Topological and Optimization Modeling for Internet Data of Online Auction Markets

Sydney C. K. Chu[1,*]   James K. Ho[2,†]

[1] Department of Mathematics, University of Hong Kong, Pokfulam Road, Hong Kong
[2] Department of Information and Decision Sciences, University of Illinois at Chicago, Il 60607, USA

Abstract   We advocate a topologically driven optimization model for the intriguing question of what "shape" the online market data is in. Our model bases on operational data from sources like eBay.com to visualize the dimensions of the markets and determine maximally the resolution of buyer/seller market efficiency or dichotomy.

Keywords   e-commerce; online auctions; market analysis; optimization modeling

1   Introduction

Online auctions have demonstrated how e-commerce can transform business, and not merely transplant conventional processes to a new medium. The resulting emergence and growth of diverse markets pose the intriguing question of what 'shape' a given market is in at a particular moment of development. This work aims to integrate optimization to our recently constructed topological model (Ho 2004) based only on operational data, without any expert knowledge of the specific auction market, or economic details from the transactions. Our topological analysis uses extensive data available on eBay.com, the foremost online auction platform to date, the dimensions for a topology are identified. A graphical model is then proposed to visualize this topology, for the purpose of comparing markets at different points of a life cycle, and whether they are efficient, or tend to flavor either buyers or sellers. The Operations Research technique of optimization modeling is then integrated into this approach to achieve its maximum resolution of such a buyers/sellers market dichotomy. Its currently on-going implementation will take the form of a decision support system.

2   Modeling of online auction markets

The Internet is a new medium of communication connecting potential partners in trade worldwide. The initial frenzy over the internet's promises led to grossly

---

*E-mail: schu@hku.hk
†E-mail: jimho@uic.edu

exaggerated valuations of business models that were mere transplantations of existing processes to the alternative channel. Now that the bubble has burst (Perkins and Perkins 2001), more sensible and critical thoughts can be turned to true transformations that are creating and nurturing markets of the future. Online auction is one of the very few cases that has held a steady course (Klein 1997, Lucking-Reiley 2000, Turban 1997), as evidenced in the success to date of eBay.com.

Founded in September 1995, eBay has become a global trading platform where on any given day, there are more than 16 million items listed across 27,000 categories. In 2003, at least 30 million people bought and sold well over $20 billion in merchandise, so that the entire culture it engenders is now being described as the eBay economy (Hof 2003). It is also the richest source of data for online auction markets, as records of all its transactions are available to the public (on a 2-week rolling basis).

The objective of our investigation is to pose the intriguing question of what "shape" a given market is in at a particular moment of development, and to produce models with tools for the answer. The key issues of our modeling work include:

- To construct a topological model based only on operational data, without any expert knowledge of the specific auction market, or economic details from the transactions.
- To identify the dimensions for a topology, using extensive analysis of eBay data to extract the information desired.
- To propose a graphical model to visualize this topology, also for the purpose of comparing markets at different points of a life cycle.
- To integrate optimization modeling into this topological analysis to further achieve its maximum resolution of any such buyers/sellers market dichotomy.

With online auctions removing conventional constraints on time and space, their activities and impact on e-commerce can only be expected to grow exponentially. Therefore the availability of operational data from eBay presents unprecedented challenges and opportunities to gain insight into online auction markets, and provide decision supports for a correct course of actions. Our project's relevance lies in this innovative auction markets that are already upon us in this modern days of e-commerce — an economy that is clearly expected to last.

2.1   Shape of markets

In a most recent work (Ho 2004), we pose the intriguing question of the "shape" of a given market at a particular moment or over its development. The work is a topological model based only on operational data, with extensive analysis of eBay data in order to identify the dimensions for a topology. Excel macro programs are designed and used to perform the substantial data mining necessary to extract the information desired. A graphical model is then proposed to visualize this topology, for the purpose of comparing markets at different points of a life cycle. Examples of application of the model are presented there.

