

Creating causal representations from ontologies and Bayesian networks

D. Cutic^{1,2}, G. Gini¹

¹ *DEIB, Politecnico di Milano, Milan, Italy, e-mail:gini@elet.polimi.it*

² *University of Zagreb, Computer Science*

Abstract. Inducing Bayesian network structure has been an active research topic since they were introduced, especially when considering their causal interpretation. There have been several attempts at guiding the learning process by providing additional knowledge, usually supplied by experts. In the recent years, ontologies have gained popularity in terms of publishing domain knowledge in a formal, systematic and machine understandable way. In this paper we focus on how to use the knowledge present in publicly available ontologies to assist the construction of causal Bayesian networks, considering the robotics scenario. We consider the needed assumptions to make the process sound and discuss about performances. We will conclude that the integration of ontologies and Bayesian nets is valid and has many potential directions for future extension.

Key words: Bayes nets, ontologies, causality.

1 Introduction

The popularity of Bayesian networks (BN) for representing uncertain knowledge, and the development of ontologies for storing machine readable and structured knowledge have made people to consider combined approaches of them. The main motivation for such integration is the high similarity of their underlying structures; this similarity allows for knowledge to be transferred both ways, and thus two main roads have been considered. One tries to incorporate uncertainty into ontologies while the other wants to exploit knowledge present in ontologies to guide the construction of the Bayesian network structure. In this paper we explore the latter one.

Interest in the causal interpretation of BNs is due to the stability of this relation; once we know there exist a causal relationship between two variables we know it to be an objective and physical constraint in the world. This comes at a price; in fact the task of finding and justifying such a relation has proven to be non-trivial, especially when the goal is to learn it from data. In that case, even under restrictive assumptions,

the process is driven by co variation and does not guarantee causality. On the other hand we can expect ontologies to comprise in themselves, among others, also causal knowledge. This is exactly the knowledge we want to exploit. We say “assist” because the approach is based around a standard structure learning algorithm which uses the ontology for tuning up its results. This approach leaves enough space for considering different interactions between the ontology and the structure learning algorithm.

Today robotics is approaching more and more the problem of knowledge representation and reuse to improve the capabilities of robot learning and social interactions. While the robotics problems are still dominated by unpredictable interactions with the real world, the societal robot is expected to interact with people sharing their reasoning and beliefs. To manage real data Bayesian networks have been traditionally used in robotics, while to share knowledge with people and the World Wide Web ontologies are considered as a basis tool. Notably projects are under way to make ontologies a way to improve robot learning and sharing of robotic skills [28]. For instance, physical relevant concepts, as affordances, can be modelled by Bayesian Networks [4]. The next step can be causal reasoning,

Going back to the tradition of action representations in AI, causality is implicit in all action representation in STRIPS-like formalisms. So causal reasoning can be the next step to make robots think more as humans

2. Causality in Bayes networks

Bayes networks (BN) were developed for prediction and abduction in artificial intelligence. In these tasks it is necessary to find a coherent interpretation of incoming observations that is consistent with both the observations and the prior information. BN are directed acyclic graphs (DAG) composed of nodes that correspond to random variables (either discrete or continuous) [12] and directed edges between nodes that indicate a direct influence of the parent to the child node [23]. The set of nodes and the set of edges define the *structure* of the BN; it represents the *conditional independence* relationships that hold in the modelled domain. Along with the structure to fully specify a BN we need to know the probability distributions $P(X_i | \text{Parents}(X_i))$ for each random variable X_i , to define the degree of influence in a quantitative way.

When considering conditional independences, represented in the BN by means of directed edges, we are not interested in the edge directionality. The graphs $X \rightarrow Y$ and $X \leftarrow Y$ both imply the same set of conditional independencies. Therefore, more than one graph can imply exactly the same set of independencies even though their structures differ in the orientation of some edges. We call such graphs *Markov equivalent* and the set of all equivalent graphs is then an *equivalence class* [12], that can be uniquely represented by a partially directed acyclic graph (PDAG).

A long tradition in philosophy has investigated the principles of causal understanding. We understand causation to be a relation between events in which the presence of

some events causes the presence of others. We assume that causation is transitive, non-reflexive, and anti-symmetric. That is:

1. if A is a cause of B and B is a cause of C , then A is also a cause of C ,
2. an event A cannot cause itself, and
3. if A is a cause of B then B is not a cause of A .

