Structural Information in Probabilistic Framework: A Long Term Portfolio Performance Analysis

Chiu-Che Tseng[1]

[1]Department of Computer Science and Information Systems[1]
Texas A&M University – Commerce
P.O. Box 3011
Commerce, Texas 75429
Email: chiu-che_tseng@tamu-commerce.edu

Abstract

Based upon the raw S&P 500 portfolio performance in eight consecutive years, this paper gives an overview analysis of the application of influence diagram to portfolio management. Our findings provide a ground for the question: Is the structure information provide by the expert superior to the one generated by the data? In this study we use an artificial intelligence system called Influence Diagram for portfolio selection. The system was constructed using both expert opinion and an automated system called BN Power Constructor and we found that the expert constructed system outperform human portfolio manager and market in the years of 1995 to 2003.

Keywords: Bayesian networks, Portfolio Management.

Introduction

The investment domain, like many other domains, is a dynamically changing, stochastic and unpredictable environment. Take the stock market as an example; there are more than two thousand stocks available for a portfolio manager or individual investor to select. This posts a problem of filtering all those available stocks to find the ones that are worth investment. There are also vast amounts of information available that will affect the market to some degree.

These problems motivate us to investigate ways for equipping our system with decision support mechanisms to be applicable in complex situations. The decision support system is to provide the investor with the best decision support under time constraints. For this purpose, we propose a system that uses the Influence Diagram as the framework to create a system that will provide the user with decision recommendations.

Related Work

We explored the way to reduce the complexity of the investment decision deliberation that might cause the investor to lose money under urgent situations, and, at the same time, to provide the highest quality investment recommendations possible. .

For portfolio management, there is related work by Sycara, et al. [11] that focused on using distributed agents to manage investment portfolios. Their system deployed a group of agents with different functionality and coordinated them under case-based situations. They modeled the user, task and situation as different cases, so their system activated the distributed agents for information gathering, filtering and processing based on the given case. Their approach mainly focused on portfolio monitoring issues and has no mechanism to deal with uncertainty. Our system on the other hand reacts to the real-time market situation and gathers the relevant information as needed. Other related research on portfolio selection problems has received considerable attention in both financial and statistics literature [1, 2].

It is assumed thus far that the structure and the conditional probabilities necessary for characterizing the network were provided externally by a domain expert or an intelligent agent capable of encoding real world experience in such terms. This section deals with the problem of constructing a network automatically from statistic observations, thus bypassing the human link in the knowledge acquisition process. The learning task of the Bayesian network is separated into two subtasks: learning the numeric parameters (conditional probabilities) for a given network topology and learning the structure of the network topology. Combining the two subtasks with complete and incomplete observation data, the learning task of the Bayesian network is divided into four categories.

1. Known structure, complete data

This is the category when the structure of the network is known and the observation data is complete. Complete data means there is no missing data in the observation. The goal here is to learn the parameter by using statistical parameter estimation.

2. Known structure, incomplete data

Observed data from the real world applications are often incomplete. This scenario happens when there are missing values and hidden values in the observed data. An example of missing values is medical records for patients since not all patients undergo every possible test, therefore there are bound to be some missing values in the records. Hidden values are the ones that are never observed. One of the algorithms used for this type of learning is the gradient ascent algorithm proposed by Binder et al.

3. Unknown structure, complete data

The goal of learning the structure of the Bayesian is to find a good network that is representative of the observed data. The operations include adding, reversing and deleting edges to search over the space of the network structures. For each candidate network, fill the parameters using the algorithm described in the above categories and evaluate the network using the scoring function. Cooper and Herskovits present a Bayesian scoring metric for choosing among alternative networks. They also present a heuristic search algorithm called K2 for learning network structure when the data is observable.

4. Unknown structure, incomplete data

Given the complexity of the problem, this is the hardest learning case of all four categories and it is still under investigate by researchers. One of the algorithms proposed by Friedman is the Structural EM algorithm.