2.2   High-dimensional data visualization

Visualization has been a fast developing approach in data-mining (Hoffman and Grinstein 2001) in which graphical models are constructed to provide visual cues for pattern recognition and knowledge discovery from complex data. In the study of financial markets (stock and commodity), the dimension of interest is primarily prices, or the fluctuation thereof. Complexity arises from the large number of instruments involved. The best known examples of visualization models for stock markets are based on the tree-map method (Shneiderman 1992, Wattenberg 1999), and the minimum-spanning-tree method (Vandewalle et al 2001). For auction markets, the game-theoretic dynamics itself gives rise to higher dimensional complexity. And with online auctions removing conventional constraints on time and space, their activities and impact on e-commerce can only be expected to grow exponentially.

In this respect, the availability of operational data from eBay presents unprecedented challenges and opportunities for insight into online auction markets. In our current work cited before (Ho 2004), we define a topological model for an online auction market to be a simultaneous graphical display of all the dimensions of its relevant database, which provides a geometrical shape as a descriptive, visual statistics of the market. The purpose is to identify the dimensions from available data, towards constructing a specific topological model that can help discern market efficiency. We seek a visual cue for whether a market is favorable to buyers or sellers without expert knowledge of the items involved or the prices attained.

Regarding visualization (Hoffman 2001, Keim 2002), our starting basis is the simplest yet visually most power display of multi-dimensional data known as the star plot, or glyph, (Chambers et al 1983). Such an illustration is given in Fig. 1 below. (The labellings of the axes are explained in a later section below.)
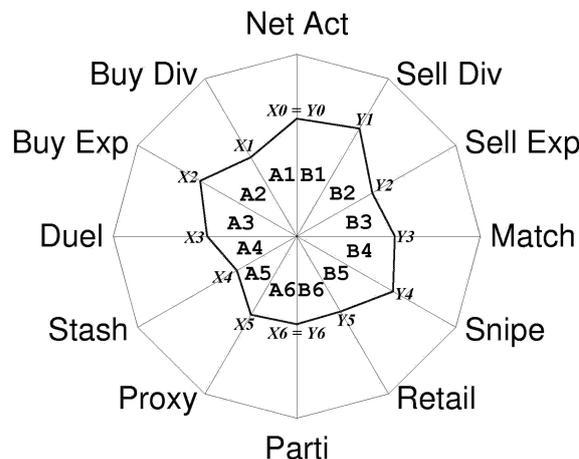
Figure 1: Star plot of 12 attribute averages from eBay data of 34 product categories

2.3 Optimal topology

As for methodology for analysis, many methods in the literatures have been explored and well studied for the problem of this aggregating multiple and incommensurate attribute functions. The approach developed in this work will be a departure from classical methodology in the following sense. While an aggregate measure is sought, it is to discern the relative dominance of the two parts of a dichotomy and no pair-wise comparison of outcomes, or preference evaluation under uncertainty is invoked. The aggregate measure is prepared for an optimization model which is to provide maximum resolution in discerning dominance.

The integration of optimization modeling approach (Chu et al 2003) with its implementation of decision support system is adopted. The aim is to pinpoint the highest possible resolution of buyers- versus sellers-market whenever such a dichotomy exists on its topology (i.e. its glyph).

2.4 Dimensions of markets

While it is obvious and well known that the shape of a star glyph depends on the configuration of the attributes along the radial dimensions of the plot, classical Exploratory Data Analysis (EDA: Tukey 1977) to date simply encourages analysts to investigate the variations on an ad hoc basis. Our first step toward a more systematic approach is motivated by the topological analysis of online auction markets (Ho 2004). Twelve dimensions (i.e. attributes) are identified, with their (initial) configurations illustrated in Fig. 1 given above:

1. NET ACTIVITY (auctions with bids)
2. PARTICIPATION (average number of bids per auction)
3. SELLER DIVERSITY (distribution of offers)
4. SELLER EXPERIENCE (distribution of sellers' ratings)
5. MATCHING (auctions ending with a single bid)
6. SNIPING (last minute winning bids)
7. RETAILING (auctions ending with the Buy-It-Now option)
8. BUYER DIVERSITY (distribution of bidder participation)
9. BUYER EXPERIENCE (distribution of buyers' ratings)
10. DUELING (evidence of competitive bidding)
11. STASHING (evidence of stock-piling)
12. PROXY (use of proxy bidding as evidence of true valuation)

The dichotomy results from attributes 3 through 7 reflecting market conditions for buyers, and attributes 8 through 12 reflecting that for sellers. Attributes 1 and 2 are common parameters of the market and considered neutral to buyers and sellers.

