Your Knowledge Partner™

Call Center Data Analysis

Megaputer Case Study in Text Mining

Merete Hvalshagen

Megaputer Intelligence, Inc.
120 West Seventh Street, Suite 310
Bloomington, IN 47404, USA
+1 812-330-0110

www.megaputer.com

Contents

Background

This project was carried out by Megaputer Intelligence for a leading U.S. manufacturer of major home appliances. The company has manufacturing plants in a number of countries, and sells its products worldwide under a portfolio of different brand names.

Situation

The home appliance company wanted to investigate what most customers complain about. Consumers with complaints contact the company's call center where representatives type in important complaint information they receives over the phone.

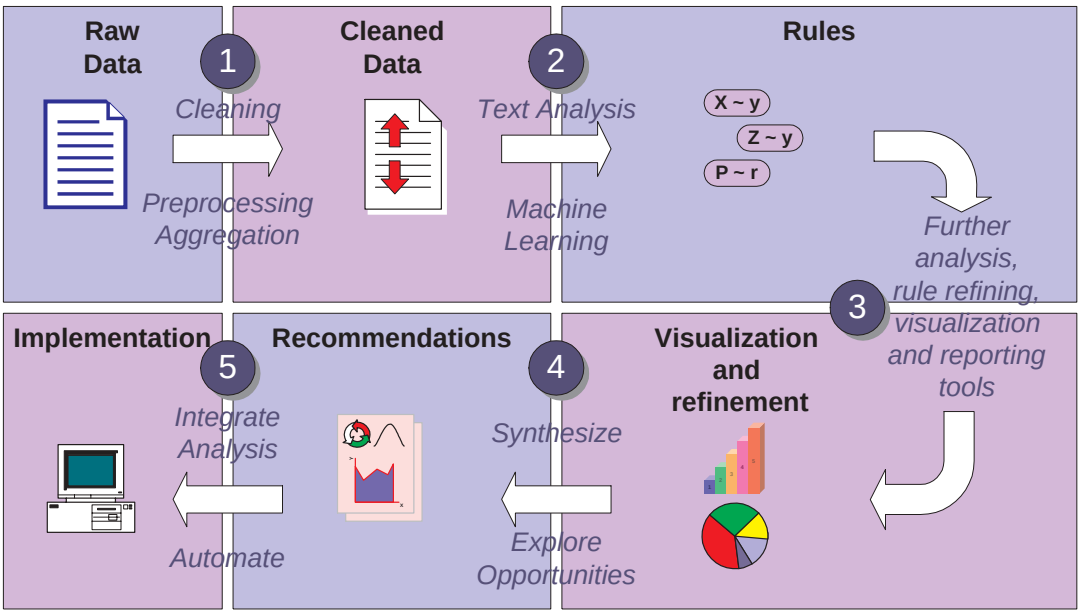
For the analysis presented here, complaints concerning two models of the same type of appliance, an older model and a newer model, were collected during a limited time period. The company was particularly interested in learning about the differences in problems between the new and old models. The company also wanted to know if the complaint text from the call center records could efficiently be used to categorize complaints according to predefined component and condition groups. The final goal was to automatically extract information from the complaint text and store it in a standardized data form.

Business Value

Quantifiable information extracted from customer complaints can be valuable to the appliance company in several ways:

- Improve products by identifying common problems with current product lines
- Reduce costs by discovering ways to process complaints and repairs more efficiently
- Increase customer satisfaction by detecting and correcting services deficiencies
- Build partner loyalty by providing better support to partner network

Approach



To obtain the results described below, the data was first cleansed and aggregated. The main task of the cleansing process was to identify and weed out frequently encountered phrases that were of no interest to the analysis. In addition, domain specific lingo had to be translated into terms more suitable for analytical purposes. The data cleansing and final analysis were carried out by utilizing the text analysis exploration engine of Megaputer Intelligence's PolyAnalyst 4.5 which includes text mining capabilities in addition to a large selection of machine learning algorithms. The Text Analysis exploration engine extracted the most important words and word combinations into textual rules. These rules were then used to tag each individual record according to the presence of the extracted terms. The obtained findings were employed by the different visualization tools Link Chart and Snake Diagrams available in PolyAnalyst to convey the results in an actionable format to business users.

Link Chart Analysis

The purpose of the Link Chart Analysis was to obtain a visual representation of the results outlining the main categories of problems and their causes. One useful chart that was created showed the problematic parts associated with the old model (called *J*) versus the new model (called *K*). This link chart creates a clear picture of the differences between the two models with regards to technical problems.

