Estimating Second-Order Arguments in Dialogical Settings (Extended Abstract)

Seyed Ali Hosseini
Department of Informatics
King's College London
ali.hosseini@kcl.ac.uk

Sanjay Modgil
Department of Informatics
King's College London
sanjay.modgil@kcl.ac.uk

Odinaldo Rodrigues
Department of Informatics
King's College London
odinaldo.rodrigues@kcl.ac.uk

ABSTRACT

This paper proposes mechanisms for agents to model other agents' arguments, so that modelling agents can anticipate the likelihood that their interlocutors can constructs arguments in dialogues. In contrast with existing works on "opponent modelling" which treat arguments as abstract entities, the likelihood that an agent can construct an argument is derived from the likelihoods that it possesses the beliefs required to construct the argument. We therefore also address how a modeller can quantify the certainty that its interlocutor possesses beliefs based on previous dialogues, and membership of interlocutors in communities.

Keywords

Opponent modelling, Argumentation-based dialogue

1. INTRODUCTION

In real world models of computational dialogue involving computational and or human participants, one needs to address the strategic choice of locutions and use of enthymemes (i.e. arguments with incomplete logical structures [1]). This requires *opponent modelling* capabilities i.e. agents modelling their interlocutors' arguments. Hence an agent can anticipate the possible counter-arguments in a given dialogue, thus moving an argument that is least susceptible to attack, and avoid sending information in enthymemes that is already known to others. In such setups, in addition to constructing its own arguments from its knowledge-base (*first-order* arguments), an agent can generate a model of its interlocutor's arguments (*second-order* arguments).

In existing works on opponent modelling, agents assign values $[0, 1]$ to abstract arguments, representing the likelihood that they can be constructed by other agents. However, models of second-order *abstract* arguments are incomplete as they do not account for all second-order arguments that can be constructed from constituent beliefs. Furthermore, most existing works do not address the provenance of these values. Here we propose two sources for these values: the information exchanged in dialogues an agent participates in, and (when dialogical data is insufficient) a quantitative measure of similarity amongst all agents in the environment, using the notion of *agent communities*.

Appears in: *Proceedings of the 15th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2016), J. Thangarajah, K. Tuyls, C. Jonker, S. Marsella (eds.), May 9–13, 2016, Singapore.*

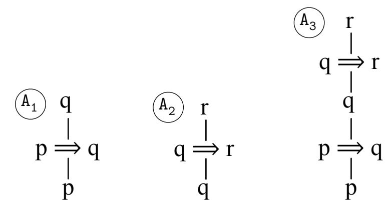

Figure 1: Several arguments

2. THE MODEL

Consider an agent i determining the likelihood that another agent j can construct a certain argument. Unlike existing works [3], we argue that this issue cannot be addressed without explicit access to the constituents of arguments.

Example 1. Suppose j submits only arguments $\mathtt{A}_1$ and $\mathtt{A}_2$ in Figure 1, in distinct dialogues with i. If i treats arguments only as abstract entities, it would believe that j only has arguments $\mathtt{A}_1$ and $\mathtt{A}_2$, and not $\mathtt{A}_3$. It is however clear that j can also construct $\mathtt{A}_3$ because it has the required beliefs to do so.

Hence access to the constituents of arguments is required when valuating the likelihoods of their construction. One common approach, though studied in the context where values denote likelihoods of truth [2], is to derive the values associated with arguments from the those associated with their constituent beliefs. In our context, this would mean that the likelihood that j can construct an argument $\mathtt{A}$ is derived from the likelihoods that it has the necessary beliefs to construct $\mathtt{A}$. The next step thus is to address the likelihoods that are associated with second-order beliefs.

