Visualization of Possibilistic Potentials

Matthias Steinbrecher and Rudolf Kruse

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

Abstract. The constantly increasing capabilities of database storage systems leads to an incremental collection of data by business organizations. The research area of Data Mining has become a paramount requirement in order to cope with the acquired information by locating and extracting patterns from these data volumes. Possibilistic networks comprise one prominent Data Mining technique that is capable of encoding dependence and independence relations between variables as well as dealing with imprecision. It will be argued that the learning of the network structure only provides an overview of the qualitative component, yet the more interesting information is contained inside the network parameters, namely the potential tables. In this paper we introduce a new visualization technique that allows for a detailed inspection of the quantitative component of possibilistic networks.

1 Introduction

The ongoing advance in the development of database systems enables today's business organizations to acquire and store huge amounts of data. However, the more data are collected, the stronger is the requirement for sophisticated analyzation methods to extract hidden patterns. The research area of Data Mining addresses these tasks and offers intelligent data analysis techniques such as classification, prediction or concept description, just to name a few.

The latter technique of concept description tries to identify 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.

P. Melin et al. (Eds.): IFSA 2007, LNAI 4529, pp. 295–303, 2007.

In this paper, we assume the database entries to have nominal attributes with one distinguished attribute designating the class of each data sample. We will use 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.

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. Finally, a new visualization method for these potential tables is presented and evaluated on real-world data.

The remainder of this paper is structured as follows: Section 2 presents a brief review of possibilistic networks. 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 paper concludes with section 5, giving an outlook of intended further investigations.

2 Background

A database D, interpreted as a table, shall contain a certain number of tuples (rows) t_h ($1 \leq h \leq N$), each of which exposes a fixed number of attributes (columns) $\{A_1, \ldots, A_n\}$ with respective domains $\mathrm{dom}(A_i) = \{a_{i1}, \ldots, a_{ir_i}\}$, i. e. $|\mathrm{dom}(A_i)| = r_i$. We allow D to contain multiple identical tuples which is modeled by a weight function $w : D \to \mathbb{N}^+$ that assigns to each distinct tuple $t \in D$ the number of occurrences in D.

In the case of precise tuples, each cell of this table contains exactly one attribute value, i. e. each tuple t assumes one distinct value a_{ik} for each attribute A_i: $\forall t \in D : A_i(t) = a_{ik}$, $i = 1, \ldots, n$, $1 \leq k \leq r_i$. From such a database (or relation) a joint probability distribution can be estimated for each tuple: $\forall t \in D : p(t) = \frac{w(t)}{N}$. Each attribute can be seen as a random variable:

$$P(A_i = a_{ik}) = \frac{|\{t \in D \mid A_i(t) = a_{ik}\}|}{N}, \quad i = 1, \ldots, n, \quad k = 1, \ldots, r_i$$

Imprecision now enters through the absence of some of these table entries, i. e. there are tuples that have one or more values missing. Since we do not know the specific value of such cells (usually designated by a '?' or '$*$' in the dataset) we have to take into consideration *all possible* values of the corresponding attribute. Thus, the absence of a specific value of attribute A of tuple t is modeled as $A(t) = \mathrm{dom}(A)$. Of course, this approach can be used as well to model partial ignorance, i. e. we can allow the attribute A to assume any subset of $\mathrm{dom}(A)$. Let us consider the imprecise database depicted in table 1. The first column shows the tuple as it may appear in a data file, the second and third column depict the values of the binary attributes A and B, respectively.

Formally, we allow each attribute A_i to be a *random set* [1], rather than a random variable. Let Ω be the finite set of all possible precise tuples over the Cartesian product of all attributes' domains, i. e. $\Omega = \times_{i=1}^{n} \mathrm{dom}(A_i)$. Then, we can define a mapping $\gamma : D \to 2^{\Omega}$ that assigns to each (possibly imprecise)

Table 1. An imprecise example table. Note, that tuples t_3, t_4 and t_5, t_6 are identical

	A	B	$\gamma(t_i)$
$t_1 = (a_1, *)$	$\{a_1\}$	$\{b_1, b_2\}$	$\{(a_1, b_1), (a_1, b_2)\}$
$t_2 = (a_1, b_2)$	$\{a_1\}$	$\{b_2\}$	$\{(a_1, b_2)\}$
$t_3 = (a_1, b_1)$	$\{a_1\}$	$\{b_1\}$	$\{(a_1, b_1)\}$
$t_4 = (a_1, b_1)$	$\{a_1\}$	$\{b_1\}$	$\{(a_1, b_1)\}$
$t_5 = (*, b_2)$	$\{a_1, a_2\}$	$\{b_2\}$	$\{(a_1, b_2), (a_2, b_2)\}$
$t_6 = (*, b_2)$	$\{a_1, a_2\}$	$\{b_2\}$	$\{(a_1, b_2), (a_2, b_2)\}$
$t_7 = (a_2, b_2)$	$\{a_2\}$	$\{b_2\}$	$\{(a_2, b_2)\}$
$t_8 = (*, *)$	$\{a_1, a_2\}$	$\{b_1, b_2\}$	$\{(a_1, b_1), (a_1, b_2), (a_2, b_1), (a_2, b_2)\}$

tuple $t \in D$ all (definitely precise) tuples $\omega \in \Omega$ that are covered by t. These sets are shown in the fourth column of table 1.

