Advances in Debating Technologies: Building AI That Can Debate Humans

Roy Bar-Haim Liat Ein-Dor Matan Orbach Elad Venezian Noam Slonim
IBM Research
{roybar, liate, matano, eladv, noams}@il.ibm.com

1 Tutorial Description

1.1 Background and Goals

Argumentation and debating are fundamental capabilities of human intelligence. They are essential for a wide range of everyday activities that involve reasoning, decision making or persuasion. *Computational Argumentation* is defined as "the application of computational methods for analyzing and synthesizing argumentation and human debate" (Gurevych et al., 2016). Over the last few years, this field has been rapidly evolving, as evident by the growing research community, and the increasing number of publications in top NLP and AI conferences.

The tutorial focuses on *Debating Technologies*, a sub-field of computational argumentation defined as "computational technologies developed directly to enhance, support, and engage with human debating" (Gurevych et al., 2016). A recent milestone in this field is *Project Debater*, which was revealed in 2019 as the first AI system that can debate human experts on complex topics.[1] Project Debater is the third in the series of IBM Research AI's grand challenges, following Deep Blue and Watson. It has been developed for over six years by a large team of researchers and engineers, and its live demonstration in February 2019 received massive media attention. This research effort has resulted in more than 50 scientific papers to date, and many datasets freely available for research purposes.

In this tutorial, we aim to answer the question: "what does it take to build a system that can debate humans"? Our main focus is on the scientific problems such system must tackle. Some of these intriguing problems include argument retrieval for a given debate topic, argument quality assessment and stance classification, identifying relevant principled arguments to be used in conjunction with corpus-mined arguments, organizing the arguments into a compelling narrative, recognizing the arguments made by the human opponent and making a rebuttal. For each of these problems we will present relevant scientific work from various research groups as well as our own. Many of the underlying capabilities of Project Debater have been made freely available for academic research, and the tutorial will include a detailed explanation of how to use and leverage these tools.

A complementary goal of the tutorial is to provide a holistic view of a debating system. Such a view is largely missing in the academic literature, where each paper typically addresses a specific problem in isolation. We present a complete pipeline of a debating system, and discuss the information flow and the interaction between the various components. We will also share our experience and lessons learned from developing such a complex, large scale NLP system. Finally, the tutorial will discuss practical applications and future challenges of debating technologies.

1.2 Contents

In this section we provide more details about the contents of the tutorial. The tutorial outline and estimated schedule are listed in Section 3.

Introduction. The tutorial first provides an introduction to computational argumentation. It then introduces the Project Debater grand challenge and provides a high-level view of the building blocks that comprise a debating system.

The next parts of the tutorial describe each of these building blocks in depth.

Argument mining. The core of a debating system is *argument mining* – finding relevant arguments and argument components (claim/conclusion, evidence/premise) for a given

[1] https://www.research.ibm.com/artificial-intelligence/project-debater/

Proceedings of the Joint Conference of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing: Tutorial Abstracts, pages 1–5, August 1st, 2021.

debate topic, either in a given article, or in a large corpus.

Argument evaluation and analysis. The next tasks in the pipeline involve analysis of the extracted arguments. *Argument quality assessment* aims to select the more convincing arguments. *Stance classification* aims to distinguish between arguments that support our side in the debate and those supporting the opponent's side.

Modeling human dilemma. A complementary source for argumentation that is widely used by professional human debaters is *principled arguments*, which are relevant for a wide variety of topics. A common example is the *black market* argument, potentially relevant in the context of debates on banning a specific product or a service (e.g., *"we should ban alcohol"*). By this argument, imposing a ban leads to the creation of a black market, which in turn makes products or services obtained therein less safe, leads to exploitation, attracts criminal elements, and so on. We discuss recent work on creating a taxonomy of common principled arguments and automatically matching relevant arguments from this taxonomy to a given debate topic.

Listening comprehension and rebuttal. In addition to presenting one side of the debate, engaging in a competitive debate further requires a debating system to effectively rebut arguments raised by the human opponent. The system must listen to an argumentative speech in real-time, understand the main arguments, and produce persuasive counter-arguments.

The nature of the argumentation domain and the characteristics of competitive debates make the understanding of such spoken content challenging. Expressed ideas often span multiple, non-consecutive sentences and many arguments are alluded to rather than explicitly stated. Further difficulty stems from the requirement to identify and rebut the most important parts of a speech that is several minutes long. This contrasts with today's conversational agents, which aim at understanding a single functional command from short inputs.