We consider two complementary mechanisms for i to evaluate the likelihood that j has a certain belief α. The first is i's history of dialogical interactions. For this, an assignment d_{ij} for any two agents i, j is defined, where given i's history of dialogues and a belief α, it returns a number $[0, 1]$ representing the likelihood that j has α. Additionally, i records the case where dialogical evidence suggests that j does not have α (in which case $\mathsf{d}_{ij}(\alpha) = \bot$). By default, for every two agents i, j and belief α, $\mathsf{d}_{ij}(\alpha) = 0$, as i has no dialogical data regarding j having α. However, once i receives such evidence in a dialogue, $\mathsf{d}_{ij}(\alpha)$ is updated. For example, when j directly commits to α in a dialogue with i, in which case $\mathsf{d}_{ij}(\alpha) = 1$, or when i is indirectly informed of j's belief in α, e.g. by another agent k, in which case $\mathsf{d}_{ij}(\alpha)$ could correspond to i's level of trust in k. Failed information-seeking dialogues from i to j regarding α would trigger $\mathsf{d}_{ij}(\alpha) = \bot$.

In case dialogical data is insufficient to determine whether j has α, i can refer to its dialogical data regarding other

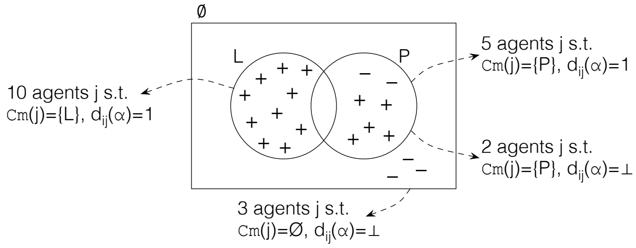

Figure 2: The figure corresponding to Example 2

agents knowing α, while factoring in their relationship with j. This relationship is modelled through the notion of *agent groups* and *communities*.

A *group* of agents is a set of agents sharing a certain property (e.g. organisational roles). A general assumption underpinning this model is that the shared property between the members of a group licenses their sharing of a specific set of beliefs. As agents may have multiple properties, agent groups may intersect, and each of these intersections may themselves license the sharing of a separate set of beliefs between its members. For example, assume a group A of agents all having beliefs $\mathcal{B}_A$ and a group B of agents with $\mathcal{B}_B$. Assuming a monotonic relationship between group membership and having beliefs, for the set $\mathcal{B}_{AB}$ of beliefs shared between agents in AB (i.e. agents in $A \cap B$), we have $\mathcal{B}_{AB} \supseteq \mathcal{B}_A \cup \mathcal{B}_B$, where the set $\mathcal{B}_{AB} \setminus \{\mathcal{B}_A \cup \mathcal{B}_B\}$ is the unique set of beliefs shared amongst AB's members because of their membership to "both A and B". Therefore, given the set $\mathcal{G}$ of all groups, we are interested in $2^{\mathcal{G}}$ where each of its elements is called a *community*. Given communities $\{A\}$ and $\{A, B\}$ (henceforth denoted A and AB), the latter is more *specific* than the former – due to their members having more properties – and the former is more *general* than the latter.[1]

Example 2. Let $\mathcal{G} = \{\mathrm{L}, \mathrm{P}\}$, where L and P respectively denote "lawyers" and "paralegals", and let α be technical legal information. The experience of i after consulting with several legal firms is summarised in Figure 2 which shows agents' community memberships and whether they believe $(+)$ or do not believe $(-)$ α, according to i. In this context, community $\emptyset$, containing all agents, represents "those working in a legal firm" and $\mathtt{Cm}(j)$ is the community j belongs to.

For any belief α and agent i, a set $\mathsf{c}_i(\alpha)$ of tuples $\langle \kappa, p \rangle$ is determined where κ is a community, and $p \in [0, 1]$, called a p-score, is the likelihood that a member of κ believes α according to i. The p-scores can be determined based on standard conditional probability $\sum_{x \in \kappa} \mathsf{d}_{ix}(\alpha) \,/\, |\{x \in \kappa \,|\, \mathsf{d}_{ix}(\alpha) \neq 0\}|$. For community $\emptyset$ in Example 2, the p-score would be 15/20, and so $\mathsf{c}_i(\alpha) = \{\langle \mathrm{L}, 1 \rangle, \langle \mathrm{P}, 0.71 \rangle, \langle \emptyset, 0.75 \rangle\}$. However in this approach, some communities (e.g. LP) may not be assigned any p-score, which could for example be due to incomplete data (i.e. agent i not having dialogues with any of their members regarding α). Moreover, some communities (e.g. $\emptyset$) might be assigned superficially high values due to being too general. In the case of $\emptyset$ (i.e. those working in a legal firm), the p-score suggests that there is 0.75 probability that another member of $\emptyset$ believes α. However, a large proportion of this value (10 out of 15 agents contributing to 0.75) is due to agents being lawyers.

