



Argumentación Computacional

Herramientas y Aplicaciones de la IA

Curso 2023 - 2024

Stella Heras – stehebar@upv.es



Índice

1. Introducción
2. Definiciones
3. Tareas de la Argumentación
4. Aplicaciones
5. Marcos de Argumentación
6. Semánticas de Aceptabilidad
7. Prácticas
8. Trabajos

5. Marcos Abstractos de Argumentación [Dung, 1993]

- $AF = (Args, R)$, dónde:
 - **Args** es un conjunto de argumentos.
 - $R \subseteq Args \times Args$ es una relación de ataque entre argumentos.
 - Args y R son términos abstractos que no dependen de ningún formalismo de representación.

5. Marcos Abstractos de Argumentación

- Ejemplo: Lógica Proposicional [Amgoud, 2002]
- Δ es un conjunto de fórmulas en lógica proposicional.
- $\text{Args} = \{\Gamma, \alpha\}$ dónde:
 - Γ es el **conjunto soporte**: menor conjunto consistente de Δ que implica
 - la conclusión α

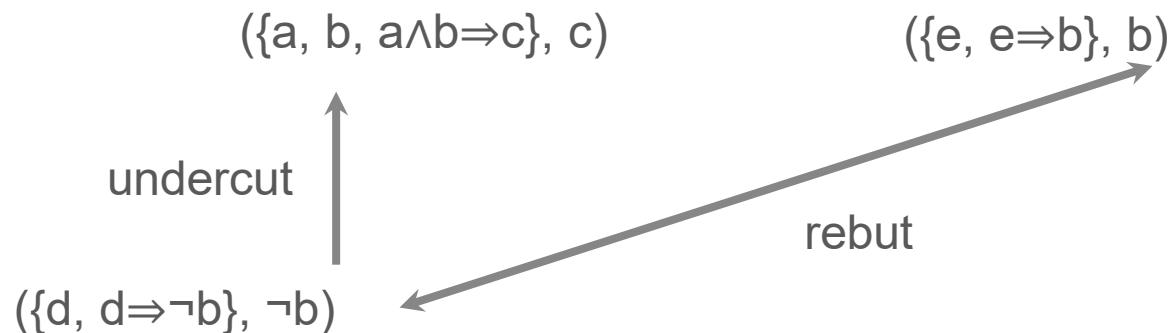
5. Marcos Abstractos de Argumentación

■ Tipos de ataques:

□ **Rebut:** $A_1 = \{\Gamma_1, \alpha_1\}$ rebuts $A_2 = \{\Gamma_2, \alpha_2\}$ si $\alpha_1 = \neg\alpha_2$

□ **Undercut:** $A_1 = \{\Gamma_1, \alpha_1\}$ undercuts $A_2 = \{\Gamma_2, \alpha_2\}$ si $\alpha_1 = \neg\alpha$ para algún $\alpha \in \Gamma_2$

$\Delta = \{a, b, a \wedge b \Rightarrow c, d, d \Rightarrow \neg b, e, e \Rightarrow b\}$



5. Marcos Abstractos de Argumentación

- Marcos Abstractos de Argumentación Basados en Valores [Bench-Capon, 2003]
- $AVAF = (Args, R, Valores, val, Audiencia)$
 - **val**: $Args \rightarrow Valores$
 - **Audiencia**: orden total sobre el conjunto de valores
 - $(A, B) \in R$ si A ataca B y $val(A) < val(B)$ no es cierto para la Audiencia considerada

6. Semánticas de Aceptabilidad

[Dung, 1993]

- Dado $AF = (Args, R)$ y $S \subseteq Args$:
 - S es **libre de conflicto** si $\forall A, B \in Args, (A, B) \notin R$ y $(B, A) \notin R$
 - $A \in Args$ es **aceptable con respecto a S** sii $\forall B \in Args / (B, A) \in R, \exists C \in S / (C, B) \in R$
 - Intuitivamente, un argumento A es aceptable con respecto a un set de argumentos S si, para todo argumento B que ataca a A existe un argumento C en el set S que ataca a B (C defiende a A).
- Se llama **extensión** a un conjunto de argumentos de $Args$, que pueden “convivir”

