INTEGRATION OF DATA MINING WITH GAME THEORY

Yi Wang
Manufacturing Engineering center (MEC), Cardiff University, 24CF 3AA Cardiff, UK; Email: sceyw2@cf.ac.uk

Abstract: Game Theory studies strategic situations where agents select different actions to maximize their returns. Game Theory has recently drawn attention from computer scientists because of its use in artificial intelligence and cybernetics. This paper presents a frame work of integrating Data Mining with Game Theory. Due to the reason of huge amount of data, it is hard for Game Theory alone to perform the modeling analysis. Data mining assists Game Theory to deal with the large amount of data and finds hidden rules to improve game analysis.

Key words: Classification; Clustering; Association; Data Mining; Game Theory; Strategy Selection.

1. INTRODUCTION

Classical Game Theory (GT) was pioneered by von Neumann, who formalized the concepts Game Theory (Neumann and Morgenstern, 1947). It was further improved by Nash, Selten, etc (Rasmussen, 1994), So far Artificial Intelligent (AI) have been incorporated in combinatorial Game Theory (Berlekamp, Conway and Guy, 2001), but it was mostly focus on computer game and board games. Combinatorial Game Theory tries to analyze the outcomes of "perfect-information" games such as poker game and chess.

In this paper, a new approach that integrates Data Mining (DM) with Game Theory attempts to discover and extract new knowledge from the

Please use the following format when citing this chapter:
Wang, Yi, 2006, in International Federation for Information Processing (IFIP), Volume 207, Knowledge Enterprise: Intelligent Strategies In Product Design, Manufacturing, and Management, eds. K. Wang, Kovacs G., Wozny M., Fang M., (Boston: Springer), pp. 275-280.

recorded data and information. These data are normally stored in databases, and can be all sorts of nature such as name and ages of the agent or payoff of return. The learned knowledge is represented in forms of rules, such as classification rules, prediction rules, association rules or clusters of rules. These results can be often used in a support for decision making for operational improvement.

In section 2 we briefly introduce the general framework combining Data Mining with Game Theory. Section 3 describes in detail how the combination could be accomplished and what has been previously done in the different area of Data Mining. The conclusion is drawn in section 4.

2. DATA MINING AND GAME THEORY

Data Mining is a combination of analysis, search, and modeling technologies. Data Mining is often defined as the process of extracting valid, previous unknown, comprehensible information from large data bases in order to improve and optimize business decision (Fayyad, et al, 1996). Data mining techniques involves different disciplines, such as databases, enterprise information systems, statistics, Machine Learning (ML), and Artificial Intelligence (AI).

Data Mining identifies new phenomena and uncovering patterns (J. Han and M. Kamber 2000), which are useful in enhancing our understanding of target systems represented by the datasets. In this paper, the functions of Data Mining are categorized as classification, clustering and association rules.

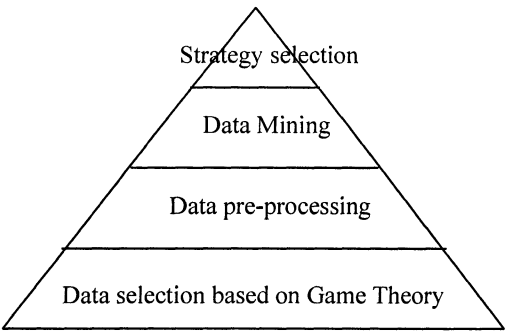

Figure 1. The process of integration of Data Mining with Game Theory.

Fig. 1 illustrates the knowledge discovery process of the integration of Data Mining with Game Theory. The process consists of 4 activities: (1). data selection based on Game Theory, (2). data pre-processing, (3). data mining and 4. Strategy selection. Knowledge discovery process can be

iterative and each of these activities may be revisited multiple times in order to improve the knowledge discovery process. A good understanding of the whole process is important for any successful application. No matter how powerful the DM algorithm is, the resulting model will not be valid if the data based on GT are not selected and pre-processed correctly.

Game Theory (Rasmussen, 1994) is concerned with the formal analysis of situations called *"games"* where agents can choose different strategies that determine their actions under particular conditions. Conditions and outcomes unfold through the interactions of the agents' strategies.

