Probabilistic Graphical Models: On Reasoning, Learning, and Revision (Extended Abstract)

Rudolf Kruse
Faculty of Computer Science, Otto von Guericke University, Magdeburg, Germany

Keywords: Bayesian Network, Markov Network, Item Set Planning.

Abstract: Probabilistic Graphical Models are of high relevance for complex industrial applications. The Bayesian network and the Markov network approach are the most prominent representatives and an important tool to structure uncertain knowledge about high dimensional domains. This extended abstract serves to highlight that the decomposition of the underlying high dimensional spaces turns out to be useful to make reasoning, learning and revision in such domains feasible. The methods are explained by using a real-world industrial application from automotive industry.

1 INTRODUCTION

In the automotive industry, customers prefer to opt for individual vehicle specifications. For this reason, some manufacturers prefer a marketing policy that offers maximum freedom in choosing individual vehicle specifications. This means that a customer can select from a large number of options, according to his personal preferences. One can choose the body variant, engine, circuit, door layout, seat cover, radio and navigation system, and this only reflect a small part of the entire product family. In the case of a typical popular car, there are about 200 such variables, each typically having 4-8 values and a total range of cardinalities from 2 to 150. Of course, not all possible instantiations of these so-called item variables result in valid vehicle configurations, as technical rules, manufacturing restrictions and selling requirements induce a common rule system that restricts the acceptable ways of item combination. However, with more than 10,000 technical rules in a class and many more rules supplied by the sales programs for the specific needs of different countries, there remains a huge number of correct vehicle specifications.

2 ITEM SET PLANNING

The main goal of item planning is the development and implementation of a software system that supports item planning, parts requirements calculation and capacity management with the aim of short and medium-term range forecasts for future vehicle production. In order to achieve high quality planning results, all relevant sources of information must be considered, namely rules for the right combination of items to form complete vehicle specifications, samples of produced vehicles reflecting customer preferences, market forecasts leading to changed specifications, item sets for planning intervals, capacity constraints and production programs, which determine the number of planned vehicles.

From a logistical point of view, the most essential result of the item planning process is the evaluation of the rates of all item combinations that are known to be relevant for the parts requirements calculation, always related to a specific vehicle class in a specific planning interval. The importance of these item combinations derives from the fact that a vehicle can be interpreted as a large set of fitting locations, each characterized by a set of alternative fitting parts for that location. Within the framework of a typical passenger car class, a total of around 70,000 different article combinations are required as installation conditions for all of the installation locations. The task of predicting the total demand for a specific part in relation to a future planning interval is to add up the demands across all of its installation locations. The need for any installation location is obtained by multiplying the rate of the combination of items that represents its installation condition by the installed

Kruse, R.
Probabilistic Graphical Models: On Reasoning, Learning, and Revision (Extended Abstract).
DOI: 10.5220/0011598200003335
In *Proceedings of the 14th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2022) - Volume 3: KMIS*, pages 9-10
ISBN: 978-989-758-614-9; ISSN: 2184-3228

one in quantity and the total number of vehicles intended to be produced in the respective planning interval.

3 PROBABILISTIC NETWORKS

The domain and expert knowledge in this application about installation rates can be formally represented by a probability distribution over the set of relevant item families or attributes. Conditional independences are used to decompose this distribution into lower dimensional distributions. Since it is possible to connect the concepts of conditional independence with the separation concept in graphs, graphical models turn out to be extremely helpful for the problem of item set planning. Two well-know models are Bayesian networks and Markov networks (Borgelt 2009).

A Bayesian networks is a directed acyclic graph (DAG's), representing a set of random variables and the dependencies between these random variables. A Markov network is an undirected conditional independence graph G = (V,E) of a probability distribution together with a family of conditional probabilities of the factorization induced by the graph.

Probabilistic graphical models allow for an efficient knowledge representation as well as an integration of new evidence via conditioning. The basic idea is to distribute (propagate) the evidence through the network to reach all attributes.

Probabilistic graphical models can also be learned from given data. Classical statistical techniques for parameters learning as well as other methods for learning the network structure are useful (Drton 2017). Approaches for learning graphical models typically fall into one of two categories: score-based approaches and constraint-based approaches. Score-based approaches consist of two elements: a score function to evaluate how well graph candidates fit the database, and some search heuristic (possibly guided by the scores) to traverse the set of graphs. The goal of constraint-based approaches is to use conditional (in)dependence tests to construct a graphical model which is a perfect map (or independence map) of the data-generating distribution. Several efficient and user-friendly commercial tools such as Hugin (HuginExpert 2022) are available for this task.

In practice, however, there is also a need to revise the probability distribution represented by a graphical model in such a way that it satisfies the given framework conditions, for example given marginal distributions. Pure evidence propagation methods such as join tree propagation and bucket elimination are unsuitable for this task. We present a knowledge-based probabilistic formalization and solution of the fundamental revision problem for Markov networks, constrained to a set of unconstrained single-variable boundary conditions (Gebhardt 2005). This probabilistic approach avoids all concepts offered by calculi with deviating semantic foundations, for example to minimize probabilistic difference measures that could be inherited from information theory. From multivariate statistics, iterative proportional fitting gives a convenient algorithm to fit the marginal distributions of a given joint distribution to desired values.

4 CONCLUSIONS

The probabilistic graphical network approach has proven to be very successful for assistance systems., Thousands of Markov networks for different planning scenarios and different model groups are in use every day for item set planning.

The methodology used for item set planning can be easily transferred to other areas. In the monograph (Kruse 2022) a tutorial introduction to this type of knowledge representation, updating, revision and learning is given.

In the item planning project we have mainly benefited from the decomposition aspect of probabilistic graphical networks. We are convinced that the concept of causality (Pearl 2018) will play a central role in many future applications.

REFERENCES

Borgelt, C., Steinbrecher, M., Kruse, R. (2009) *Graphical Models, Representations for Learning, Reasoning and Data Mining*, Wiley, Chichester, 2nd edition

Drton, M., Maathuis M. (2017) *Structure Learning in Graphical Modeling*, Annual Review of Statistics and its Application Vol.4, 365-393

Gebhardt, J., Kruse, R. (2005). *Knowledge-Based Operations for Graphical Models in Planning*. In ECSQARU 2005, Springer LNAI 3571, pp 3-14.

HuginExpert (2022), *Bayesian network software*, http://www.hugin.com

Kruse, R., et al (2022). *Computational Intelligence, A Methodological Introduction*, Springer, London, 3rd edition.

Pearl, J., Mackenzie, D. (2018), *The book of why: the new science of cause and effect*, Basic Books, New York.