Application of Graphical Models in the Automotive Industry

Matthias Steinbrecher, Frank Rügheimer, and Rudolf Kruse

Department of Knowledge Processing and Language Engineering
Otto-von-Guericke University of Magdeburg
Universitätsplatz 2, 39106 Magdeburg, Germany
{msteinbr,ruegheim,kruse}@iws.cs.uni-magdeburg.de

1 Introduction

The production pipeline of present day's automobile manufacturers consists of a highly heterogeneous and intricate assembly workflow that is driven by a considerable degree of interdependencies between the participating instances as there are suppliers, manufacturing engineers, marketing analysts and development researchers. Therefore, it is of paramount importance to enable all production experts to quickly respond to potential on-time delivery failures, ordering peaks or other disturbances that may interfere with the ideal assembly process. Moreover, the fast moving evolvement of new vehicle models require well-designed investigations regarding the collection and analysis of vehicle maintenance data. It is crucial to track down complicated interactions between car components or external failure causes in the shortest time possible to meet customer-requested quality claims.

To summarize these requirements, let us turn to an example which reveals some of the dependencies mentioned in this chapter. As we will later see, a normal car model can be described by hundreds of variables each of which representing a feature or technical property. Since only a small number of combinations (compared to all possible ones) will represent a valid car configuration, we will present a means of reducing the model space by imposing restrictions. These restrictions enter the mathematical treatment in the form of dependencies since a restriction may cancel out some options, thus rendering two attributes (more) dependent. This early step produces qualitative dependencies like "engine type and transmission type are dependent". To quantify these dependencies some uncertainty calculus is necessary to establish the dependence strengths. In our cases probability theory is used to augment the model, e. g. "whenever engine type 1 is ordered, the probability is 56% of having transmission type 2 ordered as well". There is a multitude of sources to estimate or extract this information from. When ordering peaks occur like an increased demand of convertibles during the Spring, or some supply shortages arise due to a strike in the transport industry, the model is used to predict vehicle configurations that may run into delivery delays in order to forestall such a scenario by e. g. acquiring alternative supply

D. Prokhorov (Ed.): Comput. Intel. in Automotive Applications, SCI 132, pp. 103–115.
springerlink.com

chains or temporarily shifting production load. Another part of the model may contain similar information for the aftercare, e. g. “whenever a warranty claim contained battery type 3, there is a 30% chance of having radio type 1 in the car”. In this case dependencies are contained in the quality assessment data and are not known beforehand but are extracted to reveal possible hidden design flaws.

These examples — both in the realm of planning and subsequent maintenance measures — call for treatment methods that exploit the dependence structures embedded inside the application domains. Furthermore, these methods need to be equipped with dedicated updating, revision and refinement techniques in order to cope with the above-mentioned possible supply and demand irregularities. Since every production and planning stage involves highly specialized domain experts, it is necessary to offer intuitive system interfaces that are less prone to inter-domain misunderstandings.

The next section will sketch the underlying theoretical frameworks, after which we will present and discuss successfully applied planning and analysis methods that have been rolled out to production sites of two large automobile manufacturers. Section 3 deals with the handling of production planning at Volkswagen. The underlying data is sketched in section 3.1 which also covers the description of the model structure. Section 3.2 introduces three operations that serve the purpose of modifying the model and answering user queries. Finally, section 3.3 concludes the application report at Volkswagen. The Daimler AG application is introduced in section 4 which itself is divided to explain the data and model structure (section 4.1, to propose the visualization technique for data exploration (section 4.2 and finally to present empirical evidence of the usability in section 4.3.

2 Graphical Models

As motivated in the introduction, there are a lot of dependencies and independencies that have to be taken into account when to approach the task of planning and reasoning in complex domains. Graphical models are appealing since they provide a framework of modeling independencies between attributes and influence variables. The term “graphical model” is derived from an analogy between stochastic independence and node separation in graphs. Let $V = \{A_1, \ldots, A_n\}$ be a set of random variables. If the underlying probability distribution $P(V)$ satisfies some criteria (see e. g. [5; 13]), then it is possible to capture some of the independence relations between the variables in V using a graph $G = (V, E)$, where E denotes the set of edges. The underlying idea is to decompose the joint distribution $P(V)$ into lower-dimensional marginal oder conditional distributions from which the original distribution can be reconstructed with no or at least as few errors as possible [12; 14]. The named independence relations allow for a simplification of these factor distributions. We claim, that every independence that can be read from a graph also holds in the corresponding joint distribution. The graph is then called an independence map (see e. g. [4]).

