Paper 11861-2016

Lenovo Success Story: Early Detection of Pervasive Quality Issues through Voice of Customer Analytics

Mohammed Chaara, Lenovo; Yingrui Liu Olesiuk, Lenovo; Rob Carscadden, SAS Institute Inc.; Murali Pagolu, SAS Institute Inc.

ABSTRACT

Quality issues can easily go undetected in the field for months, even years. Why? Customers are more likely to complain through their social networks than directly through customer service channels. The Internet is the closest friend for most users and focusing purely on internal data isn't going to help. Issues surfaced through the internet have the potential to be amplified and go "viral" to the user community than issues going through traditional channels. Customer complaints are often handled by marketing or customer service and aren't shared with engineering. This lack of integrated vision is due to the silo mentality in organizations and proves costly in the long run. Information systems are designed with business rules to find "out of spec" situations or to trigger when a certain threshold of reported cases is reached before engineering is involved. Complex business problems demand sophisticated technology to deliver advanced analytical capabilities. Organizations can react only to what they know about, but unstructured data is an invaluable asset that we must hatch on. So what if you could infuse the voice of the customer directly into engineering and detect emerging issues months earlier than your existing process? Lenovo just did that. Issues were detected in time to remove problematic parts, eventually leading to massive reductions in cost and time. Lenovo observed a significant decline in call volume to its corporate call center. It became evident that customers' perception of quality relative to the market is a voice we can't afford to ignore. Lenovo Corporate Analytics unit partnered with SAS® to build an end-to-end process that constitutes data acquisition and preparation, analysis using linguistic text analytics models, and quality control metrics for visualization dashboards. Lenovo executives consume these reports and make informed decisions. In this paper, we talk in detail about perceptual quality and why we should care about it. We also tell you where to look for perceptual quality data and best practices for starting your own perceptual quality program through the voice of customer analytics.

INTRODUCTION

Perceptual or perceived quality is defined by the customer's opinion. An old adage that stands unperturbed through tides of time, "the customer is always right". It is certainly influenced by many other types of quality, such as

- Actual or Objective Quality: The extent to which the product or service delivers superior service.
- Materials or Product-Based Quality: The nature and quantity of ingredients, features, or services included.
- Manufacturing Quality: Conformance to specification, the "zero defect" goal.

Why is perceptual quality so important? Perceptual quality includes a high degree of subjective measures defined by each and every customer. When it comes down to a purchasing decision, perceived quality trumps all of the other quality definitions. This is the reason why understanding it is so critical. Furthermore, understanding what is the most important to the customer can help with issue detection and prioritization, and can start us down the road to resolution, whether that means a product or process improvement, user training, or a new marketing campaign. In a survey by Maritz Research, more than 2/3 of customers who complained directly to a company, were flat out ignored. And that can come with a major penalty. 9 out of 10 customers indicate they will at least consider switching to a competitor's product based on a bad experience. Furthermore, of those whose complaints went ignored, 63% would dislike or hate it if the company contacted them about something other than their complaint. So any attempt by marketing to advertise or extend an offer would only make matters worse. But on the upside, of the 30% of customers that did get a response, the vast majority, and 83%, said they liked or loved hearing from the company. So a complaint can become a brand positive when it is actively addressed. It's not just a B2C issue. B2B and B2C respondents are included here. It's even worse for B2B customers. The bottom line is, customers want their voices heard. They have a low tolerance for sub-par experiences.

The balance of power has clearly shifted to the consumer. What can the quality and reliability department do about it? Traditional approaches for assessing consumer sentiment and measuring quality have definite flaws:

- Customer surveys are often not well designed or executed. Even if they are, the likelihood of receiving reasonable responses from the consumer base is very low.
- There is almost always a significant delay (usually in months) from the time a new product is delivered and receiving warranty claims with information on the defects.

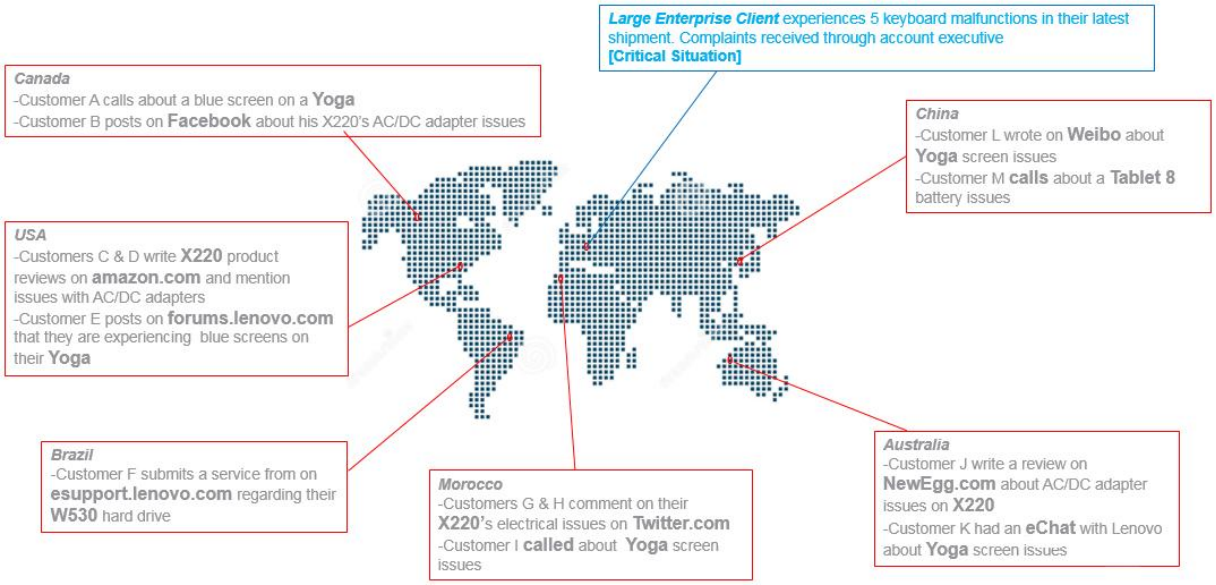
It is extremely difficult to delve into heaps of text information and comprehensively capture a customer's experience with products. The need to understand the consumers' feedback in a timely manner calls for an improved approach to understand voice of customer.

VOICE OF CUSTOMER

Lenovo didn't set out to assess the sentiment around obscure bloggers or discover new user discussion forums. Lenovo wanted to inform quality, product development and product innovation by studying data (internal and external alike). Lenovo mainly focused on supply chain optimization, cross-sell/up-sell opportunities and pricing and packaging of services. Any improvements Lenovo makes in these areas are based on listening to the customer.