2.5 Topologically driven optimization model

To further the dimensional analysis, we adopt the topological modeling approach (Vandewalle et al 2001). Given the star glyph of a multi-attribute dichotomy,

it will be both visually and intuitively appealing if the areas covered by the two parts can be used as a meaningful aggregate measure of their relative dominance. A larger area on the left side of the glyph means dominance by the left part, and vice versa. Fig. 1 above gives such a schematic illustration. The data (being the values of the attributes) X0-X6 and Y0-Y6 lead to the respective areas A1-A6 and B1-B6. Note that X0=Y0 and X6=Y6 are the two "neutral" attributes (NET ACTIVITY and PARTICIPATION).

In the case of online auction markets, this can be interpreted as market conditions being advantageous to either buyers or sellers. In mathematical terms, the aggregate value function (Dyer 1979, Kirkwood 1980) takes the form of the sum of pairwise products of adjacent attributes. In the standard star plot of Fig. 1, the data values (X's and Y's) are computed averages of 34 selected product categories of eBay auction markets (e.g. digital cameras). The area measure in flavor of buyers is only around 10% above the buyers+sellers average, indicating a slight edge towards the buyers market. Since the difference is relatively small, this collection of 34 markets taken as a whole (being a small representative of eBay) can be considered fairly efficient otherwise.

We remark here that our methodology needs to render the glyph unit and content free by harmonizing the dimensions. Spline functions (Cline 1974), in particular simple second-order (piecewise linear) splines are the tools for treating the data values.

The critical potential benefit in the concept of using area as aggregate measure arises from the degrees of freedom allowed by the topology of the glyph, namely, the configuration of the attributes, and the angles between adjacent pairs thereof. For any given arrangement of the attributes, the standard star plot produces a glyph along symmetrically spaced radial axes. Variations from this symmetry imply a feasible set of shapes and areas, which along with permutations of the configuration, offer the choice of topologies that may suit further criteria for a meaningful aggregate measure function.

Subject to the constraints of preserving the dominance in the reference standard star glyph of dichotomies, an optimal topology (configuration of attributes and angles between adjacent pairs) is sought that maximizes the discriminating power (sum of absolute differences in left and right areas) for the reference dataset. This results in a Linear Programming (LP) formulation, with decision variables representing the angles, with respective to a specific configuration (of attributes). As the number of either side of the ("left" or "right") attributes is usually small - up to a maximum of 8-10 in most applications, total enumeration of each corresponding LP solution is feasible to compute the overall optimal, or Maximum Resolution Topology (MRT). The details of its mathematical formulation are given in the next section.

However, our concept of MRT tacitly requires a full-dimensional star glyph spanned by all the attributes of the dichotomy. This MRT-LP, by allowing unlimited trade-offs among the attributes, and with its extreme-point solutions, may at times not produce the desired results. To circumvent this difficulty, we need to extend the

MRT-LP model into its goal programming version of MRT-GP by the addition of the usual deviation variables with total smoothing variation and/or bounded deviation (Chu 2001). Its mathematical details are given in our working paper (Chu 2004).

As a numerical illustration instead, the set of eBay data of the 34 online auction markets considered above was analyzed using out MRT model. Its improvement showed some 45% increase in discriminating power over its classical (or "standard") star glyph, with a dichotomy of 29 leaning towards buyers versus 5 sellers market.