2.1 Bayesian Networks

If we are dealing with an acyclic and directed graph structure G, the network is referred to as a Bayesian network. The decomposition described by the graph consists of a set of conditional distributions assigned to each node given its direct predecessors (parents). For each value of the attribute domains (dom), the original distribution can be reconstructed as follows:

$$\forall a_1 \in \mathrm{dom}(A_1) : \cdots \forall a_n \in \mathrm{dom}(A_n) :$$
$$P(A_1 = a_1, \ldots, A_n = a_n) = \prod_{A_i \in V} P\Big(A_i = a_i \mid \bigwedge_{(A_j, A_i) \in E} A_j = a_j\Big)$$

2.2 Markov Networks

Markov networks rely on undirected graphs where the lower-dimensional factor distributions are defined as marginal distributions on the cliques $C = \{C_1, \ldots, C_m\}$ of the graph G. The original joint distribution $P(V)$ can then be recombined as follows:

$$\forall a_1 \in \mathrm{dom}(A_1) : \cdots \forall a_n \in \mathrm{dom}(A_n) :$$
$$P(A_1 = a_1, \ldots, A_n = a_n) = \prod_{C_i \in C} \phi_{C_i}\Big(\bigwedge_{A_j \in C_i} A_j = a_j\Big)$$

For a detailed discussion on how to choose the functions ϕ_{C_i}, see e. g. [4].

3 Production Planning at Volkswagen Group

One goal of the project described here was to develop a system which plans parts demand for the production sites of the Volkswagen Group [8]. The market strategy is strongly customer-focused—based on adaptable designs and special emphasis on variety. Consequently, when ordering an automobile, the customer is offered several options of how each feature should be realized. The result is a very large number of possible car variants. Since the particular parts required for an automobile depend on the variant of the car, the overall parts demand can not be successfully estimated from total production numbers alone. The modeling of domains with such a large number of possible states is very complex. Therefore, decomposition techniques were applied and augmented by a set of operations on these subspaces that allow for a flexible parts demand planning and also provide a useful tool to simulate capacity usage in projected market development scenarios.

3.1 Data Description and Model Induction

The first step towards a feasible planning system consists of the identification of valid vehicle variants. If cars contain components that only work when combined with specific versions of other parts, changes in the predicted rates for one

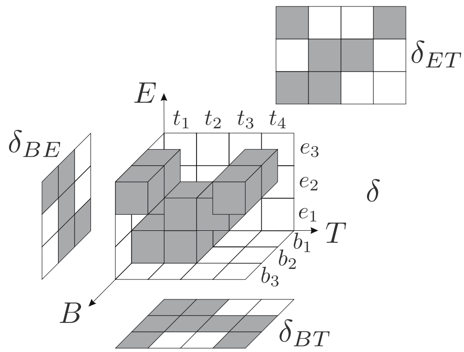

Fig. 1. The 3-dimensional space $\mathrm{dom}(E) \times \mathrm{dom}(T) \times \mathrm{dom}(B)$ is thinned out by a rule set, sparing only the depicted value combinations. Further, one can reconstruct the 3-dimensional relation δ from the two projections δ_{ET} and δ_{BT}.

component may have an influence on the demand for other components. Such relations should be reflected in the design of the planning system.

A typical model of car is described by approximately 200 attributes, each consisting of at least 2, but up to 50 values. This scaffolds a space of possible car variants with a cardinality of over 10^{60}. Of course, not every combination corresponds to a valid specification. To ensure only valid combinations, restrictions are introduced in form of a rule system. Let us assume we are dealing with three variables E, T and B representing engine type, transmission type and brake type with the following respective domains:

$$\mathrm{dom}(E) = \{e_1, e_2, e_3\}, \quad \mathrm{dom}(T) = \{t_1, t_2, t_3, t_4\}, \quad \mathrm{dom}(B) = \{b_1, b_2, b_3\}$$

A set of rules could for example contain statements like

$$\text{If } T = t_3 \text{ then } B = b_2$$

or

$$\text{If } E = e_2 \text{ then } T \in \{t_2, t_3\}$$

A comprehensive set of rules cancels out invalid combinations and may result in our example in a relation as depicted in figure 1.