6. Semánticas de Aceptabilidad

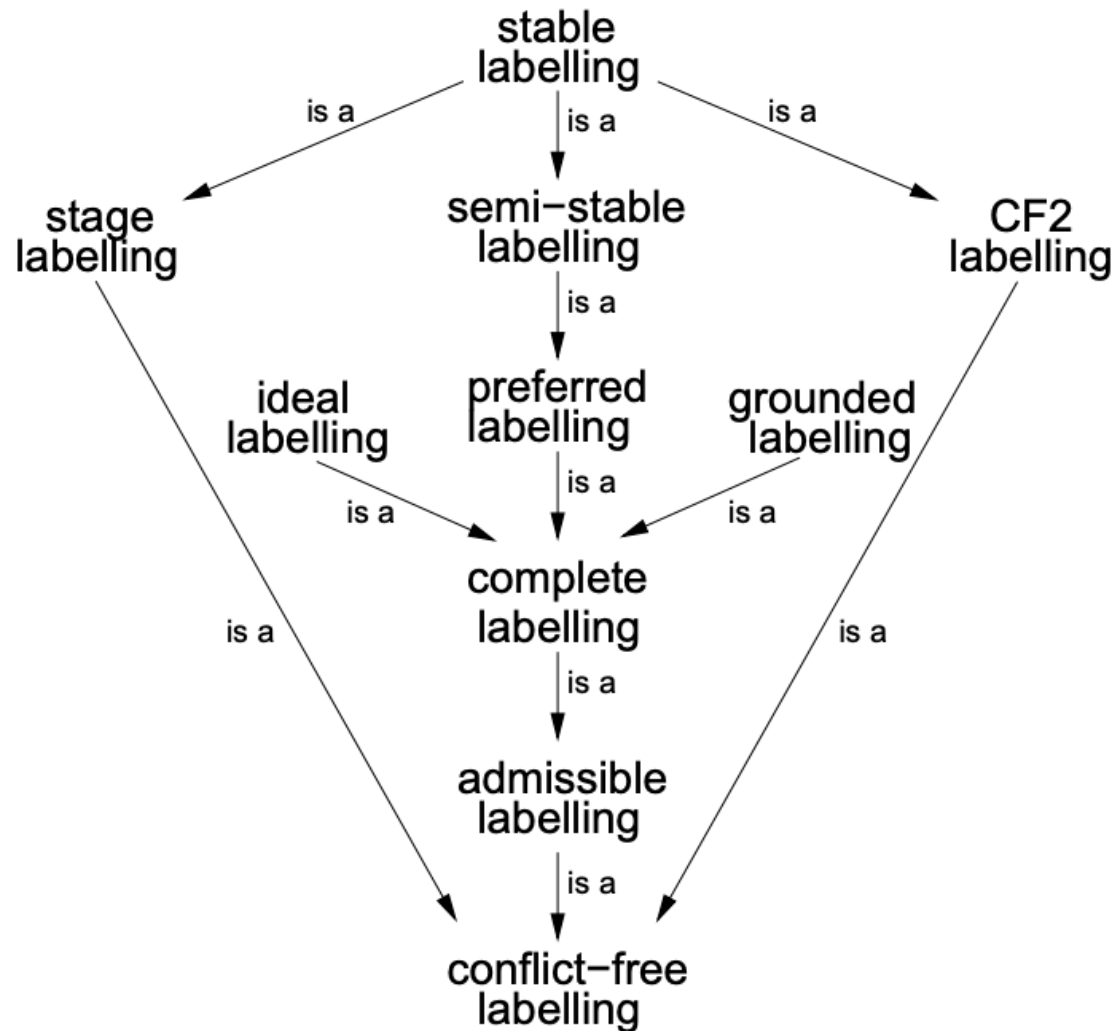
- Dado $AF = (Args, R)$ y dada la extensión libre de conflicto $S \subseteq Args$:
 - S es **admisibile** si todo argumento de S es aceptable con respecto a S
 - S es **completa** si todo argumento aceptable con respecto a S está incluido en S
 - S es **sólida (grounded)** si es la extensión completa más pequeña de $Args$
 - S es **preferida** si es una extensión completa más grande posible de $Args$
 - S es **estable** si todo argumento que no está en S es atacado por algún argumento de S

