Introducing the Second International Competition on Computational Models of Argumentation

Sarah Alice GAGGL [a], Thomas LINSBICHLER [b], Marco MARATEA [c] and
Stefan WOLTRAN [b]

[a] *Computational Logic Group, Technische Universität Dresden, Germany*
[b] *Institute of Information Systems, TU Wien, Austria*
[c] *DIBRIS, University of Genoa, Italy*

Abstract. Argumentation is a major topic in the study of artificial intelligence. In particular, the problem of solving certain reasoning tasks of Dung's abstract argumentation frameworks is central to many advanced argumentation systems. The fact that problems to be solved are mostly intractable requires efficient algorithms and solvers, that are to be evaluated on meaningful benchmarks. In this report, we introduce the Second International Competition on Computational Models of Argumentation (ICCMA'17), which is jointly organized by TU Dresden (Germany), TU Wien (Austria), and the University of Genoa (Italy), in affiliation with the 2017 International Workshop on Theory and Applications of Formal Argumentation (TAFA'17).

Keywords. Abstract argumentation, computational models, competition

1. Introduction

Argumentation [1] is a major topic in the study of artificial intelligence. In particular, the problem of solving certain reasoning tasks of Dung's abstract argumentation frameworks [2] is central to many advanced argumentation systems. The fact that problems to be solved are mostly intractable requires efficient algorithms and solvers, that are to be evaluated on meaningful benchmarks. Another unique feature of abstract argumentation is the fact that solvers are expected to handle different semantics. This makes the design of competitions quite different to other comparable events, for instance in the field of Propositional logic (SAT) or Answer-Set Programming (ASP).

In this report, we introduce the Second International Competition on Computational Models of Argumentation (ICCMA'17), which is jointly organized by TU Dresden (Germany), TU Wien (Austria), and the University of Genoa (Italy), in affiliation with the 2017 International Workshop on Theory and Applications of Formal Argumentation (TAFA'17). ICCMA'17 will be conducted in the first half of 2017, and comes two years after the first edition, ICCMA'15 [3].[1]

---

[1] http://argumentationcompetition.org/2015

The general goal of this competition is to consolidate and strengthen the ICCMA series, which in its first edition had very good outcomes in some respects, e.g. in terms of the number of submitted solvers (18). The second edition will maintain some of the design choices previously made. In particular, we will keep the I/O formats as well as the basic reasoning problems. With a slight modification to the first edition, the competition will be organized into *tasks* and *tracks*, where a *task* is a reasoning problem under a particular semantics, and a *track* collects different tasks over a semantics. On the other hand, ICCMA'17 introduces some novelties:

- a new scoring scheme is implemented for better reflecting the solvers' behavior,
- three new semantics will be included, namely semi-stable, stage and ideal semantics,
- a special "Dung's Triathlon" track is added, where solvers are required to deal with different problems simultaneously; here, the goal is to test the solvers' capability of exploiting interrelationships between semantics, and
- a "call for benchmarks" has been announced, to enrich the suite of instances for the competition.

The present report is structured as follows. First, Section 2 introduces needed preliminaries about abstract argumentation frameworks. Then, Section 3 presents the design of ICCMA'17 in terms of tracks definition and scoring scheme, while Section 4 discusses about the benchmarks that will be used in the competition. Some final remarks are given in Section 5.

2. Background

An *abstract argumentation framework* (AF, for short) [2] is a tuple $\mathscr{F} = (A, \rightarrow)$ where A is a set of arguments and $\rightarrow$ is a relation $\rightarrow \subseteq A \times A$. For two arguments $a, b \in A$ the relation $a \rightarrow b$ means that argument *a attacks* argument *b*. An argument $a \in A$ is *defended* by $S \subseteq A$ (in $\mathscr{F}$) if for each $b \in A$ such that $b \rightarrow a$ there is some $c \in S$ such that $c \rightarrow b$. A set $E \subseteq A$ is *conflict-free* (in $\mathscr{F}$) if and only if there are no $a, b \in E$ with $a \rightarrow b$. E is *admissible* (in $\mathscr{F}$) if and only if it is conflict-free and each $a \in E$ is defended by E. Finally, the range of E (in $\mathscr{F}$) is given by $E_{\mathscr{F}}^+ = E \cup \{a \in A \mid \exists b \in E : b \rightarrow a\}$.

