A Seller's Perspective Characterization Methodology for Online Auctions

Arlei Silva, Pedro Calais,
Adriano Pereira, Fernando Mourão,
Jussara Almeida and Wagner Meira Jr.
Federal University of Minas Gerais
Computer Science Department
Av. Antônio Carlos 6627 - ICEx - 31270-010
Belo Horizonte, Brazil
{arlei, pcalais, adrianoc, fhmourao,
jussara, meira}@dcc.ufmg.br

Paulo Góes
University of Connecticut
School of Business
2100 Hillside Road, 06269-1041
Storrs, CT, USA
paulo.goes@business.uconn.edu

ABSTRACT

Online auction services have reached great popularity and revenue over the last years. A key component for this success is the seller. Few studies proposed analyzing how the seller and the auction configuration affect the negotiation results. In this work we propose a methodology to characterize online auctions by the seller's perspective. This methodology is based on: (1) recognizing the characteristics of the variables related to the auction results and (2) capturing the correlation among these variables to identify seller profiles and selling strategies. We applied our methodology to a real case study, using an eBay dataset, to validate two hypotheses about sellers and their practices. These results are useful to understand the complex mechanisms that guide ending prices, success (or failure), and the attraction of bids in online auctions, which can support decision strategies for buyers and sellers.

Categories and Subject Descriptors

J.4 [**Computer Applications**]: Social and Behavioral Sciences—*Economics*

General Terms

Business-to-Consummer - B2C, e-Business, auctions

Keywords

e-commerce, online auctions, data mining, eBay

1. INTRODUCTION

Over the last years, online auctions, such as eBay [10] and Yahoo! Auctions [23], have reached great popularity and revenue. These results turned electronic auctions on one of the most relevant scenarios of Business-to-Consumer (B2C) and Consumer-to-Consumer (C2C) e-commerce models. Online auctions combine features from conventional auctions with Web technologies, establishing a new dimension of the world economy that has not been well understood yet [5].

An auction can be divided into three different parts: before negotiation begins (when the input parameters have to be specified), during negotiation time (when the bidders place bids and compete with each other), and after ending date (when the auction results are generated, also called outputs). Auction inputs can vary according to the service provider and the negotiation model. Among these data, we can mention the description of the product being auctioned, the minimum starting price of a bid, the seller's registration date in the market, among others.

One of the biggest research challenges in online auctions is the understanding of the complex mechanism that guides the results of the auction negotiation. In order to improve our understanding about these mechanisms, it is essential to assess how the auction inputs are correlated with the auction results. Previous work has focused on analyzing how different input factors are related to the success of the auction, the ending price and the attraction of bidders [2, 6, 3]. However, identifying consistent and robust correlation patterns is a challenge. In this context, we call a correlation pattern a frequent behavior observed between some inputs and outputs. An example would be the relation between the seller's percentage of positive feedback and the ending price reached. Some previous work has studied how the positive feedback raises the ending prices obtained by sellers. It is an important information to understand the auction results.

Some studies conclude that buyers do not act completely rational [20] and this phenomenon limits the identification of any valid correlation between inputs and outputs. Moreover, many different factors may affect the auction results increasing the complexity of find these patterns.

Important factors related to the auction results are, for example, the starting price, the auction duration, among others. Understanding how these factors affect the auction results is useful for bidders, sellers and online auction's providers. Sellers can make decisions that increase the chances of achieving higher selling prices. On the other

10th Int. Conf. on Electronic Commerce (ICEC) '08 Innsbruck, Austria

hand, buyers can choose to participate on auctions that exhibit characteristics that generally lead to lower selling prices. Finally, the electronic auction system can provide specific services that will help buyers and sellers, increasing its popularity and revenues.

In this paper we propose a novel methodology for identifying and validating relationships among online auction inputs and outputs. Initially, we have to understand how these inputs correlate with each other. Intuitively, the isolated study of each effect of the inputs over the outputs reduces the chances of finding robust and consistent patterns. The two main reasons for this difficulty are that: (1) the input variables may present complex interactions and (2) there are possible confounds between the variables. Moreover, it is important to notice that variables that form auction inputs can have distinct purposes (e.g., create trust, describe the product being negotiated).

We consider that some relations between the variables that compose the auction inputs are so complex for most of the proposed methodologies. For example, we can mention the effect of interaction between the starting bid and the reputation over the auction results. Some previous work concludes that setting a low starting bid is important to attract a high amount of bids and raise the probability of success. But, on the other hand, some studies show that buyers are willing to pay high prices to products offered by sellers with high reputations. Therefore, setting a low starting bid could be a good strategy for sellers with low reputation, to increase the chances of success, but for sellers with high reputation, who already have higher probability of success, it could be better setting a medium or high starting bid to ensure a high ending price.

In our methodology, auction inputs are divided into three categories, according to their functionality: (1) seller characteristics, (2) auction configuration, and (3) product characteristics. Correlating inputs according to their functionality is useful for deeply understanding the mechanisms that drive the auction results. By correlating seller's characteristics we may identify seller profiles. In a similar way, correlating auction configuration inputs we can determine selling strategies. The product characteristics are useful to separate different strategies according to the item being negotiated. We will explain better the concept of seller profile and seller strategies in our characterization methodology (Section 4). In this work, we are specifically interested in investigating 2 hypotheses:

1. The seller profile affects the choice of the strategies used to configure auction inputs.

2. The effect of the selling strategy on auction results depends on the seller profile. A given selling strategy may be effective to lead to good results for some profiles, but not to others.

We chose these hypotheses to investigate because they address important questions about sellers and their practices in online auctions and previous methodologies have not already provided enough information to study them. We will test these hypotheses by a seller's perspective characterization methodology for online auctions using a real case study. Applications of our methodology include support decision strategies for buyers and sellers in online auctions. Also, studying the seller, the configuration and the product separately allows the online auction providers to offer specific services for their users, improving quality of service.

