decision-making process. The goal is to choose some subset of attributes essential for this. It leads to the definition of *reducts*, that is, minimal subsets of attributes that keep the characteristics of the full dataset. In the context of action rule discovery an *action reduct* is a minimal set of attribute values distinguishing a favorable object from another. In our application area, it is of interest to find unique characteristics of the satisfied customers that can be used by the company to improve the customer satisfaction of 'Detractors'. We need to find a set of distinct values or unique patterns from the 'Promoter' group that does not exist in the 'Detractor' group in order to propose an actionable knowledge. Some of attribute values describing the customers can be controlled or changed, which is defined as an action. Before defining formally a *reduct*, it is necessary to introduce a *discernibility relation*.

Definition 2. Let objects $x, y \in U$ and set of attributes $B \subset A$. We say that x, y are *discernible* by B when there exists $a \in B$ such that $a(x) \neq a(y)$. x, y are *indiscernible* by B when they are identical on B, that is, $a(x) = a(y)$ for each $a \in B$. $[x]_B$ denotes a set of objects indiscernible with x by B.

Furthermore, following statements are true:

- for each objects x, y either $[x]_B = [y]_B$ or $[x]_B \cap [y]_B = \emptyset$,
- *indiscernibility relation* is an equivalence relation,
- each set of attributes $B \subset A$ determines a partition of a set of objects into disjoint subsets.

Definition 3. A set of attributes $B \subset A$ is called *reduct of the decision table* if and only if:

- B keeps the discernibility of A, that is, for each $x, y \in U$, if x, y are discernible by A, then they are also discernible by B,
- B is irreducible, that is, none of its proper subset keeps discernibility properties of A (that is, B is minimal in terms of discernibility).

The set of attributes A appearing in every reduct of information system (decision table DT) is called *the core*.

In order to allow for local and temporal analysis we divide the yearly global NPS data (2011, 2012, 2013, 2014, 2015) into separate company data (38 datasets for each year in total). For each such extracted dataset we perform a corresponding feature selection, transformation and then run an attribute reduction algorithm (in the RSES system [rse]).

4.1.2 Classification

To guarantee mining high quality action rules, we need to construct the best classifiers first. Also, the results of classification provide an overview of the consistency of knowledge hidden in the dataset. The better the classification results, the more consistent and accurate knowledge stored in the data. Also, better ability of the system to recognize promoters/passives/detractor correctly is a foundation for the system to give accurate results.

To track the accuracy of the models built on the yearly company data we performed classification experiments for each company's dataset and for each year. Evaluation was performed with 10-fold cross-validation on decomposition tree classifiers.

Decomposition trees are used to split data set into fragments not larger than a predefined size. These fragments, represented as leaves in decomposition tree, are supposed to be more uniform. The subsets of data in the leaves of decomposition tree are used for calculation of decision rules.

We saved results from each classification task: accuracy, coverage, confusion matrix.

Decision rule

The decision rule, for a given decision table, is a rule in the form: $(\phi \rightarrow \delta)$, where ϕ is called *antecedent* (or *assumption*) and δ is called *descendant* (or *thesis*) of the rule. The antecedent for an atomic rule can be a single term or a conjunction of k elementary conditions : $\phi = p_1 \wedge p_2 \wedge ... \wedge p_n$, and δ is a decision attribute. Decision rule describing a class K_j means that objects, which satisfy (match) the rule's antecedent, belong to K_j.

In the context of prediction problem, decision rules generated from training dataset, are used for classifying new objects (for example classifying a new customer for NPS category). New objects are understood as objects that were not used for the rules induction (new customers surveyed). The new objects are described by attribute values (for instance a customer with survey's responses). The goal of classification is to assign a new object to one of the decision classes.

4.1.3 Clustering

It is believed that companies can collaborate with each other by exchanging knowledge hidden in datasets and they can benefit from others whose hidden knowledge is similar. In order to recommend items (actions to improve in the service, products), we need to consider not only historical feedback of customers for this company, but we also propose looking at companies who are similar in some way, but perform better. We use concept of semantic similarity to compare companies, which is defined as similarity of their knowledge concerning the meaning of three concepts: promoter, passive, and detractor. Clients who are semantically close to each other can have their datasets merged and the same considered as a single company from the business perspective (customers have similar opinion about them).

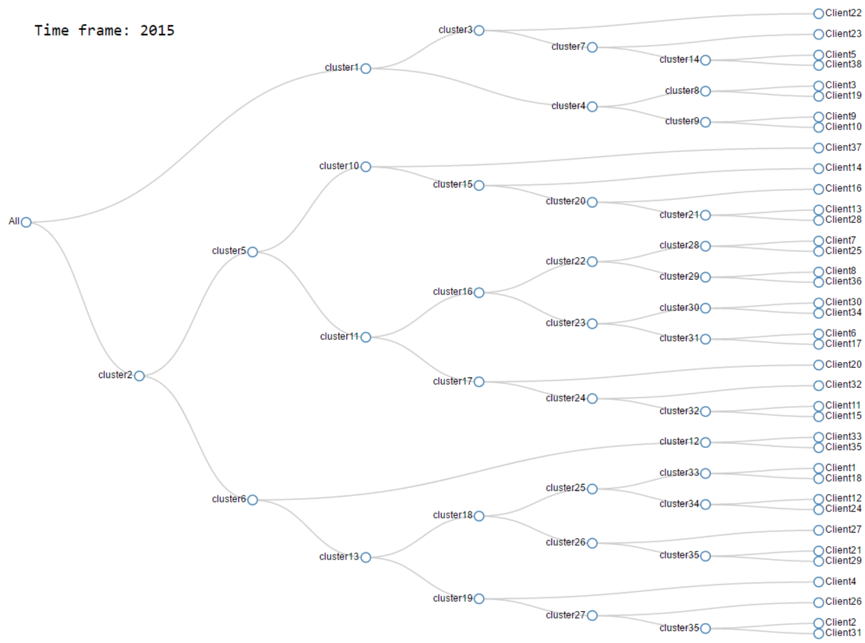

Fig. 2 *Sample hierarchical clustering dendrogram for company clients based on 2015 Service data (part of the visualization system.)*

We use hierarchical clustering algorithm to generate the nearest neighbors of each company. Given the definition of semantic similarity, the distance between any pairs of companies are quantified in a semantic way and the smaller the distance is, the more similar the companies are. A semantic similarity-based distance matrix is built on top of the definition. With the distance matrix, a hierarchical clustering structure (dendrogram) is generated by applying an agglomerative clustering algorithm.

