Representing Abductive Practical Reasoning as an Action-Based Alternating Transition System

Floris BEX [a], Katie ATKINSON [b,1] and Trevor BENCH-CAPON [b]
[a] *Faculty of Law, University of Groningen, The Netherlands.*
[b] *Department of Computer Science, University of Liverpool, UK.*

Abstract. In this report we present an approach to abductive reasoning by examining it in the context of an argumentation scheme for practical reasoning. We present a particular scheme, based on one for practical reasoning, that can be used to reason abductively about how an agent might have acted to find itself in a particular sceanrio, and its motivations for doing so. We provide a formal representation of the scheme, and its associated critical questions, in terms of Action-based Alternating Transition Systems to allow for the automatic generation of arguments.

1. An Argumentation Scheme for Abductive Practical Reasoning

In [3] a formalism is presented to describe practical reasoning in terms of an Action-based Alternating Transition System (AATS). The starting point is a previously specified account of practical reasoning that treats reasoning about what action should be chosen as presumptive argumentation using an argumentation scheme and associated critical questions defined in [2]. The underlying formalism used to ground the representation is one given in [6] in which Wooldridge and van der Hoek define a normative system in terms of constraints on actions that may be performed by agents in any given state. This underlying normative system serves as the basis for the representation of arguments about action in [3]. The formalism provides a well-specified basis for addressing the problems of practical reasoning as presumptive argumentation in a multi-agent context. In this report we examine how abductive reasoning – reasoning to the best explanation – can be represented as a species of practical reasoning using an AATS. We do this by first providing an informal description of an argument scheme and critical questions for abductive practical reasoning, then we show how this can be represented in terms of an AATS to allow for the automation of the reasoning involved.

1.1. Informal Description

Argument scheme for abductive practical reasoning:

[1]Corresponding Author: Department of Computer Science, University of Liverpool, UK. E-mail: katie@liverpool.ac.uk

The current circumstances S
are explained by the performance of action A
in the previous circumstances R
with motive M

We now turn to the critical questions for the abductive scheme. Below each critical question, the answer to that question which would attack the original argument is given. We distinguish between two types of critical questions: questions pertaining to the problem formulation phase and questions that pertain to the phase where the different explanations are chosen. In the problem formulation phase the propositions and motivations relevant to the particular situation are identified, and the AATS is constructed. In the choice of explanation phase, the appropriate arguments and counter arguments are developed, in terms of applications of the argument scheme and critical questions, and the status of the arguments is determined with respect to other arguments and the orderings of the motivations. These stages may be carried out sequentially, or they may iterate if the critical questioning leads to a reformulation of the problem. The critical questions associated with the scheme are as follows.

Critical questions for choice of explanation:

CQ1 Are there alternative ways of explaining the current circumstances S?
a) Could the preceding state R have been different?
answer: action A was done in a different preceding state R
b) Could the action A have been different?
answer: a different action A′ was done in preceding state R

CQ2 Assuming the explanation, is there something which takes away the motivation?
answer: doing action A in R to reach S demotivates M

CQ3 Assuming the explanation, is there another motivation which is a deterrent for doing the action?
answer: some other motivation M′ deters from doing action A in R to reach S

CQ4 Can the current explanation be induced by some other motive?
answer: there is another motivation M′ which motivated doing A in R to reach S

CQ5 Assuming the previous circumstances R, was one of the participants in the joint action trying to reach a different state?
answer: in R, even though one agent performed his part of A with motive M, the joint action was actually A′ which led to S′, where $A' \neq A$ and $S' \neq S$

Critical questions for problem formulation:

CQ6 Are the current circumstances true?
answer: the current state is not S

CQ7 Could the action have had the stated preconditions?
answer: A cannot be performed in R

CQ8 Were the previous circumstances the same as the current circumstances?
answer: for all propositions in S and R:
if a proposition p is true in S then p_a was already true R
if a proposition p is false in S then p_a was already false R

CQ9 Could the explanation for the current state provide the motivation?
answer: doing action A cannot be motivated by M

CQ10 Assuming the previous circumstances, would the action have the stated consequences?
answer: doing action A in in R does not bring about S

CQ11 Assuming the previous circumstances, would the action have any consequences?
answer: doing action A in R does not get you to a new state

CQ12 Are the current circumstances S possible?
answer: there is no state S (S is impossible)

CQ13 Is the joint action possible?
answer: A is not a joint action

CQ14 Are the previous circumstances R possible?
answer: there is no state R (R is impossible)

CQ15 Is the motivation indeed a legitimate motivation?
answer: M is not a motivation

