

Knowledge Revision in Markov Networks

J. Gebhardt

Intelligent Systems Consulting (ISC)
Leonhardtstraße 19, D-29227 Celle, Germany
e-mail: ISC-Gebhardt@t-online.de

C. Borgelt, R. Kruse

Dept. of Knowledge Processing and Language Engineering
Otto-von-Guericke-University of Magdeburg
Universitätsplatz 2, D-39106 Magdeburg, Germany

H. Detmer

Volkswagen Group, K-DOB-11
D-38436 Wolfsburg, Germany

Abstract

A lot of research in graphical models has been devoted to developing correct and efficient evidence propagation methods, like join tree propagation or bucket elimination. With these methods it is possible to *condition* the represented probability distribution on given evidence, a *reasoning process* that is sometimes also called *focusing*. In practice, however, there is the additional need to *revise* the represented probability distribution in order to reflect some *knowledge changes* by satisfying new frame conditions. Pure evidence propagation methods, as implemented in the known commercial tools for graphical models, are unsuited for this task. In this paper we develop a consistent scheme for the important task of revising a Markov network so that it satisfies given (conditional) marginal distributions for some of the variables. This task is of high practical relevance as we demonstrate with a complex application for item planning and capacity management in the automotive industry at Volkswagen Group.

1 Introduction

In the last decade graphical models have become one of the most popular tools to structure uncertain knowledge about high-dimensional domains [20, 17, 1] in order to make reasoning in such domains feasible [19, 14]. Their most prominent representatives are Bayesian networks [19], which are based on directed graphs and conditional distributions, and Markov networks [16], which refer to undirected graphs and marginal distributions or factor potentials.

For both types of networks several clear, correct, and efficient evidence propagation methods have been developed, with join tree propagation [16, 14] and bucket elimination [4] being among the most widely known. With these methods it is possible to *condition* the probability distribution represented by a graphical model on given evidence, i.e. on observed values for some of the variables, a reasoning process that is also called *focusing*. Efficient and user-friendly commercial tools for this task, like HUGIN [11] and NETICA [18], are widely available.

In practice, however, the need also arises to *revise* the probability distribution that is represented by a graphical model so that it satisfies given frame conditions, for example, given marginal distributions. For this task pure evidence propagation methods like join tree propagation and bucket elimination are unsuited. The available commercial tools, however, only provide these standard methods, and thus do not support the revision operation.

The research we report about here was triggered by a real-world application at the automobile manufacturer Volkswagen Group, where Markov networks are used for item planning and capacity management [5, 7, 8, 9]. In this application one has to adapt a Markov network representing the probabilities of combinations of equipment items so that given (conditional) marginal distributions are satisfied. They result from, for instance, given capacity limits, technical rules, sales programs, and the dynamics of selection frequencies of various equipment items.

The paper is structured as follows: Section 2 introduces to the complex item planning problem, to the choice of Markov networks as the appropriate representation structure, and to the fundamentals of the revision operator. In section 3 we state generally the knowledge-based problem of revising the marginals of the probability distribution represented by a Markov network. In order to specify what we want to preserve from the original distribution in the revision process, we need probabilistic dependence measures, which are discussed in section 4. In section 5 we apply iterative proportional fitting and its properties, which provides us with a solution of the revision problem. Finally, section 6 presents the resulting revision algorithm, and section 7 briefly describes some aspects of software development.

2 Motivating Real-World Application

2.1 Item Planning at Volkswagen Group

In contrast to many competing car manufacturers, Volkswagen Group favours a marketing policy that provides a maximum degree of freedom in choosing individual specifications of vehicles. That is, considering personal preferences, a customer may select from a large variety of options, each of which is taken from a so-called item family that characterizes a certain line of equipment. Body variants, engines, gearshifts, door layouts, seat coverings, radios, and navigation systems reflect only a small subset of the whole range of item families. In case of the VW Golf – being Volkswagen’s most popular car class – there are about 200 families with typically 4 to 8 values each, and a total range of cardinalities from 2 up to 150.

Of course, not all of the possible instantiations of item variables lead to valid

vehicle configurations, since technical rules, restrictions in manufacturing and sales requirements induce a common rule system that limits the acceptable ways of item combination. Nevertheless, dealing with more than 10,000 technical rules in the Golf class and even more rules delivered by the sales programs for the special needs of the different countries, there remains a giant number of correct vehicle specifications. In fact, compared to a total of 650,000 Golf cars in year 2002, one can find only a small number of vehicles within the whole production line that have identical specifications.

The major aim of the project EPL (EigenschaftenPLANung, the German word for item planning) at Volkswagen Group is the development and implementation of a software system that supports item planning, parts demand calculation, and capacity management with the aim of short-term as well as medium-term forecasts up to 24 months of future vehicle production.