The remainder of this paper is organized as follows. Section 2 discusses some related work. Section 3 defines some basic concepts related to the proposed methodology. Section 4 describes our methodology, and Section 5 shows a real case study. Finally, Section 6 presents our conclusions.

2. RELATED WORK

This section describes some related works about online auctions, sellers, selling strategies and auctions results. Online auctions present several aspects that violate the common assumptions made by the traditional economic auction theory. The auction duration is typically much longer than in traditional auctions; bidders may come and exit at any time; bidders are geographically dispersed all over the world; they have very distinct backgrounds and it is hard to predict how many bidders will end up participating in the auction.

Online auctions have been studied extensively lately. Many studies focus on validating concepts from the classic economic theory of auctions in the online environment. For example, Lucking-Reiley [17] checks the validity of the well-known results of revenue equivalence. Bajari and Hortacsu [4] address how the starting bid, set by the seller, affects the winner's course. Gilkeson and Reynolds [11] show the importance of a proper starting bid price to attract more bidders and make an auction successful.

Studies about sellers have focused on reputation systems and trust in online auctions. Some of them have analyzed the importance of reputation in auction outputs, mainly in final prices. In [3], the authors investigate the effectiveness of reputation systems and how reputation correlates to auction results. They conclude that reputation plays an important role in trust and leads to higher ending prices. Resnick et al. [21] show that sellers with high reputation are more capable to sell their products, but the gains in final prices are reduced. Using a controlled experiment, Resnick et al. [22] study more accurately the reputation's impact on the auction outputs. The results show that, in general, bidders pay higher prices to sellers with higher reputation.

Some studies address how some factors are correlated with the auction results, a very related approach to our research. In [18], Bryan et al. analyze the impact of some seller characteristics and selling strategies in results of auctions negotiating collectible United States coins on eBay. Using regression analysis, the authors found out that the negative feedback is more informative than the positive feedback. In addition, they conclude that longer auctions lead to higher final prices and that the starting bid has low effect to the final prices.

In [2] the authors analyze the interrelationships between different variables of the auction, using correlation coefficients, for sales of the Palm Vx on eBay. They categorize sellers by their negotiation frequency during data collection. Two sellers with high amount of sales are defined as retailers. The results show that retailers set low starting bids and may attract more bids than any other type of seller. Moreover, they found out that sellers with high reputation are more able to describe their products. Becherer and Halstead [6] sent e-mail questionnaires to some sellers of eBay. Using factor analysis they study seller profiles and selling strategies. Besides the high costs of the experiment, the results show the diversity of sellers and business practices on eBay.

Our work differs from these previous studies because we propose a new methodology to characterize and evaluate the factors that guide the auction results. This methodology is based on the correlation among sets of auction inputs, before analyzing the auction results. Previous work used methods, like regression analysis and correlation coefficient that would not be able to control some factors while analyzing others, as emphasized in [22]. We believe some auction variables are strongly correlated. Our methodology goes towards capturing these correlations, identifying seller profiles and selling strategies, what enable us to find more meaningful patterns and providing a better understanding of the auction results.

3. DEFINITIONS

The set of variables that affects the auction results can be large and varied. Therefore, understanding how the auction variables are correlated with the negotiation results is a complex task. To deal with this complexity we distinguish the auction inputs by their characteristics and functionalities. In this section we define the categories of auction inputs we identify, what is useful for understanding the methodology proposed in this work.

- **Seller's characteristics**

  The inputs related to the seller provide information about the person who is willing to sell the product. Unlike conventional auctions, in which buyers can make direct contact with the seller, in online auctions this contact is restricted to this set of inputs. An auction system can provide a variety of information about the seller, such as its registration date on the system, a reputation metric, or a forum where buyers would comment their experiences with such seller. This information may provide trust to the buyer about the fidelity of the product description or if it will be actually delivered. Moreover, auction sites can offer some type of information sharing about sellers among buyers, limiting the action of bad sellers on the system with little or none intervention from system administrators.

- **Auction configuration**

  The set of variables directly related to a given negotiation is called auction configuration. As components of the auction configuration, we can cite the minimum starting bid value and the number of pictures about the product being auctioned. Differently from seller information, the auction configuration is a free choice of the seller. The online auction system may allow various information to be filled by the seller during the configuration step. Sellers may become experts on generating attractive configurations for their products, while other sellers may face difficulties during this task, due to lack of experience, available time or interest.

- **Product information**

  Online auction sites sell many different products, from different brands and conditions (new and used products, refurbished and even broken devices). It is important to analyze the product's features as one of the factors that affects auction results. We believe that product characteristics affect the strategies adopted by sellers. Moreover, some seller characteristics may have different impacts on the auction results depending on the product characteristics. Previous work tried to identify different correlation patterns between auction inputs and outputs, both for new and used products. One of the findings shows that the number of pictures provided is more important to determine auction success and ending prices for used than for new products. They argue that the information asymmetry regarding product quality is more pronounced for used items than for new items in electronic marketplaces due to a wide range of quality variation of used items [12].

  The next section describes our characterization methodology, based on these definitions.

4. CHARACTERIZATION METHODOLOGY

This section presents our methodology for characterizing online auctions, aiming at identifying selling patterns and correlating them with auction results. In order to achieve this goal, our methodology is based on three key points:

1. Instead of identifying correlations among inputs and outputs, we propose to first identify significant patterns among the inputs and then to correlate these patterns with the auction outputs.

2. Since some inputs may have different roles (e.g., create trust, describe the product being negotiated), it is necessary to group them into separate sets and consider each set separately to identify meaningful patterns.

3. The results obtained need to be validated using statistical techniques. Online auction results are guided by humans (buyers), which brings an inherent noise to the data. It is necessary to check whether the differences observed are significant with a certain degree of confidence.

The next subsections describe each step of our methodology.