It was decided to employ a probabilistic Markov network to represent the distribution of the value combinations. Probabilities are thus interpreted in terms of estimated relative frequencies. Therefore, an appropriate decomposition has to be found. Starting from a given rule base R and a production history to estimate relative frequencies from, the graphical component is generated as follows: We start out with an undirected graph $G = (V, E)$ where two variables F_i and F_j are connected by an edge $(F_i, F_j) \in E$ if there is a rule in R that contains both variables. To make reasoning efficient, it is desirable that the graph has hypertree structure. This includes the triangulation of G, as well as the identification of its cliques. This process is depicted in figure 2. To complete the model, for every clique a joint distribution for the variables of that clique has to be estimated from the production history.

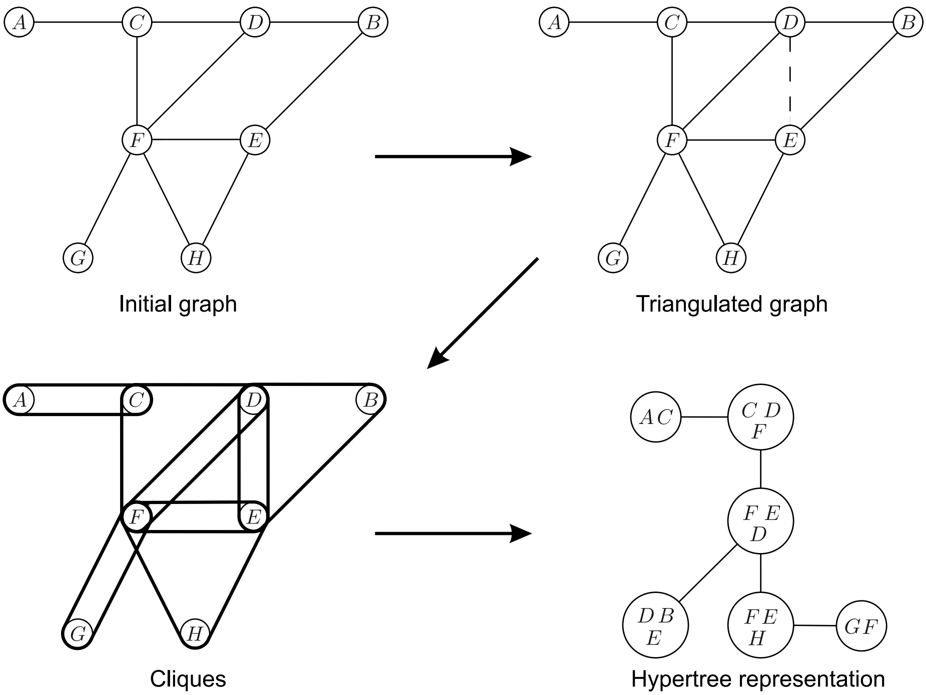

Fig. 2. Transformation of the model into hypertree structure. The initial graph is derived from the rule base. For reasoning, the hypertree cliques have to have the running intersection property which basically allows for a composition of the original distribution from the clique distributions. See [5] for details. This property can be asserted by requiring the initial graph to be triangulated.

3.2 Operations on the Model

A planning model that was generated using the above method, usually does not reflect the whole potential of available knowledge. For instance, experts are often aware of differences between the production history and the particular planning interval the model is meant to be used with. Thus, a mechanism to modify the represented distribution is required. Planning operators have been developed [10] to efficiently handle this kind of problem, so modification of the distribution and restoration of a consistent state can be supported.

Updating

Consider a situation where previously forbidden item combinations become valid. This can result for example from changes in the rule base. The relation in figure 1 does not allow engine type 2 to be combined with transmission type 1 because $(e_2, t_1) \notin E \times T$. If this option becomes valid probability mass, it has to be transferred to the respective distribution. Another scenario would be the advent of a new engine type, i. e. a change in the domain itself. Then, a multitude of new probabilities have to be assessed. Another related problem arises when subsets of cliques are altered while the information of the remaining network is retained. Both scenarios are addressed with the updating operation.

