DebateWEL: An interface for Debating With Enthymemes and Logical formulas*

Julien Balax Florence Dupin de Saint-Cyr
David Villard
IRIT, Université Paul Sabatier, 31062 Toulouse Cedex 9, France
julienbalax@gmail.com, florence.dupin@irit.fr,david0271@ymail.com

September 17, 2012

Abstract

The DebateWEL G.U.I. allows two players to exchange logical formulas and enthymemes in a persuasion debate game. DebateWEL protocol, described in [1], ensures consistency of each player with respect to himself and common-knowledge thanks to a SAT solver [2] which is called directly from the SAToulouse interface [3]. The game ends either because the players have found an agreement or because the limit of time has been reached (failure of the debate).

Being persuasive is very useful in many situations. Indeed good orators are considered to be very clever, hence an interesting challenge for AI is to be able to design artificial persuasive orators. Before proposing such artificial agents, it is necessary to define the framework in which they are going to play.

Persuasion dialogs are particular dialogs in which one agent aims at convincing others that a first assertion (called the subject of the dialog) holds. In order to design a framework in which a persuasion dialog can take place, it is necessary to define the moves that each agent is allowed to make. This definition is called a *protocol*. In persuasion dialogs, the possible types of moves are mainly assertions and challenges while in other kinds of dialogs (*e.g.* negotiations dialogs) moves like requirements or proposals are allowed. Apart from its type, a move is also characterized by its content. In a persuasion dialog this content is an assertion or an argument (*i.e.*, a rational utterance explaining the reasons for a given claim). Here are the specificities of DebateWEL protocol:

- There is no information about the state of mind of the agents, only what is publicly said is considered.
- The contents of the moves are based on classical propositional logic.
- Arguments are not abstract entities but explicit pairs (S, φ) where the support S is a set of formulas and the claim φ is a formula. Moreover, pairs in which S is not a proof of φ are allowed. If something is missing to prove φ from S or if φ is too vague then (S, φ) is called *enthymeme for* a given perfect argument that should have a more complete support or a more precise claim. As explained in [1, 4, 5]), enthymemes are very common in human dialogs.
- At the beginning of the dialog the agents agree on *a common knowledge* on which they can base their future statements, it is expressed under the form of formulas and enthymemes (they can agree on a set of approximate arguments that they consider as sufficient "proofs"). This common knowledge may only increase during the dialog.
- The protocol is *flexible*: any agent can take is turn as soon as the previous move is finished (even by himself), and can utter *any* move provided that it does not introduce any *self-contradiction* or

*This work was funded by the ANR project LELIE on risk analysis and prevention (http://www.irit.fr/recherches/ILPL/lelie/accueil.html). This is a draft version, the article was published In : JELIA'2012, The original publication is available at www.springerlink.com.

repetition, and that, before the end, *all the commitments* induced by the moves of his adversary are fulfilled.

- The dialog ends either with the victory of one agent when his adversary has agreed with him, or with a failure when they did not succeed to fulfill their commitment in time (either expressed in seconds or in terms of number of symbols used). DebateWEL is designed for inciting agents rather to agree than to reach a failure of the dialog: the score is high for a persuasive player, middle for a persuaded player and bad if the dialog fails.

Let us recall definitions from [1]: $\mathscr{L}$ being a propositional language, an *approximate argument* is a pair (S, φ) where $S \subseteq \mathscr{L}$ and $\varphi \in \mathscr{L}$. A *logical argument* (S, φ) is such that: (1) $S \subset \mathscr{L}$, (2) $S \nvdash \bot$, (3) $S \vdash \varphi$, and (4) $\nexists S' \subset S$ s.t. $S' \vdash \varphi$. (S, φ) and (S', φ') being approximate arguments, (S', φ') *completes* (S, φ) iff (1) $S \subset S'$ and $\varphi = \varphi'$ or (2) $S \subseteq S'$ and $\{\varphi'\} \cup S' \vdash \varphi$ and $\varphi \neq \varphi'$. a is an *enthymeme for* a' iff a' is a logical argument and a' completes a (a is said *incomplete*)