When we have two events of which one is the immediate cause of the other we say causation is direct. When there is a chain of causally connected events for which A is the immediate cause and C the immediate effect then A and C are said to be in an indirect causal relationship. In such a relationship once an event has happened it *screens off* the events that are its direct and indirect causes from its direct and indirect effects to which we refer as the *causal Markov assumption*. By means of causal relationships we can construct a *causal network* representing a causal process [23].

Under the causal Markov assumption we can interpret causal networks as causal Bayes networks (CBNs), since the causal network satisfies the Markov independencies of the corresponding Bayesian network. The main difference between CBN and BN is a stricter interpretation of the meaning of edges as direct causal relationships, with parent nodes being causes and child nodes effects. Moving from a probabilistic model to a causal one we get a model that is much more informative. While the joint distribution tells us how probable events are and how probabilities with subsequent observations change, a causal model also tells us how these probabilities would change as a result of external *interventions*. By means of interventions it is possible to test whether variable X causally influences variable Y . To do so we compute the marginal distribution of Y under the action $\text{do}(X=x)$, namely $P_x(y)$, for all values x of X and test whether that distribution is sensitive to x . This explains why causal relationships are more “stable” than probabilistic ones; they are *ontological*, describing objective physical constraints in our world, whereas probabilistic relationships are *epistemic*, reflecting what we know about the world. However, it is well understood that the independence assumption carried by BN does not necessarily imply causation.

2.1 Causal structure learning in Bayes nets

Looking at the world as consisting of a collection of causal systems, each consisting of a set of observable *causal variables*, we can translate such systems into CBN. To learn such a network we observe causal systems on a set of trials [13, 23], collecting data to perform statistical analysis driven by co variation instead of causation.

Learning causal relationships from raw data has been on philosophers' wish list since the 18th century. Human inference of causal relationships is taken to rely primarily on universal cues such as spatiotemporal contingency or reliable co variation between effects and their causes as well as on domain-specific knowledge [20]. Accordingly, most theories of causation invoke an explicit requirement that a cause precedes its effect in time. Yet temporal information alone cannot distinguish genuine causation from spurious associations.

In order to learn the structure of a CBN from raw data we need assumptions. First we assume that causal networks can provide reasonable models of the domain. Then we assume that there are no *latent* or hidden variables that affect the observable variables (this assumption does not hold in all domains). Under these assumptions we assume that one of the possible structures over the domain variables is the “true” causal network. However, from observations alone it is not possible to distinguish between causal networks that belong to the same equivalence class. In consequence different approaches have been developed: constraint-based learning, score-based learning, and Bayesian model averaging.

Constraint-based learning methods view a BN as a representation of independencies [8]. They test for conditional dependence in the data and then find an equivalence class of networks that best explain these dependencies and independencies. Constraint-based methods are quite intuitive; unfortunately they can be sensitive to failures in individual independence tests. In principle, we could use any linear test, but datasets in many domains have a high number of non-linear dependencies, making the use of this test inappropriate. The most commonly used tests are: Pearson’s chi-squared test, Fisher’s Z test and mutual information [25].

Score-based methods, also known as search-based, view a BN as specifying a statistical model and then address learning as a model selection problem. They define a hypothesis space of potential models — the set of possible network structures — and a scoring function measures how well the model fits the observed data. The task is to find the highest-scoring network structure. The space of BNs is a combinatorial space, consisting of a super-exponential number of structures $2^{O(n^2)}$. Therefore the problem of finding the highest-scoring network is NP-hard, and we resort to heuristic search, either score based or search based.

Finally, instead of attempting to learn a single structure the Bayesian model averaging methods generate an ensemble of possible structures and try to average the prediction of all possible structures [18].

2.2 Metrics used

Score-based methods consider the whole structure at once; they are therefore less sensitive to individual failures. They use one of the following metrics:

Maximum likelihood - measures the strength of the dependencies between variables and their parents. The maximum likelihood network will exhibit a conditional independence only when that independence holds exactly in the empirical distribution. Due to statistical noise, exact independence almost never occurs, and therefore the maximum likelihood network will be a fully connected one [5, 18].