6. Semánticas de Aceptabilidad

- Dado $AF = (Args, R)$ y dada la extensión libre de conflicto $S \subseteq Args$:
 - S es **semiestable** si S es completa y tiene un rango mínimo (respecto a la inclusión de conjuntos) entre las extensiones completas (la extensión completa más pequeña)
 - S es **escalonada (stage)** si S está libre de conflicto y tiene un rango mínimo (respecto a la inclusión de conjuntos) entre las extensiones libres de conflictos;
 - S es **ideal** si S es la máxima extensión admisible (respecto a la inclusión de conjuntos) que es un subconjunto de cada extensión preferida.

6. Semánticas de Aceptabilidad

[Baroni, et. al.. 2004]



6. Semánticas de Aceptabilidad

Ejemplos

- A y B son argumentos que defienden la aplicación de los medicamentos A y B respectivamente
- **A** y **B** promueven el valor **salud** del paciente
- **C** es un argumento que ataca a A porque A es costoso y promueve el valor **ahorro**
- **D** es un argumento que ataca a B porque B tiene efectos secundarios adversos y promueve el valor **seguridad**
- Audiencia: **seguridad** > **salud** > **ahorro**



7. Prácticas

PRÁCTICA 1: Análisis Argumentativo

- Utiliza la herramienta OVA para analizar un artículo de opinión sobre el cambio de horario en España:
 - Texto del artículo en PoliformaT
 - Sigue las guías de anotación publicadas en PoliformaT
 - Comparte en PoliformaT el grafo argumentativo generado en JSON
 - Descarga los grafos de análisis argumentativo generados por cada equipo (espacio compartido en PoliformaT) y crea un Script en Python para calcular métricas que midan el acuerdo entre los diferentes anotadores. Por ejemplo, mide el porcentaje de acuerdo en cuanto a que partes del texto argumentativo tienen una relación de “ataque (CA)” o de “soporte (RA)”
 - Tutorial práctico Krippendorff Alpha:
<https://www.surgehq.ai/blog/inter-rater-reliability-metrics-an-introduction-to-krippendorffs-alpha>

7. Prácticas

PRÁCTICA 2: Argument Solvers

- Utiliza la herramienta TOAST ([The Online Argument Structure Tool](#)) para crear un marco de argumentación y evaluarlo para determinar la aceptabilidad de los argumentos
- Utiliza los argumentos que etiquetaste sobre el debate del cambio horario en España en la tarea anterior para definir las premisas y reglas necesarias para poder evaluar el marco de argumentación
- Se debe probar si se encuentran argumentos a favor de la *query* que se realice tanto para mantener el cambio horario como para eliminarlo.
- Se pueden probar diferentes semánticas y ver las diferencias en el grafo de argumentos que se genera.
- Subir el link que se genera al evaluar una query con una semántica concreta a la tarea de PoliformaT, y describir en un párrafo los resultados obtenidos en distintas queries y semánticas.

8. Trabajos

1. Participa en la tarea DialAM 2024 (<https://dialam.arg.tech/>): **Paper**
Deadline 20 de Mayo de 2024
2. Escribir un artículo de revisión sobre “Argumentación Computacional para luchar contra la Desinformación” o cualquier otro tema relacionado con esta parte de la asignatura (acordar con la profesora)

Recursos

- Modelos ML Open Source: <https://huggingface.co/>
- Bases de datos:
 - https://en.wikipedia.org/wiki/List_of_datasets_for_machine-learning_research
 - **Zenodo**: repositorio abierto multidisciplinar mantenido por el CERN.
 - A través del motor de búsqueda de Zenodo se pueden localizar conjuntos de datos, documentos y otros materiales de investigación:
<https://www.zenodo.org/>
 - AIFDb Corpora: <http://corpora.aifdb.org/>
 - TU Datalib: <https://tudatalib.ulb.tu-darmstadt.de/handle/tudatalib/1359?locale-attribute=en>
 - Argumentative Microtexts: <https://github.com/peldszus/arg-microtexts>
 - CDCP (Cornell eRulemaking Corpus):
<https://paperswithcode.com/dataset/cdcp>
 - Debatebase: <https://idebate.net/resources/debatabase>
 - DebateSum: <https://github.com/Hellisotherpeople/DebateSum>
 - ReCAP Corpus: <https://basilika.uni-trier.de/nextcloud/s/JePuLMGdZNBjMUK>

