Q-learn Argumentation Schemes for Car Sales Dialogues

Adrian Groza
Department of Computer Science
Baritiu 28, RO-3400 Cluj-Napoca, Romania
adrian@cs-gw.utcluj.ro

Abstract

Agents need to argue with other agents many times, developing persuasion strategies that are effective over repeated situations. Applying reinforcement learning (RL) to the design of argumentation policies is appealing to dialogues where the counterpart can be modelled as a probability distribution. The idea of this research is to apply RL to speech acts in order to learn which discourse pattern is best to be conveyed during an argumentation game. Empowered by this learning mechanism, the persuasive agents gradually become more skillful through repeated argumentation.

1 Introduction

Over the past years, the research on argumentation theory has focused on identifying the most adequate technical instrumentation for modelling argumentation - defeasible logic seems to be the answer to these attempts. Recently, in the context of large scale argumentation of World Wide Argument Web (WWAW) [5], the interest has been shifted towards frameworks where all different inference mechanisms are able to co-exist under one umbrella, presently given by the Argument Interchange Format (AIF) ontology.

Learning and argumentation are related, sharing the same motivatations of increasing efficiency or the range of solvable problems [4]. Agents have to adapt their persuasion policies depending on the failure or success of past dialogues, using a learning mechanism. The RL theory is quite advanced, but applications have been limited to problems of control, operatios research, game playing. Two advandages of applying RL to argumentation sustain this research: i) the persuasive agent (lawyer, seller, teacher), can learn from its own experience; and ii) he can learn from simulation games [2].

2 Argumentation Framework

Extending the AIF ontology. We introduce a new node type, namely *context node* ($CO-node$) in the AIF ontology [5]. We argue the neccessity of this node, due to the fact that context exists independent of any object in the system. Thus, one context may be used to evaluate different arguments, whilst the same argument can be evaluated in different contexts. Because of the separation of the argument structure from contexts, more power to re-use arguments and decide when to apply a specific argument in a given state is provided.

The extended-AIF ontology has five types of nodes: An *information node* ($I-node$) represents passive information of an argument, such as claim, premise, data, locution. A *scheme node* ($S-node$) captures active information or domain-independent patterns of reasoning. The schemes are split in three disjoint sets, whose elements are: rule of inference schemes ($RA-node$), conflict application node ($CA-node$), preference application node ($PA-node$). *Forms of arguments* ($F-nodes$) model argumentation schemes, by defining premises descriptor, conclusion descriptor, presumptions and exceptions. *Protocol interaction nodes* ($PIA-node$) are used to constrain the dialogue moves within an argumentation process. $CO-nodes$ encapsulate all that is considered relevant to model the state of the argument, among the following contextual dimensions: *dialectical*, *intentional*, and *social*.

Argumentation schemes as protocol. Formally, an argumentation scheme (AS), encapsulated as a $F-node$, is composed of a set of premises A_i, a conclusion C, and a set of critical questions CQ_i, which enumerate specific ways to defeat the derivation of the consequent. One desiderata of the ASs is to simplify the argumentation process by hiding secondary premises and encapsulating them as CQs (figure 1). Based on the main premises A_1, A_2, and A_3 the consequent C is defeasibly inferred. During the process of gradually re-

978-1-4244-2673-7/08/$25.00 ©2008 IEEE

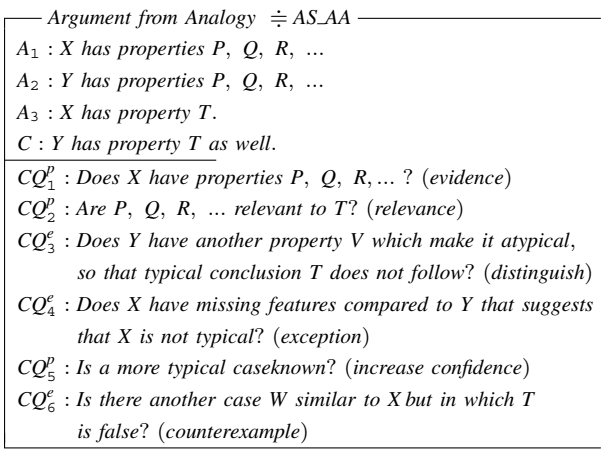

Figure 1. Simplifying argumentation.

vealing information in a debate, the conclusion might be retracted in the light of new information. Each *AS* sustaining a claim provides the correspondent *CQ*s that the opponent may use to challange the pleading. When a *CQ* is conveyed, the conclusion is suspended, until the subject of the dispute is clarified. Whoever is responsible for this clarification, in other words who has the burden of proof, depends on the type of the *CQ* [1].