The sample dendrogram (for year 2015) is shown in Figure 2 as a part of visualization system. The figure shows the hierarchical clustering of 38 companies with respect to their semantic similarity. In the following analysis, companies' names are replaced by numbers based on their alphabetical order, rather than using exact actual names due to the confidentiality.

A dendrogram is a node-link diagram that places leaf nodes of the tree at the same depth. If describing it using tree-structure-based terminology, every leaf node in the dendrogram represents the corresponding company as the number shown, and the depth of one node is the length of the path from it to the root, so the lower difference of the depth between two leaf nodes, the more semantically similar they are to each other. With the dendrogram it is easy to find out the groups of companies which are relatively close to each other in semantic similarity. In this example, the companies (leaf nodes) are aligned on the right edge, with the clusters (internal nodes) to the left. Data shows the hierarchy of companies clusters, with the root node being "All" companies.

4.1.4 Action rules

An *action* is understood as a way of controlling or changing some of attribute values in an information system to achieve desired results [IRT11]. An *action rule* is defined [RW00] as a rule extracted from an information system, that describes a transition that may occur within objects from one state to another, with respect to decision attribute, as defined by the user. In nomenclature, action rule is defined as a term: $[(\omega) \wedge (\alpha \rightarrow \beta) \rightarrow (\Phi \rightarrow \Psi)]$, where ω denotes conjunction of fixed condition attributes, $(\alpha \rightarrow \beta)$ are proposed changes in values of flexible features, and $(\Phi \rightarrow \Psi)$ is a desired change of decision attribute (action effect).

Let us assume that Φ means 'Detractors' and Ψ means 'Promoters'. The discovered knowledge would indicate how the values of flexible attributes need to be changed under the condition specified by stable attributes so the customers classified as detractors will become promoters. So, action rule discovery applied to customer data would suggest a change in values of flexible attributes, such as benchmarks, to help "reclassify" or "transit" an object (customer) to a different category (Passive or Promoter) and consequently, attain better overall customer satisfaction.

An action rule is built from *atomic action sets*.

Definition 4. *Atomic action term* is an expression $(a, a_1 \rightarrow a_2)$, where a is attribute, and $a_1, a_2 \in V_a$, where V_a is a domain of attribute a.

If $a_1 = a_2$ then a is called stable on a_1.

Definition 5. By *action sets* we mean the smallest collection of sets such that:

1. If t is an atomic action term, then t is an action set.
2. If t_1, t_2 are action sets, then $t_1 \wedge t_2$ is a candidate action set.
3. If t is a candidate action set and for any two atomic actions $(a, a_1 \rightarrow a_2)$, $(b, b_1 \rightarrow b_2)$ contained in t we have $a \neq b$, then t is an action set. Here b is another attribute $(b \in A)$, and $b_1, b_2 \in V_b$.

Definition 6. By an *action rule* we mean any expression $r = [t_1 \Rightarrow t_2]$, where t_1 and t_2 are action sets.

The interpretation of the action rule r is, that by applying the action set t_1, we would get, as a result, the changes of states in action set t_2.

The ultimate goal of building an efficient recommender system is to provide actionable suggestions for improving a company's performance (improving its NPS efficiency rating). Extracting action rules is one of the most operative methods here and it has been applied to various application areas like medicine - developing medical treatment methods [TRJ17], [WRW08], [ZRJT10] sound processing [RW10], [RD11] or business [RW00].

In our application, we first extend the single-company dataset (by adding to it datasets of its semantic neighbors). Next we mine the extended datasets for finding action rules, that is, rules which indicate action to be taken in order to increase Net Promoter Score. The first step of extracting action rules from the dataset is to complete the initialization of mining program by setting up all the variables. The process of initialization consists of selecting stable and flexible attributes and a decision attribute, determining the favorable and unfavorable state for a decision attribute, and defining a minimum confidence and support for the resulting rules. The results of action rule mining are used in the process of generating recommendations in our approach.

4.1.5 Meta actions and triggering mechanism

Recommender system model proposed by us is driven by action rules and meta-actions to provide proper suggestions to improve the revenue of companies. Action rules, described in the previous subsection, show minimum changes needed for a client to be made in order to improve its ratings so it can move to the promoter's group. Action rules are extracted from the client's dataset.

Meta-actions are the triggers used for activating action rules and making them effective. The concept of *meta-action* was initially proposed in Wang et al.[WJT06] and later defined in Ras et al. [TR10]. Meta-actions are understood as higher-level actions. While an action rule is understood as a set of atomic actions that need to be made for achieving the expected result, meta-actions need to be executed in order to trigger corresponding atomic actions.

For example, the temperature of a patient cannot be lowered if he does not take a drug used for this purpose - taking the drug would be an example of a higher-level action which should trigger such a change. The relations between meta-actions and changes of the attribute values they trigger can be modeled using either an influence matrix or ontology. An example of an influence matrix is shown in Table 1. It describes the relations between the meta-actions and atomic actions associated with them. Attribute a denotes stable attribute, b - flexible attribute, and d - decision attribute. $\{M_1, M_2, M_3, M_4, M_5, M_6\}$ is a set of meta-actions which hypothetically triggers action rules. Each row denotes atomic actions that can be invoked by the set of meta-actions listed in the first column. For example, in the first row, atomic actions $(b_1 \rightarrow b_2)$ and $(d_1 \rightarrow d_2)$ can be activated by executing meta-actions M_1, M_2 and M_3 together.

