Association for Information Systems
AIS Electronic Library (AISeL)

December 2004

Decision Trees for Understanding Trading Outcomes in an Information Market Game

Bartel van de Walle
Tilburg University, the Netherlands

Ajay Thomas
New Jersey Institute of Technology

Follow this and additional works at: http://aisel.aisnet.org/amcis2004

Decision Trees for Understanding Trading Outcomes in an Information Market Game

Bartel Van de Walle
Information Systems and Management Department
Tilburg University, the Netherlands
bartel@uvt.nl

Ajay Thomas
Information Systems Department
New Jersey Institute of Technology, USA
thomas_ajay77@yahoo.com

ABSTRACT

We present an experimental information market designed to aggregate IT job related information distributed among the traders in the market. The payoffs of the shares in this market were tied to true IT job demand in the real world. This paper focuses on the market outcomes of profit or loss made in this market, and more specifically the factors that may lead to either outcome. We explore the use of decision trees to predict the outcome of the market based on three different predictors: the individual traders' personal preferences, the collective ranking of the shares, and the market prices resulting from the traders' interaction during the market game. The decision tree illustrates that traders who agree with the market ranking of the best share incur losses, whereas those who don't agree have a higher chance of making a profit provided they then agree with the collective ranking of the best share or remain consistent in their personal preferences.

Keywords

Information market; decision making; trading behavior; decision trees.

INTRODUCTION

Marketplaces govern our economic decision making. They enable the exchange of information, goods, services and payments among people, and thereby create value for buyers, sellers, market intermediaries and for society at large (Bakos 1998). *Information markets* are markets designed with the specific purpose of aggregating information that is distributed among the traders in the market. Well-known instances of information markets are the so-called *Election markets*, which are information markets conceived to investigate how information about elections distributed in the market leads to security prices whose payoffs are tied to the actual election outcome. The Iowa Electronic Markets (IEM) host various local, national and international election markets since more than a decade, attracting thousands of traders (Forsythe et al., 1992). IEM election markets were found to forecast the actual outcome of elections with a higher precision than any of the traditional opinion polls (Pennock et al., 2001). More generally, *decision markets* are information markets in which the securities are tied to possible decision alternatives or policy options (Hanson 1998). Decision markets have recently been implemented for the aggregation of business-critical information within large corporations, for instance to obtain sales forecasts at Hewlett-Packard (Plott and Chen 2002). Decision market experiments confirm the predictive accuracy of election markets: forecasts obtained by small groups of traders are more accurate than forecasts obtained from traditional information aggregation methods. While many aspects of the information market and the behavior of the traders as they exchange and acquire information are topics of fascinating research, our key objective here is to investigate particular indicators that may lead to profit or loss in the market. In particular, we apply a well-known data mining technique (the C4.5 decision tree algorithm, Quinlan 1993) to discover the impact of three pre-defined factors for predicting the making of profits and losses that we have observed in the market. In the following Section, we introduce the information market game we have conducted at regular weekly sessions for seven consecutive weeks. We explain the experimental set-up, the rank order of the shares, and the resulting profits and losses in the market. Next, we develop our application of the decision tree algorithm and motivate the choice of the predictors used. The resulting decision tree is presented, followed by a discussion of the most significant observations. We conclude by summarizing the results of this particular application of a data mining technique in behavioral decision making research.

THE INFORMATION MARKET GAME

Game set-up

Thirty-five computer science and information systems students taking a course in advanced information systems at NJIT (New Jersey Institute of Technology) were invited to participate in the game and received course credit for participating. The participants were told that 7 consecutive market games with a duration of 30 minutes would be organized on a weekly basis. The task information specified that the 10 shares in the market were tied to as many specific jobs in the IT industry: Client-Server Programmer (with stock symbol CSPA), Database Analyst (DBAA), Internet Architect (INTA), Network Administrator (NETA), Object-Oriented Developer (OOGD), Quality Assurance Specialist (QUAA), System Administrator (SADU), Security Specialist (SESP), Systems Analyst (SYAN) or Web Developer (WEBD). These jobs were selected in a previous class discussion on IT industry jobs. The task set asked the participants to make a profit in the market by speculating on the demand for these ten different job positions at the end of the market game, so *after* the market had closed, in the larger New York City metropolitan area.

