

Document downloaded from:

<http://hdl.handle.net/10251/36114>

This paper must be cited as:

Heras Barberá, SM.; Botti Navarro, VJ.; Julian Inglada, VJ. (2012). Argument-based agreements in agent societies. *Neurocomputing*. 75(1):156-162.
doi:10.1016/j.neucom.2011.02.022.



The final publication is available at

<http://dx.doi.org/10.1016/j.neucom.2011.02.022>

Copyright Elsevier

Argument-based Agreements in Agent Societies

S. Heras^{a,*}, V. Botti^a, V. Julián^a

^a*Departamento de Sistemas Informáticos y Computación
Universitat Politècnica de València
Camino de Vera s/n. 46022 Valencia (Spain)
Telf. (+34) 96 387 73 50 - Fax. (+34) 96 387 73 59*

Abstract

In this paper, we present an abstract argumentation framework for the support of agreement processes in agent societies. It takes into account arguments, attacks among them, and the social context of the agents that put forward arguments. Then, we define the semantics of the framework, providing a mechanism to evaluate arguments in view of other arguments posed in the argumentation process. We also provide a translation of the framework into a neural network that computes the set of acceptable arguments and can be tuned to give more or less importance to argument attacks. Finally, the framework is illustrated with an example in a real domain of a water-rights transfer market.

Keywords: Argumentation, Agreement Technologies, Neural Networks

1. Introduction

The new paradigm of computing as interaction views large systems in terms of the set of entities that interact by offering and consuming services in the system [1]. Multi-Agent Systems (MAS) seem to be a suitable technology to implement this paradigm. In these systems, agents can enter in or leave the system (open MAS), interact and dynamically form agent coalitions or organisations. However, the high dynamism of MAS requires agents to have a way of reaching agreements that harmonise the conflicts that come out when they have to collaborate or coordinate their activities. Moreover, open MAS also invalidate common assumptions about the agents of most MAS, such as honesty, cooperativeness and trustworthiness. In addition, agents in MAS form societies that link them via dependency relations that emerge from agent interactions or are predefined by the system. The dependencies among agents define their social context, which has an important influence in the way agents can argue and reach agreements.

*Corresponding author

Email addresses: sheras@dsic.upv.es (S. Heras), vbotti@dsic.upv.es (V. Botti), vinglada@dsic.upv.es (V. Julián)

From ancient times, philosophy has studied the way in which humans produce arguments as defeasible proofs to support their positions in debate. Argumentation is probably the most natural way of harmonising knowledge inconsistencies and reaching agreements. It provides a productive means of dealing with non-monotonic and defeasible reasoning. During the last decade, this important property has made many Artificial Intelligence (AI) researchers place their attention on argumentation theory. Specifically, much research has been done on defining and analysing the properties and complexity of abstract argumentation frameworks [2] (Part I). Most of them are extensions of Dung’s seminal work in [3], which analysed the acceptability semantics of arguments. Research on argumentation is at its peak in the MAS community, since it has proven to be very successful in implementing agents’ internal and practical reasoning and in managing multi-agent dialogues [4].

However, little work has been done to study the effect of the social context of agents in the way that they argue and analyse arguments. Commonly, the term *agent society* is used in the argumentation and AI literature as a synonym for an *agent organisation* [5] or a *group of agents* that play specific roles, follow some interaction patterns and collaborate to reach global objectives [6]. In addition to the dependency relations between agents, we also consider values to be an important element of their social context. These values can be individual values that agents want to promote or demote (e.g. solidarity, peace, etc.) or also *social values* that are inherited from the agents’ dependency relations.

Thus, we endorse the view of abstract value-based argumentation frameworks [7], which stress the importance of the audience in determining whether an argument is persuasive or not. Starting from this approach, we extend it to consider the social context of agents in the notion of defeat between arguments and the underlying semantics of the framework. In addition, an argumentation network representing a value-based argumentation framework can be translated into a neural network by using the *neural network algorithm* presented in [8]. In this way, we are able to execute a sound computation of the prevailing arguments in the argumentation network. Thus, a hybrid artificial intelligence model in which the learning of arguments can be combined with reasoning capabilities is provided. Hybrid artificial intelligence systems (HAISs) [9] combine both symbolic and subsymbolic paradigms to build more robust and trustworthy problem-solving models. In this paper, we adapt the neural algorithm to work with argumentation frameworks for agent societies. Therefore, the underlying model can take into account the social context of agents.

To our knowledge, no research has yet been done to extend abstract argumentation frameworks to represent and manage arguments in agent societies taking into account their social context. Nevertheless, this social information plays an important role in the way agents can argue and in the acceptability semantics of arguments. Depending on their social relations with other agents, an agent can accept arguments from a member of its society that it would never accept before acquiring social dependencies with this member. For instance, in a company subordinates must sometimes accept arguments from their superiors that go against their own ideas and that they would never accept without this

power relation. Also, trade unionists would accept arguments when they are acting as representatives of the trade union that they would never accept when defending their own interests. These are major considerations that should be studied to apply argumentation social domains modelled by means of MAS.

In this paper, we present an abstract argumentation framework for the support of agreement processes in agent societies. This framework takes into account arguments, attacks among them, and the social context of the agents. Section 2 defines our notion of agent societies. Section 3 defines the semantics of the framework, providing a mechanism to evaluate arguments in view of other arguments posed in the argumentation process. Section 4 presents the neural argumentation approach that is followed. Section 5 illustrates the framework with an example. Finally, Section 6 summarises the contributions of this paper.