This operation marks these combinations as valid by assigning a positive near-zero probability to their respective marginals. Due to this small value, the quality of the estimation is not affected by this alteration. Now instead of using the same initialization for all new combinations, the proportion of the values is chosen in accordance to an existing combination, i. e. the probabilistic interaction structure is copied from reference item combinations.

Since updating only provides the qualitative aspect of the dependence structure, it is usually followed by the subsequent application of the revision operation, which is used to reassign probability mass to the new item combinations.

Revision

The revision operation, while preserving the network structure, serves to modify quantitative knowledge in such a way that the revised distribution becomes consistent with the new specialized information. There is usually no unique solution to this task. However, it is desirable to retain as much of the original distribution as possible so that the principle of minimal change [7] should be applied. Given that, a successful revision holds a unique result [9]. As an example for a specification, experts might predict a rise of the popularity of a recently introduced navigation system and set the relative frequency of this respective item from 20% to 30%.

Focusing

While revision and updating are essential operations for building and maintaining a distribution model, it is much more common activity to apply the model for the exploration of the represented knowledge and its implications with respect to user decisions. Typically users would want to concentrate on those aspects of the represented knowledge that fall into their domain of expertise. Moreover, when predicting parts demand from the model, one is only interested in estimated rates for particular item combinations. Such activities require a focusing operation. It is implemented by performing evidence-driven conditioning on a subset of variables and distributing the information through the network. Apart from predicting parts demand, focusing is often employed for market analyses and simulation. By analyzing which items are frequently combined by customers, experts can tailor special offers for different customer groups. To support planning of buffer capacities, it is necessary to deal with the eventuality of temporal logistic restrictions. Such events would entail changes in short-term production planning so that consumption of the concerned parts is reduced.

3.3 Application

The development of the planning system explained was initiated in 2001 by the Volkswagen Group. System design and most of the implementation is currently done by Corporate IT. The mathematical modeling, theoretical problem solving, and the development of efficient algorithms have been entirely provided by Intelligent Systems Consulting (ISC) Gebhardt. Since 2004 the system is being rolled out to all trademarks of the Volkswagen Group. With this software, the increasing planning quality, based on the many innovative features and the appropriateness of the chosen model of knowledge representation, as well as a considerable reduction of calculation time turned out to be essential prerequisites for advanced item planning and calculation of parts demand in the presence of structured products with an extreme number of possible variants.

4 Vehicle Data Mining at Daimler AG

While the previous section presented techniques that were applied ahead-of-time, i. e., prior and during the manufacturing process, we will now turn to the area of assessing the quality of cars after they left the assembly plant. For every car that is sold, a variety of data is collected and stored in corporate-wide databases. After every repair or checkup the respective records are updated to reflect the technical treatment. The analysis scenario discussed here is the interest of the automobile manufacturer to investigate car failures by identifying common properties that are exposed by specific subsets of cars that have a higher failure rate.

4.1 Data Description and Model Induction

As stated above, the source of information consists of a database that contains for every car a set of up to 300 attributes that describe the configuration of every car that has been sold.

The decision was made to use Bayesian networks to model the dependence structure between these attributes to be able to reveal possible interactions of vehicle components that cause higher failure rates. The induction of a Bayesian network consists of identifying a good candidate graph that encodes the independencies in the database. The goodness of fit is estimated by an evaluation measure. Therefore, usual learning algorithms consist of two parts: a search method and the mentioned evaluation measure which may guide the search. Examples for both parts are studied in [6; 11; 2].

Given a network structure, an expert user will gain first insights into the corresponding application domain. In figure 3 one could identify the road surface conditions to have a major (stochastic) impact on the failure rate and type. Of course, arriving at such a model is not always a straightforward task since the available database may lack some entries requiring the treatment of missing values. In this case possibilistic networks [3] may be used. However, with full information it might still be problematic to extract significant statistics since there may be value combinations that occur too scarcely. Even if we are in the favorable position to have sufficient amounts of complete data, the bare network structure does not reveal information about which *which* road conditions have *what kind* of impact on *which* type of failure. Fortunately, this information can be

