CONTROL CHARTS FOR CUSTOMER COMMENTS: A CASE STUDY AND A RESEARCH AGENDA

Triss Ashton
University of North Texas
College of Business
1155 Union Circle #311160, Denton, TX, 76203-5017
Phone: 940-369-8379
E-mail: Triss.Ashton@unt.edu

Nick Evangelopoulos
University of North Texas
College of Business
1155 Union Circle #311160, Denton, TX, 76203-5017
Phone: 940-565-3056
E-mail: Nick.Evangelopoulos@unt.edu

ABSTRACT

Big data increasingly includes large collections of textual data that describe various facets of the customer relationship. These datasets consist of unstructured data and represent the customer voice on key business topics including customer satisfaction and perceptions of product quality. We propose a method of converting this qualitative data to a quantitative format and monitor the resulting construct parameters by control charts. A research agenda is also proposed.

INTRODUCTION

Although there is extensive literature on Customer Relationship Management and the measurement of service quality, most service quality studies either look at snapshots of service quality or, when they adopt a longitudinal approach, they rely on structured data and use standard methodological approaches from the Quality Control and Improvement literature. In this approach, we examine customer feedback in its native form of unstructured text data and illustrate an approach for assessing process control and treating process change.

As a result of growth in the use of the internet and personal communication devices such as PCs and smart phones, organizations regularly collect customer relationship data in a variety of formats, including free text format. In a typical customer segmentation scenario, the object is to cluster customers that share similar characteristics. These characteristics are converted by scales to a quantitative form. Customer statements, however, as open-ended text responses, cannot be readily represented quantitatively. In this paper, we present a case study on a large retailer, involving a large collection of customer comments, and propose a methodological approach to convert textual data into numerical format, extract latent concepts, and monitor effects of business processes as perceived by customers. Traditional approaches to processing such unstructured data, which include manual reading of selected comments and keyword-based ad-hoc queries of the customer comment database, can be expensive, tedious, or highly biased

(Leech and Onwuegbuzie 2008). However, with the help of Latent Semantic Analysis (LSA), a text-mining method, we examine how we can efficiently reduce the data to a quantitative form, while also minimizing human bias. The research questions, therefore, are as follows:

RQ1. *Can unstructured customer comment data be automatically coded into numerical data without resorting to manual content analysis?*

RQ2. *What tools can be employed in order to monitor, statistically, the conceptual content in customer comments?*

In addressing these research questions, we follow the proof of concept approach and explore the feasibility of the proposed methods in a case study setting.

CASE STUDY

Data for this case study comes from a Fortune-500 specialty retailer offering an online subscription service to its customers in the U.S. Approximately 820,000 customers canceled their subscription over a 115-week period in the mid-2000s. As part of the cancellation process, the customers were asked to complete a survey. At the end of the structured part of the survey, the customers were provided with an open-ended text box in which to respond to the question "is there anything we could have done to keep you from terminating your contract?" Answers to this question were stored in the form of natural or raw text. Because of the large size of the data set, a 1-percent random sample of 8,222 comments was extracted for this analysis.

The company agreed to provide us with this data set on condition of anonymity. Because of the highly competitive environment the company operates in, alterations are used to protect the firm's identity. Services, products, and numerical values will be replaced by a descriptive term, for example, "[product]" will be used in lieu of the actual names of the products provided. Numerical values are replaced by ratio values to preserve a proportional representation allowing numbers to be compared.

LATENT SEMANTIC ANALYSIS FOR THE ANALYSIS OF TEXTUAL DATA

Analysis of the data was performed using LSA. LSA was first introduced by (Deerwester et al., 1990) in an information retrieval context but later evolved into a family of text mining methods, as well as a psychology and cognitive science theory of meaning (Landauer 2007). A detailed illustration of LSA as applied to information retrieval can be found in Deerwester et al. (1990). An illustration of LSA as a method of topic extraction can be found in the appendix of Sidorova et al. (2008). Methodological recommendations are provided in Evangelopoulos et al. (2011).