Influence Diagram

An influence diagram is a special type of Bayesian network, one that contains the decision node and the utility node to provide a decision recommendation from the model. Influence diagrams are directed acyclic graphs with three types of nodes—chance nodes, decision nodes and utility nodes. Chance nodes, usually shown as ovals, represent random variables in the environment. Decision nodes, usually shown as squares, represent the choices available to the decision-maker. Utility nodes, usually of diamond or flattened hexagon shape, represent the usefulness of the consequences of the decisions measured on a numerical utility scale. The arcs in the graph have different meanings based on their destinations. Dependency arcs are the arcs that point to utility or chance nodes representing probability or functional dependence. Informational arcs are the arcs that point to the decision nodes implying that the pointing nodes will be known to the decision-maker before the decision is made.

When using an influence diagram for decision support problems, there are some fundamental characteristics of the influence diagram that one must take into consideration. These characteristics influence the data requirements and the choice of the appropriate influence method. The first characteristic is the granularity of the values for each node. This characteristic affects the memory requirement for storing the probabilities and the computational time required for updating the probabilities. The more values within each node, the larger the memory required and the longer it will take to propagate the probability update. The second characteristic is the integration of the user's preference into the utility node. This characteristic will affect the decision outcome of the model. Given different preferences among users, the model might return a different decision recommendation. Another issue of this characteristic is how to model the user's preference into a set of values for the utility node. Different fields of research have suggested different approaches for this problem. Some suggest learning from the user's behavior, while some suggest obtaining data from a user survey and some simply query the expert and assign subjective values.

The third characteristic to consider is the availability of the knowledge about the structure, probabilistic knowledge for the prior and the conditional probabilities. There are many variables in a specific problem domain and there might exist several concepts in the problem domain that are observationally equivalent, which means they are not distinguishable even with infinite data. To find out which of those are relevant to the problem and the casual relationships among them present a challenge to the knowledge engineer. There has been much research and many tools devoted to the learning of the model structure from the data. [4] For the probability distribution for the node, there are two methods to obtain the probabilities. First, the probability distributions can be based on frequency by obtaining the data from gathered statistics. The second method is to obtain the probability distributions through knowledge acquisition sessions from the domain experts, who convey their subjective beliefs. In both cases, the probabilities can be refined through a feedback mechanism. Finally, the size, topology and connectivity of the model should also be considered. Applying good knowledge engineering techniques [7] throughout the construction of the model will help keep the network manageable.

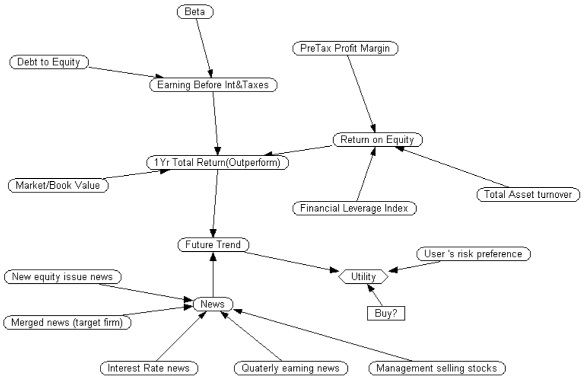

Figure 1. Influence diagram structure for stock portfolio selection.

Model Creation

The influence diagram can be created using expert opinions though knowledge engineering process or by automation algorithm using the training data. Both approaches have been implemented in different domains. Some domains such as medical domain will usually require human expert involved in the construction of the influence diagrams due to the fact that it is more trustworthy to the patients and the doctors that utilized the system. Other domains that are not as critical have been utilizing both the manual and automated processes.

We want to see how well both approaches can be applied in the portfolio selection domain. We created two model one used the expert's opinions though knowledge engineering process. And the other used a software package called BN Power Constructor. Unfortunately the BN Power Constructor was not able to find any causal-effect links for the nodes from our dataset. The main reason is due to the large variation among the values in our dataset. Even with some tweaking using expert's knowledge the system still cannot find any links to build a influence diagram.

For this phase of investigation we decided to perform the experiment without the automated solution and we will investigate further on using other automated algorithms.

