Social norms

Daniel Tang

May 13, 2024

We define a social norm as a constraint on behaviour that is shared among a group of agents. These are learned from experience, from observation of others and from culture (stories, art, media).

Social norms help make living within a society computationally tractable by reducing the number of options for action open to us at any moment (or at least making a small number salient) and by helping us predict other's behaviour by assuming they will conform to some set of norms.

We assume agents live within a Poisson ABM where each agent has a public and private state. Each event is either a unitary action or a binary interaction. Every agent has a fixed set of unitary actions (e.g. move left, eat, etc) and a set of "initiate dialogue with agent with property P" interactions based on the properties of the public state of the other agent. In a given state, the physical environment defines a mask dictating which of these actions are currently possible. An interaction event is a dialogue (turns-based encounter) with another agent. An agent's decision space is the Poisson rate it assigns to each physically possible action/interaction, and the next action it takes during a dialogue.

1 Zero intelligence societies with social norms

Given a set of social norms, we can define the zero intelligence society as one where the majority of agents follow social norms, but act randomly to the extent of freedom within the norms. To this we can add a small percentage of deviant agents that break social norms to a greater or lesser degree, acting randomly when deviant. Note that the zero inelligence agent requires no reward function.

If we show that a zero intelligence society robustly exhibits a social observable, then we can have some confidence that this observable is a stable consequence of the norms. However, we can't show that the set of norms are themselves stable or discoverable.

2 Selective copying

If agents are able to observe other's behaviour (both unitary actions and binary interactions) and have a propensity to copy it, we can define the social norm as the distribution of behaviour given the context summed over the (perhaps time-weighted) history of all historical events. The context is defined as the smallest set such that any addition to the context makes no statistically significant difference to the distribution. In this way, then the social norm can be thought of as a probabilistic policy.

If agents also learn a Q-function then they can choose to deviate from the norm, or at least add a bias to the norm, based on the expected quality of each action (temptation). In practice, for a given context, an agent will likely have a small number of samples from the norm, so can decide which (if any) of these to copy given the Q value.

2.1 Emergence of social norms from copying

Suppose there is a stochastic function $F(Z)$ which takes a PDF, Z, over the space of behaviours and returns a new PDF so that we have a discrete-time dynamic system on the sapce of PDFs over the behaviour space. Here, Z represents the social norms or "zeitgeist" and $F(Z)$ represents the generation of an observed behaviour. [in reality each agent has their own set of samples of (real and imagined) behaviours. Z is an approximation of P(behaviour of other—all context)]

If the mind has a bias towards copying more probable behaviours (i.e. probability of copying is greater than the probability of observing), then the dynamics of such copying is to tend toward a delta functions in the space of behaviours. In this way roles may emerge from minds that have a propensity to copy.

[Do norms only apply to intra-dialogue behaviour, or are there norms on choosing which dialogues to enter into? If I am married and go to work, it would seem to be a social norm that I return home after work. So, norms (via roles) extend beyond dialogues and even dictate unitary action.]

3 Social roles

At any point in time, an agent's context and history dictate which social norms are active (although it seems likely that disagreement between agents may exist on which norms are active). We call a commonly recurring set of norms a role. An agent must be aware of the roles played by other agents it interacts with (in order to calculate expectations of others' behaviour) and the role others believe it is playing (in order to calculate other's expectations of its own behaviour). [need more detail on how roles are determined]

Agents may be quite fluid in their roles, wearing different "hats" for different situations, and can simultaneously be involved in playing distinct roles with different agents.

Agents have access to a shared set of norms and roles through personal experience, observation of others and culture (i.e. stories, theatre, TV...). New roles should emerge in an evolutionary manner, through innovation and adoption. Given a vocabulary of norms and roles, a society can reach an equilibrium w.r.t.

this vocabulary if no perturbed norm or role can become adopted. So, norms and roles can be thought of as the genes and DNA of evolution, they compete for survival against other norms and roles in the vocabulary within the environment of the social interactions of the agents. Our interest is the dynamics of the evolution of norm vocabularies.

Given a set of roles applicable to a given situation, the choice of role becomes a meta-game.

What if we add survival of the agent? An agent's reward function is arbitrary in this view (defines the fitness of the role), but while an agent aims to maximise its reward, the reward function defines an agent's fitness to survive...[does this fall within our interest here?]

4 Social rewards

In addition to physical reward (eating, sex), agents are intrinsically rewarded by social acceptance and scalded by social exclusion. Agents can also generate internal reward via self esteem.

5 Representation of norms

Each norm is associated with a model of the world. The model defines the actions available to the agent as an abstraction of its physical actions, and defines the roles played by the agents involved in the model. Among the vocabulary of norms there may be multiple models and at any one time an agent may be playing roles in multiple incomensurable models.