Bayesian information criterion (BIC) - The score exhibits a trade-off between fit to data and model complexity: the stronger the dependence of a variable on its parents, the higher the score; the more complex the network, the lower the score [5, 24].

Akaike information criterion (AIC) – to identify an optimum model in a class of competing models, it measures the lack-of-fit of the chosen model and the increased unreliability of the chosen model due to the increased number of model parameters. The best approximating model is the one which achieves the minimum AIC in the class of the competing models [5, 24].

Bayesian metric with Dirichlet priors and equivalence (BDe) – used with discrete data, evolved from the search for a network with the largest posterior probability given priors over network structures and parameters. It is based on the concept of sets of likelihood equivalent network structures, where all members in a set of equivalent network are given the same score [5, 26].

Bayesian metric with Gaussian priors and equivalence (BGe) – BDe counterpart for continuous data [18].

Mutual information tests (MIT) - measures the degree of interaction between each variable and its parents. This measure is, however, penalized by a term related to the Pearson X^2 test of independence. This term attempts to re-scale the mutual information values in order to prevent them from systematically increasing with the number of variables [5].

2.3 Searching methods

The rules of the searching process can be either local (atomic), such that only one edge is added, removed or changes directionality, or global when the structure can change substantially [22].

The simplest is a greedy algorithm which at each step looks for the change in the structure with the best score. This procedure may get stuck in local minima/maxima, so more complex solutions, such as using a metaheuristic, have been proposed [14]. Unlike the score-based methods, which always return a BN or a set of them, the constraints-based methods return an equivalence class as a single PDAG or a set of BN. The learning process operates in two steps: firstly the algorithm looks for (in) dependencies and outputs the network skeleton¹, then it tries to orient as many edges as possible [1, 2, 3, 7, 18, 19, 23, 25]. The most commonly used algorithms are:

IC (Inductive Causation) – it starts with a graph containing all the nodes and no edges, and for each pair of variables X and Y searches for a subset of conditionally independent nodes S_{XY} . If no such subset exists it adds an undirected edge between X and Y . Once the undirected graph has been constructed it orients the edges. First it looks for all nonadjacent pairs of variables X and Y that have a common neighbour Z and checks if S_{XY} contains Z . If not, it orients the edges to get the V-structure $X \rightarrow Z \leftarrow Y$. It ends after orienting as many undirected edges as possible such that any alternative orientation would yield a new V-structure or a directed cycle [23, 25].

¹ skeleton = graph with the same nodes and edges, but all edges being undirected.

SGS (Spirtes, Glymour and Scheines) – same as IC except it starts with a fully connected graph and proceeds by removing edge by edge [23, 25].

PC (*Peter and Clark*) – it starts with a fully connected graph and continues with a systematic search for the sets S_{XY} . First it starts with S_{XY} of cardinality zero, then cardinality 1, and so on; meanwhile edges are removed from a complete graph as soon as separation is found. This refinement enjoys polynomial time complexity in graphs of finite degree because, at every stage, the search for a separating set can be limited to nodes adjacent to the two taken into consideration for independence [23, 25].

TPDA (*Three Phase Dependency Analysis*) –the algorithm has three phases. In the first phase the algorithm computes mutual information of each pair of nodes as a measure of closeness, and creates a draft as a connected graph. In the second phase, adds edges to the current graph when the pairs of nodes cannot be separated using a group of CI tests. The result of the second phase contains all the edges of the underlying dependency model given that the underlying model is *monotone DAG-faithful*². In the third phase, each edge is examined using a group of CI tests and it will be removed if the two nodes of the edge are conditionally independent. The result contains exactly the same edges as those in the underlying model when the model is monotone DAG-faithful [6]. This procedure may not be able to orient all the edges.

RAI (*Recursive autonomy Identification*) - starting from a complete undirected graph and proceeding from low to high cardinality of separation sets, the RAI algorithm performs the following operations: test of CI between nodes, removal of edges related to independences, edge direction according to orientation rules (as in IC), and graph decomposition into autonomous sub-structures. RAI is recursively applied to each sub-structure, while increasing the order of CI testing [27].

3 Causal relationships in ontologies

Over the last years ontologies have emerged as a way to provide a structured representation of knowledge in various domains. They represent also the basis for deductive reasoning [10, 11, 16, 17, 21]. The purpose of ontological representations is to capture concepts in a given domain in order to provide a shared common understanding of this domain, enabling interoperability, knowledge reuse, and reasoning through inference. They are deterministic in nature, consisting of concepts and facts about a domain and their relationships.