In order to reach high quality of planning results, all relevant information sources have to be considered, namely rules for the correct combination of items into complete vehicle specifications, samples of produced vehicles as a reflection of customers' preferences, market forecasts that lead to stipulations of modified item rates for planning intervals, capacity restrictions, and production programs that fix the number of planned vehicles.

W.r.t. the logistics view, the most essential result of the item planning process is to assess the rates of all those item combinations that are known to be relevant for the demand calculation of parts, always related to a certain vehicle class in a certain planning interval. The importance of these item combinations arises from the fact that a vehicle can be interpreted as a large set of installation points, each of which is characterized by a set of alternative built-in parts for the corresponding location. Which one of the alternative parts has to be chosen at an installation point, depends on its installation condition that can be specified by an item combination. Of course, at a selected installation point, all occurring installation conditions have to be disjoint, and their disjunction has to form a tautology. That is, given any correct vehicle specification, for each installation point we obtain a unique decision which of its alternative parts has to be used.

In the context of the Golf class, we find a total of about 70,000 different item combinations required as installation conditions for the whole set of installation points. The data structure that lists all installation points, their installation conditions, and the built-in quantities of the referenced parts, is called a variants-related bill of material. The task of predicting the total demand of a certain part with respect to a future planning interval is to sum up the demands over all of its installation points. The demand at any installation point results from multiplying the rate of the item combination that represents its installation condition with the built-in quantity and the total number of vehicles intended to be produced in the respective planning interval.

We conclude that calculating parts demand reduces to a simple operation, whenever the rates of all involved item combinations can be computed.

2.2 Markov Network Model

The first step in the project EPL was to search for an appropriate planning model that supports a decomposed representation of the qualitative and quantitative dependency structure between item families. We had to take into account that we deal with a finite set of discrete item variables, as well as that we get conditional independences induced by the given rule systems and customers' preferences.

Since logical rule systems can be transformed into a relational setting, and rates for item combinations may be identified as (frequentistic or subjective) occurrence probabilities within a large sample of historical or predictably valid vehicle specifications, Markov networks turned out to be the most promising environment to satisfy the given modelling purposes.

Let $U = \{A_1, \dots, A_n\}$ be a set of (discrete) random variables with respective domains $\text{dom}(A_1), \dots, \text{dom}(A_n)$. A *Markov network* is an undirected graph $G = (U, E)$ of these random variables together with a set of functions on the spaces induced by variables underlying the maximal cliques¹ of the graph. The structure of the graph encodes conditional independence statements between (sets of) random variables that hold in the joint probability distribution on the space induced by the variables in U . This is done by u -separation: Two (disjoint) sets X and Y of variables are conditionally independent given a third set Z if all paths from a variable in X to a variable in Y contain a variable in Z . Conditional independence of X and Y given Z , written $X \perp\!\!\!\perp Y \mid Z$, means that

$$p_{XY|Z}(x, y \mid z) \equiv p_{X|Z}(x \mid z) \cdot p_{Y|Z}(y \mid z),$$

where x , y and z are (variable) value vectors from the spaces induced by the random variables in X , Y , and Z . It can be shown that if the graph encodes only correct (i.e. actually obtaining) conditional independences by u -separation, then the joint probability distribution p_U factorizes according to [10]

$$p_U(u) \equiv \prod_{C \in \mathcal{C}} \phi_C(c).$$

The set \mathcal{C} is the set of all sets of variables underlying the maximal cliques of the graph. u and c are (variable) value vectors over the variables in U and C , respectively. The ϕ_C are the abovementioned functions on the spaces induced by the variables in these sets. They are called *factor potentials* [2] and can be defined in different ways from the corresponding marginal distributions.

2.3 Revision Operator

Coming back to our real-world application, once a basic prior Markov network for a certain planning interval has been generated, it becomes subject to a variety of planning operations which involve marketing and sales stipulations (e.g., installation rate of comfort navigation system increases from 20 % to 35 %) and capacity

¹A clique is a complete (fully connected) subgraph and it is called maximal if it is not contained in another complete subgraph.

restrictions from logistics (e.g., maximum availability of seat coverings in leather is 5,000). These quantitative input data are strongly related to the planning interval itself and therefore not learnable, neither from historical data nor from the non-probabilistic rules system. They typically consist of predicted installation rates or absolute demands for single items, sets of items, or (sets of) item combinations, and are frequently related to refined planning contexts (e.g., VW Golf with all-wheel drive for the US market).

In mathematical terms, this sort of additional information leads to competing partial or total changes of selected (conditional) low-dimensional probability distributions in the Markov network. They can be interpreted as the basis for a revision operation, where a prior state of knowledge (represented by the initial Markov network) in the light of new information (which is the new set of probability distributions) is revised to a posterior state of knowledge that incorporates the new information in the sense of the *principle of minimal change* [6].

