Preference Learning in Automated Negotiation Using Gaussian Uncertainty Models

Extended Abstract

Haralambie Leahu
Centrum Wiskunde & Informatica
Amsterdam, The Netherlands
H.Leahu@cwi.nl

Michael Kaisers
Centrum Wiskunde & Informatica
Amsterdam, The Netherlands
M.Kaisers@cwi.nl

Tim Baarslag
Centrum Wiskunde & Informatica
Amsterdam, The Netherlands
T.Baarslag@cwi.nl

ABSTRACT

In this paper, we propose a general two-objective Markov Decision Process (MDP) modeling paradigm for automated negotiation with incomplete information, in which preference elicitation alternates with negotiation actions, with the objective to optimize negotiation outcomes. The key ingredient in our MDP framework is a stochastic utility model governed by a Gaussian law, formalizing the agent's belief (uncertainty) over the user's preferences. Our belief model is fairly general and can be updated in real time as new data becomes available, which makes it a fundamental modeling tool.

KEYWORDS

Automated Negotiation; Preference Elicitation; Gaussian Process.

ACM Reference Format:
Haralambie Leahu, Michael Kaisers, and Tim Baarslag. 2019. Preference Learning in Automated Negotiation Using Gaussian Uncertainty Models. In *Proc. of the 18th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2019), Montreal, Canada, May 13–17, 2019,* IFAAMAS, 3 pages.

1 INTRODUCTION

Automated (agent-based) negotiation is a broadly applicable research topic in AI, with notable applications in high-frequency trading [6], cloud computing [12], pervasive computing [11], smart grids [15], supply chain management [7]. In such system architectures, agents can successfully substitute humans in making decisions, provided that the user's goals are accurately known [10].

In many situations agents do not have access to all information required for taking optimal decisions and need to elicit relevant information by interacting with the user, striking a balance between negotiation and preference elicitation. In particular, the agent needs to decide what information is relevant for the next negotiation step and how to integrate it into the decision process.

Therefore, designing agents that can efficiently learn and integrate user's preferences into decision making processes is a key challenge in automated negotiation. While accurate knowledge of the user's preferences is highly desirable, eliciting the necessary information might be rather costly, since frequent user interactions may cause inconvenience. Therefore, developing efficient elicitation strategies (minimizing elicitation costs) for inferring relevant information is of critical importance in automated negotiation.

Proc. of the 18th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2019), N. Agmon, M. E. Taylor, E. Elkind, M. Veloso (eds.), May 13–17, 2019, Montreal, Canada.

2 A TWO-OBJECTIVE MDP APPROACH

Partially Observable Markov Decision Process models have been proposed for both (automated) negotiation [9] and preference elicitation [2, 4]. Preference elicitation models were further adapted to negotiation processes in which agents may elicit absolute utility values by submitting queries to the user [1, 8]. We propose here a two-objective MDP approach, in which the agent aims to optimize the expected reward, obtained only when reaching an agreement. To this end, two types of actions are available to the agent: *queries* (submitted to the user) and *offers* (submitted to the opponent).

The lack of information on the user/opponent's reasoning is impeding a proper MDP formulation. For instance, neither the probability of reaching a certain agreement nor the size of the corresponding reward are known. To cope with this uncertainty, the agent resorts to *beliefs* about the user's and opponent's goals. As new actions are undertaken, the agent receives feedback and updates its beliefs accordingly, as indicated in Fig. 1.

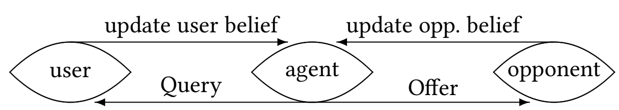

Figure 1: Agent's interactions and belief updates.

The agent's beliefs are formalized as probability laws on relevant spaces and further used to define the relevant MDP elements, e.g. transition probabilities and rewards, as random variables. The beliefs are updated sequentially, according to Bayesian rules, based on the observed reactions, i.e. answers from the user, resp. counter-offers from the opponent. Desirable features of belief models are:

- *generality*, to account for a wide range of possible scenarios;
- *tractability*, to allow easy integration of new information.

A typical way of deriving the quantitative elements defining the MDP model of interest is by means of (stochastic) *utility functions*, quantifying the user/opponent's preference for each negotiation agreement [2, 4], in which the stochasticity (randomness) is interpreted as the agent's belief. MDP *states* include the current beliefs, *rewards* are defined as random (under the current user belief) utilities of future negotiation agreements, whereas the future opponent moves are predicted based on its (belief-based) utility function. A desirable feature of a stochastic utility function is *data-measurability*; that is, the utility function must be completely determined by the observable data. Indeed, introducing non-measurable randomness in the utility model might have detrimental effect, resulting in inconsistent results depending on (subjective) prior beliefs [16].