The challenging question is how ontologies can be used in learning BN structures. The question does not have a single answer. We argue that bringing additional knowledge could be useful to guide the structure learning process. We know that causality cannot be inferred from data alone, thus we will seek help from the additional information about variables and their relationship as present in ontology.

² The mutual information between a pair of variables is a monotonic function of the number of active paths between those variables.

Except for “is a”, which is implied from the subclass statements, relationships in ontologies are user defined and domain specific. Relationships that we are interested in are those that could imply some kind of causal relationship between the ontology terms, and which could be translated into directed edges of a BN. See an example of action ontology in Figure 1.

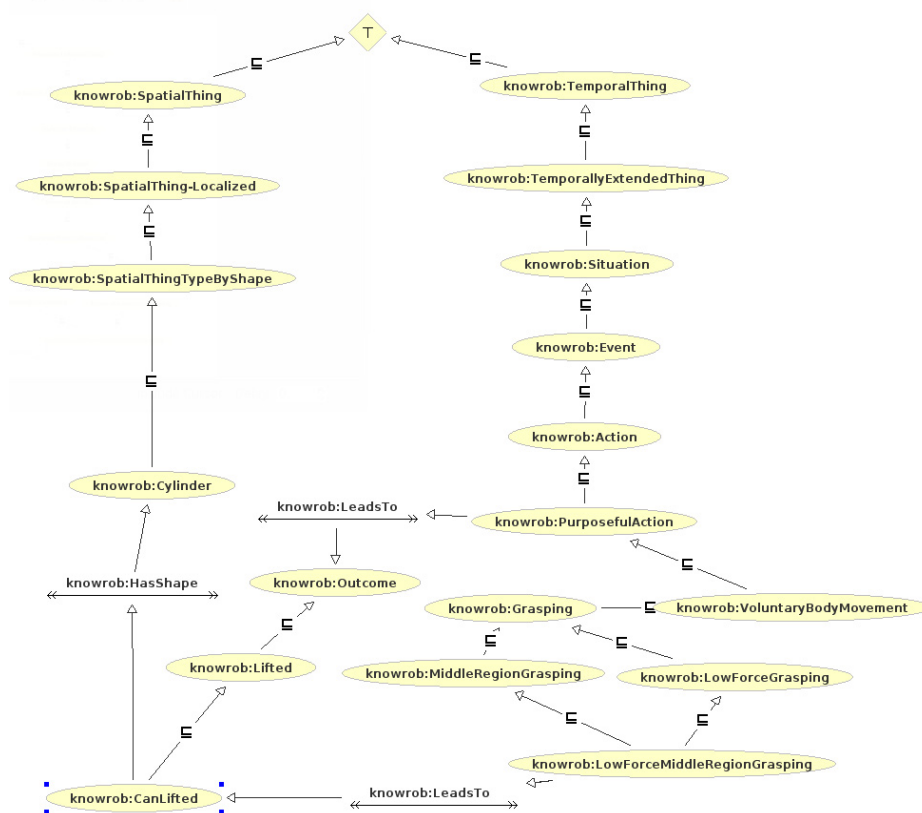


Figure 1 –An OWL ontology extending the KnowRob ontology ³.

The predefined “is_a” relationships, or *subclass relationship*, follows from the class subsumption statements $Subclass \subseteq Superclass$. This relation is reflexive, anti-symmetric and transitive; and is not a causal one. The user-defined “LeadsTo” relation is irreflexive, asymmetric and intransitive; the properties we want for a causal relationship. Therefore, when the ontology states that an action has an outcome it means that the outcome is a consequence of (it is *caused by*) the action.

³ <http://www.knowrob.org/knowrob>

In the above discussion we argued that a relationship can be regarded as causal. But in the ontology the relationships are not defined for each two terms for which the relationship holds but only for their most specific ancestors. If their connection is not explicitly stated it can be easily inferred using chain rules...

