

Abstract Argumentation for Hybrid Intelligence Scenarios

Loan Ho¹, Victor de Boer¹, M. Birna van Riemsdijk², Stefan Schlobach¹ and Myrthe L. Tielman³

¹Vrije Universiteit Amsterdam, The Netherlands

²University of Twente, The Netherlands

³Interactive Intelligence Group, TU Delft, The Netherlands

Abstract

Hybrid Intelligence (HI) is the combination of human and machine intelligence, expanding human intellect instead of replacing it. Information in HI scenarios is often inconsistent, e.g. due to shifting preferences, user's motivation or conflicts arising from merged data. As it provides an intuitive mechanism for reasoning with conflicting information, with natural explanations that are understandable to humans, our hypothesis is that Dung's *Abstract Argumentation (AA)* is a suitable formalism for such hybrid scenarios. This paper investigates the capabilities of Argumentation in representing and reasoning in the presence of inconsistency, and its potential for intuitive explainability to link between artificial and human actors. To this end, we conduct a survey among a number of research projects of the Hybrid Intelligence Centre¹. Within these projects we analyse the applicability of argumentation with respect to various inconsistency types stemming, for instance, from commonsense reasoning, decision making, and negotiation. The results show that 14 out of the 21 projects have to deal with inconsistent information. In half of those scenarios, the knowledge models come with natural preference relations over the information. We show that Argumentation is a suitable framework to model the specific knowledge in 10 out of 14 projects, thus indicating the potential of Abstract Argumentation for transparently dealing with inconsistencies in Hybrid Intelligence systems.

Keywords

Hybrid Intelligence, Argumentation, Explainability, Inconsistency, Preferences,

1. Introduction

Artificial Intelligence (AI) is being applied in a variety of real-life situations. In recent years, AI applications are starting to go beyond machine reasoning by creating what is now called *Hybrid Intelligence (HI)* systems which combine human and artificial intelligence, and attempt to *integrate* human and machines rather than use AI to replace human intelligence [1]. The idea is that artificial and human agents collaborate in complex, and often dynamic, environments. For example, preferences can shift, user's motivation or external conditions (available resources and environment) can vary over time and in different contexts. Also, in many cases data might

¹<https://www.hybrid-intelligence-centre.nl/>

1st International Workshop on Argumentation for eXplainable AI (ArgXAI, co-located with COMMA '22), September 12, 2022, Cardiff, UK

✉ loanthuyho.cs@gmail.com (L. Ho); v.de.boer@vu.nl (V. de Boer); m.b.vanriemsdijk@utwente.nl (M. B. van Riemsdijk); k.s.schlobach@vu.nl (S. Schlobach); M.L.Tielman@tudelft.nl (M. L. Tielman)



© 2022 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

be integrated from different, heterogeneous, sources. In such environments, it is likely that the information available to, and about, human and artificial agents is conflicting. This is problematic, as it is well known that classical logical approaches to reasoning fail when dealing with inconsistency so that errors in the data, or conflicting information, lead to undesired decisions or predictions. Explanation and handling of inconsistent dynamic information has thus become an important challenge for Knowledge Representation and Reasoning in such Hybrid Intelligence environments.

Recently, several works have focused on inconsistency handling [2, 3, 4, 5]. Early solutions to this problem have been developed w.r.t databases, where the areas of *database repairing* and *consistent query answering* (CQA) have gained much attention [2]. Database repairing provides the model-theoretic construct of a *database repair* –a consistent database that “minimally” differs from the original (inconsistent) database instance–, while CQA yields the set of tuples (atoms) that appear in the answer to the query over every possible repair. In [3], the author summarizes various approaches for handling inconsistent data in Ontology-Mediated Query Answering (OMQA), which adapt and extend techniques initially proposed for databases. The approach focuses on inconsistency handling, where inconsistencies are due to errors in the data (i.e., we assume the ontology has been properly debugged) and mainly discuss how inconsistency-tolerant semantics can be used to obtain meaningful information from inconsistent knowledge bases (KBs). One such approach to inconsistency handling uses *belief revision*, which proposes AGM¹ axioms to revise the KB with the main goal of preserving consistency [4]. In order to deal implicitly with inconsistencies, Tielman et. al presented a method to derive specific norms for behavior from the information on actions, values and context [5]. However, these approaches still lack transparency because of not providing explanations for users to understand *why* an event started, or what led to decisions, predictions or query answers.

Abstract Argumentation², as introduced by Dung [6], has become an important paradigm for Knowledge Representation. It is especially useful for reasoning with contradictory information, for formalizing the argument exchange between agents in, e.g., negotiation and for commonsense reasoning, logic programming, legal reasoning and decision making. The advantages of Argumentation are that it can suitably represent anything from input data, e.g. categorical data and pixels in an image, to knowledge, e.g. rules, to problem formalisations, e.g. planning, scheduling or decision making models, to outputs, e.g. classifications, recommendations, or logical inference [7, 8]. Argumentation also has strong *explainability* capabilities that allow users to understand the rationale of a decision, predictions or query answers [9]. This flexibility and wide-ranging applicability has led to a multitude of methods for the application of argumentation in AI systems. From the arguments above, it is obvious that we consider argumentation to be a suitable formalism for Hybrid Intelligence. This paper investigates this hypothesis in a more systematic way.