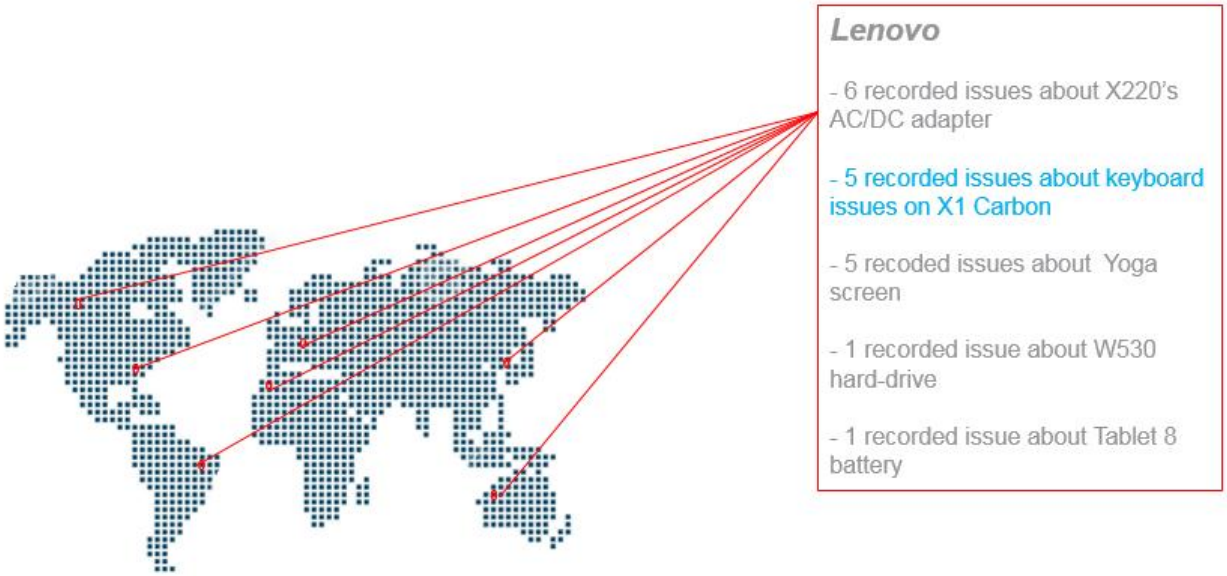
Lenovo started off by sifting through tons of unstructured data primarily by crawling the web pages and searching for content with mentions of its unique brand name – "Lenovo". This simple analysis unearthed breeding grounds of customer data such as user discussion forums previously unknown. Golden nuggets of data were found where one user for example, wrote a staggering six-page review of the current design of a Lenovo laptop, specifically the keyboard. As if it wasn't enough, that review has attracted about 2,000 comments! Traditional pre-production design reviews do not provide such level of detailed analysis or feedback. This further assured us the importance of leveraging the voice of customers' data.

In addition, the volume and the variety of the voice of customer data is rewarding and challenging. For example, Lenovo sells its product packaged with software it doesn't produce, and customers use a variety of accessories (docking stations and mouse devices) that might or might not be Lenovo products. Also, Lenovo as a company operates in 165 countries and supports more than 30 languages. Manual evaluation of the commentary can be inconsistent, can take too much time and is not scalable. It's also not possible for an individual to sense nuances among different native languages. For example, Australians describe things differently than Americans. Realizing the value of the voice of customer data and the challenge of utilizing the data fortified Lenovo's commitment to seize this information as an opportunity and started planning for its Early Alert Detection (LED) system. As shown in Display 1, Lenovo is getting data from many countries worldwide and analyze more than 7 different languages.



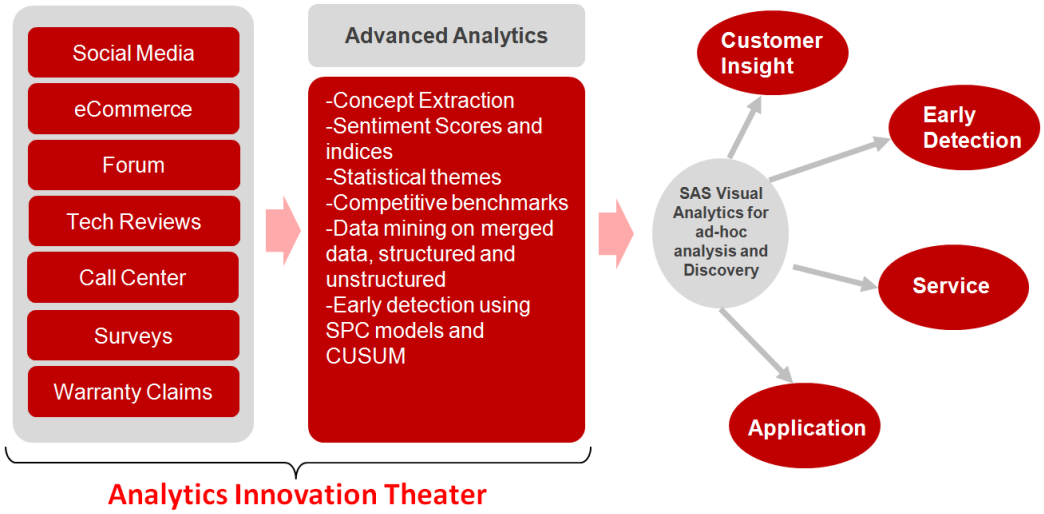
Display 1. A global snapshot showing how customer feedback comes in many forms

Our goal is to capture the four-dimension information and integrate the information into a single view (Display 2) to understand what customers are saying all over the world and how we can help them have excelled user experience with Lenovo products.



Display 2. LED Vision – A single view of quality, one voice

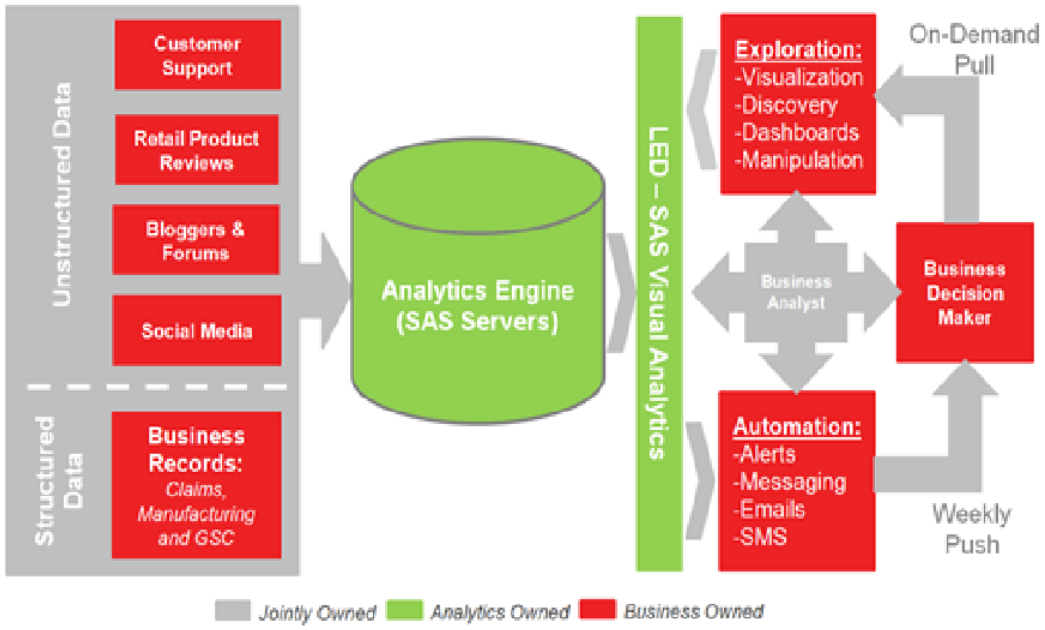
The integrated SAS eco-system (Display 3) of tools enabled us to move fast and provide innovative solution to the business. The system included BASE SAS®, SAS® Enterprise Guide, SAS® Text Miner, SAS® Content Categorization, SAS® Sentiment Analysis Studio and SAS® Visual Analytics. Our stand alone applications are easily scalable and require minimal effort to maintain.



Display 3. Lenovo's Integrated SAS Ecosystem

LENOVO EARLY DETECTION SYSTEM – OVERVIEW

Lenovo's early detection (LED) system was our pilot project using the integrated SAS system. The objective is to detect pervasive quality issues at an early stage. Integrated SAS system was designed to mine customer's perception of quality for Lenovo's and its peers' products to give us early insights into what products or parts are of concern. Using this information, we were able to look into the issue and decide if we could remove the problematic parts and reduce the cost and time for repair, as well as evaluate the potential for incorporating customer's suggestion into our product design. Display 4 shows various components that work together to generate the required customer insights.