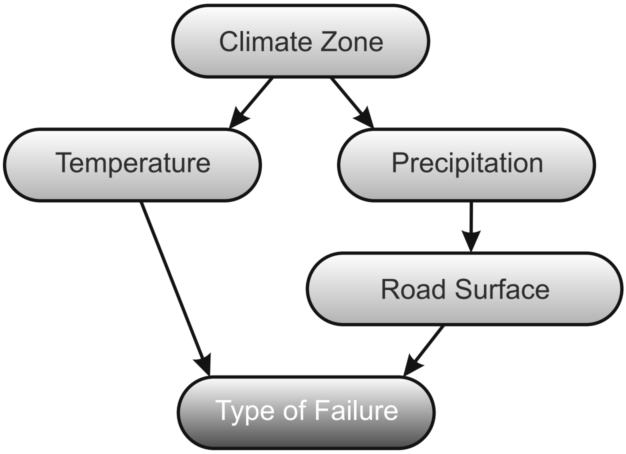

Fig. 3. An example of a Bayesian network illustrating qualitative linkage of components

retrieved easily in form of conditional probabilities from the underlying dataset, given the network structure. This becomes clear if the sentence above is restated: Given a specific road surface condition, what is the failure probability of a randomly picked vehicle?

4.2 Model Visualization

Every attribute with its direct parent attributes encodes a set of conditional probability distributions. For example, given a database D, the sub-network consisting of Failure, RoadSurface and Temperature in figure 3 defines the following set of distributions:

$$P_D(\mathsf{Failure} \mid \mathsf{Temperature}, \mathsf{RoadSurface})$$

For every distinct combination of values of the attributes RoadSurface and Temperature, the conditional probability of the attribute Failure is estimated (counted) from the database D. We will argue in the next section, that it is this information that enables the user to gain better insight into the data under consideration [15].

Given an attribute of interest (in most cases the class variable like Failure in the example setting) and its conditioning parents, every probability statement like

$$P(\mathsf{Failure} = \mathsf{Suspension} \mid \mathsf{RoadSurface} = \mathsf{rough}, \mathsf{Temperature} = \mathsf{low}) \quad = \quad p^*$$

can be considered an association rule:[1]

> If RoadSurface = rough $\wedge$ Temperature = low, then there will be a suspension failure in $100 \cdot p^*\%$ of all cases.

[1] See [1] for details.

The value p^* is then the confidence of the corresponding association rule. Of course, all known evaluation measures can be applied to assess the rules. With the help of such measures one can create an intuitive visual representation according to the following steps:

- For every probabilistic entry (i. e., for every rule) of the considered conditional distribution $P(C \mid A_1, \ldots, A_m)$ a circle is generated to be placed inside a two-dimensional chart.
- The graylevel (or color in the real application) of the circle corresponds the the value of attribute C.
- The circle's area corresponds to the value of some rule evaluation measure selected before displaying. For the remainder of this chapter, we choose this measure to be the support, i. e., the relative number of vehicles (or whatever instances) specified by the values of C and $A_1, \ldots, A_m$. Therefore, the area of the circle corresponds to the number of vehicles.
- In the last step these circles are positioned. Again, the value of the x- and y-coordinate are determined by two evaluation measures selected in advance. We suggest these measures to be recall[2] and lift.[3] Circles above the darker horizontal line in every chart mark subsets with a lift greater than 1 and thus indicate that the failure probability is larger given the instantiation of $A_1, \ldots, A_n$ in contrast to the marginal failure probability $P(C = c)$.

With these prerequisites we can recommend to the user the following heuristic in order to identify suspicious subsets:

> *Sets of instances in the upper right hand side of the chart may be good candidates for a closer inspection.*

The greater the y-coordinate (i. e., the lift value) of a rule, the stronger is the impact of the conditioning attributes' values on the class variable. Larger x-coordinates correspond to higher recall values.

4.3 Application

This section illustrates the proposed visualization method by means of three real-world datasets that were analyzed during a cooperate research project with a automobile manufacturer. We used the K2 algorithm[4] [6] to induce the network structure and visualized the class variable according to the given procedure.

---

[2] The recall is definded as $P(A_1 = a_1, \ldots, A_k = a_k \mid C = c)$.

[3] The lift of a rule indicates the ratio between confidence and the marginal failure rate: $\frac{P(C=c \mid A_1=a_1,\ldots,A_k=a_k)}{P(C=c)}$.

