Commented Bibliography

Étienne Houzé

1 Read

- (Marcus 2018) is a discussion on the article presenting AlphaZero as a *tabula rasa* method (Silver et al. 2017). It objects that since it has been engineered by experts in the go game, its very architecture already carries some specificities aimed at solvig this particular problem.
- (Russell & Norvig 2016) is a good handbook presenting the main techniques in AI. In particular, it describes knowledge-based systems and logic representation of the environment (Part III). Other parts are interesting too, but not as much linked to this work.
- (Dessalles 2015) presents an approach to define and name concepts in a geometric space (Contrast). It comes from the works in (Gärdenfors 2004) which introduces the existence of dimensions in the conceptualisation of the environment.
- (Dessalles 2008) defines a conflict as a situation when an event happens but is not desired, or when a desired event is not observed. It also presents a conflict resolution based on CAN. This approach can be used to generate dialogue between the user and the system in case of a conflict.
- (Došilović et al. 2018) is a quick global survey of the state of explanation in AI today. It offers a succint view but states well the current struggles of technologies used in most state-of-the-art solutions.
- (Zimmermann et al. 2017) is a short surver on the desires of users in Germany about the smart home. It concludes that, with Germany being the main market in Europe ($1.3B), it still has room for increase (for instance, in the US, smart homes business is worth $13B). It also shows on a very small sample that it seems users tend to put energy savings as the first incentive to use a smart home, with a predominance of independence for the older users (above 50).
- (Kaptein et al. 2017) presents the role of emotions in explanations. They can influence in three different ways :
  - Emotions can be used to understand which kind of explanation the user wants at time t.

  - Emotions can be used by the system to better describe the phenomena occuring and explain them to the user.
  - It can also be interseting for the system to look for the causes of emotions and then explain them.

- (Mohamed et al. 2017) exposes the issues of conflicts arising when several agents have different behaviours. Its work is not that much related to ours since it focuses on multi-resident activities, which is not the key part of our project.

- (Amarasinghe et al. 2018) examines a framework to explain decision made by DNNs in case an anomaly happens. This work being centered around deep learning technologies, it is not extremely relevant to ours. It would however be nice to see if a connection is possible between this conflict detection with neural nets and an explainable approach.

- (Blier & Ollivier 2017) is a master's thesis work recommended by David. It is about using complexity theory and applying it to deep learning models to sutdy whether the DL approach produces simple models in this redards. (Blier & Ollivier n.d.) is a short version of this, in an article style.

- (Olson et al. 2018) studies the possibility of combining AI to explain ML in medical applications. The method can be interesting for us, especially for their knowledge storage in a database structure.

- (Lalanda et al. 2014) describes iCasa and therefore should be cited in our technical presentation of the simulation.

- (Ehrlinger & Wöß 2016) proposes to precisely define knowledge graphs and present possible applications. The definition of a KG is then the following : *A knowledge graph acquires and integrates information into an ontology and applies a reasoner to derive new knowledge.*

2 To read

- Read or ask about Pierre Alexandre work on knowledge transfer and analogies, since our method may rely on similar techniques to create rules and explain things

- See if complexity theory can be a hint to evaluate the relevance of an explanation following Occam's razor. In fact, the most simple cause should explain and most certainly be the most relevant to the human user.

References

Amarasinghe, K., Kenney, K. & Manic, M. (2018), Toward explainable deep neural network based anomaly detection, *in* '2018 11th International Conference on Human System Interaction (HSI)', IEEE, pp. 311–317.

Blier, L. & Ollivier, Y. (2017), 'Universal compression bounds and deep learning'.

Blier, L. & Ollivier, Y. (n.d.), 'The description length of deep learning models'.

Dessalles, J.-L. (2008), 'A computational model of argumentation in everyday conversation: A problem-centred approach', *FRONTIERS IN ARTIFICIAL INTELLIGENCE AND APPLICATIONS* **172**, 128.

Dessalles, J.-L. (2015), From conceptual spaces to predicates, *in* 'Applications of Conceptual Spaces', Springer, pp. 17–31.

Došilović, F. K., Brčić, M. & Hlupić, N. (2018), Explainable artificial intelligence: A survey, *in* '2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)', IEEE, pp. 0210–0215.

Ehrlinger, L. & Wöß, W. (2016), Towards a definition of knowledge graphs., *in* 'SEMANTiCS (Posters, Demos, SuCCESS)'.

Gärdenfors, P. (2004), *Conceptual spaces: The geometry of thought*, MIT press.

Kaptein, F., Broekens, J., Hindriks, K. & Neerincx, M. (2017), The role of emotion in self-explanations by cognitive agents, *in* '2017 Seventh International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW)', pp. 88–93.

Lalanda, P., Hamon, C., Escoffier, C. & Leveque, T. (2014), icasa, a development and simulation environment for pervasive home applications, *in* '2014 IEEE 11th Consumer Communications and Networking Conference (CCNC)', pp. 1142–1143.

Marcus, G. (2018), 'Innateness, alphazero, and artificial intelligence', *arXiv preprint arXiv:1801.05667* .

Mohamed, R., Perumal, T., Sulaiman, M. N., Mustapha, N. & Razali, M. N. (2017), Conflict resolution using enhanced label combination method for complex activity recognition in smart home environment, *in* '2017 IEEE 6th Global Conference on Consumer Electronics (GCCE)', pp. 1–3.

Olson, R. S., Sipper, M., La Cava, W., Tartarone, S., Vitale, S., Fu, W., Orzechowski, P., Urbanowicz, R. J., Holmes, J. H. & Moore, J. H. (2018), A system for accessible artificial intelligence, *in* 'Genetic Programming Theory and Practice XV', Springer, pp. 121–134.

Russell, S. J. & Norvig, P. (2016), *Artificial intelligence: a modern approach*, Malaysia; Pearson Education Limited,.

Silver, D., Schrittwieser, J., Simonyan, K., Antonoglou, I., Huang, A., Guez, A., Hubert, T., Baker, L., Lai, M., Bolton, A. et al. (2017), 'Mastering the game of go without human knowledge', *Nature* **550**(7676), 354.

Zimmermann, G., Ableitner, T. & Strobbe, C. (2017), User needs and wishes in smart homes: What can artificial intelligence contribute?, *in* 'Pervasive Systems, Algorithms and Networks & 2017 11th International Conference on Frontier of Computer Science and Technology & 2017 Third International Symposium of Creative Computing (ISPAN-FCST-ISCC), 2017 14th International Symposium on', IEEE, pp. 449–453.