2. Modelling Agent Societies

In this work, we follow the approach of [10], who define an *agent society* in terms of a set of *agents* that play a set of *roles*, observe a set of *norms* and a set of *dependency relations* between roles, and use a *communication language* to collaborate and reach the global objectives of the *group*. This definition can be adapted to any open MAS where there are norms that regulate the behaviour of agents, roles that agents play, a common language that allows agents to interact by defining a set of permitted locutions and a formal semantics for each of these elements. Moreover, the set of norms in open MAS defines a *normative context* (covering both the set of norms defined by the system itself as well as the norms derived from agents' interactions)[11].

However, we consider that the values that individual agents or groups want to promote or demote and the preference orders over them are also of crucial importance in the definition of an argumentation model for agent societies. These values could explain the reasons that an agent has to give preference to certain beliefs, objectives, actions, etc. Also, dependency relations between roles could imply that an agent must change or violate its value preference order. For instance, agents of higher hierarchy could impose their values on subordinates or an agent could have to adopt a certain preference order over values in order to be accepted in a group. Therefore, we endorse the view of [7], who stress the importance of the audience in determining whether an argument (e.g. for accepting or rejecting someone else's beliefs, objectives or action proposals) is persuasive or not. Thus, in the above definition of agent society we have included the notion of values and the preference orders among them. Next, we provide a formal definition for the model of society that we have adopted:

Definition 2.1 (Agent Society). *An Agent society in a certain time t is defined as a tuple $S_t = \langle Ag, Rl, D, G, N, V, Roles, Dependency, Group, val, Valpref_Q \rangle$ where:*

- $Ag = \{ag_1, ag_2, \dots, ag_I\}$ is the set of I agents of S_t in a certain time t .
- $Rl = \{rl_1, rl_2, \dots, rl_J\}$ is the set of J roles that have been defined in S_t .

- $D = \{d_1, d_2, \dots, d_K\}$ is the set of K possible dependency relations in S_t .
- $G = \{g_1, g_2, \dots, g_L\}$ is the set of groups that the agents of S_t form, where each $g_l = \{a_1, a_2, \dots, a_M\}$, $M \leq I$ consist of a set of agents $a_i \in A$ of S_t .
- N is the defined set of norms that affect the roles that agents play in S_t .
- $V = \{v_1, v_2, \dots, v_P\}$ is the set of P values predefined in S_t .
- $\text{Roles} : \text{Ag} \rightarrow 2^{Rl}$ is a function that assigns an agent its roles in S_t .
- $\text{Dependency}_{S_t} : <_D^{S_t} \subseteq Rl \times Rl$ defines a reflexive, transitive and asymmetric partial order relation over roles.
- $\text{Group} : \text{Ag} \rightarrow 2^G$ is a function that assigns an agent its groups in S_t .
- $\text{val} : \text{Ag} \rightarrow V$ is a function that assigns an agent its set of values.
- $\text{Valpref}_Q \subseteq V \times V$, where $Q = \text{Ag} \vee Q = G$, defines an irreflexive, transitive and asymmetric preference relation $<_Q^{S_t}$ over the values.

That is, $\forall r_1, r_2, r_3 \in R, r_1 <_d^{S_t} r_2 <_d^{S_t} r_3$ implies that r_3 has the highest rank with respect to the dependency relation d in S_t . Also, $r_1 <_d^{S_t} r_2$ and $r_2 <_d^{S_t} r_1$ imply that r_1 and r_2 have the same rank with respect to d in S_t . Finally, $\forall v_1, v_2, v_3 \in V, \text{Valpref}_{ag_i} = v_1 <_{ag_i}^{S_t} v_2 <_{ag_i}^{S_t} v_3$ implies that agent ag_i prefers value v_3 to v_2 and value v_2 to value v_1 in S_t . Similarly, $\text{Valpref}_{g_j} = v_1 <_{g_j}^{S_t} v_2 <_{g_j}^{S_t} v_3$ implies that group g_j prefers value v_3 to v_2 and value v_2 to value v_1 .

3. Framework Formalisation

Most abstract argumentation frameworks (AFs) are based on Dung's framework [3], which is defined as a pair $\langle A, R \rangle$ where A is a set of arguments and $R \subseteq A \times A$ is a binary *attack* relation on A . For two arguments A and B , $R(A, B)$ means that the argument A attacks the argument B . AF abstract the structure and meaning of arguments and attacks between them and focus their research efforts on analysing generic properties and *argumentation semantics*. This semantics is the formal definition of the method by which arguments are evaluated in view of other arguments [12]. Semantics can be either *extension-based*, which determines the *extensions* or sets of arguments that can be collectively acceptable, or *labelling-based*, which labels each argument of A with a specific state in a predetermined set of possible states of an argument.

Based on Dung's AF, we define an Argumentation Framework for an Agent Society (AFAS) as:

Definition 3.1 (Argumentation Framework for an Agent Society). *An argumentation framework for an agent society is a tuple $AFAS = \langle A, R, S_t \rangle$ where: A is a set of arguments; R is an irreflexive binary attack relation on A ; and S_t is a society of agents as presented in Definition 2.1.*

Then, we specialise *AFAS* considering them for a specific agent since each agent of an open MAS can have a different preference order over values. Thus, an *audience* is defined as a preference order over values. For the definition of our Agent specific Argumentation Framework for Agent Societies, we start from the definition of Audience specific Value-based Argumentation Frameworks (AVAF) [7]. This is also based on Dung’s framework, and we will extend and adapt it to take into account the social context of agents.