[4] It is a greedy approach that starts with a single attribute (here: the class attribute) and tries to add parent attributes greedily. If no addition of an attribute yields a better result, the process continues at the just inserted parent attributes. The quality of a given network is measured with the K2 metric (a Bayesian model averaging metric).

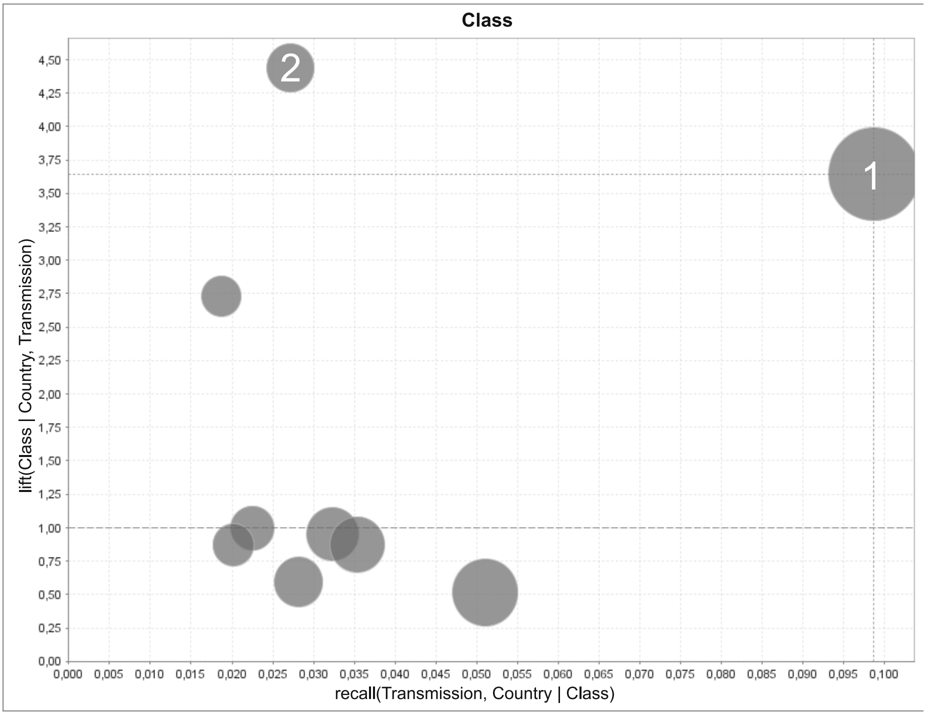

Fig. 4. The subset marked 1 corresponds to approx. 1000 vehicles whose attributes values of Country and Transmission yielded a causal relationship with the class variable. Unfortunately, there was not found a causal description of subset 2. The cluster of circles below the lift-1 line corresponds to sets of cars that fail less often, if their instantiantions of attributes become known.

Example 1

Figure 4 shows the analysis result of 60000 vehicles. The chart only depicts failed cars. Attributes Transmission and Country had most (stochastic) impact on the Class variable. The subset labeled 1 was re-identified by experts as a problem already known. Set 2 could not be given a causal explanation.

Example 2

The second dataset consisted of 300000 cars that exposed a many-valued class variable, hence the different gray levels of the circles in figure 5. Although there was no explanation for the sets 3, the subset 4 represented 900 cars the increased failure rate of which could be tracked down to the respective values of the attributes Mileage and RoadSurface.