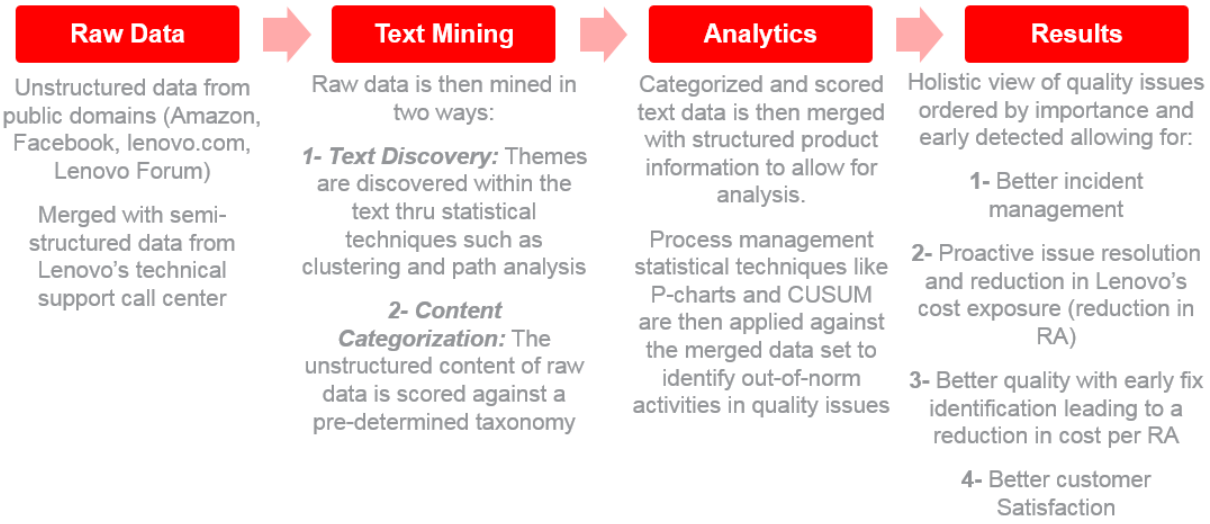


Display 4. LED Process Overview – Voice of Customer to Business Decisions

Data is collected from a number of data sources including call centers, user discussion forums, social media sources (such as Facebook, Twitter, and Weibo), e-commerce sites (such as Amazon.com, newegg.com, BestBuy.com), professional sites and blogs and survey data (Bazaarvoice.com). We

leveraged on SAS® Text Analytics suite to develop the linguistic analytical models which is the core of this entire process. SAS® Content Categorization is used for analyzing and identifying the product mentioned in the text using a Lenovo corporate standard product hierarchy classification taxonomy. SAS® Sentiment Analysis is used to analyzing the sentiment associated with the product components and/or features as described in the data.

Once we enrich the data using these analytical models, we determine how the comments and topics discussed are related to the quality issues and how to use the voice of customer output for measuring it. The measure is then used to determine the norm of the general sentiment. Control charts are deployed to measure the outcomes and determine if the voice of customer data should trigger a quality alert. In the following sections, we will discuss the details how the project is developed using our integrated SAS system. Display 5 shows the step-by-step process we put together in developing the LED system.



Display 5. LED Process Overview – Path to Value

DATA PREPARATION

Data external to the organization is gathered using web crawling. Internal data is accessed from the corporate databases using SAS® Access Engines. Data from the web includes the product information, date, review text, customer ratings and other product related contents that are available on the web pages. The type of available information may vary slightly for each data source. Internal data is primarily call center data maintained in a database and survey data that is distributed to consumers by Lenovo's survey partners.

The data sources are extracted, transformed and loaded into a standardized Lenovo Quality data mart. that transforms the unique information in the individual data source into common fields and captures the shared information in SAS (Display 6). For some data sources, we are able to access the location from where the particular product review, tweet, post, comment or survey response is originated. Using this information, we enriched the data set with the geographical co-ordinates (latitude and longitude). This information is required to develop visualizations and understand quality by geographies.

comment_date	customer_rating	origin_longitude	origin_latitude	comment_context	comment_id	comment_text	data_source
01SEP13	1.0	-122.2919000000	47.6606000000	Context={1-6 months,Window...	BZR-130901-415FA0C6F514...	Review_Text={I purchased the tablet earlier t...	Lenovo Reviews
01SEP13	1.0	-122.2919000000	47.6606000000	Context={1-6 months,Window...	BZR-130901-415FA0C6F514...	Review_Text={I purchased the tablet earlier t...	Lenovo Reviews
01SEP13	5.0	36.6499940000	-1.3666992000	Context={1-6 months,Window...	BZR-130901-7D62DC3B57A...	Review_Text={most pleasant deal. I'd recom...	Lenovo Reviews
02SEP13	5.0	-122.3042000000	47.9129940000	Context={1-6 months,Window...	BZR-130902-125537548E320...	Review_Text={I had purchased my first unit a...	Lenovo Reviews
02SEP13	5.0	-122.3042000000	47.9129940000	Context={1-6 months,Window...	BZR-130902-125537548E320...	Review_Text={I had purchased my first unit a...	Lenovo Reviews
02SEP13	5.0	-122.3042000000	47.9129940000	Context={1-6 months,Window...	BZR-130902-125537548E320...	Review_Text={I had purchased my first unit a...	Lenovo Reviews
02SEP13	5.0	-122.3042000000	47.9129940000	Context={1-6 months,Window...	BZR-130902-125537548E320...	Review_Text={I had purchased my first unit a...	Lenovo Reviews
02SEP13	5.0	-122.3042000000	47.9129940000	Context={1-6 months,Window...	BZR-130902-125537548E320...	Review_Text={I had purchased my first unit a...	Lenovo Reviews
02SEP13	5.0	-122.1134000000	47.6051030000	Context={Less than a month,...	BZR-130902-2968775E070D...	Review_Text={My only real complaint is the b...	Lenovo Reviews

Display 6. LED Process Overview – Path to Value

TEXT ANALYTICS

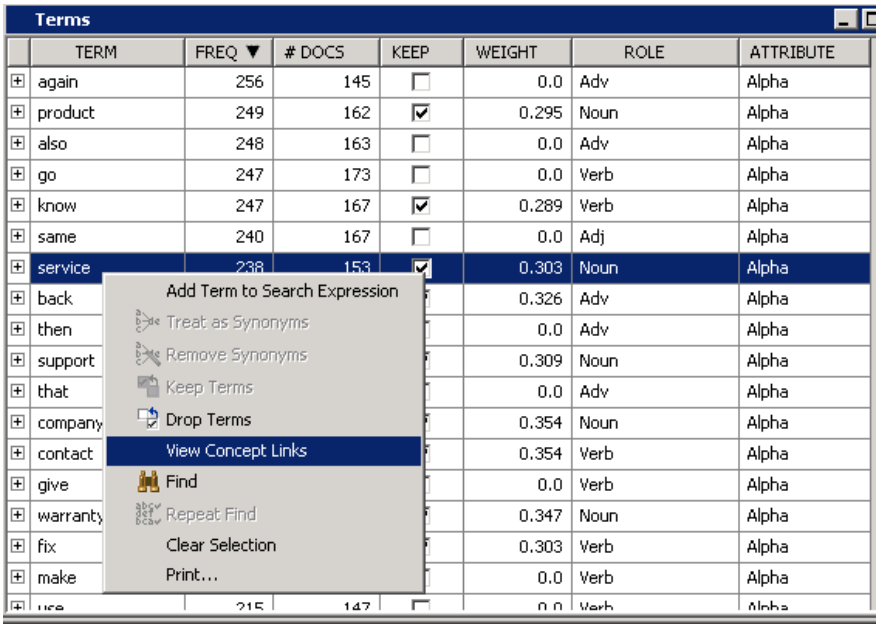
The core component of LED Voice of Customer system is the text analytics engine. It comprises of linguistic models developed in two phases. In the first phase, we performed statistical analysis (data exploration) using SAS® Text Miner to generate insights from the data to reveal underlying themes and patterns. In the second phase, Lenovo's business team which consists of Subject Matter Experts (SMEs) helped develop the hierarchical taxonomies to classify the text documents based on products (Product Hierarchy) and components (Components Hierarchy).