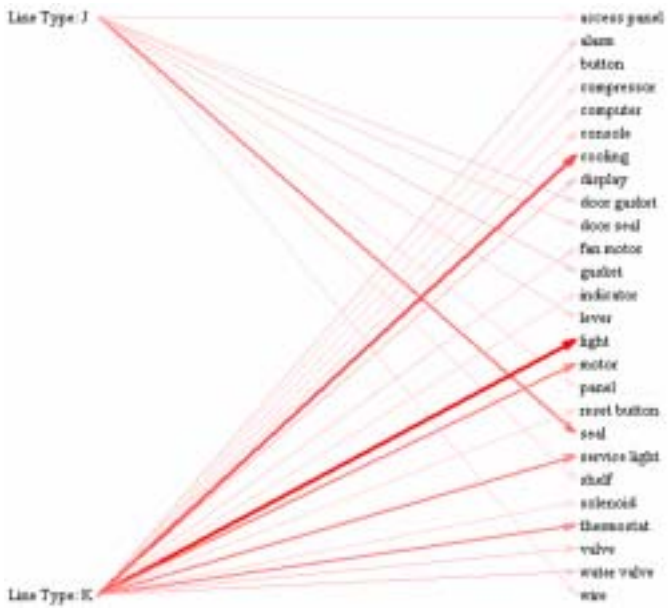


Figure 1: *Link Chart*

The lines indicate a positive correlation between the models on the left side, and problems with particular parts on the right side. A thick line conveys a stronger positive correlation, a thinner line correspond to a weaker correlation. As we can see from *Figure 1*, *Door* and *seal* complaints are strongly correlated with old model *J*, and *cooling*, motor, *light* and *thermostat* problems are strongly correlated with the new model *K*. By drilling down the company can pin point the specific problems of the models components. Listed below are a few of the actual customer complaints dealing with model J's door problems. These complaints indicate that improper alignment and closing issues are door problems.

> DOOR PROBLEMS - 4/23 KN: FREEZER DOOR IS NOT ALIGNED PROPERLY. UNIT HAS SCRATCHES ON PANEL WHERE TE ICE & WAER DISPENSER IS. . 4
>
> DOOR PROBLEMS - 7/16 YMS:FF DOOR IS STOPPING 1/2 INCH BEFORE CLOSING.HAS TO BE PUSHED CLOSED. */*2 MAN 2 CALL WAITS LET ONLY DISPA
>
> DOOR PROBLEMS - 6/12 FLC: LARGE GAP IN DOORS/ TOP OF DOORS ALMOST TOUCH AND THERE IS A LARGE GAP AT THE BOTTOM. . 6/12 KD: CALLED

By carefully examining the link chart and drilling down, the company gains valuable insights about its product line and operations. The common door problem associated with the old model *J* seems to be solved for the new model *K*. On the other hand, the company introduced some new problems with model K. These findings point to what areas the company should thoroughly look into. The next step would be to identify possible causes for the problems in order to work out feasible solutions. This might include investigating how the difficulties can be related to manufacturing processes, specific vendors, certain manufacturing plants and/or installation procedures.

Snake Chart Analysis

The Snake Chart in PolyAnalyst gives the user a more quantitative view of the results. A Snake Chart simultaneously compares two different datasets on all selected attributes on a quantitative scale.

