Paper 147-2008

A SAS® Text Mining Approach to Predicting the Resolvability of Disputes between eBay's Sellers and Buyers

Kevin Henning & Zhangxi Lin, Texas Tech University, Lubbock, Texas

ABSTRACT

A well functioning reputation and feedback system is foundational in consumer-to-consumer electronic commerce, an example of which is the electronic auction. Our interest in this paper is the analysis of the predictive power of both buyer and seller comments in determining the resolvability of transaction disputes in online auctions. Using data gathered from the eBay, Inc. reputation system, we analyze buyer and seller comments using the versatile SAS® Enterprise Miner™ software Text Miner node. In our analysis we employ a binary target variable "Resolvable" to indicate whether an auction dispute has the possibility of being resolved to the satisfaction of buyer and seller. The results suggest that textual analysis of seller comments is consistently more predictive than either content analysis based on human coding of feedback text or textual analysis of buyer feedback alone. The implications and impact of this exploratory analysis are important for potential buyers and sellers in online auctions. First, if a buyer is interested in determining whether a dispute will be resolvable, he/she should spend the most time analyzing the comments from the seller, as the seller seems to have the most power over the manner in which a dispute is resolved, indeed, if it will be resolved at all. Second, when given the choice of examining the final auction price only, or the final auction price and seller comments, a potential buyer should opt for examining seller comments because they are much more predictive of the final outcome of the dispute.

Keywords: SAS; ELECTRONIC COMMERCE; TEXT MINING; EBAY; CONTENT ANALYSIS; LOGISTIC REGRESSION

INTRODUCTION

The role of a reputation and feedback system is foundational in online auction-type e-commerce. Such systems, which allow both buyers and sellers to give positive, negative, or neutral feedback "points" and include short comments, serve to reduce information asymmetry and build up trust, thereby providing two critical characteristics of a trust-enabled online relationship: insurance (in a partial sense) and accountability (Friedman, Kahn, & Howe, 2000). In short, they provide an incentive for dealing fairly (e.g., Resnick, et al., 2000). Perhaps the most well-known online reputation system employed in e-commerce today is the one used for eBay.com, the world's foremost consumer-to-consumer auction website. Due to the sheer size and prominence of eBay (Tedeschi, 2007), the functionality of its reputation and feedback system can be considered something of a de-facto standard for how such systems should operate. As such, this paper will discuss the eBay reputation system throughout, although we contend that the principles presented herein should generalize readily to similar reputation and feedback systems.

Briefly, the eBay reputation system functions like a scoreboard. The parties involved in a given transaction are allowed to give a positive feedback point (+1), a negative feedback point (-1) or neutral feedback (a 0 value). Any single buyer or seller is only allowed to affect another seller or buyer's reputation score one time, regardless of the number of times the parties transact business (eBay, Inc., 2007). The reputation score that appears as part of the eBay user's profile is the sum of all unique feedback points that the user has received in the history of using a particular user ID. A recent addition to eBay's reputation system allows buyers to provide more detailed ratings about sellers in several dimensions: item as described, communication, shipping time, and shipping and handling charges (eBay Inc., 2007). Buyers may rate the seller using a "1 to 5 star" system, with 5 stars interpreted as an excellent rating. The ratings in each dimension are aggregated and averaged over a 12-month rolling basis. The average ratings are displayed in the user's feedback profile. An exemplar of the information provided to any person who clicks on a particular user name in the eBay database is shown in Figure 1.

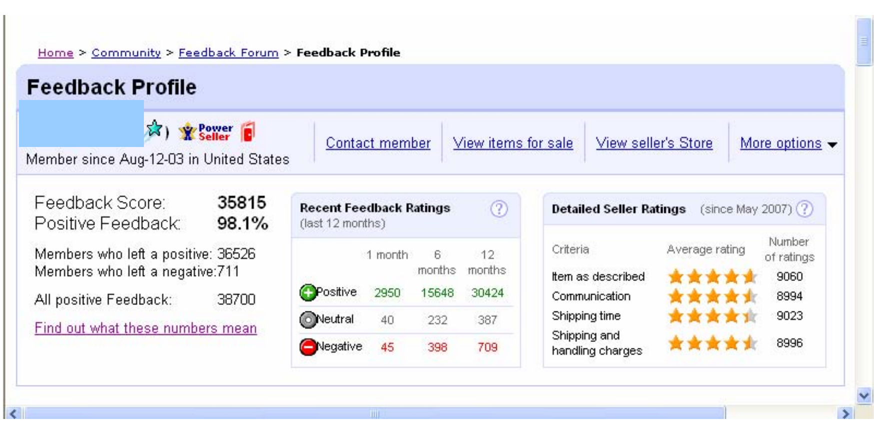

Figure 1--A typical eBay user feedback profile. Username is redacted to protect privacy.

Prior research has shown that the reputation system employed by eBay is theoretically attractive (e.g., Resnick, et al,2000; Laureti, Slanina, Yu, & Zhang, 2002; Cameron & Galloway, 2005) in that such systems tend to “filter out” incompetent and/or deceptive buyers and sellers by providing information about whether they are to be trusted. Trust is a vital component of all human interaction, and is particularly challenging to develop in an online context (Cofta, 2006; Lee, Im, & Lee, 2000). The reputation system has emerged as a proxy for the transactional and interpersonal history that would be developed between buyers and sellers over the course of repeated interaction (Resnick, et al., 2000), as well as a significant predictor of the price of the item and the likelihood that the transaction will be successful (MacInnes, Li, & Yurcik, 2005; MacInnes, 2005; Laureti, Slanina, Yu, & Zhang, 2002; Cameron & Galloway, 2005).

