Summary Report of the Second International Competition on Computational Models of Argumentation

Sarah A. Gaggl, Thomas Linsbichler, Marco Maratea, Stefan Woltran

We give an overview of design and results of the Second International Competition on Computational Models of Argumentation (ICCMA'17). Following the first edition in 2015, the competition evaluates the performance of submitted solvers on computational problems within abstract argumentation. In addition to the four original semantics, ICCMA'2017 includes three additional prominent semantics. Moreover, a dedicated call for benchmarks allowed for introducing a sophisticated instance selection process.

Introduction

Argumentation is a major topic in the study of artificial intelligence (Bench-Capon and Dunne 2007; Atkinson et al. 2017). In particular, the problem of solving certain reasoning tasks of Dung's abstract argumentation frameworks (Dung 1995) 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 briefly present design and results of the Second International Competition on Computational Models of Argumentation (ICCMA'17)[1], which has been 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 has been conducted in the first half of 2017, and comes two years after the first edition, ICCMA'15 (Thimm et al. 2016).[2]

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 maintains some of the design choices previously made, e.g. the I/O formats and the basic reasoning problems. With a slight modification to the first edition, the competition is 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. ICCMA'17 also introduces several novelties: (*i*) a new scoring scheme is implemented for better reflecting the solvers' behavior, (*ii*) three new semantics are included, namely semi-stable, stage and ideal semantics, (*iii*) a special "Dung's Triathlon" track is added, where solvers are required to deal with different problems simultaneously, with the goal of testing the solvers' capability of exploiting interrelationships between semantics, and (*iv*) a "call for benchmarks" has been performed, to enrich the suite of instances for the competition, followed by a novel instance selection stage.

Background and Format

An *abstract argumentation framework* (AF, for short) (Dung 1995) is a tuple $\mathscr{F} = (\mathsf{A}, \rightarrow)$ where A is a set of arguments and $\rightarrow \subseteq \mathsf{A} \times \mathsf{A}$ is the attack relation. *Semantics* are used to determine sets of jointly acceptable arguments by mapping each AF to a set of *extensions* $\sigma(\mathscr{F}) \subseteq 2^{\mathsf{A}}$ (see (Baroni, Caminada, and Giacomin 2011) for an overview). The main underlying concepts of semantics are conflict-freeness and admissibility. The semantics considered in the competition are *grounded*, *complete*, *preferred*, *stable* (Dung 1995), *semi-stable* (Caminada, Carnielli, and Dunne 2012), *stage* (Verheij 1996), and *ideal* (Dung, Mancarella, and Toni 2007), the last three being considered for the first time in the ICCMA series.

Following ICCMA'15, we consider four reasoning problems: *skeptical* and *credulous acceptance*; and computing a *single* and *all extensions*. The complexity of these problems under the considered semantics ranges from polynomial time to intractability in the second level of the polynomial hierarchy (Dunne and Wooldridge 2009; Dvořák 2012).

The competition features seven main tracks, one for each semantics. Each of these tracks is composed of 4 (resp. 2 for grounded and ideal semantics, given they are single-status) tasks, one for each reasoning problem. The combination of reasoning problems with semantics amounts to a total number of 24 tasks.

A special, 8th, track, the Dung's Triathlon, is conducted in order to enumerate three types of extensions, namely grounded, stable, and preferred, simultaneously.

Participants

Sixteen solvers participate in the competition, 10 of which are new entries compared to the previous edition. The solvers originate from 14 different teams from Austria, China, Finland, France, Germany, Italy, Jordan, and the UK. Each solver can compete in an arbitrary set of tasks. If a solver supports all tasks of a track, it also participates in the track. This results in each task featuring at least 9 participating solvers, and 8 solvers participating in all tracks. The solvers participating in this second event are based on a wide variety of solving approaches, ranging from direct approaches to (different forms of) reductions to SAT, ASP, CSP, and circumscription.

All solver submissions are accompanied with a system description and the full source code, in order to ensure maximal transparency and accessibility to the community.