Translated into the probabilistic framework, the task is to calculate a posterior Markov network that meets the new distribution conditions, only accepting a minimal change of the qualitative and quantitative interaction structures of the underlying prior distribution.

Preserving the qualitative structure means to leave the network structure (independence relations) unchanged, which, of course, is only possible if the revision conditions themselves do not contradict this assumption. That is, all new distributions fit the clique structure of the Markov network.

Preserving the quantitative structure means to hold the cross-product ratios of the distributions, as far as they are not affected by the revision information.

Using some concepts of multivariate statistics, especially the method of *proportional iterative fitting* [20], and formalizing the concept of revision and the principle of minimal change to Markov networks, it is possible to prove that there is a unique solution of the above mentioned revision problem. The only natural side condition to get this result is that the new marginal probability distributions do not contradict themselves, and do not contradict to the rules system that is integrated in the prior network. An in-depth examination of such a clean *knowledge-based solution* of the revision problem can be found in [7]. The obtained results have the advantage that they are also useful for the development of *decision support algorithms* for the recognition, description, avoidance, and goal-oriented correction of inconsistent revision input data.

In the following sections, we present a knowledge-based probabilistic formalization and solution of the basic revision problem for Markov networks, restricted to a set of non-conditional marginals for single variables. This clean probabilistic approach avoids any concepts offered by calculi with deviating semantic foundations, for example, to minimize probabilistic difference measures that could be adapted from information theory.

3 Formalization of the Revision Problem

Suppose that we are given a Markov network which represents a joint probability distribution p_U on a set U of variables. In addition, let J be a set of indices of variables together with marginal probability distributions $p_{A_j}^*$ on the domains of the variables A_j for all $j \in J$. The task is to *revise* the Markov network in such a way that the marginal probability distributions p'_{A_j} of the distribution p'_U represented by the new Markov network coincide with the distributions $p_{A_j}^*$ for all $j \in J$.

W.r.t. the principle of minimal change, p'_U should also preserve as much as possible the properties of the old probability distribution p_U . This means that the structure of the graph underlying the Markov network must remain unchanged, i.e., the new probability distribution should exhibit (at least) the conditional independences that hold in the old one. Furthermore, we require the (conditional) dependences between variables, as they can be assessed by some dependence measure, to stay constant or to change as little as possible.

It is quite easy to see that p'_U in case of existence is uniquely determined: although we may find many distributions that fit the conditions on the new marginals, there can be only one distribution that maximally preserves the available knowledge on qualitative and quantitative dependences adapted from the prior probability distribution.

The next section investigates probabilistic dependence measures in order to reach the intended probabilistic formalization of the revision problem.

4 Probabilistic Dependence Measures

There is a large variety of dependence measures for random variables (or random vectors), both continuous and discrete. However, for the task at hand two are especially useful, namely the well-known mixed derivative measure and the closely related cross-product ratio.

4.1 The Mixed Derivative Measure

Two (continuous) random variables X and Y are called conditionally independent given a third vector Z iff their joint density function factorizes according to

$$f_{XY|Z}(x, y | z) \equiv f_{X|Z}(x | z) \cdot f_{Y|Z}(y | z)$$

or equivalently iff

$$\log f_{XY|Z}(x, y | z) \equiv \log f_{X|Z}(x | z) + \log f_{Y|Z}(y | z).$$

Since only in this case the mixed derivative² w.r.t. x and y of the logarithm of the joint density function of X and Y given Z vanishes, one can also characterize the

²We assume implicitly that the density function is differentiable sufficiently often.

conditional independence of X and Y given Z by

$$\frac{d^2}{dx dy} \log f_{XY|Z}(x, y | z) \equiv 0.$$

It should be noted that this characterization is not limited to continuous random variables, but can also be applied to discrete ones, if the derivative is replaced by the so-called forward difference $\nabla_x = g(x+1) - g(x)$. This way of characterizing conditional independence also provides us with a way to measure the conditional dependence of two variables given a third, namely by evaluating the mixed derivative of their joint density function. Note, however, that this measure does not provide a single value, but is a function of the values x , y , and z .

Preserving the underlying graph (i.e. conditional independences) as well as preserving the structure of (conditional) dependences between random variables (or vectors) can now be defined formally as preserving the mixed derivative measure (as a function on the joint domain of the random variables, of course).

