2M3-2

Predicting Future Investment based on Agent-based Simulation

Mhd Irvan[*1] Takao Terano[*2]
[*1]Tokyo Institute of Technology [*2] Tokyo Institute of Technology
irvan@trn.dis.titech.ac.jp terano@dis.titech.ac.jp

Abstract: In recent years, many enterprises prefer to invest in new or smaller companies that are creating new service for emerging market rather than building the service by themselves. This often lead to less risks and win-win situations for both parties. Predicting which companies are getting the attention by the enterprises in the near future or which enterprise is most likely going to invest in those companies can give valuable insights to how future market would look like. This paper propose a method to accomplish such goals by running agent-based simulation, trained with machine learning techniques through artificially generated big data.

1. Introduction

Recently, many new venture and IT startup companies are coming out to offer new web services to mainstream users. They are typically building services that have yet to be offered by bigger and more established enterprises, giving birth to new emerging market and bringing uncovered potentials. A common examples of these services include, but not limited to, social networks, social gaming platforms, or community sharing services.

Large IT enterprises often hesitate to build these kind of services due to the high risk and untested market [Hochberg 2007]. Instead, they feel more comfortable to put investment in the new startup companies to either kickstart or accelerate the development process of the new web services. This often lead to a less risky and win-win situation for both sides. The enterprises get to see where the potentials grow to or how the market develop, and the IT startups is able to bring their new ideas into reality.

However, it is hard to predict which enterprise are likely to invest in which startup companies [Fried 1994]. Often investments come from, or be invested to, unexpected companies. Although some of these investments may seem to be a surprise, there are patterns hidden in those investment directions.

Predicting which new small startup companies are getting more attentions, or which enterprise is likely to invest in those companies, can give valuable insights to how the future market will shape into [Fried 1994].

This short paper introduce an on-going research proposing a method to accomplish such goal. The method takes advantage of an multi-agent-based simulation and implement genetic algorithms to search for patterns within the investment data.

2. Proposed Method

Multi-agent simulation is a method implementing multiple intelligent agents that interact autonomously with each other. The way they interact represents actual social world. In this research the agents represent both large enterprises and new startup companies. They interact with each other in the form of giving or receiving investments (Figure 1)

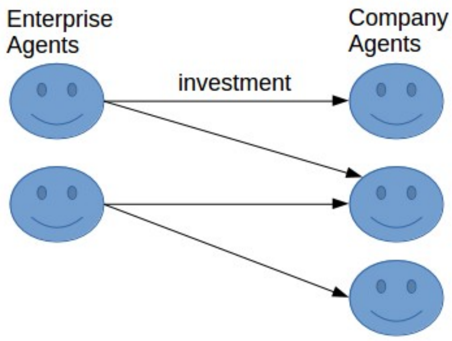

Figure 1: Interaction between agents

The agents try to predict what makes an enterprise (either its own or rivals) invest in a startup company. As the bases of reasoning, the agents consider the properties of both sides and find the patterns in the investment moves. For this short paper, as an initial approach, we consider that each of the enterprises simply has the following properties:

1. Budget (possible values: small, medium, or big)
2. Type of Company (General IT, Database, or Networking)
3. Market Reach (Local, Multi-national, Global)

Furthermore, each of the smaller startup companies simply has the following properties:

1. Team size (small or medium)
2. Type of Company (Social Web, e-Commerce, or Game)
3. Number of customers (small, medium, or big)

The agents built prediction classifiers [Wilson 1995] based these properties. Each of these classifiers are essentially a condition-action classifiers [Irvan 2013]. For example, a classifier may be read as “If a global big budget networking company invested in a medium-sized game startup company with big number of customers, it is most likely to invest in startup company B”.

The goal of the simulation is to let the agents accumulate accurate classifiers throughout the learning process. A population

Contact: Mhd Irvan, Tokyo Institute of Technology, irvan@trn.dis.titech.ac.jp

of accurate classifiers would mean that the agent could predict any hidden or future transactions, which plays crucial role in the market.

