Causal Reasoning in Machine Learning

Pier Paolo Ippolito
Student ID: 28415981
Email: ppi1u16@ecs.soton.ac.uk

***Abstract*—Current supervised Machine Learning techniques are designed to exploit possible relationships/correlations between features and labels in order to produce reliable estimates. Use of this kind of technologies in ambit such as Medicine, Finance and Law are now raising increasing concerns due to the lack of ability in such systems to correctly identify causal relationships and provide explanations about their decisions.**

I. Introduction

Thanks to recent advancements in Artificial Intelligence (AI), we are now able to leverage Machine Learning and Deep Learning technologies in both academic and commercial applications. Although, relying just on correlations between the different features, can possibly lead to wrong conclusions since correlation does not necessarily imply causation.

Two of the main limitations of nowadays Machine Learning and Deep Learning models are:

- Robustness = trained models might not be able to generalise to new data and therefore would not be able to provide robust and reliable performances in the real world.
- Explainability = complex Deep Learning models can be difficult to analyse in order to clearly demonstrate their decision making process.

Developing models able to identify cause-effect relationships between different variables, might ultimately offer a solution to solve both of these problems. This idea, has also been supported by researchers such as Judea Pearl, which advocated how having models able to reason in uncertainties could not be enough to enable researchers to create machines able to truly express intelligent behaviour [1].

II. Causality

A. Concepts of Causality

Nowadays Machine Learning models, are able to learn from data by identifying patterns in large datasets. Although, humans might be able to perform a same task after just examining a few examples. This is possible thanks to the inherit humans ability to understand causal relationships and use inductive inference [2] in order to assimilate new information about the world [3]. Creating models able to demonstrate causal reasoning, would therefore open to us a whole new world of opportunities in AI research.

Causality arises naturally in our daily life every-time we ask ourselves any type of interventional or retrospective question (eg. What if I take this action? What if I would have acted differently?).

As shown in Figure 1, Causal Reasoning can be divided in three different hierarchical levels (Association, Intervention, Counterfactuals). At each level, different types of questions can be answered and in order to answer questions at the top levels (eg. Counterfactuals) are necessary as basis knowledge from the lower levels [4]. In fact, in order to be able to able to answer retrospective questions, we would expect to first be able to respond to intervention and association type of questions.

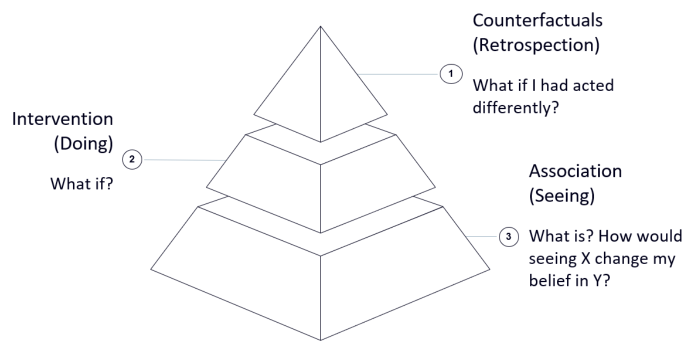

Fig. 1: Causality Hierarchy

Currently, Machine Learning models are only able to answer the probabilistic type of questions related to the Association level.

Thanks to the rising interest in this topic, a mathematical framework able to represent causal relationships has been constructed (Structural Causal Models (SCM) [4]). Using this type of framework, causal expressions can then be formulated and used in conjunction with data in order to make predictions.

B. Linear and Non-Linear Causality

Causality, can be divided into two main types: linear and non-linear (Figure 2) [5]:

- In linear causality, connections between the variables can be in a single direction and every effect can be originated by a limited number of causes. Causes always linearly precedes effects (time precedence).
- In non linear causality, connections between variables can be bi-directional and effects can possibly be originated by an unlimited number of causes.