Table 1 Sample meta-actions influence matrix

	a	b	d
$\{M_1, M_2, M_3\}$		$b_1 \rightarrow b_2$	$d_1 \rightarrow d_2$
$\{M_1, M_3, M_4\}$	a_2	$b_2 \rightarrow b_3$	
$\{M_5\}$	a_1	$b_2 \rightarrow b_1$	$d_2 \rightarrow d_1$
$\{M_2, M_4\}$		$b_2 \rightarrow b_3$	$d_1 \rightarrow d_2$
$\{M_1, M_5, M_6\}$		$b_1 \rightarrow b_3$	$d_1 \rightarrow d_2$

In our domain, we assume that one atomic action can be invoked by more than one meta-action. A set of meta-actions (can be only one) triggers an action rule that consists of atomic actions covered by these meta-actions. Also, some action rules can be invoked by more than one set of meta-actions.

If the action rule $r = [\{(a, a_2), (b, b_1 \rightarrow b_2)\} \implies \{(d, d_1 \rightarrow d_2)\}]$ is to be triggered, we consider the rule r to be the composition of two association rules r_1 and r_2, where $r_1 = [\{(a, a_2), (b, b_1)\} \implies \{(d, d_1)\}]$ and $r_2 = [\{(a, a_2), (b, b_2)\} \implies \{(d, d_2)\}]$. The rule r can be triggered by the combination of meta-actions listed in the first and second row in Table 1, as meta-actions $\{M_1, M_2, M_3, M_4\}$ cover all required atomic actions: (a, a_2), $(b, b_1 \rightarrow b_2)$, and $(d, d_1 \rightarrow d_2)$ in r. Also, one set of meta-actions can potentially trigger multiple action rules. For example, the mentioned meta-action set $\{M_1, M_2, M_3, M_4\}$ triggers not only rule r, but also another rule, such as $[\{(a, a_2), (b, b_2 \rightarrow b_3)\} \implies \{(d, d_1 \rightarrow d_2)\}]$, according to the second and fourth row in Table 1, if such rule was extracted.

The goal is to select such a set of meta-actions which would trigger a larger number of actions and the same bring greater effect in terms of NPS improvement. The effect is quantified as following: supposing a set of meta-actions $M = \{M_1, M_2, M_n : n > 0\}$ triggers a set of action rules $\{r_1, r_2, \ldots, r_m : m > 0\}$ that covers objects in a dataset with no overlap. We defined the coverage (support) of M as the summation of the support of all covered action rules. That is, the total number of objects that are affected by M in a dataset. The confidence of M is calculated by averaging the confidence of all covered action rules:

$sup(M) = \sum_{i=1}^{m} sup(r_i)$

$conf(M) = \frac{\sum_{i=1}^{m} sup(r_i) \cdot conf(r_i)}{\sum_{i=1}^{m} sup(r_i)}$

Fig. 3 *Extracted meta actions are visualized as recommendations with different NPS impact scores. Each meta action can be assigned different feasibility.*

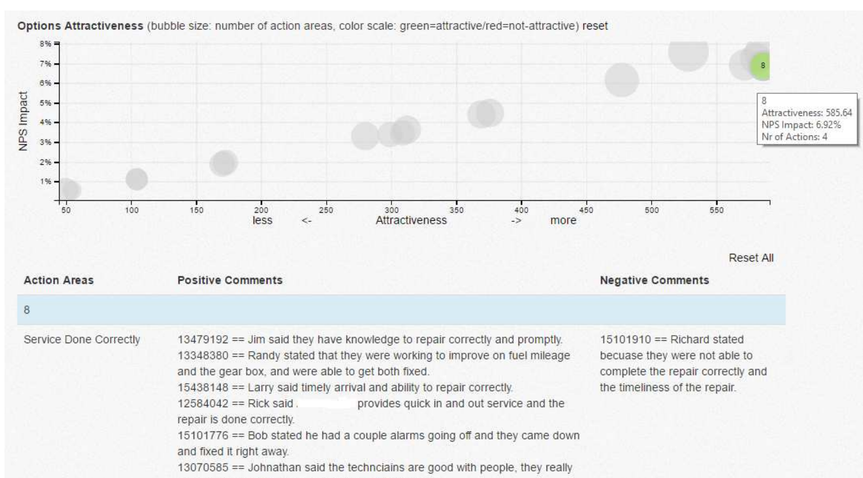

Fig. 4 *Each recommendable item (set of meta-actions) can be further analyzed by means of its attractiveness, feasibility and associated raw text comments.*

The effect of applying M is defined as the product of its support and confidence: $(sup(M) \cdot conf(M))$, which is a base for calculating the increment of NPS rating.

The increment in NPS associated with different combinations of meta-actions (recommendations) are visualized in an interactive recommender system (Figures 3, 4)

4.2 *Visualization techniques*

This subsection is related to the visualization system developed within this research and some visualization work related to solving similar problems.

4.2.1 Related work

The work in "Visualizing Multiple Variables Across Scale and Geography" [GDST16] from the 2015 IEEE VIS conference attempted a multidimensional attribute analysis varying across time and geography. Especially interesting is approach for analyzing attributes in terms of their correlation to the decision attribute, which involves both global and local statistics. The sequence of panels (Figure 5) allows for a more fine-grained, sequential analysis by discovering strongly correlated variables at a global level, and then investigating it through geographical variation at the local level. The paper's presented approach supports a correlation analysis in many dimensions, including geography and time, as well as, in our case - particular company, which bears similarity to the problems we tried to solve. The visualization proposed in the paper helps in finding more discriminative profiles when creating geo-demographic classifiers.

