Paper 074-31

Eighty Ways to Spell Refrigerator

Joshua P. Becker, Manager of Reliability, Sub-Zero Freezer Company, Inc.
John Wallace, Principal, Business Researchers Inc.

Abstract

A methodology that combines text mining, statistical process control and a balanced scorecard eliminated inaccurate manual claim coding and a significantly reduced the time required to identify root causes of field failures. As a result annual warranty claim costs have decreased, product reliability has increased, and customer satisfaction has improved.

Business Case

Sub-Zero accepted that their legacy system of collecting, analyzing and learning from corporate warranty and call center data was inadequate. It was obvious to data analysts and engineers at Sub-Zero that the keys to improving metrics such as product quality, product reliability, and operating costs associated with each, resided in the time required to learn from the data. This time period, referred to as the *definition timeline,* is explained in Figure 1 as the time associated with identification and characterization of a valuable concept within the data when one exists.

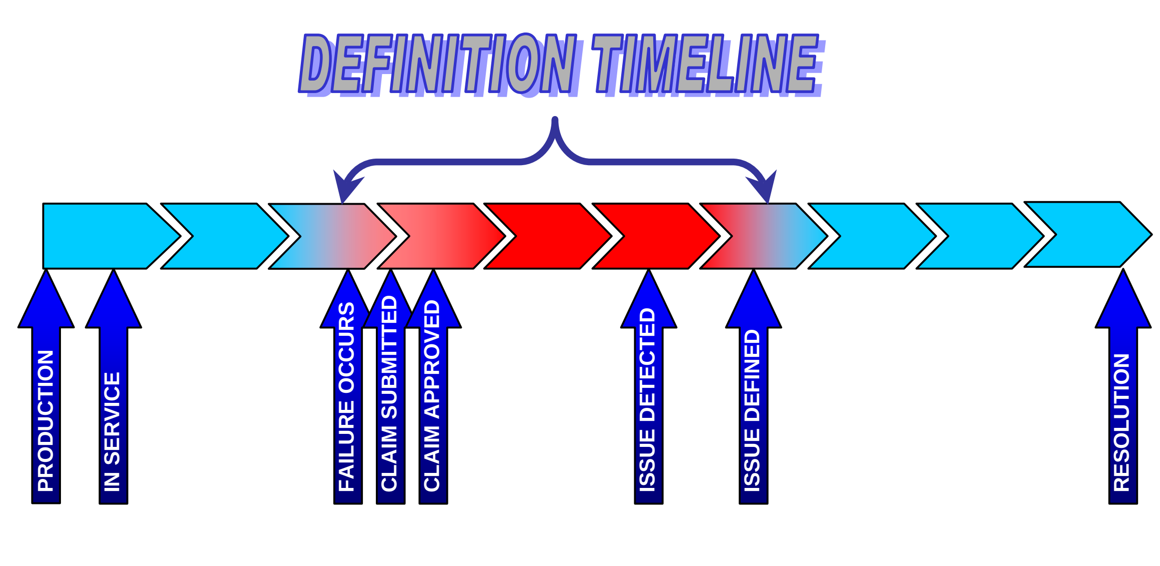

Figure 1: The Definition Timeline

Legacy Claims Processing

Conceptually, the legacy process in place at Sub-Zero was very common to most consumer durables manufacturers. Both paper and electronic claims were "unpacked", keyed in, and read. Next a non-technical claims coordinator made two business decisions: 1) a manual codification decision and 2) a payment decision. The codification decision was based on reading two text fields and the parts listed on the claim. The coordinator selected a code or codes from a list of approximately 250 warranty claim failure codes. Finally, the claims processing software applied a series of programmatic algorithms to determine if the claim was paid or rejected. At that point, the claims data existed in the corporate data warehouse and was available for analysis.