Definition. *A presumption-type CQ^p is a hidden premise encapsulated as a CQ, at the instantiation of which the burden of proof is shifted to the proponent of the argument. An exception-type CQ^e is a hidden premise encapsulated as a CQ, at the instantiation of which the burden of proof remains to the opponent.*

Enacting the *Argument from Analogy* scheme as a $F-node$ is depicted in figure 2. Based on the activation of a $CA-node$, the derivation of the conclusion is blocked until the subject is clarified. The activation of an exception leads to the instantiation of an *AS* that supports that exception. The state variables of the dialogue are encapsulated as $CO-nodes$.

The presumptions CQ^p_1, CQ^p_2, CQ^p_5 request from the opponent more justification for suporting the claim, while the exceptions CQ^e_3, CQ^e_4, CQ^e_6 challange the argument by attacking the link between the premises and the conclusion. By instantiating the CQ^p_1, the opponent requests evidence that the case *X* is actually characterized by the mentioned properties *P*, *Q*, *R*. The burden of proof is shifted back to the proponent of argument who has to provide evidence supporting its claim. The CQ^p_2 represents an undercutting defeater by attacking the link between the premises and the conclusion: the premises are not relevant enough to infer the consequent. When the CQ^p_5 is uttered, the opponent does not challenge the arguments provided by the proponent; it accepts them, but they are not considered sufficient for the consequent to be accepted. This *CQ* sugests to the proponent to engage in a persuation dialogue, where the accrual of different arguments supporting the same conclusion might hold. In case of an exception-type *CQ*, the oponent identifies evidence that *Y* is an exception (CQ^e_3), *X* is an exception (CQ^e_4), or there have been similar situations where the same premises lead to the opposite conclusion (CQ^e_6), the opponent being the one responsible for providing data to distinguish between the case *X* and *Y*.

3 Learning from Experience

Learning how to argue is based on two ideas: i) if an argument in a given situation causes something bad to happen (at the end of the game the argument is not accepted by the opponent), the agent must learn not to convey that argument in a similar context; and ii) if all arguments in a given situation cause something bad to happen, the agent should avoid utterances that would lead the dialogue in that state. The *RL* model for argumentation consists of a set of environment states s_i, a set of actions a_i, and a set of rewards r_i.

The *states* of the dialogue game are defined as a collection of state variables, encapsulated as $CO-nodes$, which include all the information that is relevant in determining what *AS* to be conveyed next. The *actions* represent a finite number of speech acts. The agent learns which locution is best to be conveyed in a given context. The communication language L_c consists of a set of *AS*s used when arguing from experience: analogical argument, argument from example, argument from generalization, argument from correlation to cause, argument from common cause [6].

The *reward function* models the goal of the proponent, representing the cost of the argumentation game defined in terms of (i) the cost to obtain information O_I (the cost to populate *AS*s by querying the dataset); (ii) the cost of information disclosure R_I (cost of sharing its own experiences in order to alter the opponent's cognitive state); and (iii) the solution cost (defined either in terms of time left and opportunity cost in case the persuasion fails, or in terms of reward or expected profit if the argument is accepted). According to the $\epsilon-greedy\ learning\ policy$, the agent must decide between chosing an *AS* that is most likely to influence the dispute, according to past experience (*exploit step*), or to instantiate a scheme or a *CQ* that stimulates discussion and extracts information (*exploration step*).