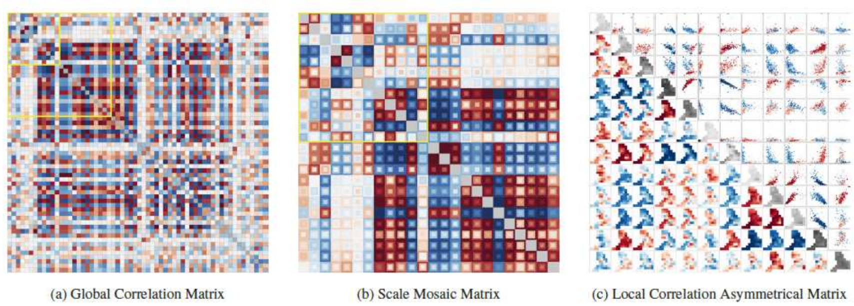

Fig. 5 *Multivariate comparison across scale and geography showing correlation in [GDST16]*

Other papers are more related to visualizing recommendations, as the visual system developed by us is mainly to support recommender system and its output (a list of recommendable items). The first such paper was "Interactive Visual Profiling of Musicians" [JFS16], from the 2015 VIS conference. The visualization system supports interactive profiling and allows for multidimensional analysis. It introduces a concept of similarity and how similar items (musicians) can be presented on visualization. We are dealing with analogous problem, as we want to compare companies (clients) based on similarity as well. The related work presented in that paper led us to the other work on recommender systems visualizations. For example, Choo et al. [CLK$^+$14] present a recommendation system for a vast collection of academic papers.

The developed VisRR is an interactive visual recommendation system which shows the results (recommended documents) in the form of a scatterplot. Problem area is similar to our project as it combines visualizing recommendations and information retrieval results.

4.2.2 Heatmap

The implemented visual system supports a feature analysis of 38 companies for each year between 2011-2015 in the area of customer service (divided into shop service / field service) and parts.

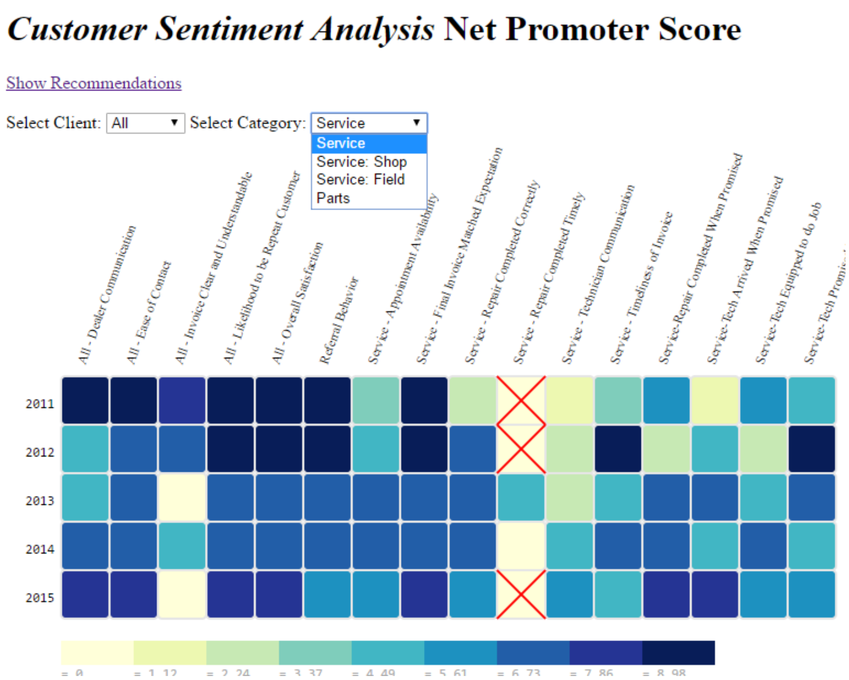

Fig. 6 *Visual design of a correlation matrix based on a heatmap.*

It serves as a feature selection method showing the relative importance of each feature in terms of its relevance to the decision attribute - Promoter Status, and it is based on the algorithm that finds minimal decision reducts. The reducts were discovered using RSES (Rough Set Exploration System [rse]). The results were saved to files from which the data-driven web-based visualization was built. Only these attributes were used that occurred in reducts and therefore, have occurrence percentage assigned as a metric displayed by the cell's color, were displayed on the heatmap (as columns).

The visualization allows the user to interactively assess the changes and implications onto predictive characteristics of the knowledge-based model. It is supported by visual additions in form of charts showing accuracy, coverage and confusion matrix of the model built on the corresponding, user-chosen dataset.

For a basic attribute analysis we propose heatmap (Figure 6) which bears similarity to a correlation matrix. However, attribute relevance to the decision attribute (Promoter Score) is defined here by means of reducts strength (a percentage occurrence of an attribute in benchmarks).

The analyst can choose the Company Category (analysis supported for 38 companies and All the companies) and Survey Type category (Service / Service:Field / Service:Shop / Parts). The columns correspond to the benchmarks (that is surveys' attributes) found in reducts. The rows represent years, in which customer satisfaction assessment was performed (current version supports datasets from years 2011-2015).

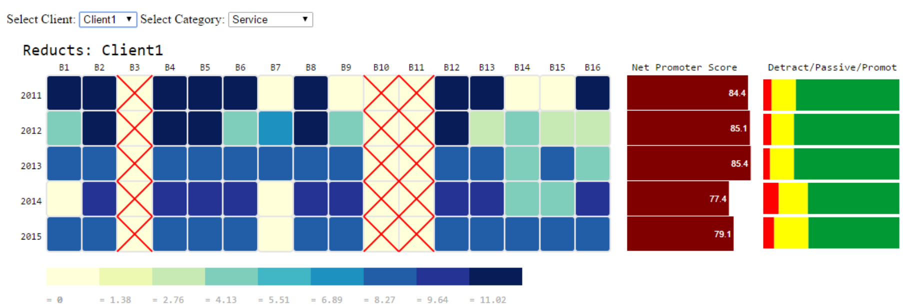

Fig. 7 *Attribute heatmap with NPS row chart and NPS category distribution chart.*

The cells represent benchmark strength in a given year - the color linear scale corresponds to an occurrence percentage. The darker cells indicate benchmarks that belong to more reducts than benchmarks represented by the lighter cells - the darker the cell, the stronger the impact of the associated benchmark on promoter score (color denotes the strength of what 'statement of action rules are saying'). The average benchmark score is visible after hovering over a cell. Additionally, we use red-crossed cells to denote benchmarks that were not asked as questions in a given year for a chosen company so that we do not confuse the benchmarks' importance with the benchmarks' frequency of asking.

