Visualization of Local Dependencies of Possibilistic Network Structures

Matthias Steinbrecher and Rudolf Kruse

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

Summary. In this chapter an alternative interpretation of the parameters of a Bayesian network motivates a new visualization method that allows for an intuitive insight into the network dependencies. The presented approach is evaluated with artificial as well as real-world industrial data to justify its applicability.

1 Introduction

The ever-increasing performance of database systems enables today's business organizations to collect and store huge amounts of data. However, the larger the data volumes grow the need to have sophisticated analyzation methods to extract hidden patterns does alike. The research area of Data Mining addresses these tasks and comprises intelligent data analysis techniques such as classification, prediction or concept description, just to name a few.

The latter technique of concept description tries to find common properties of conspicuous subsets of given samples in the database. For example, an automobile manufacturer may plan to investigate car failures by identifying common properties that are exposed by specific subsets of cars.

Good concept descriptions should have a reasonable length, i. e., they must not be too short in order not to be too general. Then again, long descriptions are too restrictive since they constrict the database samples heavily, resulting in only a few covered sample cases. Since we have to assume that the database entries expose hundreds of attributes, it is essential to employ a feature selection approach that reduces this number to a handy subset of significant attributes.

In this chapter, we assume the database entries having nominal attributes[1] with one distinguished attribute designating the class of each data sample. We will use probabilistic and possibilistic network induction methods to learn a dependence network from the database samples. Further, we only draw our attention to the class attribute and its conditioning attributes, which are its direct parents in the network, i. e., the subset of attributes that have a direct arc connecting it with the class attribute. Since most network induction algorithms allow

[1] For the treatment of metric attributes, a discretization phase has to precede the analysis task.

R. Bello et al. (Eds.): Granular Computing, STUDFUZZ 224, pp. 93–104, 2008.
springerlink.com

for the restriction of the number of parent attributes to some upper bound, we are in a favorable position to control the length of the concept descriptions to be generated.

We then show that the network structure alone does not necessarily provide us with a detailed insight into the dependencies between the conditioning attributes and the class attribute. Emphasis is then put on the investigation of the network's local structure, that is, the entries of its potential tables. Finally, a new visualization method for these potential tables is presented and evaluated.

The remainder of this chapter is structured as follows: Section 2 presents a brief review of the methods of probabilistic and possibilistic networks, mostly for introducing the nomenclature used in the following sections. In section 3 arguments for the importance of visualizing the network parameters are produced. This will lead to a concrete application and analysis in section 4. The chapter concludes with section 5, giving an outlook of intended further investigations.

2 Background

For the formal treatment of sample cases or objects of interest, we identify each sample case with a tuple t that exposes a fixed number of attributes $\{A_1, \ldots, A_n\}$, each of which can assume a value with the finite respective domain $\mathrm{dom}(A_i) = \{a_{i1}, \ldots, a_{ir_i}\}, i = 1, \ldots, n$. Let Ω denote the set of all possible tuples, then we can model a database D, which constitutes the starting point of analysis, as a weight function $w_D : \Omega \to \mathbb{N}$ that assigns to each tuple $t \in \Omega$ the number of occurences in the database D. The total number of tuples or sample cases in D is $N = \sum_{t\in\Omega} w_D(t)$. The fact $w_D(t) = 0$ states, that the tuple t is not contained in D. With this definition, the weight function can be considered an extended indicator function: The respective indicator function $\mathbb{1}_D$ would be defined as

$$\forall t \in \Omega : \mathbb{1}_D(t) = \min\{w_D(t), 1\}.$$

From w_D we can derive the following probability space $\mathbb{P}_D = (\Omega, \mathcal{E}, P)$ with the components defined as follows:

$$\mathcal{E} = 2^{\Omega},$$

$$\forall t \in \Omega : p(t) = \tfrac{w_D(t)}{N} \quad \text{and} \quad \forall E \in \mathcal{E} : P(E) = \textstyle\sum_{t\in E} p(t)$$

In the following, we only have one database at the time, so we drop the index D and refer simply to w as the source of all information and assume the space $\mathbb{P}_D$ to be the implicit probability space underlying all consequent probabilistic propositions. Therefore, a given database of sample cases represents a joint probability distribution. Even though the number of tuples in the database is small compared to $|\Omega|$, we have to look for means of further reducing the size of the joint distribution.

