Visualization of Statistical Information in Concept Lattice Diagrams

Jana Klimpke[0000−0001−5074−3089] and Sebastian Rudolph[0000−0002−1609−2080]

Computational Logic Group, TU Dresden, Dresden, Germany

Abstract. We propose a method of visualizing statistical information in concept lattice diagrams. To this end, we examine the characteristics of support, confidence, and lift, which are parameters used in association analysis. Based on our findings, we develop the notion of *cascading line diagrams*, a visualization method that combines the properties of additive line diagrams with association analysis. In such diagrams, one can read the size of a concept's extent from the height of the corresponding node in the diagram and, at the same time, the geometry of the formed quadrangles illustrates whether two attributes are statistically independent or dependent and whether they are negatively or positively correlated. In order to demonstrate this visualization method, we have developed a program generating such diagrams.

1 Introduction

Formal concept analysis (FCA) is a mathematical approach for analyzing conceptual hierarchies arising from relationships between objects and attributes. By means of an order relation, hierarchically grouped sets of entities can be sorted by set inclusion and later visualized by means of *line diagrams*, from which *qualitative*, crisp dependencies between the examined attributes, called *implications*, can be read off easily.

Association analysis is a data mining technique, used to discover and evaluate *quantitative* relationships and dependencies in a data set. It offers ways of characterizing the strength of these relationships, using the statistical measures *support*, *confidence* and *lift*.

The use of association analysis to extract "imperfect implications" – referred to as *association rules* – has been explored widely in data mining, but also specifically in FCA [11,8,9,3]. Unfortunately, the same cannot be said about the visualization part, i.e., representation of statistical information in line diagrams. One 2004 paper introduced a method of lattice drawing where concepts were placed at positions that were related to their support [10]. However, the authors stated that this method sometimes created nearly horizontal lines in the diagram. To address this problem, they introduced a spring-based lattice drawing method in a follow-up paper [7]. Additionally, the authors proposed to generate a lattice diagram in $\mathbb{R}^3$ and then allow the user to find a "best" projection into $\mathbb{R}^2$ by rotating the lattice around a central axis.

Our work aims at an approach for visualizing concept lattice diagrams in a way that, on top of displaying all perfect relationships between attributes, also reflects the frequencies of the depicted concepts as well as correlation strength between attributes. To this end, we propose *cascading line diagrams*, realized by means of a positioning rule that is inspired by the notion of additive line diagrams but adjusts the height of the concept nodes according to their extent's cardinality. By choosing a logarithmic scale for the latter, statistic independence between attributes manifests itself in perfect parallelograms, whereas positive and negative correlations lead to obtuse or acute deviations from this parallelogram shape. We present an open-source prototypical implementation for drawing cascading line diagrams, which also allows for an intuitive interactive adjustment of its parameters along the remaining degrees of freedom.

2 Preliminaries

We start by briefly introducing the basic notions of FCA [6]. A *formal context* is a triple (G, M, I), consisting of a set G of *objects*, a set M of *attributes* as well as a binary *incidence relation* $I \subseteq G \times M$ between G and M. As usual, the fact that I relates an object g to an attribute m will be written as $(g, m) \in I$.

The set of all object-attribute relationships of a formal context can be written down by means of an incidence matrix. An example of such a matrix, also known as cross table, can be seen in Table 1.
The set of attributes, shared by a set $A \subseteq G$ of objects can be derived by

Table 1. Cross table displaying the relation between objects and attributes of a given formal context.

$G \setminus M$	a	b	c	d	e
T1		×		×	
T2		×			×
T3			×		
T4	×	×	×		
T5				×	
T6		×	×		
T7					×

$$A' := \{m \in M \mid \forall\, g \in A : (g, m) \in I\}. \quad (1)$$

Dually, the set of all objects that have each of the attributes in a set $B \subseteq M$ can be obtained by

$$B' := \{g \in G \mid \forall\, m \in B : (g, m) \in I\}. \quad (2)$$

A pair $(A,\ B)$, with $A \subseteq G$ and $B \subseteq M$ will be called a *formal concept* of a context (G, M, I) if $A' = B$ and $B' = A$. The set A will be called the *extent* and the set B the *intent* of the concept. The set of all formal concepts of a formal context (G, M, I) is denoted by $\mathfrak{B}(G, M, I)$.