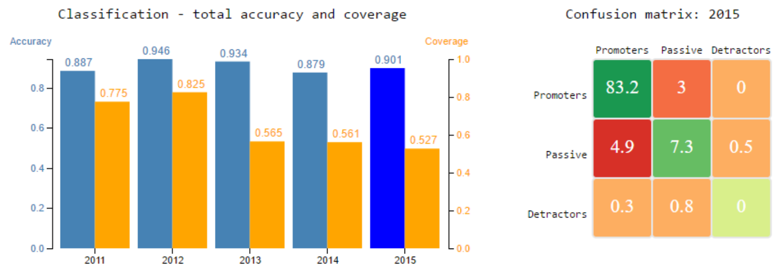

Fig. 8 *Visualizations for the classifier's results - Accuracy, Coverage and Confusion Matrix.*

The design allows to track which new benchmarks were found in reducts in relation to the previous year, which disappeared, which gained/lost in strength.

The benchmark matrix should be analyzed together with the NPS row chart to track how the changes in benchmarks' importance affected NPS (decision attribute) changes (Figure 7). The row chart showing yearly NPS changes per company is complementary to the reducts' matrix. Further, we added "Stacked area row chart" to visualize distribution of different categories of customers (detractors, passives, promoters) per year. It is complementary to the two previous charts, especially to the row NPS chart and helps to understand changes in Net Promoter Score (Figure 7). The distribution is shown based on percentage values, which are visible after hovering over the corresponding area on the chart. We used colors for differentiating categories of customers as analogy to traffic lights: red means detractors ("angry' customers), yellow for passives, and green for promoters.

4.2.3 Dual scale bar chart and confusion matrix

For the purpose of visualizing classification results we have used dual scale bar chart (Figure 8). It allows to additionally track knowledge losses by means of both: "Accuracy" and "Coverage" of the classifier model trained on single-company datasets. It is interactive and updates after the user chooses a "Client" from the drop-down menu.

Additionally, we use the confusion matrix to visualize the classifier's quality for different categories (Figure 8). From the recommender system point of view it is important that the classifier can recognize customers (covered by the model), because it means we can change their sentiment. The chart updates after interacting with the dual scale bar classification chart (hovering over the bar for the corresponding year) and shows detailed information on classifier's accuracy per each year. The rows in this matrix correspond to the actual decision classes (all possible values of the decision), while columns represent decision values as returned by the classifier in discourse. The values on diagonal represent correctly classified cases. Green colors were used to represented correctly classified cases, while red scale to represent misclassified cases. The color intensity corresponds to the extent of the confusion or the right predictions.

4.2.4 Multiple views

We use multiple views to enable a multidimensional problem analysis. However, this was quite challenging to design, as these views require both a sophisticated coordination mechanism and layout. We have tried to balance the benefits of multiple views and the corresponding complexity that arises. The multiple view was designed to support deep understanding of the dataset and the methodology for recommendations to improve NPS. The basic idea behind it is based on hierarchical structure of the system and a dendrogram was also used as an interface to navigate to more detailed views on a specific company. All other charts present data related to one company (client) or aggregated view ("All"):

- A reducts matrix is a way to present the most relevant attributes (benchmarks) in the NPS recognition process in a temporal aspect (year by year),
- NPS and Detractor/Passives/Promoters charts are complementary to reducts matrix and help track changes in Net Promoter Score in relation to benchmark importance changes,
- A Detractor/Passives/Promoter distribution chart helps understand the nature behind the changes in the NPS and customer sentiment,
- A classification and confusion matrix chart help in tracking the quality of gathered knowledge of NPS in datasets within different years and in understanding the reasons for knowledge loss (seen as decreases in classification model accuracy) in terms of benchmark changes (survey structure changes),
- A confusion matrix is a more detailed view of a classification accuracy chart and helps in understanding the reasons behind accuracy yearly changes along with the customer's distribution chart.

In conclusion, the problem of changing datasets per year and per company (client) and therefore model and algorithm can be fully understood after considering all the different aspects presented on each chart.

5 Evaluation Results

Here we present evaluation of the system on the chosen use-case scenarios.

5.1 Single client data (local) analysis

Let us present a case study of analysis process on the example of the chosen Client1 (see Figure 9). In the first row of the reduct matrix only dark colors can be observed, which means that all the colored benchmarks are equally high important. Customers' sentiment is pretty certain and defined on benchmarks. Next year, im-