Definition 3.2 (Audience-specific Value-based AF). *An audience-specific value-based argumentation framework is a 5-tuple $AVAF_a = \langle A, R, V, val, Valpref_a \rangle$ where: A, R, V and val are as defined for a Value-based Argumentation Framework (VAF) [7]; $a \in P$ is an audience of the set of audiences P ; and $Valpref_a \subseteq V \times V$ is a transitive, irreflexive, and asymmetric preference relation that reflects the value preferences of the audience a .*

Then, we extend AVAFs and define our abstract Agent-specific Argumentation Framework in an Agent Society (AAFAS) as follows:

Definition 3.3 (Agent-specific AF for an Agent Society). *An agent specific argumentation framework for an agent society is a tuple $AAFAS = \langle Ag, Rl, D, G, N, A, R, V, Role, Dependency_{S_t}, Group, Values, val, Valpref_{ag_i} \rangle$ where:*

- *$Ag, Rl, D, G, N, A, R, V, Dependency_{S_t}, Group$ and $Values$ are defined as in Definition 2.1.*
- *$Role(ag, a) : Ag \times A \rightarrow Rl$ is a function that assigns an agent the specific role that it plays (from its set of roles) when it has put forward a specific argument.*
- *$val(ag, a) : Ag \times A \rightarrow 2^V$ is a function that assigns an agent’s argument the value(s) that it promotes.*
- *$Valpref_{ag_i} \subseteq V \times V$, defines an irreflexive, transitive and asymmetric relation $\prec_{ag_i}^{S_t}$ over the agent’s ag_i values in the society S_t .*

The aim of *AAFAS* is to determine which agent’s argument attacks another agent’s argument in an argumentation process performed in a society of agents and, in each case, which argument would defeat the other. To do that, we not only have to consider the values that arguments promote and their preference relation as in AVAFs, but also the dependency relations between agents. These relations could be stronger than value preferences in some cases (depending on the application domain). For the time being, as in [10], we only consider the following dependency relations:

- *Power*: when an agent has to accept a request from another agent because of some pre-defined domination relationship between them. For instance, in a society S_t that manages the water-rights transfer of a river basin (as explained in the example of Section 5), $Farmer \prec_{Pow}^{S_t} BasinAdministrator$, since farmers must comply with the laws announced by the basin administrator.

- *Authorisation*: when an agent has committed itself to another agent for a certain service and a request from the latter leads to an obligation when the conditions are met. For instance, in S_t , $Farmer_i <_{Auth}^{S_t} Farmer_j$, if $Farmer_j$ has contracted a service that $Farmer_i$ offers.

- *Charity*: when an agent is willing to answer a request from another agent without being obliged to do so. For instance, in S_t , by default $Farmer_i <_{Ch}^{S_t} Farmer_j$ and $Farmer_j <_{Ch}^{S_t} Farmer_i$.

Thus, we can now define the agent-specific defeat relation of AAFAS as:

Definition 3.4 (Defeat). *An agent's ag_1 argument $a_1 \in AAFAS$ that is put forward in the context of a society S_t defeats $_{ag_1}$ another agent's $ag_2 \in AAFAS$ argument a_2 iff $attack(a_1, a_2) \wedge (val(ag_1, a_1) <_{ag_1}^{S_t} val(ag_1, a_2) \notin Valpref_{ag_1}) \wedge (Role(ag_1) <_{Pow}^{S_t} Role(ag_2) \vee Role(ag_1) <_{Auth}^{S_t} Role(ag_2) \notin Dependency_{S_t})$*

Therefore, we express that from the ag_1 point of view the argument a_1 defeats $_{ag_1}$ the argument a_2 as $defeats_{ag_1}(a_1, a_2)$ if a_1 attacks a_2 , ag_1 prefers the value promoted by a_1 to the value promoted by a_2 , and ag_2 does not have a power or authority relation over ag_1 . Thus, based on Dung's acceptability semantics, we can define some acceptability concepts. Note that, in these definitions, we compare arguments of different agents. However, since dependency relations are partial order relations (reflexive, asymmetric and transitive), an agent has equal power, authorisation, and dependency relations over itself ($ag \leq ag$ (reflexivity) $\rightarrow ag = ag$ (antisymmetry)) and, in the case of comparing arguments of the same agent, the AAFAS would be equivalent to an AVAF and the acceptability criteria of this AVAF would apply. Let $a_i, a_j, a_k \in A$ be the arguments of agents $ag_i, ag_j, ag_k \in Ag$, respectively, and let $a \in A$ be the argument of a generic agent.

Definition 3.5 (Conflict-free). *A set of arguments $ARG \in A$ is conflict-free $_{ag_i}$ for an agent ag_i in the society S_t if $\nexists a_i, a_j \in ARG / (attacks(a_i, a_j) \vee attacks(a_j, a_i)) \wedge ((val(ag_i, a_i) <_{ag_i}^{S_t} val(ag_i, a_j) \notin Valpref_{ag_i}) \wedge (val(ag_i, a_j) <_{ag_i}^{S_t} val(ag_i, a_i) \notin Valpref_{ag_i}) \wedge (Role(ag_i) <_{Pow}^{S_t} Role(ag_j) \notin Dependency_{S_t}) \wedge (Role(ag_j) <_{Pow}^{S_t} Role(ag_i) \notin Dependency_{S_t}) \wedge (Role(ag_i) <_{Auth}^{S_t} Role(ag_j) \notin Dependency_{S_t}) \wedge (Role(ag_j) <_{Auth}^{S_t} Role(ag_i) \notin Dependency_{S_t}))$.*

That is, a set of arguments is *conflict-free* if the following condition hold: there is no pair of arguments that attack each other, without a value preference relation or a dependency relation that invalidates the attack. Note that agent ag_i and agent ag_j can be the same in order to consider the case of arguments put forward by the same agent.