Linear causation systems are characterised by proportional relationships between cause and effects variables (eg. Deterministic Systems). Instead, in non-linear causation systems disproportionate effects can take place (eg. Non-deterministic Systems). For example, small changes in input conditions would then result in different consequences (eg. "Butterfly Effect").

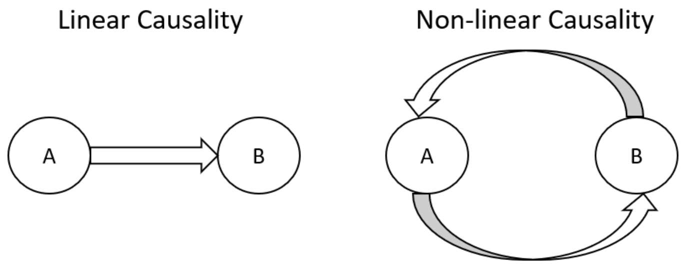

Fig. 2: Linear vs Non-Linear Causality

C. Case Study: Recommendation Systems

One of the main weakness of most Machine Learning models is the assumption that the data fed in is independent and identically distributed (IID). When this assumption holds, convergence to the lowest possible loss is achievable but when this constrain is violated the model might perform poorly even when attempting simple tasks (eg. poisoning attacks) [6].

As an example, let us consider an e-commerce recommendation system. Nowadays systems, are able to offer recommendations mainly based on products correlated to the ones we are planning to buy, although this cannot always lead to accurate estimates. For instance, we might have recently bought a new phone and we are now looking for a phone case. While browsing for phone cases, although our recommendation system might try to suggest us other items such as phones (just because they are correlated) instead of more cause-effect related items like screen protectors.

III. Techniques

One of the main technique used in order to try to discover causal relationships are Graphical Methods such as Knowledge Graphs and Bayesian Belief Networks. These two methods form in fact the basis of the Association level in the Causality Hierarchy, enabling us to answer questions such as: What different properties compose an entity and how are the different components related each other?

A. Knowledge Graphs

Knowledge Graphs, are a type of Graphical Technique commonly used in order to concisely store and retrieve related information from a large amount of data. Knowledge Graphs, are currently widely used in applications such as querying information from search engines, e-commerce websites and social networks. Following on our Recommendation System case study outlined before, Knowledge Graphs have been recently applied in causality (Yikun Xian et. al. [7]) in order to generate causal inference based recommendations.

As a simple example, let us consider what happens if we use a search engine in order to find out who Leonard Nimoy is (the actor who played the part of Spock in Star Trek). Once entered our query, the search engine will automatically build a Knowledge Graph similar to the one shown in Figure 3 taking as starting point our search query and then expanding from it to fetch any related information.

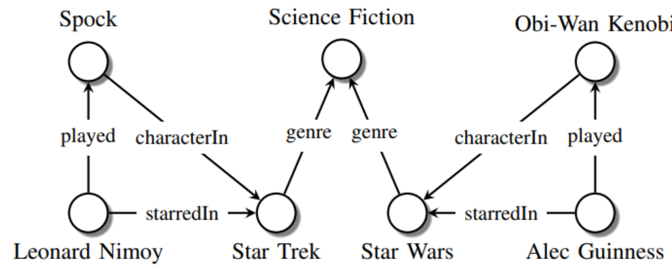

Fig. 3: Simple Knowledge Graph. Image reproduced from [8].

One of the most promising applications of Knowledge Graphs, is to create Machine Learning models able to learn from causality. Knowledge Graph Convolutional Networks (KGCN), represent one of the first successful applications in this ambit [9]. Graph Convolutional Networks, are in fact designed to create a vector (embedded) representation of a Knowledge Graph which can then be fed into a Machine Learning model to generate inference paths and provide evidences for the model predictions [10]. KGCN can be potentially used for either supervised or unsupervised tasks (eg. multi-class classification and clustering).