We consider a set of eleven speech acts, the six usual speech acts used in persuasion dialog: Accept for accepting a formula, Argue for uttering an argument (which maybe incomplete), Assert for uttering a formula, Challenge for asking for an argument explaining a formula, Close to end the dialog and Retract for removing an asserted formula. Two speech acts that are specific for enthymemes are added to this set, namely Quiz and Agree (proposed by [6]) for asking for a completion of an argument and for agreeing with the fact that a given pair (S, φ) is sufficient to convince that φ holds. We also add three more moves: Quizlink enabling to ask for a completed argument that relates its claim to the subject of the dialog, Replace allowing to complete an argument and Dismantle for retracting an argument. In [1], it is assumed that speech acts commit either the utterer (to be consistent and to be able to explain himself when challenged) and the hearer (to acknowledge that he agrees with the utterances of his adversary (unless they have been retracted)).

DebateWEL interface handles these commitments with a commitment store [7] divided into three parts (see Figure 1). On the center, the common knowledge part is divided into formulas and arguments. On each side of the screen, the commitment stores of each agent are divided into three parts: propositional formulas and approximate arguments asserted by the agent and commitments towards the other agent, *i.e.*, the requests to which he should answer. The preconditions of each move are translated into consistency checks (done by SAT4j) within the commitment store while the effects of the moves are addition to or removal from the commitment store (following the precise definitions of each move given in [1]).

In the screen shot of Figure 1, the first player is named Schopenhauer, he has asserted $drama \rightarrow supreme$ meaning "In drama, English are supreme", the common knowledge is $\{opera \rightarrow music, opera\}$ meaning that the two players agree on the fact that "opera is music" and that "opera exists". After this first move, the second player named "Adversary" is commited to Accept the formula $drama \rightarrow supreme$ (as long as Schopenhauer has not retracted this assertion). Schopenhauer has been permitted to do it because it is consistent with what he had already said (nothing) and with common knowledge. Checking consistency is done by using SAT4j. The interface for entering formulas is based on SAToulouse as shown on Figure 2 which represents the architecture of DebateWEL.

DebateWEL has several software features:

- The hand is materialized by a buzzer, when it is green the agent can buzz. After buzzing, the buzzer becomes orange and the current agent can play its move while his adversary must wait (the moves that are not possible are grayed button, if the formula or argument entered is not consistent or redundant then an error message is delivered). In Figure 1, the two agents can buzz, the quicker to buzz will be allowed to play.
- In order to help the players, they can choose to load one or several files containing formulas concerning a given topic, it is then possible to select directly these formulas instead of entering them, moreover it is also possible to select among formulas that have already been asserted in the dialog.
- A concise view of arguments is proposed: by default, only the claim is shown (the support can be fold and unfold simply by clicking on it).

- The type of the move of a request is symbolized by a colored bullet and the meaning of the colors is displayed in the center of the screen (see Figure 1).
- The score at the end of the dialog is based on the number of the player's formulas present in the common knowledge at the end of the dialog, a big penalty is associated to a player who has not fulfilled his commitments.
- There are two options for the limitation of the dialog: either the time taken by the agent or the number of symbols used in his utterances.
- At the end of the game the common-knowledge is stored in a file.
- DebateWEL enables the user to load an old game, thus it is possible to stop the dialog and save it (in order to continue later).

In the near future, we plan to develop an artificial player on the basis of an A* algorithm, and we also want to make this application available on the web. Indeed if this application is used, then it can help to collect new arguments (hence enrich a set of argument benchmarks, something that is really needed in the field of Argumentation theory).

References

[1] Dupin de Saint-Cyr, F.: Handling enthymemes in time-limited persuasion dialogs. In Benferhat, S., Grant, J., eds.: Int. Conf. on Scalable Uncertainty Management - SUM, Dayton (Ohio). Number 6929 in LNAI, Springer-Verlag (2011) 149–162

[2] Le Berre, D., Parrain, A.: The Sat4j library, release 2.2. Journal on Satisfiability, Boolean Modeling and Computation (2010) 59–64

[3] Gasquet, O., Schwarzentruber, F., Strecker, M.: SAToulouse: the computational power of propositional logic shown to beginners. In Blackburn, P., van Ditmarsch, H., Manzano, M., Soler-Tosca, F., eds.: International Congress on Tools for Teaching Logic. Volume 6680 of LNCS., Springer (2011) 77–84