During the simulation, agents are fed with 80% of the complete investment data set to learn and to search for patterns. Afterwards, they try to predict other investment happening within the remaining 20% of the data set. To further see whether any accuracy is not a coincidence, several other simulations are conducted but with different 80% parts of the complete data set, forming a 5-fold cross validation (Figure 2)

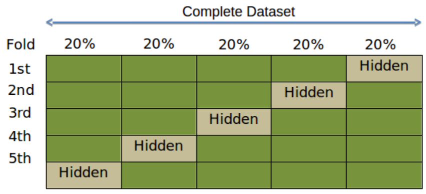

Figure 2: Five-Fold Cross Training and Validation

Genetic algorithms [Goldberg 1988] are used to search for patterns within the investment data and reinforcement learning [Kaelbling 1996] methods are used to give feedback to agents informing them the quality of the decision they are making. The agents start sensing the environment (seeing available data as inputs), then generate classifiers with genetic algorithms and prediction [Irvan 2013], and send prediction output back into the environment, receive reinforcement learning reward back from the environment (Figure 3) .

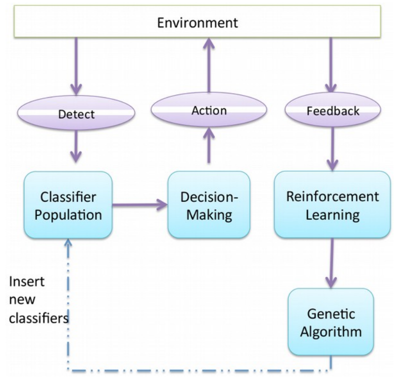

Figure 3: Architecture of the Proposed Method

3. Early Experiment

To test the feasibility of the early-phase of our proposed method, we initially randomly generated an artificial data set for the simulation to run. The data set consists of 100 enterprises and 500 startup companies. Each enterprise is set to be investing from between 5 to 10 startup companies. The simulation is run and validated by a five-fold validation, explained at the previous section.

4. Result

Table 1 summarizes the accuracy of the predictions the agents made during the simulation, in terms of precision and recall considering the relativity between true positive (tp), false positive (fp), and false negative (fn):

1. Precision

$$\frac{tp}{tp + fp}$$

2. Recall

$$\frac{tp}{tp + fn}$$

Table 1: Precision and Recall Values of Each Fold

	Precision	Recall
Fold-1	71.4%	73.6%
Fold-2	70.2%	72.1%
Fold-3	68.9%	70.3%
Fold-4	70.6%	72.7%
Fold-5	71.1%	73.2%

Our early phase of experiment shows that at the end of the simulation, the agents were able to generate mostly accurate classifiers within their knowledge population. These classifiers were able to find patterns within the data, regardless to which part of the data was hidden from their view.

5. Conclusion and Future Work

The results from the experiment suggest that our proposed method can work as a feasible solution for this kind of problem. However, since in reality, companies have much more different properties than the ones considered for this experiment, further research will implement much wider and improved classifiers by including those different properties. We are also planning to test our next improved proposed method with real data.

References

[Fried 1994] Fried, V.H., and Hisrich, R. D., Toward a Model of Venture Capital Investment Decision Making, *Financial Management*, Vol. 23, No. 3, Venture Capital Special Issue, pp. 28-37, Wiley, 1994

[Goldberg 1988] Goldberg, D.E., and Holland, J.H., Genetic Algorithms and Machine Learning, *Machine Learning*, Vol. 3, Issue 2-3, pp. 95-99, 1988

[Hochberg 2007] Hochberg, Y.V., Ljungqvist, A., and Lu, Y., Whom You Know Matters: Venture Capital Networks and Investment Performance, *The Journal of Finance*, Vol. 62, Issue 1, pp. 251-301, Wiley, 2007

[Irvan 2013] Irvan, M., Yamada, T., and Terano, T., Influence of Organizational Learning for Multi-Agent Simulation based on an Adaptive Classifier System, *IEEJ Transactions on Electronics, Information and Systems*, Vol. 133, No. 9, pp. 1752-1761, 2013

[Kaelbling 1996] Kaelbling, L.P., et al., Reinforcement Learning: A Survey", *Journal of Artificial Intelligence Research*, Vol. 4, pp.237-285, 1996

[Wilson 1995] Wilson, S. W., Classifier Fitness Based on Accuracy, *Evolutionary Computation*, Vol 3. No. 2, pp. 149-175, 1995