4.1 Identifying Auction Inputs

The first step of the methodology is to identify the inputs that will be part of the characterization process. As previously explained, input attributes are predefined before the beginning of the auction negotiation. There are many techniques for attribute selection that can be applied to this context [16]. However, part of this responsibility may also be assigned to an expert, who has knowledge to consider semantic aspects about the online auction features.

As explained in the previous section, auction inputs may have different functionalities, which makes the process of identifying correlations with auction results more difficult. To address this problem, we divide the set of input variables into different groups: (1) seller's characteristics, (2) auction configuration and (3) product's characteristics.

The set of seller's characteristics leads to the identification of seller profiles. In addition, auction configuration analysis results in the identification of selling strategies. Product characteristics turn these analysis more homogeneous, since they give us some criteria for separating different products. It is important to take into account product characteristics and evaluate selling strategies for similar products.

4.2 Identifying Auction Results

After identifying the inputs of interest, it is necessary to define which auction outputs will be evaluated. Different outputs may be selected according to the goals of the characterization. Examples of auction outputs (results) are the ending price obtained for the product being auctioned, the success (or failure) of the auction and the number of bids attracted.

4.3 Choosing Clustering Technique

In order to identify seller profiles and selling strategies, we adopt a data mining technique called clustering [7], which can be used to identify clusters (groups) with similar characteristics in terms of their attributes.

Many clustering algorithms have been proposed by literature [8, 15, 1]. It is very important to choose the best algorithm based on the dataset characteristics (i.e., dimensionality, number of transactions). Our methodology is independent of the clustering algorithm chosen.

4.4 Data Treatment

It may be necessary to perform some data treatment routines [9, 14] before executing the clustering algorithm for identifying profiles and strategies. The appropriate data treatment technique may depend on the data characteristics and the clustering algorithm chosen. For example, some algorithms can deal with continuous attributes only. Examples of data treatment include discretization, normalization and others.

4.5 Determining Seller Profiles

The identification of seller profiles is based on the seller's characteristics. Homogeneous groups of characteristics identified through a clustering technique define seller profiles. After running the clustering algorithm, it is necessary to understand the profiles by their attributes. We use statistical metrics, such as the average, median and dispersion metrics (standard deviation, co-variance) to analyze the characteristics of each profile. Determining seller profiles can help us understand details of sellers, we will apply these details to study more accurately the results achieved by the selling strategies.

As an example, suppose we identify a given seller profile P_a, which exhibits high reputation and has been selling products for a long period. Sellers that present this profile may obtain a success rate higher than others of another hypothetical profile P_b, which presents low reputation and is composed by newcomers. We could infer possible reasons for a higher success rate for P_a as the impact of their reputation and experience.

4.6 Identifying Selling Strategies

Selling strategies are identified by grouping the set of inputs related to auction configuration. As described in Section 3, it is important to consider the product characteristics. Many product characteristics may be used for selling analysis, some examples include the type of product (video games or Cd's), the brand (Nintendo or Sony) and the condition (new, used, or broken). As explained in the last step (Section 4.5), a clustering technique is employed. The attribute values that define each strategy may be analyzed using statistical metrics, such as the average, median and dispersion metrics.

An example of a selling strategy would be a strategy S_a for new products, characterized by low starting bid and long duration. This strategy can be more successful than a hypothetical strategy S_b, employed for new products, characterized by higher starting bids and shorter negotiation periods. A possible reason for this result would be the fact that S_a attracts more bidders and the longer negotiation period allows the winning price to be gradually increased, resulting in higher ending prices than the auctions from S_b.

4.7 Correlating Seller Profiles and Strategies

After identifying seller profiles and selling strategies, it is important to analyze how each seller profile behaves, that is, which selling strategies they choose and at what frequency. This is a first step towards understanding the mechanisms that guide the results of online auctions. In this step, we may test our first hypothesis: Does the seller characteristics affect the strategies employed by her/him? Moreover, some sellers may choose more effective strategies than others. An example of this would be expert sellers that choose a combination of duration and starting bid that improves chances of getting high ending prices.

4.8 Analyzing the Results Obtained by the Profiles Applying the Selling Strategies

Assuming that seller profiles and selling strategies have been found, considering the product characteristics, and that seller profiles and selling strategies have been correlated, the next step is to analyze these profiles, strategies and auction results in order to identify aspects from these correlations that may explain auction results. By this analysis, we test our second hypothesis: Do auctions results obtained by the selling strategies depend on the seller characteristics?

In order to compute these correlations, we need to combine seller profiles with selling strategies to analyze each of the results of interest. The analysis of average value of the results are not enough for establishing final conclusions. It is necessary to validate results through the use of statistical methods. We suggest the use of the t-test [14] for this step.

The next section describes in detail our case study, where we apply the proposed methodology.

5. CASE STUDY

This section presents our case study, which evaluates the methodology presented in Section 4 using an eBay dataset. We aim to validate the methodology for distinguishing the actors (sellers) from their practices (selling strategies). Moreover, we are interested in analyzing the wealth of details provided about the online auction mechanisms. To guide the evaluation process, we test two hypotheses about sellers and their selling strategies.

We apply our methodology to a real dataset that consists of 1014 eBay auctions for Nintendo GameCube from 05/25/2005 to 08/15/2005. There are 102 auctions for new items and 912 for used. In order to obtain homogeneity, we do not consider auctions for broken products, which set 'reserve price' or the 'buy it now' [10] option.

The next subsections describe the results of applying each step of our methodology to this case study.

5.1 Identifying the Auction Inputs

As we described in Section 4.1, the first step of our methodology is to identify auction inputs. Next subsections present the selected auction inputs:

5.1.1 *Seller's characteristics*

Out of all the pieces of information about sellers provided by eBay, we have selected the following subset of meaningful inputs:

- **SERA**: Seller's reputation score on eBay. It is the sum of the rates given by the bidders who negotiate with such seller on eBay. The given rates can be -1 (negative), 0 (neutral) or +1 (positive).
- **POFB**: Percentage of positive rates given to the seller.
- **TREG**: Time (days) since the seller has been registered on eBay.
- **POSE**: Status given by eBay to the most successful sellers. Its value can be 1 or 0, if the seller is a power seller or not, respectively. The criteria used for being a power seller include consistent sales volume and high positive feedback.