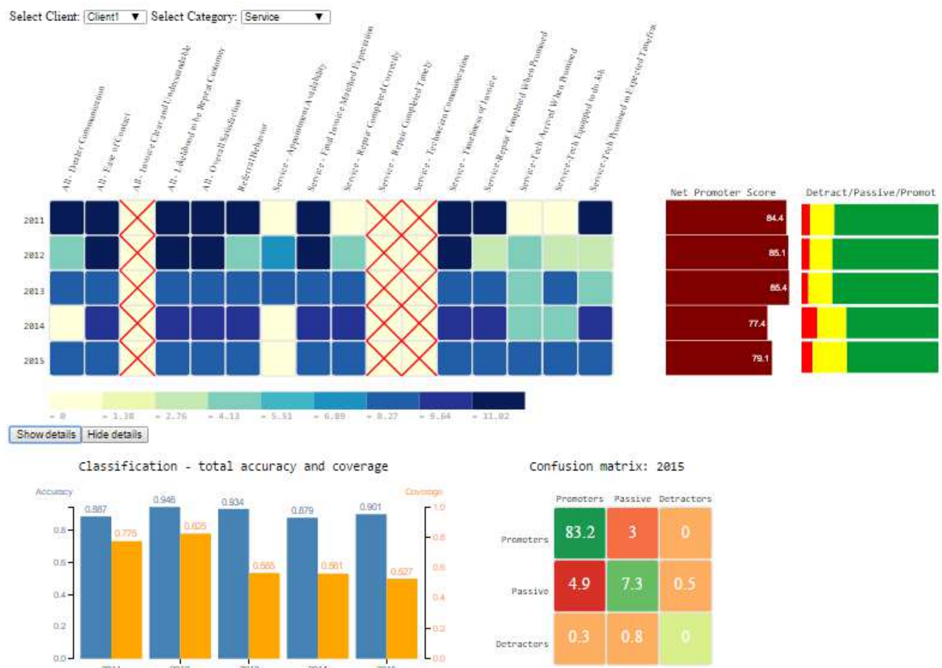

Fig. 9 *Visual analysis for a single client ("Client1").*

portance of benchmarks dropped a lot, the values decreased. For some reason, customers changed their opinion. Average score of these benchmarks says that customers are little less satisfied. New benchmarks are showing up, but they are not that important. Some benchmarks lost importance changing the color from the darkest to lighter. We can observe movement of people from Detractors to Promoters, in connection with benchmark analysis it provides much wider view. Checking confusion matrix allows to go deeper into what customers are thinking. In the third year all benchmark lost in their way. Customers do not have strong preference on which benchmarks are the most important, but NPS is still going up. In the fourth year 2 benchmarks disappeared, in 5th year all of benchmarks are again getting equal strength. We could observe a huge movement into Detractor group in 2014, but at the same time customers got more confidence which benchmarks are the most important. In 2015 all of the benchmarks have the same strength.

Number of Passives increases from 2014 to 2015, and the move to this group is from Detractor group (we could observe a huge movement from detractors to passives). Net Promoter Score was moving up and later the customers were not certain (about benchmarks preference) and then there followed a huge drop in NPS.

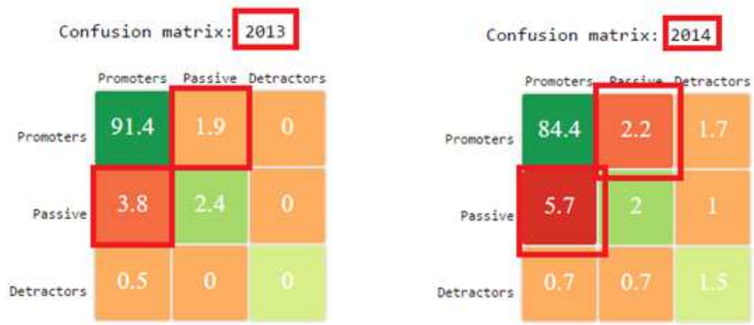

Fig. 10 *Classification results analysis - confusion matrices.*

Looking at the classification model, coverage jumped down from 2013 to 2014 incredibly that is number of customers which system can classify. Accuracy went down, and by looking at the confusion matrix (see Figure 10) we can conclude that more people who are similar in giving their rates are confused by the system between Promoters and Passives. Some of such customers end up in Promoter category, while the others in Passives.

5.2 Global customer sentiment analysis and prediction

By looking at the “Reducts” heatmap, we can draw the conclusion that the customers’ sentiment towards the importance of benchmarks is apparently changing year by year. Some benchmarks are getting more important, some less. By looking at the benchmarks for “All” companies (see Figure 6), "Overall Satisfaction" and "Likelihood to be Repeat Customer" seem to be the winners. Clearly, it will differ when the personalization goes to the companies’ level. It might be interesting to personalize further - going to different seasons (Fall, Spring,....) and seeing if the customers’ sentiment towards the importance of benchmarks is changing. Our recommendations for improving NPS are based on the customers’ comments using the services of semantically similar companies (clients) which are doing better than the company we want to help. So, it is quite important to check which benchmarks they favor. It is like looking ahead how our customers most probably will change their sentiment if the company serving these customers will follow the recommendation of the system we build. In other words, we can somehow control the future sentiment of our customers and choose the one which is the most promising for keeping NPS improving instead of improving it and next deteriorating.

From the proposed and developed charts, we can see how customers' sentiment towards the importance of certain benchmarks is changing year by year. Also, we can do the analysis trying to understand the reasons behind it with a goal to prevent them in the future if they trigger the decline in NPS.

The advantage of using recommender system the way it is proposed - based on hierarchical structure (dendrogram), assuming that companies use its recommendations, will be the ability to predict how customers' sentiment towards the importance of certain benchmarks will probably change with every proposed recommendation. So, we can not only pick up a set of benchmark triggers based on their expected impact on NPS but also on the basis of how these changes most probably will impact the changes in customers' sentiment towards the importance of benchmarks. This way we can build a tool which will give quite powerful guidance to companies. It will show which triggers are the best, when taking into account not only the current company's performance but also on the basis of the predicted company's performance in the future assuming they follow the system recommendations.

5.3 Recommendations for a single client

The first step is to find a desired client on the map (Figure 11) and click the associated dot. One can see, Client16 was found semantically similar to 2 other clients (called here extensions): (1) - most similar, (2) - secondly similar. Since these two clients were performing better than Client16 in 2015 by means of NPS, also the customers' feedback from these two clients is used to suggest recommendations for improving NPS as it shows how customers would evaluate Client16 if its performance was better but similar to its extensions. This way the base for discovering recommendations for Client16 is richer and the same should guarantee its better improvement in NPS. By similarity we mean similarity in customers' understanding of three concepts including promoter, detractor, passive.

After clicking the point for Client16, the recommendation chart for this client displays (Figure 12). We can see that in total 18 combinations of different changes recommended to implement were generated, from which 17th seems to be the most attractive (top right)-assuming all actions have the same default feasibility of 10. The generated actions are related to 8 different areas:

- service done correctly,
- price competitiveness,
- proactive communication,
- technician's knowledge,
- invoice accuracy and clarity,
- need for more technicians,

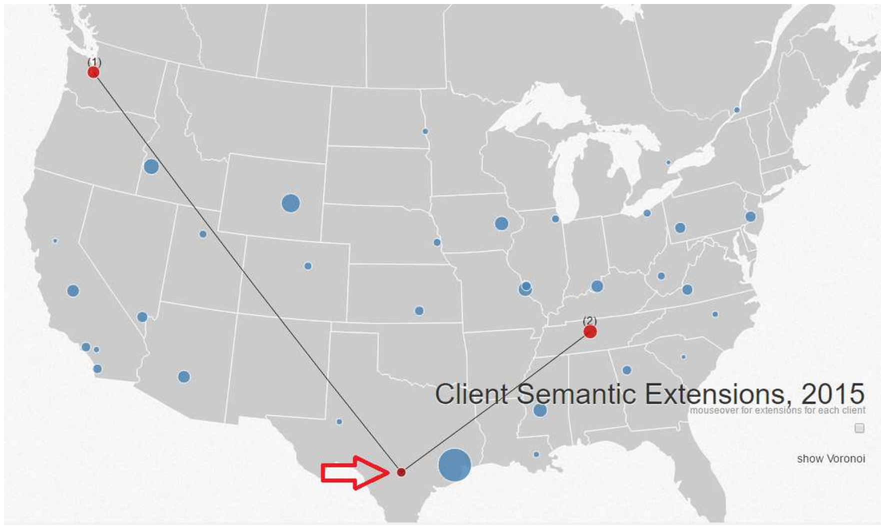

Fig. 11 *Visualization of semantic extensions of Client16.*

- care and respect from technician,
- staff care and respect to customers.