With this interpretation, each tuple $t \in D$ can be considered a *context*. The example contexts are shown in figure 1. A precise tuple obviously only describes a context that contains itself. Note, that due to the presence of multiple identical tuples ($t_3 \equiv t_4$ and $t_5 \equiv t_6$), we obtain identical contexts as well. The degree of possibility of any precise tuple $\omega \in \Omega$ is the probability of the set of contexts that contain ω:

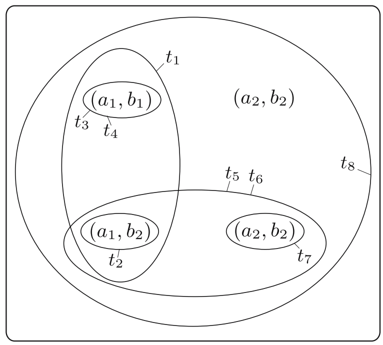

Fig. 1. The contexts induced by table 1

$$\pi_D : \Omega \to [0, 1] \quad \text{with}$$
$$\pi_D(\omega) = P_D(\{t \in D \mid \omega \in \gamma(t)\})$$

This coincides with the *one-point coverage* [2] of ω under D. The probability function P_D belongs to the random set and is part of the probability space $(D, 2^D, P_D)$, where in our study each tuple $t \in D$ has the same elementary probability $p(t) = \frac{1}{N}$. In the interpretation from [3] we can derive a possibility measure Π from the distribution π_D in the following way:

$$\Pi : 2^\Omega \to [0, 1] \quad \text{with} \quad \Pi(E) = \max_{\omega \in E} P_D(\{t \in D \mid \omega \in \gamma(t)\})$$

2.1 Possibilistic Networks

Even though the database D will be much smaller than Ω in practice, we need methods to further reduce the size of the joint possibility distribution induced by D. One idea is to exploit certain independency conditions within π_D such as the *possibilistic non-interactivity*, 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:

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

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

$$
\begin{aligned}
&\Pi(A_1 = a_1, \ldots, A_k = a_k \mid B_1 = b_1, \ldots, B_l = b_l) \\
&\quad = \max\{\pi_D(\omega) \mid \omega \in \Omega \ \wedge \ \bigwedge_{i=1}^{k} A_i(\omega) = a_i \ \wedge \ \bigwedge_{i=1}^{l} B_i(\omega) = b_i\}
\end{aligned} \tag{2}
$$

The graph nodes coincide with the attributes. Let parents(A) denote the set of all nodes that have an edge pointing to node A. With these prerequisites we can use a *directed acyclic graph (DAG)* to encode such independencies in the following way: Given an instatiation of the attributes in parents(A), attribute A is conditional independent of the remaining attributes. Such a DAG is said to carry the *structural* or *global* or *qualitative* information of a possibilistic network.

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 2. Each column (like the one shaded in gray) corresponds to one specific parent attribute instantiation Q_{ij}. Each entry θ_{ijk} is read as

A_i	Q_{i1}	$\cdots$	Q_{ij}	$\cdots$	Q_{iq_i}
a_{i1}	θ_{i11}	$\cdots$	θ_{ij1}	$\cdots$	$\theta_{iq_i 1}$
$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$
a_{ik}	θ_{i1k}	$\cdots$	θ_{ijk}	$\cdots$	$\theta_{iq_i k}$
$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$
a_{ir_i}	θ_{i1r_i}	$\cdots$	θ_{ijr_i}	$\cdots$	$\theta_{iq_i r_i}$

Fig. 2. A general potential table

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

These conditional distributions encode the *parametrical* or *local* or *quantitative* component of the network. The usual learning task of a possibilistic network consists of two components: a search heuristic and an evaluation measure. Examples for the former can be found in [4,5,6], examples for the latter are studied in [7].

3 Visualization of Potential Tables

After the learning task for a possibilistic network is completed, we are given a DAG that is encoding the detected (in)dependencies in the above-mentioned manner. A sample network is depicted in figure 3.

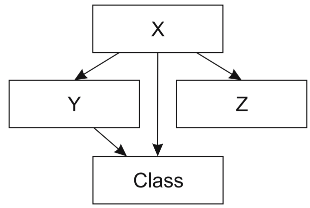

Fig. 3. A possibilistic network example

Since we are interested in the impact that certain attribute (values) have on the class attribute, we concentrate our attention on the direct ancestors of the class node, i. e., its parent nodes.