Definition 3.6 (Acceptability). *An argument $a_i \in A$ is acceptable $_{ag_i}(a_i)$ in a society S_t wrt a set of arguments $ARG \in A$ iff $\forall a_j \in A \wedge defeats_{ag_i}(a_j, a_i) \rightarrow \exists a_k \in ARG \wedge defeats_{ag_i}(a_k, a_j)$.*

That is, if the argument is *defeated $_{ag_i}$* by another argument of A , some argument of the subset ARG defeats $_{ag_i}$ this other argument.

Definition 3.7 (Admissibility). A conflict-free set of arguments $ARG \in A$ is admissible for an agent ag_i iff $\forall a \in ARG \rightarrow \text{acceptable}_{ag_i}(a)$.

Definition 3.8 (Preferred Extension). A set of arguments $ARG \in A$ is a preferred – extension $_{ag_i}$ for an agent ag_i if it is a maximal (wrt set inclusion) admissible $_{ag_i}$ subset of A .

Then, for any $AAFAS = \langle Ag, Rl, D, G, N, A, R, V, Role, Dependency_{S_t}, Group, Values, val, Valpref_{ag_i} \rangle$ there is a corresponding $AFAS = \langle A, R, S_t \rangle$, where $R = \text{defeats}_{ag_i}$. Thus, each attack relation of $AFAS$ has a corresponding agent specific defeat_{ag_i} relation in $AAFAS$. These properties are illustrated in the example of Section 5. The next section introduces the neural network algorithm that we use to compute the prevailing arguments (the preferred extensions) of our argumentation framework for agent societies.

4. Neural Argumentation Algorithm

Once we have proposed our argumentation framework for agent societies, we aim to perform a sound and efficient computation of the preferred extensions of the framework. The work proposed in [8] presents an interesting approach that translates value-based argumentation networks into computable neural networks. Thus, we extend this approach to work with the proposed argumentation framework for agent societies.

A neural network consists of a set of nodes (*neurons*) and a set of edges (*connections*) from neuron i to neuron j labelled with a weight $W_{ij} \in \mathfrak{R}$. Each neuron is characterised at a certain time t by its *input vector* $I_i(t)$, its *input potential* $U_i(t)$, its *activation state* $A_i(t)$ and its *output vector* $O_i(t)$. The input potential of neuron i at time t is calculated as a weighted sum such that $U_i(t) = \sum_j W_{ij} I_j(t)$. The activation state $A_i(t)$ is given by the neuron *activation function* h_i such that $A_i(t) = h_i(U_i(t))$. In addition, a weight θ_i with input fixed at 1 is the *threshold* of neuron i such that i is active at time t if $A_i(t) > \theta_i$.

A relation between argumentation networks and neural networks by representing arguments as neurons and connections as attacks (with negative weights) and supports (with positive weights) between arguments is established. Then, in order to compute the prevailing arguments of the network (the preferred extension), the relative strength of the attacks is considered. Following our approach for an agent-specific argumentation framework in an agent society and our notion of defeats_{ag} from the point of view of agent ag , we define the following function to represent the strength of an argument:

Definition 4.1 (Strength of an Argument). The strength of an argument is represented by the function $v_{ag} : A \times A \rightarrow \{0, 1\}$ where $ag \in Ag$ represents an agent. Thus, given two arguments $a_i, a_j \in A$, if $\text{defeats}_{ag}(a_i, a_j)$, then $v_{ag}(a_i, a_j) = 1$ and a_i is said to be stronger than a_j for agent ag . Otherwise, $v_{ag}(a_i, a_j) = 0$, and a_i is said to be weaker than a_j for agent ag .


```

Input : Given an Argumentation Framework for an Agent Society AFAS
           with arguments  $A = \{a_1, a_2, \dots, a_n\}$ 
1.1 Create the input and output layers of a neural network  $\mathcal{N}$  such that the
      $i$ -th neuron represents the  $i$ -th argument of the AFAS
1.2 Given  $0 < A_{min} < 1$ , calculate  $W \geq \frac{1}{A_{min}} * (\ln(1 + A_{min}) - \ln(1 - A_{min}))$ 
1.3 for each  $a_l \in A$  ( $1 \leq l \leq n$ ) do
1.4   | Add a neuron  $N_l$  to the hidden layer of  $\mathcal{N}$ 
1.5   | Connect neuron  $a_l$  in the input layer of  $\mathcal{N}$  to hidden neuron  $N_l$  and set
     | the connection weight to  $W$ 
1.6   | Connect hidden neuron  $N_l$  to neuron  $a_l$  in the output layer of  $\mathcal{N}$  and
     | set the connection weight to  $W$ 
1.7 end
1.8 for each  $(a_i, a_j) \in R^1$  do
1.9   | Connect hidden neuron  $N_i$  to output neuron  $a_j$ 
1.10  | if  $v_{ag}(a_i, a_j) = 0$  then
1.11  |   |  $W' > h^{-1}(A_{min}) - W * A_{min}$ 
1.12  |   end
1.13  | if  $v_{ag}(a_i, a_j) = 0$  then
1.14  |   |  $W' < \frac{h^{-1}(-A_{min}) - W}{A_{min}}$ 
1.15  |   end
1.16 end
1.17 Set the activation threshold  $\theta$  of each neuron in  $\mathcal{N}$  to zero
1.18 Set  $g(x) = x$  as the activation function of the neurons of the input layer of
      $\mathcal{N}^2$ 
1.19 Set  $h(x) = \frac{2}{1 + e^x} - 1$  as the activation function of the neurons in the
     hidden and output layers of  $\mathcal{N}$ 
Output: Neural Network  $\mathcal{N}$ 

```

Algorithm 1: Neural Argumentation Algorithm

The neural network algorithm generates a *single hidden layer* neural network for a value-based argumentation network in which solid connections represent positive weights $W > 0$ (supports from one argument to another), and dotted connections represent negative weights $W' < 0$ (attacks from one argument to another). The specific selection of the functions used to calculate W' and the activation functions $g(x)$ and $h(x)$ are justified in [8]. By running the neural network (setting the weights $W_r = 1$), the preferred extension(s) of the underlying argumentation network can be computed (even when the argumentation network presents cycles). An example is explained in Section 5. Algorithm 1 shows the pseudocode of the neural network algorithm that is extended to work with argumentation frameworks for agent societies.