One prominent way are *Graphical Models*, which can be destinguished further between Markov Networks [10] and Bayesian Networks [12], the latter of which is introduced in the next section.

2.1 Bayesian Networks

From the database oriented point of view, reducing one large, high dimensional database table can be accomplished by decomposing it into several lower dimensional subtables. Under certain conditions one can reconstruct the initial table using the natural join operation. These certain conditions comprise the *conditional relational independence* between the attributes in the initial table. Attributes A and B are relationally independent given a third attribute C, if once any value of C is held fixed, the values of A and B are freely combinable.

The probabilistic analog consists of decomposing the high dimensional joint probability distribution into multiple distributions over (overlapping) subsets of attributes. If these sets of attributes are *conditionally probabilistically independent* given the instantiations of the attributes contained in the overlap, a lossless reconstruction of the original joint distribution is possible via the chain rule of probability:

$$\forall \tau \in \mathcal{S}_n : P(A_1, \dots, A_n) = \prod_{i=1}^{n} P(A_{\tau(i)} \mid A_{\tau(i-1)}, \dots, A_{\tau(1)})$$

$\mathcal{S}_n$ denotes the symmetric group of permutations of n objects. The description which attributes are involved in a conditional independence relation is encoded in a directed acyclic graph (DAG) in the following way: The nodes of the graph correspond to the attributes. Let parents(A) denote the set of all those nodes that have a directed link to node A. Then, given an instantiation of the attributes in parents(A), attribute A is conditionally independent of the remaining attributes. Formal: Let $X = \{A_1, \dots, A_k\}$, $Y = \{B_1, \dots, B_l\}$ and $Z = \{C_1, \dots, C_m\}$ denote three disjoint subsets of attributes, then X and Y are conditionally probabilistically independent given Z, if the following equation holds:

$$\begin{aligned}
&\forall a_1 \in \mathrm{dom}(A_1) : \cdots \forall a_k \in \mathrm{dom}(A_k) : \\
&\quad \forall b_1 \in \mathrm{dom}(B_1) : \cdots \forall b_l \in \mathrm{dom}(B_l) : \\
&\qquad \forall c_1 \in \mathrm{dom}(C_1) : \cdots \forall c_m \in \mathrm{dom}(C_m) : \\
&\qquad\quad P(A_1 = a_1, \dots, A_k = a_k, B_1 = b_1, \dots, B_l = b_l \mid C_1 = c_1, \dots, C_m = c_m) \\
&\qquad\qquad = P(A_1 = a_1, \dots, A_k = a_k \mid C_1 = c_1, \dots, C_m = c_m) \\
&\qquad\qquad\quad \cdot P(B_1 = b_1, \dots, B_l = b_l \mid C_1 = c_1, \dots, C_m = c_m)
\end{aligned} \tag{1}$$

If a network structure is given, each attribute A_i is assigned a *potential table*, i. e., the set of all conditional distributions, one for each distinct instantiation of the attributes in parents(A_i). The general layout of such a table is shown in figure 1. Each column (like the one shaded in gray) corresponds to one specific parent attributes' instantiation Q_{ij}. Each entry θ_{ijk} is read as

$$P(A_i = a_{ik} \mid \mathrm{parents}(A_i) = Q_{ij}) \;=\; \theta_{ijk}$$

The learning of Bayesian Networks consists of identifying a good candidate graph that encodes the independencies in the database. The goodness of fit is

A_i	Q_{i1}	$\cdots$	Q_{ij}	$\cdots$	Q_{iq_i}
a_{i1}	θ_{i11}	$\cdots$	θ_{ij1}	$\cdots$	θ_{iq_i1}
$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$
a_{ik}	θ_{i1k}	$\cdots$	θ_{ijk}	$\cdots$	θ_{iq_ik}
$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$
a_{ir_i}	θ_{i1r_i}	$\cdots$	θ_{ijr_i}	$\cdots$	$\theta_{iq_ir_i}$

Fig. 1. A general potential table

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 [4, 9, 3].

