Addressing Concept Drift in Reputation Assessment

Extended Abstract

Caroline Player
University of Warwick
Coventry, CV4 7AL, UK
c.e.player@warwick.ac.uk

Nathan Griffiths
University of Warwick
Coventry, CV4 7AL, UK
nathan.griffiths@warwick.ac.uk

ABSTRACT

In this paper, we address the limitations of existing methods to select representative data for trust assessment when agent behaviours can change at varying speeds and times across a system. We propose a method that uses concept drift detection to identify and exclude unrepresentative past experiences, and show that our approach is more robust to dynamic agent behaviours.

KEYWORDS

Trust; Reputation; Stereotypes; Concept Drift

ACM Reference Format:
Caroline Player and Nathan Griffiths. 2018. Addressing Concept Drift in Reputation Assessment. In *Proc. of the 17th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2018), Stockholm, Sweden, July 10–15, 2018,* IFAAMAS, 3 pages.

1 INTRODUCTION

In multi-agent systems (MAS) agents typically must interact with others to achieve their goals. Therefore it is vital to identify *trustworthy* interaction partners, but this is challenging in open and dynamic environments where there may be malicious agents or changeable conditions. In MAS, agents typically use trust and reputation models to choose who to interact with [12, 14]. Trust is the expectation one agent has in another to do a task, typically calculated using past experiences [4]. However, when agents can change their behaviour, past experience is no longer an indication of future performance. We assume that agents' abilities can change at different times and speeds from others as a result of time sensitive features including connectivity, task load and demand [9]. Existing trust and reputation models commonly use sliding windows and forgetting factors, which are limited in their ability to select data which is representative of an agent's current behaviour.

We propose a method, the AdWin Tree, for agents to select data they assess to be representative of true behaviours for trust and reputation assessment by integrating a *concept drift* detection algorithm. Our method proves resilient to dynamic behaviours.

2 BACKGROUND

Trust and reputation models use direct and indirect past experiences for an agent to estimate another agent's future performance [11]. Many statistical models focus on filtering out inaccurate reports where witness perceptions are different or they lie [5, 10, 15, 16]. However, existing approaches do not identify how representative these past experiences are of an agent's current behaviour. In this work, we adopt Beta Reputation System (BRS), a mathematically rigorous reputation model which is general enough to suit many applications, although any model can be substituted [6].

Stereotype models associate reputation with agents' observable features if we can assume that an agent might behave similarly to others who share the same observable features [2, 8, 13]. One advantage of assuming that agents have stereotypical behaviour is that our method to select relevant data is more proactive. For example, by identifying a negative change in an agent behaviour that shares a location with another agent, we might assume that connectivity there is currently poor and therefore neither are trustworthy.

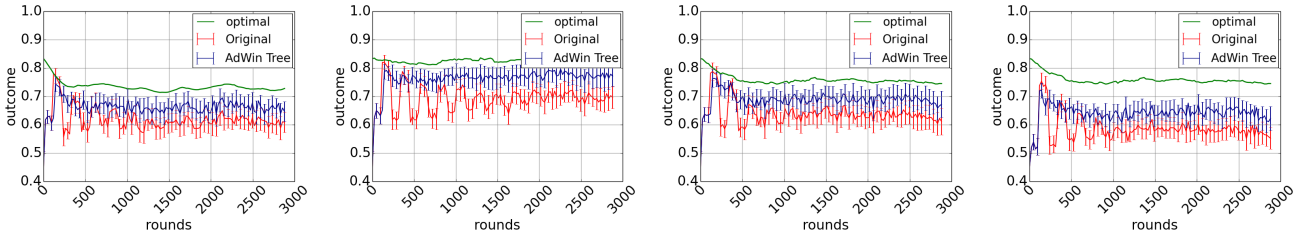
Some trust, reputation and stereotype models account for changes in behaviour over time with a sliding window or forgetting factor [5–7, 10, 15]. A sliding window of size n keeps the most recent n experiences. The intuition is that the most recent n instances capture representative data for an agent's current behaviour. A forgetting factor has a similar effect, by retaining all records but weighting recent interactions higher than older ones. There are several problems with sliding windows and forgetting factors. First, choosing the window size or rate of forgetting is challenging. If an agent forgets old instances too quickly they will lose relevant data that could increase accuracy. Conversely, if too much information is retained then agents will make assessments based on data that no longer represents current behaviours. The optimal values will depend on the application. Second, the optimal values may vary over time. Third, agents may not change their behaviour at the same rates, and so a global window or forgetting factor will not suit all agents. Finally, sliding windows and forgetting factors are only effective in coping with gradual change rather than sudden changes, which may occur at different times for different agents.