Extensión de trabajos para TFM

- **Título:** Optimización de la contratación pública mediante el uso de técnicas de IA en las fases de diseño y construcción

Directoras: Stella Heras (stehebar@upv.es) y Laura Montalbán (laumondo@upv.es)

Descripción: Las agencias públicas encargadas del desarrollo de infraestructuras están demandando el desarrollo de modelos digitales de gestión que optimicen la contratación en las fases de diseño y construcción. El objetivo de dichos modelos es reducir los sobrecostes y retrasos en el diseño y construcción de dichas infraestructuras. Estos modelos requieren minimizar la subjetividad en la identificación y evaluación de los riesgos, determinar la relación de interdependencia de los riesgos en función de la complejidad y la incertidumbre asociada con las actuaciones a llevar a cabo, y calibrar los factores utilizados para la contratación en las fases de diseño y construcción. Para hacer frente a estos retos, el proyecto pretende hacer uso de diferentes técnicas de IA (e.g. LLMs, NLP, argumentación computacional) para: 1. Desarrollar un algoritmo que extraiga la información de la licitación y adjudicación de contratos de obras 2. Desarrollar un modelo de análisis de la relación calidad-precio de las ofertas presentadas en función de las características de la contratación. 3. Desarrollar un modelo de predicción de riesgos de sobrecostes o retrasos temporales. Se dispone para ello de bases de datos con información sobre expedientes de licitación y ejecución de los contratos de la administración pública.

Temática: NLP, ML.

Extensión de trabajos para TFM

- **Título:** Técnicas de argumentación computacional para la detección y análisis de la desinformación

Directores: Stella Heras (stehebar@upv.es) y Vicente Botti (vbotti@dsic.upv.es)

Descripción: Tal y como se define en la Comunicación de la Comisión Europea (CE) sobre el Plan de Acción Europeo para la Democracia (EDAP) "la desinformación es un contenido falso o engañoso que se difunde con la intención de engañar o de obtener un beneficio económico o político y que puede causar un perjuicio público". El panorama actual de la investigación y la tecnología dedicadas a la lucha contra la desinformación revela una situación sin precedentes en la que la cantidad y calidad de las bases de datos validadas por verificadores de hechos, que contienen indicadores de contenidos etiquetados como posible desinformación en múltiples formatos, está en auge y las API que permiten la inspección de estos datos en tiempo real o casi real están evolucionando hacia la estandarización, de modo que la información pueda obtenerse de forma fiable y escalable. En este contexto, es crucial investigar cómo pueden aplicarse las tecnologías de AI (en concreto, la argumentación computacional) para capacitar a los ciudadanos para cuestionar, investigar y comprender si un hecho o una noticia es cierta. Además, también es necesario ayudar a los profesionales de los medios de comunicación y a los responsables políticos a construir elementos para crear informes de datos personalizables y fiables sobre la desinformación. El objetivo de este trabajo es investigar el papel de la argumentación computacional y la minería de argumentos para detectar los intentos de desinformación, por ejemplo la demagogia, las falacias y las amenazas persuasivas, que tratan de inducir a error a la opinión popular y/o personal hacia la aceptación de noticias o hechos alterados.

Temática: minería de argumentos, lucha contra la desinformación.