Experiment Settings

Our model of the investment domain consists of a number of stocks for the investor to construct an investment portfolio. The goal is to maximize the profit from the investment portfolio. The experiments are written in C++ and built on top of the Netica Belief network package, running on a LINUX platform.

In the experiments we ran, we selected eight financial ratio data from the S&P 500 companies as the input factors to the system. The training data is collected from the Compustat database from the period of 1985 to 2003. To test the performance, we used a sliding window technique, the data from previous ten years are used to construct the influence diagram and test it on the following year. For example, the data fro m 1993 to 2002 are used to construct the model and the data from 2003 is used for testing. The system made the decision recommendation on which of the S&P 500 companies to be included in the investment portfolio for that particular year.

Results

Our testing results are quite encouraging, even though our system do not take risk and other factor into consideration, the raw performance is still quiet impressive compared to the S&P 500 index.

Here is s a table showing the performance of our system verse the S&P500 index.

Table 1. One year total return for our system verse S&P 500 index.

Year /One year total return	Our system	S&P 500 Index
1995	44.9	32.3
1996	27.2	19.8
1997	41.6	29.2
1998	40.6	15.1
1999	60.1	20.9
2000	22.0	13.8
2001	4.61	0.33
2002	-10.2	-17.4
2003	26.3	19.3

Conclusion

Our decision support system uses the influence diagram as the decision model; the structural information of the influence diagram plays an important role on the performance of our system. We obtained the structural information from the domain expert and the information represents what the expert's opinion on the causal relationships among the nodes. From the experiment results we ran nine consecutive years, we can see that the influence diagram system works better than the index.

Given the above analysis, we could conclude that by using an artificial intelligence system for portfolio selection has performance edge over the human portfolio manager and the market. In the next phase of our study we would like to research other automated structure learning algorithms to compare with our system.

References

[1] T. M. Cover. Universal portfolios. *Mathematical Finance*, vol. 1, pages 1-29, 1991.

[2] T. M. Cover and D. H. Gluss. Empirical Bayes stock market portfolios. *Advances in Applied Mathematics*, vol. 7, pages 170-181, 1986.

[3] K. C. Chang and R. Fung. Refinement and coarsening of bayesian networks. In *Proceedings of*

the Sixth Conference on Uncertainty in Artificial Intelligence, pages 475-482, 1990.

[4] N. Friedman. The Bayesian Structural EM Algorithm, In *Proceedings of the Fifteenth International Conference on Machine Learning*, 1998.

[5] E. J. Horvitz and M. Barry. Display of Information for Time-Critical Decision Making. In *Proceedings of the Eleventh conference on Uncertainty in Artificial Intelligence*, pages 296-305, 1995

[6] R. A. Howard. Influence to Relevance to Knowledge. *Influence Diagrams, Belief Nets and Decision Analysis*, pages 3-23, 1990.

[7] K. B. Laskey and S. M. Mahoney. Network fragments: Representing knowledge for constructing probabilistic models. In *Proceedings of the Conference on Uncertainty in Artificial Intelligence*, pages 334-341, 1997.

[8] R. C. Nickerson and D. W. Boyd. The use and value of models in decision analysis. *Operation Research,* vol. 28, 1980.

[9] K. L. Poh and E. Horvitz. Reasoning about the Value of Decision Model Refinement: Methods and Application. In *Proceedings of Ninth Conference on Uncertainty in Artificial Intelligence*, pages 174-182, 1993.

[10] S. J. Russell and E. H. Wefald. Principles of metareasoning. *Artificial Intelligence*, 49: pages 361-395, 1991.

[11] K. P. Sycara and K. Decker. Intelligent Agents in Portfolio Management. *Agent Technology*, Springer-Verlag, 1997.

[12] C. Tseng, P. J. Gmytrasiewicz and C. Ching. Refining Influence Diagram For Stock Portfolio Selection. In *Proceeding of the Seventh International Conference of the Society for Computational Economics,* 2001.

[13] S. R. Watson and R. V. Brown. The valuation of decision analysis. *Journal of the Royal Statistical Society, 141, Part I*, pages 69-78, 1978.