Co-evolution of trading and culture

Jean-Marc Montanier

November 25, 2014

1 Question

This research is part of the EPNet project, and aims at the modelisation of trade and culture within the roman empire thanks to agent based methods. On the one side, the trade of products is dependent of their adoption within the culture at hand. On the other side, the traits of a culture are dependent of the communication channels (mainly used for trade) in order to spread. The interplay between culture and trade will be investigated. The main goal is to identify the elements necessary to reproduce the patterns observed in the roman network.

Within this general vision, the introduction of a new good in the trade network will be studied in depth. This modification adds new trading opportunities that the agents may seize in order to maximise their wealth. However, when the product is introduced, the culture of the agents is already in favour of others well known products. It is therefore not sure that the agents will adopt the new product. Two main mechanisms relative to the adoption of a new good will be investigated. First, some agents prefer to conform to the major culture, and others prefer to adopt different cultural train (anti-conformism). Second, the leader of a province may have a key influence in the adoption of the new product as a mark of social status.

Previous works on the evolution of culture has studied the social mechanisms that influence the composition of culture. Conformism and anti-conformism are among these mechanisms. On top of the composition of the cultures, the speed at which it varies is also studied. Within the simulations conducted the agents are strongly abstracted and do not have any economical constraint. As presented earlier the economical aspects are expected to influence the evolution of culture.

Earlier works on economical networks were based on rational agents. The goal of these works is to reach an optimal efficiency at the scale of the system. However humans are not rational agents. More realistic modelisations are based on complex agents termed “cognitive agents”. However, these approaches rely on numerous experiments in order to calibrate the parameters of the model. Moreover, due to the computational complexity of the models they do not scale well to a high number of agents.

In order to study the effect of new goods within an established network we will need a model where agents can evolve culture and trade in the same time. This model has to be able to scale-up to high number of agents in order to simulate complex dynamics. As we can not perform calibration experiments, the number of parameters of the model will have to be low. Finally the environment has to be set up so that the agents which are able to learn new behaviours are more successful than the agents with static behaviours. It can be noted that these needs have been highlighted in recent articles but not fully addressed until now.

2 Environment

In previous works two environments have been proposed which are close to address the requirements of this project.

2.1 Model of wealth inequality

The model proposed in [1] studies the wealth inequalities in a dynamic network. The agents are in a virtual environment and create a dynamic network where they can produce, consume and trade two products. Agents can negotiate the rate at which the two products will be traded. The agents have a will (hard-coded in their behaviour) to keep their wealth and number of agents within their trading network above a threshold.

This model has two major constraints which makes it unfit for the study presented here. First, the products serve only for trade purposes and are not used as cultural factors, i.e. some goods can not be accumulated so that the agent show them to others. Second, only two goods are considered and the system can not scale up to an undetermined number of products.

2.2 Model of agricultural society

The model proposed in [5] studies the settlement of primitive agricultural societies. In this model agents produce, trade and consume 2 essentials goods and 2 non-essentials goods. In this model the agents do not trade thanks to a network but rather all in the same time in a market. If an agent has enough resources it can create another agent. On the contrary, if an agent has too few of the essential resources it dies.

This model is interesting to us since certain cultures can evolve through non-essential goods. However the focus of the article was given to the presentation of the model. Therefore the outcome of the simulations have been scarcely studied and no studies have been done on the evolution of culture. This framework is constrained as it does not allow the evolution of the trading network.

2.3 Culture and trading network

The model presented in this work proposes to integrate the most interesting features from [1] and [5]. Moreover the possibility to propose new goods to trade will be introduced.

Agents are operating within a virtual world without embodiment. The agents are placed in a trading network. This network is used by the agents to trade goods of two types: essential and non-essential. For each type, new goods can be introduced, and old ones removed from the trading network. Agents can produce, trade all goods. If enough of the essential goods are found in an agent a new agent is created. If too few of the essential goods are in an agent, the agent die.

The agents can establish links with other agents autonomously. These link are uni-directional, i.e. agent a sends good to agent b but agent b do not send good to agent a. Naturally bi-directional links composed of two uni-directional one can be built. In order to build a connection, an agent first propose it to a second agent. This second agent has then the possibility to accept the link or refuse it. A third agent can also propose a connection from one agent to another, thus acting in quality of leader. Once established the link can be removed by each agent involved unilaterally.

The trade is based on a currency and not on exchanges of good, thus simplifying the introduction of new goods. Each pair of agent involved in a trade can negotiate the price of a good. Agents are aware of the amount of goods that the other agents of their network possess.

Since part of the goods are non-essential, the evolution of culture based on these products can be expected. The interplay between necessary and unnecessary goods will be closely analysed. Since the agents can communicate to each other about the goods they possess, conformism or anti-conformism strategies can be expected to be learnt.

With this setup the dynamics to ensure the survival and welfare of an agent can be potentially complex including learning of preferred good, proposing and accepting the best commercial connections. It is expected that simple agents will do poorly to ensure their survival in this environment. However cognitive agents can potentially show very rich behaviours.

3 Agents

3.1 Previous models

In the model from [1] the agents have a rather fix pre-designed behaviour. This behaviour is controlled by a set of 13 parameters. The trading strategy is learned through random modifications when an agent is not satisfied, i.e. when it has too few of the necessary resources.

In [5], the agents are planning ahead in order to maximise the profit. The details of the planning method used along with its parameters is not documented. Agents are learning an estimation of the reward they will receive, and an estimation of the result of working effort. The learning mechanism is a moving average scheme.