Concepts can be ordered hierarchically by using the order relation $\leq$. For the concepts (A_1, B_1) and (A_2, B_2) one lets $(A_1, B_1) \leq (A_2, B_2)$ iff $A_1 \subseteq A_2$. We call (A_1, B_1) a *lower neighbor* of (A_2, B_2) – and write $(A_1, B_1) \prec (A_2, B_2)$ – whenever both $(A_1, B_1) \leq (A_2, B_2)$ and there is no "intermediate concept" $(\dot{A}, \dot{B})$ satisfying $(A_1, B_1) \leq (\dot{A}, \dot{B}) \leq (A_2, B_2)$. It turns out the set of a context's formal concepts together with $\leq$ is not only an ordered set, but even a complete lattice, called the *concept lattice* of the context. As for partial orders in general, it is common to visualize concept lattices by means of *line diagrams*, where each node represents one concept of the concept lattice and, for all $(A_1, B_1) \prec (A_2, B_2)$ there is an ascending straight line connecting the node representing (A_1, B_1) to the node representing (A_2, B_2). While this requirement puts some constraints on the vertical positioning of the nodes in a lattice diagram, there is still a lot of leeway and it is a non-trivial question how to arrive at a "good" diagram.

One approach to obtain particularly well-readable diagrams, called *additive line diagrams*, will be explained in the following.

2.1 Additive Line Diagrams

An attribute $m \in M$ is called *irreducible* if there is no set $X \subseteq M$ of attributes with $m \notin X$ so that $\{m\}' = X'$. The set of all irreducible attributes is denoted by M_{irr}. The set $irr(A, B)$ of all irreducible attributes of a concept is defined by $B \cap M_{irr}$.
An additive line diagram is obtained based on a function mapping each irreducible attribute to a two-dimensional vector according to Formula (3). The position in the plane at which the node representing concept (A, B) should be drawn is determined according to Formula (4) [6].

$$vec : M_{irr} \to \mathbb{R} \times \mathbb{R}_{<0} \tag{3}$$

$$pos(A, B) := \sum_{m \in irr(A,B)} vec(m) \tag{4}$$

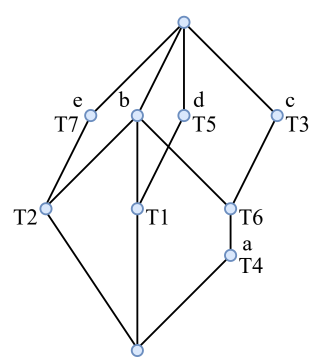


Fig. 1. Example of an additive line diagram based on the context given in Table 1.

The resulting diagrams are characterized by many parallel lines, which makes them easier to read than most other diagrams. An example is displayed in Fig. 1.

An *implication* is a rule of the form $X \to Y$, where X is the body and Y is the head of the rule. X and Y are sets of attributes. Given a formal context an implication $X \to Y$ is valid iff $Y \subseteq X''$. It is not too difficult and a standard exercise in FCA to directly read implications from a concept lattice diagram.

2.2 Association Analysis

Association rule learning is a data mining method used to discover and evaluate relationships and dependencies in databases. It is used, among others, in shopping cart analysis. Put into FCA terminology, the goal of association analysis is to find connections between individual attributes in order to be able to make statements about which of them often co-occur together in objects. For this purpose, one uses the statistical characteristics *support*, *confidence* and *lift* [4,5], which are explained in this section together with their relationship to FCA.

The basis of association analysis are *association rules*, which have the form $X \Rightarrow Y$. X and Y represent disjoint and real subsets of the set of all attributes M. An object fulfills such a rule if it has all the attributes, which occur in X and Y.

Support. The support of an attribute set X describes the relative frequency of their joint occurrence in the data (that is: the context) and is calculated by

$$supp(X) := \frac{|\{g \in G \mid X \subseteq \{g\}'\}|}{|G|} = \frac{|X'|}{|G|}. \tag{5}$$