Benchmarks and Selection

ICCMA'17 takes advantage, for the first time, of a dedicated *call for benchmarks*, which is customary in other competitions. We have received six submissions, among them AF generators as well as concrete sets of AFs. The latter include collections of (a) AFs instantiated from assumption-based argumentation, (b) AFs translated from planning problems, and (c) AFs obtained from traffic network graphs. The submitted generators allow to produce AFs that are crafted to be challenging for (d) strong admissibility and (e) semi-stable semantics, as well as (f) AFs from well-known graph classes from the literature (Barabasi-Albert, Erdös-Rényi, and Watts-Strogatz). With the generators, we produced instances aiming to cover a possibly broad range of difficulty.

Together with the generators from ICCMA'15, namely GroundedGenerator, SccGenerator, and StableGenerator (see (Thimm and Villata 2017)), these sets contribute to the benchmark suite of ICCMA'17, for a total of 3990 instances in 11 domains. The benchmark suite includes a heterogeneous set of benchmarks, i.e. random, crafted, and application-oriented.

Starting from this suite, a benchmark selection process has been applied to select the instances that are indeed run in the competition. Following related competitions, e.g. SAT and ASP competitions (see (Belov et al. 2014; Gebser, Maratea, and Ricca 2017) for details), but for the first time in the ICCMA series, we have selected instances based on their expected hardness, in order to have a benchmark suite covering a wide variety of expected difficulties. Given the high number of tasks and tracks in the competition, we have grouped the tasks already evaluated in ICCMA'15 into three groups, based on the complexity of the tasks. For each of these groups, we have classified all instances into 5 hardness categories, from "very easy" to "too hard", according to the performance of three among the best solvers from ICCMA'15 in a "representative" task within the group. The selection of these solvers ensures that they implement different solving approaches in order not to have biased results. For the tasks related to newly introduced semantics and for Dung's Triathlon, we considered the classification obtained by the group containing the tasks of highest complexity. Then, after the classification, 350 instances per group distributed over the hardness categories are selected, at the same time ensuring that instances are also distributed over domains.

Finally, for the acceptance tasks, and considering that the number of instances has to be constant among tasks, we select only one argument for each instance, with the exception that we drop the "very easy" instances for acceptance tasks, and select two arguments to be queried for the "too hard" instances.

Results

The winner of each track has been awarded. For each track, the score of a solver is obtained by the sum of scores over all tasks of the track, each of them obtained by the sum of points over all instances. For each instance, a solver gets (i) 1 point if it delivers the *correct* result within 600 (resp. 1800 for Dung's Triathlon) seconds CPU time, (ii) -5 points if it delivers an incorrect result, and (iii) 0 points otherwise. It has to be noted that the points assigned have changed: to incorrect results we now assign a negative reward, while in ICCMA'15 those were assigned 0 points. This change has been applied to put focus on correctness of solvers and to prevent solvers from guessing answers. We think in this way the final score better reflects the solvers' behavior. Correctness of results was verified by comparing the results to reference solutions by ASPARTIX (a reliable solver from ICCMA'15, see (Egly,

Track	Solver	Points
Complete	pyglaf	1229/1400
Preferred	ArgSemSAT	1146/1400
Stable	pyglaf	1183/1400
Semi-Stable	argmat-sat	1164/1400
Stage	argmat-sat	1065/1400
Grounded	CoQuiAAS	695/700
Ideal	pyglaf	585/700
Dung's Triathlon	argmat-dvisat	276/350

Table 1: Award winners.

Gaggl, and Woltran 2010)), dedicated ASP encodings for checking single extensions, and comparing solutions between solvers.

Overall, the winner of a track is the solver that gets the highest score. Ties are broken by the total time it took the solver to return correct results.

Table 1 lists the award winner of each track, by showing the semantics in the first column, the winner of the related track in the second column, and the number of points achieved by the winner in the third column. The exception is the last row, where the winner of the special Dung's Triathlon track is presented. The list of winners reflects the diversity of the solving approaches in the competition already noticed before and ultimately confirms the usefulness of having such a wide variety of approaches. Indeed, the 5 winning solvers implement concepts that take advantages of SAT (under different integration schemas), circumscriptions, and CSP techniques. For detailed results, see `http://argumentationcompetition.org/2017/results.html`.

Conclusions and Outlook

The fact that two thirds of tracks have been won by solvers newly introduced at ICCMA'17 shows that the field of computational models of argumentation is not only vibrant but also highly amenable for further improvements and innovation. Moreover, considering that pyglaf (winner of 3 tracks) uses a novel approach based on reduction to circumscription indicates that even more variety of solving techniques can be fruitful for the development of the field.