¹The *attack* relation R .

²In this way, $I_n(t) \in \{-1, 1\}$, where 1 represents the Boolean value *true* and -1 represents the boolean value *false*.

The Algorithm 1 takes a value-based argumentation framework and outputs a single hidden layer neural network \mathcal{N} . It uses a semi-linear activation function $h(x) = \frac{2}{1 + e^x} - 1$ and a parameter $A_{min} \in (0, 1)$ that establishes a threshold over which a neuron can be considered *active*. Then, a neuron is considered active whenever $A_\alpha(t) > A_{min}$ and inactive when $A_\alpha(t) < -A_{min}$. Also, the algorithm establishes a support weight $W > 0$ and an attack weight $W' < 0$ for the connections. The exact value for weights can be determined by each agent. In this way, an agent can assign more or less importance to support or attack arguments, depending on its preferences or the application domain. Furthermore, the weights can be modified by training the algorithm to converge to an expected output (a desired preferred extension or set of prevailing arguments). This is an interesting approach to test the semantics of the framework under different settings. The next section provides an example of the proposed framework and its corresponding neural network.

5. Application of the Framework to the Management of Water-Right Transfer Agreements

To exemplify our framework, let us propose a scenario of an open MAS that represents a water market [13], where agents are users of a river basin, belong to a society S_t , and can enter or leave the system to buy and sell water rights. A water right is a contract with the basin administrator that specifies the volume that can be spent, the water price, the district where the water is located, etc. Here, suppose that two agents that play the role of farmers (F_1 and F_2) in a group (the river basin RB) are arguing to decide over a water-right transfer agreement and a basin administrator (BA) controls the process and makes a final decision. The basin has a set of norms N_{RB} and commands a charity (Ch) dependency relation between two farmers ($Farmer <_{Ch}^{S_t} Farmer$) and a power (Pow) dependency relation between a basin administrator and a farmer ($Farmer <_{Pow}^{S_t} BasinAdministrator$). In addition, the farmers prefer to reach an agreement before taking legal action to avoid the intervention of a jury (J). Also, F_1 prefers economy over solidarity ($SO <_{F_1}^{S_t} J <_{F_1}^{S_t} EC$), F_2 prefers solidarity over economy ($J <_{F_2}^{S_t} EC <_{F_2}^{S_t} SO$) and by default, BA has the value preference order of the basin, (which is $EC <_{BA}^{S_t} SO <_{BA}^{S_t} J$).

In this scenario, F_1 puts forward the argument: “I should be the beneficiary of the transfer because my land is adjacent to the owner’s land”. Here, we suppose that the closer the plots of land, the cheaper the transfers between them and then, this argument could promote economy. However, F_2 replies with the argument: “I should be the beneficiary of the transfer because there is a drought and my land is almost dry”. In this argument, we assume that crops are lost in dry lands and that helping people to avoid losing crops promotes solidarity. Also, agents know that if no agreement is reached, a jury must interfere and everyone wants to avoid that. Then, they can also put forward the following arguments: “ F_2 should allow me (F_1) to be the beneficiary of the water-right transfer to avoid the intervention of a jury (J)”, and “ F_1 should allow me (F_2) to be the

beneficiary of the water-right transfer to avoid the intervention of a jury (J)”. In addition, the *BA* knows that the jury will interfere if the agreement violates the value preferences of the river basin (which promotes solidarity over economy) and puts forward the argument: “ F_1 should allow F_2 to be the beneficiary of the water-right transfer to avoid the intervention of a jury (J)”.

In view of this context, the *BA* could generate an $AFAS = \langle A, R, S_t \rangle$. Thus, we have the following arguments, which are all possible solutions for the water-right transfer agreement process:

- A1 (posed by F_1): F_1 should be the beneficiary of the water transfer (denoted by F_1w) to promote economy (EC).
- A2 (posed by F_2): F_1 should not be the beneficiary of the water transfer (denoted by F_1nw) to promote solidarity (SO).
- A3 (posed by F_2): F_2 should be the beneficiary of the water transfer (denoted by F_2w) to promote solidarity (SO).
- A4 (posed by F_1): F_2 should not be the beneficiary of the water transfer (denoted by F_2nw) to promote saving (EC).
- A5 (posed by F_1): F_2 should allow F_1 to be the beneficiary of the water transfer ($F_1w \& F_2nw$) to avoid the intervention of a Jury (J).
- A6 (posed by F_2 and *BA*): F_1 should allow F_2 to be the beneficiary of the water transfer ($F_1nw \& F_2w$) to avoid the intervention of a Jury (J).