2.2 Possibilistic Networks

While probabilistic networks like Bayesian Networks are well-suited to handle uncertain information, they lack the ability to cope with imprecision. Imprecision in the application discussed arises when tuples in the database have missing values.

The interpretation of possibility, especially the notion of *degrees of possibility* is based on the context model [8] where possibility distributions are induced by *random sets* [11]. A random set needs a sample space that it is referencing to. In the studied case this will be Ω. Further, a random set defines a family of (neither necessarily disjoint nor nested) subsets $C = \{c_1, \dots, c_m\}$ of Ω, called *contexts*. These contexts are the sample space of a probability space $(C, 2^C, P_\Gamma)$ and are understood as the physical frame conditions under which the contained elements, namely the $\omega \in \Omega$, are considered possible. This family is defined via $\gamma : C \to 2^\Omega$. With these ingredients, the tuple $\Gamma = (\gamma, P)$ constitutes an imperfect description of an unknown state $\omega_0 \in \Omega$. The degree of possibility is then defined as the one-point coverage [11] of Γ, namely:

$$\pi_\Gamma : \Omega \to [0,1] \quad \text{with} \quad \pi_\Gamma(\omega) = P_\Gamma(\{c \in C \mid \omega \in \gamma(c)\})$$

The imperfection named above now incorporates imprecision as well as uncertainty: imprecision enters via the set-valued context definitions while uncertainty is modeled by the probability space over the contexts. Relations and probability distributions can be seen as the two extremes of a possibility distribution: if there is no imprecision, i. e., all contexts contain only one element, a possibility distribution becomes a probability distribution. In contrast to this, when there is only one context $c^\star$ with $\gamma(c^\star) = R \subseteq \Omega$ then for each $\omega \in \Omega$ we have

$$\pi_\Gamma(\omega) = \begin{cases} 1 & \text{if } \omega \in R \\ 0 & \text{otherwise} \end{cases}$$

and thus the uncertainty disappears.

In the interpretation from [2] we can derive a possibility measure Π from the distribution π_Γ in the following way:

$$\Pi : 2^\Omega \to [0,1] \quad \text{with} \quad \Pi(E) = \max_{\omega \in E} P_\Gamma(\{c \in C \mid \omega \in \gamma(c)\})$$

A possibilistic analog for the conditional probabilistic independence constitutes the *possibilistic non-interactivity*[5], which is defined as follows: Let $X = \{A_1, \ldots, A_k\}$, $Y = \{B_1, \ldots, B_l\}$ and $Z = \{C_1, \ldots, C_m\}$ denote three disjoint subsets of attributes, then X and Y are conditionally possibilistically independent given Z, if the following equation holds:

$$\begin{aligned}
&\forall a_1 \in \mathrm{dom}(A_1) : \cdots \forall a_k \in \mathrm{dom}(A_k) : \\
&\quad \forall b_1 \in \mathrm{dom}(B_1) : \cdots \forall b_l \in \mathrm{dom}(B_l) : \\
&\quad\quad \forall c_1 \in \mathrm{dom}(C_1) : \cdots \forall c_m \in \mathrm{dom}(C_m) : \\
&\quad\quad\quad \Pi(A_1 = a_1, \ldots, A_k = a_k, B_1 = b_1, \ldots, B_l = b_l \mid C_1 = c_1, \ldots, C_m = c_m) \\
&\quad\quad\quad\quad = \min\{\Pi(A_1 = a_1, \ldots, A_k = a_k \mid C_1 = c_1, \ldots, C_m = c_m), \\
&\quad\quad\quad\quad\quad\quad \Pi(B_1 = b_1, \ldots, B_l = b_l \mid C_1 = c_1, \ldots, C_m = c_m)\}
\end{aligned} \tag{2}$$

where $\Pi(\cdot \mid \cdot)$ denotes the conditional possibility measure defined as follows:

$$\begin{aligned}
&\Pi(A_1 = a_1, \ldots, A_k = a_k \mid B_1 = b_1, \ldots, B_l = b_l) \\
&\quad = \max\{\pi_\Gamma(\omega) \mid \omega \in \Omega \;\wedge\; \bigwedge_{i=1}^{k} A_i(\omega) = a_i \;\wedge\; \bigwedge_{i=1}^{l} B_i(\omega) = b_i\}
\end{aligned} \tag{3}$$