As seller reputation is cumulative, and most of the feedback given is positive, we can consider it in conjunction with percentage of positive feedback's to analyze the seller experience [21]. Thus a seller owning a high reputation and a high percentage of positive feedback should be considered an experienced seller. The seller's reputation and feedback are two traditional attributes, studied by previous work. We added the time that seller has been registered because this attribute, in conjunction with the seller's reputation, provides a measure of the seller activity (approximating the number of auctions in a given time interval) on eBay. Analyzing the power seller attribute we expect to distinguish this more successful group of sellers and study their practices.

5.1.2 *Auction Configuration*

From the items that compose the auction configuration on eBay we selected:

- **STBI**: Starting bid, that is the minimum value for a bid for the item offered.
- **TIME**: Auction duration (in days).
- **FILESIZE**: Size of the file used by the seller to describe the product, in KBytes.
- **PIC**: Number of pictures used by the seller to show the product he is willing to sell.

Starting bid and auction duration are two critical decisions for sellers on eBay. We analyze the file size and number of figures to capture the effects of the description over the results obtained by the selling strategies.

5.2 Identifying Auction Results

After identifying the auction inputs, the next step is to select the auction results, as described on Section 4.2. The auction results analyzed in this case study are:

- **SUCCESS**: Success rate of set of auctions. Defined as the number of auctions that achieve success (i.e., finish with a selling).
- **WIBI**: Winning bid or ending price.
- **NOBI**: Number of bids attracted by the auction.

5.3 Clustering Technique

In order to identify seller profiles and selling strategies, as described in Section 4.3, we employ the k-means algorithm [13]. The reasons for choosing k-means are: (1) performance, (2) the concept of centroid, used by k-means and (3) the simplicity of the algorithm. A centroid is an imaginary point that has the average properties of a given cluster, so we can use it to represent that cluster.

The number of clusters (k) is a parameter of the k-means algorithm. We identify the best number of clusters to be generated through three clustering evaluation quality metrics, namely, the beta-cv, the beta-var and the squared error [19, 14].

5.4 Data Treatment

As we described in Section 4.4, we treated the data to improve the quality of the characterization results. Auction with inconsistent data and outliers were removed from the dataset. We also removed auctions selling more than one item to reduce the noise due to product heterogeneity.

Employing k-means on attributes with skewed distributions may result on weak results, as the difference between the values may not be representative. Therefore, distribution functions and item descriptions given by the seller played an important role in the data treatment process. For example, the probability distribution of the reputation values (SERA) shows that that there are many sellers with low reputation and a small group with high reputation. Employing k-means for clustering reputation may result on large groups with low reputation and small groups with high reputation. To address this problem, we applied the mathematical function logarithm to reduce the impact of the differences between the values of some attributes (STBI, PIC and SERA). Moreover, to set the same weight to all the attributes we normalized them in the interval (0,1).

5.5 Seller Profiles

In order to identify seller profiles we executed k-means algorithm for different values of k on the seller attributes, as described in Section 4.5. The best value found for k was 6.

Table 1 describes each group in terms of their frequency in the dataset, the average values and coefficients of variation of the attributes evaluated on the clustering process. Table 2 shows the average output results for each profile. We will then characterize and understand each identified profile. All the comparisons among the results of the profiles were validated with an 80% of confidence t-test [14]. We emphasize that, although we use just the input variables to identify the profiles through the clustering algorithm, along this section we analyze the profiles by both their characteristics and the results obtained by them.

P0: Experienced profile. It exhibits the longest average registration period, high reputation and high percentage of positive feedback. This profile achieves ending prices higher than all other profiles, except when compared to profile P1. It does not have high success rates, and just an average number of bids per auction is observed.

P1: Expert profile. This profile presents long registration time, high positive feedback and medium reputation. P1 obtains good results in terms of ending prices and success rates. It suggests that this profile is already able to obtain results as good as those achieved by P0, although they have less experience.

Table 1: Seller profiles

	log(SERA)		POFB		TREG		POSE	
	AVG	CV	AVG	CV	AVG	CV	AVG	CV
P0 (12%)	5.43	0.25	98.97	0.02	2082.46	0.15	0	-
P1 (23%)	3.52	0.32	98.47	0.03	1387.71	0.22	0	-
P2 (22%)	1.38	0.62	98.13	0.07	337.54	0.89	0	-
P3 (29%)	3.88	0.22	98.82	0.02	510.32	0.52	0	-
P4 (5%)	6.54	0.18	99.07	0,01	981.17	0.75	1	-
P5 (9%)	0.00	0.00	1.58	5.54	194.66	1.87	0	-
GLOBAL	3.54	0.50	89.51	0.32	854.89	0.82	0	-

Table 2: Average values of outputs for seller's profile

PROFILE	WIBI (US$)	SUCCESS	NOBI
P0	75.83	0.88	9.51
P1	77.19	0.91	9.51
P2	70.64	0.81	8.52
P3	61.90	0.91	9.51
P4	58.93	0.91	12.62
P5	64.58	0.63	6.34
GLOBAL	69.19	0.86	9.14

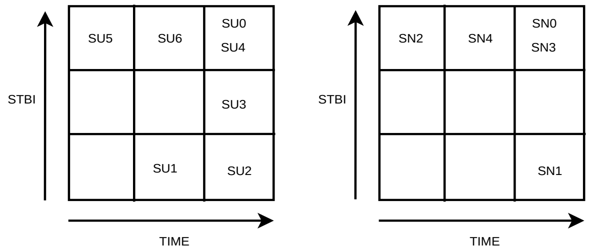
P2: Beginner profile. P2 has a low reputation, high feedback and short registration time. Although P2 represents novice sellers, its members reach higher ending prices than P3, P4 and P5 members, as we present next. However, P2 is not able to achieve similar results in success rate and number of bids attracted.