4.2 The Cross-Product Ratio

The cross-product ratio is a common probabilistic measure for the association of two random events A and B . It is defined as

$$\text{cpr}(A, B) = \frac{P(A \cap B) \cdot P(\bar{A} \cap \bar{B})}{P(\bar{A} \cap B) \cdot P(A \cap \bar{B})}.$$

However, the decision-theoretic meaning of the cross-product ratio becomes clearer if it is expressed either in conditional probabilities or using the notion of the (conditional) *odds* of an event A (given another event B), i.e.

$$\text{odds}(A) = \frac{P(A)}{P(\bar{A})} \quad \text{and} \quad \text{odds}(A | B) = \frac{P(A | B)}{P(\bar{A} | B)},$$

which yields

$$\text{cpr}(A, B) = \frac{P(A | B) \cdot P(\bar{A} | \bar{B})}{P(\bar{A} | B) \cdot P(A | \bar{B})} = \frac{\text{odds}(A | B)}{\text{odds}(A | \bar{B})}.$$

$\text{cpr}(A, B)$ quantifies the balance between whether A or \bar{A} is more probable in case that B or \bar{B} is observed. It is easy to see that the cross-product ratio is 1 if and only if the two events A and B are independent (i.e., knowledge on whether B or \bar{B} is true does not influence the probabilities of A and \bar{A} , respectively).

It should be noted that the cross-product ratio is closely related to the mixed derivative measure. To see this, consider two binary variables X and Y , i.e. variables with values 0 and 1. Then the joint probability distribution of these two variables can be represented as [20]

$$p_{XY}(x, y) = p_{XY}(0, 0)^{(1-x)(1-y)} p_{XY}(0, 1)^{(1-x)y} p_{XY}(1, 0)^{x(1-y)} p_{XY}(1, 1)^{xy}.$$

Therefore it is

$$\log p_{XY}(x, y) = \log p_{XY}(0, 0) + x \log \frac{p_{XY}(1, 0)}{p_{XY}(0, 0)} + y \log \frac{p_{XY}(0, 1)}{p_{XY}(0, 0)} + xy \log \frac{p_{XY}(1, 1)p_{XY}(0, 0)}{p_{XY}(0, 1)p_{XY}(1, 0)}$$

and consequently

$$\frac{d^2}{dx dy} \log p_{XY}(x, y) \equiv \log \frac{p_{XY}(1, 1)p_{XY}(0, 0)}{p_{XY}(0, 1)p_{XY}(1, 0)} = \log \text{cpr}(X = 1, Y = 1).$$

In this special case the mixed derivative measure, which, in general, is a function, reduces to a constant. Due to this relationship between the cross-product ratio and the mixed derivative measure, it is already plausible that preserving the one entails preserving the other.

With respect to the structural needs of Markov networks, we generalize the cross-product ratio from events to random variables by defining

$$\text{cpr}_{XY|Z}(x_1, x_2, y_1, y_2, z) = \frac{f_{XY|Z}(x_1, y_1 | z) \cdot f_{XY|Z}(x_2, y_2 | z)}{f_{XY|Z}(x_2, y_1 | z) \cdot f_{XY|Z}(x_1, y_2 | z)}.$$

As with the mixed derivative measure, preserving conditional independences as well as the structure of (conditional) dependences between random variables can now be defined formally as preserving this generalized cross-product ratio (as a function on $\text{dom}(X)^2 \times \text{dom}(Y)^2 \times \text{dom}(Z)$, of course).

5 Solution of the Revision Problem

In order to solve the revision problem specified in sections 3 and 4, we need an approach that supports the construction of the distribution p'_U that we are searching for. From multivariate statistics, *iterative proportional fitting* is a well-known algorithm for adapting the marginal distributions of a given joint distribution to desired values [20]. It consists in computing the following sequence of probability distributions:

$$\begin{aligned} p_U^{(0)}(u) &\equiv p_U(u) \\ \forall i = 1, 2, \dots: \quad p_U^{(i)}(u) &\equiv p_U^{(i-1)}(u) \frac{p_{A_j}^*(a)}{p_{A_j}^{(i-1)}(a)} \end{aligned}$$

where j is the $((i - 1) \bmod |J| + 1)$ -th element of J , the index set that indicates the variables for which marginal distributions are given (cf. section 3). $p_{A_j}^*$ is the desired marginal probability distribution on the domain of the variable A_j , $p_{A_j}^{(i-1)}$ the corresponding distribution as it can be computed from $p_U^{(i-1)}$ by summing over the values of all variables in U except A_j .

In each step the probability distribution is modified in such a way that the resulting distribution satisfies one given marginal distribution (namely the distribution $p_{A_j}^*$). However, this will, in general, change the marginal distribution for a variable A_k , which has been processed in a previous step. Therefore the adaptation has to be iterated, traversing the set of variables several times.

In the following we show that iterative proportional fitting has the desired properties that we listed in section 3: If the revision problem has a solution, it converges to a (uniquely determined) probability distribution that has the desired marginals, has the same structure of (in)dependences as the original distribution as well as the same conditional probability distribution of the non-revised variables given the revised ones.

Furthermore, at the end of these investigations, it turns out that the revised distribution differs minimally from the original distribution with respect to Kullback-Leibler information divergence, so that one can get an additional information-theoretic confirmation of the probabilistic solution of the revision problem.

