Explaining Bayesian Networks in Natural Language

Conor Hennessy [1]

Abstract. The advent of Artificial Intelligence has been one of the great achievements in Computer Science. The increase in usage of AI systems has given birth to a new challenge; explainability. This extended abstract serves to highlight the importance of explainability in AI, and to examine the challenge it poses. The abstract will have a particular focus on Bayesian Networks and their usage. I will discuss the state of the art in approximate reasoning, namely Bayesian Reasoning and Fuzzy Syllogistic Reasoning, as well as describe a proposed approach to the Natural Language Explanation of Bayesian Networks.

1 INTRODUCTION

In recent times there has been an obvious explosion in the usage of AI models in various situations. In the case of highly complex models, the reasoning or methodology behind the decisions of a model may remain unknown to the end user. It is true that there may exist certain use cases where such a scenario of a lack of understanding about the decision process of a model is acceptable. What is clear, however, is that the potential inability to explain the reasoning behind a model's results serves as a roadblock to the future usage of machine learning in general. In fact, the inability to produce an explanation of the decision of a model may even impinge on the rights of a European citizen in certain circumstances [15] .

A natural example of the conundrum of explainability can be drawn from the medical field. Take for example a model which has learned to diagnose chest pathoses such as lung nodules or cancer, pulmonary embolism or airway diseases [10]. If such a tool were being used by a medical professional to diagnose a patient based on the patient's X-ray, the question of the accuracy would, of course, be a crucial one. Explainability, however, would also be highly important. The medical professional should feel confident in the reasoning of the model and the patient should have peace of mind that the diagnosis is reliable and logical. A method of generating Natural Language explanations of such a model reasoning would alleviate this problem, and is the motivation for this extended abstract.

The model that is the focus of this abstract is the Bayesian Network (BN). In addition to medical applications [7], BN's have found use in evidential reasoning in Law [2] and in DNA profiling in criminal forensics [5], amongst other uses. I will focus on a use case of predictive inference in this abstract. While Using BN's as a model carries certain advantages, which shall be discussed, its use also carries its own challenges. The graphical nature of a BN can aid in the intuition of a system, and there are several graphical tools, such as GeNIe, that can provide understanding for the user. Graphical aids, however, do not provide sufficient knowledge about the model for users not familiar with probabilistic reasoning. Graphs can be misleading, and conditional probability tables are not digestible for an average user. Bayesian reasoning in particular is challenging and often not intuitive.

With an automatic Natural Language Generation method to explain the BN, there can be much more widespread utilization of BN's and higher level of clarity in their usage. I will outline the inner workings of a predictive BN, how the information of a BN can be condensed into Fuzzy quantified syllogisms, and how this could then be used in the Natural Language Explanation.

[1] Centro Singular de Investigacion en Tecnoloxías Intelixentes (CiTIUS), Universidade de Santiago de Compostela, email: conor.hennessy@usc.es

2 KNOWLEDGE REPRESENTATION AND REASONING

Before an approach to Natural Language Generation can be used to explain a BN, the knowledge and reasoning of a BN must be represented so that a computer system can understand them. Bayesian Reasoning is what is represented in the BN. The starting point for my research will be to model the BN as a quantified syllogism and to solve this syllogism. To do this, a method of representing Bayesian Reasoning must be used. As a first step, I will be investigating Fuzzy Quantified Syllogism.

2.1 Bayesian Reasoning and Bayesian Networks

In classical Bayesian Inference, the Posterior Probability is calculated by including a prior parameter, designed to be a measure of the prior information that is known, which can then be updated as more evidence is gathered [6]. Bayesian networks are Directed Acyclic Graphs (DAG's), representing a set of Random Variables and the dependencies between these Random Variables [11]. The Random Variables are shown as nodes in the DAG, while the edges between these nodes demonstrate the dependencies. The state of a node might affect the probability of another node, depending on the relationship between the nodes. Similarly, the probability that a node is in one state depends on the state of another node according to prior information about the relationships among the nodes. The direction of the arrow on each edge indicates the direction of causality [4], though for certain models, such as those predicting risk, edges may not indicate causality.

Using BN's as a model is advantageous in several ways. They are excellent in modelling potential cause and effect [4] and in risk prediction [1]. They also allow for the encoding of prior knowledge in the model, and the explicit definition of dependencies between nodes leads to more compact graphs [1].

In [1] a predictive BN is introduced, outlining the relationship between patient characteristics, disease and test results. Using the model, predictions can be made for the patient regarding the disease

Proceedings of the 1^{st} Doctoral Consortium at the European Conference on Artificial Intelligence (DC-ECAI 2020), pages 59–60, Santiago de Compostela, Spain, August 29-30, 2020.

in question. It serves as an excellent example for visualising the process, as the predictors, the hierarchy of dependencies and the conditional probability table are clearly presented. It is important to note that the methodology for representing a predictive BN introduced in this abstract is also a potential methodology for diagnostic and intercasual BN's. The first step I will be investigating is how to represent this information in quantified statements.

2.2 From Bayesian Reasoning to Fuzzy Syllogistic Reasoning

There have been several proposals of methodologies to represent information contained in a BN, such as the qualified support graph method discussed in [3]. The use of Fuzzy syllogisms as a method to represent BN's, however, carries the following potential advantages.