During the first steps of the analysis, LSA reduces a set of documents to a terms-by-documents weighted matrix **X**. Next, Singular Value Decomposition (SVD) decomposes the **X** matrix to its eigenvectors and singular values, given in matrix form by

$$\mathbf{X} = \mathbf{TSD}^{\mathrm{T}}, \tag{1}$$

where **T** is a term-by-factor eigenvectors matrix, **S** is the factor-by-factor singular values (i.e., square roots of eigenvalues) diagonal matrix, and $\mathbf{D^T}$ is a factors-by-documents eigenvector matrix (Deerwester et al., 1990). By keeping only the first k singular values ($s_1,\ldots,s_k$), the original term frequency matrix $\mathbf{X}$ is projected in a lower k-dimensional space as

$$\widehat{\mathbf{X}} = \mathbf{T}_\mathrm{k}\mathbf{S}_\mathrm{k}\mathbf{D}_\mathrm{k}^\mathrm{T}. \tag{2}$$

The objective is for the resulting $\widehat{\mathbf{X}}$ matrix to be the best k-dimensional approximation of the $\mathbf{X}$ matrix in the least-squares sense. $\widehat{\mathbf{X}}$ can, therefore, serve as a representation of $\mathbf{X}$ at a higher level of conceptual abstraction. Selection of an optimal value of k is still an open research problem (Evangelopoulos et al. 2011). The SVD of $\mathbf{X}$ produces two factor loading matrices, term loadings $\mathbf{L_T}$, and document loadings $\mathbf{L_D}$. The term loadings are produced by considering the term variance-covariance matrix $\widehat{\mathbf{X}}\widehat{\mathbf{X}}^\mathrm{T}$ as

$$\widehat{\mathbf{X}}\widehat{\mathbf{X}}^\mathrm{T} = \mathbf{T}_\mathrm{k}\mathbf{S}_\mathrm{k}(\mathbf{T}_\mathrm{k}\mathbf{S}_\mathrm{k})^\mathrm{T} = \mathbf{L_T}(\mathbf{L_T})^\mathbf{T}. \tag{3}$$

Similarly, the document loadings are produced by considering the document variance-covariance matrix $\widehat{\mathbf{X}}^\mathrm{T}\widehat{\mathbf{X}}$ as

$$\widehat{\mathbf{X}}^\mathrm{T}\widehat{\mathbf{X}} = \mathbf{D}_\mathrm{k}\mathbf{S}_\mathrm{k}(\mathbf{D}_\mathrm{k}\mathbf{S}_\mathrm{k})^\mathrm{T} = \mathbf{L_D}(\mathbf{L_D})^\mathbf{T}. \tag{4}$$

Once the $\mathbf{L_T}$ and $\mathbf{L_D}$ matrices are extracted, they can be rotated in order to yield equivalent factors that are interpretable by humans. Factor rotations are performed through a multiplication by an orthonormal rotation matrix $\mathbf{M}$. The rotated term and document loadings then become $\mathbf{L_T^*} = \mathbf{T}_\mathrm{k}\mathbf{S}_\mathrm{k}\mathbf{M}$ and $\mathbf{L_D^*} = \mathbf{D}_\mathrm{k}\mathbf{S}_\mathrm{k}\mathbf{M}$. A common approach for identifying $\mathbf{M}$ is through a rotation procedure such as varimax or orthomax.

Using matrices $\mathbf{L_T^*}$ and $\mathbf{L_D^*}$, through co-examination of high-loading terms and high-loading documents, we are able to label the extracted factors. Since the terms are used exclusively to articulate factors, we follow the generally accepted standards established in multivariate analysis and focus on terms with loading values greater than 0.6 while associating terms to factors (Hair, 2006). The documents, however, are expected to include a far larger content than the high-level conceptual factors extracted here, such as customer stories that go on a tangent, language usage patterns, etc. Therefore, we expect factors to have a looser relationship with documents as compared to the strong association between factors and terms. As a result, we may need to apply smaller loading thresholds. This observation is in line with Evangelopoulos et al. (2011).