Reputation scores alone, however, cannot capture the “whole story” behind a transaction; the impressions of buyers and sellers are typically nuanced. To clarify, assume that there are two buyers, one only moderately satisfied with his purchase and another ecstatic about hers. Both buyers will no doubt leave a positive feedback point because, overall, the transaction was probably a positive one. But does this feedback point truly capture the essence of the transaction? It should be clear that it does not. However, one may get a better understanding of the experiences of these buyers by examining their *feedback comments*, short freeform textual comments that users are allowed to leave in addition to a feedback point. It is here that we may understand how each buyer truly feels. Buyer A, who is only moderately happy, may write: “product ok no major probs. delivery sort of slow.” The ecstatic buyer, B, meanwhile might leave a comment such as “sooooo happy with this product. This seller is the best. A+!.” Using a traditional numerical reputation system, Buyer A and Buyer B are deemed identical in terms of their experience, which is clearly not the case.

We take the view that the type of nuanced information available in textual comments from auction participants is of great value to potential buyers and sellers. With the explosion of user-generated content on the Internet, there is growing interest in tapping into the information provided by the interactions of users with one another. Further, prior research has found that individuals are highly influenced by the experiences of others and tend to trust the opinions of other consumers similar to them over the opinions of experts or other sources (Huang & Chen, 2006). Exploratory research by Weinberg and Davis (2003) indicates that potential buyers in an online auction context are highly influenced by the comments left by others, particularly negative comments.

This tendency of individuals to “home in” on negative comments or information and give much greater weight to negative information in decision making has been well documented (e.g., Weinberger, Allen, & Dillion, 1981; Kanouse, D., 1984; Weinberger, & Lepkowska-White, (2000); Ofir & Simonson, 2001; Weinberg & Davis, 2003; Wouter & Pidgeon, 2004). Given the intense interest of individuals in so-called negative information, it should be eminently useful for potential buyers and researchers alike to understand exactly how negative comments in the context of online auctions are used and how the information embedded in these comments—which, according to prior research, is what most consumers will focus on—contributes to the resolvability of a transaction dispute. As far as we are able to tell, no significant research has addressed the analysis of negative eBay comments as a means to gain a greater understanding of the outcomes of electronic auctions.

In this paper we employ the text mining capabilities of SAS Enterprise Miner ™ software to perform some exploratory analysis on actual eBay transaction data to determine the predictive power of buyer and seller comments on the resolvability of auction disputes.

DATA

We collected data on 890 auctions on eBay involving 206 unique user IDs. The auctions involved a wide array of product types—including consumer electronics, apparel, cosmetics, and computer accessories—and a wide range of prices, with a minimum price of $0.01 and a maximum of $1,748. A description of the variables used in the dataset may be found in Table 1.

Variable Name	**Description**
SellerID	The unique eBay username of the seller in an auction
BuyerID	The unique eBay username of the buyer in an auction
Item_Name	The description of the item being auctioned as written by the seller
Price	Final auction price in $US
Buyer_Complaint	Negative comments left by the buyer in the eBay reputation system
Seller_Reply	The seller’s reply to the buyer’s negative comments in the eBay reputation system

Table 1--Data dictionary of the original, raw dataset

Our primary interest in this paper is the analysis of both Buyer and Seller comments left in the eBay reputation system. But neither Buyer nor Seller is required to leave comments after an auction, and occasionally one of the parties chooses to leave comments while the other does not. We found that this was the case for several of our original 890 observations. Because we wanted to have a complete dataset with no missing values for this exploratory research, we preprocessed the original dataset to remove 326 observations in which either the seller or the buyer chose not to leave feedback. This left us with 563 complete observations, which is quite reasonable for an exploratory study. Of these 563 observations, 114 are related to unique Seller_IDs. The maximum number of the complaints collected on the same Seller_ID is 48, and the minimum is 1.

CONTENT ANALYSIS OF BUYER AND SELLER COMMENTS

In order to ascertain whether a particular dispute between a buyer and a seller, as described in their respective feedback comments, is able to be resolved or not, we chose to employ a systematic method of coding the Buyer_Complaint and Seller_Reply variables using a simple 1 to 3 Likert-style scale that measured the emotional state of the buyer as evidenced by his/her comments and the constructiveness of the seller's reply. The coding scheme we employed is provided in the following table:

Buyer's Emotional State		
Code	Meaning	Description
1	Dissatisfied	Complaint is largely factual and addresses specific issues without personal attacks against the seller
2	Angry	Complaint still contains facts, but also clearly indicates that buyer is expressing anger, disappointment, sadness, etc.
3	Irate	Complaint contains few, if any, facts and is geared towards attacking the seller personally with no mention of specific remedy requested
Seller's Reply		
Code	Meaning	Description
1	Direct	Seller addresses the problem directly and professionally
2	Defensive	Seller takes offense at negative feedback and may accuse buyer of causing the problem. May include mild personal attacks.
3	Unreasonable	Seller shows little or no interest in addressing the problem, instead opting to attack the buyer personally

Table 2--Textual analysis coding scheme

To lessen the impact of bias introduced by the inherently subjective nature of assessing meaning and implication from text, the two authors coded the data separately, avoiding discussion until both had finished. At the end of the individual coding, the code assigned to the Buyer_Complaint and the Seller_Reply by each of the two coders was compared. We considered the coding to match if both codes were identical; we did not consider fuzzy concepts of "near-matches" in this research. The match rate for the Buyer coding was 63.59%, while the Seller coding produced a match rate of 68.38%. The overall match rate was just under 66%. For exploratory research using only two coders, we considered this a sufficiently high to proceed. We also computed two additional variables, simple averages of the two codes each assigned to Buyer_Complaint and Seller_Reply called "ACode_B" and "ACode_S".