Unlike Boolean Logic, where the output is either true or false, Fuzzy Logic can represent levels of uncertainty or vagueness, as the output can be any $x \in [0, 1]$, where $x \in R$. Natural Language also involves this uncertainty and vagueness. In the case of quantitative linguistic variables, such as with the example of height in [13], they too can be often be fuzzy variables. In the given example of height, tall, very tall, and very very tall all signify varying levels of precision. In the same way, probabilistic statements can be considered fuzzy, for example "approximately .." or "highly unlikely..". As fuzzy logic mimics the human decision making process more closely, it has found use in AI as a form of approximate reasoning [12]. For this reason, it could serve as an excellent tool to bridge the gap between the BN reasoning process and formulating a human language explanation.

Though the connection between probability theory and Fuzzy Logic is a long established one [14], there has been renewed interest in the relationship between Bayesian Reasoning and Fuzzy Syllogistic Reasoning. In order to connect the two concepts, I will take as a starting point the enhanced Syllogistic Reasoning Schema introduced in [8], which allows for any number of variables and several quantifier types, allowing the conditions for the BN to be represented. In [9] a blueprint is then laid out for extracting quantified statements from a BN to form a Fuzzy Syllogism, and compiling a knowledge base for use in an approach to creating a natural language explanation of the BN.

3 EXPLAINABILITY

Once the Bayesian Reasoning process is translated to a Fuzzy Quantified Syllogism, and the Knowledge Base is generated, this would form the first stage in the process of the automatic Natural Language explanation of the BN.

In [2], a 4-step architecture for the Natural Language explanation of a BN is outlined. The Fuzzy Syllogistic approach to BN explanation discussed in [9] focuses on Step 1 from [2], namely the Content Determination phase. Content Determination amounts to the extraction of the information required for the Natural Language Explanation, utilizing the quantified statements.

In the case of template based Natural Language Generation, steps 2 - 4 are normally combined into a single step, namely Linguistic Representation [9]. The Fuzzy Syllogistic approach described above would form the first step in a template based automatic description system, followed by the Linguistic Representation step. In the Linguistic Representation step, a full Natural Language Explanation is generated using the quantified statements compiled in the previous step.

4 FUTURE WORK

As this abstract serves only as a presentation of the problem of explainability of BN's, I will be studying the topics introduced here in more depth. Following this, I hope to define and implement a methodology for explaining in Natural Language the various types of BN reasoning, including predictive, diagnostic and inter-causal reasoning. As BN's grow, scalability becomes an issue, with both the graphical and probability table components of the BN rapidly expanding to cumbersome levels. This issue must also be addressed in future work. The Natural Language Generation methodology can then be implemented in both academic and industrial use cases, with the testing/validation process being assessed by a human user.

ACKNOWLEDGEMENTS

This extended abstract is associated with the NL4XAI project. This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 860621.

REFERENCES

[1] Paul Arora, Devon Boyne, Justin Slater, Alind Gupta, Darren Brenner, and Marek Druzdzel, 'Bayesian networks for risk prediction using real-world data: A tool for precision medicine', *Value in Health*, **22**, (03 2019).
[2] Jeroen Keppens, 'Explaining bayesian belief revision for legal applications', in *JURIX*, (2016).
[3] Jeroen Keppens, 'Explainable bayesian network query results via natural language generation systems', pp. 42–51, (06 2019).
[4] Lexin Liu, 'A software system for causal reasoning in causal bayesian networks', (2008).
[5] Julia Mortera, Alexander Dawid, and Steffen Lauritzen, 'Probabilistic expert system for dna mixture profiling', *Theoretical population biology*, **63**, 191–205, (06 2003).
[6] Svein Nyberg, *Bayes' Theorem*, 107–135, 08 2018.
[7] Bo Pang, David Zhang, Naimin Li, and Kuanquan Wang, 'Computerized tongue diagnosis based on bayesian networks', *IEEE Transactions on Biomedical Engineering*, **51**, (11 2004).
[8] Martin Pereira, Juan Vidal, F. Díaz-Hermida, and Alberto Bugarín, 'A fuzzy syllogistic reasoning schema for generalized quantifiers', *Fuzzy Sets and Systems*, **234**, 79–96, (11 2014).
[9] Martín Pereira-Fariña and Alberto Bugarín, 'Content determination for natural language descriptions of predictive bayesian networks', in *11th Conference of the European Society for Fuzzy Logic and Technology (EUSFLAT 2019)*, pp. 784–791. Atlantis Press, (2019/08).
[10] Edwin J. R. van Beek and John T. Murchison, *Artificial Intelligence and Computer-Assisted Evaluation of Chest Pathology*, 145–166, Springer International Publishing, Cham, 2019.
[11] Davy Weissenbacher, 'Bayesian network, a model for nlp?', (01 2006).
[12] John Yen and Reza Langari, *Fuzzy logic: intelligence, control, and information*, volume 1, Prentice Hall Upper Saddle River, NJ, 1999.
[13] L. A. Zadeh, 'Outline of a new approach to the analysis of complex systems and decision processes', *IEEE Transactions on Systems, Man, and Cybernetics*, **SMC-3**(1), 28–44, (1973).
[14] Lotfi A. Zadeh, 'Discussion: Probability theory and fuzzy logic are complementary rather than competitive', *Technometrics*, **37**(3), 271–276, (1995).
[15] Scott Zoldi. Explainable ai: Implications for compliance with gdpr and beyond. https://www.linkedin.com/pulse/explainable-ai-implications-compliance-gdpr-beyond-scott-zoldi/. Accessed: 2020-06-20.

Proceedings of the 1^{st} Doctoral Consortium at the European Conference on Artificial Intelligence (DC-ECAI 2020), pages 59–60, Santiago de Compostela, Spain, August 29-30, 2020.