We used SAS® Sentiment Analysis Studio for the development of Components Hierarchy due to the fact that we intend to classify the content and at the same time understand customers' sentiment of the specific product component/feature that the customers mentioned. This is a better practice than maintaining two similar taxonomies for components in both SAS® Content Categorization and SAS® Sentiment Analysis. For example, if a problem with the laptop battery is mentioned in the review, then Sentiment Analysis project will tag battery as a component and assign negative sentiment to the review. Sentiment Analysis projects are created for different languages including English, Japanese, Spanish and Mandarin so far as we move towards the global of achieving a truly global LED system.

In the next few sections, we explain the list of steps we went through to create the text analytic process. We should always keep in mind that text mining and analysis requires a mix of intuition, subject matter expertise and some understanding of how the tool works.

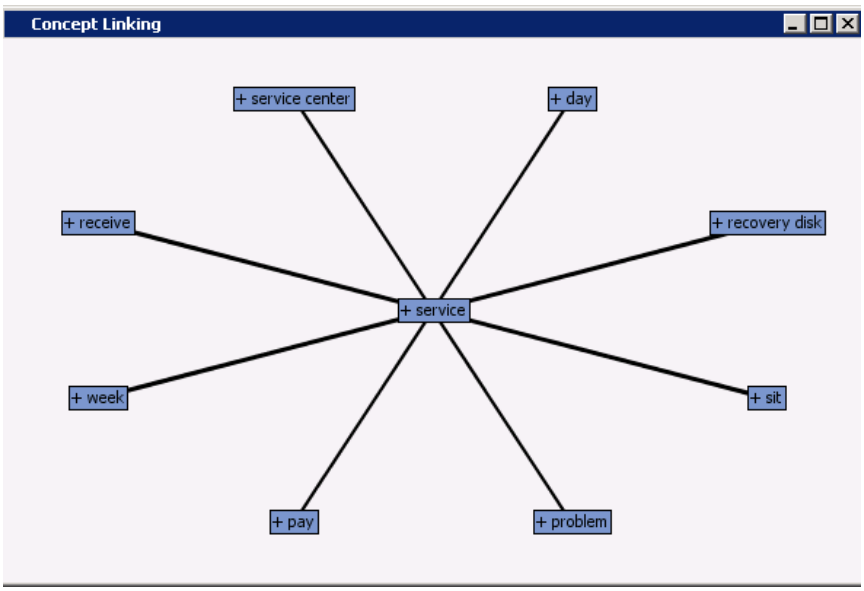
EXPLORE THE HIERARCHY STRUCTURE

We used concept links (Display 7) and text topic to explore our component and product hierarchies. Concept links feature provide a view of the concept links within the Text Filter node. It helps us to understand what the high frequency terms are in the discussion.



Display 7. Terms list in Text Filter node

For example, in the concept linking diagram below (Display 8) for the term "service", the related terms which appear most frequently along with this term appear in the diagram with connecting bars. The bars with the highest thickness represents that those two terms are strongly associated. The analysis helps us understand the root causes of issues and sometimes even more two or more issues occurring together which may not be known prior in time.



Display 8. Concept Linking in Text Filter node

Text topics are also identified in the document collection (Display 9). Text topics are different from text clusters in a sense that text topics can be regarded as the distinctive characteristic features of document collection where each document can exhibit one or more of those characteristics. A single document can have one or more text topics mapped to it. Text clusters on the other hand restricts each document to be eligible for one and only cluster. Hence a document assigned to a particular cluster is automatically ineligible for another cluster membership. The basis for both the clustering and topic extraction are the Singular Value Decomposition units generated based on the term-by-document matrix.

Topics

Category	Topic ID	Document Cutoff	Term Cutoff	Topic	Number of Terms	# Docs
Multiple	1	0.066	0.012	+order,+day,+order,+ship,+time	389	6811
Multiple	2	0.061	0.011	lenovo,+love,+lenovo,+laptop,+pc	46	3709
Multiple	3	0.056	0.012	+la,+el,+en,+para,se	394	2498
Multiple	4	0.058	0.011	+good,+luck,+good luck,+product,+lenovo	66	3634
Multiple	5	0.055	0.011	<,=,>,+look,+window	41	621
Multiple	6	0.055	0.011	+yz,+help,+private message,+detail,+hear	121	3799
Multiple	7	0.058	0.011	+love,lenovo,+lenovo,<,+product	26	2159
Multiple	8	0.049	0.011	+yoga,11s,+love,+keyboard,+pro	50	1376
Multiple	9	0.064	0.011	+lenovo,+india,+colleague,+ibm,=	72	7512
Multiple	10	0.060	0.011	+laptop,+lenovo laptop,lenovo,+lenovo laptop,+help	80	5555
Multiple	11	0.053	0.011	+want,lenovo,+helix,+check,+money	39	2971
Multiple	12	0.048	0.011	+awesome,+guy,+look,+lenovo,+wow	33	933
Multiple	13	0.055	0.011	lenovo,+thinkpad,+product,+ideapad,+love	146	5255
Multiple	14	0.046	0.011	+nice,+wow,+want,+work,+phone	24	820
Multiple	15	0.050	0.011	+tablet,+thinkpad,+thinkpad tablet,+phone,ics	111	2064
Multiple	16	0.051	0.012	+window,+install,+driver,+help,+problem	228	3149
Multiple	17	0.048	0.011	thinkpad,+thinkpad,+edge,+ibm,+keyboard	154	3134
Multiple	18	0.054	0.011	+computer,+buy,+work,+fix,+lenovo	166	2988
Multiple	19	0.056	0.011	+buy,+product,+price,+look,$	149	5754
Multiple	20	0.048	0.011	+know,+guy,+product,^,+good	117	3141
Multiple	21	0.051	0.011	+email,+gavin,gavinoh@lenovo,+help,+look	163	3849
Multiple	22	0.043	0.011	+:d,+yes,<,+wait,thinkpad	49	1012
Multiple	23	0.050	0.012	+check,+forum,+help,+find,+issue	180	4709
Multiple	24	0.052	0.012	+look,+great,+cool,+thinkpad,+yes	293	6298
Multiple	25	0.056	0.012	+problem,+work,+fix,+product,+issue	306	5810

Display 9. Text Topics generated in Text Miner

These text topics are useful for creating categories/sub-categories for the hierarchy structure where the rules can be developed to classify documents into components and sub-components. These text topics can comprehensively provide a big picture of what are the recurring issues, problems or themes that are under the discussion on public forums, Facebook or even Amazon. It also requires the analyst to perform two or three iterations of the text mining by working on the stop list, synonyms and stemming to get refined results.

LANGUAGE DETECTION

In reality, we have many data sources where we can find customer comments and reviews. To analyze such large volumes of data especially if they are from social media websites (such as Facebook, Twitter

etc.), we expect a diverse range of users from across the world and their posts in different languages other than English. An efficient way to handle that is by introducing a language detection mechanism in the process flow where we identify the language used in a text document by analyzing its content. Then we can segregate the data by language and feed them through the text analytics models (developed and maintained by each language) separately. This is a much more efficient process than to dump all data in one data set and run it through multiple text analytics model by language. For the purpose of language detection, we can leverage on **%tmfilter** macro which comes with SAS® Text Miner. It performs with reasonable accuracy in scenarios where document text is composed of terms predominantly from a single language and when those documents have at least 200 characters. SAS® Enterprise Miner Help documentation has detailed explanation on how to use %tmfilter macro for language detection.