The final step in the preprocessing of the data involved defining a target variable to predict using SAS Enterprise Miner™ software. Reflecting extant literature on negative feedback in eBay's reputation system (Li & Lin, 2004), we defined a binary variable, "Resolvable," to indicate whether the dispute has the possibility of being resolved to the satisfaction of buyer and seller. In this research, we only consider two possible conditions: the dispute is able to be resolved (Resolvable = 1) or it is not able to be resolved (Resolvable = 0). Because Resolvable is defined as a binary variable, we could not use the simple average of two coders' opinions in this case, but rather, we discussed each dispute and agreed upon the resolvability of the dispute (i.e., whether to assign a 1 or 0 value to Resolvable for that particular dispute). The following tables list the criteria we used to classify the disputes:

Criteria for Resolvable = 1 (Dispute is able to be resolved)	
Criteria	Example Situation
Buyer and seller appear willing to work toward a resolution, and are not merely interested in trading snide remarks.	A buyer does not receive an item he/she bought after a reasonable (in the buyer's mind) time. Seller agrees to look investigate the problem and/or refund the buyer's payment
After receiving a negative comment the seller describes how he/she is working to resolve the issue	A buyer claims that the product he/she received is not the one that he/she was expecting to receive based on the auction listing description. Seller responds by agreeing to ship the correct product.
The dispute is a current one in which buyer and seller still appear to be open to solving the problem, not merely describing how the other party failed to fulfill their part of the transaction.	A buyer claims he/she received the correct item, but that it was damaged or stopped working after a short period of time for no apparent reason. The seller agrees to refund the purchase price or send a replacement.
The general tone of the dispute is one of a simple misunderstanding/miscommunication and not one involving a fundamental difference of opinion	A buyer claims he/she is unable to contact the seller. Seller responds by informing the buyer that on occasion, the email from the seller gets diverted to spam folders.

Table 3--A list of criteria to assign a '1' value to Resolvable variable

Criteria for Resolvable = 0 (Dispute is not able to be resolved)	
Criteria	Example Situation
The buyer and/or seller do not seem concerned at all about resolving the problem and instead trade harsh comments without mention of a possible solution	The buyer offers nothing factual in his/her complaint and instead berates the seller in a number of ways, including using profanity and/or accusing the seller of outright fraud. Seller counters the buyer's claim by contending that the buyer is at fault, and also uses profanity and/or harsh words.
The buyer and/or seller discuss the dispute in the past tense, indicating that both have resigned the fact that the dispute is not resolvable	A buyer says, "Never got the product." The seller responds, "You didn't give us a chance to fix it. You should have contacted us."
The seller gives a generic response that does not indicate an active interest in solving the problem at hand	A buyer claims he/she never received a product. The seller responds, "Service is #1 priority! We guarantee everything!"
The seller and buyer fundamentally disagree about some aspect of the transaction	An international buyer complains that the seller does not seem interested in shipping to the buyer's home country. The seller responds by saying that he/she does not offer international shipping and such was stated in the original auction listing.

Table 4--A list of criteria to assign a '0' value to Resolvable

The distribution of non-resolvable and resolvable disputes is approximately 50%-50%, which is not inconstant with the findings of Li and Lin (2004). The distribution is shown in Figure 2 below.

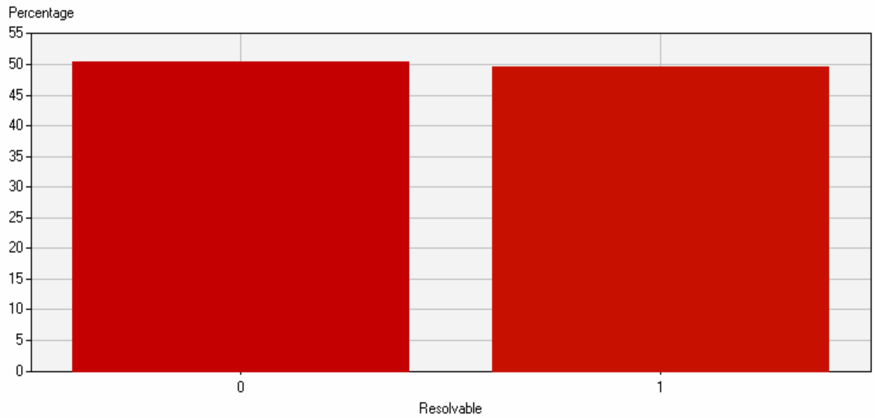

Figure 2--Distribution of the variable Resolvable.

MODELING

Using SAS Enterprise Miner ™ software, we created the following model for our analyses.