Process and Analytical Deficiencies

Deficiencies in the legacy process stemmed from the early stages of claims processing and propagated through the entire decision making channel. The root of imperfection was the manual codification process. A large margin of error was introduced at this juncture due to ambiguity and misinterpretation of technician's comments. In addition, the possibility existed that numerous parts were replaced on a claim required the coordinator to make a decision as to which part or parts should drive the codification. Moreover, selecting a code from a list of 250 possible codes bred inconsistency in the choices that coordinators made. Claims were often miscoded or lumped together in the general failure codes that were easier to remember and interpret, essentially masking the valuable failure information. Lastly, a substantial lag between the time the claim arrived and when it was available for an engineer to analyze was generated due to the time consuming nature of keying in, reading, and manually coding warranty claims.

The legacy process was also analytically inefficient. True engineering issues were confounded in a mix of miscoded and generalized warranty claims. Increasing trend lines or elevated metrics related to any given failure code only provided a meager warning of a potential issue. The root cause identification process was extremely time consuming, laborious, and cost prohibitive. Engineers needed to spend their time problem solving, testing, and implementing expeditious resolutions to field failures, rather than sifting through countless rows of data in an effort to define anomalous field issues.

Objectives

In order to improve upon the process described above, Sub-Zero embraced the idea of using analytical software, automation, and text mining to drive learning from their data. In terms of a business objective, the new system needed to:

1) Significantly shorten the definition timeline
2) Eliminate the manual coding process
3) Sharply reduce the need for engineers to read claims

The analytical objectives were to:

1) use a bottom-up analysis approach
2) derive meaningful groups of like claims (clusters) through data-driven categorization of text records
3) capture each claim's parts and failure modes
4) automatically track changes within these clusters and generate statistically significant alerts to emerging field issues

Approach

The approach used on this problem combines three analytical techniques: Singular Value Decomposition (SVD), Expectation-Maximization Clustering and Statistical Early Warning Methodologies. This paper focuses on the text mining portions over the early warning efforts of the project. SVD is the data preparation used to add structure to unstructured text. It uses linear algebra to decompose a m x n sparse matrix into its least squares best fit m x K matrix where K is a user specified number of dimensions. Those dimensions can then be used for tasks like document clustering and document classification. SVD is not used for tasks like part-of-speech tagging or entity extraction. Given the quality of our data, no natural language processing tasks were employed.

Related Work

Berry details the evolution of extracting information from textual data since the 1970s. One particular widespread technique for analyzing textual data is Latent Semantic Indexing (LSI). Deerwest et al. presented a seminal work on LSI. The technique is a text mining algorithm based on Singular Value Decomposition (SVD).

Given the advances in extracting information, some researchers began to analyze the extracted information. The National Institute of Standards in Technology (NIST) sponsored competitions on trend detection in textual newsfeeds called TDT-3 for Topic Detection and Tracking. Feldman and Dagan presented their work at the KDD conference in 1995. Lent and Aggrawal applied their previous work in sequential pattern mining to text data in 1997.

Meanwhile, other field reliability researchers were making advances in analyzing observational data. Wu and Meeker published their approach to analyzing the fields in warranty data using a multivariate model that manages three timelines simultaneously.

Data Overview

Warranty Claims

This data mainly consists of free form text. Field technicians are required to fill out several fields on a standard warranty invoice. Two of these fields are named “Customer Request” and “Service Performed”, and are filled in using the technician’s own words. These fields are meant to capture the complaint the customer had about the product and the actions taken by the technician to resolve the issue, respectively. Clearly the potential for misspellings, slang terminology, industry jargon, and improper grammar is high. An example of these fields is provided in Figure 2.