In the same way, the support of an association rule $X \Rightarrow Y$ describes the frequency with which the union $X \cup Y$ occurs in the data, that is

$$supp(X \Rightarrow Y) := supp(X \cup Y) = \frac{|(X \cup Y)'|}{|G|}. \tag{6}$$

Confidence. The confidence is a measure of how often a rule is fulfilled in relation to X. It is calculated by

$$conf(X \Rightarrow Y) := \frac{supp(X \Rightarrow Y)}{supp(X)}. \tag{7}$$

Lift. Since the confidence does not refer to the frequency with which the head of the formula occurs “normally”, it cannot be used to make a statement about how strong the body of an association rule really “promotes” the head. In order to account for this fact, the lift is used, which is calculated by

$$lift(X \Rightarrow Y) := \frac{conf(X \Rightarrow Y)}{supp(Y)} = \frac{supp(X \cup Y)}{supp(X) \cdot supp(Y)}. \tag{8}$$

As is obvious from that formula, $lift(X \Rightarrow Y) = lift(Y \Rightarrow X)$ always holds and the corresponding value provides information regarding the correlation between attribute occurrences X and Y. The following correspondences apply:

- $lift(X \Rightarrow Y) > 1$: X and Y positively correlated
- $lift(X \Rightarrow Y) = 1$: X and Y not correlated
- $lift(X \Rightarrow Y) < 1$: X and Y negatively correlated

3 Weight-Dependent Positioning

This section describes how support, confidence, and lift can be read from a line diagram where each concept node's y-coordinate is chosen according to the size of its extent. Refining this idea, we then introduce the diagram type *cascading (additive) line diagram*, where dependencies between attributes, can be read using the *parallelogram method*, which is also presented.

3.1 Weighted Formal Contexts

Often, formal contexts representing large real-world data sets contain many objects that coincide in terms of their attributes. To represent such data in a succinct but statistically faithful[1] manner, we endow formal contexts with weights. A *weighted formal context* is a quadruple (G, M, I, mult) extending a formal context (G, M, I) by a mapping mult, which assigns a *weight* (or *multiplicity*)

$$\mathit{mult} : G \to \mathbb{N}^+ \tag{9}$$

Table 2. Example of a weighted context.

	mult	a	b	c	d
O1	10	×	×		
O2	10			×	×
O3	1	×			
O4	1		×		
O5	1			×	
O6	1				×

to every object. In our setting, $\mathit{mult}(g) = n$ means that object g occurs n times in our data set.[2] In order to reflect this in cross tables, we extend them by a column *mult*, which contains the weight of each object (cf. Table 2, where O1 and O2 are taken to appear 10-fold).

Based on this, the weight associated to a concept (A, B) is defined as the sum of the weights of all objects contained in its extent A. We define:

$$\mathit{wgt} : \mathfrak{P}(G) \to \mathbb{N} \tag{10}$$

$$\mathit{wgt}(A) := \sum_{g \in A} \mathit{mult}(g) \tag{11}$$

3.2 Linear Vertical Positioning

In this paper, the nodes in concept lattice diagrams are positioned in a 2-dimensional Cartesian coordinate system. Thereby, in order to implement our goal that statistical information be readable from the diagram, we first investigate the approach where we let the y-position of a node be defined by its weight. In the lattice diagram of a given weighted formal context (G, M, I, mult) defined this way, the highest node has the y-coordinate $\mathit{wgt}(G)$. We note that such a positioning always creates admissible lattice diagrams, since the function defined in Formula (11) is a monotonic mapping from $(\mathfrak{P}(G), \subset)$ to $(\mathbb{N}, <)$.

[1] clarifying the context would prune duplicates but distort the statistical information

[2] Note, however, that the notion easily generalizes to settings where the weight expresses other qualities that justify to assign more statistical importance to certain objects (in which case one might rather choose $\mathbb{Q}$ or $\mathbb{R}$ as codomain).

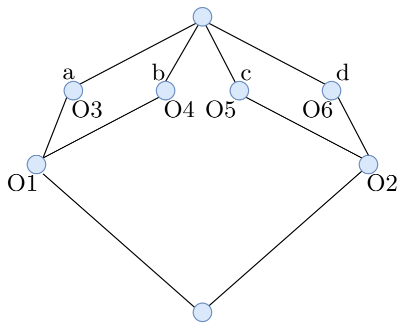


Fig. 2. Additive line diagram.

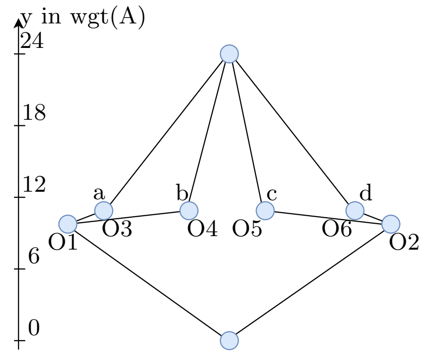


Fig. 3. Y-weighted, additive line diagram.