3 GAUSSIAN BELIEF MODELS

Gaussian Processes are a rather popular choice in belief-modeling in Machine Learning [3]. Their popularity stems from the fact that they cover a wide range of (utility/preference) functions, outperforming (in terms of generality) competing (e.g. linear) models, where extra assumptions are typically required (e.g. existence of trade-offs in multi-issue negotiation, for linear models). In addition, the Laplace Approximation of the Bayesian posterior (resulting in relatively simple update rules) makes GP belief models tractable.

For the sake of generality, we focus on *ordinal* utility functions, i.e. measurable w.r.t. ordinal data (pairwise comparisons between alternatives), in contrast to [4], where *cardinal* utility functions, which require sampling absolute utility values, are used. Ordinal utilities provide arguably more realistic models than cardinal ones, since absolute values are rarely available in practice and - even then - less consistent over time (than ordinal preferences) [5]. Therefore, we assume that the user's preferences define a partial order (i.e. reflexive, anti-symmetric and transitive relation) on the negotiation space (of all possible agreements), which is consistent over time.

To handle ordinal utility functions (thus, assuming that only ordinal data is observable), we adopt the *instance preference learning* framework [13, 14], where a belief is formalized by a Gaussian law over the class of functions on negotiation space, parametrized by a pair of mean, resp. covariation, functions. A random sample is then a GP (to be understood as a stochastic preference function), based on which preferences between pairs of (different) outcomes are derived by direct comparison of the corresponding GP values. GP's are not measurable w.r.t. ordinal data (since the magnitude of the utility value of some particular outcome can not be inferred by comparing it to all other outcomes) and need to be converted into ordinal utility functions, before being integrated into our model.

Beliefs are subject to Bayesian updating, given (a set of) ordinal data. Exact formulae for sequential updating, i.e. for the new mean and covariation functions, can be derived from the general update procedure [14] and integrated into a belief-based preference learning scheme, compatible with the MDP model in Section 2. Initial beliefs are chosen based on prior knowledge and assumptions on how utility values are correlated. Monotonicity assumptions can be also included (as ordinal data) in the initial belief(s).

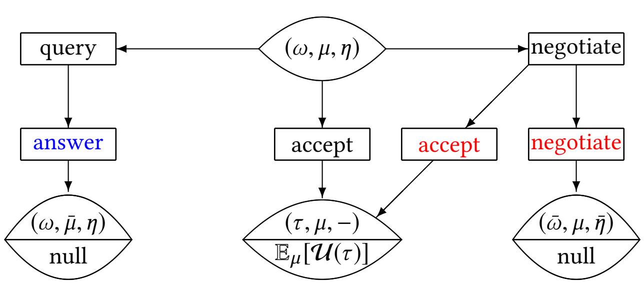

Figure 2: The belief MDP dynamics. The rectangles denote actions (black text for the agent, blue for the user and red for the opponent) and the rounded shapes represent MDP states with respective (expected) immediate rewards.

4 THE GP-BASED BELIEF MDP MODEL

To integrate the Gaussian belief model described in Section 3 in the two-objective MDP model in Section 2, we consider generic MDP (transitory) states in the form $S = (\omega, \mu, \eta)$, where:

- ω denotes the actual state of the negotiation and encodes all relevant information about the negotiation history;
- μ and η denote (Gaussian) beliefs about the user's, resp. opponent's, preferences.

Terminal negotiation states (corresponding to negotiation agreements) are generically denoted by τ. MDP states (τ, μ, η) are terminal, hence they are the only states 'paying off' a reward $\mathcal{U}(\tau)$ (if reached by the MDP), where $\mathcal{U}$ denotes a stochastic ordinal utility function under the (current) belief μ; in particular, the η-component becomes irrelevant in terminal states.

The dynamics of the belief MDP are graphically illustrated in Fig. 2. Assume that the agent is in the belief state (ω, μ, η). In line with the interaction scheme in Fig. 1, should it pose a (comparative) query to the user, it updates the user belief μ based on the corresponding answer (translated into ordinal data), resulting in a new user belief $\bar{\mu}$. On the other hand, should the agent accept the offer currently on the table (specified by ω), the MDP moves in a terminal state τ and a reward $\mathcal{U}(\tau)$ is obtained. Finally, should the agent submit an offer to the opponent, it can either get accepted, resulting in a terminal negotiation state τ' with corresponding reward $\mathcal{U}(\tau')$, or be negotiated, resulting in a counter-offer from the opponent, which is used in two-ways: to update the negotiation state (resulting in a new state $\bar{\omega}$) and to update the opponent belief (resulting in a new belief $\bar{\eta}$), based on the assumption that the counter-offer is preferred by the opponent to the submitted offer.