Customer Request	Service Performed
DOORS ARE NOT ALIGNED PROPERLY THE RESSZER HAS A DIFFERENT SIZE GASKET THAN THE REFER	FF DOOR STICKS OUT FARTHER THAN THE FREEZER DOOR. FZR DOOR 2 1/2"; FF DOOR 2 3/4". ADJ'D DOORS. DIDN'T MAKE ANY DIFFERENCE. WILL CALL TECHLINE & WILL CALL CUST.
ICEMAKER DOESN'T MAKE ICE ICE MAKER NOT MAKING ICE AND THE WATER IS FREEZING ON THEBACK PANEL INSIDE OF FREEZER.	DEFROSTED FILLTUBE ALSO INSTALLED FILL TUBE ALL THE WAY INTO SOCKET HEATER FOR FILL TUBE REASSEMBLED UNIT TEST RAN AND VERIFIED PROPER OPERATION. Total Time = 1
i/m makes too much ice	wty pts order complete installed i/m extension arm test ok. it was broken
i/m arm broken	installed rudder arm

Figure 2: Warranty Free Form Text Fields

A third field included for analysis, shown in Figure 3, is one that contains the structured text names of the parts replaced during the service call. The data contains limited misspellings, slang, jargon, and grammar issues. However, historical part naming conventions within the organization make the part names fairly convoluted and difficult to text mine.

Partname	Partname
VALVE WATER 3090020	BLADE FAN 5-PROP 6-3/4"
MOLDING ASSY SS HANDLE	SWITCH LIGHT
TERMINATOR DEFROST	MOTOR ASSY REPLACEMENT FRE FAN

Figure 3: Warranty Structured Parts Field

Call Center

Call center data also contains two free form text fields that are entered at the time of contact. These fields, called "Summary" and "Description" are similar to those in the warranty data. In the case of a field technician calling in for assistance, the summary and description fields are completed at the time of contact to capture a summary of the issue, and a more detailed description of the root problem and resolution. An example of these fields is shown in Figure 4. Misspellings and improper grammar run rampant within these text fields. However, slang and industry jargon are not as pervasive as in the warranty claims data.

Summary	***Description***
PL to replace cond fan motor, and/or comp.	Humming noise from comp or cond. Try fan motor first.
IM problem. On RO. Refrigerator crisper drawers stick.	Customer will contact Red and White to look at crispers. Drawer jamed so that door couldn/t close.
unit refer down autho service can't come out for >>>	>> over a week cust may use unautho company we would send ss parts if they are what needed.
Gasket is ok - some whistling noise	Left mess. that it is normal. Cust is still upset with whitling, will have Quan inspect. Left mess for Quan. Quan checked door - door is normal.

Figure 4: Call Center Free Form Text Fields

Data Preparation

As seen in Figure 5, the data preparation activities dominate the modeling process. Inspection of the warranty claims data revealed that the distinction between the two text fields was not always respected. We ultimately concatenated the two fields together into a larger one. Another source of text for each record came from the descriptions of the replaced parts. The part description text was particularly difficult as it was not written using proper sentence structure. For example “COMP, REF.” actually stands for “refrigerator compressor”.