The columns in the $\mathbf{L_D}$ matrix (factor loadings) associate documents (customer comments) to factors (specific issues raised by the customers). As the documents are time-ordered, moving through the elements of each column vector in the $\mathbf{L_D}$ matrix corresponds to following the evolution of a certain factor (an issue raised by the customers) through time. A measure of magnitude or length of a factor vector is given by the vector's length, equal to the square root of the sum of the squared components of the factor (Bronson and Costa 2009, p.33):

$$||\boldsymbol{y}|| = \sqrt{\sum_{\boldsymbol{i}=\mathbf{1}}^{\boldsymbol{n}} \boldsymbol{y}_{\boldsymbol{i}}^{\mathbf{2}}}. \tag{5}$$

Since the component values are measures of length, not only do they correlate, they also quantify the relationship between the document and the factor.

EXTRACTION OF TOPIC FACTORS

LSA revealed 20 factors in the customer comment data set. The factors were independently labeled by two researchers without controversy. A closer review of the document loadings concludes that the loadings are not normally distributed and that they are autocorrelated. Autocorrelation occurs because of cross loadings that result when a customer has more than one reason for closing an account.

An illustration of factor loading distribution is presented in Figure 1. These distributions consist of two distributions, the data of interest somewhere out in the far tail, and then a mixture of unrelated data plus noise data clustered about zero. As one moves up (or down for negative sloped data) through the distribution from zero, many documents unrelated to the factor of interest share one or more terms associated with a given factor resulting in that document having an elevated loading value. As documents increase the number of terms they share with the factor, their quantity decreases, but their loading values increase. Eventually, as one moves out into the distribution's tail, the documents collect enough of the terms associated with the factor that the semantic nature of the comment shifts and it becomes aligned with the factor definition.

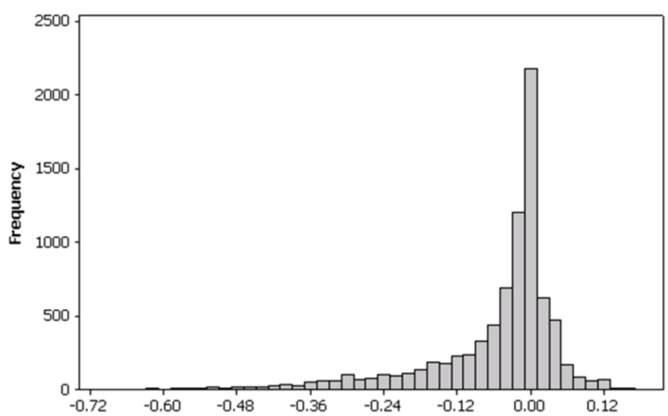

Figure 1: Histograms of loading data from $\mathbf{L_D}$ matrix for factor F7. Frequencies are on the y-axis and loading (eigenvector component value) on the x-axis.

To split the two distributions, we determined the cut-off point where documents ceased to relate to the factors. Each factor in the $\mathbf{L_D}$ matrix was sorted in descending order and a manual reading of each document occurred. In factor 1, below a component value of 0.1815, documents ceased to relate to the factor identify and were simply noise observations. In factor 2, a manual review found that the cutoff point was 0.2284, below which the documents are simply noise. In each of these factors, the customer comment writer coincidentally using terms associated with the factor as they discussed other topics. A breakdown of the loadings on factor F7 separates the mixed loadings shown in Figure 1 into "insignificant loadings" (noise), shown in Figure 2, and “significant loadings”, shown in Figure 3, left graph. The dividing point (threshold) for factor F7 was found to be equal to -0.1800.

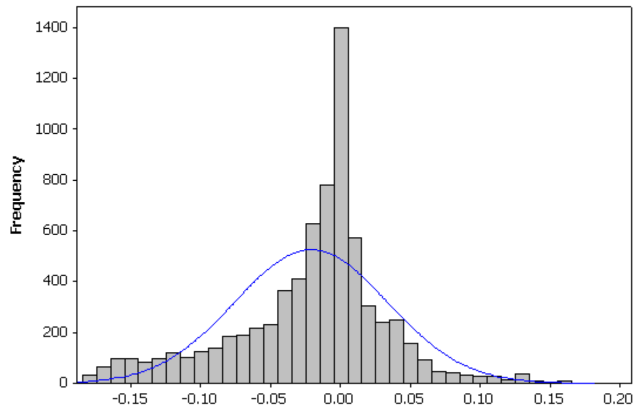

Figure 2: Histogram of factor F7 data about zero. Data includes those component scores representing zero and noise data with a normal curve overlay. Frequency on the y-axis and eigenvector component value on the x-axis.