3.1 Reasoning in ontologies

In order to infer if some relationship holds between two terms, either directly or indirectly, we need to perform reasoning over the ontology by means of an inference engine usually referred to as the *reasoner*. Reasoning is needed because knowledge in an ontology might not be explicit and a reasoner is required to deduce implicit knowledge. We would need a reasoner to find out if there exists a causal relationship between terms. Unfortunately, the available reasoners do not provide built-in methods for performing such inference; therefore, we need to think of a procedure that relies on the available functionalities, mainly subclass and superclass retrieval.

Now let us introduce another class expression, the *qualified existential restriction*, indicated as \exists relation.class. It is a class denoting the set of all objects of the universe that are in relation “relation” with the objects from “class” For example, \exists parentOf.Female is the set of all objects that have female children.

Now we can specify the steps of the process for inferring the presence of a (causal) relationship between two terms that we refer to as *Cause* and *Effect*.

1. Find the superclasses of *Effect*.
2. Find all the classes that have a causal relation to the classes retrieved in step 1.
3. Find if the set of the classes retrieved in steps 2 contains *Cause*.
4. If it does there exists a causal relationship between *Cause* and *Effect*.

In steps 1 and 2 we find all the classes that are (causally) related to the *Effect* class or any of its ancestors since if a class is related to an ancestor the relation also influences that class which is just a more specific instance of the ancestor. If the *Cause* class is among the classes that are causally related to the *Effect* or some of its ancestors we can infer that there exists a causal relationship between the two.

It might be tempting to consider taking into account subsets of the *Effect* class or subsets/supersets of the *Cause* class, but such inferences are not sound. In case of subsets of the *Effect* class, a cause that causally influences a subset of the descendants this relation does not tell us anything about its relation to the *Effect* class. Each descendant has additional information which might be the reason for the presence of the causal relation. Even if all the descendants would be causally related to the *Cause* it would not be enough to justify the existence of the relation for the *Effect* class because there might be other subclasses not yet present in the ontology. The same reasoning can be used when considering subclasses of the *Cause* class.

On the other hand, if we consider including the superclasses of the *Cause* class when looking for relations to the *Effect* class, two problems arise. The first is that a class is causally related to its own ancestor; such a relation is a tautological statement. A general rule would be to disregard relationships that are “too high” in the hierarchy. But it is not possible neither to state how high is too high nor to know during reasoning where the terms reside since the ontologies are represented as DAG without ordering of terms. The other problem is more severe. Let us consider four classes X, Y, W, Z, where Y and Z are subclasses of X and W respectively; moreover X and Y have relation “relation” on W and Z. If we were to infer a causal relationship between Y and any sibling of Z by looking at both the causal relations of Y and X, considering the superclasses of Y we would find it because of the relation between X and W. But such a conclusion is wrong because we know that Y causally influences only Z while the conclusion drawn would be that it is causally related not only to Z’s ancestors but also to all its siblings.

4 Combined approaches

Combining knowledge from ontologies and Bayes networks is done in [15] using a manually constructed ontology which gets translated into a BN. The ontology is used to make it easier for experts to model the domain knowledge. [10] continued to devise an algorithm in the telecommunications domain. Later, an extension of the OWL language was proposed by [11] in order to incorporate probabilistic knowledge to allow probabilistic reasoning over the constructed BN.

In the medical domain [17] proposed a semi-automatic algorithm which extracted nodes from an ontology and let the expert draw the causal relationships between them. For the same domain [29] proposed another way for incorporating uncertainty in ontologies; their approach had the both the goal to add uncertainties in an ontology and to allow for the probability distribution to be updated by adding new data.

[16] dealt with the translation of the ontology into Objective Oriented BN (OOBN). The advantage of OOBN is that nodes can be assigned properties and be represented in a hierarchy making them more similar to ontologies than regular BNs. It is straightforward to see that by means of the hierarchy it is possible to translate the ontologies’ “is_a” relationship into the OOBN.

[21] used the knowledge and functionalities present in both models to transfer knowledge both directions. First they use the knowledge in the ontology to constrain the possible BN structures and guide the learning process. Then the BN structure is used to update the ontology structure by adding newly found causal relationships. They impose the following constraints: each causal graph node must be modelled by a corresponding concept in the domain ontology, and the causal relations have to be defined between all elements of the ontology for which it holds. These constraints do not allow using existing ontologies but require designing the needed ones.