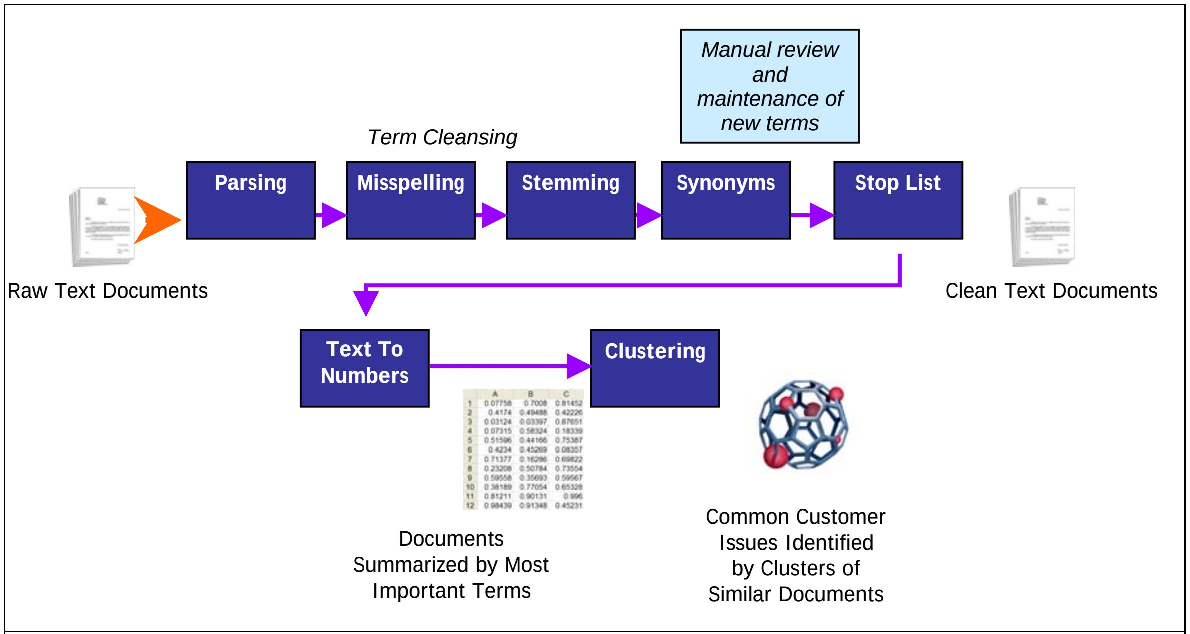

Figure 5: The Text Mining Process

Misspellings

As previously indicated, the quality of the text provided in warranty claims is quite different from the English found in a novel. For example, one common word, compressor, had 52 different spellings. Here a misspelling is defined as a term that is not in the English dictionary. These unmatched terms were separated from the terms that are in the dictionary. Then each misspelled term was compared to all of the correctly spelled terms. The SAS function SPEDIS, provided the fuzzy matching necessary to suggest which correctly spelled word was the best match to the incorrectly spelled one. At this point, the matched list was reviewed by an analyst. For some close matches, the recommended closest term was always taken, for others the analyst would override the recommendation. This review process was conducted in Excel where automation was written in Visual Basic.

Synonyms

One very important step of the data preparation process is the development and application of a domain specific synonym list. This aids the subsequent model by removing the need to rationalize the relationship between words that have similar meaning by providing the relationship ahead of time. For example, in the case of refrigeration, both “gas” and “Freon” have the same meaning.

The synonym list is also used to store the relationship of adjacent words that are really one term. These compound nouns (i.e. compressor bracket) are also referred to as noun groups. This step is required because a term is defined as a string within a larger string that is separated by spaces. With technical data, noun groups are very common. We used a first order Markov transition matrix to find terms that co-occurred sequentially. This technique speeds up identification of compound nouns considerably.

In the end, the 400,000 documents chosen for analysis had 30,000 distinct terms. When the terms that appeared more than once were removed, this left 20,000 terms. Roughly half of those terms were deemed misspelled and matched against the other half. Both the final misspelling list and the final synonym list have 10,000 terms. Applying these two lists and the stop list leaves about 3,500 terms left to analyze.

Model Building

The modeling strategy employed was a descriptive model based on clustering. The objective was to create groups of records that were all about the same topic (homogenous). In this case, the usefulness of a cluster is subjective but focuses on the homogeneity of the documents within it. The domain expert assesses the homogeneity of a cluster as to how well that cluster describes a particular engineering problem. For example, if every document in a cluster refers to the door gasket, it is more homogenous then a cluster with documents about the door gasket or the compressor.

Since a document collection may have hundreds of different topics discussed in a one-year period, tens or hundreds of clusters may be necessary in order to achieve homogeneity within a cluster. It was recognized early on that the default clustering model would not meet expectations. In fact, it was determined that a *hierarchy* of models would be required to satisfy our needs. Based on domain expertise, clustering models were tuned for homogeneity using two techniques: subclustering and merging clusters.