Modeling and reasoning in inconsistent knowledge with respect to HI systems is challenging, especially in the context of human interaction, dynamic knowledge and preferences. Moreover, to what extent providing argumentation-based explanations make an HI system more transparent and more trustworthy to users has not yet been thoroughly investigated [10]. This study provides a first overview over KRR formalisms in HI scenarios, and the potential role of Argumentation,

¹Named after the authors of Alchourron, Gardenfors, and Makinson.

²We will often simply refer to the formal framework of Abstract Argumentation as "Argumentation".

in the presence of inconsistencies.

The main contributions of this paper are as follows:

- We present results from a **survey** among several sub-projects of the Hybrid Intelligence Centre (<https://www.hybrid-intelligence-centre.nl/>). Based on using qualitative data analysis methods, for each scenario, we show whether inconsistencies exist in the HI scenarios and explain the reasons leading to the inconsistencies.
- We **demonstrate the capabilities** of Argumentation in representation and reasoning inconsistent KB of HI scenarios. For this purpose, based on analysis results from the survey, we study whether the representation of an inconsistent KB within a HI scenario can be mapped into the Argumentation Framework (AF). Particularly, for each scenario, we show that arguments represent facts of KB and attack relations between the arguments capture conflicting information. Argumentation trees can then be presented as *dialectical dispute trees*, which provides the user with dialogues of explanations to better understand the rationale of a decision, predictions or query answers.
- We show how Argumentation can **enable explainability** in the HI systems, for solving various types of problems in decision-making, justification of an opinion, and dialogues.

2. Preliminaries

2.1. Argumentation overview

We shall briefly recall the Abstract Argumentation Framework (AAF) (Dung [6]).

Abstract Argumentation Framework. An abstract argumentation is a pair $\mathcal{A} = (Arg, Att)$ where Arg is the set of arguments, and $Att \subseteq Arg \times Arg$ is the *attacking relation* i.e., $(A, B) \in Att$ means A attacks B . Let $\mathcal{M} \subseteq Arg$, we say that: (1) \mathcal{M} attacks an argument A if there exists an $A' \in \mathcal{M}$ s.t. $(A', A) \in Att$. (2) \mathcal{M} defends an argument A if \mathcal{M} attacks every argument attacking A . \mathcal{M} is *conflict-free* if there are no arguments $A, A' \in \mathcal{M}$ s.t. $(A, A') \in Att$.

Extensions. Semantics of AAF are specific subsets of arguments, which are defined from the aforementioned properties. Let $\mathcal{A} = (Arg, Att)$ be an AAF and $\mathcal{M} \subseteq Arg$. We say that: (1) \mathcal{M} is an *admissible extension* if \mathcal{M} is conflict-free and defends each argument in it. (2) \mathcal{M} is a *complete extension (cmp)* if \mathcal{M} is an admissible extension containing all arguments that it defends. (3) \mathcal{M} is a *preferred extension (prf)* if \mathcal{M} is a maximal (w.r.t. set inclusion) admissible extension. (4) \mathcal{M} is a *stable extension (stb)* if \mathcal{M} is a conflict-free and attacks every argument which is not in it. (5) \mathcal{M} is a *grounded extension (grd)* if \mathcal{M} is a minimal (w.r.t. set inclusion) complete extension.

Acceptance. In order to evaluate the arguments, two types of acceptance are introduced in terms of their extensions: the *sceptical* and *credulous* acceptance. We say that an argument A is *sceptically (credulously)* accepted w.r.t. a semantic S iff it is in all extensions (at least one extension) under S .

In AAF, any information which may be in dialectical relationships of disagreement (attack) with other information may be considered to be an argument, and arguments (according to this loose interpretation of the term) typically have a negative or positive impact on the acceptability of arguments they attack. In this spirit, we recall the notion of the argumentation tree introduced

in [11]. An *argumentation tree* called a dispute tree is a description of how these arguments are defended or attacked. Based on argumentation trees, a dialogical process of explanation is described as follows:

Dialogical explanation: A dialogical process of explanation is a two-person argument game between a proponent and an opponent. The proponent and the opponent are engaged in an argumentation dialogue of a sequence of moves. The dialogue starts by the proponent with an initial argument. Then, the opponent presents an argument (or a set of arguments) that attacks the initial argument of the proponent. Next, the proponent tries to avoid this attack and reinstate the query by using another argument which is not attacked by the opponent. The opponent tries to extend the previous set of attackers so that it attacks all the initial arguments advanced so far. When the opponent fails to extend the set, it retraces back and chooses another set of attackers and continues the dialogue from thereafter. By doing so, the opponent is trying to construct a set of arguments that attacks all the initial argument.

We recall the notion of *Structured Argumentation Framework (SAF)* as an extension of Abstract Argumentation Framework, which is proposed [12]. SAF represents the arguments in the form of logical rules and the attack relations capture the contrary information between the arguments. First, we need to define some concepts. A KB $\mathcal{K} = \{\mathcal{R}, \mathcal{C}, \mathcal{F}\}$ where \mathcal{F} is a set of facts, \mathcal{C} is set of negative constraints, and \mathcal{R} is a set of rules which follow the general form: $R : a \leftarrow b_1, \dots, b_n$ (claim \leftarrow premise).