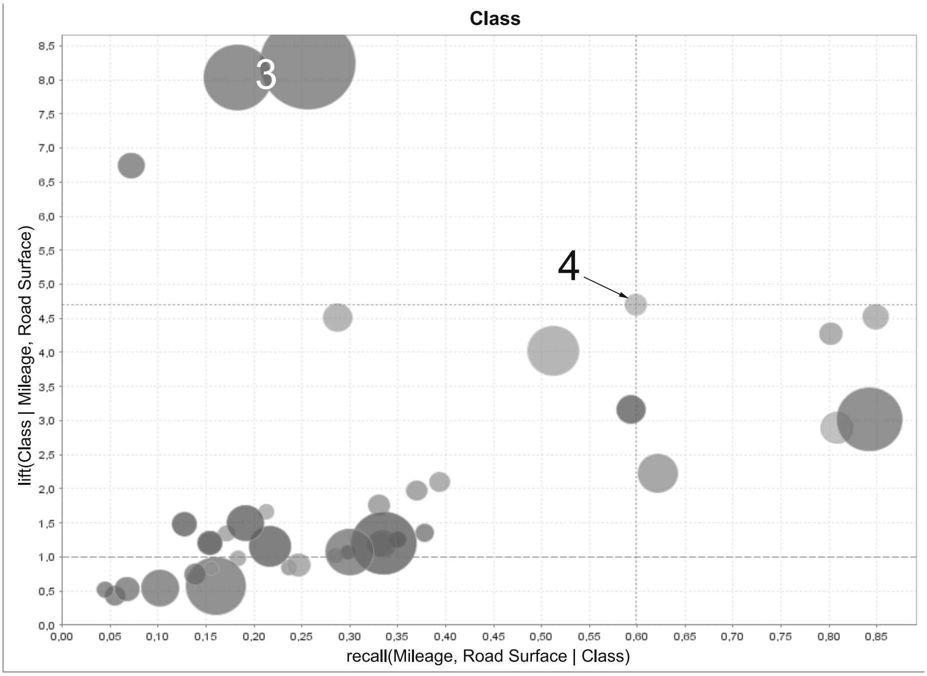

Fig. 5. Although it was not possible to find a reasonable description of the vehicles contained in subsets 3, the attribute values specifying subset 4 were identified to have a causal impact on the class variable.

Example 3

As a last example, the same dataset as in example 2 yielded the result as shown in figure 6. Here, an expert user changed the conditioning attributes manually and identified the set 5 which represented a subset of cars whose failure type and rate were affected by the respective attribute values.

User Acceptance

The proposed visualization technique has proven to be a valuable tool that facilitates the identification of subsets of cars that may expose a critical dependence between configuration and failure type. Generally, it represents an intuitive way of displaying high-dimensional, nominal data. A pure association rule analysis needs heavy postprocessing of the rules since a lot of rules are generated due to the commonly small failure rate. The presented approach can be considered a visual exploration aid for association rules. However, one has to admit that the rules represented by the circles share the same attributes in the antecedence, hence the sets of cars covered by these rules are mutually disjoint, which is a considerable difference to general rule sets.

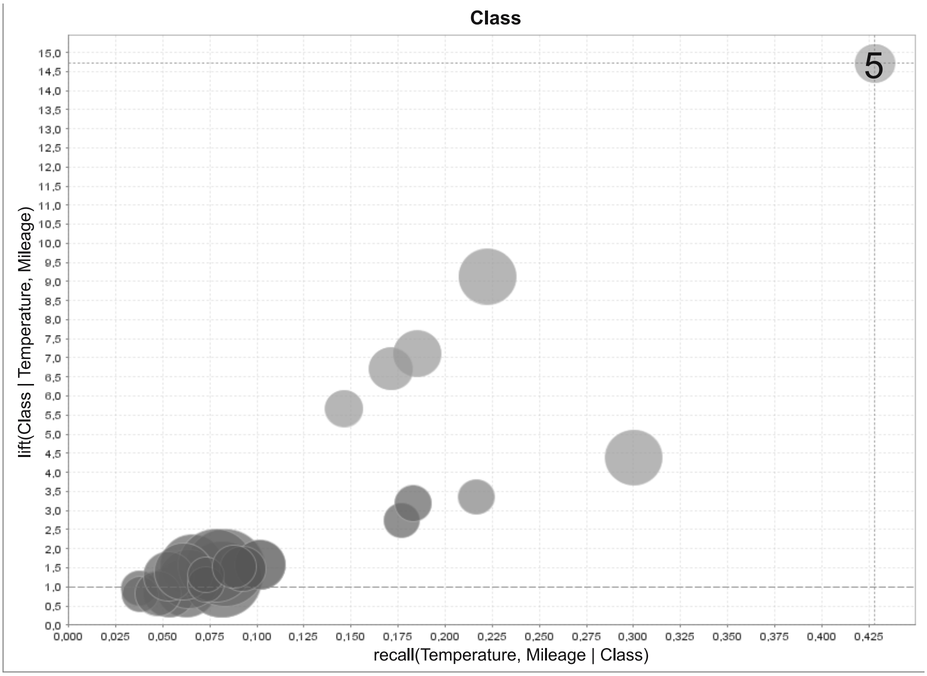

Fig. 6. In this setting the user selected the parent attributes manually and was able to identify the subset 5, which could be given a causal interpretation in terms of the conditioning attributes Temperature and Mileage