5.1 Preservation of the Dependence Structure

For the generalized cross-product ratio we have to distinguish two cases. Either the attribute A_j , the marginal distribution of which is revised in step i , is equal to X or Y (case 1) or to neither of the two (case 2). Suppose first that A_j is equal to X or equal to Y . Then

$$\begin{aligned}
 p_{XY|Z}^{(i)}(x, y | z) &= \frac{\sum_{\substack{\forall A_k \in U - \{X, Y, Z\}: \\ a_k \in \text{dom}(A_k)}} p_U^{(i)}(u)}{\sum_{\substack{\forall A_k \in U - \{Z\}: \\ a_k \in \text{dom}(A_k)}} p_U^{(i)}(u)} \\
 &= \frac{\sum_{\substack{\forall A_k \in U - \{X, Y, Z\}: \\ a_k \in \text{dom}(A_k)}} p_U^{(i-1)}(u) \frac{p_{A_j}^*(a_j)}{p_{A_j}^{(i-1)}(a_j)}}{\sum_{\substack{\forall A_k \in U - \{Z\}: \\ a_k \in \text{dom}(A_k)}} p_U^{(i-1)}(u)} \\
 &= p_{XY|Z}^{(i-1)}(x, y | z) \cdot \frac{p_{A_j}^*(a_j)}{p_{A_j}^{(i-1)}(a_j)}
 \end{aligned}$$

where the somewhat sloppy notation w.r.t. the sums is intended to indicate that the sum is taken over all values of all attributes in U except X , Y , and Z , or except Z alone, respectively. The factor $p_{A_j}^*(a_j)/p_{A_j}^{(i-1)}(a_j)$ is missing in the denominator, because in the denominator it is summed over all values of X and Y and thus over all values of A_j (since either $X = A_j$ or $Y = A_j$). This consideration also makes it clear that if A_j is equal to neither X nor Y , then

$$p_{XY|Z}^{(i)}(x, y | z) \equiv p_{XY|Z}^{(i-1)}(x, y | z),$$

because in this case the summation removes the factor from both the numerator and the denominator. As a consequence we have immediately that if A_j is equal to neither X nor Y , the generalized cross-product ratio stays the same, because the conditional distributions it is computed from are unchanged. On the other hand, if A_j is equal to X or equal to Y , we have

$$\begin{aligned}
& \text{cpr}_{XY|Z}^{(i)}(x_1, x_2, y_1, y_2, z) \\
&= \frac{p_{XY|Z}^{(i)}(x_1, y_1 | z) \cdot p_{XY|Z}^{(i)}(x_2, y_2 | z)}{p_{XY|Z}^{(i)}(x_2, y_1 | z) \cdot p_{XY|Z}^{(i)}(x_1, y_2 | z)} \\
&= \frac{p_{XY|Z}^{(i-1)}(x_1, y_1 | z) \cdot \frac{p_{A_j}^*(a_1)}{p_{A_j}^{(i-1)}(a_1)} \cdot p_{XY|Z}^{(i-1)}(x_2, y_2 | z) \cdot \frac{p_{A_j}^*(a_2)}{p_{A_j}^{(i-1)}(a_2)}}{p_{XY|Z}^{(i-1)}(x_2, y_1 | z) \cdot \frac{p_{A_j}^*(a_2)}{p_{A_j}^{(i-1)}(a_2)} \cdot p_{XY|Z}^{(i-1)}(x_1, y_2 | z) \cdot \frac{p_{A_j}^*(a_1)}{p_{A_j}^{(i-1)}(a_1)}} \\
&= \frac{p_{XY|Z}^{(i-1)}(x_1, y_1 | z) \cdot p_{XY|Z}^{(i-1)}(x_2, y_2 | z)}{p_{XY|Z}^{(i-1)}(x_2, y_1 | z) \cdot p_{XY|Z}^{(i-1)}(x_1, y_2 | z)} \\
&= \text{cpr}_{XY|Z}^{(i-1)}(x_1, x_2, y_1, y_2, z),
\end{aligned}$$

where a_k is the value that attribute A_j has in the vector x_k for $k = 1, 2$.

5.2 Preservation of the Conditional Distribution

In the previous section we obtained the result that

$$p_{XY|Z}^{(i)}(x, y | z) \equiv p_{XY|Z}^{(i-1)}(x, y | z)$$

if the attribute A_j is equal to neither X nor Y . The straightforward generalization to random vectors X and Y yields as a special case that the conditional distribution of the variables in $U - V$, for which no marginal distribution is fixed, given the variables in $V = \{A_j | j \in J\}$, for which a marginal distribution is fixed, stays unchanged. That is,

$$p_{U-V|V}^{(i)}(w | v) \equiv p_{U-V|V}^{(i-1)}(w | v).$$

All we have to do is to choose $Z = V$ and X and Y in some arbitrary way satisfying $X \cup Y = U - V$.

