The Thirty-Second International Florida
Artificial Intelligence Research Society Conference (FLAIRS-32)

Towards Interactive Causal Relation Discovery Driven by an Ontology

Melanie Munch,[1] Juliette Dibie,[1] Pierre-Henri Wuillemin,[2] Cristina Manfredotti[1]
[1]UMR MIA-Paris, AgroParisTech, INRA, University of Paris-Saclay, 75005 Paris, France
[2]Sorbonne University, UPMC, Univ Paris 06, CNRS UMR 7606, LIP6, 75005 Paris, France
melanie.munch@agroparistech.fr, juliette.dibie@agroparistech.fr,
pierre-henri.wuillemin@lip6.fr, cristina.manfredotti@agroparistech.fr

Abstract

Discovering causal relations in a knowledge base represents nowadays a challenging issue, as it gives a brand new way of understanding complex domains. In this paper, we present a method to combine an ontology with a probabilistic relational model (PRM), in order to help a user to check his/her assumption on causal relations between data and to discover new relationships. This assumption is important as it guides the PRM construction and provide a learning under causal constraints.

Introduction

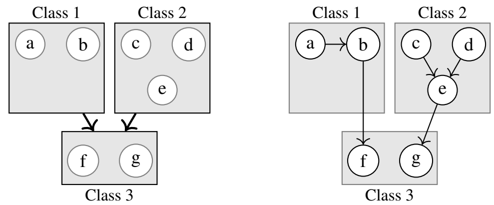
In order to analyze and understand complex domains, a good representation of the causal relations between the different variables considered is valuable. In this article, we introduce a method that offers a probabilistic reasoning over a knowledge base structured by an ontology in order to discover new causal relations. Ontologies allow data and expert knowledge to be gathered and semantically organized, thus allowing a better understanding of complex domains. However, they cannot provide complex probabilistic reasoning. We propose to achieve this by using probabilistic relational models (PRMs) (Friedman et al. 1999). PRMs extend BNs with the notion of classes from the domain of relational databases, thus allowing a better representation between the different attributes. However, due to this specificity their learning can be tricky. Using the semantic and structural knowledge contained in a knowledge base structured by an ontology, this learning can be greatly eased and, thus, be guided toward a learned model close to the reality described by the ontology (Munch et al. 2017). However, different PRMs can be defined from a same knowledge base. Thus, in order to select one, we consider a causal assumption given by a user (a domain expert) of the form "*Does attribute A have a causal influence over attribute B?*" that he wants to be checked as true or false. The first section of this paper presents the background and state of the art, especially on PRM and causal discovery. The second section presents our approach to learn a PRM from an ontology guided by a user's causal assumption, and an experiment on a transformation process. The last section concludes this paper.

Background and State of the Art

A BN is the representation of a joint probability over a set of random variables that uses a Directed Acyclic Graph (DAG) to encode probabilistic relations between variables. However, in this paper we need to group attributes by specific causal relations and BN lack the notion of modularity. PRMs extend the BN representation with a relational structure between potentially repeated fragments of BN called classes (Torti, Wuillemin, and Gonzales 2010). PRMs are defined by two parts: a high-level, qualitative description of the structure of the domain that describes the classes and their attributes (i.e. the relational schema RS as shown Fig. 1 (a)), and a low-level, quantitative information given by the probability distribution over the different attributes (i.e. its relational model RM as shown in Fig. 1 (b)). Once instantiated the classes are equivalent to a BN.

An essential graph (EG) is a semi-directed graph associated to a BN. They both share the same skeleton, but the orientation of the EG's edges can vary. If the orientation of an edge is the same for all the BNs in the same Markov equivalence class, then it is also oriented in the EG (they are called essential arcs (Madigan et al. 1996)); if not, it remains unoriented. This way the EG expresses whether an orientation between two nodes can be reversed without modifying the probabilistic relations encoded in the graph: whenever the constraint given by an essential arc is violated, the conditional independence requirements are changed and the structure of the model itself has to be changed.

