Motivating explanations in Bayesian networks using MAP-independence

Johan Kwisthout[1 [0000−0003−4383−7786]]

Donders Institute for Brain, Cognition, and Behaviour, Radboud University, Nijmegen, The Netherlands `johan.kwisthout@donders.ru.nl`
http://www.socsci.ru.nl/johank/

Abstract. This is an extended abstract of [4], enriched with a discussion of follow-up results and algorithmic considerations on the topic of MAP-independence in Bayesian networks.

Keywords: Bayesian Networks · Most Probable Explanations · Relevance · Explainable AI · Computational Complexity

1 Motivation

In decision support systems the motivation and justification of the system's diagnosis or classification is crucial for the acceptance of the system by the human user. In Bayesian networks a diagnosis or classification is typically formalized as the computation of the most probable joint value assignment to the hypothesis variables, given the observed values of the evidence variables (generally known as the MAP problem). While solving the MAP problem gives the most probable explanation of the evidence, the computation is a black box as far as the human user is concerned and it does not give additional insights that allow the user to appreciate and accept the decision. For example, a user might want to know to whether an unobserved variable could potentially (upon observation) impact the explanation, or whether it is irrelevant in this aspect. In this paper ([4]) we introduce a new concept, MAP-independence, which tries to capture this notion of relevance, and explore its role towards a potential justification of an inference to the best explanation.

2 Formal definition

Given a Bayesian network $\mathcal{B} = (\mathbf{G}, \Pr)$ with observation variables $\mathbf{E}$ and hypothesis variables $\mathbf{H}$, the MAP problem establishes, for some observation $\mathbf{E} = \mathbf{e}$, $\mathbf{h}* = \text{argmax}_{\mathbf{h}} \Pr(\mathbf{H} = \mathbf{h}, \mathbf{E} = \mathbf{e})$, i.e., the most probable explanation for the evidence. Given $\mathbf{h}*$ and a set of potential (unobserved) variables $\mathbf{R} \subseteq \mathbf{V}(\mathbf{G}) \setminus \{\mathbf{E} \cup \mathbf{H}\}$, the MAP-independence problem decides whether any observation $\mathbf{r}$ to $\mathbf{R}$ may change the most probable explanation (making $\mathbf{R}$ *relevant* for explaining the evidence); the Maximum MAP-independence problem seeks to find the largest subset of $\mathbf{R}$ which is MAP-independent relative to the explanation.

3 Algorithms and analysis

In general, these problems are intractable, more in particular $\mathsf{NP}^{\mathsf{PP}}$-hard or co-$\mathsf{NP}^{\mathsf{PP}}$-hard, as we prove in the original paper (section 4.1). There we also present brute-force approaches, with as primary goal to establish fixed-parameter tractability results. Bounding $\mathbf{R}$ in addition to the constraints needed to render MAP tractable (see section 4.2 of the full paper) is necessary for tractability.

Maximum MAP-independence is reduced from a novel satisfiability variant, named Partition-free A-Majsat, that introduces 'partition selection' as part of the problem definition: Is there a non-trivial partition $\{\mathbf{X_A}, \mathbf{X_M}\}$ of the variables $\mathbf{X}$ of a formula ϕ, such that for all truth assignments to $\mathbf{X_A}$, the majority of truth assignments to $\mathbf{X_M}$ satisfies ϕ? This is to be contrasted with traditional canonical complete SAT variants in the polynomial hierarchy where such a partition is part of the input. The paper conjectures $\mathsf{NP}^{\mathsf{NP}^{\mathsf{PP}}}$-completeness of this problem (and thus of Maximum MAP-independence) but a proof thereof was out of reach as the absence of a partition makes a Cook-style proof challenging.

Perhaps more interesting from an application side of view are the follow-up results by [7, 8], consolidated in a jointly authored journal paper [9]. Here, algorithms are presented to find singleton variables that are MAP-independent, and then explore the search space of super-sets of these variables that are potentially part of the maximum MAP-independent set $\mathbf{R}$. This approach uses 'computational by-products' of the original MAP computation, and is further augmented with search tree pruning techniques.

4 Follow-up results

In the concluding section of [4] we briefly hinted at future work where the relevance of the *existing evidence* for MAP is assessed. This naturally complements the MAP-independence of *unobserved* variables with the dual question: Rather than 'If I were to observe this (yet unobserved) variable, might that change the MAP?' the question now becomes 'Had I not observed this variable, would that have changed the MAP?' In general, like the complement problem, this question turns out to be intractable in general (i.e., $\mathsf{NP}^{\mathsf{PP}}$-hard); proof of this claim is given in the supplementary material[1]. Obviously, in specific cases one can (e.g.) establish that an evidence node E is d-separated from the rest of the evidence, and thus irrelevant for establishing the MAP explanation [5].

5 Future work

In addition to the further exploration of *quantifying* MAP-independence, as proposed in [4] discussing [1, 2, 8], the follow-up result invites for a more elaborate complexity analysis of the decisiveness of some (potential) observation for an explanation and the formal framework necessary to be able to prove completeness for appropriate complexity classes.

[1] http://www.socsci.ru.nl/johank/material/bnaic23/supplement.pdf

References

1. van Elteren, M.: Comparing quantified MAP-dependence to other measures of relevance in Bayesian networks. BSc. thesis in Artificial Intelligence, Radboud University (2022)
2. Jacobs, L.: What factors influence the perceived optimal threshold value in the visualization of quantified MAP-dependence? BSc. thesis in Artificial Intelligence, Radboud University (2022)
3. Kwisthout, J.: Most Probable Explanations in Bayesian networks: Complexity and tractability. International Journal of Approximate Reasoning **52**(9) (2011)
4. Kwisthout, J.: Motivating explanations in Bayesian networks using MAP-independence. International Journal of Approximate Reasoning **153**, 18–28 (2023). https://doi.org/10.1016/j.ijar.2022.11.012
5. Meekes, M., Renooij, S., van der Gaag, L.: Relevance of evidence in Bayesian networks. In: European Conference on Symbolic and Quantitative Approaches to Reasoning and Uncertainty. pp. 366–375. Springer (2015)
6. Park, J., Darwiche, A.: Complexity results and approximation settings for MAP explanations. Journal of Artificial Intelligence Research **21**, 101–133 (2004)
7. Renooij, S.: Relevance for robust Bayesian network MAP-explanations. In: Salmerón, A., Rumí, R. (eds.) Proceedings of The 11th International Conference on Probabilistic Graphical Models. Proceedings of Machine Learning Research, vol. 186, pp. 13–24. PMLR (2022)
8. Valero-Leal, E., Bielza, P.L..C.: Extending MAP-independence for Bayesian network explainability. In: Proceedings of the first international workshop on Heterodox Methods for Interpretable and Efficient AI (2022)
9. Valero-Leal, E., Bielza, C., Larrañaga, P., Renooij, S.: Efficient search for relevance explanations using MAP-independence in Bayesian networks. International Journal of Approximate Reasoning **160**, 108965 (2023). https://doi.org/10.1016/j.ijar.2023.108965