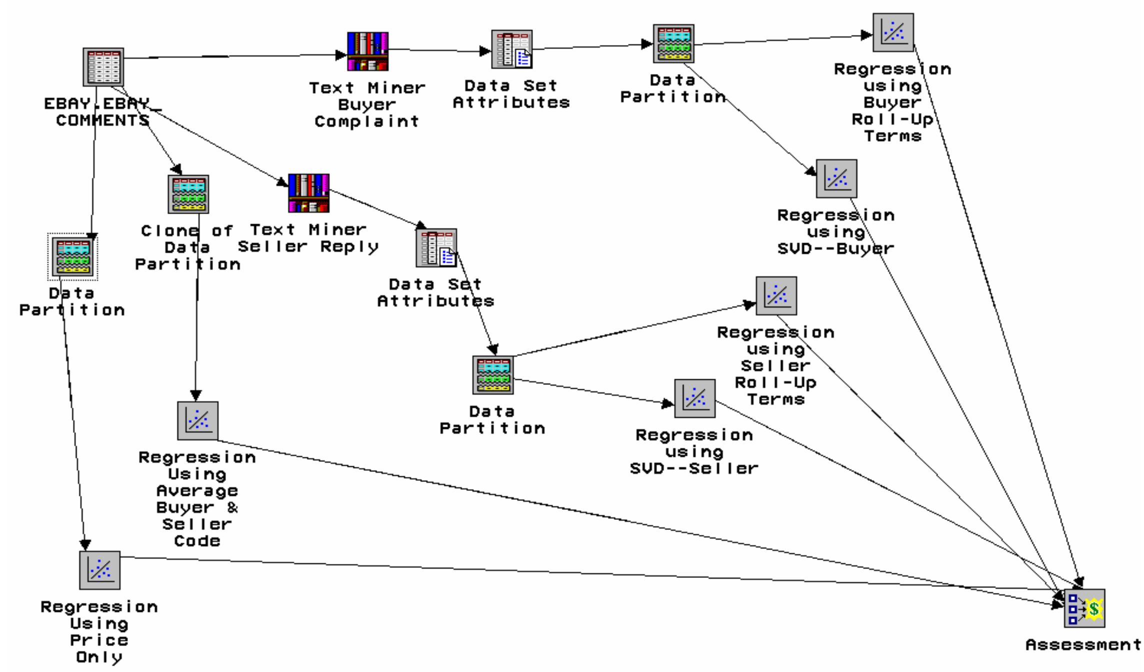

Figure 3--Model diagram

Several elements of the diagram are worth mentioning. First, since we are interested in the effect that buyer and seller feedback comments have on the probability of a dispute being resolvable, we choose to employ a logistic regression model with Resolvable as the response variable. This is appropriate because Resolvable is a binary-type variable indicating whether a dispute was, in the coders' opinion, able to be resolved. The predictor variables we use vary depending on the particular model.

We note that we have included two regression nodes that do not directly use buyer or seller comments. We do so in order to be able to compare models using direct textual information to those using indirect or no textual information. One such regression node uses Price as the sole predictor variable and the other uses the ACode_B and ACode_S.

The remaining four regression nodes use the textual information extracted by the Text Miner node. Each of the two text fields, Buyer_Complaint and Seller_Reply, is analyzed using Text Miner with the following options selected:

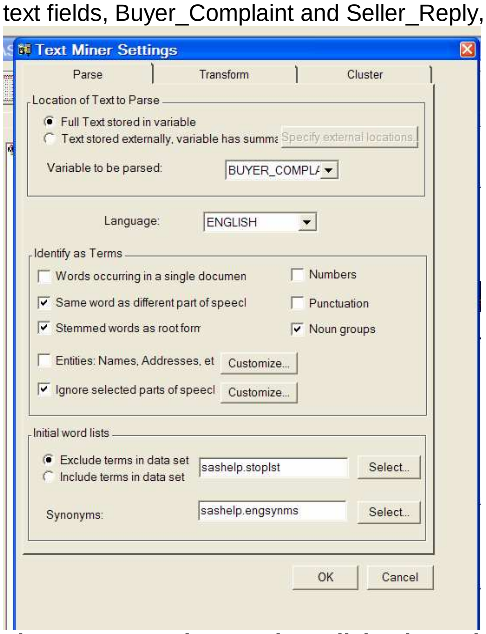

Figure 4--Text Miner settings dialog box. The setup is identical for the Seller_Reply text.

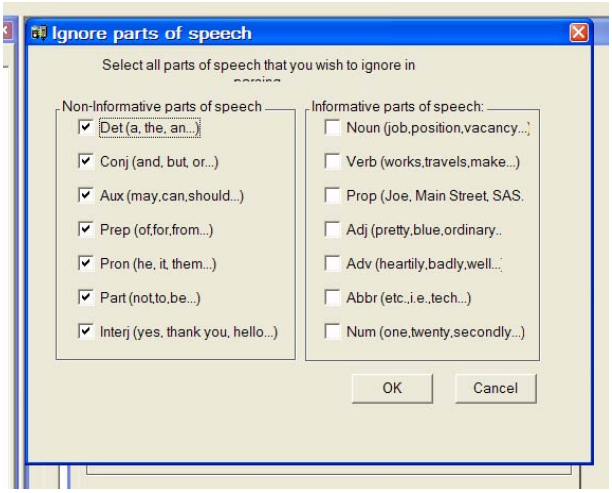

Figure 5--Dialog box with parts of speech that we wish to exclude from analysis

Figures 4, 5, and 6 depict the setup dialog box in SAS Enterprise Miner™ software. For the most part, we use the default settings, with two notable exceptions. First, as shown in Figure 5, we choose to ignore some of the more common and less informative parts of speech such as prepositions, conjunctions, and so forth. Second, we choose to calculate both singular value decomposition values and 100 roll-up terms (i.e. the terms given the most weight or significance by the text mining algorithm). This is easily accomplished by clicking on the "Transform" tab in the Settings dialog box, as shown here in Figure 6:

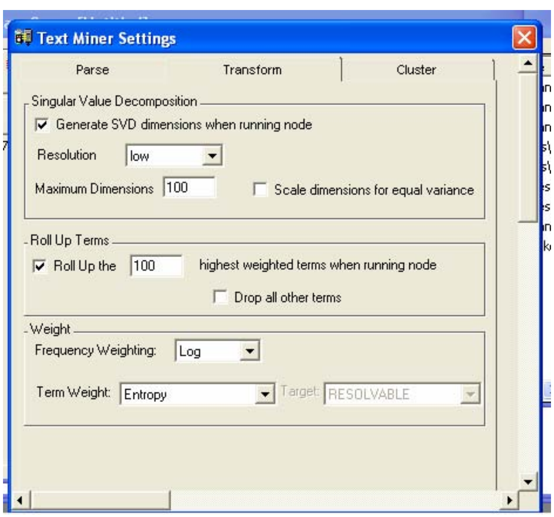

Figure 6--Transform tab of the Text Miner Settings dialog box. We choose to keep the 100 highest-weighted terms and generate a maximum of 100 SVD dimensions.