Extensión de trabajos para TFM

- **Título:** Técnicas de argumentación computacional para la gestión de sesgos de género en modelos de IA generativos

Directores: Stella Heras (stehebar@upv.es)

Descripción: La argumentación computacional ofrece una vía para examinar críticamente tanto los datos de entrenamiento como las salidas de los modelos generativos de IA, ayudando a identificar, explicar y corregir sesgos de género, lo cual es crucial para el desarrollo de tecnologías de IA justas y responsables. Algunas posibles líneas de actuación son: - Análisis Crítico de Datos de Entrenamiento; - Evaluación de Resultados del Modelo; - Argumentación Basada en Reglas y Principios Éticos; - Interpretación y Explicación de Decisiones del Modelo; - Entrenamiento y Retroalimentación Iterativa; - Desarrollo de Contraargumentos. El presente trabajo pretende explorar una o varias de estas líneas, en un área de investigación todavía incipiente y con gran atractivo tanto para la academia como para la industria dada la reciente proliferación de este tipo de modelos.

Temática: minería de argumentos, NLP.

Referencias

- Congresos y Workshops: COMMA, CMNA, ...
- Revista: Argument and Computation (Taylor and Francis)
- Centros de Investigación:
 - ARG-Tech Centre for Argument Technology: <https://arg-tech.org>
 - LUH-AI: <https://www.ai.uni-hannover.de/en/institute/research-groups/nlp>
 - AIG-Hagen: <https://www.fernuni-hagen.de/aig/en/>
- Libros:
 - Argumentation in AI. Iyad Rahwan y Guillermo Simari. Springer 2009.
 - Argumentation Schemes. D. Walton et al. Cambridge University Press 2008
 - Fundamentals of Critical Argumentation. D. Walton. Cambridge University Press 2005
 - Fundamentals of Argumentation Theory. F. van Eemeren et al. Lawrence Elbaum Associates 1996
 - Navigating Fake News, Alternative Facts, and Misinformation in a Post-Truth World. Kimiz Dalkir, Rebecca Katz. IGI Global 2020

Referencias

■ Artículos (Computational Argumentation):

- **Prakken, H.** (2005). Coherence and flexibility in dialogue games for argumentation. *Journal of logic and computation*, 15(6), 1009-1040.
- **Dung, P. M.** (1993). On the Acceptability of Arguments and its Fundamental Role in Nonmonotonic Reasoning and Logic Programming. In *IJCAI* (Vol. 93, pp. 852-857).
- **Amgoud, L., & Cayrol, C.** (2002). A reasoning model based on the production of acceptable arguments. *Annals of Mathematics and Artificial Intelligence*, 34(1-3), 197-215.
- **Baroni, P., Caminada, M., & Giacomin, M.** (2011). An introduction to argumentation semantics. *The knowledge engineering review*, 26(4), 365-410.
- **Bench-Capon, T. J.** (2003). Persuasion in practical argument using value-based argumentation frameworks. *Journal of Logic and Computation*, 13(3), 429-448.
- **Walton, D., & Krabbe, E. C.** (1995). *Commitment in dialogue: Basic concepts of interpersonal reasoning*. SUNY press.
- **McBurney, P., & Parsons, S.** (2001). Dialogue games in multi-agent systems. *Informal Logic*, 22(3).

Referencias

■ Artículos (Computational Argumentation):

- **Mirko, L. E. N. Z., Sahitaj, P., Kallenberg, S., Coors, C., Dumani, L., Schenkel, R., & Bergmann, R. (2020).** Towards an argument mining pipeline transforming texts to argument graphs. In *Computational Models of Argument: Proceedings of COMMA* (Vol. 326, p. 263)
- **Schneider, J., et al. (2013).** A review of argumentation for the social semantic web. *Semantic Web*, 4(2), 159-218.
- **Reed, C. (2014).** OVA+: An argument analysis interface. In *Computational Models of Argument: Proceedings of COMMA* (Vol. 266, No. 2014, p. 463).
- **Ruiz-Dolz, R., Nofre, M., Taulé, M., Heras, S., & García-Fornes, A. (2021).** Vivesdebate: A new annotated multilingual corpus of argumentation in a debate tournament. *Applied Sciences*, 11(15), 7160.
- **Ruiz-Dolz, R., & Iranzo-Sánchez, J. (2023).** VivesDebate-Speech: A Corpus of Spoken Argumentation to Leverage Audio Features for Argument Mining. *arXiv preprint arXiv:2302.12584*.
- **Ruiz-Dolz, R., Taverner, J., Heras, S., Garcia-Fornes, A., & Botti, V. (2022, July).** A Qualitative Analysis of the Persuasive Properties of Argumentation Schemes. In *Proceedings of the 30th ACM Conference on User Modeling, Adaptation and Personalization* (pp. 1-11).
- **Ruiz-Dolz, R., Alemany, J., Barberá, S. M. H., & García-Fornes, A. (2021).** Transformer-based models for automatic identification of argument relations: a cross-domain evaluation. *IEEE Intelligent Systems*, 36(6), 62-70.