Core NLP capabilities. This section describes several core NLP capabilities developed as part of Project Debater, including thematic clustering, highly scalable Wikification and semantic similarity for phrases and Wikipedia concepts.

From arguments to narrative. A debating system must arrange the arguments obtained from various sources (arguments mined from a corpus, principled arguments, and counter arguments for rebuttal) into a coherent and persuasive narrative that would keep the audience's attention for several minutes. This section describes the various steps in the narrative generation pipeline. We also discuss the role of humor in keeping a debate lively.

Moving forward – applications and implications. In this part we discuss possible applications and future directions for debating technologies. As an example, we present *Speech by Crowd*, a platform for crowdsourcing decision support. This platform collects arguments from large audiences on debatable topics and generates meaningful narratives summarizing the arguments for each side of the debate. We also discuss *Key Point Analysis*, a novel method for extracting the main points in a large collection of arguments, and quantifying the prevalence of each point in the data.

Demo session - using debating technologies in your application. Many of the Project Debater components presented in this tutorial have been recently released as cloud APIs, and are freely available for academic use.[2] In the final part of the tutorial, we provide an overview of these APIs, and demonstrate their use for building practical applications.

1.3 Relevance to the Computational Linguistics Community

The tutorial is relevant to a broad audience of NLP researchers and practitioners, working on problems related to argumentation mining, stance classification, discourse analysis, text summarization, NLG, dialogue systems, and more.

2 Tutorial Type

This is a *cutting-edge* tutorial. The main difference between this tutorial and previous tutorials on computational argumentation or argument mining (Slonim et al., 2016; Budzynska and Reed, 2019) is that we focus on the science behind *debating systems* — systems that can engage in a live debate with humans. Accordingly, a large portion of the tutorial's topics, e.g., listening comprehension, rebuttal, narrative generation and modeling

[2] https://early-access-program.debater.res.ibm.com/academic_use

human dilemma, was not covered in previous tutorials. Some of the topics, like argument mining, argument quality and stance classification were previously discussed in the tutorial of Slonim et al. (2016), however we will mostly focus on more recent advancements in these areas. The tutorial of Budzynska and Reed (2019) focused on argument structure parsing based on argumentation theory, which can be viewed as complementary to the content of the current tutorial.

3 Outline and Estimated Schedule

Part 1: Introduction (20 min)

- What is Computational Argumentation?
- Project Debater - AI that can debate human experts; outside the AI comfort zone
- Building blocks: decomposing the grand challenge

Part 2: Argument Mining (25 min)

- What is argument mining?
- Identification of argument components
- Document-level vs. sentence level approach
- Corpus-wide argument mining
- Debate topic expansion
- Token-level argument mining

Part 3: Argument Evaluation and Analysis (25 min)

- Argument stance classification
- Argument quality

Part 4: Modeling Human Dilemma (15 min)

- Common principled arguments
- When do principled arguments apply?

Coffee break

Part 5: Listening Comprehension and Rebuttal (25 min)

- Debate vs. classical conversation systems
- Understanding the gist of long, spontaneous speech
- From understanding to rebuttal

Part 6: Core NLP capabilities (10 min)

- Thematic clustering
- Wikification
- Multi-word and concept-level similarity

Part 7: From Arguments to Narrative (10 min)

- Narrative generation pipeline: argument filtering, redundancy removal, clustering, theme extraction, rephrasing and speech generation
- Keeping a live debate lively: the importance of humor

Part 8: Moving Forward – Applications and Implications (20 min)

- Possible applications
- Speech by crowd
- Key point analysis
- Future directions

Part 9: Demo Session - Using Debating Technologies in Your Application (30 min)

- Overview of Project Debater APIs
- Usage examples

4 Prerequisites

The tutorial will be self-contained. We assume basic knowledge of NLP and machine learning, at the level of introductory courses in these areas.