Note: **You may contact the authors for the sample SAS® macro which can execute a language detection algorithm on an input SAS dataset and enrich the data with an additional language column in its output.**

CONTENT CLASSIFICATION

Using the SAS® Content Categorization Studio, we built classification taxonomies with Boolean Linguistic rules to analyze the incoming data from all the data sources and tag the content Product Hierarchies (L1-L4). Similarly, the Component Hierarchies (C1-C3) are generated from the Sentiment Analysis model (Display 10) along with the evaluated sentiment (positive, negative or neutral).

comment_date	customer_rating	sentiment	positive	negative	neutral	l1	l2	l3	l4	c1	c2	c3	comment_text
01SEP13	1.0	Neutral	0	0	1	NOTEBOOK	THINK	THINKPAD Tab...	THINKPAD ...	HARDWARE	BATTERY	Charge	Review_Text=[I purchased the tablet earlier t..
01SEP13	1.0	Positive	1	0	0	NOTEBOOK	THINK	THINKPAD Tab...	THINKPAD ...	OVERALLPRODUCT	PRODUCT	Overview	Review_Text=[I purchased the tablet earlier t..
01SEP13	5.0	Positive	1	0	0	NOTEBOOK	LENOVO	G SERIES	G580	OVERALLPRODUCT	PRODUCT	Recommend	Review_Text=[most pleasant deal. I'd recom...
02SEP13	5.0	Neutral	0	0	1	DESKTOP	THINK	M SERIES	M72E	PERSON	RELATION	Parents	Review_Text=[I had purchased my first unit a..
02SEP13	5.0	Positive	1	0	0	DESKTOP	THINK	M SERIES	M72E	OVERALLPRODUCT	PRODUCT	Overview	Review_Text=[I had purchased my first unit a..
02SEP13	5.0	Positive	1	0	0	DESKTOP	THINK	M SERIES	M72E	SERVICE	WARRANTY	Warranty	Review_Text=[I had purchased my first unit a..
02SEP13	5.0	Neutral	0	0	1	DESKTOP	THINK	M SERIES	M72E	USAGE	TYPE	Video	Review_Text=[I had purchased my first unit a..
02SEP13	5.0	Neutral	0	0	1	DESKTOP	THINK	M SERIES	M72E	PERSON	GENDER	Female	Review_Text=[I had purchased my first unit a..

Display 10. Text Topics generated in Text Miner

PRODUCTS HIERARCHY

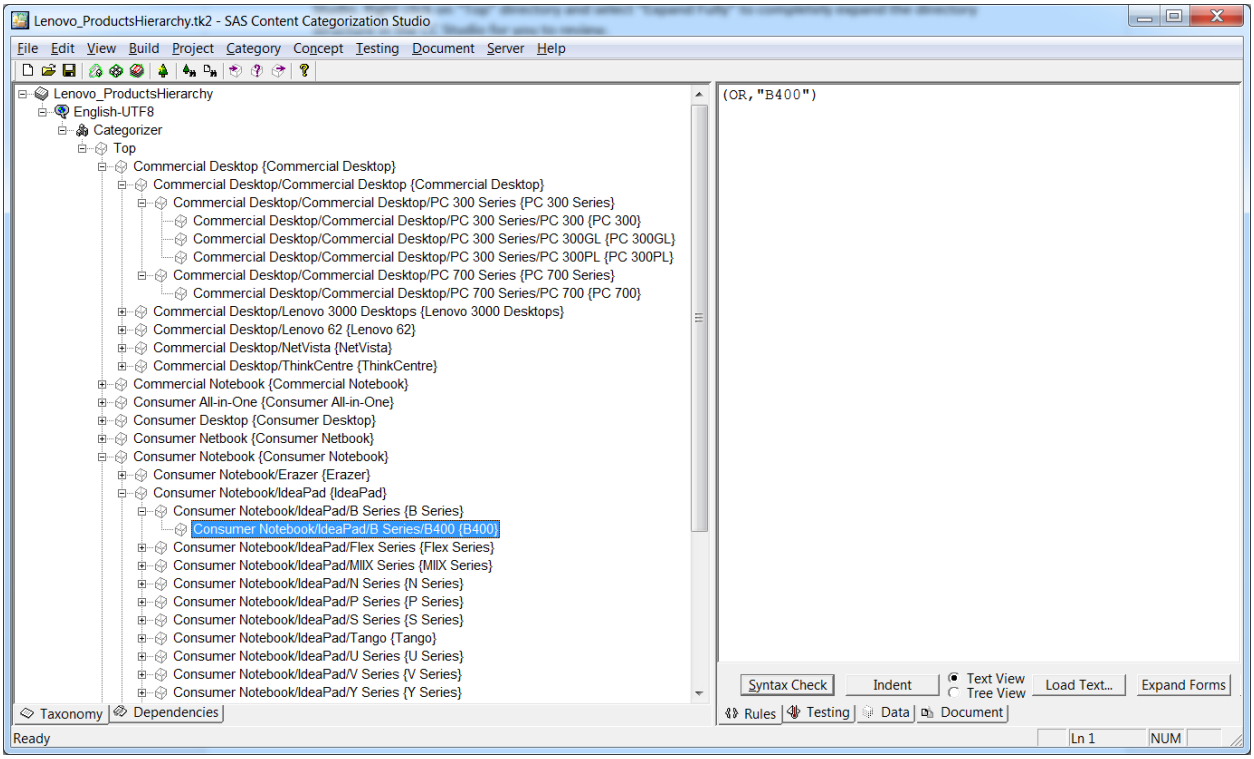
Lenovo corporate call center stores and maintains the list of Lenovo product offerings by means of a product hierarchy which constitutes approximately 17 Product Groups, 47 Brands, 262 Series and 1,337 Sub-Series (Display 11). Exact numbers may vary now since some of the non-performing or outdated brands/series are replaced by new models with latest designs and capabilities from time to time. Products hierarchy content categorization project can be created through the automated approach. In this method, we will need to have the hierarchies stored in excel spread sheets. For example, the product hierarchies for Lenovo are provided by the call center in an excel spreadsheet with 5 columns (Product Group, Brand, Series, Sub-Series and Machine ID). For the purpose of LED project, we didn't use the Machine ID since it is an internal number for Lenovo and is not necessarily known to the users thus we do not expect such data coming through from any data source. Hence, Levels 1 through 4 (also known as L1-L4) is the product hierarchy we need to use for classifying the unstructured data.