2. Action-Based Alternating Transition systems

In order to be able to reason rigorously about actions and their effects, we need a well-defined structure in which we can represent how the actions of an agent will lead to transitions from one state to another. In particular we need to be able to contextualise these transitions so that the effects of actions can be made dependent on the action of other agents, and other events in the environment. One such structure is provided by Alternating Transition Systems (ATS), originally developed to underpin the Alternating-time Temporal Logic of [1]. These structures have also been used by van der Hoek et al. [6] to explore the social laws paradigm for describing coordination in multi-agent systems introduced largely through the work of Shoham, Tennenholtz and Moses (e.g. [5]). Like [6] we give the notions of actions and their pre-conditions a central role, so we adopt their version of ATS in which actions and pre-conditions are first class entities. This version is called an *Action Based Alternating Transition Systems* (AATS) in [6], and it has been used in [3] to provide formal definitions for an argument scheme and critical questions for practical reasoning. In this report we also use this structure to represent our argument scheme and critical questions for abductive practical reasoning. We first provide the definition of an AATS, as given in [6].

Assume first that the systems we wish to model may be in any of a finite set Q of possible *states*, with some $q_0 \in Q$ designated as the *initial state*. Systems contain a set Ag of *agents* and each agent $i \in Ag$ is associated with a set Ac_i of possible actions. It is assumed that these sets of actions are pairwise disjoint (i.e., actions are unique to agents).

A joint action j_C for set of agents C (termed a *coalition* is a tuple $\langle \alpha_1, ..., \alpha_k \rangle$, where for each α_j (where $j \leq k$) there is some $i \in C$ such that $\alpha_j \in Ac_i$. Moreover, there are no two different actions α_j and $\alpha_{j'}$ in j_C that belong to the same Ac_i. The set of all joint actions for coalition C is denoted by J_C, so $J_C = \prod_{i \in C} Ac_i$. Given an element j of J_C and an agent $i \in C$, i's action in j is denoted by j_i.

An *Action-based Alternating Transition System* (AATS) is an $(n + 7)$-tuple $S = \langle Q, q_0, Ag, Ac_1, ... , Ac_n, \rho, \tau, \Phi, \pi \rangle$, where:

- Q is a finite, non-empty set of *states*;

- $q_0 \in Q$ is the *initial state*;
- $Ag = \{1,...,n\}$ is a finite, non-empty set of *agents*;
- Ac_i is a finite, non-empty set of actions, for each $i \in Ag$ where $Ac_i \cap Ac_j = \emptyset$ for all $i \neq j \in Ag$;
- $\rho : Ac_{Ag} \rightarrow 2^Q$ is an *action pre-condition function*, which for each action $\alpha \in Ac_{Ag}$ defines the set of states $\rho(\alpha)$ from which α may be executed;
- $\tau : Q \times J_{Ag} \rightarrow Q$ is a partial *system transition function*, which defines the state $\tau(q, j)$ that would result by the performance of j from state q - note that, as this function is partial, not all joint actions are possible in all states (cf. the pre-condition function above);
- Φ is a finite, non-empty set of *atomic propositions*; and
- $\pi : Q \rightarrow 2^\Phi$ is an interpretation function, which gives the set of primitive propositions satisfied in each state: if $p \in \pi(q)$, then this means that the propositional variable p is satisfied (equivalently, true) in state q.

In addition to the elements of an AATS given in [6], we need to provide an extension to enable the representation of motivations from the underlying argument scheme for abductive practical reasoning. Firstly, we have a set *Am* of motivations for each agent (which are a subset of a set *M* of motivations). Every transition between two states from the set *Q* is either promoted, demoted, or is neutral, with respect to each motivation. Note that motivations are not unique to agents: individual agents may or may not have motivations in common. Whether a motivations is promoted or demoted by a given action will be determined by comparing the state reached with the state left. More formal definitions of these elements are given below:

- Am_i is a finite, non-empty set of motivations $Am_i \subseteq M$, for each $i \in Ag$.
- $\delta : Q \times Q \times Av_{Ag} \rightarrow \{+, -, =\}$ is a *valuation function* which defines the status (promoted (+), demoted (–) or neutral (=)) of a motivation $m_u \in Am_{Ag}$ ascribed by the agent to the transition between two states: $\delta(q_x, q_y, m_u)$ labels the transition between q_x and q_y with one of $\{+, -, =\}$ with respect to the motivation $m_u \in Am_{Ag}$.