5.3 Convergence

An exact proof that iterative proportional fitting converges to a (uniquely determined) probability distribution with the desired marginals and having minimal Kullback-Leibler information divergence from the original one is difficult and beyond the scope of this paper. Therefore we confine ourselves to indicating why such a proof is possible, by considering the sequence of values of the Kullback-Leibler information divergence [20]. Let p_U^* be any probability distribution on the universe

of discourse that satisfies the revision settings (that is, it has the desired marginal distributions). Then

$$\begin{aligned}
 & I_{\text{KL}}\left(p_U^*, p_U^{(i)}\right) \\
 &= \sum_{u \in \text{dom}(U)} p_U^*(u) \log_2 \frac{p_U^*(u)}{p_U^{(i)}(u)} \\
 &= \sum_{u \in \text{dom}(U)} p_U^*(u) \log_2 \frac{p_U^*(u)}{p_U^{(i-1)}(u)} + \sum_{u \in \text{dom}(U)} p_U^*(u) \log_2 \frac{p_U^{(i-1)}(u)}{p_U^{(i)}(u)} \\
 &= I_{\text{KL}}\left(p_U^*, p_U^{(i-1)}\right) + \sum_{u \in \text{dom}(U)} p_U^*(u) \left(\log_2 p_U^{(i-1)}(u) - \log_2 p_U^{(i)}(u) \right) \\
 &\stackrel{(*)}{=} I_{\text{KL}}\left(p_U^*, p_U^{(i-1)}\right) + \sum_{u \in \text{dom}(U)} p_U^*(u) \left(\log_2 p_{A_j}^{(i-1)}(a) - \log_2 p_{A_j}^*(a) \right) \\
 &= I_{\text{KL}}\left(p_U^*, p_U^{(i-1)}\right) + \sum_{a \in \text{dom}(A_j)} p_{A_j}^*(a) \left(\log_2 p_{A_j}^{(i-1)}(a) - \log_2 p_{A_j}^*(a) \right) \\
 &= I_{\text{KL}}\left(p_U^*, p_U^{(i-1)}\right) - \sum_{a \in \text{dom}(A_j)} p_{A_j}^*(a) \left(\log_2 p_{A_j}^*(a) - \log_2 p_{A_j}^{(i-1)}(a) \right) \\
 &= I_{\text{KL}}\left(p_U^*, p_U^{(i-1)}\right) - I_{\text{KL}}\left(p_{A_j}^*, p_{A_j}^{(i-1)}\right)
 \end{aligned}$$

The step marked (*) follows, because according to the revision formula of the iterative proportional fitting procedure it is

$$\begin{aligned}
 p_U^{(i)}(u) &= p_U^{(i-1)}(u) \frac{p_{A_j}^*(a)}{p_{A_j}^{(i-1)}(a)} \\
 \Leftrightarrow \log_2 p_U^{(i)}(u) &= \log_2 p_U^{(i-1)}(u) + \log_2 p_{A_j}^*(a) - \log_2 p_{A_j}^{(i-1)}(a) \\
 \Leftrightarrow \log_2 p_U^{(i-1)}(u) - \log_2 p_U^{(i)}(u) &= \log_2 p_{A_j}^{(i-1)}(a) - \log_2 p_{A_j}^*(a).
 \end{aligned}$$

Combining the above derivation with the well-known fact that the Kullback-Leibler information divergence is non-negative and zero only if the two distributions coincide³, we arrive at a highly suggestive result: The Kullback-Leibler information divergence between p_U^* and $p_U^{(i)}$ decreases monotonically. Since

$$I_{\text{KL}}\left(p_{A_j}^*, p_{A_j}^{(i-1)}\right) \geq 0, \quad \text{it is} \quad I_{\text{KL}}\left(p_U^*, p_U^{(i)}\right) \leq I_{\text{KL}}\left(p_U^*, p_U^{(i-1)}\right).$$

It is also bounded by zero. Therefore it must have a limit, suggesting that the process converges to a limiting probability distribution. Furthermore, since the increments in this sequence of Kullback-Leibler information divergences must go to zero (otherwise it would obviously not be convergent), it is intuitively clear that this limiting distribution will have the desired marginal distributions. However, this is only a plausible argument. A rigorous proof, which depends crucially on the notion of convex sets, can be found in [3].

³A proof of this property can be found, for instance, in [1].

6 The Revision Algorithm

Up to now we considered revision with iterative proportional fitting on the whole probability distribution p_U represented by the given Markov network. In practice, however, this cannot be done, because one of the main reasons for using a Markov network in the first place is that the distribution p_U cannot be represented as a whole due to its size. Therefore we have to consider how the revision has to be performed on the parameters of the given Markov network, that is, the factor potentials of the factorization. However, in implementations usually not the factor potentials are stored with the cliques, but the corresponding marginal distributions, because this has advantages w.r.t. evidence propagation.