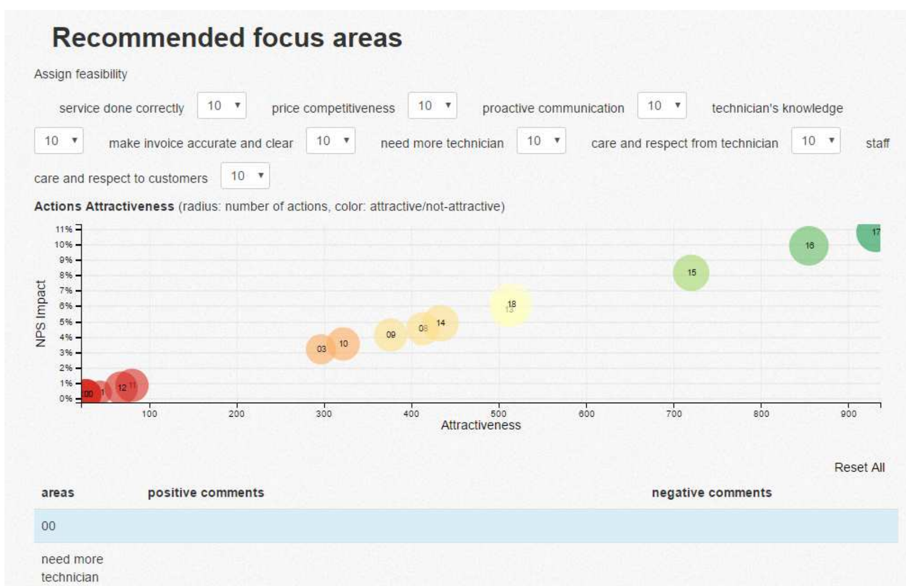

Fig. 12 *Visualizaed recommendations for the Client16.*

We can tweak with different feasibilities for improving each of this areas after consulting with the client to help choose the optimal combination of changes so that to both: maximize NPS Impact and facilitate implementing the changes. The bubble chart adjusts according to the chosen feasibility; choosing feasibility of 0 means we do not want to consider a set containing this area.

When we hover the mouse over the bubble we can see the quantified details of the changes (Figure 13):

- Ordering number (ID) - 15, 16, etc.,
- Number of changes (actions)-the greater the bubble the more changes it involves,
- NPS Impact-the predicted impact on NPS,
- Attractiveness-computed as Feasibility times Impact.

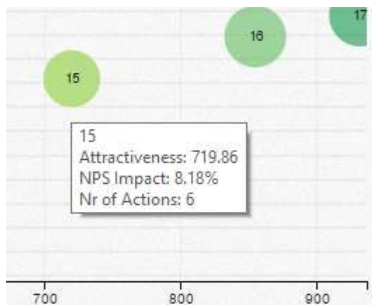

Fig. 13 *Details of the chosen recommendation.*

When we click on the bubble we can explore details in the data table (Figure 14):

- Areas of focus,
- Positive comments about this client associated with this area,
- Negative comments about this client associated with this area.

6 Conclusions

Within this research, we proposed approach (Section 4) and designed, implemented and evaluated a visualization data-driven system that supports a real-world recommendation problem in the area of business, as defined in Section 3. The problem of customer's satisfaction is multidimensional and varies with different characteristics of customer, service and repairing company. The approach presented in this chapter uses interactive visual exploration to identify variables whose correlation varies geographically and at different scales. It also allows for sensitivity analysis of variables relating to customer satisfaction, which is relevant to the repair company's consulting industry.

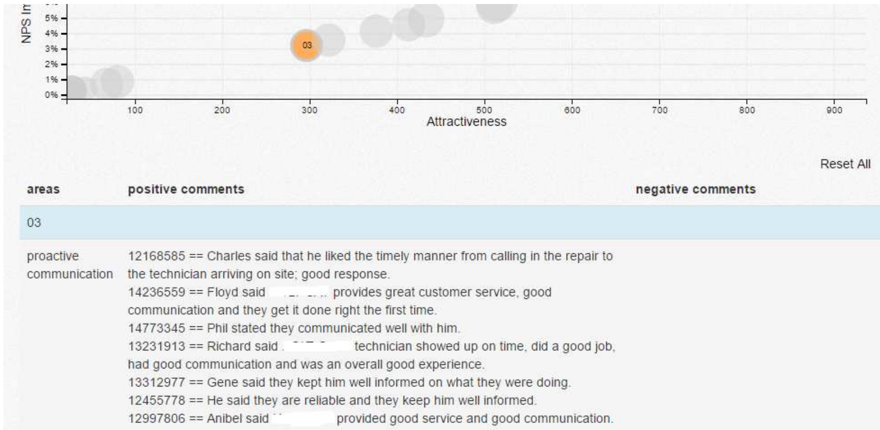

Fig. 14 *Raw text comments associated with recommendation (meta-actions) nr 3 for the Client16.*

We performed a number of preprocessing, transformation and data mining experiments on the originally raw dataset which contained superfluous and incomplete information. We proposed a wide spectrum of different visualization techniques to support analytics in the given problem area looking at different analytical aspects: locality, temporal and spatial. We attempted to balance the cognitive load of multiple views with the amount of information we needed to convey to enable effective problem insights. The developed system can also serve businesses for further customer feedback methodology's enhancement. The results in the form of NPS improvement can be explored in an interactive web-based recommender system. The results have been validated with the future users and they are convincing and satisfying from the business point of view.