It should be noted that setting the lower threshold at the minimally observed levels is problematic in that there is no way to automate the data cutoff. All subsequent projects would require manual review of each factor in order to determine the lower threshold. Manual reviews can quickly and easily generate biased results. Further, while setting the threshold at the lowest observed level does include all documents that reflect the semantic nature of the factor, it also increases the number of noise items. In order to automate this process, and to minimize the introduction of noise, we accepted a 5-percent probability tail given that 20 factor are retained as the minimal threshold for this analysis. This is a heuristic based on the assumption that, *on average, each customer comment raises exactly one issue* (main reason for cancellation of service). The same heuristic was used in Sidorova et al. (2008). For the data set under consideration, 5-percent probability results in documents with eigenvector component values greater than 0.2701 retained as items of interest for each factor and documents with eigenvector values less than or equal to 0.2701 discarded as noise and zero data. The resulting distribution of final loading data is presented in Figure 3, right graph.

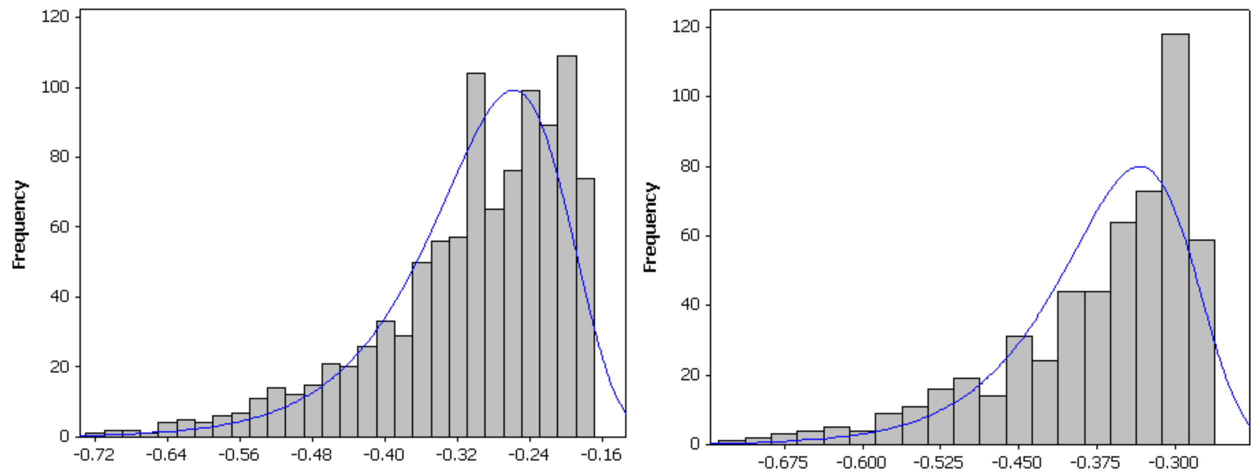

Figure 3: Histograms of component values of interest for factor F7. Data are presented with three-parameter Weibull distribution overlay. Frequency on the y-axis and eigenvector component value on the x-axis. Left graph are observations with values greater than |0.1800| and right pane are observations with values greater than the heuristically derived |0.2701|.

The 5-percent probability criterion performed reasonably well. Some data was lost because the lower threshold observed in each factor was below 0.2701, but noise data was minimized in the analysis set. Still, above the 0.2701 level, there are observations that are simply noise. During the manual data review, the tipping point, that point where noise began to occur was also observed. There are a few noise documents between the tipping points and the 5-percent probability point (0.2701), but there are a lot more between the 5-percent probability point and the minimum observed thresholds. For example, factor F5 began experiencing noise as early as 0.6770. Yet factors F2, F12, and F20 were noise free up until the semantic nature of the factor ceased at the lower thresholds. For the 20 factors extracted, the mean tipping point, the point where cross loading began to appear, was 0.3709. Based on these observations, the assumption of documents loading, on average, to one factor appears to be supported. However, further research on this topic is warranted.