B. Bayesian Belief Networks

Bayesian Belief Networks, are a type of probabilistic model which makes use of simplifying assumptions so that to reliably define connections between different elements and calculate their probabilities relationships efficiently. By analysing interactions between the different elements, we can finally make use of these type of models in order to discover causal relationships. In a Bayesian Network, nodes represent variables while edges report the probabilistic connections between the different elements. A simple example of a three variables Bayesian Belief Network, is available in Figure 4.

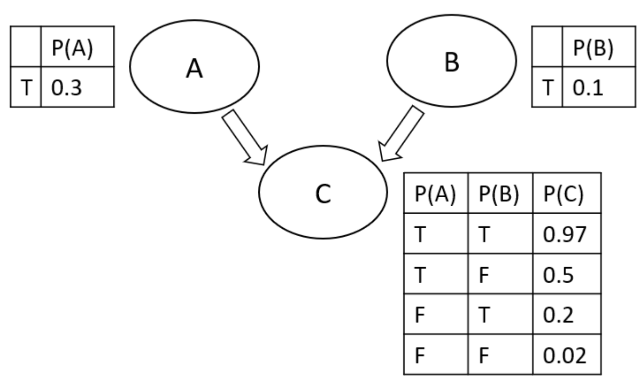

Fig. 4: Bayesian Belief Network

Bayesian Belief Networks, are able to express both conditional dependent and independent variables connections. These type of networks, follow additionally the Markov condition [11] (provided the parents of every node in a network, each node is conditionally independent of their nondescendent nodes). Finally, using Bayes probabilistic approach (Equation 1), we can be able to update the connection probabilities iteratively based on new gathered evidence.

$$P(A|B) = \frac{P(B|A) \times P(A)}{P(B)} \quad (1)$$

where:

A, B = Events
$P(B|A)$ = Likelihood
$P(A)$ = Prior
$P(B)$ = Normalizing Constant
$P(A|B)$ = Posterior

Great research focus by companies such as DeepMind is currently put into using Bayesian Belief Networks as starting point in order to create Causal Bayesian Networks (CBN) [12]. Causal Bayesian Networks, are nowadays used in order to visually identify and quantitatively measure unfairness patterns in datasets (elements in the data which can lead to Machine Learning models biased towards specific sub-categories). Additionally, researches also demonstrated the possibility to use Causal Bayesian Networks in order to identify if not just the data but also the Machine Learning models itself are biased or not towards specific classes [13].

C. Extras

Graphical Methods have been of great importance in the last few years in order to start applying causality to Machine Learning. Although, in order for us to move from the Association to the Intervention level in the Causality Hierarchy, alternative approaches might be necessary. Some additional techniques which are commonly used in Explainable AI and Causality in order to answer Intervention types of questions (eg. What if?) are:

- Feature Selection techniques (eg. Recursive Feature Elimination, Shapley Values).
- Global and Local Model Surrogate (eg. Local Interpretable Model-Agnostic Explanations).
- Bias in AI (eg. Pre-processing, In-processing, Post-processing algorithms).
- Modelling Hidden Variables (eg. Hidden Markov Models, Boltzmann Restricted Machine).

These and other further approaches will be considered throughout this research study. Finally, apart from Machine Learning, Causal Inference can also be applied to other fields of Artificial Intelligence such as Reinforcement Learning. In fact, in order for agents to achieve good performances in an environment, they need to be able to think about what consequences would their action lead to [14], therefore requiring causal abilities belonging to the Counterfactual hierarchical level. Additionally, causality can be used in this ambit also to create causal partial models to predict high dimensional future observations in lower dimensional spaces [15].

D. Comparison

From a statistical and a research point of view, Graphical Methods and modelling techniques to identify hidden variables and biases are now representing an area of growing interest since they are related to areas which have not been explored yet as much as Machine Learning in past decade (although being still possibly able to be integrated with these techniques). On the other hand, Feature Selection and Global/Local Model Surrogate techniques are methods commonly used nowadays in Deep Learning problems in order to make complex models more easy to analyse to understand their decision making process (eg. finding out which features had greater weight when making a prediction and using surrogate models to create linear models at a local scale for non-linear problems).