Concept drift techniques statistically detect when the relationship between features X and class Y in a dataset change between two time points i.e $P_t(XY) \neq P_{t'}(XY)$ [3]. This occurs when the posterior probability changes i.e. $P_t(Y|X) \neq P_{t'}(Y|X)$. For example, when the connectivity in a location changes, that location no longer correlates with the trust originally learned from data about that area. We use the Adaptive Windowing (AdWin) model which adjusts the size of a window of data by retaining only records which are statistically assessed to be relevant [1]. Unlike some other techniques, AdWin specifically identifies the point of change and therefore it is compatible with managing memory for trust models.

3 AGENT MODEL

A set of agents is divided into *trustors* and *trustees*, and are connected in a complete bipartite graph where a trustor represents a consumer, i, trying to identify a trustworthy service provider, a trustee, j.

Proc. of the 17th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2018), M. Dastani, G. Sukthankar, E. André, S. Koenig (eds.), July 10–15, 2018, Stockholm, Sweden.



(a) Only gradual drift. $p_{Gr} = 0.2$

(b) Only sudden drift. $p_{Su} = 0.001$

(c) $p_{Gr} = 0.2$, $p_{Su} = 0.005$ p_{te} **= 0.1**

(d) $p_{Gr} = 0.2$, $p_{Su} = 0.005$ p_{te} **= 0.6**

Figure 1: Results

Agent j can be described by their ID and the set of their *observable features*, $\overrightarrow{\tau} j$. Each feature represents a characteristic of an agent, for example technical specifications of a device in a network. Some features correlate with behaviour, known as *relevant features*, while *irrelevant features* are random to represent a realistic assumption that not all observable characteristics of an agent affect behaviour and the model needs to be robust to this.

The behaviour of agent i is represented as a real value $B_i \in [0, 1]$ indicating the proportion of interactions they behave well in. For example, if $B_i = 0.7$, agent i will behave well in 7 out of 10 interactions on average. Trust, reputation and stereotype systems aim to estimate the behaviour of agents and identify the agent most likely to behave well in its next interaction.

Agents' behaviour and relevant features are defined by the profiles they belong to, and we evaluate using 5 profiles. All the agents of a profile experience the same changes in their behaviour value. Agents are undergoing concept drift when they exhibit the same observable features as before but their behaviour changes. This change can occur slowly or abruptly, known as *gradual* and *sudden* drift respectively. We define environmental parameters, p_{Gr} and p_{Su}, to describe the probability of these changes occurring.

A trustor selects one trustee to interact with per round. Trustors maintain a history of interactions where a record contains the partner's ID, observable features, time of the interaction and number of good and bad experiences with that partner. As we model open systems, at the end of a round trustees can leave the network with a probability p_{te} to be replaced by a new agent of the same profile, thus maintaining an equal distribution of agent behaviours.

4 ESTIMATING EXPECTED BEHAVIOUR

Using BRS, trustors calculate trust in a trustee using their own and witnesses' reports of direct interactions. BRS requires an *a priori* estimate of behaviour, which is determined by the output of the stereotype model and is given less precedence over time as direct experiences are collected. The stereotype model used is Burnett's decision tree [2]. Each leaf of a trustor's decision tree represents an estimate of an agent profile i.e a stereotype. We build an AdWin model at each leaf to monitor for behaviour changes at different rates and times in profiles [1].

After an interaction, the trustor saves the outcome to the AdWin model for the appropriate stereotype by classifying the trustee's observable features with the decision tree. The AdWin model concatenates the binary outcome to the end of a variable size window, which contains outcomes from interactions with other agents of that detected stereotype. To detect if drift has occurred, the window is divided into every possible split of two subwindows. The distributions found to fit the data of each subwindow are compared, and if they are outside a margin of similarity then concept drift is assumed to have occurred. In such cases, the older of the two subwindows is deleted, as are the entries those instances correspond to in the interaction history. The M5 decision tree is then rebuilt with the remaining data. Thus, both trust and stereotype assessments use appropriate information

5 EVALUATION AND DISCUSSION

We compare the AdWin Tree to the original stereotype tree, evaluated with a fixed window size of 100 [2]. The performance metric is the average utility trustors receive at each time, since this indicates of how well trustors identified the best behaving partners[1].