5 Conclusion

This paper presented an empirical evidence that graphical models can provide a powerful framework for data- and knowledge-driven applications with massive amounts of information. Even though the underlying data structures can grow highly complex, both presented projects implemented at two automotive companies result in effective complexity reduction of the methods suitable for intuitive user interaction.

References

[1] Agrawal, R., Imielinski, T., Swami, A.N.: Mining Association Rules between Sets of Items in Large Databases. In: Buneman, P., Jajodia, S. (eds.) Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data, May 26-28, pp. 207–216. ACM Press, Washington (1993)

[2] Borgelt, C., Kruse, R.: Some experimental results on learning probabilistic and possibilistic networks with different evaluation measures. In: 1st International Joint Conference on Qualitative and Quantitative Practical Reasoning (EC-SQARU/FAPR 1997), Bad Honnef, Germany, pp. 71–85 (1997)

[3] Borgelt, C., Kruse, R.: Probabilistic and possibilistic networks and how to learn them from data. In: Kaynak, O., Zadeh, L., Turksen, B., Rudas, I. (eds.) Computational Intelligence: Soft Computing and Fuzzy-Neuro Integration with Applications. NATO ASI Series F, pp. 403–426. Springer, New York (1998)

[4] Borgelt, C., Kruse, R.: Graphical Models — Methods for Data Analysis and Mining. John Wiley & Sons, United Kingdom (2002)

[5] Castillo, E., Gutiérrez, J.M., Hadi, A.S.: Expert Systems and Probabilistic Network Models. Springer, Heidelberg (1997)

[6] Cooper, G.F., Herskovits, E.: A Bayesian Method for the Induction of Probabilistic Networks from Data. Machine Learning 9, 309–347 (1992)

[7] Gärdenfors, P.: Knowledge in the Flux—Modeling the Dynamics of Epistemic States. MIT Press, Cambridge (1988)

[8] Gebhardt, J., Detmer, H., Madsen, A.L.: Predicting Parts Demand in the Automotive Industry — An Application of Probabilistic Graphical Models. In: Proc. International Joint Conference on Uncertainty in Artificial Intelligence (UAI 2003), Bayesian Modelling Applications Workshop, Acapulco, Mexico, August 4-7 (2003)

[9] Gebhardt, J., Borgelt, C., Kruse, R., Detmer, H.: Knowledge Revision in Markov Networks. Journal on Mathware and Soft Computing, Special Issue From Modelling to Knowledge Extraction XI(2–3), 93–107 (2004)

[10] Gebhardt, J., Kruse, R.: Knowledge-Based Operations for Graphical Models in Planning. In: Godo, L. (ed.) ECSQARU 2005. LNCS (LNAI), vol. 3571, pp. 3–14. Springer, Heidelberg (2005)

[11] Heckerman, D., Geiger, D., Chickering, D.M.: Learning Bayesian Networks: The Combination of Knowledge and Statistical Data. Technical Report MSR-TR-94-09, Microsoft Research, Advanced Technology Division, Redmond, WA, 1994 (Revised February 1995)

[12] Lauritzen, S.L., Spiegelhalter, D.J.: Local Computations with Probabilities on Graphical Structures and Their Application to Expert Systems. Journal of the Royal Statistical Society, Series B 2(50), 157–224 (1988)

[13] Pearl, J.: Aspects of Graphical Models Connected with Causality. In: 49th Session of the International Statistics Institute (1993)

[14] Pearl, J.: Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Morgan Kaufmann, San Mateo (1988)

[15] Steinbrecher, M., Kruse, R.: Visualization of Possibilistic Potentials. In: Melin, P., Castillo, O., Aguilar, L.T., Kacprzyk, J., Pedrycz, W. (eds.) IFSA 2007. LNCS (LNAI), vol. 4529, pp. 295–303. Springer, Heidelberg (2007)