5 CONCLUSIONS

We propose an MDP model for automated negotiation with incomplete information, based on stochastic ordinal utility functions, in which the uncertainty is formalized by a Gaussian belief. Gaussian belief models are both general and tractable, thus providing an attractive alternative for preference modeling.

A key step in this formalism is converting preferences into utility values, without introducing 'extra' randomness in the utility model. This requires a 'standardized' method for deriving absolute (utility) values from ordinal data. A general approach is to assign a deterministic 'weight' to each pair of outcomes, which is added to the utility value of the preferred one (if any).

In our model, the agent aims to maximize the expected utility of the (future) negotiation agreement, with respect to all possible sequences of actions and user/opponent's reactions. A one-step look-ahead strategy on agent actions and expected response provides a first-hand, heuristic solution that future work may further refine, by designing efficient (maximum) search algorithms.

ACKNOWLEDGMENTS

This work has received funding from the ERA-Net Smart Energy Systems' focus initiative Smart Grids Plus, within the European Union's Horizon 2020 research and innovation programme, under grant No. 646039. It was also supported by the Veni research programme of the Dutch Organisation for Scientific Research (NWO), under the project No. 639.021.751.

REFERENCES

[1] Tim Baarslag and Enrico H. Gerding. 2015. Optimal Incremental Preference Elicitation during Negotiation. In *Proceedings of the Twenty-fourth International Joint Conference on Artificial Intelligence (IJCAI'15)*. AAAI Press, 3–9. http://dl.acm.org/citation.cfm?id=2832249.2832250

[2] C. Boutilier. 2002. A POMDP Formulation of Preference Elicitation Problems. *18th National Conference on Artificial Intelligence* (2002), 239–246. http://dl.acm.org/citation.cfm?id=777092.777132

[3] C.E. Rasmussen and C.K.I. Williams. 2006. *Gaussian Processes for Machine Learning*. MIT Press. www.GaussianProcess.org/gpml

[4] Urszula Chajewska, Daphne Koller, and Ronald Parr. 2000. Making Rational Decisions using Adaptive Utility Elicitation. In *In Proceedings of the Seventeenth National Conference on Artificial Intelligence*. 363–369.

[5] D.C. Kingsley and T.C. Brown. 2006. Preference Uncertainty, Preference Learning and Paired Comparison Experiments. *Land Economics* 84 (2006), 530–544.

[6] F. McGroarty, A. Booth, E. Gerding, and V.L.R. Chinthalapati. 2018. High-frequency Trading Strategies, Market Fragility and Price Spikes: An Agent-based Model Perspective. *Annals of Operations Research* (2018), 1–28. https://doi.org/10.1007/s10479-018-3019-4

[7] M. Wang, H. Wang, D. Vogel, K. Kumar, and D.K.W. Chiu. 2009. Agent-based Negotiation and Decision Making for Dynamic Supply Chain Formation. *Engineering Applications of Artificial Intelligence* 22 (2009), 1046–1055. Issue 7.

[8] Yasser Mohammad and Shinji Nakadai. 2018. FastVOI: Efficient Utility Elicitation During Negotiations. In *International Conference on Principles and Practice of Multi-Agent Systems*. Springer, 560–567.

[9] P. Paruchuri, N. Chakraborty, R. Zivan, K. Sykara, M. Dudik, and G. Gordon. 2009. POMDP Based Negotiation Modeling. *IJCAI MICON Workshop* (2009).

[10] S. Kraus, J. Wilkenfeld, and G. Zlotkin. 1995. Multiagent Negotiation under Time Constraints. *Artificial Intelligence* 75 (1995), 297–345.

[11] S.D. Ramchurn, B. Deitch, M.K. Thompson, D.C.D. Roure, N.R. Jennings, and M. Luck. 2004. Minimising Intrusiveness in Pervasive Computing Environments Using Multi-agent Negotiation. *Mobile and Ubiquitous Systems: Networking and Services* (2004), 364–371. https://doi.org/10.1109/MOBIQ.2004.1331743

[12] K.M. Sim. 2011. Agent-based Cloud Computing. *IEEE Transactions on Services Computing* 5 (2011), 564–577. Issue 4.

[13] W. Chu and Z. Ghahramani. 2005. Gaussian Processes for Ordinal Regression. *Journal of Machine Learning Research* (2005), 1019–1041.

[14] W. Chu and Z. Ghahramani. 2005. Preference Learning with Gaussian Processes. *Proceedings of the 22nd Conference on Machine Learning* (2005), 137–144.

[15] W. Ketter, J. Collins, and P. Reddy. 2018. Power TAC: A Competitive Economic Simulation of the Smart Grid. *Energy Economics* 39 (2018), 262–270.

[16] Y. Ben-Haim and M. Demertzis. 2016. Decision Making in Times of Knightian Uncertainty: An Info-Gap Perspective. *The Open-Access, Open-Assessment E-Journal* 23 (2016), 1–29. Issue 10.