With these prerequisites a possibilistic network is a decomposition of a multivariate possibility distribution:

$$\forall \tau \in \mathcal{S}_n : \Pi(A_1, \ldots, A_n) = \min_{i=1}^{n} \Pi(X_{\tau(i)} \mid X_{\tau(i-1)}, \ldots, X_{\tau(1)})$$

Learning possibilistic networks follows the same guidelines as the induction of probabilistic networks. Again, a usual learning task consists of two components: a search heuristic and an evaluation measure. Examples for the former are the same as for Bayesian Networks, examples for the latter can be found in [6].

3 The Quantitative Component: Visualization

The result of the network learning task consists of a directed acyclic graph (DAG) representing the observed probabilistic or possibilistic (in)dependencies between the attributes exposed by the database samples. An example is depicted in figure 2.

This graph can be interpreted as the *structural* or *qualitative* or *global* component of such a network. This view is justified since the graph structure describes

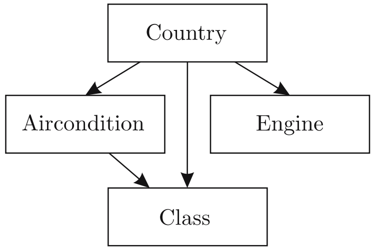


Fig. 2. An example of a probabilistic network

the identified (in)dependencies between the entirety of attributes. The graph allows us to deduce statements like the following:

- Attributes Country and Aircondition have some (statistical) influence on the Class attribute.
- Engine does not seem to have a reasonable impact on the Class attribute. It is merely governed by attribute Country.[2]

Although these statements certainly convey valuable information about the domain under consideration, some questions remain unanswered. Combined into one question, it is desirable to know *which combinations* of the conditioning attributes' *values* have *what kind of impact* on *which class values*? The emphasized words denote the entities that carry much more information about the data volume under analysis. Fortunately, this information is already present in form of the *quantitative* or *local* component of the induced networks, namely the potential tables of the nodes.

Since the goal stated in section 1 was to find concept descriptions based on concepts designated by the class attribute, we only need to consider the class attribute's potential table. Therefore, the actual problem to solve is: How can a potential table (containing either probabilistic or possibilistic values) be represented graphically, incorporating the entities mentioned above?

The remainder of this section will deal with the didactical introduction of a visualization method for probabilistic potential tables. Then, this method will be transferred to the possibilistic case.

Figure 1 shows a general potential table. In the case studied here, the attribute A_i corresponds to the class attribute C. However, we will continue to refer to it as A_i, since we can use the visualization for presenting any attribute's potential table. Each of the q_i columns of the table corresponds to a distinct instantiation of the conditioning attributes. Therefore, the database can be partitioned into q_i disjoint subsets according to these conditioning attributes instantiations. Every fragment, again, is then split according the r_i values of

[2] Since these networks are computationally induced, we refrain from using the notion *causality* here. It is for an expert to decide whether the extracted dependencies carry any causal relationships.

attribute A_i. The relative frequencies of the cardinalities of these resulting sets comprise the entries of the potential table, namely the θ_{ijk}.

We can assign to each table entry θ_{ijk} a set of database samples $\sigma_{ijk} \subseteq \Omega$ which corresponds to all samples having attribute A_i set to a_k and the parent attributes set to the j-th instantiation (out of q_i many). Since we know the entire potential table, we can compute probabilities such as $P(A_i = a_{ik})$ and $P(\text{parents}(A_i) = Q_{ij})$. With these ingredients each table entry θ_{ijk} can be considered an association rule [1]:

$$\texttt{If } \text{parents}(A_i) = Q_{ij} \texttt{ then } A_i = a_{ik} \texttt{ with confidence } \theta_{ijk}\texttt{.}$$

Therefore, all association rule measures like recall, confidence, lift,[3] etc. can be evaluated on each potential table's entry.