Semantics are used to determine sets of jointly acceptable arguments by mapping each AF $\mathscr{F} = (A, \rightarrow)$ to a set of *extensions* $\sigma(\mathscr{F}) \subseteq 2^A$. The extensions under complete, preferred, stable, semi-stable [4], stage [5], grounded and ideal [6] semantics are defined as follows. Given an AF $\mathscr{F} = (A, \rightarrow)$ and a set $E \subseteq A$,

- $E \in \mathbf{CO}(\mathscr{F})$ iff E is admissible in $\mathscr{F}$ and if $a \in A$ is defended by E then $a \in E$,
- $E \in \mathbf{PR}(\mathscr{F})$ iff $E \in \mathbf{CO}(\mathscr{F})$ and there is no $E' \in \mathbf{CO}(\mathscr{F})$ s.t. $E' \supset E$,
- $E \in \mathbf{ST}(\mathscr{F})$ iff $E \in \mathbf{CO}(\mathscr{F})$ and $E_{\mathscr{F}}^+ = A$,
- $E \in \mathbf{SST}(\mathscr{F})$ iff $E \in \mathbf{CO}(\mathscr{F})$ and there is no $E' \in \mathbf{CO}(\mathscr{F})$ s.t. $E_{\mathscr{F}}'^+ \supset E$,
- $E \in \mathbf{STG}(\mathscr{F})$ iff E is conflict-free in $\mathscr{F}$ and there is no E' such that E' is conflict-free in $\mathscr{F}$ and $E_{\mathscr{F}}'^+ \supset E$,
- $E \in \mathbf{GR}(\mathscr{F})$ iff $E \in \mathbf{CO}(\mathscr{F})$ and there is no $E' \in \mathbf{CO}(\mathscr{F})$ s.t. $E' \subset E$,
- $E \in \mathbf{ID}(\mathscr{F})$ iff E is admissible in $\mathscr{F}$, $E \subseteq \bigcap \mathbf{PR}(\mathscr{F})$ and there is no $E' \subseteq \bigcap \mathbf{PR}(\mathscr{F})$ s.t. E' is admissible in F and $E' \supset E$.

For more discussion on these semantics we refer to [1].

Note that both grounded and ideal extensions are uniquely determined and always exist [2,6]. Thus, they are also called *single-status* semantics. The other semantics introduced are *multi-status* semantics. That is, there is not always a unique extension induced by the semantics. In order to reason with multi-status semantics, usually, one takes either a credulous or skeptical perspective. That is, given an AF $\mathscr{F} = (\mathsf{A}, \rightarrow)$ and a semantics $\sigma \in \{\mathbf{CO}, \mathbf{PR}, \mathbf{ST}, \mathbf{SST}, \mathbf{STG}, \mathbf{GR}, \mathbf{ID}\}$, argument $a \in \mathsf{A}$ is

- *credulously accepted* in $\mathscr{F}$ under semantics σ if there is a σ-extension $E \in \sigma(\mathscr{F})$ with $a \in E$, and
- *skeptically accepted* in $\mathscr{F}$ with semantics σ if for all σ-extensions $E \in \sigma(\mathscr{F})$ it holds that $a \in E$.

Recall that stable semantics is the only case where an AF might possess no extension. In such a situation, each argument is defined to be skeptically accepted.

3. Competition Format & Rules

The competition will feature seven main tracks, one for each semantics. Each of these tracks is composed of 4 (resp. 2 for single-status semantics) tasks, one for each reasoning problem. A special track, Dung's Triathlon, will be conducted in order to enumerate three semantics simultaneously.

Each solver participating in the competition can support, i.e. compete in, an arbitrary set of tasks. If a solver supports all tasks of a track, it also participates in the track.

In the following we describe the rules that will be applied to the evaluation of the tasks and tracks of the competition. For detailed information about input- and output-format as well as requirements to the solver interface we refer the reader to

    http://www.dbai.tuwien.ac.at/iccma17/SolverRequirements.pdf.