Referencias

■ Artículos (NLP):

- **Yamen Ajjour, Wei-Fan Chen, Johannes Kiesel, Henning Wachsmuth, and Benno Stein (2017).** Unit Segmentation of Argumentative Texts. In Proceedings of the Fourth Workshop on Argument Mining, pages 118–128.
- **Khalid Al-Khatib, Henning Wachsmuth, Matthias Hagen, and Benno Stein (2017).** Patterns of Argumentation Strategies across Topics. In Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, pages 1362–1368.
- **Milad Alshomary, Nick Düsterhus, and Henning Wachsmuth (2020).** Extractive Snippet Generation for Arguments. In Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval, pages 1969–1972.
- **Milad Alshomary, Shahbaz Syed, Arkajit Dhar, Martin Potthast, and Henning Wachsmuth (2021).** Counter-Argument Generation by Attacking Weak Premises. In Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021, pages 1816–1827.
- **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.
- **Yonatan Bilu and Noam Slonim. Claim Synthesis via Predicate Recycling (2016).** In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, pages 525–530.
- **Roxanne El Baff, Henning Wachsmuth, Khalid Al Khatib, Manfred Stede, and Benno Stein (2019).** Computational Argumentation Synthesis as a Language Modeling Task. In Proceedings of the 12th International Conference on Natural Language Generation, pages 54–64.
- **Vanessa Wei Feng and Graeme Hirst (2011).** Classifying Arguments by Scheme. In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics, pages 987–996.

Referencias

■ Artículos (NLP):

- **Timon Gurcke, Milad Alshomary, and Henning Wachsmuth(2021).** Assessing the Sufficiency of Arguments through Conclusion Generation. In Proceedings of the 8th Workshop on Argument Mining, pages 67–77.
- **Kazi Saidul Hasan and Vincent Ng (2013).** Stance Classification of Ideological Debates: Data, Models, Features, and Constraints. In Proceedings of the Sixth International Joint Conference on Natural Language Processing, pages 1348--1356.
- **Xinyu Hua, Zhe Hu, and Lu Wang (2019).** Argument Generation with Retrieval, Planning, and Realization. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, pages 2661–2672.
- **Christian Stab (2017).** Argumentative Writing Support by means of Natural Language Processing, Chapter 5. PhD thesis, TU Darmstadt.
- **Lu Wang and Wang Ling (2016).** Neural Network-Based Abstract Generation for Opinions and Arguments. In: Proceedings of the 15th Conference of the North American Chapter of the Association for Computational Linguistics, pages 47–57.

Referencias

- Artículos (LLM + Argumentation):
 - **Ruíz-Doz, R; Lawrence, J. (2024).** Detecting Argumentative Fallacies in the Wild: Problems and Limitations of Large Language Models.
 - **Ruíz-Dolz, R., Taverner, J., Lawrence, J., & Reed, C. (2024).** NLAS-multi: A Multilingual Corpus of Automatically Generated Natural Language Argumentation Schemes. arXiv preprint arXiv:2402.14458.
 - **de Sousa, L. H. H., Trajano, G., Morales, A. S., Sarkadi, S., & Panisson, A. R. (2024).** Using Chatbot Technologies to Support Argumentation. In *16th International Conference on Agents and Artificial Intelligence (ICAART 2024)*. SciTePress.
 - **Gorur, D., Rago, A., & Toni, F. (2024).** Can Large Language Models perform Relation-based Argument Mining?. *arXiv preprint arXiv:2402.11243*.