5 Conclusions and future research

Knowledge represented in the ontology can be used at different stages:

- before using the BN structure construction algorithm in order to find connections to lower the number of possible structures,
- after using the algorithm but before assigning edge orientation,
- after producing a PDAG structure, to infer the orientation of undirected edges,
- for checking the correctness of the orientation of the directed edges.

The complexity, both space and time, seems to be quite a challenge when dealing with experimental data, that can require hundred or thousand of nodes in the BN. Just fully connecting a graph proves to be a challenge. We consider here just the complexity of the steps in the algorithm, ignoring the complexity of the reasoners [9].

We have done a complexity analysis of the PC algorithm, on a graph with n nodes and k maximal degree of any node. In the worst case in each iteration no edge will be removed. Therefore in the i -th iteration we have to check separation sets of cardinality i and for each of the $n \cdot (n-1)$ pairs of variables there are $\binom{n-2}{i}$ candidate separation sets, which for a total of

$$n \cdot (n-1) \cdot \sum_{i=0}^k \binom{n-2}{i} \leq n \cdot (n-1)^{k+1} \quad (1)$$

independency tests (such worst case is highly unlikely). The complexity is polynomial, namely $O(n^{k+1})$.

For each independency test, using the Fisher's Z test in the recursive formula, the number of arithmetic operations is exponential; however the complexity can be reduced caching the results, at the cost of increasing space complexity. In experimental trials we made on various ontologies and data, we have been able to deal with datasets of 6000 variables in less than 2 minutes. For comparison most of the publicly available constraint-based structure learning algorithms struggle with the subset of the same dataset having around 100 variables. The ideas we tried will be the basis of a new method under development.

Also the complexity of the inference by means of the ontology can be formalized. Let m be the maximum number of ontology terms mapped to any variable. Then for u undirected edges we have to make

$$2 \cdot u \cdot m^2 \cdot 3 \quad (2)$$

calls to the reasoner. "2u" is because we need to check both orientations, the third factor is the number of pairs to be sought for a causal relationship, and the last is the number of calls to the reasoner for retrieving superclasses and related terms. Even though the number of calls to the reasoner is quite low, the reasoning process is time consuming for big ontologies. It may take hours.

We can conclude that integrating ontologies and BN does provide a valid approach for the problem of CBN structure learning. The knowledge transferred from the on-

tology can be regarded as reflecting the true world, which provides the needed guarantee for interpreting the resulting BN as a causal one.

We have also found critical issues. The time required for inferring relations in the ontology is too high for practical applications. Open directions for future work are so optimizing the step for inferring the edge orientation, finding the base structure learning algorithm with best performance, or considering different types of BN. Also, the mapping of the ontology terms to variables might have to be performed manually for most of the domains. A next step will be using publicly available domain knowledge in form of ontologies to assist the learning process, as in the indicated RoboEarth. A matching dataset is also available.

Acknowledgements

We thank for partial support the EU Life+ CALEIDOS and PROSIL projects.

References

- [1] Acid, S. et al, J. (2004). A comparison of learning algorithms for Bayesian networks: a case study based on data from an emergency medical service. *Artificial Intelligence in Medicine*, 30(3), p. 215-232.
- [2] Al-Akwaa, F. M., & Alkhwilani, M.M. (2012). Comparison of the Bayesian Network Structure Learning Algorithms. *Intl J of Advanced research in computer science and software engineering*, 2(3), p. 404-408.
- [3] Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., & Sherlock, G. (2000). Gene Ontology: tool for the unification of biology. *Nature genetics*, 25(1), p. 25-29.
- [4] Barck-Holst, C., Ralph, M., Holmar, F., Kragic, D. (2009). Learning grasping affordance using probabilistic and ontological approaches. *Proc IEEE ICAR*
- [5] Carvalho, A. M. (2009). Scoring functions for learning Bayesian networks. *Inesc-id Tech. Rep.*
- [6] Cheng, J., Bell, D., & Liu, W. (1998). Learning Bayesian networks from data: An efficient approach based on information theory. <http://www.cs.ualberta.ca/~jcheng/bnpc.htm>.
- [7] Cooper, G. F., & Yoo, C. (1999). Causal discovery from a mixture of experimental and observational data. In *Proceedings of the Fifteenth conference on Uncertainty in artificial intelligence*, Morgan Kaufmann Publishers Inc, p. 116-125.
- [8] De Campos, C. P., Zeng, Z., & Ji, Q. (2009). Structure learning of Bayesian networks using constraints. In *Proceedings of the 26th Annual International Conference on Machine Learning*, p. 113-120. ACM.
- [9] Dentler, K., Cornet, R., ten Teije, A., & de Keizer, N. (2011). Comparison of reasoners for large ontologies in the OWL 2 EL profile. *Semantic Web*, 2(2), p. 71-87.