PG	Brand	Series	Sub Series
Commercial Desktop	Commercial Desktop	PC 300 Series	PC 300
Commercial Desktop	Commercial Desktop	PC 300 Series	PC 300GL
Commercial Desktop	Commercial Desktop	PC 300 Series	PC 300PL
Commercial Desktop	Commercial Desktop	PC 700 Series	PC 700
Commercial Desktop	Lenovo 3000 Desktops	Lenovo 3000 E Series	Lenovo 3000 E100
Commercial Desktop	Lenovo 3000 Desktops	Lenovo 3000 E Series	Lenovo 3000 E200
Commercial Desktop	Lenovo 3000 Desktops	Lenovo 3000 H Series	Lenovo 3000 H100
Commercial Desktop	Lenovo 3000 Desktops	Lenovo 3000 H Series	Lenovo 3000 H105
Commercial Desktop	Lenovo 3000 Desktops	Lenovo 3000 J Series	Lenovo 3000 J100
Commercial Desktop	Lenovo 3000 Desktops	Lenovo 3000 J Series	Lenovo 3000 J105
Commercial Desktop	Lenovo 3000 Desktops	Lenovo 3000 J Series	Lenovo 3000 J110
Commercial Desktop	Lenovo 3000 Desktops	Lenovo 3000 J Series	Lenovo 3000 J115
Commercial Desktop	Lenovo 3000 Desktops	Lenovo 3000 J Series	Lenovo 3000 J200
Commercial Desktop	Lenovo 3000 Desktops	Lenovo 3000 K Series	Lenovo 3000 K100
Commercial Desktop	Lenovo 3000 Desktops	Lenovo 3000 Q Series	Lenovo 3000 Q100
Commercial Desktop	Lenovo 3000 Desktops	Lenovo 3000 S Series	Lenovo 3000 S200
Commercial Desktop	Lenovo 3000 Desktops	Lenovo 3000 S Series	Lenovo 3000 S205
Commercial Desktop	Lenovo 3000 Desktops	Lenovo 3000 S Series	Lenovo 3000 S200p
Commercial Desktop	Lenovo 62	Lenovo 62 Series	Brazil 1
Commercial Desktop	NetVista	NetVista A Series	NetVista A Series
Commercial Desktop	NetVista	NetVista A Series	Netvista A20
Commercial Desktop	NetVista	NetVista A Series	NetVista A20i, PIKA A20i
Commercial Desktop	NetVista	NetVista A Series	NetVista A20i
Commercial Desktop	NetVista	NetVista A Series	NetVista A21
Commercial Desktop	NetVista	NetVista A Series	NetVista A21, NetVista A21i, NetVista A2
Commercial Desktop	NetVista	NetVista A Series	NetVista A21,NetVista A21i
Commercial Desktop	NetVista	NetVista A Series	NetVista, NetVista A22, NetVista A22p

Display 11. Lenovo Corporate Products Hierarchical Structure

In the automated process, we may use a SAS program to import the hierarchies from the excel spreadsheet to create an XML file. This XML file will hold the entire hierarchical structure contained within the spreadsheet. Once the SAS program is executed and the .xml file (English.directory.xml) is created, it can be imported into SAS® Content Categorization Studio (Display 12).SAS® Content Categorization Studio uses the imported product hierarchies to identify the products reviewed in the comments.

Note: You may contact the authors for the SAS® macro we developed to import the hierarchical structure from an excel spreadsheet into an XML file. You may need to change a few parameters as required before you can run the program.

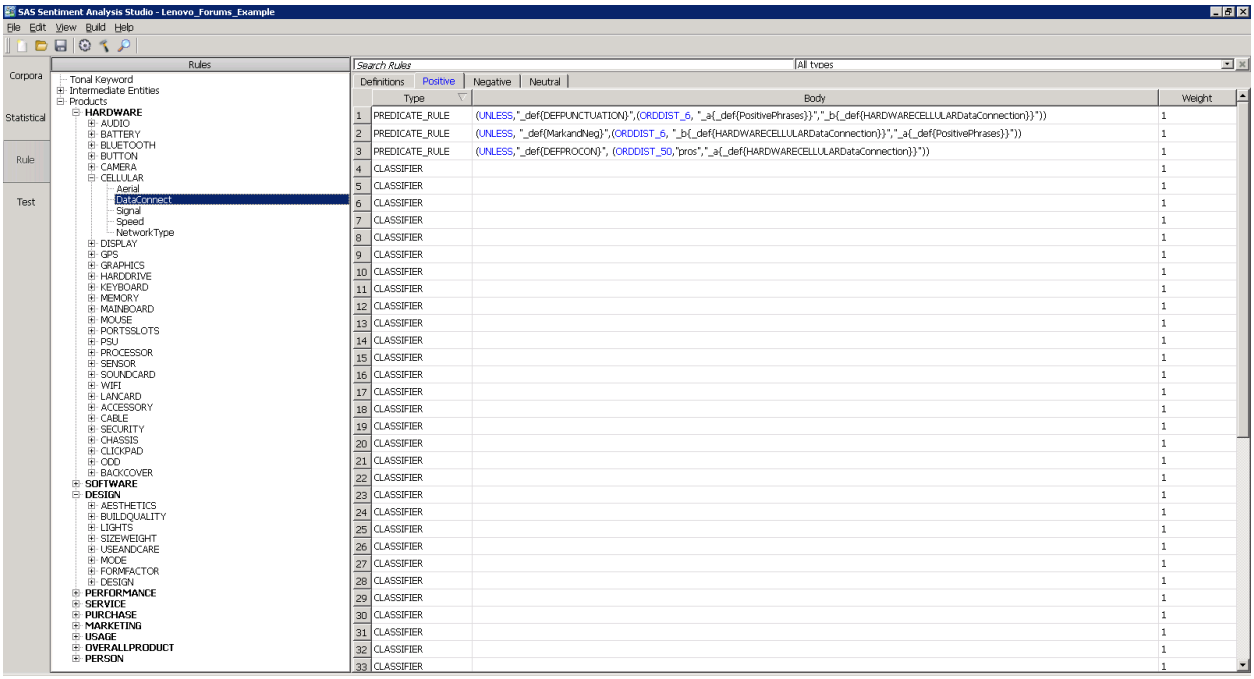


Display 12. Lenovo Product Hierarchy Taxonomy in SAS® Content Categorization Studio

SENTIMENT ANALYSIS

SAS® Sentiment Analysis Studio (SA) is used to analyze whether the consumers' feedback has a positive, negative or neutral sentiment. The Rule tab in the SAS® Sentiment Analysis shows the rules for intermediate entities, tonal keywords and the hierarchical structure for extracting the components specific sentiment and overall sentiment (Display 13). We can see that these are all Linguistic Boolean rules leveraging on the proximity and conditional occurrence of components or its synonyms along with the keywords which describe positive, negative or neutral tone in the text. Synonyms for components, phrases / symbols or slang words representing the tone of the text are maintained separately in Intermediate Entities which are used as arguments in the Boolean rules for Sentiment Analysis.

The SAS® Sentiment Analysis allows us to identify the features/components of the products that people are reviewing. It also analyzes the sentiment of the reviews. Lenovo is using this information to understand the most/least favorite features of the products and understand if there is pervasive quality issues that are revealed through the voice of the customers. This structured data is also used for additional analytic models, e.g. control chart.



Display 13. Lenovo Components Hierarchy Taxonomy in SAS® Sentiment Analysis Studio

CONTROL CHART

Product quality issues can generate significant cost for the business in terms of both lost revenue and lost reputation. Lenovo needs to find an effective way to get alerts for potential product issues at an early stage after product launch. Customers' reviews on the internet can be a good source for early quality issue detection. The challenge here is people tend to talk good and bad things freely. Consumers may not actually know the problem with the product. A complaint may not be true. Negative reviews are useful for us to foresee the product quality issues. However, we need to use it in a smart way so that we can reduce the possibility for false alerts. Control chart provides us the right tool for this.