Structured Argumentation Framework. Let $\mathcal{K} = (\mathcal{F}, \mathcal{R}, \mathcal{C})$ be a KB. The corresponding SAF is the pair (Arg', Att') such that an argument $A \in Arg'$ is a tuple (H, C) with H a non-empty \mathcal{R} -consistent subset of \mathcal{F} and C a set of facts s.t. (1) $C \subseteq SAT(H)$ and (2) there is no $H_0 \subset H$ s.t. $C \subseteq SAT(H_0)$. The support H of an argument A is denoted by $Supp(A)$ and the conclusion C by $Conc(A)$. A attacks B , denoted by $(A, B) \in Att'$ iff there exists $\alpha \in Supp(B)$ s.t. $Conc(A) \cup \{\alpha\}$ is \mathcal{R} -inconsistent, where $SAT(H)$ is called the saturation of a set of facts H .³

2.2. State of the art in Argumentation

Argumentation is becoming one of the main mechanisms for solving reasoning problems with conflicting information. Calegari et al. sketch a vision of Explainability of intelligent systems, in which they show how argumentation is suitable for explaining agent intelligent behaviours in the domain of computable law for autonomous vehicles[14]. In some recent works, authors place themselves in various argumentation frameworks to provide a useful platform for representing and reasoning with maximally consistent subsets of KBs in propositional logic [15, 16] and in inconsistent ontological KBs [12, 17]. Other argumentation-based approaches recently have centered on formalizing *legal reasoning*, *commonsense reasoning*, *decision-making* and the exchange of arguments between agents in *negotiation*. Prakken et al. provide a formalization of legal reasoning with cases into Argumentation [18]. An approach to empower *commonsense reasoning* and make it more explainable with Argumentation is given in [19]. Botschen et al. investigate the use of whether external knowledge of event-based frames and fact-based entities can contribute to decomposing an argument as stated in the Abstraction and Reasoning Corpus (ARC) task. Using Argumentation-Based Dialogues (ABD) to explain an opinion can be a method

³According to [13], the saturation of a set of facts \mathcal{F} by \mathcal{R} is the set of all possible atoms and conjunctions of atoms that are entailed, after using all rule applications from \mathcal{R} over \mathcal{F} until a fixed point.

of providing an explanation [20]. In the context where people or agents in a dialogue have an ostensible purpose, but their own goals or the goals of the other participants may not be consistent with this purpose. The proposed negotiation protocols using ABD allow the agents to perform negotiations to find the winning participant and to explain in more detail how the winner reached a decision. Argumentation is highly related to *decision-making*. Several works with applications in recommendation systems (RSs) make use of Argumentation to support explaining the result of decision-making. Several RSs have been built with DeLP as the main recommendation and explanation engine. One is that of [21] for the movie domain, handling incomplete and contradictory information and using a comparison criterion to solve conflicting situations. Another is introduced by [22], deploying DeLP to provide a hybrid RS in an educational setting, using argumentation to differentiate between different techniques for generating recommendations.

We observe that most existing works focus on use-cases (scenarios) where knowledge of human and artificial agents is static. These works have not fully investigated modelling general real knowledge in which human and artificial agents work together in HI scenarios where complex environments of the HI scenarios are rarely static (e.g. conflicting information may result from shifting preferences, user aspects or changing over time and varying in different contexts). Our work aims to fill this gap, namely, we investigate to what extent these techniques can also be used for dealing with conflicting and dynamic information in the context of HI scenarios.

3. Survey Research

We designed and performed a survey research to determine the capabilities of argumentation in representation and resolving conflicts in the HI scenarios. We describe our research methodology, participants, materials, and procedure.

3.1. Research Methodology

To investigate the concepts of *inconsistency* and *explainability capabilities of Argumentation* in HI scenarios, our research contains two part:

Part 1. We investigate how Argumentation can support representation and reasoning with inconsistent KBs in the HI scenarios. For this purpose, we conduct a survey and follow-up interview among HI project members. Based on a *qualitative data analysis method*, we analyze the survey and interview results. Then, we investigate the existence of inconsistent knowledge and the reasons for the inconsistencies in these scenarios. In particular, we analyse types of knowledge, types of formal representations and additional, or contextual, information (user's motivation, feeling, emotion, behaviour, data provenance, time, preferences). Through analysing these aspects, we find that inconsistencies can be the result of time, context, human aspects, and external conditions such as available resources and environment. The results are shown in Section 4.1.

Part 2. We show how Argumentation can support explainability. For this purpose, we perform a translation of HI scenarios into AF. In particular, based on the types of knowledge and the types of formal representations, we examine whether the translation of HI scenarios (i.e. representation of inconsistent KB in the HI scenarios) into the AF can take place immediately. For each scenario,

in the argumentation setting, we show that arguments represent facts of KB and attack relations between the arguments capture conflicting information. Then, we construct an argumentation tree as a dispute tree. Based on the argumentation tree, we provide a dialectical dialogue of explanation to allow interaction with humans. Besides, we describe how Argumentation enables explainability according to what they explain (i.e., providing explanations for various problem types such as decision-making, justification of an opinion, and explanation through dialogues). We believe that such a classification is more interesting for the reader who tries to locate which research studies are related to the solution of specific problem types. The results are shown in Section 4.1.

3.2. Participants

We conducted a survey among 26 sub-projects of the Hybrid Intelligence Centre⁴. We distributed the survey among primary contacts of these projects. Five of the participants did not respond to our survey, which resulted in a final number of 21 contributing participants.