The market game was introduced to the participating students during an information session, which was followed by a pilot market game of 30 minutes. During the information session, the nature of the market and the purpose of the experiment was introduced. At the end of the information session, every student received 10 shares in every job, with every share at that time having an introductory and equal market value of 10 dollars. In addition, every student received an account which contained 500 dollars. Seven market sessions were held at a weekly interval in the Group Decision Support System room at NJIT's Collaborative Hypermedia Lab. Right before the start of every market session, the participants were asked to privately rank order the jobs from best to worst according to their expected demand at the end of the market. Once all rank order forms were collected, the start of the market was announced and students could start trading. During the session, a student had to find another party with whom to trade shares. As soon as an agreement had been reached, both parties (buyer and seller) were instructed to come to an 'official' market trading desk, where the transaction was verified and entered into the database. The closing of the market was announced ten minutes beforehand. No transactions were entered after market close.

The shares: group ranking and market price

The fundamental data in this study are the individual rankings and the market price of the shares. The individual rankings are obtained at the start of every market session, while the market prices are obtained at the close of the market session. In contrast to the private nature of the individual ratings, the market prices emerge from the interaction of all participants in the market and are announced to all continuously. Every market participant was asked to rank order the shares fom best (first position) to worst (10th and last position), prior to the market session. Every share in this individual rank order is then scored by assigning 10 points to the best share, 9 points to the second best, …, and 1 point to the lowest ranked share. This approach is known as the Borda method, which is a well established social choice method commonly used in committee meetings (Arrow 1951). The so-called Borda group score for every share is then obtained by summing the number of points obtained by the share. This produces a collective or group ranking for each of the shares, computed from the individual rank orders.
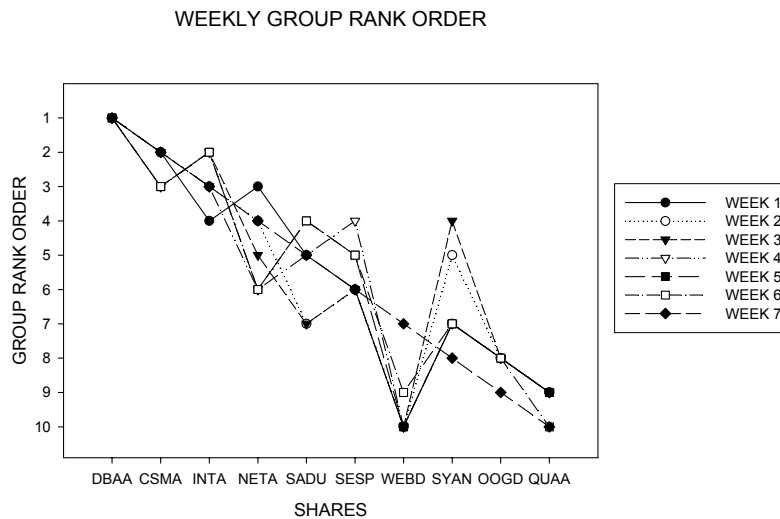
Figure 1. The weekly Group Ranking of the market shares

As soon as the forms with the individual rankings were collected, the students could start trading in the market for 30 minutes. The closing prices of the shares at every session determine the group's rank order obtained through and driven by the market activity. The weekly rank orders as derived from the group decision rule and directly from the shares' closing prices are shown in Figures 1 and 2, respectively.
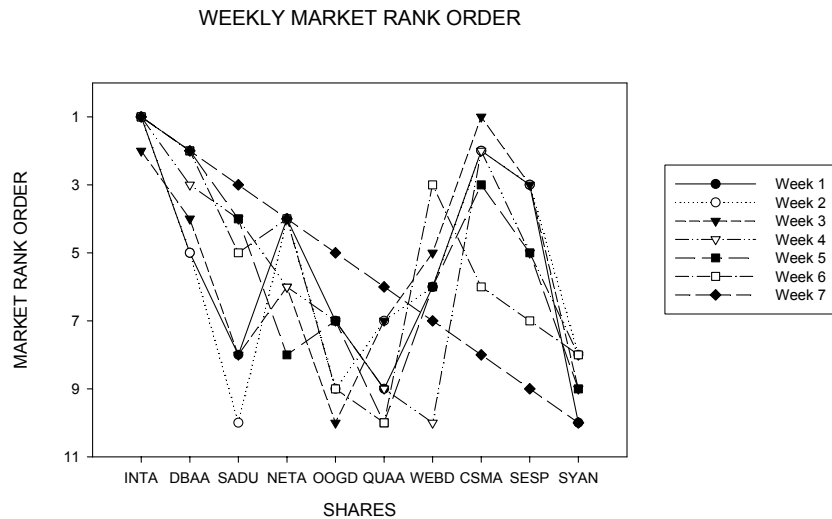
WEEKLY MARKET RANK ORDER