The idea of the algorithm is to assign each attribute, the marginal distribution of which is to be revised, to a maximal clique of the Markov network, to use steps of iterative proportional fitting to adapt the marginal distributions on the maximal cliques, and to distribute the information added by such a revision to the other maximal cliques by standard evidence propagation (preferably carried out by join tree propagation). More formally, we split the set J of indices of all variables, the marginal distributions of which are to be revised, into sets J_C of indices for all $C \in \mathcal{C}$, where \mathcal{C} is the set of sets of variables underlying the maximal cliques of the Markov network (cf. section 2.2). These sets J_C of indices must satisfy the following three conditions:

- (i) $\bigcup_{C \in \mathcal{C}} J_C = J$,
- (ii) $\forall C_1, C_2 \in \mathcal{C} : (C_1 \neq C_2) \Rightarrow (J_{C_1} \cap J_{C_2} = \emptyset)$,
- (iii) $\forall j \in J : A_j \in \mathcal{C}$.

Thus each variable A_j is assigned to the clique C , for which $j \in J_C$. Note that there are, in general, several ways to split the set J into sets J_C , $C \in \mathcal{C}$, because a variable may be contained in several cliques and thus there may be a choice into which set J_C to put its index.

In pseudocode, the resulting revision algorithm reads:

```

forall  $C \in \mathcal{C}$  do                                (* traverse the maximal cliques and *)
   $p_C^{(0)}(c) := p_C(c)$                             (* initialize marginal distributions *)
   $i := 0$ ;                                           (* initialize the step counter *)
  repeat                                           (* iterative proportional fitting *)
     $i := i + 1$ ;                                     (* increment the step counter *)
    forall  $C \in \mathcal{C}$  do begin                       (* traverse the maximal cliques *)
      forall  $j \in J_C$  do                            (* traverse the assigned attributes *)
         $p_C^{(i)}(c) := p_C^{(i-1)} \cdot \frac{p_{A_j}^*(a)}{p_{A_j}^{(i-1)}(a)}$ ; (* revise the marginal distribution *)
        (* on the current maximal clique *)
      do evidence propagation;                       (* adapt the marginal distributions *)
    end                                             (* on all other maximal cliques *)
  until convergence;                               (* until limiting distribution reached *)

```

Convergence may be checked, for instance, by determining the maximal change of a marginal probability on a maximal clique: If this maximal change falls below a user-defined threshold, the loop is terminated. It is useful to add a check on the number of iterations done, because there may be no distribution that satisfies the revision settings and then convergence cannot be achieved. That is, if a user-defined threshold on the number of iterations is reached without meeting the convergence criterion, the loop is terminated and a failure reported. In this case the revision settings (i.e. the desired marginal distributions) have to be reconsidered and adapted.

Investigations on the theory and the development of a decision support tool that helps in finding and correcting inconsistent revision data shows one of the topics of current research. As an alternative to adapting the revision settings, it is also possible to take one of the competing revision distributions that level off during the iteration process as an approximation that meets at least a part of the desired constraints.

That what is done inside the loop over the maximal cliques in the above algorithm, namely applying some iterative proportional fitting steps only to the marginal distribution on a maximal clique and distributing the revision information afterwards by standard evidence propagation, is indeed equivalent to performing the analogous iterative proportional fitting steps on the whole probability distribution, can easily be seen by considering how join tree propagation [16, 14] works. Intuitively, the conditional probability distribution of all variables not contained in the current maximal clique given the variables in this maximal clique is factored out, the marginal distribution of the maximal clique is adapted, and the two factors are put together again.

7 Concluding Remarks

We presented a consistent scheme for revising marginal distributions of a probability distribution that is represented by a Markov network and showed that it has all desirable properties. Interestingly, many of the ingredients of this method could be collected from multivariate statistics. Nevertheless, except from some limited theoretical considerations (see [13]), the possibility of their application to adapting the marginal probabilities of a Markov network has been neglected or seems to be largely unknown in the AI community. Otherwise it would not be understandable why market-leading implementations of Bayesian network algorithms, like HUGIN [11] and NETICA [18], do not provide tools for such a revision operation.

Note that revision is only one of the many operations on Markov networks needed in order to develop and implement an innovative software system that is focused on, but not restricted to item planning and the prediction of parts demand in the automotive industry. Other necessary functionalities consist in the fusion of historical random samples with rules systems for a future planning interval into a Markov network, different kinds of updating operations, high-speed propagation and information retrieval techniques, and so forth. The importance of performance aspects is obvious, since it turns out that item planning, even when reduced to

a single vehicle class within a single planning week, requires handling Markov networks of about 200 cliques and maximum dimensionalities of at least 12 for individual cliques. As a consequence, domains with a cardinality of more than 100,000,000 elements have to be considered.