In additive line diagrams, the position of a node is determined by summing up the vectors associated to all irreducible nodes that define it and adding them to a normalization vector. The lattice diagrams represented in this way are characterized by many parallel lines, which increases readability. By simply adjusting the y-coordinate in the way described above, this advantage is lost. To illustrate this, consider the context from Table 2.

Figure 2 shows a possible additive line diagram which can be derived from the context. Fixing the y-coordinate of each node with its weight results in Fig. 3. The parallel sides of the diagram are lost and some of the lines have become almost horizontal, which is obstructing readability. In this work, therefore, the x-coordinate is not determined as an unweighted sum of the irreducible nodes, as is the case with additive line diagrams. Instead, the x-position is determined by compressing or stretching the vector resulting from the summation of the vectors of all upper neighbor nodes. Figure 4 illustrates this approach. Figure 5 shows the resulting line diagram without auxiliary lines.

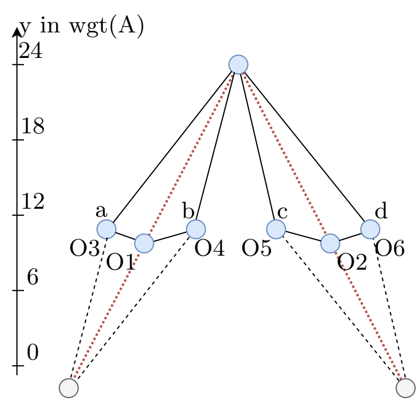


Fig. 4. Weighted line diagram with auxiliary lines.

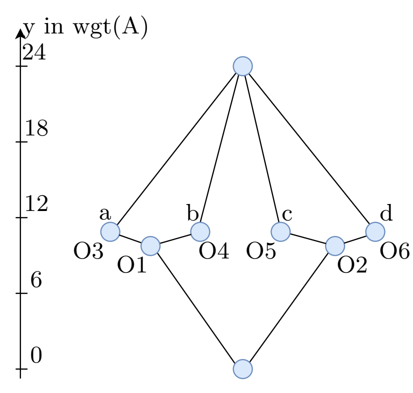


Fig. 5. X-Y-weighted, additive line diagram.

Support. Having in mind the meaning of multiplicities, the support of a weighted context's concept (A, B) is defined analogous to Formula (5) by its weight as

$$supp(B) := \frac{wgt(B')}{wgt(G)} = \frac{wgt(A)}{wgt(G)} =: supp(A) \tag{12}$$

Since $wgt(G)$ is constant and $wgt(A)$ was used as the weight of the concept to define the y-coordinate of the corresponding node, the support can also be read from the previously defined lattice diagram. This only requires an adjustment of the labeling of the y-axis.

In order to illustrate this, we introduce a new example, which will be used in the following to explain how to read off the confidence, the lift, and the correlation. The used weighted context is shown in Table 3. It reflects the statistical distribution by height and gender in Germany in 2006, differentiated by height <175 and ≥175 centimeters. The *mult* column shows the percentage of the respective gender for each height.

Table 3. Distribution by height and gender in Germany 2006 [2].

	mult	male	female	<175	≥175
M1	31	×		×	
M2	69	×			×
F1	91		×	×	
F2	9		×		×

Figure 6 shows a line diagram that can be derived from the given context. On the y-axis, in addition to the weight $wgt(A)$, the support $supp(A)$ is shown. The support is calculated by Formula (12). This line diagram allows for reading the weight as well as the support of the displayed concepts from the y-axis.

The natural way of defining the support of an implication in a weighted context is by the weight of all objects that fulfill this implication, relative to the weight of the set of all objects G. It can be formalized in two ways using the intents X and Y or the extents X' and Y'.

$$supp(X \to Y) := \frac{supp(X' \cap Y')}{supp(G)} = \frac{supp(X \cup Y)}{supp(G)} \tag{13}$$