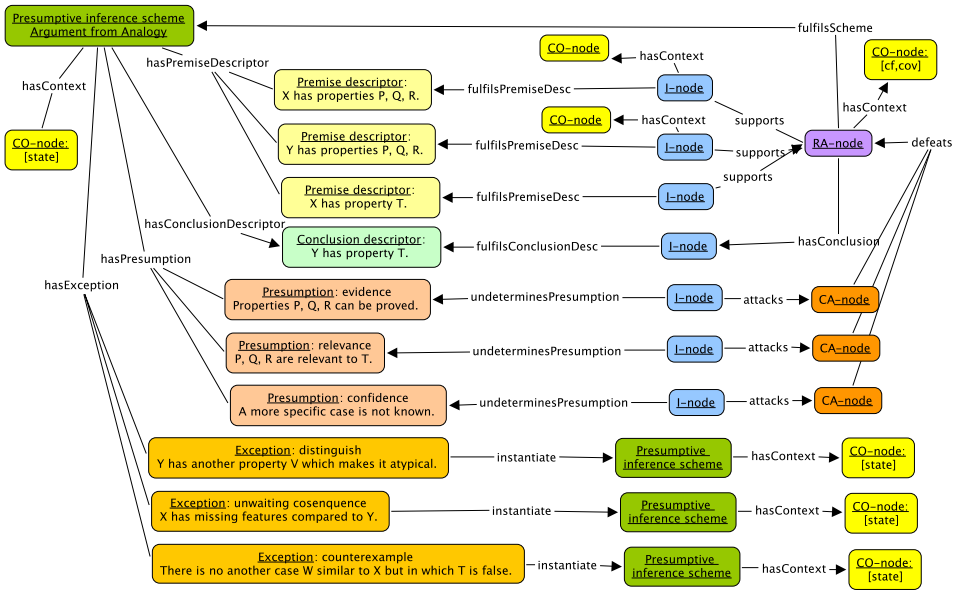

Figure 2. Enacting argumentation schemes in AIF ontology.

4 Models of RL-based Argumentation

The running scenario illustrates a car sales agent for online persuasion of potential clients. To mimic different types of human agents, the seller and buyers have different perspectives and amount of experience related to the car domain, usually not balanced, meaning that the available dataset[1] is distributed among agents in different percentage and may be redundant. The agents extract a set of association rules that can be used to instantiate the pre-defined *AS*s. The buyers accept arguments based on an *accept function*: the argument must fit the cognitive state of the agent within a threshold τ. If an instantiation of the *analogy AS* is characterised by the confidence cf_p and the coverage cov_p conveyed by the proponent, we define $\Delta cf = 0$, when $cf_p \leq cf_o$, otherwise $\Delta cf = cf_p - cf_o$, and $\Delta cov = 0$, when $cov_p \leq cov_o$, otherwise $\Delta cov = cov_p - cov_o$, where cf_o and cov_o represent the certainty factor and coverage of the current argument from the opponent viewpoint. An argument is accepted if $w\Delta cf + (1-w)\Delta cov \leq \tau$, where w measures the importance the buyer is giving to the accuracy and to the coverage.

Yes/No feedback. Consider the seller and the buyer's theories include the following association rules: $\prec r_1^s(0.9, 0.2) : a \rightarrow acc, r_2^s(0.8, 0.3) : a, b \rightarrow acc \succ$, and respectively $\prec r_1^b(0.7, 0.3) : a \rightarrow acc, r_2^b(0.6, 0.2) : a, b \rightarrow acc \succ$, where the pair of values represent the *cf* and the *cov*, and the consequent *acc* means that the buyer accepts to buy the car. The action set of the seller is: $L_c = \{AS\_AA, share\_example, provide\_fact\}$. The first model considers that the following state variables are relevant: $State = \{P, Q, R, T, cf, cov\}$. The parameters of the reward function are: $Q_t = 0.1$, representing the cost of instantiating a new *analogy AS*; $Q_i = 0.3$, the cost of sharing a past experience with the buyer, while the wasted time receives a penalty of -1; otherwise he gets a reward of 10. the time spent receives a penalty of -1 if the buyer does not accept the argument, respectively a reward of 10 otherwise. The accept function is computed with $w = 0.5$ and $\tau = 0.1$.

No.	P	Q	R	T	cf	cov	Action	F	R
1	a	-	-	acc	0,7	0.2	AS_AA	NO	-1.1
2	a	b	-	acc	0.5	0.3	AS_AA	NO	-1.1
3	a	b	-	acc	0.5	0.3	share	YES	9.7

Table 1. Trace in a dialogue run.

In the first move of the dialogue (table 1), the seller rises the argument $r_1^s(0.9, 0.2) : a \rightarrow acc$. Based on the cognitive state of the buyer, because the coverage is satisfied ($cov_o = 0.3 \geq cov_p = 0.2$), his acceptance function depends only on the certainty factor: $A_1 = cf_p - cf_o = 0.9 - 0.7 = 0.2 > \tau$. Therefore, the argument is not accepted. In this state, the received reward is given by the cost of instantiating the *AS* and time spent