3.3. Materials and procedure

Part 1 of the study was divided into two sessions. For the first session, we conducted a survey by asking the participants for information through a questionnaire, which was distributed in an online video call to the participants. The survey questions consisted of two parts: The first part included general questions with regard to use-case (or scenario) descriptions that the project members are working on. The second part consisted of specific questions with regard to knowledge (data) considered in the projects. For the second session, we conducted interviews (both online and face to face) focused on the projects that most clearly deal with inconsistencies. Particularly, we conducted an interview with 7 respondents of selected projects after analyzing the responses to the survey. The complete material can be found in a link⁵ for detail.

In the following sections, we discuss the results of the survey and the interviews we conducted.

4. Results

In this section, we describe main outcomes of the study for a selection of sub-projects. Due to space limit of this paper, we cannot discuss all projects but refer the reader to an (online) appendix⁶ where the other projects are presented. In that online appendix we also show a table summarizing the main survey results. Here, for each project, we give an overview of the use-cases (scenarios) and investigate the capabilities of argumentation in representation of inconsistent knowledge and providing explanations. We discuss the survey results.

⁴Since the survey was conducted this number has increased to 32 projects.

⁵<https://forms.gle/i55LgTHdXQr6FRL36>

⁶<https://drive.google.com/file/d/1dGD7TH7PlqMtF5eLDOPHXkn2pzanNcDI/view?usp=sharing>

4.1. Analysing HI Scenarios

After analysing the responses to the survey, we find that 14 out of 21 projects have conflicting information in their use-cases (scenarios). We categorize 14 projects based on the type of problem that Argumentation can address in their use-cases. These problems consists of *decision-making*, *justification* of an opinion, and *dialogues* between Human–System and System–System scenarios. For each use-case, we analyse the reasons why the inconsistencies can occur. Moreover, we explore the use of argumentation by showing how argumentation represents conflicting information and what they explain when solving these problems in decision-making, justification, and dialogues. For Project 2, 10, 11, 12, 13, 16 and 27⁷, we do not analyze in more detail since either conflicting information is not available in their scenarios (Project 2, 10, 11, 12, 16 and 27) or the project currently does not use formalised data/knowledge (Project 13)

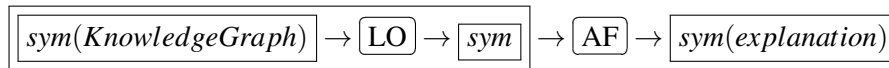
We divide this section based on the most important practical problem types that Argumentation can solve, such as decision-making, justification, and explanation through dialogues. Before continuing, we introduce the notations that we use through this section. We use ovals to denote representative methods to perform knowledge representation in the HI scenarios, and boxes to denote their input and output (i.e. data structures). Accordingly, we distinguish two types of components (ovals): those that perform some form of logical formalisms (labelled as the "LO" components) and those that perform forms of argumentation (labelled as the "AF" components): $\boxed{\text{LO}}$, $\boxed{\text{AF}}$. Based on two aspects (i.e. types of data (knowledge), types of formal representations) based distinction discussed in Research Methodology (Section 3.1), we use two kinds of input and output-boxes: those that contain "model-based" (symbolic, relational) structures, those that contain "model-free" data (model-free data such as images, text or numbers): $\boxed{\text{sym}}$, $\boxed{\text{data}}$.

4.2. Decision-making with Argumentation

The contribution of Argumentation is highly related to decision-making, in fact Argumentation was originally proposed to facilitate decision-making [23]. The contributions of Argumentation are support or opposition of a decision, reasoning for a decision, tackling KBs with inconsistency, and recommendations.

Project 9: AutoAI for dynamic data. Argumentation can also be used to explain a decision-making for calendar scheduling of the agents in a **Digital Assistant** application. *Project 9* aims to construct a hybrid system to assist employees in calendar scheduling within a company. The system has multiple agents that are independently operating, and each agent is assigned a task to setup a meeting and manage their calendar. The agents work independently and attempt to set meetings through bargaining games. The environment in the system is rarely static: Agents are added and removed, other agents can change their calendar, agents have preference over the offer of the user that they are making and the preference can change over time. Conflicting information among agents may result from dynamic environment. For example, other agents can change their calendar, the system might get conflicting proposals over time. In such system, actions like rescheduling or denying a meeting must be explained to the user.

⁷See <https://www.hybrid-intelligence-centre.nl/projects/> for more detail.



Mapping the use-case to the AF is illustrated as follows: *Imagine we have the options (1) book this meeting, (2) do not book this meeting at 10am. The agent schedules the meeting at 10am. Then, we have an argument for booking this meeting at 10am. Unfortunately, the manager gets sick, and he will not be able to join the meeting. He postpones the meeting. This means we have a formal argument not to book the meeting at 10am.* The system should explain why to postpone the meeting. Temporal Datalog ([24]) can be used to model this scenario: Consider $\mathcal{H}_1 = \{\mathcal{R}_1, \mathcal{C}_1, \mathcal{F}_1\}$, in which:

$$\begin{aligned} \mathcal{R}_1 &= \{R_1 : \text{manager}(x) \rightarrow \text{bookMeeting}(x, y, t_1), \\ \mathcal{R}_2 &= \{R_2 : \text{manager}(x) \wedge \text{gotSick}(x, \text{sick}) \rightarrow \text{cancelMeeting}(x, y, t_2)\}; \\ \mathcal{C}_1 &= \{C : \forall x, y \text{ bookMeeting}(x, y, t_1) \wedge \text{cancelMeeting}(x, y, t_2) \wedge t_1 = t_2 \rightarrow \perp\}; \\ \mathcal{F}_1 &= \{f_1 : \text{manager}(\text{Tim}), f_2 : \text{bookMeeting}(\text{Tim}, \text{meetingA}, 10\text{am}), \\ &f_3 : \text{gotSick}(\text{Tim}, \text{sick}), f_4 : \text{cancelMeeting}(\text{Tim}, \text{meetingA}, 10\text{am})\}. \end{aligned}$$