P3: Intermediate profile. P3 has intermediate values for reputation and registration time. Although the members of this profile have a half of the registration time of the members of P1, they have a similar reputation. Members of P3 have high positive feedback. They reach low ending prices but attract an average number of bids and achieve a high success rate.

P4: Power sellers. They do not have a long registration time, but present the highest reputation and highest positive feedback. P4 is also a very active profile (i.e., they sell frequently on eBay). P4 achieves the lowest ending prices, except P3. The success rate of these sellers is high, comparable to P1 and P3, although it is important to notice that P4 sellers attract more bids than P1 and P3 sellers.

P5: Unsuccessful profile. P5 has the highest heterogeneity of attributes among the identified profiles, which indicates it can be a mixture of profiles. P5 has the shortest registration time and the lowest reputation and positive feedback. Sellers from P5 also achieve low prices. Success rate and number of bids attracted by P5 are the lowest among the identified profiles.

The characterization of the seller profiles lead us to some interesting conclusions. P4 (power sellers) has the highest reputation among the identified profiles, but obtains low ending prices, while P0, other profile with high reputation, achieves high ending prices. As previously described, the criteria to be a power seller include a consistent sales volume, but do not include high ending prices. P4 reaches a success rate higher than P0, with 75% of confidence. Some previous work [2, 21], while studying another database collected from eBay, also found a low correlation between seller reputation and ending price. We believe that this is a consequence of the different interests of sellers on eBay. P4 sellers present high volume of sales, but reach the lowest ending prices, which can be explained by the fact that they probably bargain for good prices with their suppliers since the high volume of items they used to buy.



(a) Used products (b) New products

Figure 1: STBI X TIME diagrams

5.6 Selling Strategies

In order to identify selling strategies, k-means algorithm was executed for different values of k over the auction configurations. As described in Section 4.6, we consider the product characteristics to split the set of configurations. We use the following product characteristics from our database: (1) the product (video game), (2) the brand (GameCube) and (3) the condition (new or used). We selected one product of one brand in this case study. The number of products and brands can vary according to the goals of the characterization. Both used and new items were considered. The best value found for k was 5 and 7 for new and used products, respectively.

Tables 3(a) and 3(b) describe each group by average values and the coefficient of variation of the evaluated attributes for used and new items, respectively. Since we apply a *log* function to attribute STBI, the negative values correspond to very low prices. Table 4(a) contains the average values of the auction results of each strategy for used items, and Table 4(b) for new items.

To simplify analyzing the selling strategies we construct two classification systems. The first classification is about the interaction between the starting bid (STBI) and the auction duration (TIME). The second one is about how the description resources are employed (PIC and FILESIZE). Then, we classified the strategies into the STBI X TIME diagrams (Figure 1). About the use of the description resources, we analyze the number of pictures of the product and the description file size to evaluate if a strategy is descriptive. All the comparisons among the results of the strategies were validated with an 85% of confidence t-test. Next, we characterize each identified strategy for both of them.

Table 3: Selling strategies

(a) Used products

	log(STBI(US$))		TIME(days)		FILESIZE(KBytes)		log(PIC)	
	AVG	CV	AVG	CV	AVG	CV	AVG	CV
SU0 (27%)	3.75	0.21	7.10	0.08	42.35	0.04	1.25	0.15
SU1 (5%)	-1.04	-1.92	2.36	0.44	43.73	0.06	1.41	0.18
SU2 (11%)	-1.13	-1.75	6.74	0.16	44.96	0.07	1.32	0.18
SU3 (4%)	2.40	0.63	5.79	0.33	56.80	0.07	2.41	0.30
SU4 (15%)	3.23	0.25	6.84	0.14	47.58	0.05	1.47	0.16
SU5 (13%)	3.96	0.19	0.72	0.89	44.92	0.08	1.37	0.21
SU6 (24%)	3.49	0.23	3.85	0.25	43.49	0.06	1.28	0.18
GLOBAL	3.00	0.64	4.86	0.51	46.13	0.09	1.36	0.23

(b) New products

	log(STBI(US$))		TIME(days)		FILESIZE(KBytes)		log(PIC)	
	AVG	CV	AVG	CV	AVG	CV	AVG	CV
SN0 (18%)	3.64	0.23	6.55	0.13	46.89	0.04	1.30	0.17
SN1 (14%)	-0.99	-1.90	5.42	0.31	45.76	0.06	1.20	0.14
SN2 (20%)	3.75	0.28	1.43	0.9	43.82	0.05	1.32	0.17
SN3 (28%)	3.73	0.18	6.96	0.11	42.82	0.03	1.30	0.16
SN4 (20%)	3.42	0.25	3.50	0.51	52.95	0.06	1.65	0.27
GLOBAL	2.79	0.75	5.09	0.48	44.65	0.09	1.36	0.25

5.6.1 Used Products

SU0 and SU4 - Cautious strategies: Employ high starting bids and long auctions. They are usual strategies to reach high ending prices and maximize the chances of selling. In terms of the results, these strategies achieve different ending prices and success rates. Part of this difference can be explained by the starting bids employed by SU0 and SU4. The lower starting bid of SU4 leads it to a lower ending price but a higher success rate in comparison to SU0. The success rate achieved by SU0 is the lowest one among the selling strategies for used products.

SU5 - Risky strategy: Employs high starting bids and short auctions. This combination has high risk of failure, but, if successful, ensures a high ending price. The number of bids attracted by this strategy is the lowest one among the strategies for used items.

SU2 - Force selling strategy: Applies low starting bids and long auctions, a combination that encourages the participation of bidders and allows the price to raise during the auction. The number of bids attracted is the highest among the strategies for used products. The success rate obtained is high but the ending price is low.