[4] Besnard, P., Hunter, A.: A logic-based theory of deductive arguments. Artificial Intelligence **128**(1-2) (2001) 203 – 235

[5] Hunter, A.: Real arguments are approximate arguments. In: Proceedings of the 22nd National Conference on Artificial intelligence, AAAI Press (2007) 66–71

[6] Black, E., Hunter, A.: Using enthymemes in an inquiry dialogue system. In Padgham, L., Parkes, D., eds.: Proceedings of the 7th international joint conference on Autonomous agents and multiagent systems - AAMAS, International Foundation for Autonomous Agents and Multiagent Systems Richland, SC (2008) 437–444

[7] Hamblin, C.: Fallacies. Methuen, London (1970)

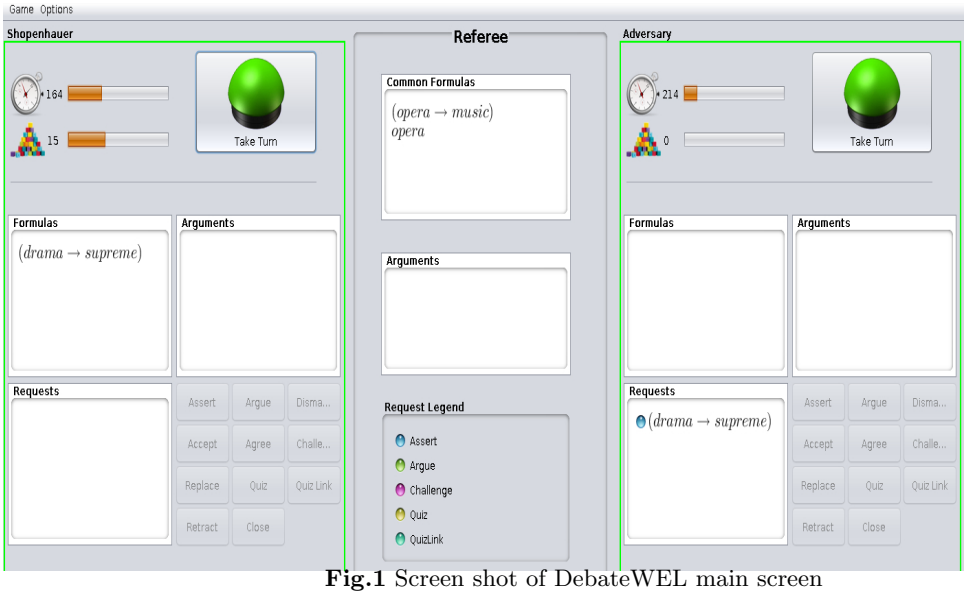

Fig.1 Screen shot of DebateWEL main screen

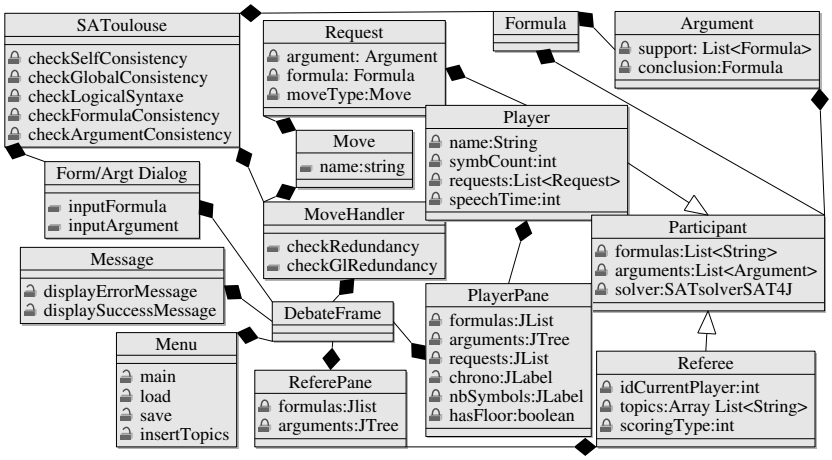

Fig.2 Simplified Architecture of DebateWEL