The setup of each regression node is straightforward. With reference to the model diagram in Figure 1 above, the following table describes the setup of each regression node in terms of the predictor variables used. The response variable is always Resolvable.

Regression Node Description	Predictors Used
Regression Using Price Only	Price
Regression Using Buyer and Seller Code	ACode_B; ACode_S (averages of human coding)
Regression Using Buyer Roll-Up Terms	The 100 roll-up terms calculated by Text Miner node
Regression Using SVD—Buyer	50 of the 100 SVD dimension values calculated by Text Miner node
Regression Using Seller Roll-Up Terms	The 100 roll-up terms calculated by Text Miner node
Regression Using SVD—Seller	50 of the 100 SVD dimension values calculated by Text Miner node

Table 5--Setup of regression nodes. Response is always the value of Resolvable

We partition the dataset into training and scoring datasets using a stratified method based on the variable Resolvable. This allows us to evaluate the performance of each regression model in terms of the predictive power of the predictors identified in Table 5. The setup screen of the Data Partition node is shown in Figure 7:

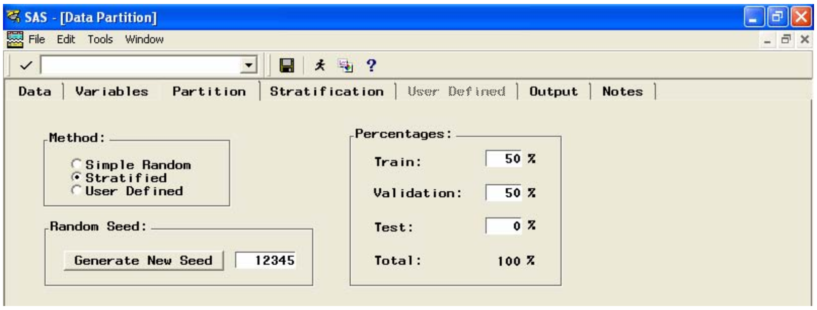

Figure 7--Data partition setup

As a final setup step, we logarithmically transformed the interval-level variable Price in order to correct for the extreme right skew of the raw distribution. The untransformed distribution is depicted in Figure 8 and the transformed distribution is shown in Figure 9. We use the transformed data in the subsequent modeling.

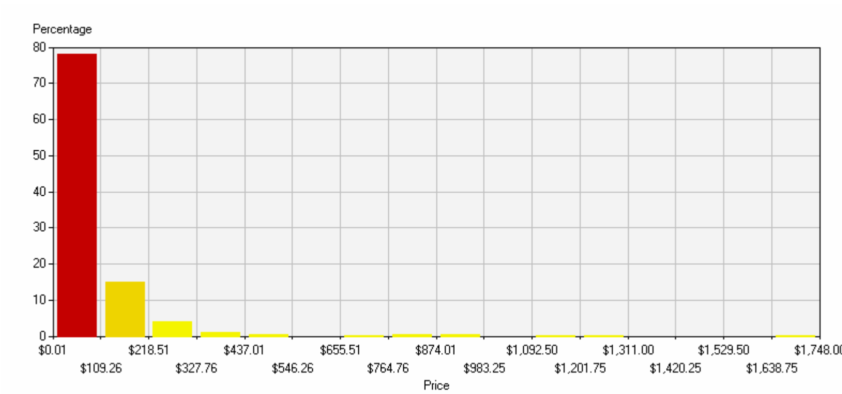

Figure 8--Distribution of untransformed Price variable

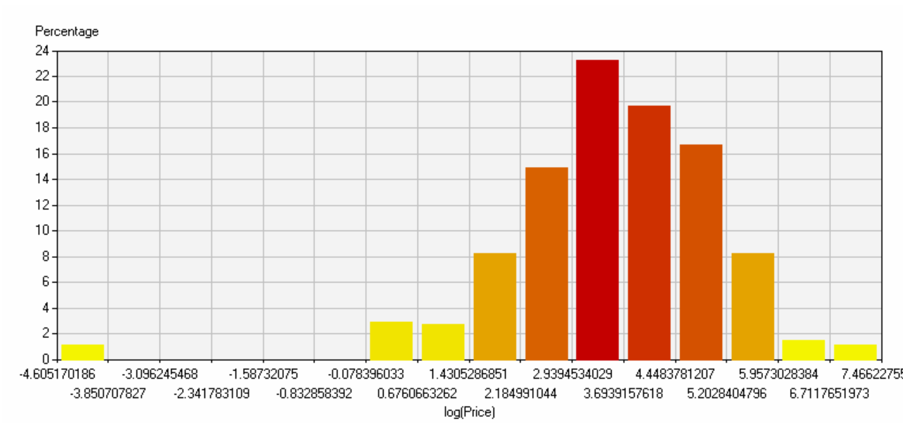

Figure 9--Distribution of transformed Price variable

RESULTS

We report our results using lift charts generated by the Assessment node (see Figure 3 above). We first examine how the models that do not include any direct textual information perform by themselves. Figure 10 illustrates that compared to a price-only regression model (R_PriceO), the regression model that incorporates the averages of the human coding for both buyer and seller (i.e., ACode_B and ACode_S) provides about 55% greater predictive power for the response variable Resolvable. This suggests that price *per se* is not a good predictor of whether a dispute can be resolved.