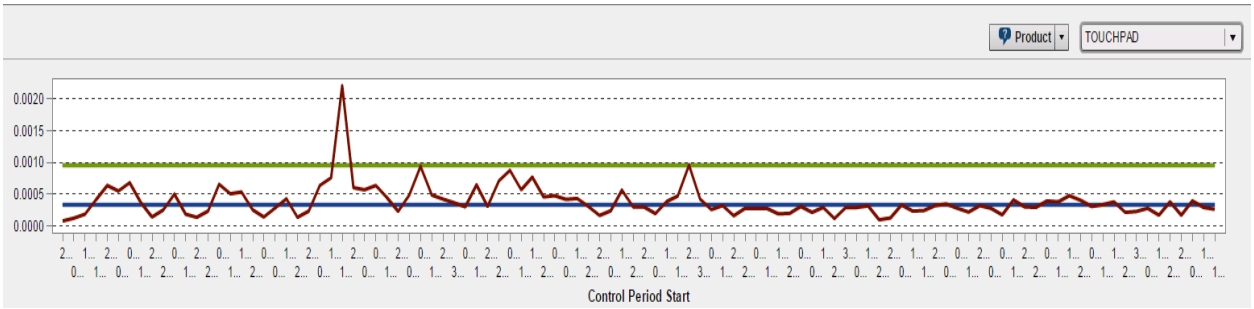
The control chart got its popularity through its application in the manufacturing industry post World War II and have been a standard quality control tool ever since. It is used to monitor if a manufacturing process is stable or in control. For example, consider a textile company that wants to monitor number of defects per square meter of fabric. The fabric is spooled onto roll as it is inspected for defects. Each piece is one meter wide and varying meters in length. The number of defects for each piece is denoted. The average defect per square meter is calculated. Then the upper bound is determined by $\bar{u} + k\sqrt{\bar{u}/n_i}$ where n_i is number of inspection units (the meters) in each piece, $\bar{u}$ is average number of nonconformities per unit (meter) taken across subgroups and *k* is the number of deviations from the mean (usually set to a default value of 3). The U-chart process is used here if the process is out of control. That is, number of defects exceeds the upper limit. In addition, we also used Cumulative Sum (CUSUM) control chart as another measure to capture the upward trends. In statistical quality control, the CUSUM is a sequential analysis technique developed by E. S. Page of the University of Cambridge. It is typically used for monitoring change detection.

In Lenovo's case, after we analyzed the product and component structure and the consumers' sentiment, we used control charts to generate alerts for quality issues. The benchmark was created by using a selected group of products within a specified period after the product launch. The upper bound is determined by using negative comment ratios and calculating its deviation from the norm. An alert is generated if the negative comment ratio of a product exceeds the upper bound.

Additionally, the control algorithm captured means of the hierarchy that are larger than the subseries average. For example, the average percentage of negative comments is calculated for each level of the subseries Yoga keyboard. If an average for a particular level (example European Market YOGA XYZ keyboard) of the hierarchy, say is over this, an alert is triggered also.

The comments can have either a positive or negative sentiment. We need to apply solid metrics to help avoid two common traps when determining if the number of negative comments is benign or potentially troubling. The first trap is that there will always be negative comments and there will always be negative comments because of human nature – some people will always complain, how do we weigh the negative against the positive. The second trap is to selectively discount the negative comments and only remember the positive. Luckily there is a methodology that quantifies negative versus positive comments and can provide early signal of huge shift to the negative and allow proactive action to address the negative comments. The process captures a number of different trend patterns that have negative comments trending upward (out of control). These levels can be proactive addressed.

U-charts (Display 14) provide weekly means to 'detect' the number of negative comments if the number is shifting upwards (trending upwards over a select number of weeks) or spiking negatively (outside 3 standard deviations of the norm for the particular level of the hierarchy) or outside a predefined limit.



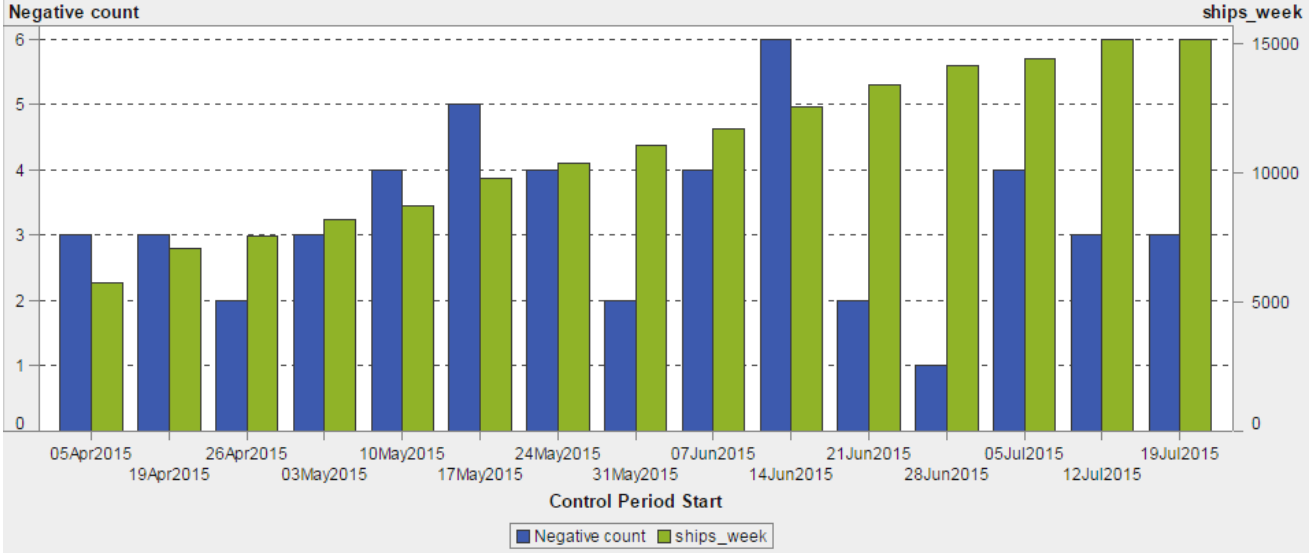
Display 14. A sample U-chart which shows the trends for number of negative comments

Typically, the following rules (often referred to as Western Electric rules) are used to decide if an alert is generated.

We selectively picked rule 1, 2(above mean) and 3(increasing only) in our case to focus on extreme shifts in patterns.

Test Index	Pattern Description
1	One point beyond Zone A (outside the control limits)
2	Nine points in a row in Zone C or beyond on one side of the central line
3	Six points in a row steadily increasing or steadily decreasing
4	Fourteen points in a row alternating up and down
5	Two out of three points in a row in Zone A or beyond
6	Four out of five points in a row in Zone B or beyond
7	Fifteen points in a row in Zone C on either or both sides of the central line
8	Eight points in a row on either or both sides of the central line with no points in Zone C

We also calculate the cumulative sum (CUSUM) of negatives comments over time and forecast the expected sum in the future (Display 15). If the number of negative comments strays (above) outside a limit (upper confidence limit of forecast), this signals that the numbers of negative comments is potentially trend out of control. The CUSUM is useful to capture data quality issues with sparser data than the traditional control charts.