2.6 A GP formulation for MRT

With the star glyph as the topology, we now introduce its optimization model. A (k^{th}) data *record* consists of ("left" or) X-attribute data $\{x_i^k\}$ and ("right" or) Y-attribute data $\{y_j^k\}$, such that x_i^k denotes the kth record's i^{th} X-attribute value and y_j^k denotes the kth record's j^{th} Y-attribute value, with (unit interval data range)

$$\{0 \leq x_i^k \leq 1 | i \in I\} \quad \text{and} \quad \{0 \leq y_j^k \leq 1 | j \in J\},$$

where

$$I \equiv \{i = 1, 2, \cdots, m | \bar{x}_1 \geq \bar{x}_2 \geq \cdots \geq \bar{x}_{m-1}\}$$

and

$$J \equiv \{j = 1, 2, \cdots, n | \bar{y}_1 \geq \bar{y}_2 \geq \cdots \geq \bar{y}_{n-1}\}.$$

The $\bar{x}_i$ and $\bar{y}_j$ above are the average attribute-values over the records. The index sets I and J are defined as one convention to standardize a "default" configuration of the attributes. We assume the natural boundary conditions of

$$x_0^k = y_0^k \quad (\geq) \quad x_m^k = y_n^k \quad \forall k \in K.$$

These are either neutral attributes, or dummy attributes "Top" and "Bottom" assigned to the vertical axes as displayed on the star glyph. The records are indexed by $k \in K$ in the manner to be explained below. When referring to the ith X-attribute as a variable rather than its value for a particular data instance, we shall write x_i instead of x_i^k; and similarly y_j instead of y_j^k.

In the star glyph, denote the *angles* between attributes x_{i-1} and x_i by α_i, $i \in I$, and that between attributes y_{j-1} and y_j by β_j, $j \in J$, for all $|K|$ records. Let the *weights* be

$$a_i \equiv \sin \alpha_i, \ i \in I; \quad \text{and} \quad b_j \equiv \sin \beta_j, \ j \in J.$$

And we have $0 \leq a_i \leq 1$, $i \in I$ and $0 \leq b_j \leq 1$, $j \in J$.

Hence for the kth record, the sector area between attributes x_{i-1} and x_i, and that between attributes y_{j-1} and y_j, are given respectively by

$$A_i^k (= A_i^k(x)) \equiv \frac{1}{2} x_{i-1}^k x_i^k \sin \alpha_i$$
$$= \frac{x_{i-1} x_i}{2} a_i, \quad i \in I;$$

and

$$B_j^k(= B_j^k(y)) \equiv \frac{1}{2} y_{j-1}^k y_j^k \sin \beta_j$$
$$= \frac{y_{j-1} y_j}{2} b_j, \quad j \in J.$$

Next, we partition the data records according to their left-right dominance under equal weighting in a standard star plot. Using $a_i = 1/|I|$, and $b_j = 1/|J|$ in the above formulas for $A_i^k, \forall i$, and $B_j^k, \forall j$, respectively, we group the $|K|$ records into two partitions such that $K \equiv K^+ \cup K^-$, with

$$K^+ = \{k \in K | A^k \geq B^k\};$$

$$K^- = \{k \in K | B^k > A^k\},$$

where

$$A^k = \sum_{i \in I} A_i^k \quad \text{and} \quad B^k = \sum_{j \in J} B_j^k.$$

Note that A^k is the area of the left part of the dichotomy in the star glyph spanned by its X-attributes, and B^k is the area of the right part of the dichotomy spanned by its Y-attributes of record k, $k \in K$. K^+ indicates records with X-attribute dominance (or left dominance), and K^- indicates records with Y-attribute dominance (or right dominance).

With these notations, we formulate the *Maximum Resolution Topology* Linear Program (or MRT-LP for short) as follows.

$$\text{Max} \sum_{k \in K^+} [A_k - B_k] + \sum_{k \in K^-} [B_k - A_k]$$

Subject to

$$A^k - B^k \geq 0, \quad k \in K^+$$
$$B^k - A^k \geq 0, \quad k \in K^-$$
$$\sum_{i \in I} a_i = 1; \quad \sum_{j \in J} b_j = 1$$
$$a_i \geq 0, \ i \in I; \quad b_j \geq 0, \ j \in J.$$

The objective function above, with each term constrained to be non-negative, measures the gross dominance by the larger part in the dichotomy. By maximizing, we seek the highest discriminating power, in an aggregate sense, to separate the parts.