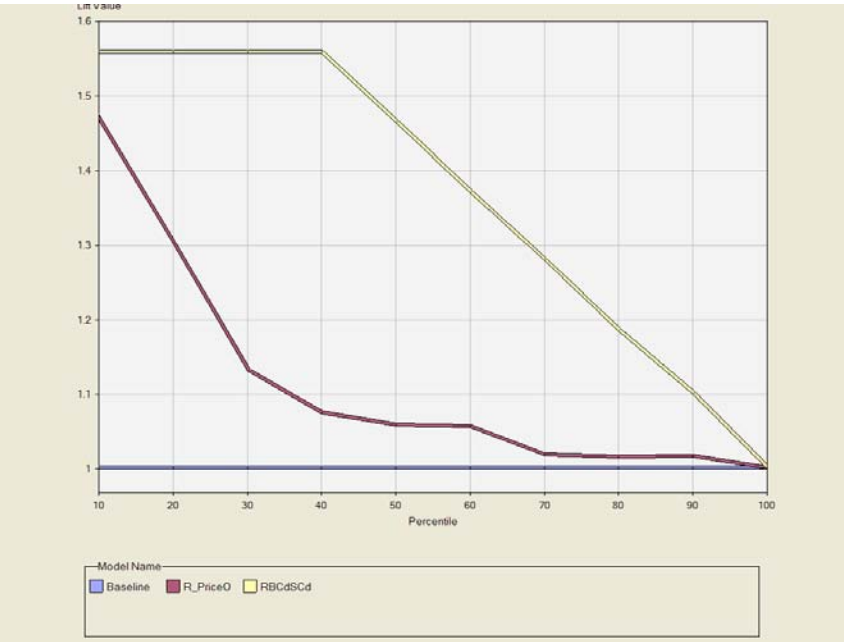

Figure 10—Lift chart comparing the Price Only regression model (R_PriceO) to the model employing the averages of the buyer and seller coding (RBCdSCd)

We next compared the performance of the model using the averages of the coding to the model using only the 100 roll-up terms from the Text Miner analysis of buyer's complaints. We see again in Figure 11 that the human coding provides a superior prediction of resolvability compared to using only buyer comments.

Figure 11--Comparison of regression using coding input (RBCdSCd) versus model using only 100 roll-up terms from buyer comment analysis (RRoll_Bu)

The situation is dramatically different, however, when we compare the performance of the human coding model to that of the regression model employing roll-up terms from the analysis of seller comments. Here, we see that the roll-up-term model outperforms the human coding model in the first decile before being overtaken by the human coding model in the second decile of the testing dataset (see Figure 12). Further, the superior predictive power of the seller comments over buyer comments (using roll-up terms) is clearly seen in Figure 13.

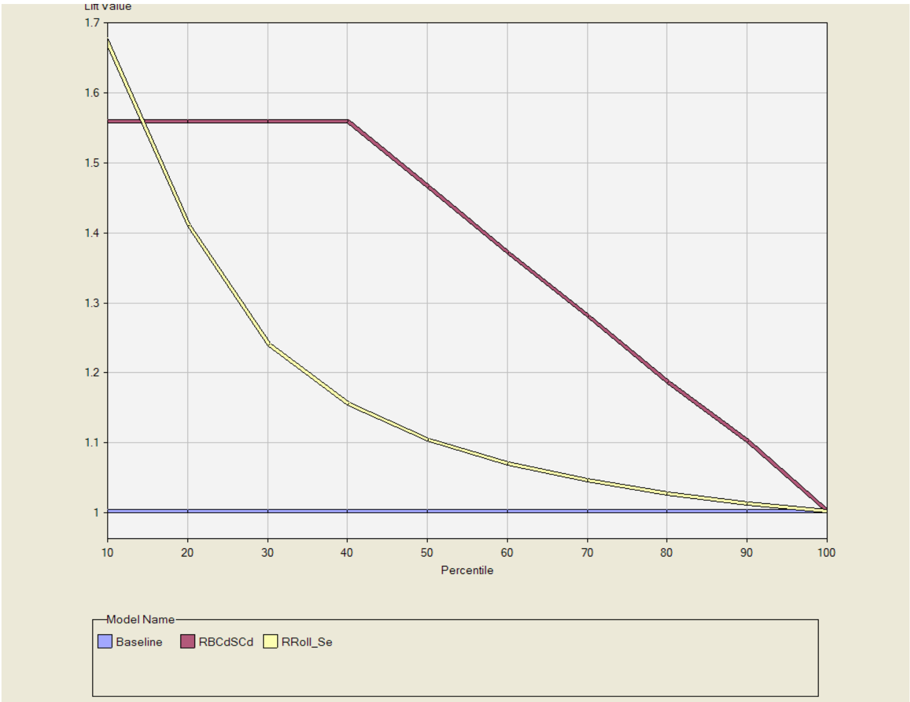

Figure 12--The human coding model (RBCdSCd) is slightly outperformed by the seller roll-up term model (RRoll_Se) for in the first decile of the testing dataset