6 Social status

Is social status something more than the playing of a role? Being a king requires not only that I play the role of a king, but also that a number of people play the role of subject when interacting with me. In this way I cannot unilaterally decide to be king. A king remains king as long as the ruler-subject relationships endure. Once a stable set of What is the dynamics of the emergence of these social relationships and how do they remain stable and how do they crumble?

This requires agents to be able to identify each-other, in order to learn and identify the social status of other during an interaction [it also implies that during an interaction, I'm not only interacting with a single agent, but rather a node in a whole social network].

7 Social network

This would come naturally out of a social network topology: an edge is a social tie that joins two agents in a well defined pair of roles. Social status comes in an agent's relationship to other agents through social ties. In this picture, an agent can then decide to make or break social ties, or enter into episodes with existing socially tied agents. Social tie types become the units of evolution, but these can be embeded within wider social networks which may themselves be considered units of replication (which may outlive the constituent agents)...[is this part of the explanationn for the success of capitalism?]...this could lead to interesting dynamics. A social structure that can expel agents that don't fulfil their requirements and recruit new ones would be stable over a very wide range of individual agent behaviours, given a diversity of agents in society. [Could we expand capitalism to zero-intelligence even if we include the formation and management of companies?]

A good test of a social network representation, for our needs, is: can it represent capitalism? Followed by: is it flexible enough to represent, e.g. feudalism, communism, sub-cultures, government, law-enforcement, Elinor Ostrom's cooperative groups... this is a pre-requisite to asking how structures can change.

The edges in this graph can be thought of as social contracts: mutual agreements that each participant will play their role, including when the contract may end, and perhaps even with restrictions on what other social contracts an agent may enter into (monogamy, employment, conflict of interest). If the representation of contracts are expressive enough, then agents can innovate by offering new contract types as well as forming novel structures using existing contract types.

A contract between two objects consists of an public interface for each object (this defines the set of "actions" each may take on the other) and a constraint on the strategy of each object (i.e. two sets of socially acceptable strategies, one for each participant). At its simplest, this can be a single strategy for each agent, but more realistic seems to be very simple contracts, but using quite complicated abstractions. Part of the evolution of society, in this case, becomes the evolution of the abstractions available for creating contracts.

[What if the contract is expressed as code (smart contract) and is seen as injecting code into the participating agents. In this way, it is the smart-contract code that is the unit of evolution]

[Is it important to have rewards for social status, or is it already contained in the rewards that come from status? In reality, there seems to be a "double counting" of rewards: If i have a job I hate and win the lottery, I'm happy about winning, presumably because my expectation of future rewards has increased, but this is in addition to the future rewards themselves. Does this, along with selective memory, explain why people are willing to make negative expectation gambles?]

Given a set of roles, agents could be aware of the need for a given tie, which could direct behaviour to fulfill that need.

Can two agents be linked by two different ties? i.e. play different roles in respect to each-other: it would seem so (perhaps my wife is also my dentist). In this case, how do agents distinguish roles?

8 Emergence of roles

If a society is defined by its available roles and its state of social tie, how does a society develop new roles? Once agents are identifiable, roles can be encoded in a policy, so the existance and evolution of social roles is really just an abstract way of talking about policy learning. So, what are the necessary/sufficient conditions on learning that lead to the emergence of social roles?

Can roles emerge without agent identifiability? [if roles are made clear at the start of the episode: i.e. there needs to be a negotiation of roles until agents agree on a social tie that would be Nash equilibrium for them. Can agents learn this? It would seem that if agents can learn from eachother and are commonly placed in the same position (i.e. there are a finite number of games they can play) then the Nash equilibria would emerge. In-fact the negotiation is part of the Nash equilibrium. Given a decision tree, an agent can use MCTS to verify that it is a solution.]

When agents are identifiable, roles can be encoded in "folk psychology": Agents assume other agents have an unknown "mind state" which is an input to other's policy. Over a longitudinal social-tie each agent gets to know the others folk psychological mind state.

Roles may emerge out of a lack of ability to track complex probability distributions over other's internal state. If we're both only able to track a small sample of states then the establishment of social roles helps the aganes to cooperate.

Stability of roles: If I live in a society of people who know and respect a set of roles,and I enter an epiosode and know what role other is playing and what role he expects me to play...I can use self-play (against the social roles) to form my policy, irrespective of my roles. If my policy coincides with the role I was supposed to play anyway (i.e. the roles identify a Nash equilibrium), then there is stability.

If we're using MCTS then the role can be used as the off-tree policy.

Learning roles: Suppese agents are able to learn from the exhibited behaviour of episodes of other pairs of agents, then roles just identify common Nash equilibria... perhaps we also need to see "lifestyles", i.e. how do multiple social ties fit together into the meta-game?