References

[cli] *Client Heartbeat: Customer Satisfaction Software Tool*. https://www.clientheartbeat.com/, . – Accessed: 2016

[CLK+14] CHOO, Jaegul ; LEE, Changhyun ; KIM, Hannah ; LEE, Hanseung ; LIU, Zhicheng ; KANNAN, Ramakrishnan ; STOLPER, Charles D. ; STASKO, John ; DRAKE, Barry L. ; PARK, Haesun: VisIRR: Visual analytics for information retrieval and recommendation with large-scale document data. In: *2014 IEEE Conference on Visual Analytics Science and Technology (VAST)*, IEEE, Oktober 2014. – ISBN 978–1–4799–6227–3, 243–244

[cus] *Customer Sure Customer Feedback Software | CustomerSure*. http://www.customersure.com/, . – Accessed: 2016

[flo] *Online Survey and Benchmarking Application | Floq*. http://floqapp.com/,

[GDST16] GOODWIN, S. ; DYKES, J. ; SLINGSBY, A. ; TURKAY, C.: Visualizing Multiple Variables Across Scale and Geography. In: *IEEE Transactions on Visualization and Computer Graphics* 22 (2016), Jan, Nr. 1, S. 599–608.

http://dx.doi.org/10.1109/TVCG.2015.2467199. – DOI 10.1109/TVCG.2015.2467199. – ISSN 1077–2626

[IRT11] IM, Seunghyun ; RAS, Zbigniew W. ; TSAY, Li-Shiang: Action Reducts. In: *Foundations of Intelligent Systems - 19th International Symposium, ISMIS 2011, Warsaw, Poland, June 28-30, 2011. Proceedings*, 2011, 62–69

[JFS16] JANICKE, Stefan ; FOCHT, Josef ; SCHEUERMANN, Gerik: Interactive Visual Profiling of Musicians. In: *IEEE Transactions on Visualization Computer Graphics* 22 (2016), Nr. 1, S. 200–209. http://dx.doi.org/doi.ieeecomputersociety.org/10.1109/TVCG.2015.2467620. – DOI doi.ieeecomputersociety.org/10.1109/TVCG.2015.2467620. – ISSN 1077–2626

[Paw85] PAWLAK, Zdzislaw: *Rough sets and decision tables*. 1985

[PM81] PAWLAK, Zdzislaw ; MAREK, Wiktor: *Rough sets and information systems*. 1981

[qua] *The World's Leading Research and Insights Platform | Qualtrics.* https://www.qualtrics.com/,

[RD11] RAS, Zbigniew W. ; DARDZINSKA, Agnieszka: From data to classification rules and actions. In: *Int. J. Intell. Syst.* 26 (2011), Nr. 6, 572–590. http://dx.doi.org/10.1002/int.20485. – DOI 10.1002/int.20485

[rse] *RSES 2.2 User's Guide*. http://logic.mimuw.edu.pl/ rses,

[RW00] RAS, Zbigniew W. ; WIECZORKOWSKA, Alicja: Action-Rules: How to Increase Profit of a Company. In: *Principles of Data Mining and Knowledge Discovery, 4th European Conference, PKDD 2000, Lyon, France, September 13-16, 2000, Proceedings*, 2000, 587–592

[RW10] RAS, Zbigniew W. (Hrsg.) ; WIECZORKOWSKA, Alicja (Hrsg.): *Studies in Computational Intelligence*. Bd. 274: *Advances in Music Information Retrieval*. Springer, 2010. http://dx.doi.org/10.1007/978-3-642-11674-2. http://dx.doi.org/10.1007/978-3-642-11674-2. – ISBN 978–3–642–11673–5

[sat12] *SATMETRIX. Improving your net promoter scores through strategic account management*. 2012

[sura] *Customer Satisfaction Surveys: Questions and Templates*. https://www.surveymonkey.com/mp/csat/, . – Accessed: 2016

[surb] *SurveyGizmo | Professional Online Survey Software and Tools.* https://www.surveygizmo.com/,

[tem] *Temper - Find out how your customers feel about every aspect of your business*. https://www.temper.io/,

[TR10] TZACHEVA, Angelina A. ; RAS, Zbigniew W.: Association Action Rules and Action Paths Triggered by Meta-actions. In: *2010 IEEE International Conference on Granular Computing, GrC 2010, San Jose, California, USA, 14-16 August 2010*, 2010, 772–776

[TRJ17] TARNOWSKA, Katarzyna A. ; RAS, Zbigniew W. ; JASTREBOFF, Pawel J.: *Studies in Computational Intelligence*. Bd. 685: *Decision Support System for Diagnosis and Treatment of Hearing Disorders - The Case of Tinnitus*. Springer, 2017. http://dx.doi.org/10.1007/978-3-319-51463-5. http://dx.doi.org/10.1007/978-3-319-51463-5. – ISBN 978–3–319–51462–8

[WJT06] WANG, Ke ; JIANG, Yuelong ; TUZHILIN, Alexander: Mining Actionable Patterns by Role Models. In: LIU, Ling (Hrsg.) ; REUTER, Andreas (Hrsg.) ; WHANG, Kyu-Young (Hrsg.) ; ZHANG, Jianjun (Hrsg.): *ICDE*, IEEE Computer Society, 2006. – ISBN 0–7695–2570–9, 16

[WRW08] WASYLUK, Hanna A. ; RAS, Zbigniew W. ; WYRZYKOWSKA, Elzbieta: Application of Action Rules to HEPAR Clinical Decision Support System. In: *Experimental and Clinical Hepatology* 4 (2008), Nr. 2, S. 46–48

[ZRJT10] ZHANG, Cynthia X. ; RAS, Zbigniew W. ; JASTREBOFF, Pawel J. ; THOMPSON, Pamela L.: From Tinnitus Data to Action Rules and Tinnitus Treatment. In: *2010 IEEE International Conference on Granular Computing, GrC 2010, San Jose, California, USA, 14-16 August 2010*, 2010, 620–625