SU3 - Descriptive strategy: Employs large description files and many figures. Moreover, it maximizes the chances of selling, by long auctions, but avoids low ending prices (as the prices achieved by SU2), by setting a medium starting bid. The results obtained are a high success rate but a ending price even higher than SU2.

SU1 - "Rushed" strategy without ensuring about ending prices: Applies medium starting bids and auction durations. The results obtained are an intermediate success rate and the lowest ending price among the strategies for used products. Comparing the results of SU1 and SU2, we can see that the longer duration employed by SU2 allows price developing during the auction, leading to higher ending prices.

SU6 - "Rushed" strategy ensuring the ending price: Employs medium durations, but sets a high starting bid, to ensure that ending prices will not be very low. The results are intermediate ending prices and success rate.

5.6.2 New Products

SN0 and SN3 - Cautious strategies: They are similar to SU0 and SU4, for used products. However, the results obtained by these strategies are different. We believe that it is because some factors are not been considered in the selling identification and analysis. Furthermore, we have a small sample of new products, which reduces the chances of finding consistent and significant patterns for these products.

SN1 - Force selling strategy: This strategy is similar to SU2, to used products. The results achieved by this strategy are low ending prices, one of the most successful strategies for new products and the highest number of bids attracted among the strategies for new products.

SN2 - Risky strategy: This strategy is similar to SU5, for used products, high starting bid and short duration. The results achieved are few bids attracted, low success rate and an intermediate ending price.

SN4 - Descriptive strategy: This strategy is similar to SU3, in terms of description, but it employs high starting bid and medium durations, similar to SU6. The results achieved are intermediary ending prices and success rate.

The next section correlates the presented strategies with the seller profiles.

5.7 Correlating Seller Profiles and Strategies

In this section we will study the strategies applied by each profile, as described in Section 4.7. Our objective is to understand better the results achieved by each profile by their practices, and test the hypothesis that there is a strong correlation between the seller profile and the strategies employed by it.

For new items (see Figure 2(a)) there are not predominant strategies, different profiles act in diverse ways for products in this condition. However, according to Figure 2(b), there are two predominant strategies for used items: SU0 and SU6. As we have already stated, SU0 and SU6 do not obtain high success rates or attract many bids, but SU0 achieves high ending prices. Starting from the configuration point of view, we can see that both strategies are very simple, they make

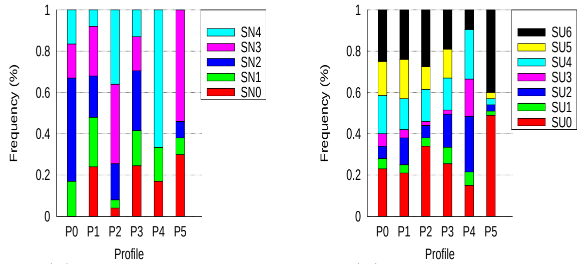
Table 4: Average values of outputs for selling strategy

(a) Used products

STRATEGY	WIBI(US$)	SUCCESS	NOBI
SU0	71.57	0.74	6.67
SU1	52.03	0.83	13.83
SU2	68.17	1.00	20.54
SU3	82.99	1.00	14.86
SU4	67.95	0.97	10.72
SU5	75.83	0.81	2.19
SU6	63.65	0.81	7.63
GLOBAL	68.69	0.85	9.14

(b) New products

STRATEGY	WIBI (US$)	SUCCESS	NOBI
SN0	95.34	0.89	8.67
SN1	65.64	1.00	16.64
SN2	68.55	0.90	7.57
SN3	64.99	0.83	7.45
SN4	67.87	1.00	7.70
GLOBAL	71.66	0.91	9.00



(a) New products (b) Used products

Figure 2: Frequency of selling strategy per seller profile

little use of description resources and employs high starting bids. In terms of duration, SU0 is based on long auctions and SU6 adopts intermediate durations.

P0 and P1 obtain high ending prices, high success rates and attract an intermediate number of bids. They have similar behavior patterns, mainly for used items. Both profiles apply, frequently, strategies to ensure high ending prices. While P1 sellers uses strategy SU2 more often than P0, the latter group chooses strategy SU0 and SU4 more than P1. SU2 achieves better results than SU4, in terms of ending prices and success rate. Comparing these profiles leads us to believe that, although P0 has superior characteristics, mainly higher reputation and longer registration time, P1 applies good strategies, reaching similar results.

In the case of profile P2, the predominant strategies for new products are SN3 and SN4. Both strategies are based on ensuring high ending prices, although SN4 is a descriptive strategy and SN3 apply longer auctions. For used items, we can notice that the dominance of the strategies SU0 and SU6 is higher than we observed for P0 and P1 profiles. The strategies SU3 and SU5, which achieve high ending prices and success rates, are less often used by P2 than by the others, except P5. The strategies applied by P2 lead us to understand some reasons that explain the ending price and success rate achieved by this profile, which are worse than the previous profiles.

P3 applies a diverse set of strategies. The frequency of the strategy SU6 is lower for it in comparison with the previous profiles, although the strategies SU1 and SU2 are applied more frequently. Both strategies apply low starting bids. SU1 obtains very low ending prices and SU2 gets medium ending prices, what indicates, in part, why P3 achieves one of the lowest ending prices among the identified profiles.

P4 presents a singular behavior pattern. While selling new and used products, P4 applies strategies that achieve high success rate and attract many bids. The strategy SN4, for new items, is one that obtains the highest success rates. SU2, SU3 and SU4 are the strategies with highest success rates for used items. Moreover, SU2, which attracts more bids than any identified strategy, is the most common strategy of this profile. The strategies applied by this profile usually do not ensure high ending prices, but set long durations to increase the chances of success. As can be seen, the strategies employed by P4 explain the results achieved by this profile: low ending prices, high success rates, and attraction of many bids.