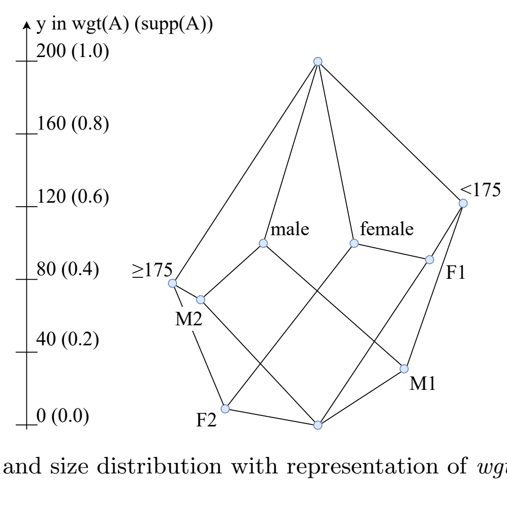


Fig. 6. Gender and size distribution with representation of $wgt(A)$ and $supp(A)$.

Confidence. As explained in Section 2.1, implications can be read off from the concept lattice diagram. For example, the line diagram displayed in Fig. 6 can be used to find the implication $\{male\} \rightarrow \{\geq 175\}$.

According to Section 2.2, the confidence of an association rule can be determined using Formula (7). For an implication $X \rightarrow Y$, the confidence is defined analogously as:

$$conf(X \rightarrow Y) := \frac{supp(X \rightarrow Y)}{supp(X)} \tag{14}$$

3.3 Logarithmic Vertical Positioning

Since the confidence is defined as a fraction, it is difficult to read it from the previously defined representation of the concept lattice diagram. With the help of the logarithmic law $\log_a(x/y) = \log_a(x) - \log_a(y)$, Formula (14) can be transformed and displayed as subtraction.

$$\log_a(conf(X \rightarrow Y)) := \log_a\left(\frac{supp(X \rightarrow Y)}{supp(X)}\right) \tag{15}$$

$$= \log_a(supp(X \rightarrow Y)) - \log_a(supp(X)) \tag{16}$$

The logarithmized confidence could be easily read from a line diagram, where the logarithmic support is shown on the y-axis. There are several possibilities for the choice of the base a of the used logarithm, but they only lead to a linear vertical scaling. In the following, the base 2 was chosen, since $\log_2(0.5) = -1.0$ applies and thus the results can be easily estimated. The concept lattice with logarithmically scaled y-axis is shown in Fig. 7.

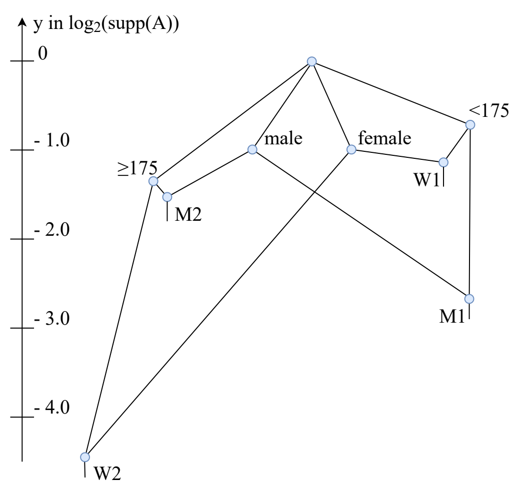


Fig. 7. Gender and size distribution with representation of the logarithmized support of each node on the y-axis.

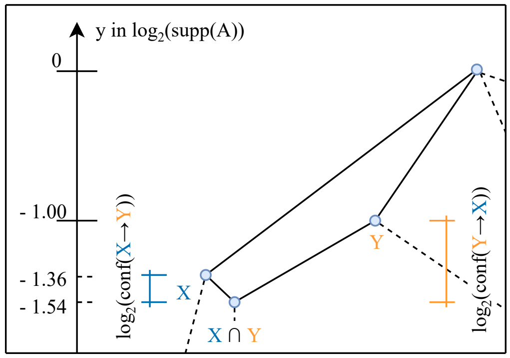


Fig. 8. Example for reading the logarithmic confidence for the implications $X \to Y$ and $Y \to X$.

Since nodes with weight zero would have the value $-\infty$ with this scaling, they are not displayed. To be able to see if a bottom element with weight 0 exists, short vertical auxiliary lines are attached to those nodes, that share an edge with it. Those can be seen in Fig. 7. Figure 8 illustrates the process of reading the logarithmic confidence for the implications $X \to Y$ and $Y \to X$. It shows an excerpt of Figure 7.