Rule R_1 states that a meeting is booked by a manager at a certain time t_1 . Rule R_2 represents that if a manager got sick, the meeting is not booked by the manager. Rule C expresses the contrary information when booking the meeting. Facts f_1, f_2, f_3, f_4 represent instances of the KB. We define an attack relation between two arguments A_1 and A_2 to model contradicting information when booking the meeting, where:

$$\begin{aligned} A_1 &= (\{\text{manager}(\text{Tim})\}, \{\text{bookMeeting}(\text{Tim}, \text{meetingA}, 10\text{am})\}), \\ A_2 &= (\{\text{manager}(\text{Tim}), \text{gotSick}(\text{Tim}, \text{sick})\}, \{\text{cancelMeeting}(\text{Tim}, \text{meetingA}, 10\text{am})\}). \end{aligned}$$

Two conflicting arguments represent: A_1 states that the meeting A is booked by the manager Tim at 10am, A_2 states that the manager Tim got sick and the meeting A is not booked by him. We construct an argumentation tree as follows.

$$\begin{array}{c} A_1 = (\{\text{manager}(\text{Tim})\}, \{\text{bookMeeting}(\text{Tim}, \text{meetingA}, 10\text{am})\}) \\ \uparrow \\ C : \forall x, y \text{ bookMeeting}(x, y, t_1) \wedge \text{cancelMeeting}(x, y, t_2) \wedge t_1 = t_2 \rightarrow \perp \\ A_2 = (\{\text{manager}(\text{Tim}), \text{gotSick}(\text{Tim}, \text{sick})\}, \{\text{cancelMeeting}(\text{Tim}, \text{meetingA}, 10\text{am})\}) \end{array}$$

Based on the argumentation tree, a dialogical process between the assistant agent and the user can be constructed as follows:

User: Why not
bookMeeting(Tim, meetingA, 10am)
 given that A_2 ?

Reasoner: Because

cancelMeeting(Tim, meetingA, 10am)
⁸

the following constraint is violated:

$C \quad : \quad \forall x, y \text{ bookMeeting}(x, y, t_1) \wedge$
 $\text{cancelMeeting}(x, y, t_2) \wedge t_1 = t_2 \rightarrow \perp$

User: I understood the reason "why
 the meeting A is not booked at
 10am"!

4.3. Dialogues and Argumentation for Explainability

Argumentation-based dialogues can also be used to provide explanations for *opinions*, which is typical for HI scenarios. These dialogues occur between two parties that collaborate to decide what actions to adopt in some situation, or where one tries to persuade the other participant to adopt their point of view.

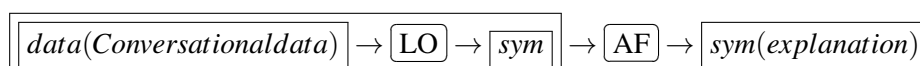
Project 3: Mining texts for perspectives for human-machine deliberation. Argumentation can be used to try to persuade the other participant to adopt their point of view in a deliberation platform. *Project 3* is about constructing a deliberation platform where knowledge takes the form of text-based discussions from a variety of online sources. In online discussions where we can assume participants have stakes, finding, representing and summarizing perspectives is a useful tool to increase the scale of the discussions from relatively small to (hopefully) crowd-scale. In these scenarios, personal preferences, values, real-life context, the topic of the discussion, etc., can be cause for contradictory information. An example can be an anti-vaccination advocate versus a medical doctor trying to have a discussion on vaccination strategy. While they need to decide on some set of actions to take, preferences for what type of action (if any) to make can be due to any of the reasons described above. This is a typical project for the use of the AF to explain decision-making to users. The AF can be used to explain an appropriate opinion through the dialogue where one tries to persuade the other participant to adopt their point of view. Since the project has textual data to represent two-way conversations, mapping the deliberation framework to the AF is very natural.

$$\boxed{data(text)} \rightarrow \boxed{AF} \rightarrow \boxed{sym(explanation)}$$

Project 14: Interactive Machine Reasoning for Responsible HI. In the context of dialogue between human and agent, *Project 14* considers a behaviour support application in the healthy lifestyle domain. The project considers user models that can be defined as the system's representation of the user's knowledge. The user models are constructed through direct interactions at run-time. Based on this user model the agent can derive what it deems to be appropriate support

⁸ $A_2 = (\{manager(Tim), getsick(Tim, sick)\}, \{cancelMeeting(Tim, meetingA, 10am)\})$ is a counter-argument of A_1 .