- [10] Devitt, A., Danev, B., & Matusikova, K. (2006). Ontology-driven Automatic Construction of Bayesian Networks for Telecommunication Network Management. In 2nd Int. Workshop: Formal Ontologies Meet Industry (FOMI 2006).
- [11] Ding, Z., Peng, Y., & Pan, R. (2006). BayesOWL: Uncertainty modeling in semantic web ontologies. In *Soft Computing in Ontologies and Semantic Web*, p. 3-29. Springer Berlin.
- [12] Friedman, N., Linial, M., Nachman, I., & Pe'er, D. (2000). Using Bayesian networks to analyze expression data. *Journal of computational biology*, 7(3-4), p. 601-620.
- [13] Goodman, N.D., Ullman, T.D., & Tenenbaum, J.B. (2011). Learning a theory of causality. *Psychological review*, 118(1), 110.
- [14] Harwood, S., & Scheines, R. (2002). Genetic algorithm search over causal models.
- [15] Helsper, E.M., & van der Gaag, L. C. (2002). Building Bayesian networks through ontologies. In *ECAI*, Vol. 2002, p. 15th.
- [16] Ishak, M.B., Leray, P., & Amor, N.B. (2011). Ontology-based generation of object oriented Bayesian networks. In *Proceedings of the 8th Bayesian Modelling Applications Workshop*, p. 9-17.
- [17] Jeon, B.J., & Ko, I.Y. (2007). Ontology-based semi-automatic construction of bayesian network models for diagnosing diseases in e-health applications. In *Frontiers in the Convergence of Bioscience and Information Technologies. FBIT 2007*, p. 595-602. IEEE.
- [18] Koller, D., & Friedman, N. (2009). *Probabilistic graphical models: principles and techniques*. The MIT Press.
- [19] Korb, K.B., & Nicholson, A. E. (2003). *Bayesian artificial intelligence*. CRC press. 59
- [20] Lucas, C.G., & Griffiths, T.L. (2010). Learning the form of causal relationships using hierarchical Bayesian models. *Cognitive Science*, 34(1), p. 113-147.
- [21] Messaoud, M.B., Leray, P., & Amor, N.B. (2011). SemCaDo: a serendipitous strategy for learning causal bayesian networks using ontologies. In *Symbolic and Quantitative Approaches to Reasoning with Uncertainty*, p. 182-193. Springer Berlin Heidelberg.
- [22] Pe'er, D. (2003). *From Gene Expression to Molecular Pathways* (Doctoral dissertation, Hebrew University of Jerusalem).
- [23] Pearl, J. (2000). *Causality: models, reasoning and inference*. Cambridge: MIT press.
- [24] Sakamoto, Y., Ishiguro, M., & Kitagawa, G. (1986). *Akaike information criterion statistics*. Dordrecht, The Netherlands: D. Reidel.
- [25] Spirtes, P., Glymour, C., & Scheines, R. (2000). *Causation, prediction, and search* (Vol. 81). The MIT Press
- [26] Yang, S., & Chang, K.C. (2002). Comparison of score metrics for Bayesian network learning. *IEEE Trans SMC, Part A*, 32(3), p. 419-428.
- [27] Yehezkel, R., & Lerner, B. (2009). Bayesian network structure learning by recursive autonomy identification. *The Journal of Machine Learning Research*, 10, p. 1527-1570.
- [28] Waibel, M. et al, (2011) A World Wide Web for robots: RoboEarth. *IEEE Robotics&Automation Magazine*, June 2011, p. 69-82.
- [29] Zheng, H.T., Kang, B.Y., & Kim, H.G. (2008). An ontology-based Bayesian network approach for representing uncertainty in clinical practice guidelines. In *Uncertainty Reasoning for the Semantic Web I*, p. 161-173. Springer Berlin Heidelberg.