Indeed, a numerical experiment with $|K| = 16$ records, $|K^+| = 8, |K^-| = 8, m = 3, n = 4$, and random input data for x_i and y_j over [0, 1], gives the following results. The maximum objective function value (=TtlMRT) is equal to 1.859, compared to the equal-weight total area difference given by

$$\sum_{k \in K^+} [A^k - B^k] + \sum_{k \in K^-} [B^k - A^k] \ \left| \ \left\{ a_i = \frac{1}{|I|}, \forall i; b_j = \frac{1}{|J|}, \forall j \right\} = 1.278, \right.$$

or an improvement of over 45%.

However, the optimal weights (by a Lingo code) are:

$a^* = (0.796, 0.000, 0.204)$ and

$b^* = (0.000, 0.777, 0.223, 0.000)$.

The extreme point solution by linear programming collapses three out of the seven inter-attribute angles, and cannot produce a proper star glyph displaying all the attributes. To circumvent this difficulty, we extend the MRT-LP model into its goal programming version (Chu 2001) of MRT-GP by the addition of the usual deviation variables with total smoothing variation (=TtlVar) and/or bounded deviation (=TtlDev). This MRT-GP with decision variables a_i, $i \in I$ and b_j, $j \in J$ is then given by

Max	TtlMRT
Min	TtlDev
Min	TtlVar

Subject to

$$A^k - B^k = D^k, \quad k \in K^+$$

$$B^k - A^k = E^k, \quad k \in K^-$$

$$\sum_{i \in I} a_i = 1; \quad \sum_{j \in J} b_j = 1;$$

$$a_i \geq 0, \, i \in I; \quad b_j \geq 0, \, j \in J$$

$$TtlMRT = \sum_{k \in K^+} D^k + \sum_{k \in K^-} E^k$$

$$TtlDev = \frac{1}{|I|} \sum_{i \in I} [PDevX(i) + NDevX(i)] + \frac{1}{|J|} \sum_{j \in J} [PDevY(j) + NDevY(j)]$$

$$TtlVar = \frac{1}{|I|} \sum_{i \in I} [PVarX(i) + NVarX(i)] + \frac{1}{|J|} \sum_{j \in J} [PVarY(j) + NVarY(j)]$$

(Deviation bounds)

$$\frac{1}{|I|} - F * \frac{1}{|I|} \leq a_i \leq \frac{1}{|I|} + F * \left(1 - \frac{1}{|I|}\right), \quad i \in I$$

$$\frac{1}{|J|} - F * \frac{1}{|J|} \leq b_i \leq \frac{1}{|J|} + F * \left(1 - \frac{1}{|J|}\right), \quad j \in J$$

(Deviation goal constraints)

$$a_i - PDevX(i) + NDevX(i) = \frac{1}{|I|}, \quad i \in I$$

$$b_j - PDevY(j) + NDevY(j) = \frac{1}{|J|}, \quad j \in J$$

(Variation/smoothing goal constraints)

$$a_i - a_{i+1} - PVarX(i) + NVarX(i) = 0, \quad i \in I$$

$$(a_{m+1} \equiv a_0)$$

$$b_j - b_{j+1} - PVarY(j) + NVarY(j) = 0, \quad j \in J$$

$$(b_{n+1} \equiv b_0)$$

The $0 < F < 1$ parameter in the MRT-GP formulation above is the fraction of deviation from equal weights allowed for the weight variables a and b. An empirical trial, using weighted rather than pre-emptive GP with a deviation allowance of 50% ($F = 0.5$) returns 1.657 for TtlMRT, which is about a 30% improvement over 1.278. The weights are $a = (2/3, 1/6, 1/6)$ and $b = (1/8, 9/16, 3/16, 1/8)$.

This numerical instance illustrates only one of the two aspects of degrees of freedom of the topology — the "angles". The other aspect — "configuration" — can be simply treated by considering all the 144 (= 3! × 4!) cases of complete permutations of $\{i \in I\}$ and $\{j \in J\}$ from the two sets of attributes.