[1] In this context, agents in $\emptyset \in 2^{\mathcal{G}}$ do not need to have any specific properties, essentially all agents in the environment, and the beliefs shared amongst them is *common knowledge*.

To remedy these issues, the p-scores are calculated iteratively, which ensures that **a)** those communities without a p-score value inherit the highest value of the most specific communities that are more general (e.g. L and P in the case of LP), and **b)** that each agent contributes only to the p-score of the community whose membership is the most likely reason why it believes α (i.e. the community with the highest $\mathsf{c}_i(\alpha)$ value). In the case of Example 2, in iteration 1 the p-score of L is set to 1 because it has the highest $\mathsf{c}_i(\alpha)$ value. Then from the remaining agents and communities, P will receive 0.71 and $\emptyset$ will receive 0.25, thus at iteration 2, P's p-score is set to 0.71. In iteration 3, from the remaining agents, $\emptyset$ will be assigned 0, as all of its members are in other communities with higher p-score values.

These values are then aggregated in a function $\mathsf{c}_{ic}(\alpha)$ for every agent i, community c, and belief α, representing the likelihood that a member of c believes α according to i. Together with $\mathsf{d}_{ij}(\alpha)$, these functions provide an approximate model of j's beliefs, according to i.

3. CONCLUSION

With our proposed mechanisms, an agent is able to utilise its dialogue history and its knowledge of agents' distribution in communities to determine the likelihood that another agent has certain beliefs, and subsequently derive the likelihood that it can construct arguments from those beliefs.

Suppose now that to persuade j to accept ϕ, i wants to send an argument for ϕ. By anticipating the arguments that j can construct and submit as counter-arguments, i can strategically choose from amongst all its arguments for ϕ, those that are least susceptible to be attacked by j. Let $\mathtt{Poss}(\phi)$ denote the set of all possible arguments claiming ϕ that i can construct. For each argument $\mathtt{A}$ in $\mathtt{Poss}(\phi)$, i must first identify every possible counter-argument to $\mathtt{A}$ along with the likelihoods associated with j being able to construct them. Then for each argument in $\mathtt{Poss}(\phi)$, i would consolidate the number of its counter-arguments and their likelihood values, allowing it to select the argument which is least likely to be attacked by j.

Another application is the use of enthymemes in dialogues. For this, i examines all sub-arguments $\mathtt{A}'$ of an argument $\mathtt{A}$ it wants to send to j, and removes $\mathtt{A}'$ from $\mathtt{A}$ if its likelihood value is higher than a certain threshold.

In the future, we will consider enabling agents to not only anticipate the arguments of other agents, but also what arguments they deem acceptable, which would allow for devising even more sophisticated strategies in dialogues.

REFERENCES

[1] S. A. Hosseini, S. Modgil, and O. Rodrigues. Enthymeme construction in dialogues using shared knowledge. In *Computational Models of Argument*, volume 266 of *Frontiers in Artificial Intelligence and Applications*, pages 325 – 332. IOS Press, 2014.

[2] A. Hunter. A probabilistic approach to modelling uncertain logical arguments. *International Journal of Approximate Reasoning*, 54(1):47 – 81, 2013.

[3] H. Li, N. Oren, and T. Norman. Probabilistic argumentation frameworks. In *Theories and Applications of Formal Argumentation*, volume 7132 of *Lecture Notes in Computer Science*, pages 1–16. 2012.