[1] Autos dataset from machine learning repository: www.ics.uci.edu/ mlearn/

so far: $R_1 = -0.1 - 1 = -1.1$. In the next step the seller conveys another argument $r_2^s(0.8, 0.3) : a, b \rightarrow acc$. The acceptance value being $A_2 = 0.5(0.8 - 0.6) + 0.5(0.3 - 0.2) = 0.15 > \tau$, the argument is not accepted. The resulting penalty is the same as in the previous step $R_2 = -1.1$. In the step 3, the seller keeps the same association rule as argument, but he decides to alter the cognitive state of the buyer by providing a past experience. If the buyer trusts the example, he incorporates it in his theory. The association rule r_2^b will be altered, resulting $r_{22}^b(0.69, 0.21) : a, b \rightarrow acc$. The accept value being $A_3 = 0.5(0.8 - 0.69) + 0.5(0.3 - 0.22) = 0.095 < \tau$, the buyer accepts the argument r_2^s. Thus, the received reward in this state is $R_3 = -0.3 + 10 = 9.7$. The reward represents the immediate value of an action in a given state. After a number of dialogue games, the long running value of each state will be estimated based on:

$$Q_n(s,a) = (1-\alpha_n)Q_{n-1}(s,a) + \alpha_n[r + \gamma max_{a'} Q_{n-1}(s',a')]$$

where $\alpha_n = 1/(1 + visits_n(s,a))$ and $visits(s,a)$ is the number of visits that the state-action pair has received so far. The learning rule considers a weighted average of the current Q value and the revised estimate in order to avoid the repeated alternations of the values of Q in the nondeterministic case[2]. Thus, the revision of the Q value is made more smoothly, updates becoming smaller as training takes place [3].

CQs-based feedback. One advantage of *RL* is that it can accomodate different levels of feedback amount. For instance, the buyer can share its viewpoint in terms of its own certainty factor cf_o and coverage cov_o. If this is the case, it is much easier for the seller to decide which argumentation policy is better to follow. If the provided values are close to their own, he anticipates if sharing an example impacts on the cognitive state of the buyer at the required level. Otherwise, the seller may decide to change the argument type and to approach the buyer from a different perspective. The argumentation process is a trajectory in the chosen state space driven by the conveyed *CQ*s and *AS*s. When planning for this trajectory, the first idea is to propose rules having the subject of debate as consequent of the rule. If this fails, the seller has to get away from the goal by starting to prove the premises of the above rules. The *RL* provides agents exactly with this mechanism: *learning when it is better to get away from the goal*. For instance, if the CQ_1^p is uttered, the seller can start to prove the requested properties or it can instantiate another rule having the same initial consequent. The main drawback is given by the number of interactions needed for the agent to learn a persuasion strategy.

[2]To identical pairs of state and arguments, buyers react in different ways.

5 Related Work and Conclusion

Synergy of argumentation and machine learning is an open research issue. The work in [2] adopts *RL* to adjust negotiation strategies for a sales agent. The persuasion dialogue is based on Dung's abstract argumentation framework and the status of arguments according to the dialectical approach of Vreeswijk and Prakken. By using *AS*s, our work is more oriented towards the interaction with human agents. Combining argumentation and datamining has been investigated in [7]. Rising arguments is based on PADUA (Protocol for Argumentation Dialogue Using Association Rules), designed to empower agents with the ability to extract arguments from a dataset. In our approach, the argumentation protocol is directly encapsulated as *CQ*s.

The current research[3] can be integrated into the area of *action language perspective*, in the larger context of newly arising *Pragmatic Web* paradigm. If the WWAW is be succesful, we anticipate an explosion of public structured arguments, available on the Internet, that can be used in the learning process to better estimate the Q-value of each *AS*. The strategy will be shifted towards re-using arguments in different contexts and learning from others' experiences.

References

[1] T. F. Gordon, H. Prakken, and D. Walton. The Carneades model of argument and burden of proof. *Artificial Intelligence*, 171(10-15):875–896, 2007.

[2] S. Huang and F. Lin. The design and evaluation of an intelligent sales agent for online persuasion and negotiation. *Electron. Comm. Rec. Appl.*, 6(3):285–296, 2007.

[3] T. M. Mitchell. *Machine Learning*. McGraw-Hill, 1997.

[4] S. Onta and E. Plaza. Learning and joint deliberation through argumentation in multiagent systems. In *AAMAS '07*, pages 1–8, New York, NY, USA, 2007. ACM.

[5] I. Rahwan, F. Zablith, and C. Reed. Laying the foundations for a World Wide Argument Web. *Artificial Intelligence*, 171(10-15):897–921, 2007.

[6] C. Reed and G. Rowe. Araucaria: Software for argument analysis, diagramming and representation. *Intern. J. on Artif. Intell. Tools*, 13(4):961–979, 2004.

[7] M. Wardeh, T. J. M. Bench-Capon, and F. Coenen. PADUA protocol: Strategies and tactics. In K. Mellouli, editor, *ECSQARU*, volume 4724 of *LNCS*, pages 465–476. Springer, 2007.

[3]This work was supported by the grant TD7 CNCSIS 534 from the National Research Council of the Romanian Ministry for Education and Research.