3.1. Tasks

A *task* is a reasoning problem under a particular semantics. Following ICCMA'15 we consider four different problems:

DC-σ *Given* $\mathscr{F} = (\mathsf{A}, \rightarrow)$ and $a \in \mathsf{A}$, *decide* whether a is credulously accepted in $\mathscr{F}$ under σ.

DS-σ *Given* $\mathscr{F} = (\mathsf{A}, \rightarrow)$ and $a \in \mathsf{A}$, *decide* whether a is skeptically accepted in $\mathscr{F}$ under σ.

SE-σ *Given* $\mathscr{F} = (\mathsf{A}, \rightarrow)$, *return* some set $E \subseteq \mathsf{A}$ that is a σ-extension of $\mathscr{F}$.

EE-σ *Given* $\mathscr{F} = (\mathsf{A}, \rightarrow)$, *enumerate* all sets $E \subseteq \mathsf{A}$ that are σ-extensions of $\mathscr{F}$.

for the seven semantics $\sigma \in \{\mathbf{CO}, \mathbf{PR}, \mathbf{ST}, \mathbf{SST}, \mathbf{STG}, \mathbf{GR}, \mathbf{ID}\}$.

For single-status semantics (**GR** and **ID**) some problems collapse, i.e. **SE** and **EE** require to compute the unique extension; and **DC** and **DS** are equivalent. Thus, for **GR** and **ID** only the problems **SE** and **DC** are considered.

The combination of problems with semantics amounts to a total number of 26 tasks. Each solver participating in a task will be queried with a fixed number[2] of instances corresponding to the task with a timeout of 10 minutes each. For each instance, the solvers get

- 1 point, if it delivers the *correct* result;
- -5 points, if it delivers an *incorrect* result (note that in contrast to ICCMA'15, wrong answers are now penalized);
- 0 points otherwise.

The terms *correct* and *incorrect* for the different reasoning problems are defined as follows.

- **DC**-σ (resp. **DS**-σ): if the queried argument is credulously (resp. skeptically) accepted in the given AF under σ, the result is *correct* if it is YES and *incorrect* if it is NO; if the queried argument is not credulously (resp. not skeptically) accepted in the given AF under σ, the result is *correct* if it is NO and *incorrect* if it is YES.
- **SE**-σ: the result is *correct* if it is a σ-extension of the given AF and *incorrect* if it is a set of arguments that is not a σ-extension of the given AF. If the given AF has no σ-extensions, then the result is *correct* if it is NO and *incorrect* if it is any set of arguments.
- **EE**-σ: the result is *correct* if it is the set of all σ-extensions of the given AF and *incorrect* if it contains a set of arguments that is not a σ-extension of the given AF.

Intuitively, a result is neither correct nor incorrect (and therefore gets 0 points) if (i) it is empty (e.g. the timeout was reached without answer) or (ii) it is not parsable with respect to the required output format (e.g. due to some unexpected error message). For **EE**-σ there is also the case that the result (iii) contains σ-extensions, but not all of them.

The *score* of a solver for a particular task is the sum of points over all instances. The ranking of solvers for the task is then based on the scores in descending order. Ties are broken by the total time it took the solver to return correct results.

For semi-stable and stage semantics, we recall that those semantics coincide with stable for AFs that possess at least one stable extension. Benchmarks will be selected such that the majority of AFs does not have a stable extension.

3.2. Tracks

All tasks for a particular semantics constitute a *track*. Therefore there is one track for each semantics.

The ranking of solvers for a track is based on the sum of scores over all tasks of the track. Again, ties are broken by the total time it took the solver to return correct results.

Note that in order to make sure that each task has the same impact on the evaluation of the track, all tasks for one semantics will have the same number of instances.

The winner of each track will be awarded.

---

[2]The exact number will be determined later.

3.3. Dung's Triathlon

In this special track, different enumeration tasks are joined together. The aim of this track is to evaluate solvers also with respect to their capability of exploiting interrelationships between different semantics.

The problem to solve in this track is defined as follows.