Figure 2. The weekly Market Ranking of the shares

Gains and losses in the market

At the start of the game, it was explained to the participants that the 'real-world' demand of the jobs at the end of the market game would be determined by looking up the number of postings for these jobs on some well-known internet job sites. Based upon this assessment, we would calculate a 'real world' market value to every job, which would be the 'true' monetary value of the corresponding share in the laboratory market. It was carefully explained that it was this 'true' value that would determine the ultimate portfolio value of every participant, and that this true value could be different from the value that is realized by trading in the market game. The shares' real world values was determined as follows. We consulted three well-known job sites to determine the demand for jobs on these sites on a random day in the week following the closing of the market for the New York city – New Jersey larger Metropolitan area. Using these data, we were able to calculate the real market share for each of the jobs for each of the job sites as the weighted sum of the market shares on every job site, with the weight of a site being the normalized ratio of the total number of jobs on that site to the total number of jobs on all sites.
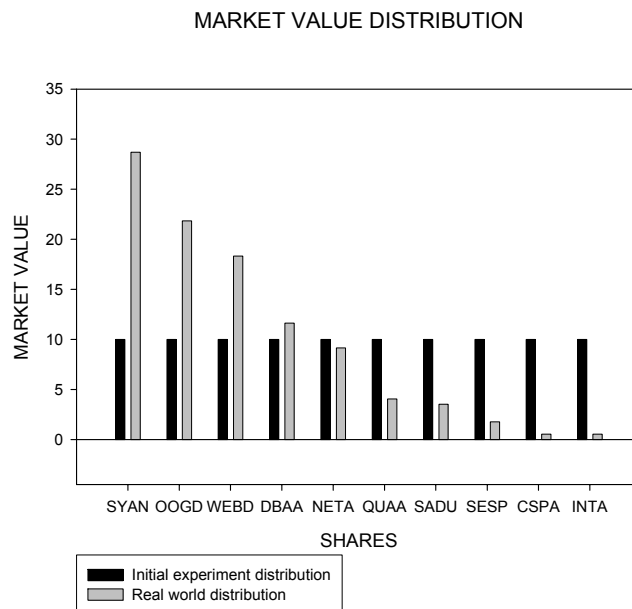
MARKET VALUE DISTRIBUTION

Figure 3. Initial and final market value of the shares

Figure 3 shows the resulting market share values compared to the original share values of 10 US dollars each. Clearly, the true value of some shares differs significantly from the share value at the introduction of the game, either in positive or negative direction.

The question now is: can we predict profit or loss in the market, based on the traders' group ranking and/or the market verdict? In the next section, we will use the well known data mining technique of decision trees to address this question.

DECISION TREES FOR EXPLAINING MARKET OUTCOMES

The decision tree is a well known data mining technique to classify the available data and discover patterns from the analysis. For our purposes, we made use of the established C4.5 decision tree algorithm developed by Ross Quinlan (Quinlan 1993). The resulting decision trees graphically represent how the various decisions made by the traders in the market influence the traders' final financial outcome, which can be a profit or a loss.

The decision tree is designed based on a particular variable known as the output variable. In our case, the output variable is the profit or loss made in the market by the traders. The variables which interact with each other and influence the output are called *predictors*. The decision tree algorithm takes into account the amount of information gain from each variable for the classification. The information gain *I* from the output is defined by

$$I(p,n) = -(p/p+n)\log_2 (p/p+n) - (n/p+n)\log_2(n/p+n),$$

where *p* and *n* stand for the different possible outcomes, i.e., profit or loss. It should be noted that in our experimental market, a trader has made profit in the market if her final portfolio value is more than \$750; otherwise, the trader has made a loss. For the sake of simplicity, the profit and loss values are not further classified according to their magnitudes.

The information gain from each of the predictor variables is then calculated. This value is called the entropy for that variable. The formula for calculating the entropy *E* for a predictor *A* is:

$$E(A) = \sum (p_i /p_i+n_i) \, I \, (p_i, n_i)$$

The difference between *I* and *E* determines the gain for a particular predictor. Similarly the gain for each predictor is calculated. The predictor with the highest gain is selected as a node in the decision tree. The process is then repeated to find the next node, and in this way the complete decision tree is computed.

Clearly, the selection of the predictor variables is crucial as these determine the accuracy and relevance of the resulting decision tree. We have selected three predictors we believe characterize the decision making process of the individual trader, and potentially influence the final outcome of the trading.