We can now extend the original specification of an AATS to accommodate the notion of motivations and re-define an AATS as a $(2n + 8)$ tuple $S = \langle Q, q_0, Ag, Ac_1, ... , Ac_n, Am_1, ..., Am_n, \rho, \tau, \Phi, \pi, \delta\rangle$

3. Formal Definitions for the Argument Scheme for Abductive Practical Reasoning

We now present the formal definitions of the argument scheme and critical questions in terms of an AATS. The critical questions can be grouped into two categories: those concerned with choice explanation, and those concerned with problem formulation. We present the formal definitions of the critical questions as grouped into these categories. We begin by presenting the formal version of the argument scheme:

ABS1: The current circumstances $q_0 = q_y$
are explained by agent i participating in joint action j_n where ${j_n}^i = \alpha_i$,
in the previous circumstances q_x, where $\tau(q_x, j_n)$ is q_y
and $\exists p_a \in \Phi$

such that either $p_a \in \pi(q_y)$ and $p_a \notin \pi(q_x)$, or $p_a \notin \pi(q_y)$ and $p_a \in \pi(q_x)$
such that for some $m_u \in M$, $\delta(q_x, q_y, m_u)$ is +.

We now present the formal version of the critical questions that can be used to challenge instantiations of the above argument scheme.

3.1. *Critical Questions for Choice of Explanation*

CQ1a: The previous circumstances were not q_x and were actually $q_z \in Q$, in which agent $i \in Ag$ could have participated in joint action $j_n \in J_{Ag}$, such that $\tau(q_z, j_n)$ is q_y.

CQ1b: In the previous circumstances $q_x \in Q$, agent $i \in Ag$ could have participated in joint action $j_m \in J_{Ag}$, where $j_n \neq j_m$, such that $\tau(q_x, j_m)$ is q_y.

CQ2: There is a p_b, where $p_a \neq p_b$, such that either $p_b \in \pi(q_y)$ and $p_b \notin \pi(q_x)$, or $p_b \notin \pi(q_y)$ and $p_b \in \pi(q_x)$, such that $\delta(q_x, q_y, m_u)$ is –.

CQ3: There is a p_b, where $p_a \neq p_b$, such that either $p_b \in \pi(q_y)$ and $p_b \notin \pi(q_x)$, or $p_b \notin \pi(q_y)$ and $p_b \in \pi(q_x)$, such that $\delta(q_x, q_y, m_w)$ is –, where $m_u \neq m_w$.

CQ4: There is a p_b, where $p_a \neq p_b$, such that either $p_b \in \pi(q_y)$ and $p_b \notin \pi(q_x)$, or $p_b \notin \pi(q_y)$ and $p_b \in \pi(q_x)$, such that $\delta(q_x, q_y, m_w)$ is +, where $m_u \neq m_w$.

CQ5: $j_n{}^i = j_m{}^i$, $j_n \neq j_m$ and $\tau(q_x, j_n) \neq \tau(q_x, j_m)$.

3.2. *Critical Questions for Problem Formulation*

CQ6: $q_0 \neq q_y$.

CQ7: $q_x \notin \rho(\alpha_i)$.

CQ8: $\forall\, p_j \in \Phi$, $p_j \in \pi(q_y)$ and $p_j \in \pi(q_x)$, or $p_j \notin \pi(q_y)$ and $p_j \notin \pi(q_x)$.

CQ9: $\delta(q_x, q_y, m_u)$ is not +.

CQ10: $\tau(q_x, j_n)$ is not q_y.

CQ11: $\tau(q_x, j_n)$ is q_x.

CQ12: $q_y \notin Q$.

CQ13: $j_n \notin J_{Ag}$.

CQ14: $q_x \notin Q$.

CQ15: $m_u \notin Am_i$.

The above formalism can be used by agents in scenarios where reasoning takes place to generate arguments to explain how an agent may have acted to find itself in a particular situation and its motivations for doing so. A worked example demonstrating such a scenario is set out in [4].

References

[1] R. Alur, T. A. Henzinger, and O. Kupferman. Alternating-time temporal logic. *ACM Journal*, 49 (5):672–713, 2002.

[2] K. Atkinson. *What Should We Do?: Computational Representation of Persuasive Argument in Practical Reasoning*. PhD thesis, Department of Computer Science, University of Liverpool, Liverpool, UK, 2005. Available online.

[3] K. Atkinson and T. J. M. Bench-Capon. Practical reasoning as presumptive argumentation using action based alternating transition systems. *Artificial Intelligence*, 171(10–15):855–874, 2007.

[4] F. J. Bex, T. J. M. Bench-Capon, and K. Atkinson. Did he jump or was he pushed? abductive practical reasoning. In *Legal Knowledge and Information Systems. JURIX 2008: The Twenty First Annual Conference*, 2008. *Under review*.

[5] Y. Moses and M. Tennenholtz. Artificial social systems. *Computers and AI*, 14(6):533Ű–562, 1995.

[6] M. Wooldridge and W. van der Hoek. On obligations and normative ability: Towards a logical analysis of the social contract. *Journal of Applied Logic*, 3:396–420, 2005.