D3 *Given* $\mathscr{F} = (\mathsf{A}, \rightarrow)$, *enumerate*

- all sets $E \subseteq \mathsf{A}$ that are **GR**-extensions[3] of $\mathscr{F}$, followed by
- all sets $E \subseteq \mathsf{A}$ that are **ST**-extensions of $\mathscr{F}$, followed by
- all sets $E \subseteq \mathsf{A}$ that are **PR**-extensions of $\mathscr{F}$.

For scoring and ranking the same method as for single tasks is applied, except that the timeout for each instance is 30 minutes.

The result of **D3** is *correct* if it is the set of all **GR**-extensions, followed by the set of all **ST**-extensions, followed by the set of all **PR**-extensions, and *incorrect* if the first set contains a set of arguments that is not the **GR**-extension, the second set contains a set of arguments that is not the **ST**-extension, or the third set contains a set of arguments that is not the **PR**-extension.

It is up to the system how the enumeration of the extensions of different semantics is performed (for instance, computing all complete extensions and output the subset-minimal and subset-maximal ones for the grounded, resp. preferred extensions; or alternatively, just calling three independent routines for the three semantics under consideration). However, please follow strictly the output format as outlined in `http://www.dbai.tuwien.ac.at/iccma17/SolverRequirements.pdf`.

The winner of Dung's Triathlon will be awarded.

4. Benchmarks

ICCMA'17 will take advantage, for the first time, of a dedicated *call for benchmarks*, which is customary in other competitions. The goal of this call is to enlarge the set of domains that will be considered, and thus possibly having a more heterogeneous set of benchmarks (e.g., random, crafted, application-oriented) in the evaluation. Contributors will be asked to provide an instance set for the benchmark they are submitting, and/or an instance generator, possibly with an indication about the estimated difficulty of the instances.

Submitted benchmark instances must follow the input formats as outlined in `http://www.dbai.tuwien.ac.at/iccma17/SolverRequirements.pdf`. All submitted instances, together with those of ICCMA'15, will become part of the suite of instances made available to the community (after the competition), and a selection of the suite will be used to evaluate solvers at ICCMA'17. Such selection will be (also) based on a pre-selection, whose goal is to have evaluated mostly "meaningful" instances, i.e. instances that are neither too easy nor too hard, and can thus be useful for solver performance comparisons.

---

[3]Although grounded semantics is a single-status semantics, we treat it here like a multi-status semantics for the sake of uniformity.

5. Discussion

The calls for solvers and benchmarks have been spread in July 2016. In accordance with the calls, each solver will have to be registered by solver contributors, that must also specify on which tasks and tracks the solver competes. Then, the evaluation process will consist of two phases: Initially, the competitors will be given a set of representative AFs to test their solvers on their own machines. Then, they will submit a final version of the source code of their solvers that will be run by the organizers on the actual competition machines, and instances (unknown to the competitors until this time).

We hope that the second edition of the competition will not only consolidate and strengthen the ICCMA series but also contribute to the advancement of research concerned with computation models of argumentation.

Lastly, the design of the competition is to be considered still preliminary: The organizers reserve the right to make changes if needed. Important changes will be shared and discussed with the ICCMA steering committee. Up-to-date information about the competition can be found at

    http://www.dbai.tuwien.ac.at/iccma17.

References

[1] Baroni, P., Caminada, M., Giacomin, M.: An introduction to argumentation semantics. The Knowledge Engineering Review **26** (2011) 365–410

[2] Dung, P.M.: On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. Artificial Intelligence **77** (1995) 321–358

[3] Thimm, M., Villata, S., Cerutti, F., Oren, N., Strass, H., Vallati, M.: Summary report of the first international competition on computational models of argumentation. AI Magazine **37** (2016) 102–104

[4] Caminada, M.W.A., Carnielli, W.A., Dunne, P.E.: Semi-stable semantics. Journal of Logic and Computation **22** (2012) 1207–1254

[5] Verheij, B.: Two approaches to dialectical argumentation: admissible sets and argumentation stages. In: Proceedings of the 8th Dutch Conference on Artificial Intelligence (NAIC'96). (1996) 357–368

[6] Dung, P., Mancarella, P., Toni, F.: Computing ideal sceptical argumentation. Artificial Intelligence **171** (2007) 642–674