While we think that future editions of ICCMA should stick to a guided instance selection process as described in this report, the community should aim for benchmarks from real-world domains to be included in future benchmark suites. On the technical side, changing the output format for enumeration tasks could be beneficial for the verification of large solutions.

The next edition of the competition will be conducted in 2019, for more information refer to `http://argumentationcompetition.org/`.

Acknowledgments

This work has been supported by FWF (project I2854).

References

Atkinson, K.; Baroni, P.; Giacomin, M.; Hunter, A.; Prakken, H.; Reed, C.; Simari, G. R.; Thimm, M.; and Villata, S. 2017. Towards artificial argumentation. *AI Magazine* 38(3):25–36.

Baroni, P.; Caminada, M.; and Giacomin, M. 2011. An introduction to argumentation semantics. *The Knowledge Engineering Review* 26(4):365–410.

Belov, A.; Diepold, D.; Heule, M.; and Järvisalo, M. 2014. Proceedings of SAT competition 2014: Solver and benchmark descriptions. Department of Computer Science Series of Publications B, vol. B-2014-2.

Bench-Capon, T. J. M., and Dunne, P. E. 2007. Argumentation in artificial intelligence. *Artificial Intelligence* 171(10-15):619–641.

Caminada, M.; Carnielli, W.; and Dunne, P. E. 2012. Semi-stable semantics. *Journal of Logic and Computation* 22(5):1207–1254.

Dung, P. M.; Mancarella, P.; and Toni, F. 2007. Computing ideal sceptical argumentation. *Artificial Intelligence* 171(10–15):642–674.

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

Dunne, P. E., and Wooldridge, M. 2009. Complexity of abstract argumentation. In Rahwan, I., and Simari, G., eds., *Argumentation in Artificial Intelligence*. Springer. 85–104.

Dvořák, W. 2012. *Computational Aspects of Abstract Argumentation*. Ph.D. Dissertation, TU Wien, Institute of Information Systems.

Egly, U.; Gaggl, S. A.; and Woltran, S. 2010. Answer-set programming encodings for argumentation frameworks. *Argument & Computation* 1(2):147–177.

Gebser, M.; Maratea, M.; and Ricca, F. 2017. The sixth answer set programming competition. *Journal of Artificial Intelligence Research* 60:41–95.

Thimm, M., and Villata, S. 2017. The first international competition on computational models of argumentation: Results and analysis. *Artificial Intelligence* 252:267–294.

Thimm, M.; Villata, S.; Cerutti, F.; Oren, N.; Strass, H.; and Vallati, M. 2016. Summary report of the first international competition on computational models of argumentation. *AI Magazine* 37(1):102–104.

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

Sarah A. Gaggl is a postdoctoral research assistant at the Computational Logic Group at the TU Dresden. She received her PhD in Computer Science at TU Wien in 2013. Her main research interests are in abstract argumentation, answer set programming, and knowledge representation. Webpage: `https://iccl.inf.tu-dresden.de/web/Sarah_Alice_Gaggl`

Thomas Linsbichler is a postdoctoral researcher at the Institute of Information Systems of TU Wien, Austria. His main research interests are in knowledge representation and reasoning, argumentation, and algorithms. Webpage: `http://www.dbai.tuwien.ac.at/staff/linsbich/`

Marco Maratea is an associate professor in Computer Engineering at University of Genoa, Italy. In Fall 2015, 2016, and 2017 he was University Lecturer at the Institute for Information Systems of the Faculty of Informatics at the Vienna University of Technology. His research interests include artificial intelligence, logic programming, and knowledge representation and reasoning. Webpage: `http://www.star.dist.unige.it/marco/`

Stefan Woltran is professor of Foundations of Artificial Intelligence at Vienna University of Technology. His research focuses on problems in the area of knowledge representation and reasoning, argumentation, complexity analysis in artificial intelligence, and logic programming. In Winter Term 2013, he held a deputy professorship at Leipzig University. In 2013, he also received the prestigious START award from the Austrian Science Fund (FWF). Webpage: `http://www.dbai.tuwien.ac.at/staff/woltran/`