The Education of a Crook: Reinforcement Learning in Social-Cultural Settings

(Extended Abstract)

Taranjeet Singh Bhatia
tsbhatia@cs.ucf.edu

Saad Ahmad Khan
skhan@eecs.ucf.edu

Ladislau Bölöni
lboloni@eecs.ucf.edu

Dept. of Electrical Engineering and Computer Science
University of Central Florida
Orlando, FL 32816

ABSTRACT

The ability to manipulate social and cultural values in order to achieve one's own goals is a hard-to-teach but profitable skill. In this paper we represent a complex social scenario, the Spanish Steps flower selling scam, using a social calculus framework based on culture sanctioned social metrics (CSSMs) and concrete beliefs (CBs). Then, we show how a crooked seller can learn a profitable strategy through reinforcement learning. Although the search space defined by the social calculus is large, we found that function approximation based Q-learning allows us to successfully learn efficient strategies in a relatively small number of runs. The learned strategy allows the seller to manipulate an unprepared tourist's social values of politeness and dignity, as well as his perception of the peers and crowds opinion. This allows the seller to manipulate some of his opponents to act against their own interests by purchasing an overpriced flower while well-knowing that they are being cheated.

Categories and Subject Descriptors

I.2.11 [**Computing Methodologies**]: Artificial Intelligence—*Multiagent systems*

General Terms

Social cultural models, Q-learning, Reinforcement learning

Keywords

multi-agents; simulation; social-cultural

1. INTRODUCTION

There is a widespread belief that the ability to navigate the complexities of social and cultural interactions are one of the most difficult tasks for artificial intelligence entities (agents or robots) to achieve [4, 3].

The objective of this paper is to investigate the degree to which strategies associated with successful manipulation of a social scenario can be learned. To investigate this, we

consider a relatively complex social scenario, the Spanish Steps flower selling scam. In this scenario, a crooked seller tries to sell an overpriced flower to a tourist client, typically a romantic couple. After the clients decline to buy a bouquet, the seller offers a single flower with the implication that it is a gift. After letting the clients hold the flower for a while, the seller returns and asks for money. The scenario ends either with the client paying or escalating his return efforts sufficiently to force the seller to take back the flower. The clients know that they are being cheated: the fact that the scam sometimes succeeds requires a very precise manipulation of the social sentiments by the crook. We express this scenario using the social calculus techniques outlined in [1]. This framework represents the state of the actors in a social scenario using *culture sanctioned social metrics* (CSSMs) such as wealth, politeness and dignity, and their *concrete beliefs* (CBs) of the actors or crowd about topics relevant to the scenario such as "is the seller crooked"? Actions taken by the actors affect the CSSMs and CBs through *action impact functions* (AIFs). AIFs can, in principle, take arbitrary forms, but we argue that certain forms can represent both a convenient mathematical framework as well as offer sufficient fidelity in the modeling of real world scenarios. We used reinforcement learning to allow the agent to pursue its goals (expressed in terms of CSSMs). Our concrete task is learn a strategy where the Seller actor, as in when it is turn to take its action, will take actions which manipulate the CSSMs involved in a way so as to lead the scenario to a state desired by the seller (which involves the buyer buying the flower).

2. SCENARIO MODELING

The Spanish-Step scenario [1, 2] is turn taking scenario which means only one actor is responsible for taking action in a particular state. Therefore, it can be easily represented with a progress state graph. One example of progress state graph is shown below which demonstrates unsuccessful sale transaction between active participants i.e. the `Seller` (S) and the `Client` (C).

$$S0 \xrightarrow[\text{offer flowers}]{(\alpha 1, S)} S1 \xrightarrow[\text{decline buying}]{(\alpha 4, C)} S3 \xrightarrow[\text{offer gift}]{(\alpha 5, S)} S4 \xrightarrow[\text{decline gift}]{(\alpha 8, C)[x]}$$

$$S3 \xrightarrow[\text{give up}]{\alpha 15, S} TN$$
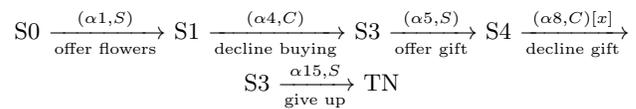
Action $\alpha 8$ is parameterized by "loudness" x which determine how many onlookers will over hear the transaction.