There are many rules and patterns extracted by Data Mining which mirror the interaction of conflict between the agents, such as strategy of struggle in an environment of limited resource. Modelling these problems using the concepts of Game Theory improves the understanding of these rules. Classical Game Theory can offer strong mathematical proof techniques to the problem-solving approaches with high computational complexity. Adopting such strong proof techniques from Game Theory will lead to further advances in Data Mining.

In Game Theory due to the uncertainty with the opponent's action, to be on the safe side, thus leads to in a strategy selection matrix of a prisoner's dilemma, an agent will choose to confess. (Rasmussen, 1994) If both agents are rational, the solution to the *Prisoner's Dilemma* is that both should confess. In real life strategy selection is based on a agent's knowledge of the opponent. If there is perfect knowledge, then the agent has control of the *Game*. (Pham, Wang, and Dimov, 2004) This knowledge is gathered from past records by applying Data Mining methods.

For an agent following steps can be followed:

1. Observation. Data is observed and gathered. Only the relevant data will be taken into account.
2. Analysis. The relevant data is put into the Data Mining algorithm to get interesting rules, which is used for the game play.
3. Strategy selection. The rules provide a better understanding of the game environment, and the agent could make a better selection of the strategies available.
4. Testing. When the selected strategies are put into the real world, it will cause expected and unexpected feedback.
5. Return to step 1 and update the observation and continue the next steps.

3. MINING DESCRIPTION

3.1 Classification

Classification is the process of sub-dividing a dataset with regard to a number of specific outcomes. For example, we might want to classify our agents into 'winner' or 'loser' categories with regard to their payoff. The category or 'class' into which each agent is placed is the 'outcome' of classification process. A classification model is said to be 'trained' on historical data, for which the outcome is known for each record. It is then applied to a new, unclassified dataset in order to predict the outcome for each record.

Major researches done here is based on game mining, (Tveit and Tveit, 2002) Data Mining applied multi-agent games networking (Smed, et al., 2001) and game usage (Srivastava, et al., 2000). The main motivation for performing classification in computer games is to improve the game, so that the agents become more satisfied and stay longer, and to increase the revenue of the game service again. Game mining is defined as three main types: (1). game content mining; (2). game structure mining; and (3). game usage mining. It was inspired by well-known types of web mining, which are corresponding to: (1). web content mining; (2). web structure mining; and (3). web usage mining. (Cooley, et al. 1997).

The game mining approach (Tveit and Tveit, 2002) has been extended further here to find a relationship between different factors, such as previous selected strategy or the nature of agents. The nature includes age, income, etc. Thus datasets are divided into classes. It is important to find the proper training and target classes because real world classification problems usually involve many different types of classes.

3.2 Clustering

Clustering is usually achieved using statistical methods, e.g. a k-means algorithm. Each record is compared with a set of existing clusters. A record is assigned to the cluster it is nearest to, and this in turn changes the value that defines that cluster. Cluster analysis is the process of identifying the relationships on the basis of their targets similarity and dissimilarity. Unlike classification, clustering does not need pre-defined target variables.

Clusters are focused around a "centre" or "centres" which is initially defined and some measure must be used to estimates the similarity and their dissimilarity. The similarity and dissimilarity can be measured as the distance from each other and from the cluster centres.

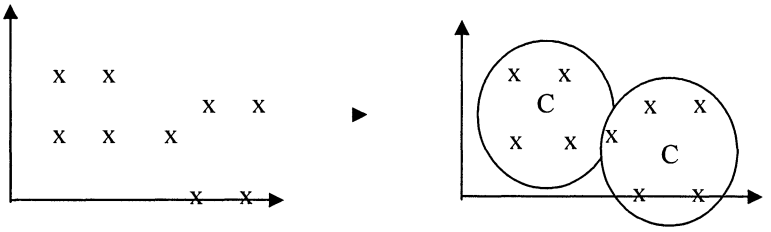

Figure 2. Clustering of data.