Subclustering can be defined as using the documents assigned to a given cluster in a first model as the modelset for a subsequent model. The subcluster model had fewer terms then the first model and therefore allowed documents that were once viewed as similar to be grouped separately. In the case of a poor cluster, subclustering was necessary because the parent cluster lacked homogeneity as in the door gasket and compressor example above. In other cases, a cluster that was clearly about noise was broken into different types: squeak, rattle, etc. Merging was used if the model identified a cluster about recharging and another cluster about adding Freon.

Multiple Dimensions

Each warranty claim has three primary dimensions to the text: 1) The part or parts that failed 2) the failure mode 3) the corrective action taken. The analysis had for its primary goal to model the failure mode and failed part dimensions. This objective was met by employing two separate clustering models to develop dual taxonomies.

Multiple Cluster Assignments

In general, clustering algorithms have a common assumption of mutual exclusivity: a case can belong to only one cluster. The Expectation-Maximization algorithm used to build these models assigns a probability of belonging to every cluster. These multiple probabilities in practice, however, did not provide assignments to multiple clusters. The probability distribution across clusters for any given document was always in favor of the best cluster (.95 or higher). The approach used to create multiple clusters was programmatic and produced results such as those in Figure 7.

As indicated in the table above, in some cases a part or failure mode was not detected and that assignment would remain null. Lastly, one may notice that in the first record in Figure 6 the door handle is bent, not the magnetic strip. Since the analysis leverages a collection of documents, the frequency of *magnetic strip – bent* documents can be calculated and easily discarded as an anomaly. On the other hand, even if the relationship did not make complete sense but occurred very often, its statistical significance would still warrant investigation. Perhaps the door handle was bent because of the magnetic strip and statistical process control would be an excellent tool to find such an outcome.

Part A	*Part B*	*Failure Mode*	*Raw Text*
MAGNETIC TRIM STRIP	DOOR HANDLE	BENT	HAVING PROB WITH MAG STRIPS FOUND FREEZER HANDLE BENT AND LAYING FLAT REPLACE HANDLE MOLDING ASSEMBLY HANDLE
HINGE	HINGE	NOISE	UNIT NOISE REPLACE PARTS HINGE TOP HINGE BOTTOM
CONTROL BOARD	ICEMAKER PUMP	INSUFFICIENT ICE PRODUCTION	NO ICE REPLACED BOARD PER TECH LINE NO CHANGE REPLACED PUMP BOARD PC PUMP COMPLETE WATER
WATER VALVE	ICEMAKER	NOISE	ICEMAKER MAKING A REAL LOUD NOISE REMOVE BAD PUMP AND INSTALL NEW CK UNIT VALVE WATER
SYSTEM TUBING		NOISE / VIBRATION-BUZZ	VERY NOISY ADJUST TUBING RATTLE

Figure 6: Multiple Part and Single Failure Mode Assignments

Cluster Naming

A cornerstone of the modeling process is applying meaningful names to the clusters that come out of the algorithm. This step is an opportunity to really understand how the model is working. When clusters are named, two or more clusters with same name are treated identically. Some example cluster names are provided in Figure 7 and Figure 8 shows some actual records form the Condenser Fan Motor cluster shown in Figure 7.

Cluster Name	*#*	*%Tot w/ Term1*	*Term1*	*Term2*	*Term3*	*Term4*	*Term5*
CONDENSER FAN MOTOR	943	51%	condenser fan motor: 485	condenser fan: 395	blade: 353	fan blade: 326	plastic: 323
DISPENSER LIGHTS	442	73%	dispenser: 322	light: 303	bulb: 178	door: 176	light bulb: 133
WIRE CONNECTION	189	41%	wire: 78	connect: 76	disconnect: 58	power: 52	cord: 35

Figure 7: Example Clusters and Terms that Define Them

Customer Request	*Service Performed*
Freezer not cooling	removed test leads from condneser fan blade locked motor down
Makes a loud noise	Adjusted condenser fan motor to eliminate rattle. Adjusted doors to meet at top
Freezer not cooling	Removed and replaced cond fan motor
GASKET IS EXTREMELY HOT	INSTALLED CONDENSER FAN MOTOR AND CHECKED FOR PROPER OP

Figure 8: Example Records from the Condenser Fan Motor

There is no guarantee that a document clustering model will be able to create rational groups out of every single record in the analytic dataset. In some cases, a record may be poorly expressed or may be an anomaly. These records accounted for 3-5% of the data.