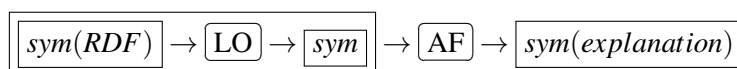
actions. We emphasize that the user models are rarely static, as the user or their context change with time. For example, the user sets *eating healthy* as a long-term goal, while unfortunately he/she behaves otherwise (i.e., he/she eat fast-food at this very moment). In this context, conflicts may occur when a human's current desires conflict with their long-term goals or values. Preferences of information come from the user themselves such as feelings, motivations or emotions. In such scenarios, explainability and transparency of the agent can help to increase the chance that the user actually does what the agent suggests them to do. Argumentation and preference reasoning are used to explain the agent's recommendations that depend on the user's feelings, motivations, emotions. We illustrate the mapping of the use-case to AF as follows. Due to space limitations we have to refer the interested reader to the online Appendix for more specific examples of this usecase.



4.4. Justification through argumentation

Justification is a form of explaining an argument in order to make it more convincing and to persuade an opposing participant. With the help of argumentation and dialogue trees, we can show whether an argument is acceptable or not.

Project 26: Knowledge Representation Formalisms for Hybrid Intelligence. *Query-Answer (QA)* systems are another type of application of the HI projects. Project 26 contains a case study conducted by the authors of this paper. The work investigates reasoning techniques for an inconsistent KB in the biographical domain. Structured biographical metadata is extracted and integrated from heterogeneous sources that are diverse and reflect changes over time [25]. The envisioned QA system would interact with a user, for example a (digital) humanities scholar, to allow that user to understand the diversity and perspectives in the source material, making this a HI scenario. A simple example of an inconsistency in the data is: *A person has multiple biographies from different sources and the biographies list different birthdays. This leads to inconsistent information about the person.* Querying in inconsistent knowledge is non-trivial. Additionally, in such QA system, the users need to understand why an answer is provided for the query and which information of the person's event conflicts with other information. Thus, proving explanation functionalities that enable the users to understand the rationale of an answer is necessary. In our project, we take into account argumentation theory to support the inconsistent-tolerant query answering in inconsistent KBs. The translation of KB to Argumentation is illustrated as follows:



We start with a very simple scenario: *A user queries "When did Johan Rudolph Thorbecke die?" which is expressed as $Q(x) = Person(Thorbecke) \wedge deathDate(x)$. The QA system returns a (credulous) answer "14th Oct 1860" for the query. The user expected that "10th Oct 1860" is also an answer and wants to understand why this is not the answer to his query*

We use Datalog[±] ([13]) to represent the knowledge of this project. Rule R models the concept of a person with a given death date. Rule C represents fundamental constraints: If a person has two death dates, the death dates coincide. Facts f_1, f_2, f_3 express that Thorbecke has death dates that are 14/10/1860 and 10/10/1860 respectively. We now translate the KB \mathcal{K} into SAF: Consider $\mathcal{K}_2 = \{\mathcal{R}_2, \mathcal{C}_2, \mathcal{F}_2\}$ where:

$$\begin{aligned}\mathcal{R}_2 &= \{R : \forall x \text{Person}(x) \rightarrow \exists y \text{deathDate}(x, y)\}; \\ \mathcal{C}_2 &= \{C : \forall x, y, z \text{Person}(x) \wedge \text{deathDate}(x, y) \wedge \text{deathDate}(x, z) \rightarrow y = z\}; \\ \mathcal{F}_2 &= \{f_1 : \text{Person}(\text{Thorbecke}), f_2 : \text{deathDate}(\text{Thorbecke}, 14/10/1860), \\ & f_3 : \text{deathDate}(\text{Thorbecke}, 10/10/1860)\}.\end{aligned}$$

In such argumentation setting, we have a set of arguments:

$$\begin{aligned}A_1 &= (\{\text{Person}(\text{Thorbecke})\}, \{\text{deathDate}(\text{Thorbecke}, 14/10/1860)\}), \\ A_2 &= (\{\text{Person}(\text{Thorbecke})\}, \{\text{deathDate}(\text{Thorbecke}, 10/10/1860)\}).\end{aligned}$$

A_1 attacks A_2 since the argument $\text{deathDate}(\text{Thorbecke}, 14/10/1860)$ is in conflict with $\text{deathDate}(\text{Thorbecke}, 10/10/1860)$, as A_1 and A_2 model conflicting death dates. We construct an argumentation tree:

$$\begin{array}{c} A_2 = (\{\text{Person}(\text{Thorbecke})\}, \{\text{deathDate}(\text{Thorbecke}, 10/10/1860)\}) \\ \uparrow \\ C: \forall x, y, z \text{Person}(x) \wedge \text{deathDate}(x, y) \wedge \text{deathDate}(x, z) \rightarrow y = z \\ \uparrow \\ A_1 = (\{\text{Person}(\text{Thorbecke})\}, \{\text{deathDate}(\text{Thorbecke}, 14/10/1860)\}) \end{array}$$

Next, a dialogical process that explains to the user is the following:

User: Why not $\text{deathDate}(\text{Thorbecke}, 10/10/1860)$ given that A_2 ?⁹

Reasoner: Because we know that $\text{deathDate}(\text{Thorbecke}, 10/10/1860)$ ¹⁰, the following constraint is violated:
 $\forall x, y, z \text{Person}(x) \wedge \text{deathDate}(x, y) \wedge \text{deathDate}(x, z) \rightarrow y = z,$

User: I understood the reason "why 10/10/1860 is not Thorbecke's death date"

The above example shows the potential for explanation facilities to help the user to understand why an answer of the query is (skeptical or credulous) accepted, or not accepted. This example shows the potential of Argumentation to support more natural interaction between humans and systems in HI scenarios.