The already mentioned project EPL was initiated in 2001 by Corporate IT, Sales, and Logistics of the Volkswagen Group. The aim was to establish for all trademarks a common item planning system that reflects the modelling approach based on Markov networks.

System design and most of the implementation work of EPL is currently done by Corporate IT. The mathematical modelling, theoretical problem solving, and the development of efficient algorithms, extended by the implementation of a new software library called MARNEJ (MARKov Networks in Java) for the representation and the above-mentioned functionalities on Markov networks have been entirely provided by ISC Gebhardt.

The worldwide rollout of the system EPL to all trademarks of the Volkswagen Group will be started during the year 2004. Up to 15 system developers implement the client-server architecture in Java. The planned configuration uses 6 to 8 Hewlett Packard machines with 4 AMD Opteron 64-Bit CPUs and 16 GB of main memory each, and a terabyte storage device. The system is running Linux and an Oracle database system.

References

- [1] C. Borgelt and R. Kruse. *Graphical Models — Methods for Data Analysis and Mining*. J. Wiley & Sons, Chichester, United Kingdom 2002
- [2] E. Castillo, J.M. Gutierrez, and A.S. Hadi. *Expert Systems and Probabilistic Network Models*. Springer-Verlag, New York, NY, USA 1997
- [3] I. Csiszar. I-Divergence Geometry of Probability Distributions and Indirect Observations. *Studia Scientiarum Mathematicarum Hungarica* 2:299–318. Hungarian Academy of Sciences, Budapest, Hungary 1975
- [4] R. Dechter. Bucket Elimination: A Unifying Framework for Probabilistic Inference. *Proc. 12th Conf. on Uncertainty in Artificial Intelligence (UAI'96, Portland, OR, USA)*, 211–219. Morgan Kaufmann, San Mateo, CA, USA 1996
- [5] H. Detmer and J. Gebhardt. Markov-Netze für die Eigenschaftsplanung und Bedarfsvorschau in der Automobilindustrie. *Künstliche Intelligenz* (Themenheft Unsicherheit und Vagheit) 3/01:16–22. arendtap, Bremen, Germany 2001
- [6] P. Gardenfors. *Knowledge in Flux – Modeling the Dynamics of Epistemic States*. MIT Press, 1988
- [7] J. Gebhardt. The Revision Operator and the Treatment of Inconsistent Stipulations of Item Rates. Project EPL: Internal Report 9. ISC Gebhardt and Volkswagen Group, K-DOB-11, Wolfsburg, Germany 2001

- [8] J. Gebhardt and H. Detmer. Revision and Updating of Probabilistic Networks for Item Planning and Capacity Management in the Automotive Industry. *Proc. 9th Int. Conf. on Information Processing and Management of Uncertainty in Knowledge-based Systems (IPMU'02, Annecy, France)*, 35–38. Université de Savoie, Annecy, France 2002
- [9] J. Gebhardt, H. Detmer, and A.L. Madsen. Predicting Parts Demand in the Automotive Industry – An Application of Probabilistic Graphical Models. *Proc. Int. Joint Conf. on Uncertainty in Artificial Intelligence (UAI'03, Acapulco, Mexico)*, Bayesian Modelling Applications Workshop, 2003
- [10] J.M. Hammersley and P.E. Clifford. *Markov Fields on Finite Graphs and Lattices*. Unpublished manuscript, 1971. Cited in: [12]
- [11] Hugin Expert. HUGIN – Bayesian Network Software. Aalborg, Denmark 1989–2003. <http://www.hugin.dk>
- [12] V. Isham. An Introduction to Spatial Point Processes and Markov Random Fields. *Int. Statistical Review* 49:21–43. Int. Statistical Institute, Voorburg, Netherlands 1981
- [13] R. Jirousek. Solution of the Marginal Problem and Decomposable Distributions. *Kybernetika* 27:403–412, 1991
- [14] F.V. Jensen. *An Introduction to Bayesian Networks*. UCL Press, London, United Kingdom 1996
- [15] S. Kullback and R.A. Leibler. On Information and Sufficiency. *Annals of Mathematical Statistics* 22:79–86. Institute of Mathematical Statistics, Hayward, CA, USA 1951
- [16] S.L. Lauritzen and D.J. Spiegelhalter. 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. Blackwell, Oxford, United Kingdom 1988
- [17] S.L. Lauritzen. *Graphical Models*. Oxford University Press, Oxford, United Kingdom 1996
- [18] Norsys Software Corp. NETICA – Bayesian network software. Vancouver, BC, Canada 1995–2003. <http://www.norsys.com/>
- [19] J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference (2nd edition)*. Morgan Kaufman, San Mateo, CA, USA 1992
- [20] J. Whittaker. *Graphical Models in Applied Multivariate Statistics*. J. Wiley & Sons, Chichester, United Kingdom 1990