Ontology-based XAI - Annotated Bibliography

Alberto D. López

March, 2021

References

[1] Mohammad Al Hasan, Vineet Chaoji, Saeed Salem, and Mohammed Zaki. Link prediction using supervised learning. In *SDM06: workshop on link analysis, counter-terrorism and security*, volume 30, pages 798–805, 2006.

[2] Erik Brynjolfsson and Tom Mitchell. What can machine learning do? workforce implications. *Science*, 358(6370):1530–1534, 2017.

General XAI - Check

[3] 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.

[4] 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 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 787–795, 2017.

[5] 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.

[6] Healthcare Cost, Utilization Project, et al. Clinical classifications software (ccs) for icd-9-cm. *Available at: www. hcup-us. ahrq. gov/toolssoftware/ccs/ccs. jsp. Accessed February*, 27, 2018.

[7] Arun Das and Paul Rad. Opportunities and challenges in explainable artificial intelligence (xai): A survey. *arXiv preprint arXiv:2006.11371*, 2020.

General XAI - Check

[8] Derek Doran, Sarah Schulz, and Tarek R Besold. What does explainable ai really mean? a new conceptualization of perspectives. *arXiv preprint arXiv:1710.00794*, 2017.

General XAI - Check

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

[10] Valentina Ghidini, Alan Perotti, and Rossano Schifanella. Quantitative and ontology-based comparison of explanations for image classification. In *International Conference on Machine Learning, Optimization, and Data Science*, pages 58–70. Springer, 2019.

[11] Hyeon-Woo Kang and Hang-Bong Kang. Prediction of crime occurrence from multi-modal data using deep learning. *PloS one*, 12(4):e0176244, 2017.

[12] Phung Lai, NhatHai Phan, Han Hu, Anuja Badeti, David Newman, and Dejing Dou. Ontology-based interpretable machine learning for textual data. *arXiv preprint arXiv:2004.00204*, 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, et al. An explainable deep-learning algorithm for the detection of acute intracranial haemorrhage from small datasets. *Nature Biomedical Engineering*, 3(3):173, 2019.

Clinical XAI - Check

[14] Zachary C Lipton. The mythos of model interpretability. *Queue*, 16(3):31–57, 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*, 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.

[17] Cecilia Panigutti, Alan Perotti, and Dino Pedreschi. Doctor xai: an ontology-based approach to black-box sequential data classification explanations. In *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 preprint arXiv:1710.04324*, 2017.

[19] Arne Seeliger, Matthias Pfaff, and Helmut Krcmar. Semantic web technologies for explainable machine learning models: A literature review. In *PROFILES/SEMEX@ ISWC*, 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 preprint arXiv:2005.13799*, 2020.

[21] Kacper Sokol and Peter Flach. Explainability fact sheets: a framework for systematic assessment of explainable approaches. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, pages 56–67, 2020.

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 114331U. International Society for Optics and Photonics, 2020.

[23] Michael Q Stearns, Colin Price, Kent A Spackman, and Amy Y Wang. Snomed clinical terms: overview of the development process and project status. In *Proceedings of the AMIA Symposium*, page 662. American Medical Informatics Association, 2001.