Huck, et al., (2000) mentioned that three clusters were identified with two equilibrium locations and one focal point. The observations are related to best-response in game dynamics, and to the fact that the agents rely on best-responses in particular when they are close to the equilibrium configuration.

Pavan and Pelillo (2003) used clustering for evolutionary game dynamics. The new framework for pair wise hierarchical clustering centred on the notion of a dominant set, a newly introduced graph-theoretic concept that has proven to be relevant in partitioned clustering and image segmentation problems. The approach is general and can be applied to solve a variety of problems in computer vision and pattern recognition. Its potential for the problem of image database organization has been demonstrated.

3.3 Association

Association Mining identifies links between attributes in a dataset. The task of association rule mining is to find certain association relationships among a set of items in a database.

A new way of extracting causal rules and making optimal selections has been proposed (Pham, et al. 2005). The method hinges on regarding a manufacturing strategy selection problem as a "*Game*" involving several agents. From a certain time point in the present to a certain time point in the future, any actions of any agents in the "*Game*" at any moment may influence on the actions of others. Each agent has a pattern of strategy selection, which is observable by other agents, and to which other agents have a counter pattern. A change in strategy pattern by an agent might cause other agents to change their selection pattern.

4. CONCLUSION

Game Theory studies strategic situations where agents select different actions to maximize their returns. Data Mining is the process of extracting valid, previous unknown, comprehensible knowledge from large data bases

in order to improve and optimize business decision. In this paper, we have established a framework to integrate Data Mining with Game Theory in order to benefit from the two techniques. Some of the researches in this area are reviewed. We also classified Data Mining into three sub categories and showed how these two subjects can be successfully integrated together in detail.

5. REFERENCES

1. Berlekamp, E. R., Conway, J. H., and Guy R. K., (2001), *Winning Ways for Your Mathematical Plays*, A. K. Peters Ltd.
2. Cooley, R., Mobasher, B., and Srivastava, J., (1997), Web mining: Information and pattern discovery on the world wide web. *Proceedings of the 9th IEEE International Conference on Tools with Artificial Intelligence (ICTAI'97)*. IEEE, November.
3. Pham, D.T., Wang, Y., and Dimov, S., (2005), Intelligent manufacturng strategy selection, *International Conference on Intelligent Production Machines and Systems 2005*, Cardiff. pp 363-369 july
4. Fayyad, U.,Piatetsky-Shapiro, G., Smyth, P., (1996), From data mining to knowledge discovery: An overview, in: *Advantaces Knowledge Discovery and Data Mining,* Fayyad, U. Piatetsky-Shapiro, G., Smyth, P. and Uthurusamy, Eds., Cambridge, MA: MIT Press, pp. 1-36.
5. Han, J. and Kamber, M., (2000), *Data Mining: Concepts and Techniques*, Morgan Kaufmann Publishers.
6. Huck, S., Müller, W. and Vriend, N., (2000), *The East End, the West End, and Kings Cross: On Clustering in the Four-Agent Hotelling Game,* Humboldt Universitaet, Berlin.
7. Neumann, V. J. , and Morgensten, O., (1947), *The Theory of Games and Economic Behavior*, Princeton, NJ: Princeton University Press.
8. Pavan, M. and Pelillo, M., (2003), Dominant sets and hierarchical clustering, *ICCV03*, pp. 362-369, WWW Version.
9. Rasmussen, E., (1989), *Games and Information: An Introduction to Game Theory*, Oxford: Basil Blackwell.
10. Smed, J., Kaukoranta, T. and Hakonen, H., (2001), Aspects of networking in multiagent computer games. *Proceedings of International Conference on Application and Development of Computer Games in the 21st Century*, pp. 74-81, November.
11. Srivastava, J., Cooley, R., Deshpande, M. and Tan, P. N., (2000), Web usage mining: Discovery and applications of usage patterns from web data. *SIGKDD Explorations*, 1(1):12-23, January.
12. Tveit, A. and Tveit, G. B., (2001) Game Usage Mining: Information Gathering for Knowledge Discovery Massive Multiagent Games. *Proceedings of International Conference on Application and Development of Computer Games in the 21st Century.*