P5 has the highest dominance of strategies SU0 and SU6, for used items, while for new products the more frequent strategy is SN3. These strategies do not achieve good results in terms of ending price, except SU0, and success rate, and also do not attract many bids.

The results presented in this section corroborate our first hypothesis, confirming that sellers apply different strategies and the seller profile affects the adoption of different strategies to configure auction inputs.

5.8 Analyzing the Results Obtained by the Profiles Applying the Selling Strategies

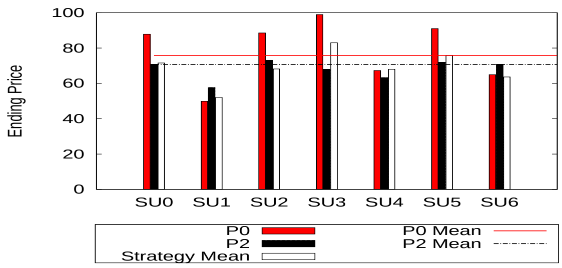
In this section we describe the last step of our methodology, which provides a way to better understand the patterns that correlate auction inputs and outputs, as described in Section 4.8.

In the last section, we analyzed how each seller profile chooses its strategies. Now, we analyze the results achieved by each seller profile when different selling strategies are applied. Due to space constraints, we will not present every combination of seller and strategy in this paper. We selected two seller profiles and performed a detailed study of how their strategies affect the auction results for used products. This analysis will lead us to verify our second hypothesis, confirming that the effect of the selling strategies depends on the seller profile.

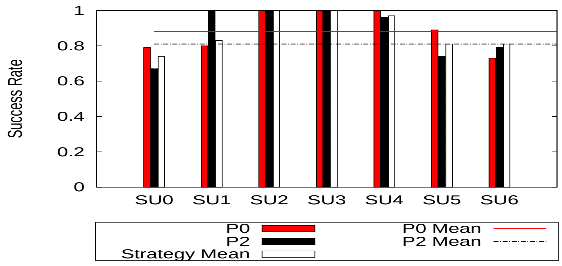
We have chosen seller profiles P0 and P2 for this analysis, since they exhibit different characteristics and very different results. Profile P0 is better than P2 in terms of reputation, longer registration date and percentage of positive feedback. Our goal is to identify the impact of these differences when different groups of sellers apply the same strategies. We may also be able to identify which strategies are more suitable for each seller profile. All the comparisons among the results of the profiles applying the selling strategies were validated with an 80% of confidence t-test.

Figure 3 shows the results achieved by profiles P0 and P2 for each strategy, in terms of ending price, success rate and number of bids attracted, respectively. Moreover, auction results for P0 and P2 are also compared with the average results of the strategies.

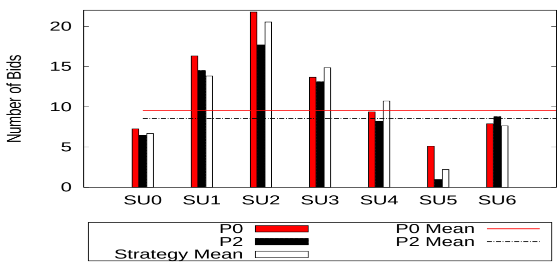
We can notice that, with few exceptions, sellers from P0



(a) Ending price



(b) Success rate



(c) Number of bids attracted

Figure 3: Results obtained by P0 and P2 applying each selling strategy for used products

achieve higher ending prices, higher success rates and attract more bids than sellers from P2. P2 beats profile P0 in terms of ending price and success rate only when its members apply strategy SU1. Considering the number of bids attracted, P2 beats P0 in just one strategy (SU3). It shows that the profile P0, which has better characteristics (higher reputation, higher positive feedback and longer registration time than P2), is more able to obtain higher ending prices, higher success rates and to attract more bids than P2.

We selected the strategy that achieves the highest average ending price when applied by P2 and compared with the results achieved by P0 applying each other strategy. Our goal is to show that, even though P2 exhibits inferior attributes (e.g., lower reputation) than P0, P2 may obtain comparable or even better results in some scenarios. P2 achieves better ending prices applying strategy SU2 than P0 applying strategy SU4. In the case of strategy SU1, we can state that P2 reaches a success rate that is comparable to the ones achieved by the best strategies of P0.

In general, the results achieved by analyzing the strategies without considering the profiles were the same for P0 and P2. For example, SU1 was identified as a strategy that achieves low ending prices on average and this trend is seen for both profiles P0 and P2. However, strategy SU3 is identified as the one that achieves the highest ending prices on average, but this result was not confirmed when applied by profile P2. It shows that it is important to consider the profile to analyze the strategies. Although SU3 achieves high ending prices in general, this does not occur when used by members of P2.

A more interesting situation occurs when we consider strategy SU5, identified as the worst strategy in attracting bids. We can notice that SU5 results are very different, for this metric, when correlated with profiles P0 and P2. Sellers from profile P2 attract fewer bids than profile P0 when strategy SU5 is applied. Moreover, success rate and ending prices achieved by P0 when applying SU5 are far better than the ones achieved by P2. This may indicate a risky strategy is more recommended for sellers from P0 than for sellers from P2. SU5 is a strategy that demands high feedback and reputation from the seller.

The results analyzed in this section show the importance of considering sellers and their strategies separately. Although we identify sellers with positive and negative characteristics in terms of reputation, feedback and others, it is important to take into consideration the selling strategies employed by them. On the other hand, successful or unsuccessful strategies (regarding some auction result indicator, e.g., ending price, number of attracted bid, etc.) may present different results depending on the seller who applies each of them. These results confirm our second hypothesis: the effect of the selling strategies depends on the seller characteristics. When we study both the sellers and their practices, we are more able to understand the auction results.