First predictor: personal preference consistency. As indicated above, all traders were asked to rank order the 10 shares (or jobs) at the beginning of each market session. If the rankings over all weeks are analyzed, we see that some traders opt for the same share as their "top job" every week. This shows a strong personal preference on the part of the trader for a particular share, which may of course influence the trader's behavior and consequently her profit or loss made in the market. The first predictor therefore analyzes the role of personal preference in determining the profit/loss made in the market. We have classified this variable as follows: if a particular share has been voted as the top share during 6 or more weeks, then he/she has a 'high' personal preference consistency. If the top vote for a particular stock is repeated at least 3 weeks, but no more than 5 weeks, the personal preference consistency is 'medium'. Else, the trader has 'low' or no preference consistency at all. Note that, here and for the following two predictors, we have chosen for only three different classification values in order not to overly complicate the resulting decision tree.

Second predictor: group ranking agreement. The individual rankings of all the traders we have obtained at the beginning of each market session can be aggregated to produce a 'group ranking' using the well known group decision rule of de Borda (Arrow 1951). The share that is collectively ranked first clearly is very popular in the group of traders, and therefore may be a 'winner' in the market. Consequently, a trader who has that favorite group share as her personal top ranked share may have an advantage in the market. The second predictor therefore analyzes how often the top group share, as computed with the Borda rule, is identical to a trader's individual top ranked share. The criteria for high, medium and low are the same as in the previous case.

Third predictor: market verdict agreement. The 'market verdict' is the best share at the end of each market session, i.e., the share that has the highest trading value in that session. This predictor analyzes how many times the individually top ranked share of each trader is identical to the best share at the end of each market session. The consistency values of high, medium and low are given based on the number of coincidences of these two factors over the seven market sessions as above.

RESULTS AND DISCUSSION

The decision tree resulting from the application of the Quinlan algorithm on the market data according to the predictor and output variables defined above is shown in Figure 4. The three predictors are noted 'Preference', 'Group Ranking' and 'Market Verdict', respectively. According to this tree, the best predictor is 'Market Verdict'. Traders who had a 'high' market verdict agreement for their top ranked share made a loss in the market; from those who had a 'medium' market verdict agreement, 83% have incurred a loss as well. Of those who had a low agreement with the market verdict, those who highly agreed with the group ranking had a chance of 2 in 3 to make a profit. If the agreement with group ranking however only was 'medium', there was a 60% chance of a loss. Those who had a low agreement with market verdict *and* with group ranking, had a more than 60% chance of a profit in case they had a high or medium strong personal preference with respect to a particular share. If they only had a low personal preference, they had a chance of 60% to make a loss.
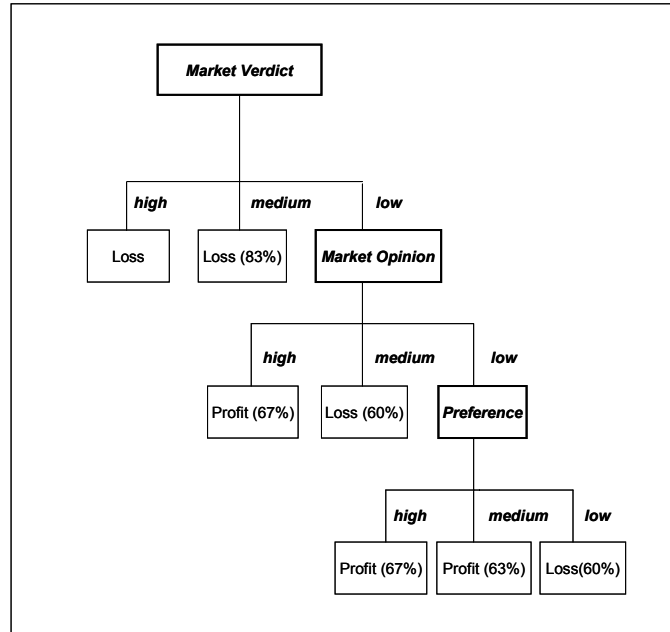
Figure 4. Decision Tree for Market at Week 7

Some interesting insights result from the decision tree analysis. First and foremost, we learn that the 'market verdict' is not a good predictor for making a profit in the market. In fact, the more a trader's top ranked share corresponds to the share that performs best in the market, the higher the chance that she is losing money in the market. Fundamentally, this indicates that the traders in the market did not identify those shares that would lead to a profit. Indeed, we can see from Figure 2 that the market's best share was mostly INTA (Internet Architect) – six weeks out of seven – or (once) CSMA (Client-Server programmer), both of which have according to Figure 3 the lowest real value. Apparently, the traders were subject to an acute case of 'group think', in which the share (INTA) that was identified during the initial trading session as the best share remained so throughout almost the entire 7 weeks of the game.