Numerous related works have established that using constraints while learning BNs brings more efficient and accurate results, for parameters learning (De Campos and Ji 2008) or structure learning (De Campos, Zhi, and Ji 2009). In this article we define structural constraints as an ordering between the different variables. The K2 algorithm (Cooper and Herskovits 1992), for instance, requires a complete ordering of the attributes before learning a BN, allowing the introduction of precedence constraints between the attributes. This particular algorithm needs a complete knowledge over all the different attributes precedences; however problems of learning with partial order have also been tackled (Parviainen and Koivisto 2013). In our case we will likewise transcribe incomplete knowledge as partial structural organization for the PRM's relational schema in order to discover new causal relations.



(a) Relational schema (b) Relational model

Figure 1: The high (a) and low (b) level structures of a PRM

Causal models are DAGs allowing one to express causality between its different attributes (Pearl 2009). Their construction is complex and requires interventions or controlled randomized interventions, which are often difficult or impossible to test. As a consequence the task of discovering causal relations using data, known as causal discovery, has been researched in various fields over the last few years. There are two types of methods for structure learning from data: independence-based ones, such as the PC algorithm (Spirtes, Glymour, and Scheines 2000), and score-based ones, such as Greedy Equivalent Search (GES) (Chickering 2003). Usually independence-based methods give a better outlook on causality between the attributes by finding the *"true"* arc orientation, while the score-based ones offer a structure that maximizes the likelihood considering the data. Finally, other algorithms such as MIIC (Verny et al. 2017) use independence-based algorithms to obtain information considered as partially causal and thus allowing to discover latent variables. In this article we propose to explore if combining ontological knowledge and a user's causal assumption with BN learning score-based algorithms allows causal discovery. Other works have already proposed the use of EG: (Hauser and Bühlmann 2014) for instance proposes two optimal strategies for suggesting interventions in order to learn causal models with score-based methods and the EG. Integrating knowledge in the learning has also been considered: (Ben Messaoud, Leray, and Ben Amor 2009) offers a method to iterative causal discovery by integrating knowledge from beforehand designed ontologies to causal BN learning, and (Amirkhani et al. 2017) proposes two new scores for score-based algorithms using experts knowledge and their reliability. While PRM offers a way to express and consider the expert knowledge in learning, to the best of our knowledge no learning causality method that combines PRM and ontological knowledge and is guided by a user's causal assumption has been proposed yet.

Causal Discovery Driven by an Ontology

In this article we present a four-steps method in order to learn a PRM using ontological kowledge guided by a causal assumption: (1) the user expresses expert knowledge in a causal form "*The attribute A has a causal influence over the attribute B"* that he/she wants to check in a given knowledge base structured by an ontology; (2) the attributes of the user's causal assumption are used to define, from the knowledge base, the attributes of two classes in the RS, the explaining and consequence classes; (3) the attributes previously defined for each class of the RS are enriched with new attributes from the knowledge base, judged as interesting by the user for the causal assumption's study; (4) using the defined RS a PRM is learned, whose analysis will help us validate the user's causal assumption and in the end uncover new causal relations.

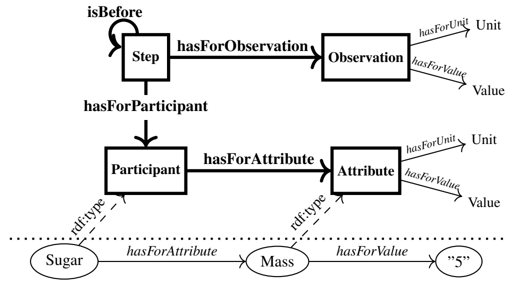


Figure 2: Excerpt of a knowledge base about transformation processes

Preliminaries Definitions

A knowledge base $\mathcal{KB}$ is defined by a couple ($\mathcal{O}$, $\mathcal{F}$) where:

- the ontology $\mathcal{O} = (\mathscr{C}, DP, OP, A)$ is defined in OWL[1] by a set of classes $\mathscr{C}$, a set of *owl:DataTypeProperty* DP in $\mathscr{C} \times T_D$ with T_D being a set of primitive datatypes (e.g. integer, string), a set of *owl:ObjectProperty* OP in $\mathscr{C} \times \mathscr{C}$, and a set of axioms A (e.g. subsumption, property's domains and ranges).
- the knowledge graph $\mathcal{F}$ is a collection of triples (s, p, o) in RDF[2], called instances, where s is the subject of the triple, p is a property that belongs to $DP \cup OP$ and o is the object of the triple; for a triple (s, p, o), we note $domain(p) = s$ and $range(p) = o$.