5 Reading List

1. A survey on argument mining: Lawrence and Reed (2019)
2. Project Debater: Slonim et al. (2021)
3. Identification of argument components within an article: Levy et al. (2014), Rinott et al. (2015), Lippi and Torroni (2015)
4. Corpus-wide argument mining: Stab et al. (2018), Ein-Dor et al. (2020)
5. Argument quality: Wachsmuth et al. (2017), Habernal and Gurevych (2016)
6. Stance classification: Bar-Haim et al. (2017)
7. Modeling human dilemma: Bilu et al. (2019)
8. Listening Comprehension: Mirkin et al. (2018)

6 Tutorial Presenters

- Roy Bar-Haim, IBM Research AI
  roybar@il.ibm.com
  https://researcher.watson.ibm.com/researcher/view.php?person=il-ROYBAR
- Liat Ein-Dor, IBM Research AI
  liate@il.ibm.com
  https://researcher.watson.ibm.com/researcher/view.php?person=il-LIATE
- Matan Orbach, IBM Research AI
  matano@il.ibm.com
  https://researcher.watson.ibm.com/researcher/view.php?person=il-MATANO
- Elad Venezian, IBM Research AI
  eladv@il.ibm.com
  https://researcher.watson.ibm.com/researcher/view.php?person=il-ELADV
- Noam Slonim, IBM Research AI
  noams@il.ibm.com
  https://researcher.watson.ibm.com/researcher/view.php?person=il-NOAMS

Roy Bar-Haim is a Research Staff Member at IBM Research AI. Since joining Project Debater in 2013, he has been leading a global research team working on stance classification, sentiment analysis, argument mining and argument summarization. He has published in leading NLP and AI conferences and journals, including ACL, EMNLP, AAAI, COLING, EACL, JAIR and JNLE. He regularly reviews for top NLP and AI conferences, and serves as a member of the TACL elite reviewer team. Roy received his Ph.D in Computer Science from Bar-Ilan University. Before joining IBM, he led NLP research teams in several startup companies. Roy delivered a one-hour talk about Project Debater at the NeurIPS 2018 Expo.

Liat Ein-Dor is a Research Staff Member at IBM Research AI. She received her Ph.D in theoretical physics from Bar-Ilan University in 2001 and has taught several courses there. In 2002 she was a postdoctoral fellow in Laboratoire de Physique Théorique de l'Ecole Normale Supérieure Paris, and from 2003 to 2006 she was a Postdoctoral Fellow and a Research Consultant at the Weizmann Institute of Science. Since 2006, Liat has been working as a research scientist in the hi-tech industry, and joined IBM's Haifa Research Lab in 2010. She has been leading research activities within Project Debater on tasks such as semantic similarity and argumentation mining. She has a diverse background in machine learning, having worked on a variety of domains including computational linguistics, computational biology, fraud detection and theoretical physics. She has publications in all these fields.

Matan Orbach is a Research Staff Member at IBM Research AI. Since joining IBM in 2014, he has worked on a diverse set of NLP tasks, recently focusing on multilingual stance detection and targeted sentiment analysis. Within Project Debater, Matan has led a team working on rebuttal generation through the use of principled arguments. Prior to joining IBM, he received his M.Sc. from the faculty of Electrical Engineering at the Technion, where his research focused on graph-based semi-supervised learning.

Elad Venezian is a Research Staff Member at IBM Research AI. He is currently the chief architect of Project Debater with a focus on making Project Debater technologies available to academia and business. Prior to this role, Elad served in different technical and leadership roles in the Project Debater grand challenge, among them, leading the speech generation team. Elad received his M.Sc. from the faculty of Electrical Engineering at the Tel Aviv University, where his research focused on non-linear systems.

Noam Slonim is a Distinguished Engineer at IBM Research AI. He received his doctorate from the Interdisciplinary Center for Neural Computation at the Hebrew University and held a post-doc position at the Genomics Institute at Princeton University. Noam proposed to develop Project Debater in 2011. He has been serving as the Principal Investigator of the project since then. Noam published around 60 peer reviewed articles, focusing on the last few years on advancing the emerging field of Computational Argumentation. Noam initiated and co-organized the ACL-2016 tutorial on NLP Approaches to Computational Argumentation and the 2015 Dagstuhl workshop on Debating Technologies. In EMNLP 2018 he co-chaired the Argument Mining workshop. Noam delivered a keynote speech on Project Debater at EMNLP 2019.

References

Roy Bar-Haim, Indrajit Bhattacharya, Francesco Dinuzzo, Amrita Saha, and Noam Slonim. 2017. Stance classification of context-dependent claims. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*, pages 251–261, Valencia, Spain. Association for Computational Linguistics.

Yonatan Bilu, Ariel Gera, Daniel Hershcovich, Benjamin Sznajder, Dan Lahav, Guy Moshkowich, Anael Malet, Assaf Gavron, and Noam Slonim. 2019. Argument invention from first principles. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 1013–1026, Florence, Italy. Association for Computational Linguistics.