The water transfer cannot be decided in favour of both water users, so $attacks(A1, A3)$ and vice versa, and we assume that a decision favouring at least one part must be taken, so $attacks(A2, A4)$ and vice versa. In addition, a water user cannot simultaneously be and not be the beneficiary of a water transfer, so $attacks(A1, A2)$ and $attacks(A3, A4)$ and vice versa. Also, $attacks(A5, A2)$, $attacks(A5, A3)$, and $attacks(A5, A6)$ and all of these arguments attack A5. Also, $attacks(A6, A1)$, $attacks(A6, A4)$, and $attacks(A6, A5)$, and all of these arguments attack A6. Then:

$$A = \{A1, A2, A3, A4, A5, A6\}$$

$$R = \{attacks(A1, A3), attacks(A3, A1), attacks(A2, A4), attacks(A4, A2), attacks(A1, A2), attacks(A2, A1), attacks(A3, A4), attacks(A4, A3), attacks(A5, A2), attacks(A5, A3), attacks(A5, A6), attacks(A2, A5), attacks(A3, A5), attacks(A6, A5), attacks(A6, A1), attacks(A6, A4), attacks(A1, A6), attacks(A4, A6)\}$$

$$S_t = \langle Ag, Rl, D, G, N, V, Role, Dependency_{S_t}, Group, Values, Valpref_Q \rangle$$

> where:

- $Ag = \{F_1, F_2, BA\}$
- $Rl = \{Farmer, BasinAdministrator\}$

- $D = \{Power, Charity\}$
- $G = \{RB\}$
- $N = N_{RB}$
- $V = \{EC, SO, J\}$
- $Role(F_1) = Role(F_2) = Farmer$ and $Role(BA) = BasinAdministrator$
- $Dependency_{S_t} = \{Farmer <_{Pow}^{S_t} BasinAdministrator, Farmer <_{Ch}^{S_t} Farmer\}$
- $Group(F_1) = Group(F_2) = Group(BA) = RB$
- $Values(F_1) = Values(F_2) = Values(BA) = \{EC, SO, J\}$
- $Valpref_{F_1} = \{SO <_{F_1}^{S_t} J <_{F_1}^{S_t} EC\}$, $Valpref_{F_2} = \{EC <_{F_2}^{S_t} J <_{F_2}^{S_t} SO\}$,
 $Valpref_{BA} = \{EC <_{BA}^{S_t} SO <_{BA}^{S_t} J\}$

Therefore, the *AFAS* for this example is shown in Figure 1a. Furthermore, Figure 2 shows the corresponding neural network obtained by applying Algorithm 1 to the *AFAS*. Solid connections (supports) from the hidden layer to the output layer are labelled with positive weights W , and dotted connections (attacks) are labelled with negative weights W' . These negative weights are omitted in Figure 2 for purposes of clarity.

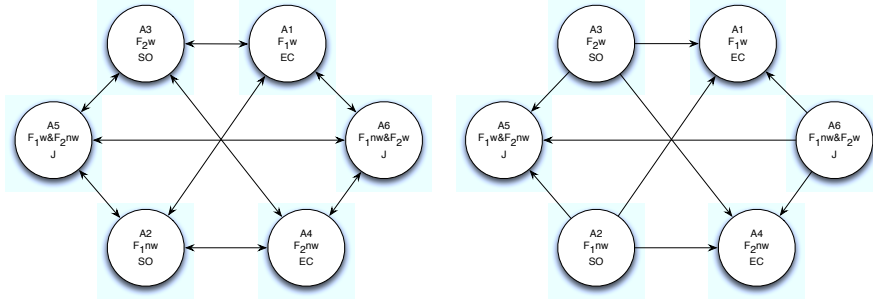


Figure 1: a) *AFAS* Example; b) *AFAS*_{F₂} Example

Also for purposes of clarity, let us take the assumption that $W = 1$, $W' = -1$ if $v_{ag}(a_i, a_j) = 1$, $W' = 0$ if $v_{ag}(a_i, a_j) = 0$, $h(x) = 1$ if $x \geq 1$, $h(x) = -1$ if $x \leq -1$ and $h(x) = x$ if $-1 < x < 1$ ¹. Thus, we give the same weight to supporting and attacking arguments (not considering those attacking arguments that do not defeat the attacked argument). The activation states for each neuron of the input (I), hidden (H) and output (O) layers are the following:

- *Input Layer*: $\forall i \in \{1..6\}$, $g_i(x) = x$ and $I_i(t) = W_r * O_i(t-1) = O_i(t-1)$
- *Hidden Layer*: $\forall i \in \{1..6\}$, $U_i(t) = W * I_i = I_i$ and $H_i(t) = h(U_i(t))$

¹This is an approximation of the standard sigmoid activation function.