The results in Figure 1 show that the AdWin Tree outperforms the original approach, suffering less severe sudden drops in performance, meaning that it is more resilient to behaviour changes. When agents have less chance for repeat interactions, the dependency on stereotypes is higher, therefore we notice a larger difference in performance between Figures 1c and 1d as p_{te} increases.

Noisy features cause existing techniques to perform badly given dynamic behaviour. However, our model retains a higher proportion of relevant data and can more accurately identify relevant features, allowing it to better estimate the agent profiles in the system. The performance of our method is bottlenecked by how accurately the decision tree can identify agent profiles as stereotypes. Each AdWin model is applied to an identified stereotype, and if these are inaccurately identified then multiple profiles will be analysed by a single AdWin model, making it harder to monitor changes in a single profile.

ACKNOWLEDGEMENTS

The lead author gratefully acknowledges funding from the UK EPSRC (grant no. EP/L016400/1), the Centre for Doctoral Training in Urban Science.

[1]All results presented were found to be statistically significant with $p < 0.001$ in paired t-tests using the area under the curve to remove the time dependency.

REFERENCES

[1] A Bifet and R Gavaldà. 2007. Learning from time-changing data with adaptive windowing. In *Proceedings of the 2007 SIAM International Conference on Data Mining*. 443–448.

[2] C Burnett, T J Norman, and K Sycara. 2010. Bootstrapping Trust Evaluations through Stereotypes. In *Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems*. 241–248.

[3] J Gama, I Žliobait, A Bifet, M Pechenizkiy, and A Bouchachia. 2014. A survey on concept drift adaptation. *ACM Computing Surveys (CSUR)* 46, 4 (2014), 44.

[4] D Gambetta. 2000. Can we trust trust? *Trust: Making and breaking coooperative relations* 13 (2000), 213–237.

[5] T Dong Huynh and N Jennings. 2004. FIRE: An integrated trust and reputation model for open multi-agent systems. In *ECAI 2004: 16th European Conference on Artificial Intelligence, August 22-27, 2004, Valencia, Spain: including Prestigious Applicants [sic] of Intelligent Systems (PAIS 2004): proceedings*, Vol. 110. 18.

[6] A Jøsang and R Ismail. 2002. The beta reputation system. *Proceedings of the 15th Bled Electronic Commerce Conference* 5 (2002), 2502–2511.

[7] Z Liang and W Shi. 2005. PET: A personalised trust model with reputation and risk evaluation for P2P resource sharing. In *Proceedings of teh 38th Annual Hawaii International Conference on System Sciences*.

[8] X Liu, A Datta, K Rzadca, and E Lim. 2009. Stereotrust: a group based personalized trust model. In *Proceedings of the 18th ACM conference on Information and knowledge management*. 7–16.

[9] R Raje, B R Bryant, A M Olson, M Auguston, and C Burt. 2002. A quality-of-service-based framework for creating distributed heterogeneous software components. *Concurrency and computation: practive and experience* 14, 12 (2002), 1009–1034.

[10] K Regan, P Poupart, and R Cohen. 2006. Bayesian reputation modelling in e-marketplaces sensitive to subjectivity, deception and change. In *Proceedings of the National Conference on Artificial Intelligence*.

[11] P Resnick, K Kuwabara, R Zeckhauser, and E Friedman. 2000. Reputation Systems. *Commun. ACM* 43, 12 (2000).

[12] N Rodrigues, P Leitão, and E Oliveira. 2015. Self-interested service-oriented agents based on trust and QoS for dynamic reconfiguration. In *Service Orientation in Holonic and Multi-agent manufacturing*. Springer, 209–218.

[13] M Sensoy, B Yilmaz, and T J Norman. 2016. Stage: Stereotypical trust assessment through graph extraction. *Computational Intelligence* 32, 1 (2016), 72–101.

[14] L Teacy, M Luck, A Rogers, and N Jennings. 2012. An efficient and versatile approach to trust and reputation using hierarchical bayesian modelling. *Artificial Intelligence* 193 (2012), 149–185.

[15] L Teacy, J Patel, N Jennings, and M Luck. 2006. Travos: Trust and reputation in the context of inaccurate information sources. *Autonomous Agents and Multi-Agent Systems* 12, 2 (2006), 183–198.

[16] L Xiong and L Liu. 2004. PeerTrust: Supporting reputation-based trust for peer-to-peer electronic communities. *IEEE transactions on knowledge and data engineering* 16, 7 (2004), 843–857.