With these prerequisites, we are able to depict each table entry as a circle, the color of which depends on the class variable. As an example we consider the class attribute C to have two parent attributes A and B. All three attributes are binary. The domain of the class attribute will be assigned the following colors: $\{c_1, c_2\} = \{\circ, \bullet\}$. The (intermediate) result is shown in figure 3(a). In the next step (figure 3(b)) we enlarge the datapoints to occupy an area that corresponds to the absolute number of database samples represented, i. e., $|\sigma_{ijk}|$.

Finally, each datapoint has to be located at some coordinate (x, y). For this example we choose

$$x = \text{recall}(\sigma_{ijk}) \quad \text{and} \quad y = \text{lift}(\sigma_{ijk})$$

The result is shown in figure 3(c). A data analysis expert can now examine the chart and extract valuable information easily in the following ways: At first, since he is likely to be interested only in sample descriptions belonging to one specific class (e.g. class=failure), his focus is put on the black (filled) circles in the diagram. If he is interested in highly conspicuous subsets of sample cases, the circles at the very top are auspicious candidates since they possess a high lift. Put briefly, the rule of thumb for an expert may read:

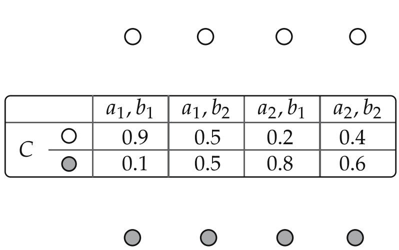
> "Large circles in the upper right corner are promising candidate subsets of samples that could most likely yield a good concept description."

An example with meaningful attributes is postponed to section 4. For the remainder of this section, we will discuss the applicability of the presented visualization that was based on probabilistic values and measures to the possibilistic domain.

---

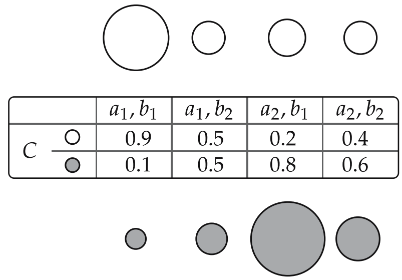
[3] These measures are defined as follows: $\forall \theta_{ijk}$:

$$\begin{aligned}
\text{recall}(\sigma_{ijk}) &= P(\text{parents}(A_i) = Q_{ij} \mid A_i = a_{ik}) \\
\text{conf}(\sigma_{ijk}) &= P(A_i = a_{ik} \mid \text{parents}(A_i) = Q_{ij}) = \theta_{ijk} \\
\text{lift}(\sigma_{ijk}) &= \frac{\text{conf}(\sigma_{ijk})}{P(A_i = a_{ik})}
\end{aligned}$$



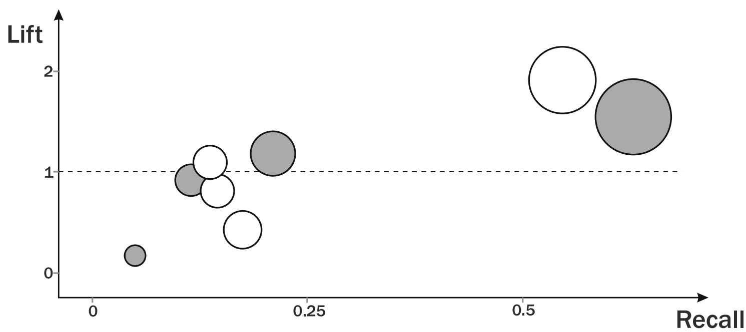
		a_1, b_1	a_1, b_2	a_2, b_1	a_2, b_2
C	○	0.9	0.5	0.2	0.4
	●	0.1	0.5	0.8	0.6

(a) Each entry is assigned a datapoint σ, the color designating the class value.



		a_1, b_1	a_1, b_2	a_2, b_1	a_2, b_2
C	○	0.9	0.5	0.2	0.4
	●	0.1	0.5	0.8	0.6

(b) The size (area) of each datapoint corresponds to the absolute number of samples described by the corresponding table entry.



(c) The location of the center of each datapoint σ is set to the coordinates $(x, y) = (\text{recall}(\sigma), \text{lift}(\sigma))$.

Fig. 3. We assume the class attribute C to have the two parent (conditioning) attributes A and B. All three attributes are binary with the respective domains $\{a_1, a_2\}$, $\{b_1, b_2\}$ and $\{c_1 = \circ, c_2 = \bullet\}$.