Display 15. A sample CUMSUM chart showing trends of cumulative negative comments

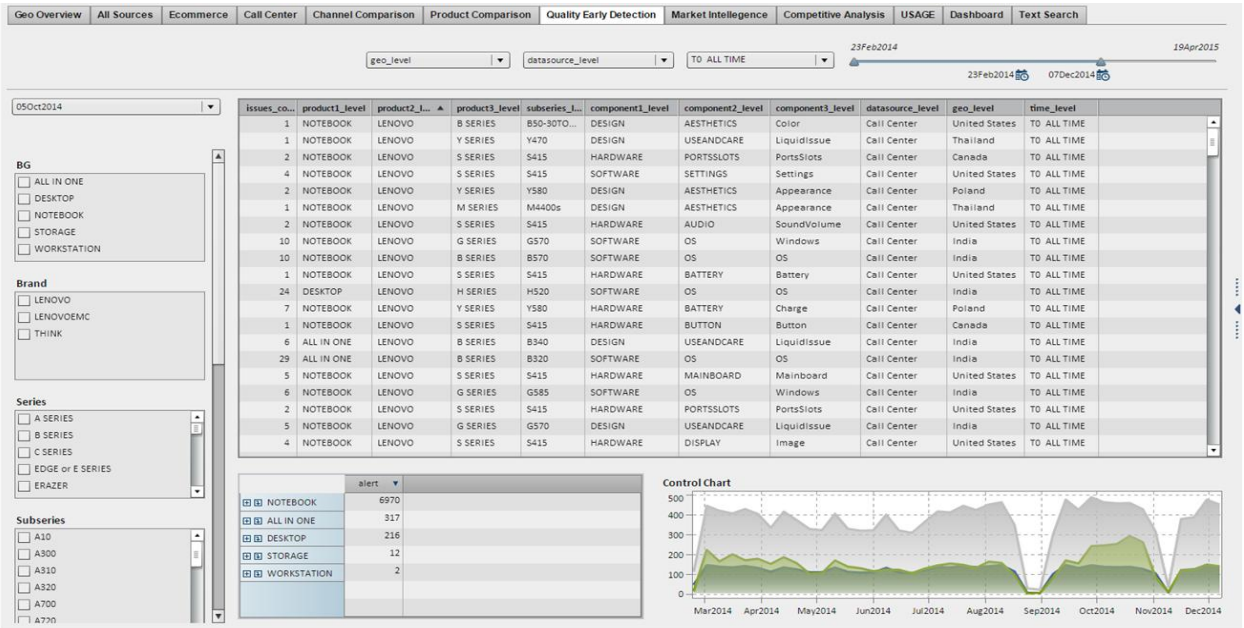
VISUAL ANALYTICS

Now that we have the result created in flat datasets, we need to be able to show case the result and get buy-in from the business users. SAS® Visual Analytics (VA) is a powerful tool to help the business users quickly adopt the analytic methodology and generate business insights. Majority of the people is able to absorb the massive information better through visualization than flat data files. VA is used to neatly demonstrate the analytic results with stories and interactive graphs.

We move the processed dataset to our LASR server in VA through Hadoop. The in-memory reading technology makes it possible to visualize large datasets with very short load time. The alerts are automatically generated and displayed on the dashboard as the data gets refreshed weekly. The business users can easily visualize the top issues associated with different products, drill down to details and ask what-if questions to gain insights to the quality problems. The users will discuss the analytic result with the quality team and decide if actions need to be taken. They will, in return, provide us feedback to further improve the visualization report and better fit business needs. An overview of the data sources and product hierarchy is provided (Display 16). The users can filter on a number of constrains to check different product combinations and timer periods.

Display 16. Lenovo Early Detection – Executive Dashboard Summary

The control chart (Display 17) shows the upper bound and mean of the Negative Comment Ratio (NCR). NCR can be defined as the total number of negative comments divided by the total number of products shipped. We can easily see from the graph that how the weekly NCR is compared to the norm and determine if there is an upward trend in the customers' feedback regarding certain product or component. The dashboard also allows the user to drill down the product hierarchy and identify the root causes of the alerts.



Display 17. Control chart showing trends of negative comments ratio (NCR)

RESULTS

The analysis-driven discovery of an issue with docking stations provided the second big win for Lenovo's LED initiative. It wasn't until Lenovo used SAS to analyze the combination of call center notes and social media posts that the word "docking" was connected to the problem. This helped quality engineers to

figure out that root cause of the issue is a software update. We were able to pick up that feedback within weeks and quickly work on the solution. It used to take 60 to 90 days because we had to wait for the reports to come back from the field. Now it takes just 15 to 30 days. That reduction in detection time has driven a 10 to 15 percent reduction in warranty costs. As warranty claims cost the company about $1.2 billion yearly, this is a significant saving.

While the call center information was crucial, the social media component was what sealed the deal. With Twitter and Facebook, people described what they were doing at that minute, 'I docked the machine and X happened.' It's raw, unbiased and powerful. An unforeseen insight was found when analyzing what customers were saying as they got their PCs up and running. Lenovo realized its documentation for its products and warranties was unclear. With the improved documentation, Lenovo saw a 30 to 50 percent reduction in calls coming in for general information. This significantly reduced the call center operating cost for the Lenovo.

SAS provides the framework to "manage the crazy amount of data" that is generated. The project's success has travelled like wildfire within the organization. Lenovo initially planned on about 15 users, but word of mouth has led to 150 users signing up to log in to the LED and other voice of customer dashboards for visual presentations.

The results have been impressive:

- Over 50 percent reduction in issue detection time.
- 10 to 15 percent reduction in warranty costs from out-of-norm defects.
- 30 to 50 percent reduction in general information calls to the contact center.

SUMMARY

The application of analytics has ultimately led us to a more holistic understanding of the concept of quality. Quality isn't just a PC working correctly. It is people knowing how to use it, getting quick and accurate help from the company, getting the non-Lenovo components to work well with the hardware and understanding what the customers like about the existing product – rather than just redesigning it because product designers think it's the right thing to do. SAS® Text Analytics tools made it possible to make use of massive internet and company data and quantify the text information. The early detection tool helps the business to narrow down the search for potential quality issues. This allows the users to better allocate resources to research key quality issues, provide timely solutions and save cost. SAS® LASR server makes it possible to visualize large dataset in SAS® Visual Analytics.

CONTACT INFORMATION

Mohammed Chaara, Director of Customer Insight & VOC Analytics, Lenovo (mchaara@lenovo.com)

Mohammed Chaara is the Director of Customer Insight & VOC Analytics at Lenovo U.S. Headquarters in Morrisville, NC. He was the lead strategist and project manager for the Lenovo Early Warning Detection – Perceptual Quality project. He has rich experience in Analytics and as a strategy leader with more than ten years of business experience specializing in advanced analytics, marketing strategy, analytics resource and vendor management.

Yingrui Liu Olesiuk, Lead Data Scientist, Lenovo (yolesiuk@lenovo.com)

Yingrui Liu Olesiuk is a lead data scientist of Customer Insight Center of Excellence (CICOE) at Lenovo. She manages text analytic projects for data acquisition, analytic processing and advance analytics. Prior to joining Lenovo, she was a senior analytic consultant at Cardinal Health. She led the development and implementation of the analytic framework for Cardinal Health's order monitoring system. She is currently working on improving customer experience utilizing voice of customer data.

Rob Carscadden, Principal Data Scientist, SAS (rob.carscadden@sas.com)

As a data scientist, Dr. Rob Carscadden has been with SAS Institute for over 15 years and has helped SAS customers get value and insights from their data using analytics. Technical areas of expertise include statistics, data mining, text mining, forecasting, quality control and predictive modeling. He has

worked with customers from the public sector (federal, state and local) and private companies tackling problems of fraud, waste, abuse, customer retention, cross/up selling, quality assurance and forecasting.

Murali Pagolu, Analytical Consultant, SAS (murali.pagolu@sas.com)

Murali Pagolu brings over 10 years of cross-industry professional consulting experience in Information Technology and Advanced Analytics. He has expertise in Data Mining, Text Mining, Consumer Segmentation, Predictive Modeling, and Statistical Analysis. He is also a SAS published author for the book “Text Mining and Analysis: Practical Methods, Examples, and Case Studies Using SAS”. He led several project implementations successfully in a variety of applications such as Anomaly Detection, Information Retrieval, Content Classification, Information Extraction, Sentiment Mining, Early Warning Detection and Voice of Customer Analytics.

REFERENCES

Chakraborty, G., M. Pagolu, and S. Garla. 2013. *TextMining and Analysis: Practical Methods, Examples, and Case Studies Using SAS*®. Cary, NC:SAS Institute Inc.

Company, W. E. (1956), *Statistical Quality Control Handbook*, available from Western Electric Company, Commercial Sales Clerk, Select Code 700-444, P.O. Box 26205, Indianapolis, IN 46226.

Grant, E. L. and Leavenworth, R. S. (1988), *Statistical Quality Control*, Sixth Edition, New York: McGraw-Hill

Kume, H. (1985), *Statistical Methods for Quality Improvement*, Tokyo: AOTS Chosakai, Ltd.

ACKNOWLEDGMENTS

Thanks to our colleagues at Lenovo Ken Seethaler, Jay Johnson, Ronald Mitchell, Haomin Zhang and Jane Jiao and at SAS Revi Listiadji, Brinda Gupta, Laura Beyer, Tony Fountain, Jeff Brown and Mike Hitmar for their contributions to make this project a success.

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.