In previous approaches the learning and planning algorithms are specific to the environment used and therefore weakly documented. This is detrimental to the field as it is difficult to understand if the results depend of the design of the environment or the particularities of the learning algorithm. The learning approach presented in this paper is therefore relying the consequent work established in the field of Artificial Intelligence. These approaches will have also the advantage to rely on few and well documented parameters.

3.2 Q-learning models

Withing AI, Reinforcement learning methods are targeting the learning of behaviours for agent based problems. These methods are of interest to this project as they are well documented. Moreover multiple documented implementations exist for the most used approaches.

One of these approaches is the Q-learning algorithm. The main idea is is to formalize the environment as sets of states, actions and rewards. At each time step the agent is presented a state from the environment and chooses an action. Consequently, a reward is given by the environment. The value of the reward depends of the action chosen. The algorithm is learning a policy which will return which action to perform in which state. The aim is to learn the policy leading to the highest cumulated reward.

However this approach is not adapted to continuous states and actions. Therefore the states and actions have to be discretized. This is challenging as new goods will constitute new dimensions in the state space of the agent and will therefore modify the best discretization.

Once discretized the size of states and actions spaces will be too large to be efficiently manipulated by the learning algorithm. This problem is even more important due to the high number of agents involved in the simulation.

The estimation of the reward to give to an agent can also be difficult. An important element in reinforcement learning is to have the agent estimate which reward it would obtain from an action. If there is a high number of possible actions it is difficult to store the expected reward for each of them.

The remaining of this section will explore the possible solutions to these problems.

3.2.1 State discretization

In order to discretize the state space efficiently, an approach developed for robotic problems in [6] can be used. The core idea of this approach is to discretize the state space during the learning of the policy. If one state leads to too many uncertainties (it is difficult to know which reward is associated to it), it will be split in two different states.

The approach has been used for visual data with a fix number of features to classify them. The proposition is to transform this approach to be used by various kind of sensors. Features suitable to describe different type of data will have to be looked for. Another problem to address is the possibility to add dynamically a new dimension in the state space.

3.2.2 Policy learning

In order to overcome the problems relative to planning when a high number of states and actions are present, it is proposed to perform a policy search. Within the policy search approaches, a policy is modelized thanks to few parameters. The search is then performed in the space of the parameters of the policy.

A policy can be implemented in multiple ways. In recent literature it has been proposed to rely on Guaussian based approaches [2, 3, 4].

3.2.3 Reward estimation

Finally, to estimate rewards, neural network can be used as general predictors as proposed in [7]. Within this approach a neural network can be trained to estimate the reward associate to the known pairs of states and actions. Once trained it can be used to generalize the reward for unknown pairs of states and actions.

4 Roadmap

Due to the complexity of the implementation it will be realised step by step. Each step has to reply to a specific objective. This roadmap is meant as temporary. It has to be modified as much as needed during the course of implementation depending on the results obtained and difficulties found.

Step 1: Environment In this step the main interest is the study of simple agents within a rich environment. The main work is to implement fully the environment. Within this environment simple agents are implemented in order to understand the first basic effects: imitation of the common behaviour, preference to the new object, imitation of a leader. These agents will not be endowed with learning abilities. Nevertheless they will be helpful to answer the following questions: can leaders promote the use of new products ?

Step 2: Fixed discretization, classic Q-learning This step constitutes the first attempt to implement a learning agent within the environment. As a consequence the algorithm is kept as standard as possible. The idea is to observe the learning dynamics and verify if the hypothesized problems (problem of discretization, size of search space) are effectively present. The following question will be investigated: can the agents learn a culture not related to survival or trade?

Step 3: change to Smart discretization The introduction of dynamic discretization should allow a faster performance of the algorithm. On top of this technical aspect it will also allow the study of the addition of new dimensions in the state space, which means new goods. From this point it will be therefore possible to study the evolution of culture based on new goods with learning agents. In which condition does the agents learn cultures based on recently introduced goods ? We expect that the imitation of leaders and the anti-conformism behaviours are at the core of the necessary conditions.

Step 4: Switch to policy learning This constitute a technical step aiming to improve the speed of simulation. It is currently placed after the step 2 and 3 but may be fully integrated within the step 2 if the speed of simulation is too slow to obtain results.

Step 5: add policy estimation The addition of policy estimation will increase the estimation window of the agents. By this mean we expect to obtain agents which are realising better estimation of their actions and are therefore selecting with a higher accuracy the best action. On top of verifying is this prediction is true in our setup we would like to investigate the following question: does the planning of actions favour the evolution of culture ?

References

[1] R. Alexander Bentley, Mark W. Lake, and Stephen J. Shennan. Specialisation and wealth inequality in a model of a clustered economic network. 32(9):1346–1356.

[2] Marc Deisenroth and Carl E. Rasmussen. PILCO: A model-based and data-efficient approach to policy search. In *Proceedings of the 28th International Conference on Machine Learning (ICML-11)*, pages 465–472.

[3] Marc Peter Deisenroth, Gerhard Neumann, and Jan Peters. A survey on policy search for robotics. 2(1):1–142.

[4] Sergey Levine and Pieter Abbeel. Learning neural network policies with guided policy search under unknown dynamics. In *Advances in Neural Information Processing Systems*, pages 1071–1079.

[5] W. Macmillan and H.Q. Huang. An agent-based simulation model of a primitive agricultural society. 39(2):643–658.

[6] Justus Piater, Sébastien Jodogne, Renaud Detry, Dirk Kraft, Norbert Krüger, Oliver Kroemer, and Jan Peters. Learning visual representations for perception-action systems. 30(3):294–307.

[7] Martin Riedmiller, Thomas Gabel, Roland Hafner, and Sascha Lange. Reinforcement learning for robot soccer. 27(1):55–73.