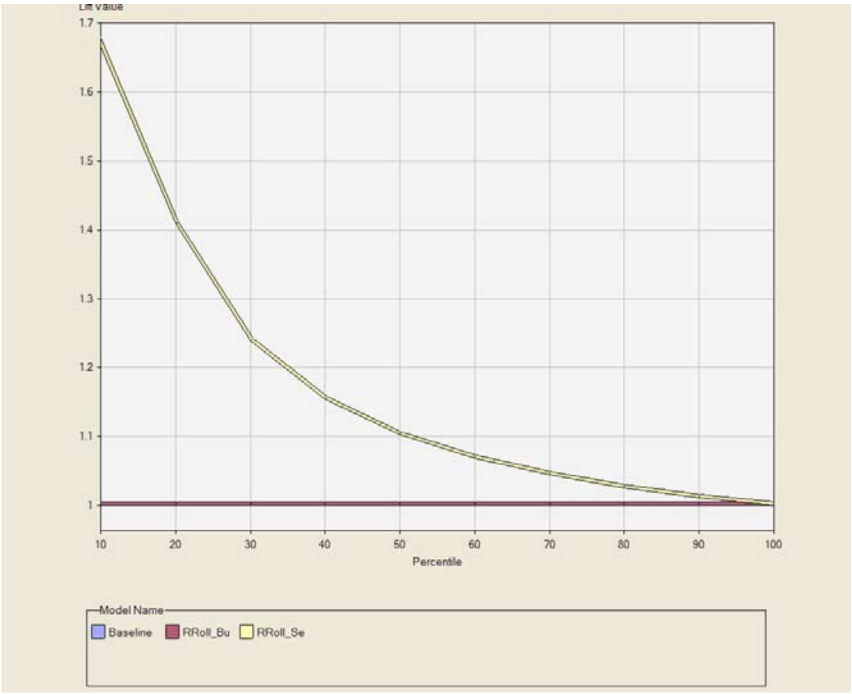

Figure 13--The seller roll-up regression model outperforms the buyer roll-up regression model

In order to determine which of the textual analysis-based models had the best predictive power, we compared the performance of the roll-up-term-based models for buyer and seller with the singular value decomposition (SVD) models for buyer and seller comments. The results are shown in Figure 14.

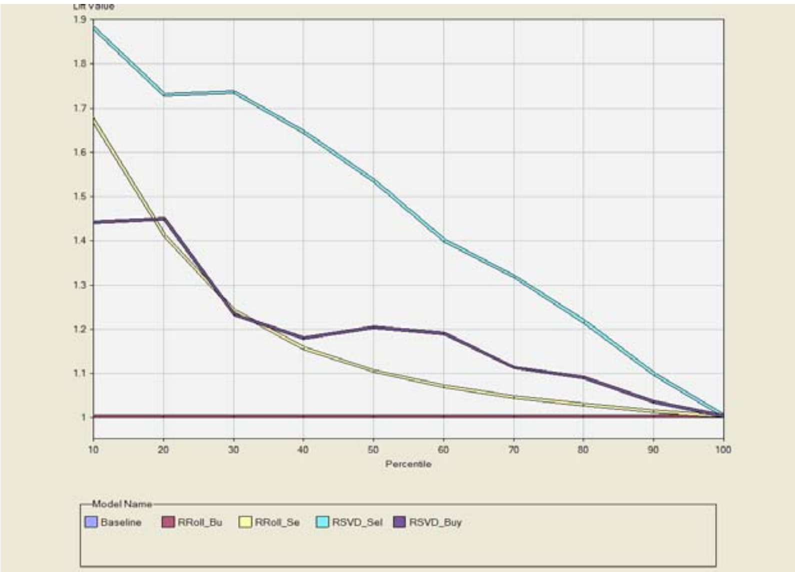

Figure 14--Analysis of all text-based models. We see that the SVD-Seller model outperforms all of the others consistently

Figure 14 shows one of the most significant findings of this research. When we compare the four possible text-based models using either buyer roll-up, seller roll-up, buyer SVD, or seller SVD as a predictor, the seller SVD model consistently outperforms the others in predictive power. The results suggest that seller-comment-based models in general are superior to buyer-comment-based models.

As a final comparison, we examine the effectiveness of the human coding model versus the SVD modes for buyer comments and seller comments. The results are interesting in that they suggest that human coding is more useful in predicting resolvability than textual analysis of buyer comments, but that textual analysis of seller comments is consistently more predictive than human coding. This situation may be seen in Figure 15.

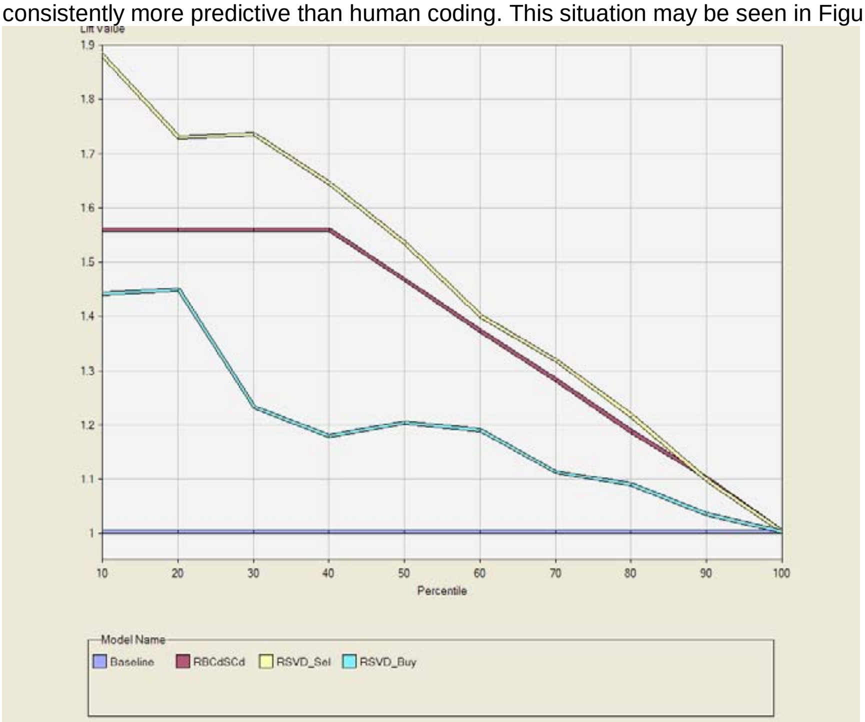

Figure 15--Performance of human coding model compared with that of SVD models

DISCUSSION AND CONCLUSION

The results from this exploratory analysis of buyer and seller comments produced some interesting and meaningful results. First, we found that models that include only final auction price are much less predictive of resolvability of disputes than models that incorporate, directly or through rater content analysis, the comments of buyers and sellers. Among those models that use only the simple average of human coding of buyer and seller comments, seller comments appear to be more predictive overall. This is not particularly surprising, given the disproportionate power of the seller versus the buyer to perform certain actions that would likely lead to a resolution to a dispute. However, our results do provide some quantitative support for this rather intuitive notion of the seller being in control of the situation.