We use four CSSM metrics in this scenario, two concrete (`worth` and `time`) and two intangibles (`dignity` and `politeness`). In brief, `worth` is the sum of the financial worth of the person, measured in real world currency. People in general will try to increase their financial worth. The `dignity` of a person decreases by insult, while a successful revenge increases it. In general, the Western culture requires people to maintain their dignity. The `politeness` is the conformance to the perceived social norms of speech and gestures. A person uttering a rude statement lowers his own politeness (and offends the dignity of the target). These CSSMs will be considered as features of the states. Therefore, we describe the state as a collection of CSSMs.

$$\begin{cases} S &=& \{CSSM_1, \ldots, CSSM_n\} \\ CSSM_i &=& F_i^{CSSM}(A, S, a) \end{cases} \quad (1)$$

The AIF for CSSMs as shown in the Eq 1, defined as $\mathcal{F}$: $\mathbf{A} \times \mathbf{S} \times \mathbf{a} \to \mathbf{S}$ describes the way in which the state of a scenario instance evolves under the impact of the action and parametrization choices of the participating social agents. We interpret $S' = \mathcal{F}(A, S, a)$ as the new full state of the system if actor A performs action $a(\alpha, A, x_1 \ldots x_n)$ in state S. In order to reduce the design space of the F^{CSSM} functions we must answer three questions: (a) the subset of the state which substantively effect a given CSSM, (b) the shape of the AIFs functions and (c) the parametrized mathematical forms which can represent these shapes in a convenient way?

In this scenario, F^{CSSM} will be a numerical function depending only on the CSSMs and CBs whose estimator agent is the same as the estimator agent of $CSSM_i$. The update of a CSSM in the form of `CSSM_i = CSSM(C,M,SA,PA,EA_x)` will be kept and maintained by the estimator actor `EA_x`, and this actor only has access to the other CSSMs and CBs in its own private state.

The shape of these functions associated with tangible CSSMs usually have simple AIFs. For instance, if an action takes time t_a then the action will add this value to the time CSSM. If the action involves paying the sum of m_a dollars, this will decrease the wealth CSSM of the payee actor and increase it for receiving actor. Things are significantly more complicated for intangible social values, whose change can be highly nonlinear and dependent on multiple factors. For instance, for expressions of dignity, we find that humans have a sensibility threshold: they ignore trifling offenses. Similarly, there is an upper saturation threshold: a level at which the offense is so big that further increasing it would not affect the dignity level. Furthermore, the update of social values often depends on the beliefs: we are less offended by the angry voice of the interaction partner if we believe that he is right to be angry. We conclude that the shape of the AIF can include various positive or negative slopes, thresholds and saturation behaviors. This leads to the conclusion that it is appropriate to model the CSSM AIFs as a combination of generalized logistic functions as shown in Eq 2.

$$\begin{cases} F_i^{CSSM} &=& \sum_k \left(L(x_k, K_k, M_k, B_k) \right) \\ L(x_k, K_k, M_k, B_k) &=& \frac{K}{1+e^{-B(x_k - M)}} \end{cases} \quad (2)$$

K signifies the upper asymptote, M stands for the location of the largest growth and B for the growth rate for the specific parameter x associated with action such as *loudness* or *offensiveness* measure.

3. EXPERIMENTAL RESULTS

An effective action selection can lead to an optimum goal state. We are using the function approximation based Q-learning reinforcement where Q-values are linear combination of weighted features i.e. CSSMs. The Q-values are not dependent upon the action because the extracted features depend only on the state and get updated with the new feature's values from the AIF. The advantage of using function approximation based Q-learning is that we need not remember the Q-Value for every state-action pair, nor for features based on the states.

$$\begin{array}{rcl} Q(s,a) &=& \sum_i w_i f_i(s) \\ Q_{new}(s,a) &=& r(s,a,s') + \gamma \cdot (\sum_i w_i f_i(s')) \\ w_i &=& w_i + \alpha \cdot (Q_{new}(s,a) - Q_{old}(s,a)) f_i(s) \end{array}$$

Our implemented learning model traces the evolution of the weights for the feature vector. The features are represented as the CSSMs of individuals as well as the perception and belief values of the crowd. For the training model, we introduced a number of different consistence policy actors for the client agent to generalize our training of the seller i.e. *casual, busy, arrogant, smart*, and *wealthy*. These five consistent policy agents for the clients not only help in generalizing our model but also effectively cross-validate it against stochastic behavior shown by the agents during simulation.

4. REFERENCES

[1] T. Bhatia, S. Khan, and L. Bölöni. A modeling framework for inter-cultural social interactions. In *Proc. of the Second Int. Workshop on Human-Agent Interaction Design and Models (HAIDM-13) at AAMAS-2013*, pages 16–31, 2013.

[2] L. Bölöni. The Spanish Steps flower scam - agent-based modeling of a complex social interaction. In *Proc. of 11th Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS 2012)*, pages 1345–1346, Jun 2012.

[3] F. Grimaldo, M. Lozano, F. Barber, and G. Vigueras. Simulating socially intelligent agents in semantic virtual environments. *Knowledge Engineering Review*, 23(4):369, 2008.

[4] S. Mascarenhas, R. Prada, A. Paiva, and G. J. Hofstede. Social importance dynamics: A model for culturally-adaptive agents. In *Intelligent Virtual Agents*, pages 325–338. Springer, 2013.