CONTROL CHARTS

Control charts are limited by the distribution they are suited for. Some charts are best suited only to normally distributed data while others are suited to other non-normal distributions. The *c*-chart and *u*-chart, for instance, are well suited to the Poisson distribution (Duncan, 1986, p. 462) while the distribution of Shewhart type charts is approximately normal (Shore, 2000). Others have unique characteristics, for example, the *p*-chart need n to be high enough to avoid alarming on a single non-conformity if $\hat{p}$, the actual process value, is low (Duncan 1986, p. 451).

Exponential weighted moving average (EWMA) control charts were selected for this demonstration because the data set under consideration is non-normal and autocorrelated. Since we will be using sampling, the full effect of the underlying distribution will be highly moderated by Central Limit Theorem. However, we use samples based on document submissions per period. The resulting samples are variable in size potentially resulting is a skewed distribution.

Additionally, we are interested in small shifts in the mean that result because of management changes. EWMA charts are known to work well with autocorrelated data (Chao-Wen Lu & Marion, 1999). Since EWMA charts are distribution-free, the chart should still be appropriate if subsequent data set exhibited different distribution characteristics (Borror & Montgomery, 1999 and Maravelakis, Panaretos, & Psarakis, 2005). In addition, EWMA charts perform well on autocorrelated Weibull data (Black, Smith, & Wells, 2011) and the underlying data appears to follow a Weibull type distribution pattern.

Figure 4 provides EWMA graph for factor F7 for the entire 115-week data collection period. The weight factor λ has been set to the upper end of the recommended working interval of 0.25 (Montgomery 1996, p. 338). Also, note that several points are out of control resulting from using a fixed centerline while the underlying system was changed by management.

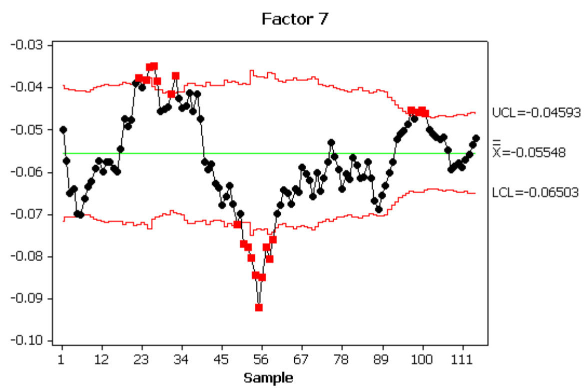

Figure 4: EWMA chart of factor F7. Weight factor λ is set to 0.25.

Note that factor F7 has loadings with negative values. Interpretation of charts with negative data can be confusing because the preferred state is zero which in negatively sloped cases will be closest the upper control limit. Negative-valued loading data will be decreasing as the observations move toward the upper control limit while positive-valued loading data will be decreasing as it moves toward the lower control limit. This potential confusion can be alleviated by a data transformation either by taking the absolute value or through multiplication by (-1). Selection of a transformation method is dependent on when the transformation is performed. If a transformation is performed prior to splitting out the noise component, use the (-1) multiplication method because data exists on both sides of zero. If the transformation occurs after splitting out the noise component, either method is satisfactory.

A FUTURE RESEARCH AGENDA

In this paper, we have attempted to address the problem of monitoring service quality issues in control charts using textual data. Our treatment of this relatively unexplored domain is far from complete. Research questions that should be addressed in future studies include the following:

RQ1. *What is the best text mining method to be used in extracting topics from a collection of documents, such as customer comments, that represent an unstructured source of data for capturing service quality characteristics?*

Discussion: In this paper, we have used Latent Semantic Analysis, but other methods, such as Non-negative Matrix Factorization or Latent Dirichlet Allocation might provide viable alternatives.