3.1 The Possibilistic Case

The above-mentioned circles are serving as visual clues for subsets of samples and were located at coordinates which are computed by probabilistic (association rule) measures. Of course, these measures can be mathematically carried over to the possibilistic setting. However, we have to check whether the semantics behind these measures remain the same. For the following considerations, we assume the following abbreviations for the sake of brevity: A subset of sample cases σ is defined by the class value a_{ik} and the instantiation of the parent attributes Q_{ij}:

$$\sigma = (Q_{ij}, a_{ik}) \stackrel{\text{Abbrev}}{=} (A, c)$$

Since the definition of the conditional possibility is symmetric, i. e., $\forall A, B$: $\Pi(A \mid B) = \Pi(B \mid A) = \Pi(A, B)$, the definitions for recall, confidence and support would coincide. Therefore, we define them as follows:

$$\mathrm{supp}^{\mathrm{poss}}(\sigma) = \Pi(A,c) \qquad \mathrm{recall}^{\mathrm{poss}}(\sigma) = \frac{\Pi(A,c)}{\Pi(c)}$$
$$\mathrm{conf}^{\mathrm{poss}}(\sigma) = \frac{\Pi(A,c)}{\Pi(A)} \qquad \mathrm{lift}^{\mathrm{poss}}(\sigma) = \frac{\Pi(A,c)}{\Pi(A)\Pi(c)}$$

The justification for this type of definition is as follows: As the degree of possibility for any tuple t, we assign the total probability mass of all contexts that contain t [7]. With this interpretation, the term $\Pi(A = a)$ refers to the maximum degree of possibility of all sets of tuples, for which $A(t) = a$ holds, i. e., $\Pi(A = a) = \max\{p(t) = \frac{w(t)}{N} \mid t \in \Omega \wedge A(t) = a\}$. This probabilistic origin allows us to look at the possibility of an event E (i. e., a set of tuples) as an upper bound of elementary events' probablitities contained in E [2].

4 Application and Results

For testing purposes, we firstly created an artificial dataset where some conspicuity was manually put into the data in order to verify whether these dependencies were found and, most importantly, whether these peculiarties become obvious in the visualization. Then, of course, the presented technique was evaluated on real-life data the (anonymized) results of which we will present as well.

4.1 Manually-Crafted Dataset

The artificial dataset was generated by a fictitious probabilistic model the qualitative structure of which is shown in figure 2. The conspicuity to be found was that a single aircondition type had a higher failure rate in two specific counties, whereas this type of aircondition accounted for the smallest proportion of all airconditions.

As learning algorithm we used the well-known K2 algorithm [4] with the K2 metric as evaluation measure. Note that this example visualizes the potential tables of a Bayesian Network (the one shown in figure 2), i. e., it represents probabilistic values.

Figure 4 shows all sets of sample cases that are marked defective by the class attribute. Since in this artificial model both attributes Aircondition and Country have a domain of five values each, there are 25 different parent instantiations and thus 25 circles in the chart. As one can cleary see, there are two circles standing out significantly. Because we chose the lift to be plotted against the y-axis, these two sets of sample cases expose a high lift value, stating that the respective parent instantiations (here: combination of Country and Aircondition) make the failure much more probable. Since both circles account for only a small portion of all tuples in the database, they have small recall, indicated by being located at the left side of the chart.

4.2 Real-Life Dataset

The real-life application which produced empirical evidence that the presented visualization method greatly enhances the data analysis process took place during a cooperative project at the DaimlerChrysler Research Center. As a