⁹ $A_2 = (\{\text{Person}(\text{Thorbecke})\}, \{\text{deathDate}(\text{Thorbecke}, 10/10/1860)\})$ is an argumentation where $Q = \text{deathDate}(\text{Thorbecke}, 10/10/1860)$ is a conclusion of A_2

¹⁰ $A_1 = (\{\text{Person}(\text{Thorbecke})\}, \{\text{deathDate}(\text{Thorbecke}, 14/10/1860)\})$ is a counter-argument of A_2

4.5. Discussion

Summary Results. Our analysis shows that 14 out of 21 prototypical Hybrid Intelligence projects have scenarios with inconsistent information. In 7 out of these 14 projects preferential information is available. For 10 out of the 14 projects (Project 3, 8, 9, 14, 19, 20, 23, 26 30 and 32) we identified how to apply Argumentation to model the specific representation knowledge. In particular, 4 out of 10 projects utilise knowledge graphs (KG) as inputs in their scenarios (Project 32, 30, 8 and 26). Project 32 uses RDF and named graphs technologies to model social dialogue between human and embodied agents in multi-modal environments. Project 30 utilises a KG (specifically OWL) to express commonsense knowledge. Knowledge Graphs (ie. DBpedia, ConceptNet,...) are used in Project 8 to construct Conversational Recommender Systems. Similarly, Project 26 considers biographical dictionaries expressed with KGs. Knowledge in Project 9 and Project 14 are also naturally formalised. This means that for 6 out of 10 projects there are already logical representation formalisation available as intermediate steps, that allow seamless mappings into Argumentation. For the remaining projects, the prior knowledge is represented by propositional logic in Project 23; the knowledge in Project 20 can be encoded in frames/FSMs created by experts, or in (PO)MDPs or hybrid solutions; Project 3 employs a text form to represent data in Deliberation platform. Those three projects (Project 20, 23 and 03) can naturally be formalised in AAF directly.

Regarding remaining 4 out of 14 projects, we could not easily identify how to map the use-case Argumentation despite the existence of inconsistent information. The main reasons being that these projects use different types of data that lend themselves less naturally to the AF model, e.g., synthetic numeric data or image data (Project 5). Project 24 uses queries (in text-form), documents (in text-form) and some kind of relevance signal, either click logs or relevance. Project 6 and Project 22, in their current setup, do not have formal representations for their knowledge (data).

For Project 2, 10, 11, 12, 13, 16 and 27, conflicting information is not available in their scenarios or the projects currently do not using data or knowledge. Therefore, the application of Abstract Argumentation to these projects is not natural.

Limitations. While the study provides various interesting insights, it also has limitations. First, we chose to only focus on projects of the HI Centre. More recently, other Hybrid Intelligence cases were introduced that we did not yet consider in this study. Nevertheless, we believe that this study shows the breadth of HI and that Argumentation often is a very suitable formalism, and very concretely, how it can play a role.

Second, we discuss the complexity of using AF in HI systems, and what makes AF so suitable for HI systems. In various scenarios of the HI projects, many dialogue types (such as deliberation and negotiation) concern what should be done in a given situation, rather than what is true. Data/knowledge from these dialogues expressed in natural language or synthetic simple numeric data or documents. The use of argumentation to model the conflict of such data/ knowledge is still challenging. In addition, there are various projects having massive data in real-world application. In such scenarios, a decision may have many (possibly infinite) argumentative claim backings, often the explainee cares only about a small subset (relevant to the context). Thus, the challenge is to select a subset of the possible explanations (based on different criteria), and the explainer and the explainee may interact and argue about these explanations.

5. Conclusion

In this paper, we investigated the topic of Argumentation in Hybrid Intelligence scenarios. Our goals were to (1) demonstrate the capabilities of Argumentation in representing and reasoning about knowledge of both human and artificial agents in the presence of inconsistency in HI, and (2) show how Argumentation enables Explainability in these use-cases (scenarios). We conducted both a survey and follow-up interview among individual projects in the Hybrid Intelligence research program representing a variety of HI scenarios. We analyse to what extent Argumentation is applicable by clarifying the practical inconsistency types of the HI scenarios that Argumentation can address. These include inconsistencies related to commonsense reasoning, decision making, and negotiation. We then model particularly the presentation of conflicting information for selected scenarios based on the form of argument representation. The results show that 14 out of 21 projects have inconsistent information occurring in their scenarios in which preferences of information are available in 7 out of 14 projects. Regarding 14 projects having inconsistent information, we identified that Argumentation Framework can be applied to model the specific tasks for 10 out of 14 projects.

As future work, we plan to focus on modelling and implementation for each scenario. Moreover, our future work may materialize human-machine dialogue from human text dialogues in the HI scenarios, which has not yet received much attention. Causality could be achieved by reasoning over each step that leads to a decision and explaining why alternatives were left out. Nevertheless, we see that not many works exist that combine Argumentation and causality for this purpose. Therefore, we plan to focus on arguments with commonsense knowledge, an interesting area that has not yet received much attention.

Acknowledgments

This work is partially supported by the Hybrid Intelligence programme¹¹, funded by a 10 year Zwaartekracht grant from the Dutch Ministry of Education, Culture and Science.