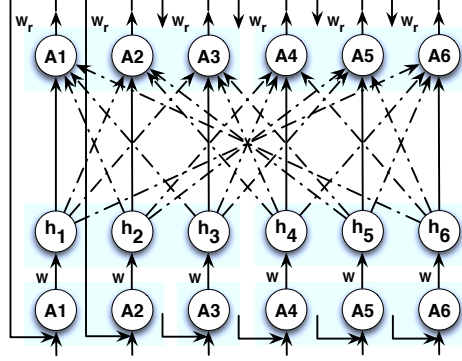


Figure 2: AFAS Neural Network Example

- Output Layer: $\forall i \in \{1..6\}$

$$U_1(t) = W * H_1(t) + W' * H_2(t) + W' * H_3(t) + W' * H_6(t)$$

$$U_2(t) = W' * H_1(t) + W * H_2(t) + W' * H_4(t) + W' * H_5(t)$$

$$U_3(t) = W' * H_1(t) + W * H_3(t) + W' * H_4(t) + W' * H_5(t)$$

$$U_4(t) = W' * H_2(t) + W' * H_3(t) + W * H_4(t) + W' * H_6(t)$$

$$U_5(t) = W' * H_2(t) + W' * H_3(t) + W * H_5(t) + W' * H_6(t)$$

$$U_6(t) = W' * H_1(t) + W' * H_4(t) + W' * H_5(t) + W * H_6(t)$$

$$\text{and } O_i(t) = h(U_i(t))$$

Now, let us consider what happens with specific agents by creating their AAFAS. For instance, recalling that F_1 prefers economy to other values and gives solidarity the lesser value ($SO <_{F_1}^{S_t} J <_{F_1}^{S_t} EC$), we have that $AAFAS_{F_1} = \langle Ag, Rl, D, G, N, A, R, V, Role, Dependency_{S_t}, Group, Values, val, Valpref_{F_1} \rangle$. Then, eliminating the unsuccessful attacks (due to value preferences of F_1), we have the equivalent $AFAS_{F_1}$ for $AAFAS_{F_1}$ as $AFAS_{F_1} = \langle A, \{attacks(A1, A3), attacks(A1, A2), attacks(A4, A2), attacks(A4, A3), attacks(A5, A2), attacks(A5, A3), attacks(A6, A5), attacks(A6, A1), attacks(A6, A4)\}, S_t \rangle$, which is shown in the graph of Figure 3a.

Now, by applying Algorithm 1 and the values for the weights and activation functions specified above, we can compute the preferred extension(s) of $AFAS_{F_1}$ by taking into account the value preferences of F_1 and the dependency relationships of the river basin. In this case, $v_{F_1}(A1, A3)=1$, $v_{F_1}(A3, A1)=0$, $v_{F_1}(A2, A4)=0$, $v_{F_1}(A4, A2)=1$, $v_{F_1}(A1, A2)=1$, $v_{F_1}(A2, A1)=0$, $v_{F_1}(A3, A4)=0$, $v_{F_1}(A4, A3)=1$, $v_{F_1}(A5, A2)=1$, $v_{F_1}(A5, A3)=1$, $v_{F_1}(A5, A6)=0$, $v_{F_1}(A2, A5)=0$, $v_{F_1}(A3, A5)=0$, $v_{F_1}(A6, A5)=1$, $v_{F_1}(A6, A1)=1$, $v_{F_1}(A6, A4)=1$, $v_{F_1}(A1, A6)=0$ and $v_{F_1}(A4, A6)=0$. We start with the input vector $[1,1,1,1,1,1]$ and obtain the output vector $[0,-1,-1,0,0,1]$. Then, we use this vector as the input vector for the next iteration of the algorithm. The sequence from the initial input vector to the output vectors are the following: $[1,1,1,1,1,1] \rightarrow [0,-1,-1,0,0,1] \rightarrow [-1,-1,-1,-1,-1,-1] \rightarrow [-1,1,1,-1,-1,1] \rightarrow [-1,1,1,-1,-1,-1]$

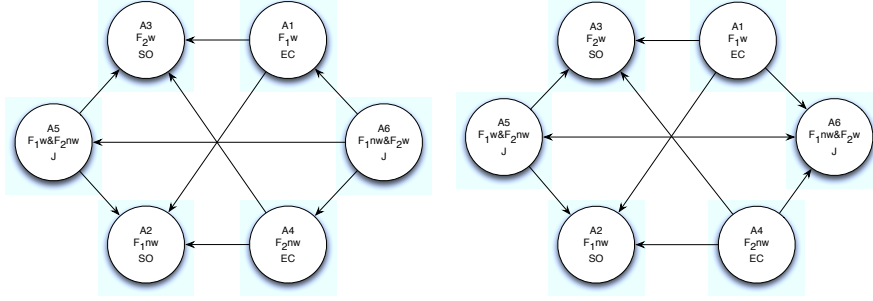


Figure 3: a) $AFAS_{F_1}$ Example; b) $AFAS_{F_1}$ Modified Example

$1,1]$, which converge to the stable state $[-1,1,1,-1,-1,1]$. In this way, the algorithm computes that $A2$, $A3$ and $A6$ should prevail.

Then, from the F_1 point of view, the $AFAS$ neural network has the preferred extension $PE_{F_1} = \{A2, A3, A6\}$, meaning that F_2 should be the beneficiary of the water-right transfer to promote solidarity and the no intervention of a jury. This demonstrates how the power dependency relation of BA prevails over farmers and their arguments. Otherwise, if we change the environment and set a charity dependency relation of basin administrators over farmers $Farmer <_{Ch}^{S_t} BasinAdministrator$, the preferences of F_1 would prevail and the graph would be like the one shown in Figure 3b.

By applying the algorithm, from the input vector $[1,1,1,1,1,1]$ we obtain the sequence of output vectors $[1,-1,-1,1,0,-1] \rightarrow [1,-1,-1,1,1,-1] \rightarrow [1,-1,-1,1,1,-1]$. Thus, the network converges to the stable state $[1,-1,-1,1,1,-1]$ and the preferred extension would be $PE_{F_1, modified} = \{A1, A4, A5\}$ that would defend F_1 as the beneficiary of the transfer agreement.

In its turn, F_2 gives the highest value to solidarity, but prefers to avoid a jury over economy ($EC <_{F_2}^{S_t} J <_{F_2}^{S_t} SO$). Therefore, $AAFAS_{F_2} = \langle Ag, Rl, D, G, N, A, R, V, Role, Dependency_{S_t}, Group, Values, val, Valpref_{F_2} \rangle$. Then, eliminating the unsuccessful attacks we have the equivalent $AFAS_{F_2}$ for $AAFAS_{F_2}$ as $AFAS_{F_2} = \langle A, \{attacks(A3, A1), attacks(A2, A1), attacks(A2, A4), attacks(A2, A5), attacks(A3, A1), attacks(A3, A4), attacks(A3, A5), attacks(A6, A5), attacks(A6, A1), attacks(A6, A4)\}, S_t \rangle$, which is shown in the graph in Figure 1b.

