Ontology-based XAI - Annotated Bibliography
Alberto D. López

References

[1] Erik Brynjolfsson and Tom Mitchell. What can machine learning do? Workforce implications: Profound change is coming, but roles for humans remain. *Science*, 358(6370):1530–1534, 2017.

General XAI - Check

[2] Lorenz Bühmann, Jens Lehmann, and Patrick Westphal. DL-LearnerA framework for inductive learning on the Semantic Web. *Journal of Web Semantics*, 39:15–24, 2016.

[3] Edward Choi, Mohammad Taha Bahadori, Le Song, Walter F. Stewart, and Jimeng Sun. GRAM: Graph-based attention model for healthcare representation learning. In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, volume Part F1296, pages 787–795, 2017.

[4] Roberto Confalonieri, Tillman Weyde, Tarek R. Besold, and Fermín Moscoso del Prado Martín. Trepan Reloaded: A Knowledge-driven Approach to Explaining Artificial Neural Networks. *arXiv preprint arXiv:1906.08362*, 2019.

[5] Arun Das and Paul Rad. Opportunities and Challenges in Explainable Artificial Intelligence (XAI): A Survey. *arXiv*, 2020.

General XAI - Check

[6] Derek Doran, Sarah Schulz, and Tarek R. Besold. What does explainable ai really mean? A new conceptualization of perspectives. *arXiv*, 2017.

General XAI - Check

[7] a Elixhauser, C Steiner, and L Palmer. Clinical Classifications Software (CCS), 2014. *U.S. Agency for Healthcare Research and Quality*, 27(November 2013):1–54, 2014.

[8] Giuseppe Futia and Antonio Vetrò. On the integration of knowledge graphs into deep learning models for a more comprehensible AI-Three challenges for future research. *Information (Switzerland)*, 11(2):122, 2020.

[9] Valentina Ghidini, Alan Perotti, and Rossano Schifanella. Quantitative and Ontology-Based Comparison of Explanations for Image Classification. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 11943 LNCS, pages 58–70. Springer, 2019.

[10] Mohammad Al Hasan, Vineet Chaoji, Saeed Salem, Mohammed Zaki, and New York. Link Prediction using Supervised Learning. In *SDM06: workshop on link analysis, counter-terrorism and security*, volume 30, pages 798–805, 2006.

[11] Hyeon Woo Kang and Hang Bong Kang. Prediction of crime occurrence from multimodal data using deep learning. *PLoS ONE*, 12(4):e0176244, 2017.

[12] Thi Kim Phung Lai, Nhat Hai Phan, Han Hu, Anuja Badeti, David Newman, and Dejing Dou. Ontology-based interpretable machine learning for textual data. *arXiv*, 2020.

[13] Hyunkwang Lee, Sehyo Yune, Mohammad Mansouri, Myeongchan Kim, Shahein H. Tajmir, Claude E. Guerrier, Sarah A. Ebert, Stuart R. Pomerantz, Javier M. Romero, Shahmir Kamalian, Ramon G. Gonzalez, Michael H. Lev, and Synho Do. An explainable deep-learning algorithm for the detection of acute intracranial haemorrhage from small datasets. *Nature Biomedical Engineering*, 3(3):173–182, 2019.

    Clinical XAI - Check

[14] Zachary C. Lipton. The mythos of model interpretability. *Communications of the ACM*, 61(10):35–43, 2018.

[15] Gianclaudio Malgieri and Giovanni Comandé. Why a Right to Legibility of Automated Decision-Making Exists in the General Data Protection Regulation. *International Data Privacy Law*, 7(4):243–265, nov 2017.

    General Data Protection Regulation justification.

[16] George A. Miller. WordNet: A Lexical Database for English. *Communications of the ACM*, 38(11):39–41, 1995.

Wordnet ontology paper

[17] Cecilia Panigutti, Alan Perotti, and Dino Pedreschi. Doctor XAI An ontology-based approach to black-box sequential data classification explanations. In *FAT* 2020 - Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, pages 629–639, 2020.

[18] Md Kamruzzaman Sarker, Ning Xie, Derek Doran, Michael Raymer, and Pascal Hitzler. Explaining trained neural networks with semantic web technologies: First steps. *arXiv*, 2017.

[19] Arne Seeliger, Matthias Pfaff, and Helmut Krcmar. Semantic web technologies for explainable machine learning models: A literature review. In *CEUR Workshop Proceedings*, volume 2465, pages 30–45, 2019.

Survey on KAs and Ontologies applications to XAI.

[20] Amitojdeep Singh, Sourya Sengupta, and Vasudevan Lakshminarayanan. Explainable deep learning models in medical image analysis. *arXiv*, 2020.

[21] Kacper Sokol and Peter Flach. Explainability fact sheets: a framework for systematic assessment of explainable approaches. In *arXiv*, pages 56–67, 2019.

Pending. Explainability survey, the authors also propose a taxonomy aimed at describe methods in five dimensions: functional, operational, usability, safety and validation.

[22] Martin Stano, Wanda Benesova, and Lukas S. Martak. Explainable 3D convolutional neural network using GMM encoding. In *Twelfth International Conference on Machine Vision (ICMV 2019)*, volume 11433, page 43. International Society for Optics and Photonics, 2020.

[23] M. Q. Stearns, C. Price, K. A. Spackman, and A. Y. Wang. SNOMED clinical terms: overview of the development process and project status. In *Proceedings / AMIA ... Annual Symposium. AMIA Symposium*, pages 662–666. American Medical Informatics Association, 2001.