Although such a network conveys valuable information about the underlying data, some important questions remain unanswered. Cut short, it is desirable to know *which combinations* of the conditioning attributes' *values* have *what kind of impact* on *which class values*? The emphasized words in the last sentence mark the entities that carry much more information about the database under consideration. We can use the potential tables — or more specific: the class attribute's potential table — to extract the demanded information. Thus, the goal is to find an intuitive way of representing a potential table graphically, incorporating the entities mentioned above.

In order to represent the entries of a potential table in a chart, we investigate the semantics of these values a little bit further. A value θ_{ijk} tells us that given the j-th instantiation of the parent nodes of attribute A_i, then it is possible to a degree of θ_{ijk} that the attribute A_i assumes the i-th value of its domain.

In a probabilistic setting, i. e., if we dealt with Bayesian Networks [8,9], the values θ_{ijk} would designate probabilities in the following way:

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

For the next considerations, we assume the following abbreviations for the sake of brevity: A subset of sample cases σ_{ijk} is defined by the class value a_{ik} and the instantiation of the parent attributes Q_{ij}: $\sigma_{ijk} = (Q_{ij}, a_{ik}) := (A, c)$. With this interpretation, each σ_{ijk} represents an association rule [10]:

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

For each probabilistic entry θ_{ijk} we would compute three different values of evaluation measures from the domain of association rules,[1] e. g.:

$$m_x = \text{recall}(\sigma_{ijk}), \qquad m_y = \text{lift}(\sigma_{ijk}), \qquad m_z = \text{supp}(\sigma_{ijk})$$

with

$$\text{supp}(\sigma_{ijk}) = P(\text{parents}(A_i) = Q_{ij}, A_i = a_{ik})$$
$$\text{recall}(\sigma_{ijk}) = \frac{\text{supp}(\sigma_{ijk})}{P(A_i = a_{ik})}$$
$$\text{lift}(\sigma_{ijk}) = \frac{\theta_{ijk}}{P(A_i = a_{ik})}$$

[1] For a detailed analysis of such measures we refer to [11].

Finally, we display each σ_{ijk} as a circle of size m_z and locate it at position (m_x, m_y) in a two-dimensional chart.

Since we intend to visualize possibilistic values, we interpret the σ_{ijk} as possibilistic association rules where the value of θ_{ijk} represent the degree of possibility rather than the confidence. The presented measures are transformed into their possibilistic counterparts. Of course, we have to check whether the semantics behind these measures remain intact. 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 [12]. 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 [3].

4 Experiments and Evaluation

The visualization technique presented here was introduced during a data mining project in cooperation with an automobile manufacturer. To justify the practical applicability, we intend to present real-world results. Since the underlying datasets are highly confidential, we are not allowed to show any attribute names or values. However, the charts will give a good insight, how suspiciuous subsets of tuples can be identified.

The dataset under analysis contained approximately 50.000 vehicle descriptions, including one class attribute designating, whether the respective car was faulty or not. A network was learned which revealed the class attribute to have two parent nodes, anonymized to X and Y. We then chose the three evaluation measures to be *recall*, *lift* and *support*, which resulted in the chart depicted in figure 4. Choosing these measures, a user can apply the following heuristic to identify possible conspicuous tuples:

"Large circles in the upper right corner are promising candidate subsets of samples that could most likely be suspicuous."

Dark circles represent faulty vehicles, white circles non-faulty ones. In figure 4 we identify a large gray circle on the right side, labeled '1'. The corresponding attribute values for X and Y belonging to this circle were identified by experts as having a causal effect on these 800 faulty vehicles.

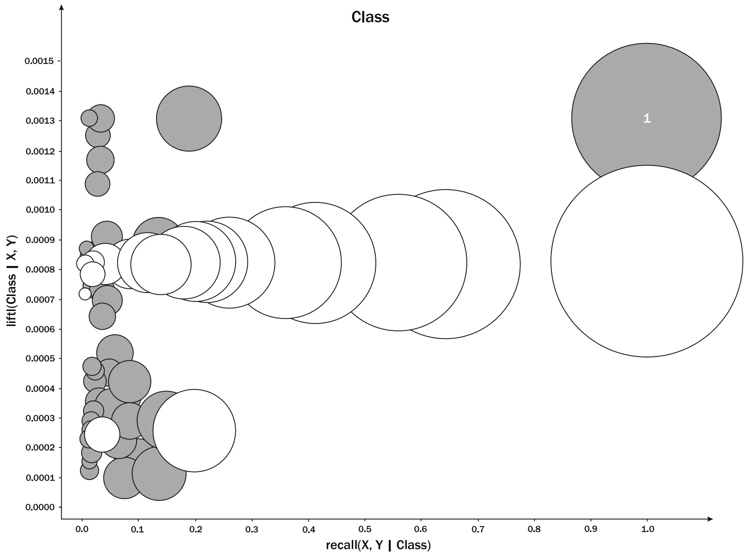

Fig. 4. The circle with label '1' represents 800 tuples having a large lift and recall