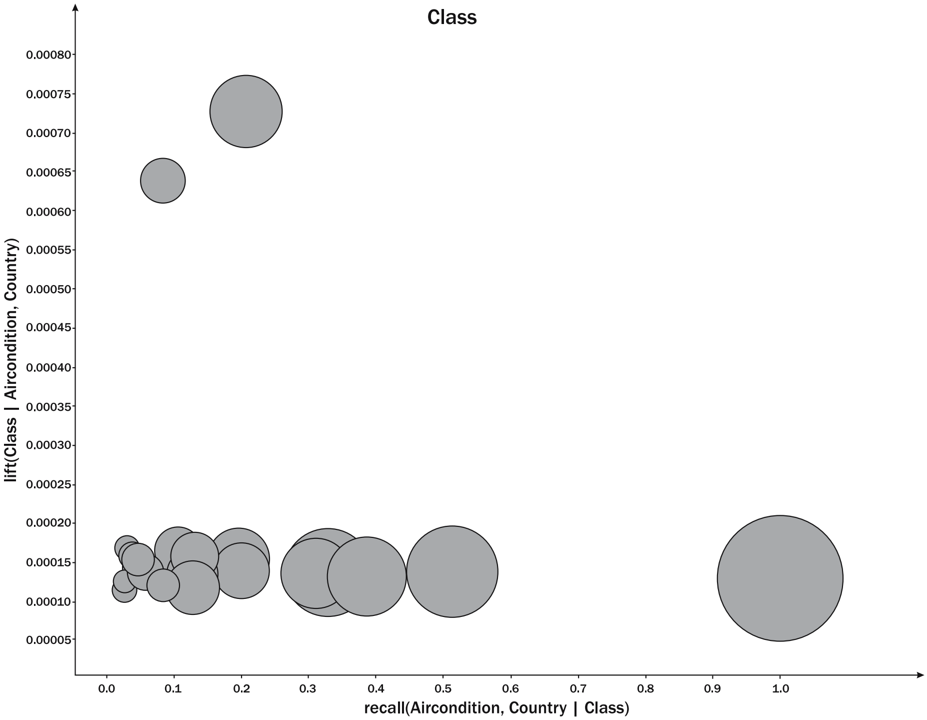


Fig. 4. The two outstanding circles at the top of the chart indicate two distinct sets of samples having a much higher failing rate than the others. They reveal the two intentionally incorporated dependencies, i. e., one specific type of arcondition is failing more often in two specific countries.

leading manufacturer of high-quality automobiles, one of the company's crucial concerns is to maintain the high level of quality and reliability of their products. This is accomplished by collecting extensively detailed information about every car sold and to analyze complaints in order to track down the fault promptly. Since these data volumes are highly confidential, we are not allowed to present specific attribute names and background information. Nonetheless, the charts generated by visualizing the induced possibilistic networks will provide a fairly good insight into the everyday usage of the presented visualization method.

Figure 5 shows a possibilistic chart of the binary class variable. In this case, the non-faulty datasets are depicted as well (unfilled circles). As one can easily see, we find a relatively large circle in the upper right corner. The size of this circle tells that it represents a reasonable number of affected cars, while the high lift states, that the selected parent instantiation should be subject of a precise investigation. In fact, the consultation of a production process expert indeed revealed a causal relationship.