Fig. 2 gives an excerpt of the PO[2] ontology[3] dedicated to transformation processes (on the top) associated with a small example of $\mathcal{F}$ in the bottom. $\mathcal{O}$ is composed of four main classes: the **step** class, that defines the different steps and how they are linked together in time; the **participant** class, that defines the different objects used during the step (e.g. electronic scale, mixtures); the **observation** class, that defines the observations made on the participants. The class **attribute** characterizes the participants. Using the previous notations, *Step* $\in \mathscr{C}$, *Unit* $\in T_D$, *hasForParticipant* $\in OP$, *hasForValue* $\in DP$. Fig. 3 gives an example of a knowledge graph using this ontology.

The **user's causal assumption** $\mathcal{H}$ is of the form: "E_1, ..., E_n *have a causal influence on* C_1, ..., C_p" with E_i an explaining attribute and C_j a consequence one. We denote the sets of the attributes of $\mathcal{H}$ as $A^{\mathcal{H}}_E = \{E_1, ..., E_n\}$ and $A^{\mathcal{H}}_C = \{C_1, ..., C_p\}$, with $A^{\mathcal{H}} = A^{\mathcal{H}}_E \cup A^{\mathcal{H}}_C$. Using the instantiated transformation process of Fig. 3, a user's assumption

[1] https://www.w3.org/OWL/

[2] https://www.w3.org/RDF/

[3] http://agroportal.lirmm.fr/ontologies/PO2

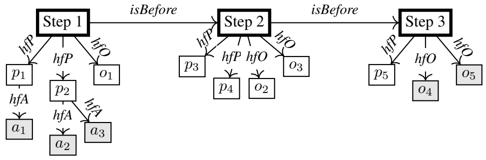


Figure 3: Example of a transformation process using the PO^2 ontology. *hfP: hasForParticipant, hfO: hasForObservation, hfA: hasForAttribute*

$\mathcal{H}_f$ over our PO^2 example could be : "*The attributes of* p_1 *and* p_2 *have an influence over* o_4, o_5", with $A_E^{\mathcal{H}} = \{a_1, a_2, a_3\}$ and $A_C^{\mathcal{H}} = \{o_4, o_5\}$ (grayed out on the figure). In order to construct a PRM as close as possible to the causal assumption provided by the user, we define a generic RS composed of two different type of classes, the **explaining** and the **consequence**, whose attributes are respectively denoted as explaining and consequence attributes. Distinguishing between them influences the causal discovery: if a relation is found between an explaining and a consequence attribute, the direction of causality is automatically determined from the explaining attributes to the consequence ones. We define this as a **causal constraint**. The class order guides the PRM learning, as we restrain our set of possible structures only to those that respect these causal constraints. Once the RS defined, we need to select attributes to fill the classes. Since the probabilistic dependencies are learned using a score-based Bayesian learning method (denoted M), then it depends on statistical evaluation. Thus not all attributes from a knowledge base can be selected: they have to fit certain conditions, and be useful for the learning. In a knowledge base $\mathcal{KB}$ we call **useful learning attribute** an attribute a that is not constant and whose set of values is bound. These useful learning attributes correspond in $\mathcal{KB}$ to datatype properties $p \in DP$. In our example (Fig. 3), if we consider that all instances of a same attributes have the same unit, then the datatype property *hasForUnit* is not useful.