Katarzyna Budzynska and Chris Reed. 2019. Advances in argument mining. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics: Tutorial Abstracts*, pages 39–42, Florence, Italy. Association for Computational Linguistics.

Liat Ein-Dor, Eyal Shnarch, Lena Dankin, Alon Halfon, Benjamin Sznajder, Ariel Gera, Carlos Alzate, Martin Gleize, Leshem Choshen, Yufang Hou, Yonatan Bilu, Ranit Aharonov, and Noam Slonim. 2020. Corpus wide argument mining - A working solution. In *Proceedings of the Thirty-Fourth AAAI Conference on Artificial Intelligence*, pages 7683–7691. AAAI Press.

Iryna Gurevych, Eduard H. Hovy, Noam Slonim, and Benno Stein. 2016. Debating Technologies (Dagstuhl Seminar 15512). *Dagstuhl Reports*, 5(12):18–46.

Ivan Habernal and Iryna Gurevych. 2016. Which argument is more convincing? analyzing and predicting convincingness of web arguments using bidirectional LSTM. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1589–1599, Berlin, Germany. Association for Computational Linguistics.

John Lawrence and Chris Reed. 2019. Argument mining: A survey. *Computational Linguistics*, 45(4):765–818.

Ran Levy, Yonatan Bilu, Daniel Hershcovich, Ehud Aharoni, and Noam Slonim. 2014. Context dependent claim detection. In *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, pages 1489–1500, Dublin, Ireland. Dublin City University and Association for Computational Linguistics.

Marco Lippi and Paolo Torroni. 2015. Context-independent claim detection for argument mining. In *Proceedings of the 24th International Conference on Artificial Intelligence*, IJCAI'15, pages 185–191. AAAI Press.

Shachar Mirkin, Guy Moshkowich, Matan Orbach, Lili Kotlerman, Yoav Kantor, Tamar Lavee, Michal Jacovi, Yonatan Bilu, Ranit Aharonov, and Noam Slonim. 2018. Listening comprehension over argumentative content. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 719–724, Brussels, Belgium. Association for Computational Linguistics.

Ruty Rinott, Lena Dankin, Carlos Alzate Perez, Mitesh M. Khapra, Ehud Aharoni, and Noam Slonim. 2015. Show me your evidence - an automatic method for context dependent evidence detection. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 440–450, Lisbon, Portugal. Association for Computational Linguistics.

Noam Slonim, Yonatan Bilu, Carlos Alzate, Roy Bar-Haim, Ben Bogin, Francesca Bonin, Leshem Choshen, Edo Cohen-Karlik, Lena Dankin, Lilach Edelstein, Liat Ein-Dor, Roni Friedman-Melamed, Assaf Gavron, Ariel Gera, Martin Gleize, Shai Gretz, Dan Gutfreund, Alon Halfon, Daniel Hershcovich, Ron Hoory, Yufang Hou, Shay Hummel, Michal Jacovi, Charles Jochim, Yoav Kantor, Yoav Katz, David Konopnicki, Zvi Kons, Lili Kotlerman, Dalia Krieger, Dan Lahav, Tamar Lavee, Ran Levy, Naftali Liberman, Yosi Mass, Amir Menczel, Shachar Mirkin, Guy Moshkowich, Shila Ofek-Koifman, Matan Orbach, Ella Rabinovich, Ruty Rinott, Slava Shechtman, Dafna Sheinwald, Eyal Shnarch, Ilya Shnayderman, Aya Soffer, Artem Spector, Benjamin Sznajder, Assaf Toledo, Orith Toledo-Ronen, Elad Venezian, and Ranit Aharonov. 2021. An autonomous debating system. *Nature*, 591(7850):379–384.

Noam Slonim, Iryna Gurevych, Chris Reed, and Benno Stein. 2016. NLP approaches to computational argumentation. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics: Tutorial Abstracts*, Berlin, Germany. Association for Computational Linguistics.

Christian Stab, Tristan Miller, Benjamin Schiller, Pranav Rai, and Iryna Gurevych. 2018. Cross-topic argument mining from heterogeneous sources. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 3664–3674, Brussels, Belgium. Association for Computational Linguistics.

Henning Wachsmuth, Nona Naderi, Yufang Hou, Yonatan Bilu, Vinodkumar Prabhakaran, Tim Alberdingk Thijm, Graeme Hirst, and Benno Stein. 2017. Computational argumentation quality assessment in natural language. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*, pages 176–187, Valencia, Spain. Association for Computational Linguistics.