References

- [1] Z. Akata, D. Balliet, M. de Rijke, F. Dignum, V. Dignum, G. Eiben, A. Fokkens, D. Grossi, K. Hindriks, H. Hoos, H. Hung, C. Jonker, C. Monz, M. Neerinx, F. Oliehoek, H. Prakken, S. Schlobach, L. van der Gaag, F. van Harmelen, H. van Hoof, B. van Riemsdijk, A. van Wynsberghe, R. Verbrugge, B. Verheij, P. Vossen, M. Welling, A research agenda for hybrid intelligence: Augmenting human intellect with collaborative, adaptive, responsible, and explainable artificial intelligence, *Computer* 53 (2020) 18–28.
- [2] L. Bertossi, Database repairs and consistent query answering: Origins and further developments, in: *Proceedings of the 38th ACM SIGMOD-SIGACT-SIGAI Symposium on Principles of Database Systems, PODS '19*, Association for Computing Machinery, New York, NY, USA, 2019, p. 48–58.

¹¹<https://www.hybrid-intelligence-centre.nl/>

- [3] M. Bienvenu, A short survey on inconsistency handling in ontology-mediated query answering, *KI - Kunstliche Intelligenz* (2020) 1–9.
- [4] L. H. Tamargo, A. J. Garcia, M. A. Falappa, G. Simari, A belief revision approach to inconsistency handling in multi-agent systems, in: *The IJCAI-09 Workshop on Nonmonotonic Reasoning, Action and Change (NRAC)*, 2009.
- [5] M. L. Tielman, C. M. Jonker, M. B. van Riemsdijk, What should i do? deriving norms from actions, values and context, in: *MRC@IJCAI*, 2018.
- [6] P. M. Dung, On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games, *Artif. Intell.* 77 (1995) 321–357.
- [7] A. Vassiliades, N. Bassiliades, T. Patkos, Argumentation and explainable artificial intelligence: a survey, *The Knowledge Engineering Review* 36 (2021) e5. doi:10.1017/S0269888921000011.
- [8] T. Miller, Explanation in artificial intelligence: Insights from the social sciences, *Artificial Intelligence* 267 (2019) 1–38.
- [9] K. Cyras, A. Rago, E. Albini, P. Baroni, F. Toni, Argumentative XAI: A survey, *CoRR abs/2105.11266* (2021).
- [10] C. Roberta, C. Giuseppe, L. Francesca, O. Andrea, S. Giovanni, Defeasible systems in legal reasoning: A comparative assessment, *Frontiers in Artificial Intelligence and Applications* (2019) 169 – 174.
- [11] P. Besnard, A. Hunter, A logic-based theory of deductive arguments, *Artificial Intelligence* 128 (2001) 203–235.
- [12] A. Arioua, M. Croitoru, S. Vesic, Logic-based argumentation with existential rules, *International Journal of Approximate Reasoning* 90 (2017) 76 – 106.
- [13] A. Cali, G. Gottlob, T. Lukasiewicz, A. Pieris, Datalog+/-: A family of languages for ontology querying, in: *Datalog Reloaded - 1st International Workshop, Datalog 2010*, volume 6702, 2011, pp. 351–368.
- [14] R. Calegari, A. Omicini, G. Pisano, G. Sartor, Arg2p: an argumentation framework for explainable intelligent systems, *Journal of Logic and Computation* 32 (2022) 369–401.
- [15] O. Arieli, A. Borg, C. Straundefinder, Prioritized sequent-based argumentation, in: *Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems*, International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC, 2018, p. 1105–1113.
- [16] J. Heyninck, C. Straundefinder, A fully rational argumentation system for preordered defeasible rules, in: *Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems, AAMAS '19*, International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC, 2019, p. 1704–1712.
- [17] M. Bienvenu, C. Bourgaux, Querying and repairing inconsistent prioritized knowledge bases: Complexity analysis and links with abstract argumentation, *CoRR abs/2003.05746* (2020).
- [18] H. Prakken, *Logics of Argumentation and the Law*, Cambridge University Press, 2017, p. 3–31.
- [19] T. Botschen, D. Sorokin, I. Gurevych, Frame- and entity-based knowledge for common-sense argumentative reasoning, in: *Proceedings of the 5th Workshop on Argument Mining*, Association for Computational Linguistics, Brussels, Belgium, 2018.

- [20] P. Pilotti, A. Casali, C. Chesñevar, A belief revision approach for argumentation-based negotiation agents, *International Journal of Applied Mathematics and Computer Science* 25 (2015).
- [21] C. E. Briguez, M. C. Budan, C. A. D. Deagustini, A. G. Maguitman, M. Capobianco, G. R. Simari, Argument-based mixed recommenders and their application to movie suggestion, *Expert Syst. Appl.* 41 (2014) 6467–6482.
- [22] P. Rodriguez, S. H. Barbera, J. Palanca, J. M. Poveda, N. D. Duque, V. Julian, An educational recommender system based on argumentation theory, *AI Commun.* 30 (2017) 19–36.
- [23] H. Mercier, D. Sperber, Why do humans reason? arguments for an argumentative theory, *Behavioral and Brain Sciences* 34 (2011) 57–74. doi:10.1017/S0140525X10000968.
- [24] A. Ronca, M. Kaminski, B. C. Grau, B. Motik, I. Horrocks, Stream reasoning in temporal datalog, *AAAI’18/IAAI’18/EAAI’18*, AAAI Press, 2018.
- [25] C. Dijkshoorn, L. Aroyo, J. V. Ossenbruggen, G. Schreiber, Modeling cultural heritage data for online publication, *Appl. Ontology* 13 (2018) 255–271.