Assumption's Attributes Identification

In order to identify the attributes of the explaining and consequence classes of the RS, we propose to build the set $S_{\mathcal{H}}^{\mathcal{KB}}$ of all useful learning datatype properties of $\mathcal{KB}$ corresponding respectively to the explaining and consequence causal assumption's attributes. To do so we start from each attribute a of $A^{\mathcal{H}}$ of the assumption $\mathcal{H}$, and construct the set S^a of its corresponding datatype properties in $\mathcal{KB}$. First we use a similarity measure (e.g. Jaccard measure) to compute for each attribute $a \in A^{\mathcal{H}}$ the similarity between its name and a $\mathcal{KB}$ entity's label. If it is higher than α (α experimentally fixed in [0,1]), we have: (i) if the entity is a datatype property, it is added to S^a; (ii) if the entity is a class, all of its datatype properties are added to S^a; (iii) if the entity is an object property, its range and domain classes are gathered, and we apply (ii). Second, for each datatype property added we check whether they are useful for the learning and, if not, we exclude them from the set. For all S^a we also verify that a connected knowledge graph can be constructed from their union, to prevent cases where each datatype property has individually enough instantiations, but not enough global instances that link them together. Last, the user checks each S^a and can choose to exclude datatype properties he judges inadequate. In the end, for each attribute a, its set S^a is either entirely checked or empty: in this last case, it means that the attribute a is not relevant for $\mathcal{KB}$, and that $\mathcal{H}$ cannot be checked. In our example, $\mathcal{H}_f$ defines three participants a_1, a_2 and a_3 considered as explaining, and two observations o_4 and o_5 considered as consequence. Only the useful *hasForValue* datatype property is selected. As a consequence $\mathcal{H}$ can be checked.

Enriching the Set of PRM Attributes

Most of the time the attributes expressed in $\mathcal{H}$ are not enough to find causal relations between data, requiring to find other useful learning attributes to improve the RS building. We make successive iterations on the knowledge graph over the properties, starting from the entities found previously, and following the other to which they are linked if there have enough instances. If we find a datatype property through a path with enough instances, it is added if it is useful for the learning and relevant. When adding a datatype property, the user has to decide in which class he wants to put it: if he doesn't know, it is put in the higher explaining class by default. In our PO^2 example the other participants' values attributes and observations are selected. The separation into steps induces the need for new classes: we want to be able to separate for each step explaining and consequence attributes. As a matter of fact, if we consider that each step happens at a distinct moment, and that attributes can only be explained by those that happen at the same time or before, then we need to define at least one explaining and one consequence class for each step (i.e. for each considered time). Fig. 4 (b) presents the RS defined to answer these constraints.

PRM Construction

We learn a PRM using our RS, a learning method M and the knowledge graph of $\mathcal{KB}$. In order for the user to check the model, we propose **an interactive and iterative method** based on the study of the EG. Considering that the PRM has been learned under causality constraints (given by the user), the EG helps to determine causal relations: if an edge is oriented in the EG, then it is said causal assuming that (1) the data we dispose is representative of the reality, (2) all the attributes interesting for the problem are represented and (3) the causal information brought by the user is considered as true. We make two verifications: a first one for the inter-classes relations, and a second one for the intra-class relations. The EG **inter-classes** relations are the first to be presented to the user since they are the one he had direct control over: if he detects a wrong orientation, it means that the RS has been badly constructed and has to be modified. The **intra-class** relations are then presented. In the case of a **non oriented** relation in the EG, the user can choose to keep its orientation as it is in the learned PRM or inverse it, which would require a modification of the current RS. Likewise, if a relation is **oriented**, then the user can also choose to keep this orientation, or declare it wrong according to his

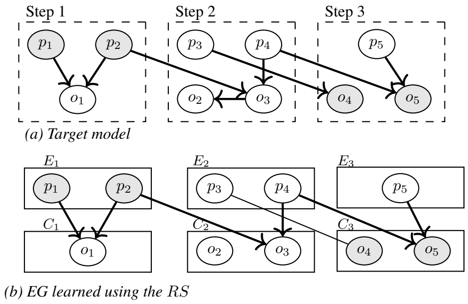


Figure 4: Comparison of the model used to generate the database (a) with the learned EG using the RS (b).

knowledge of the domain. In this last case a modification of the RS is also required. When **modifying the** RS, two cases are possible: if one of the node only needs to change class, then the same class structure are kept in the RS; otherwise new classes need to be introduced.

From the knowledge graph of Fig. 3 and the probabilistic relations that we have defined in Fig. 4 (a), we generate a dataset of 5000 different instances (165,000 RDF triplets) and apply our mmethod. Fig. 4 (b) shows the EG learned. All relations except for one inter-class are oriented, meaning that considering our knowledge base, the constraints brought both by the ontology (i.e. time constraint) and the causal assumption, only one result is possible. Using it, we can see that p_1 and p_2 do not explain o_4 and o_5: $\mathcal{H}_f$ is therefore not checked.