6. CONCLUSIONS AND FUTURE WORK

In this work we propose a new methodology to characterize important aspects of online auctions from the seller's perspective. Based on two hypotheses, we apply the proposed methodology to a case study, using a real eBay dataset. Our methodology allows identifying seller profiles and selling strategies, considering the characteristics of the product being auctioned. This analysis has lead us to a better comprehension of how seller profiles and selling strategies are correlated and how they affect the auction results.

The results validate our hypotheses that: (1) the seller profile is correlated with the selling strategies adopted by her/him, (2) the seller profile affects the results achieved by the selling strategies. The validation of the first hypothesis shows that the seller behavior is not random, that is, different sellers adopt their strategies according to their interests, capacities and experience. The second hypothesis suggests that choosing a selling strategy is not simple, since it is important to consider the seller's characteristics to evaluate the applicability of a strategy. Moreover, the second hypothesis indicates the importance of recommendation services in order to provide a support decision tool to select a proper configuration set for the auction.

Besides the validation of the hypotheses, the methodology leads us to some interesting findings. We emphasize the analysis of the power seller profile (P4) on eBay, which sells a large amount of products by low prices. These sellers adopt their strategies in a singular way, achieving high success rates and attracting many bids. Another important finding is the identification of two predominant strategies for used products on eBay. These strategies do not obtain neither high success rates nor many bids attracted, which is a motivation to provide recommendation services to sellers.

As future work we are going to validate our methodology using other datasets and apply it to design a recommendation system for sellers on eBay. Moreover, we want to understand more deeply the mechanisms that guide the auction results, analyzing the negotiation patterns attracted by sellers and their selling strategies. Another future work is to study seller's evolution, evaluating how they change their profiles and strategies over time.

7. ACKNOWLEDGEMENT

This work was partially supported by CNPq, CAPES, Finep, and Fapemig.

8. REFERENCES

[1] E. Achtert, H.-P. Kriegel, A. Pryakhin, and M. Schubert. Hierarchical density-based clustering for multi-represented objects. In *1st Int'l Workshop on Mining Complex Data MCD*, Houston, TX, November 2005. IEEE.

[2] S. Anderson, D. Friedman, G. Milam, and N. Singh. Seller strategies on ebay. Industrial Organization 0412004, EconWPA, Dec. 2004.

[3] S. Ba and P. A. Pavlou. Evidence of the effect of trust building technology in electronic markets: price premiums and buyer behavior. *MIS Quarterly*, 26(3):243–268, 2002.

[4] P. Bajari and A. Hortacsu. The winner's curse, reserve prices, and endogenous entry: Empirical insights from ebay auctions. *RAND Journal of Economics*, 34(2):329–55, 2003.

[5] R. Bapna, P. B. Goes, and A. Gupta. Insights and analyses of online auctions. *Communications of the ACM*, 44(11):42–50, 2001.

[6] R. C. Becherer and D. Halstead. Characteristics and internet marketing strategies of online auction sellers. *Int'l Journal of Internet Marketing and Advertising*, 1(1):24–36, 2004.

[7] H.-H. Bock. Data mining tasks and methods: Classification: the goal of classification. pages 254–258, 2002.

[8] S. Chelcea, A. D. Silva, Y. Lechevallier, D. Tanasa, and B. Trousse. Pre-processing and clustering complex data in e-commerce domain. In *1st Int'l Workshop on Mining Complex Data MCD'2005*, Houston, Texas, November 2005. IEEE.

[9] T. Dasu and T. Johnson. *Exploratory Data Mining and Data Cleaning*. John Wiley & Sons, 2003.

[10] ebay. http://www.ebay.com.

[11] J. Gilkeson and K. Reynolds. Determinants of internet auction success and closing price: An exploratory study. *Psychology & Marketing*, 20(6):537–566, 2003.

[12] P. Góes, Y. Tu, and Y. A. Tung. Auction segmentation and selling strategies in electronic markets. *Working Paper, University of Connecticut*, 2007.

[13] J. Hartigan. *Clustering Algorithms*. John Wiley and Sons, Inc., New York, NY, April 1975.

[14] R. Jain. *The Art of Computer Systems Performance Analysis: Techniques for Experimental Design, Measurement, Simulation, and Modeling*. Wiley-Interscience, New York, NY, April 1991.

[15] M. Kearns, Y. Mansour, and A. Y. Ng. An information-theoretic analysis of hard and soft assignment methods for clustering. In *Proc. of the 13rd Conf. on Uncertainty in Artificial Intelligence*, pages 282–293, Providence, Rhode Island, USA, 1997.

[16] H. Liu and H. Motoda. *Feature Selection for Knowledge Discovery and Data Mining*. Kluwer Academic Publishers, Norwell, MA, USA, 1998.

[17] D. Lucking-Reiley. Using field experiments to test equivalence between auction formats: Magic on the internet. *American Economic Review*, 89(5):1063–1080, 1999.

[18] D. Lucking-Reiley, D. Bryan, N. Prasad, and D. Reeves. Pennies from ebay: The determinants of price in online auctions. *Journal of Industrial Economics*, 55(2):223–233, June 2007.

[19] D. A. Menasce and A. F. A. Virgilio. *Scaling for E Business: Technologies, Models, Performance, and Capacity Planning*. Prentice Hall PTR, Upper Saddle River, NJ, USA, 2000.

[20] C. Plott. Rational choice in experimental markets. *The Journal of Business*, pages 301–327, 2001.

[21] P. Resnick and R. Zeckhauser. Trust among strangers in internet transactions: Empirical analysis of ebay's reputation system. *The Economics of the Internet and E-Commerce*, edited by M.R. Baye. Amsterdam: Elsevier Science B.V.:127–157, 2002.

[22] P. Resnick, R. Zeckhauser, J. Swanson, and K. Lockwood. The value of reputation on ebay: A controlled experiment. *School of Information, University of Michigan*, Ann Arbor, Michigan, USA:34, 2003.

[23] Yahoo! auctions. http://auctions.yahoo.com.