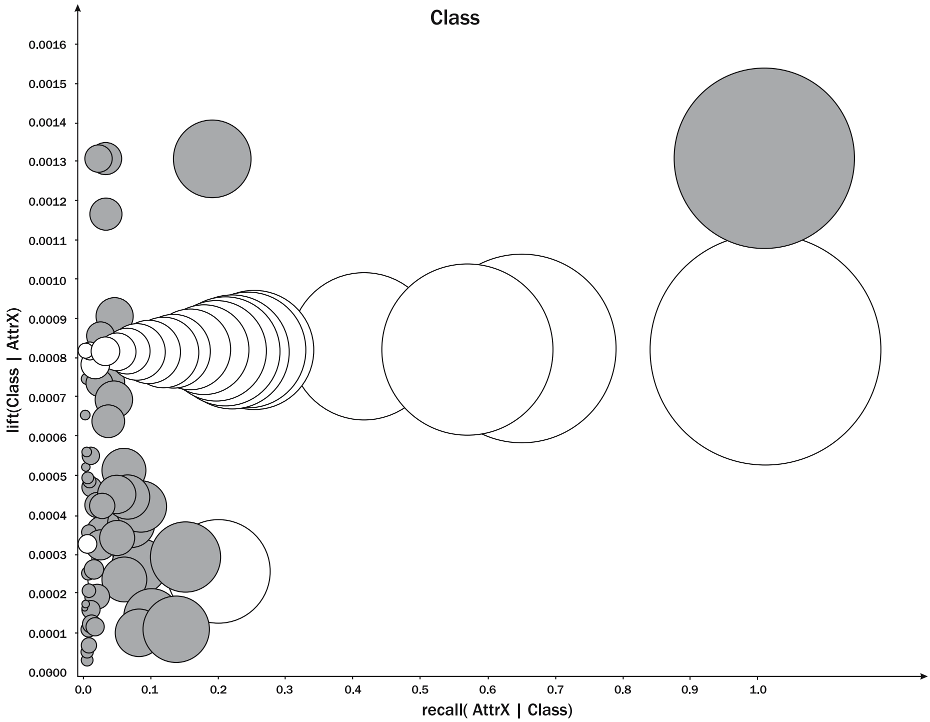


Fig. 5. The large circle in the top right corner indicates a set of vehicles whose specific parents attributes' values lead to a higher failure rate. An investigation by experts revealed a real causal relationship.

4.3 Practical Issues on the Visualization

As it can be seen from figure 5 and 4, the circles show a fairly large overlap which may lead to large circles covering and thus hiding smaller ones. In the real-world application — from which the figures are taken — there are several means of increasing the readability of the charts. On the one hand, all circles can be scaled to occupy less space while the user can zoom into a smaller range of the plot. Further, the circles can be made transparent which reveals accidentally hidden circles.

5 Conclusion and Future Work

In this chapter, we presented a brief introduction to both probabilistic and possibilistic networks, the latter due to its natural ability of handling imprecise data becoming increasingly interesting for industrial applications since real-world data often contains missing data. We argued further that the learning of such a network only reveals the qualitative part of the contained dependencies, yet the more meaningful information being contained inside the potential tables, i. e., the quantitative part of the network. Then, a new visualization technique was presented that is capable of displaying high-dimensional, nominal potential tables containing probabilistic as well as possibilistic parameters. This plotting method was evaluated in an industrial setting enabling production experts to easier identify extreme data samples.

Since the presented technique only dealt with datasets that represented the current state of the database at a specific (but fixed) moment in time, it would be interesting to extend the visualization to temporal aspects, that is, time series. Then, it would be possible not only to use the mentioned association rule measures but also their derivatives in time to make trends visible.

References

1. Agrawal, R., Imielinski, T., Swami, A.: Mining association rules between sets of items in large databases. In: Proc. of the ACM SIGMOD Conference on Management of Data, pp. 207–216 (1993)
2. Borgelt, C.: Data Mining with Graphical Models. PhD Thesis, Otto-v.-Guericke-Universität Magdeburg, Germany (2000)
3. Borgelt, C., Kruse, R.: Some experimental results on learning probabilistic and possibilistic networks with different evaluation measures. In: ECSQARU/FAPR 1997. Proc. of the 1st International Joint Conference on Qualitative and Quantitative Practical Reasoning, pp. 71–85 (1997)
4. Cooper, G., Herskovits, E.: A Bayesian method for the induction of probabilistic networks from data. Journal of Machine Learning (1992)
5. Dubois, D., Prade, H.: Possibility theory. Plenum Press, New York (1988)
6. Gebhardt, J., Kruse, R.: Learning possibilistic networks from data. In: Proc. 5th Int. Workshop on Artificial Intelligence and Statistics, pp. 233–244 (1995)
7. Gebhardt, J., Kruse, R.: A possibilistic interpretation of fuzzy sets by the context model. In: IEEE International Conference on Fuzzy Systems, pp. 1089–1096 (1992)
8. Gebhardt, J., Kruse, R.: Int. Journal of Approximate Reasoning 9, 283–314 (1993)
9. Heckerman, D., Geiger, D., Maxwell, D.: Learning Bayesian networks: The combination of knowledge and statistical data. Technical Report MSR-TR-94-09 85–96, Microsoft Research, Advanced Technology Division, Redmond, WA (1994)
10. Lauritzen, S., Spiegelhalter, D.: Journal of the Royal Statistical Society. Series B 2(50), 157–224 (1988)
11. Nguyen, H.: Information Science 34, 265–274 (1984)
12. Pearl, J.: Probabilistic reasoning in intelligent systems: Networks of plausible inference. Morgan Kaufmann, San Mateo, California (1988)