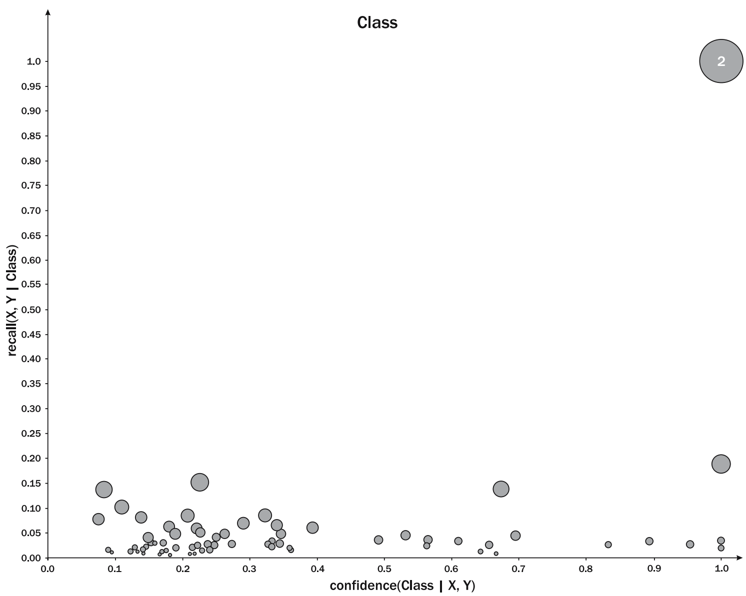

Fig. 5. Only faulty vehicles sets are shown. Again, the circle labeled '2' represents eye-catching tuples, that had a causal relationship with the corresponding values for X and Y.

Another way of looking at this dataset is by choosing different measures for the x- and y-axis. Motivated by charts from the domain of *Information Retrieval*, we assign the measures *confidence* and *recall* to the x- and y-axis, respectively. Omitting the non-faulty circles for clarity results in the chart of figure 5.

Again, the circle marked as ‘2’ represents (the same) 800 faulty vehicles.

5 Conclusion and Future Work

In this paper, we gave a short introduction to possibilistic networks and its ability of handling imprecise data which is becoming more and more a requirement for industrial applications since real-world data often contains missing data. We argued further that the more interesting information is contained inside the quantitative part of a network, namely its potential tables. Then, a new visualization technique was presented that is capable of displaying high-dimensional, nominal potential tables containing possibilistic parameters. This plotting method was evaluated in an industrial setting which produced empirical evidence that the presented visualization method greatly enhances the exploratory data analysis process. Since the presented visualization method aids to find concept descriptions and combines identical tuples (w. r. t. a subset of attributes) it may be promising to try to apply a modified version of this technique in the area of Text Mining, where several documents (again, identical w. r. t. some attribute, e. g. topic or keywords) may be grouped and displayed against other document groups.

References

1. Nguyen, H.T.: Using Random Sets. Information Science 34 (1984) 265–274
2. Gebhardt, J., Kruse, R.: The Context Model Ů An Integrating View of Vagueness and Uncertainty. Int. Journal of Approximate Reasoning 9 (1993) 283–314
3. Borgelt, C., Kruse, R.: Graphical Models - Methods for Data Analysis and Mining. John Wiley & Sons, United Kingdom (2002)
4. Cooper, G.F., Herskovits, E.: A Bayesian Method for the Induction of Probabilistic Networks from Data (1992)
5. 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.
6. 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 (ECSQARU/FAPR’97), Bad Honnef, Germany (1997) 71–85
7. Gebhardt, J., Kruse, R.: Learning Possibilistic Networks from Data (1995)
8. Pearl, J., Russel, S.: Bayesian Networks (1994)
9. Heckerman, D.: A Tutorial on Learning with Bayesian Networks. Technical Report MSR-TR-95-06, Microsoft Research, Advanced Technology Division, Redmond, WA (1995) Revised June 96.

10. Agrawal, R., Imielinski, T., Swami, A.: Mining Association Rules Between Sets of Items in Large Databases (1993).
11. Yao, Y.Y., Zhong, N.: An Analysis of Quantitative Measures Associated with Rules. In: Methodologies for Knowledge Discovery and Data Mining, Springer-Verlag Berlin (1999).
12. Gebhardt, J., Kruse, R.: A Possibilistic Interpretation of Fuzzy Sets by the Context Model. In: IEEE International Conference on Fuzzy Systems, San Diego (1992) 1089–1096.