IV. Conclusion

Causality has been researched and used for many years in statistics but not yet much in Artificial Intelligence. Identifying useful connections between these two different ambit could therefore play a vital role in making a new breakthrough towards creating intelligent systems.

References

[1] To Build Truly Intelligent Machines, Teach Them Cause and Effect Quanta Magazine. Accessed at: *https://www.quantamagazine.org/to-build-truly-intelligent-machines-teach-them-cause-and-effect-20180515*, March 2020.

[2] Inductive Inference Jan-Willem Romeijn, in Philosophy of Statistics, 2011. ScienceDirect Accessed at: *https://www.sciencedirect.com/topics/mathematics/inductive-inference*, March 2020.

[3] Human-level concept learning through probabilistic program induction Brenden M. Lake, Ruslan Salakhutdinov, Joshua B. Tenenbaum. Accessed at: *https://web.mit.edu/cocosci/Papers/Science-2015-Lake-1332-8.pdf*, March 2020.

[4] The Seven Tools of Causal Inference with Reflections on Machine Learning JUDEA PEARL, UCLA Computer Science Department, USA. Accessed at: *https://ftp.cs.ucla.edu/pub/stat*$_s er/r481.pdf$, March 2020.

[5] Systems Thinking Systems Modelling A Course for Understanding Systems and Creating Systems Models. The Sustainability Laboratory's. Accessed at: *https://systemsinnovation.io/system-dynamics-book/*, March 2020.

[6] CAUSALITY FOR MACHINE LEARNING Bernhard Schölkopf, Max Planck Institute for Intelligent Systems. Accessed at: *https://arxiv.org/pdf/1911.10500.pdf*, March 2020.

[7] Reinforcement Knowledge Graph Reasoning for Explainable Recommendation Yikun Xian et. al. Accessed at: *https://arxiv.org/pdf/1906.05237.pdf*, March 2020.

[8] A Review of Relational Machine Learning for Knowledge Graphs Maximilian Nickel, Kevin Murphy, Volker Tresp, Evgeniy Gabrilovich. Accessed at: *https://arxiv.org/pdf/1503.00759.pdf*, March 2020.

[9] KGCNs: Machine Learning over Knowledge Graphs with TensorFlow James Fletcher, Grakn Labs. Accessed at: *https://blog.grakn.ai/kgcns-machine-learning-over-knowledge-graphs-with-tensorflow-a1d3328b8f02*, March 2020.

[10] Knowledge Graphs and Causality Gianmario Spacagna, Vademecum of Practical Data Science. Accessed at: *https://datasciencevademecum.com/2019/12/19/knowledge-graphs-and-causality/*, March 2020.

[11] Introduction to Causal Inference Peter Spirtes, Department of Philosophy, Carnegie Mellon University. Accessed at: *http://www.jmlr.org/papers/volume11/spirtes10a/spirtes10a.pdf*, March 2020.

[12] Causal Bayesian Networks: A flexible tool to enable fairer machine learning Silvia Chiappa and William Isaac, DeepMind. Accessed at: *https://deepmind.com/blog/article/Causal*$_B ayesian_N etworks$, March 2020.

[13] Path-Specific Counterfactual Fairness Silvia Chiappa and Thomas P. S. Gillam, DeepMind. Accessed at: *https://arxiv.org/pdf/1802.08139.pdf*, March 2020.

[14] Causal Reasoning from Meta-reinforcement Learning Ishita Dasgupta et. al. Deep Mind. Accessed at: *https://arxiv.org/pdf/1901.08162v1.pdf*, March 2020.

[15] Causally Correct Partial Models for Reinforcement Learning Danilo J. Rezende, Ivo Danihelka et. al. Accessed at: *https://arxiv.org/pdf/2002.02836.pdf*, March 2020.