Our analysis also examined the relative predictive power of two forms of analysis based on Text Miner output. The first type of output is roll-up term output. As seen in Figures 12 and 13, models using roll-up term output from analysis of seller comments perform quite well in comparison with models using human coding input, although they do not consistently outperform the human coding model. Buyer roll-up term models do not perform well at all in comparison to seller roll-up term models.

Figures 14 and 15, however, provide the most interesting results. These lift charts indicate that models using singular value decomposition (SVD) output from analysis of seller comments provide the best predictive power among all of the models investigated. The next-best predictive model is that employing human coding input, followed by the model employing Buyer SVD dimensions.

The implications of this exploratory analysis are important for potential buyers and sellers in electronic auctions. First, if a buyer is interested in determining whether a dispute will be resolvable, he/she should spend the most time analyzing the comments from the seller, as the seller seems to have the most power over the manner in which a dispute is resolved, indeed, if it will be resolved at all. Second, when given the choice of examining the final auction price only, or the final auction price and seller comments, a potential buyer should opt to examine seller comments because the latter is much more predictive of the final outcome. Finally, because the SVD-dimension based models performed the best in terms of predictive power, it should not be difficult for future researchers to develop a recommender system for sellers in auctions that would be able to recommend certain sellers to certain buyers based, in part, on textual analysis of negative feedback comments.

We must qualify all of the above by reiterating that this research was exploratory in nature, and thus the theoretical models for the relationships we have found have not been explored in detail. Further research into the factors affecting auction resolvability needs to be conducted. In addition, our conclusions were based on a sample size of 563. A larger sample of comments would, of course, allow more accurate models to be constructed. Finally, the research conclusions are limited by the use of only two coders (the two authors of this paper) and would be far more compelling if two or three more coders (i.e. four or five in all) could be used, so that more of the coder bias could be averaged out.

ACKNOWLEDGMENTS

Authors thank Jie (Jerry) Zhang and Rongyun (Laura) Chen, currently MIS graduate students at Southwestern University of Finance and Economics, China, for their effort in data collection with quality outcomes. We also thank Dr. Peter Westfall for his helpful comments.

REFERENCES AND FURTHER READING:

Cameron, D. D. & Galloway, A. (2005). Consumer motivations and concerns in online auctions: An exploratory study. International Journal of Consumer Studies, 29(3), pp. 181-192.

Cofta, P. (2006). Convergence and trust in ecommerce. BT Technology Journal, 24(2), pp. 214-218.

eBay Inc. (2007). Feeback scores and your reputation. Retrieved August 15, 2007, from http://pages.ebay.com/help/feedback/feedback-scores.html

Friedman, B., Kahn, P. H., & Howe, D. C. Trust online. *Communications of the ACM, 43*(12), pp. 34-40.

Hu, M. & Liu, B. (2004). Mining and summarizing customer reviews.

Huang, J. and Chen, Y. (2006). Herding in online product choice. Psychology & Marketing 23(5), p. 413–428.

Laueti, P., Slanina, F., Yu, Y., & Zhang, Y. (2002). Buyer feedback as a filtering mechanism for reputable sellers. Physica A, 316, pp. 413-429.

Lee, C. C. & Hu, C. (2004). Analyzing hotel customers' e-complaints from an internet complaint forum. Journal of Travel and Tourism Marketing, 17(23), pp. 167-181.

Lee, Z., Im, I., Lee, S. J. (2000). The effect of negative buyer feedback on prices in internet auction markets. In the Proceedings of the 21st International Conference on Information Systems, Brisbane, Australia.

Li, D. & Lin, Z. (2004). Negative Reputation Rate as the Signal of Risk in Online Customer-to-Customer Transactions. Proceedings of ICEB. December 5--8, 2004, Beijing.

MacInnes, I. (2005). Causes of disputes in online auctions. Electronic Markets, 15(2), pp. 146-157.

MacInnes, I., Li, Y., & Yurick, W. (2005). Reputation and dispute in eBay transactions. International Journal of Electronic Commerce, 10(1), pp. 27-54.

Ofir, C. & Simonson, I. (2001). In search of negative customer feedback: The effect of expecting to evaluate on satisfaction evaluations. Journal of Marketing Research, 38, pp.

Resnick, P. Zeckhauser, R., Friedman, E. & Kuwabara, K. (2000). Reputation systems. Communications of the ACM, 43(12), pp. 45-48.

Tedeschi, B. (2007). eBay moves to recharge its auctions. The New York Times, 156(53979), p. C5.

Weinberg, B. D. & Davis, L. (2003). Exploring the WOW in online-auction feedback. Journal of Business Research, 58, pp. 1609-1621.

Weinberger, M. G. & Lepkowska-White, E. (2000). The effects of negative information on purchase behavior. Journal of Marketing Management, 16(5). pp. 465-481.

Weinberger, M. G., Allen, C. T. & Dillion, W. R. (1981). Negative information: Perspectives and research directions. Advances in Consumer Research, 8(1), pp. 398-404.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the authors at:

Texas Tech University
Rawls College of Business
Lubbock, TX 79409-2101, USA
Work Phone: 806-742-3167
E-mail: kevin.henning@ttu.edu
Web: http://ta.ba.ttu.edu/khenning