Second, we observe that the Group Ranking predictor has a different effect. For those who have a low agreement with the Market Verdict, if their agreement between their personal top rank and the collective top rank however is *high*, then there is a 2/3 chance that they will make a profit in the market (this drops to 40% if the agreement is *medium*). In addition, in case the agreement with Group Ranking is low, if the Personal Preference is *high*, then there is a chance of 67% change of profit, dropping to 63% in case of *medium* Preference and only 40% in case of *low* Preference. According to Figure 1, the share ranked first by the groups during all sessions was Database Administrator (DBAA), the real world value of which is only slightly higher than the introduction price. Apparently however, this was of enough importance to explain the profit in the market. The only remaining case where there is more chance of loss than profit is in case the trader is not consistent in his or her preferences, and most of the sessions changes to another 'most preferred' share. In this case of changing preferences, the trader has a 60% chance of incurring a financial loss in the market.

CONCLUSIONS

Analyzing the results of an information market experiment, we found that the 'decision-theoretical' collective or group ranking of shares in the market does not fully agree with the 'behavioral' ranking of the shares based on the market value. We were interested in the impact of both rankings on the making of profits or losses in the market, and used a decision tree algorithm to verify whether these rankings could to some degree predict the actual market outcome. We found that when traders moderately or strongly agreed with the choice of the best performing share in the market, they were very likely to incur a loss. When they only weakly agreed, chances were in favor of a profit when either they moderately or strongly agreed with the best share defined collectively, or when they were at least mostly consistent in their personal preference in case of weak agreement. The use of decision trees provided additional insights into our understanding of the traders' decision making. Clearly, much work remains to be done to substantiate our findings and investigate speculations such as groupthink in more detail. Post-experiment student interviews should be conducted, or the online class discussions that were held among the students in between the experimental sessions should be analyzed (Moldovan et al., 2003). In follow-up experiments, we intend to introduce trading groups in the market and study the internal private group decisions versus the decisions following from the interaction in the market. One important research question is whether this setup inhibits group think or, at the contrary, enhances it even more.

REFERENCES

1. Arrow, K.J. (1951) Social Choice and Individual Values. Wiley, New York.

2. Bakos, Y. (1998) The Emerging Role of Electronic Marketplaces, *Communications of the ACM*, 41, 8, 35-42.

3. Forsythe, R., F. Nelson, G.R. Neumann and J. Wright (1992) Anatomy of an experimental political stock market. *The American Economic Review* 82, 5, 1142 – 1161.

4. Gigerenzer, G. and R. Selten (2001) Bounded rationality – the adaptive toolbox. The MIT Press, Cambridge MA.

5. Hanson, R. (1999) Decision markets. *IEEE Intelligent Systems* 14, 3, 16-19.

6. Moldovan, M. A. Paliwal and B. Van de Walle (2003) Trading in the market: a classroom experiment in group decision dynamics. Proceedings of the International Conference on Information Technology: Research and Education ITRE2003 Newark, New Jersey (USA), August 2003, pp. 370 – 374.

7. Pennock, D., S. Lawrence, C. Giles and F. Nielsen (2001) The power of play: efficient and forecast accuracy in web market games. Science 291, 987 – 988, February 9 2001 [Letters].

8. Plott, C.R. and K-Y. Chen (2002) Information aggregation mechanisms: concept, design and implementation for a sales forecasting problem, California Institute of Technology Social Science Working Paper 1131 (March 2002).

9. Quinlan, J. R.. (1993) C4.5: Programs for machine learning. Morgan Kaufman, San Mateo, California.

10. Van de Walle, B. (2003) A relational analysis of decision makers' preferences. *International Journal of Intelligent Systems* 18, 775 – 791.

11. Van de Walle. B. and M. Moldovan (2003) An information market for multi-agent decision making: observations from a human experiment. *Lecture Notes in Artificial Intelligence* 2774: *Knowledge-based Intelligent Information and Engineering Systems*, pp. 66 – 72.

12. Van de Walle, B., S. Heitsch and P. Faratin (2001) Coping with one-to-many multi-criteria negotiations in an electronic marketplace. *Proceedings of the eNegotiations Workshop at the 17th International Database and Expert Systems Applications Workshop DEXIA'01*, pp. 747-751.