At this point the categorized records were ready to be aggregated and monitored using early warning techniques discussed in Wu and Meeker.

Conclusion

Since completion of the text mining model, numerous years of historical warranty claims data has been scored. The model has proven to be very effective. A comparison of the model output to the historical failure codes assigned by Sub-Zero personnel showed strong correlation. In some differences, it was the manual failure code that was wrong and the model had assigned the claim to the proper cluster. It is clear the days of manual coding and reading claim detail are numbered. Presently Sub-Zero estimates that the automated process will shorten the problem definition timeline by an estimated 64 days.

References

Berry, M. (Ed.). 2003. Survey of Text Mining: Clustering, Classification, and Retrieval. Springer-Verlag.

Lent, B.; Agrawal, R.; Srikant, R. Discovering Trends in Text Databases, 3rd Int'l Conf. on Knowledge Discovery and Data Mining, California USA, 1997.

Deerwest, S. Dumais, S.T., Furnas, G.W.; Landauer, T.K.; and Harshman, R. 1990. Indexing by latent semantic analysis. Journal of the American Society for Informations Science 41(6):391-407.

R. Feldman and I. Dagan. Knowledge discovery in textual databases. In Proceedings of the First International Conference on Knowledge Discovery (KDD-95). ACM. Aug 1995.

Kontostathis, A.; Indro D.; Holzman, L; Pottenger, W. (2004). Use of Term Clusters for Emerging Trend Detection. Preprint.

Tjoelker, R.; Poteet, S. Integrating Text Mining and Data Mining for Engineering Applications. Spring Research Conference on Statistics in Industry and Technology, The University of Dayton, June 4–6, 2003.

Wu, H.; Meeker, W (2002), Early Detection of Reliability Problems Using Information From Warranty Databases. *Technometrics*, **44**, 120-133.

Bios

Josh Becker is the Manager of Reliability for Sub-Zero Freezer Co., Inc. and Wolf Appliance Company, LLC. His core responsibilities include reliability improvement, reliability assurance, and statistical analysis of warranty and call center data. He has developed several corporate reporting tools designed to disseminate reliability data throughout the organization and empower employees to make reliability conscious decisions. He is also the system architect of the organization's current warranty analysis and service incidence rate tracking system.

His industry experience is concentrated in the manufacturing and building systems consulting industries. He has been involved in several major software system implementations with Sub-Zero including ERP, Call Center, Electronic Warranty Claim Processing, and Warranty Analysis. Most recently he is acting project manager for a large scale SAS implementation. He holds a BS in Mechanical Engineering from the University of Wisconsin - Madison, an MBA in Technology Management from the University of Phoenix, and is a Certified Reliability Engineer through the American Society for Quality.

John Wallace is a principal consultant and founder of Business Researchers, Inc. He has a track record at applying data mining techniques to business problems. His work has included designing applications to analyze customer profitability, customer loyalty, product profitability and product quality. He has worked as an analytical consultant at SAS Institute as well as a Business Intelligence analyst at UUNET.

His consulting experience has included working for clients in the automotive, financial services, ISP, grocery, wireless, retail, PC/server and consumer software industries. He has successfully managed application development teams, created system architecture, developed new analytical methodologies and estimated complex models. He has leveraged techniques including text mining, response modeling, segmentation and survival data mining. He holds an MBA in Decision Science from the George Washington University.

Josh Becker
SubZero Freezer Corporation
4717 Hammersley Rd.
Madison, WI 53711
Ph: 608-270-3395

John Wallace
Business Researchers, Inc.
2082 Union Street
San Francisco, CA 94123
Ph: 415-377-4759