Reading correlation. As described in Section 2.2, it is possible to derive from the lift of an association rule $X \Rightarrow Y$ how X and Y are correlated. Analogous to Formula (8) for association rules, the lift for implications $X \to Y$ in weighted contexts is defined by

$$lift(X \to Y) := \frac{conf(X \to Y)}{supp(Y)}. \tag{17}$$

Since statements about the polarity and approximate strength of the correlation are usually more important than the correlation coefficient's exact value, it is not necessary to calculate the lift exactly. In many cases, a quantitative estimation is sufficient. If the lift is greater than one, X and Y are positively correlated. If it is smaller than one, they are negatively correlated. As the lift is defined as a fraction, it is advisable to apply the logarithm, similar to the case of the confidence. This way, we obtain the following characterization for the case of positive correlation:

$$\frac{conf(X \to Y)}{supp(Y)} > 1.0 \tag{18}$$

$$\log_2\left(\frac{conf(X \to Y)}{supp(Y)}\right) > \log_2(1.0) \tag{19}$$

$$\log_2(conf(X \to Y)) - \log_2(supp(Y)) > 0.0 \tag{20}$$

$$\log_2(conf(X \to Y)) > \log_2(supp(Y)) \tag{21}$$

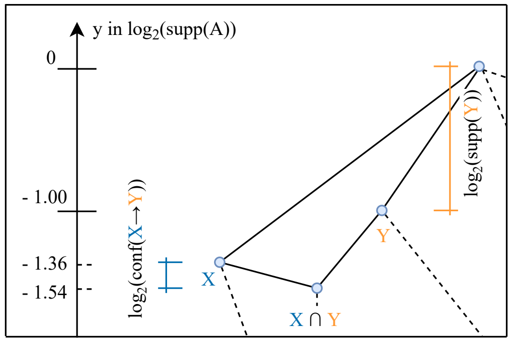


Fig. 9. Estimating the dependence of X and Y by reading off confidence and support.

To estimate whether X and Y are positively correlated, it is therefore sufficient to compare the logarithmic values of the implication's confidence and the head's support. This can be done purely graphically.

Figure 9 shows the values to be read for the estimation of the lift for the implication $X \rightarrow Y$. It shows a cutout of Fig. 7, with X being ≥ 175 and Y being *male*. It can therefore be concluded that there is a positive lift for the implication $\{\geq 175\} \rightarrow \{male\}$, since $\log_2(conf(\{\geq 175\} \rightarrow \{male\}))$ is -0.18 and $\log_2(supp(male))$ is -1.0. Comparing the two values shows

$$-0.18 > -1.0 \;\rightarrow\; \text{positively correlated} \tag{22}$$

A second possibility to read the correlation between X and Y in the logarithmic lattice diagram can be derived from the definition of the lift. This is due to the fact that the lift for the implication $X \rightarrow Y$ mathematically equals the lift for the implication $Y \rightarrow X$.

Another way to determine whether X and Y are statistically dependent or independent is the *parallelogram method* presented below. As already described, X and Y are statistically independent exactly if

$$\log_2(conf(X \rightarrow Y)) = \log_2(supp(Y)). \tag{23}$$

If the independence of X and Y can be derived from $X \rightarrow Y$, then follows

$$\log_2(conf(Y \rightarrow X)) = \log_2(supp(X)). \tag{24}$$

If X and Y are independent, i.e. Formulas (23) and (24) are satisfied, a parallelogram is formed in the diagram. However, if they are dependent, a one-sided distortion of the parallelogram along the diagonal occurs. A simple, convex square is formed. If the downward-pointing half of the parallelogram is compressed, X and Y are positively correlated. If it is stretched, they are negatively correlated. Table 4 shows all three possible correlation types together with an example.

On the basis of the deviation of the calculated parallelogram the dependence of X and Y can be read. For this purpose, it may be helpful to mark the result-

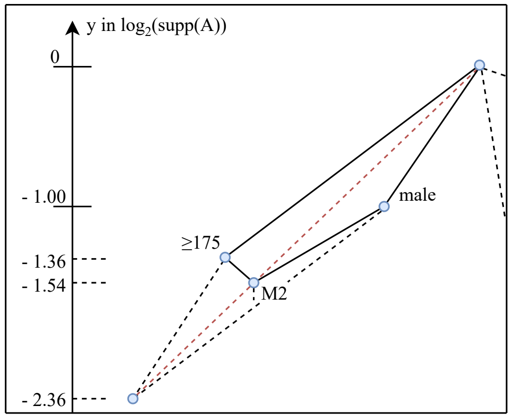


Fig. 10. Illustrating the parallelogram method. The downward-pointing half of the quadrangle is compressed, indicating that ≥ 175 and *male* are positively correlated.

ing parallelogram, which symbolizes independence, with auxiliary lines in the diagram. Figure 10 shows such an example.