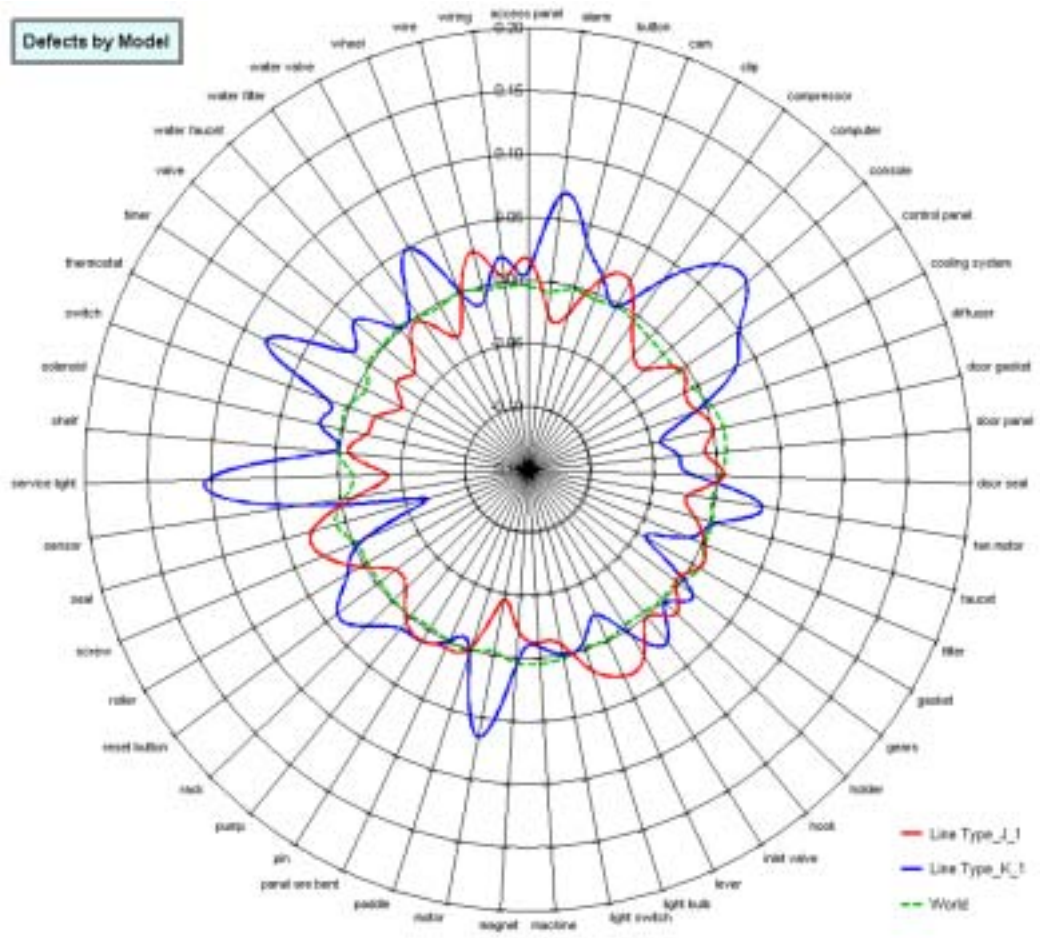


Figure 2: *Snake chart*

This chart illustrates how various types of problems are distributed differently for the new model *K* and the old model *J*. The red line represents the old model and the blue line represents the new model. The hyphenated green line is a combination of both models to indicate how the products differ from the norm. Each radial line corresponds to an identified part problem, and the Snake Chart shows the difference in the relative frequency (normalized by dispersion) of appearance of various issues in customer complaints. The further the intersection is located from the center the more often (relative to the total mean) the corresponding issue appears in customer complaints.

From *Figure 2*, we can see that Model K has a higher relative share of fan motor, cooling, noise, water valves and thermostat problems than Model J, and that Model J has a higher relative share of problems related to levers, light bulbs, seals and doors. These results are in line with those gleaned from the link chart, but provide a more detailed and informative description of the situation.

There is a number of ways a Snake Chart can be employed to provide useful information for the company. Because a Snake Chart will clearly show the differences between two or more datasets across all attributes – a visual attribute profile – it has many valuable applications. Apart from showing problem profiles for the new and old models, it can be used to illustrate how various problems develop over the life cycle of the product. Separating complaints by product age and the time the complaint was registered, the company will be able to see what new problems arise as the appliance grows older. Complaints from early sales can help the company anticipate the types of problems in the future, and thereby prepare and handle these more efficiently. Data can also be divided by manufacturing plants to see if differences in production equipment or processes influence product attributes in unexpected ways.

Classification

In many situations it can be useful to categorize complaint information into predefined "component" and "condition" codes. In order to do this, you need to define a list of possible components and conditions. Once the codes are determined, classification can be done either by matching keywords from the list, or by providing a significant amount of pre-categorized data used for training the system to find patterns of terms that determine what category should be assigned to the considered claim.

After initial training and preparation are carried out, the process can be automated and integrated into existing systems in order to deliver results to management personnel in the format suitable for making informed decisions. PolyAnalyst offers several methods for classifying such as Decision Trees and Text Categorizer.

Classification in the purpose of cross-referencing combined with term extraction can be used for populating standard forms and thereby transforming unstructured information into quantifiable terms. Once the vital information from the consumers' complaint text is available in ordinary data tables, it can easily be integrated with current reporting systems and data warehouses as a new source of business intelligence. The complaint information, for instance, can be adapted to analysis like balanced score card, vendor performance and resource allocation, or an input to product development and quality programs.

Conclusion

This case has illustrated a few text mining techniques which can be employed to provide quantifiable results as input to the decision making process. As illustrated, the analysis described above can be combined with other types of data or be employed by existing corporate analysis to obtain a more holistic picture of different operational aspects and business issues.

It's important to note that this type of textual analysis provides the company with an entirely new source of knowledge. The knowledge derived from this example could not be found elsewhere in the organization. Getting accurate feedback from critical business process is crucial for enterprises that want to improve the predictive abilities required to make optimal business decision. As in this case, turning to novel sources of knowledge can enable a company to see clearer and further into the future.

Unstructured text is widely acknowledged as a huge untapped source of valuable corporate information. Fortunately, text mining has the capability of extracting meaningful relationships from these very large quantities of unstructured data. Text mining is the keystone for enterprises that want to take business intelligence to the next level.

Corporate and Americas Headquarters
Megaputer Intelligence, Inc.
120 West Seventh Street, Suite 310
Bloomington, IN 47404, USA
TEL: **+1.812.330.0110**; FAX: **+1.812.330.0150**
EMAIL: info@megaputer.com

Europe Headquarters
Megaputer Intelligence, Ltd.
B. Tatarskaja 38
Moscow 113184, Russia
TEL: **+7.095.951.8079**; FAX: **+7.095.953.5731**
EMAIL: info@megaputer.com

Megaputer
Your Knowledge Partner™