RQ2. *What is the best metric to be used in quantifying customer feedback related to a specific service quality characteristic as captured by a certain text mining method?*

Discussion: In this paper, we have used document loadings on LSA factors, but other metrics, such as filtered loadings, transformed loadings, or high-loading document proportions should also be considered and compared.

RQ3. *What is the best control chart to use for purposes of monitoring a service quality characteristic as captured by a certain text mining method and measured by a certain metric?*

Discussion: In this paper, we have used EWMA charts as a tool for detecting whether a process related to a certain service quality characteristic is in control or of out-of-control. Other possibilities would include using p-charts where the proportions of customer comments raising a certain quality issue are recorded.

CONCLUSION

In this paper, we explored methodological approaches to monitoring service quality components using unstructured customer comment data. We identified Latent Semantic Analysis as a suitable method in quantifying raw customer comments and extracting concepts related to causes for service quality breakdown. We then explored the use of EWMA charts in plotting time-ordered factor loadings and identifying out-of-control situations. We illustrated these approaches using real customer comment data provided by a large retailer with an online store. This illustration suggests the applicability of our methodological choices.

The analysis of customer comment data is still in its early stages and it involves a number of open problems that provide fruitful directions for future research. Industry holds an abundance of raw text data and a shortage of effective analysis methods. With effective methods of analysis, the voice of the customer can be presented in a concise quantitative format such as control charts that are easy and quickly read. In contrast, survey-based research work depends on the instrument developer framing a question to which a customer responds with a numerical score or a ranking. When asked the same question in an open-ended question format, the customer can reveal unexpected information that might not be accounted for when the survey question is framed. Analysis of customer comments in the manner presented here facilitates unexpected information. Therefore, we expect strong interest in this area from researchers as well as practitioners. The research questions listed in the last section provide some possible directions for further research.

REFERENCES

Black, G., Smith, J., & Wells, S. (2011). The impact of weibull data and autocorrelation on the performance of the shewhart and exponentially weighted moving average control charts. *International Journal of Industrial Engineering Computations*, *2*(3), 575-582.

Borror, C. M., & Montgomery, D. C. (1999). Robustness of the EWMA control chart to non-normality. *Journal of Quality Technology, 31*(3), 309

Chao-Wen Lu, & Marion, R. R.,Jr. (1999). EWMA control charts for monitoring the mean of autocorrelated processes. *Journal of Quality Technology, 31*(2), 166.

Deerwester, Scott, Dumais, Susan T, Furnas, George W, Landauer, Thomas K. (1990). Indexing by latent semantic analysis. *Journal of the American Society for Information Science*, *41*(6), 391.

Duncan, A. J. (1986). *Quality control and industrial statistics* .Homewood, Il.:Irwin.

Evangelopoulos, N., Zhang, X., & Prybutok, V. R. (2011). Latent semantic analysis: Five methodological recommendations *European Journal of Information Systems*, In Press.

Hair, J. F. (2006). *Multivariate data analysis* (6th ed.). Upper Saddle River, N.J.: Pearson Prentice Hall.

Landauer, T.K. (2007). LSA as a theory of meaning. In Landauer, T.K., McNamara, D.S., Dennis, S. & Kintsch, W. (Eds.), *Handbook of Latent Semantic Analysis*. Mahwah, NJ: Lawrence Erlbaum Associates.

Leech, N., and Onwuegbuzie, A. (2008). Qualitative Data analysis: A Compedium of Techniques and a Framework for Selection for School Psychology Research and Beyond. *School Psychology Quarterly*, 23(4), pp. 587-604.

Maravelakis, P. E., Panaretos, J., & Psarakis, S. (2005). An examination of the robustness to non normality of the EWMA control charts for the dispersion. *Communications in Statistics: Simulation & Computation, 34*(4), 1069-1079.

Montgomery, D. C. (1996). *Introduction to statistical quality control* .New York, NY:Wiley.

Shore, H. (2000). General control charts for attributes. *IIE Transactions, 32*(12), 1149.

Sidorova, A., Evangelopoulos, N., Valacich, J. S., & Ramakrishnan, T. (2008). Uncovering the intellectual core of the information systems discipline. *MIS Quarterly*, *32*(3), 467-A20.