3   Concluding remarks

To facilitate successful applications, the MRT-GP model will eventually be coded into a Decision Support System (DSS), integrating the GP solution code as an optimizer engine with a front-end spreadsheet as well as back-end reporter user interface. We would adopt an approach similar to another previous implementation (Chu et al 2003), which involves codes in Visual Basic, DB-Grid, and other spreadsheet and graphics tools. Such a deliverable is user friendly and covers potentially a wide variety of other application areas as well. We can certainly expect innovative uses of the DSS as promised by this topological optimization modeling.

Finally, we emphasize that while our "target" application area of this topological analysis and optimization is online auction markets, our concept of such an MRT for multiple attribute data is entirely generic in nature. It is simply optimizing over the pairing of attributes as well as the weight on their products resulting in a maximum resolution topology with respect to a given dataset of pre-exist dichotomies. Future work can therefore include empirical studies and application of the approach in diverse fields, such as consumer market research, diffusion of innovation and technology, dynamics of online auction markets, political voting preference and medical diagnostics. Theoretical results and refinements are also expected from further analysis of data normalization and harmonization, choice of spline functions, the process of selecting reference subsets, and alternatives to the goal constraints of bounded variations on the angles in the optimal topology.

References

[1] Chambers J, Cleveland W, Kleiner B, Tukey P. Graphical Methods for Data Analysis, Belmont, CA: Wadsworth Press; 1983.

[2] Chu SCK. An Optimization Model for Maximum Resolution Topology. Working Paper; 2004. 5pp.

[3] Chu SCK. A goal programming model for crew duties generation, International Journal of Multi-criteria Decision Analysis 2001; 10(3): 143–151.

[4] Chu SCK, Lin CKY, Lam SS. Hospital Lift System Simulator: a Performance Evaluator-Predictor, European Journal of Operational Research 2003; 146(1): 156–80.

[5] Cline AK. Scalar- and planer-valued curve fitting using splines under tension, Communications of the Association for Computing Machinery 1974; 17: 218–220.

[6] Dyer JS, Sarin RK. Measurable multi-attribute value functions, Operations Research 1979; 27: 810–822.

[7] Ho JK. Topological Analysis of Online Auction Markets, International Journal of Electronic Markets 2004; 17(3): 202–13.

[8] Hof RD. The eBay Economy, BusinessWeek, 25 August, 2003.

[9] Hoffman P, Grinstein G. A Survey of Visualizations for High-dimensional Data Mining, in U. Fayyad et al. (eds), Information Visualization in Data Mining and Knowledge Discovery, Morgan Kaufmann; 2001. p. 47–82.

[10] Keim DA. Information visualization and visual data mining, IEEE Transactions on Visualization and Computer Graphics 2002; 7(1): 100–107.

[11] Kirkwood CW, Sarin RK. Preference conditions for multi-attribute value Functions, Operations Research 1980; 28: 225–232.

[12] Klein S. Introduction to Electronic Auctions, International Journal of Electronic Markets 1997; 7(4): 3–6.

[13] Lucking-Reiley D. Auctions on the Internet: What's Being Auctioned, and How? The Journal of Industrial Economics 2000; 48: 227–52.

[14] Perkins AB and Perkins MC. The Internet Bubble (rev. edn), New York: Harper-Business; 2001.

[15] Shneiderman B. Tree Visualization With Tree-maps: A 2-D Space-filling Approach, ACM Transactions on Graphics 1992; 11(1): 92–9.

[16] Tukey JW. Exploratory Data Analysis. Addison-Wesley, Reading, MA; 1997.

[17] Turban E. Auctions and Bidding on the Internet: An Assessment, International Journal of Electronic Markets 1997; 7(4): 7–11.

[18] Vandewalle N, Brisbois F, Tordoir X. Non-random Topology of Stock Markets, Quantitative Finance 2001; 1: 372–4.

[19] Wattenberg M. Visualizing the Stock Market, in Extended Abstracts of ACM CHI 99 Conference on Human Factors in Computing Systems Pittsburgh, Philadelphia 1999; 188–9. Demonstration of software available at `http://www.smartmoney.com/marketmap/` [accessed 16 March 2004].