By applying the algorithm, from the input vector $[1,1,1,1,1,1]$, we obtain the sequence of output vectors $[-1,1,1,-1,-1,1] \rightarrow [-1,1,1,-1,-1,1]$. Thus, the network converges to the stable state $[-1,1,1,-1,-1,1]$ and the preferred extension would be $PE_{F_2} = A2, A3, A6$, which means that F_2 defends its position as beneficiary of the water transfer.

In this example, we have illustrated the properties of the proposed argumentation framework for agent societies. In addition, the resulting $AFAS$ argumentation network has been translated into a neural network by using Algorithm 1.

For purposes of simplicity, we have given the same weight to support arguments and successful attacks, and we have used a simplification of the algorithm activation function $h_i(t)$. However, these weights can be fixed to give more or less importance to each type of argument.

6. Conclusions

In this paper, we have presented an abstract argumentation framework to help reach agreements in agent societies. After defining our concept of agent society, we have provided the formal definition of our argumentation framework for agent societies. This is an extension of Value-based Argumentation Frameworks [7] to include agents' values, value preference orders, and dependency relations. The influence of dependency relations among agents in argumentation discourses is a new area of research. An interesting related work that characterises hierarchies in which contradictory information can be present, and argumentation allows to solve possible inconsistencies was presented in [14].

The main properties of the framework have been presented. The argumentation network resulting from the argumentation framework has been translated into a neural network by using a neural network algorithm. In this way, the preferred extension(s) of the argumentation network can be efficiently computed. Then, the framework and its properties have been illustrated in a real scenario of a water rights transfer market, where two farmer agents argue with a basin administrator agent to be the beneficiary of a water-right transfer agreement.

The integration of neural networks and argumentation is a promising area for modelling complex systems (as in the case of agent societies) with few contributions to date. The proposal of hybrid argumentation of [15], in which the interplay between neural network argumentation and symbolic argumentation is analysed based on the pioneering work of [16] constitutes some recent related work. Further approaches should investigate new translation algorithms and perform a comprehensive analysis on the network performance, training the it to adapt weights and converge to desirable outcomes.

Acknowledgements

This work is supported by the Spanish government grants CONSOLIDER INGENIO 2010 CSD2007-00022, TIN2008-04446 and TIN2009-13839-C03-01 and by the GVA project PROMETEO 2008/051.

References

- [1] M. Luck, P. McBurney, Computing as interaction: agent and agreement technologies, in: IEEE International Conference on Distributed Human-Machine Systems, 2008.
- [2] I. Rahwan, G. Simari (Eds.), Argumentation in Artificial Intelligence, Springer, 2009.

- [3] P. M. Dung, On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming, and n -person games, *Artificial Intelligence* 77 (1995) 321–357.
- [4] I. Rahwan, Argumentation in multi-agent systems, *Autonomous Agents and Multiagent Systems*, Guest Editorial 11 (2) (2006) 115–125.
- [5] J. Ferber, O. Gutknecht, F. Michel, From Agents to Organizations: an Organizational View of Multi-Agent Systems, in: *Agent-Oriented Software Engineering VI*, Vol. 2935 of LNCS, Springer-Verlag, 2004, pp. 214–230.
- [6] E. Oliva, P. McBurney, A. Omicini, Co-argumentation artifact for agent societies, in: *5th International Workshop on Argumentation in Multi-Agent Systems*, ArgMAS-08, 2008.
- [7] T. Bench-Capon, K. Atkinson, *Argumentation in Artificial Intelligence*, 2009, Ch. Abstract argumentation and values, pp. 45–64.
- [8] A. S. d’Avila Garcez, D. M. Gabbay, L. C. Lamb, Value-based Argumentation Frameworks as Neural-Symbolic Learning Systems, *Journal of Logic and Computation* 15 (6) (2005) 1041–1058.
- [9] A. Abraham, E. Corchado, J. Corchado, Hybrid learning machines, *Neurocomputing* 72 (13-15) (2009) 2729–2730.
- [10] V. Dignum, Phd dissertation: A model for organizational interaction: based on agents, founded in logic, Ph.D. thesis (2003).
- [11] N. Criado, E. Argente, V. Botti, A Normative Model For Open Agent Organizations, in: *International Conference on Artificial Intelligence, ICAI-09*, 2009.
- [12] P. Baroni, M. Giacomin, *Argumentation in Artificial Intelligence*, Springer, 2009, Ch. Semantics of Abstract Argument Systems, pp. 25–44.
- [13] V. Botti, A. Garrido, A. Giret, P. Noriega, Managing water demand as a regulated open MAS, in: *Workshop on Coordination, Organization, Institutions and Norms in agent systems in on-line communities, COIN-09*, Vol. 494, 2009, pp. 1–10.
- [14] R. Brena, J. Aguirre, C. Chesñevar, E. Ramírez, L. Garrido, Knowledge and information distribution leveraged by intelligent agents, *Knowledge and Information Systems* 12 (2) (2007) 203–227.
- [15] W. Makiguchi, H. Sawamura, A Hybrid Argumentation of Symbolic and Neural Net Argumentation, in: *4th International Workshop on Argumentation in Multi-Agent Systems*, Vol. 4946 of LNCS, Springer, 2007, pp. 197–233.
- [16] A. d’Avila Garcez, L. Lamb, D. Gabbay, Argumentation Frameworks as Neural Networks, in: *Neural-Symbolic Cognitive Reasoning, Cognitive Technologies*, Springer, 2009, pp. 143–159.