Conclusion

In this article we present a method to integrate expert knowledge into a learning in order to discover new causal relations in a knowledge base. A user's causal assumption helps define the RS of a PRM, used for its learning. Since the RS has been defined under causal constraints we deduce that the EG's oriented arcs of the PRM transcribe potential causal relations. In this paper we presented a possible application using a knowledge base about transformation processes and data generated from the PO^2 ontology. On an other experiment we applied our method on the part of the DBPedia ontology dedicated to films (which represented 90,000 RDF triplets for 10,000 films). With our two experiments[4] we have shown that (1) this method offers an interactive and iterative solution to integrate expert knowledge to a causal discovery task; and (2) this integration of expert knowledge bring an overall improvement of the quality of the model learned. This work is a first step to tackle interactive and iterative machine learning combining ontology and PRM in order to improve the learned relational model using expert knowledge. Our future work will focus on the iterative part studying BN's introspection to find and explain causality relations in a semi-automatic way allowing an ontology's enrichment with expert rules that will improve the learning.

[4] All data and code relevant to the experiments are available at: https://bit.ly/2RYVjG8

References

Amirkhani, H.; Rahmati, M.; Lucas, P. J. F.; and Hommersom, A. 2017. Exploiting experts' knowledge for structure learning of bayesian networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 39(11):2154–2170.

Ben Messaoud, M.; Leray, P.; and Ben Amor, N. 2009. Integrating ontological knowledge for iterative causal discovery and visualization. In Sossai, C., and Chemello, G., eds., *Symbolic and Quantitative Approaches to Reasoning with Uncertainty*, 168–179.

Chickering, D. M. 2003. Optimal structure identification with greedy search. *J. Mach. Learn. Res.* 3:507–554.

Cooper, G. F., and Herskovits, E. 1992. A bayesian method for the induction of probabilistic networks from data. *Machine Learning* 9(4):309–347.

De Campos, C. P., and Ji, Q. 2008. Improving bayesian network parameter learning using constraints. In *2008 19th International Conference on Pattern Recognition*, 1–4.

De Campos, C.; Zhi, Z.; and Ji, Q. 2009. Structure learning of bayesian networks using constraints. In *Proceedings of the 26th Annual International Conference on Machine Learning*, ICML '09, 113–120. New York, USA: ACM.

Friedman, N.; Getoor, L.; Koller, D.; and Pfeffer, A. 1999. Learning probabilistic relational models. In *Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence, IJCAI 99, Stockholm, Sweden, July 31 - August 6, 1999. 2 Volumes, 1450 pages*, 1300–1309.

Hauser, A., and Bühlmann, P. 2014. Two optimal strategies for active learning of causal models from interventional data. *Int. J. Approx. Reasoning* 55:926–939.

Madigan, D.; Andersson, S. A.; Perlman, M. D.; and Volinsky, C. T. 1996. Bayesian model averaging and model selection for markov equivalence classes of acyclic digraphs. *Communications in Statistics–Theory and Methods* 25(11):2493–2519.

Munch, M.; Wuillemin, P.-H.; Manfredotti, C.; Dibie, J.; and Dervaux, S. 2017. Learning probabilistic relational models using an ontology of transformation processes. In *On the Move to Meaningful Internet Systems. OTM 2017 Conferences*, 198–215.

Parviainen, P., and Koivisto, M. 2013. Finding optimal bayesian networks using precedence constraints. *Journal of Machine Learning Research* 14:1387–1415.

Pearl, J. 2009. *Causality: Models, Reasoning and Inference*. New York, USA: Cambridge University Press, 2nd edition.

Spirtes, P.; Glymour, C.; and Scheines, R. 2000. *Causation, Prediction, and Search*. MIT press, 2nd edition.

Torti, L.; Wuillemin, P.-H.; and Gonzales, C. 2010. Reinforcing the Object-Oriented Aspect of Probabilistic Relational Models. In *PGM 2010 - The Fifth European Workshop on Probabilistic Graphical Models*, 273–280.

Verny, L.; Sella, N.; Affeldt, S.; Singh, P. P.; and Isambert, H. 2017. Learning causal networks with latent variables from multivariate information in genomic data. *PLOS Computational Biology* 13(10):e1005662.