3.4 Cascading Additive Line Diagrams

In order to achieve that the parallelogram method can also be used in situations where the nodes representing X and Y do not share an edge with the top node, we refine our idea leading to the definition of *cascading additive line diagrams*[3], detailed below.

The y-coordinate of each node in a cascading additive line diagram is defined by the logarithm of the support of the concept associated with the corresponding node, just as it was described in Section 3.1. The x-coordinate, on the other hand, is defined taking into account the positions of all upper neighbor nodes, leading to a recursive definition – hence the name 'cascading'. The upper neighbors (also called direct predecessors) of the node associated to the concept (A, B) are determined according to the order relation as defined in Section 2 and Formula (25).

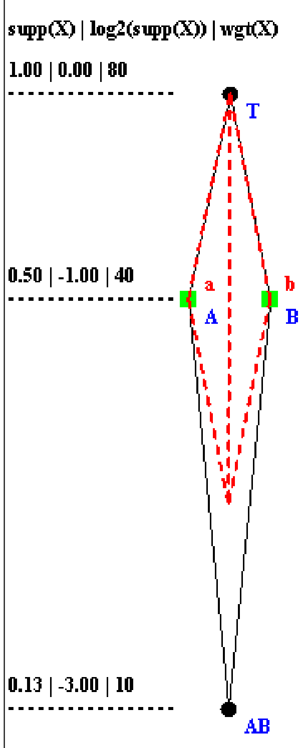
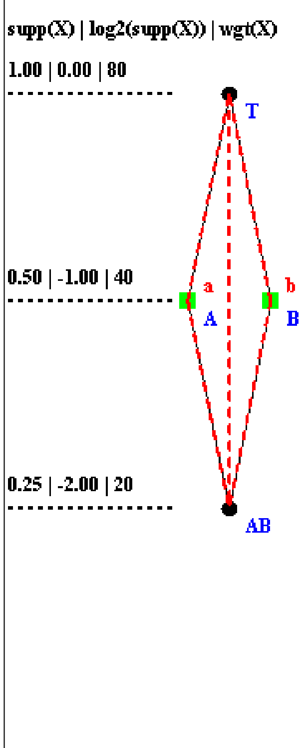
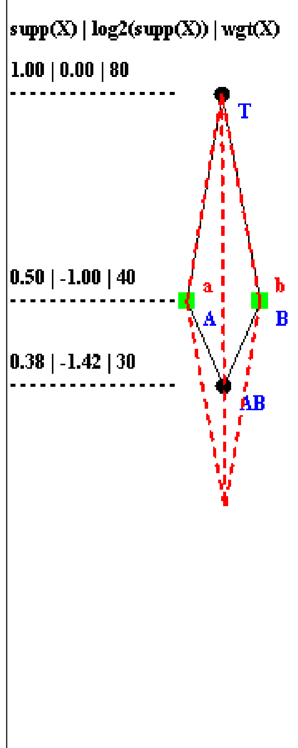
$$pre(A, B) := \{(C, D) \in \mathfrak{B}(G, M, I) \mid (A, B) \prec (C, D)\} \tag{25}$$

Remember that in the diagram, these are just the nodes that have a higher y-position and share an edge with the node in question. The *unscaled position* $upos(A, B)$ of each concept node is defined by adding the positions of all predecessors:

$$upos(A, B) := \sum_{(C,D)\in pre(A,B)} pos(C, D) \; + \sum_{m\in M_{irr},\, A=\{m\}'} vec(m) \tag{26}$$

[3] short: cascading line diagrams

Table 4. Minimal examples showcasing the three different types of correlation. The diagrams have been created with the help of the program described in Section 4.

<table>
<tr><td>negatively correlated</td><td>independent</td><td>positively correlated</td></tr>
<tr>
<td>
<table>
<tr><th></th><th>mult</th><th>a</th><th>b</th></tr>
<tr><td>T</td><td>10</td><td></td><td></td></tr>
<tr><td>A</td><td>30</td><td>×</td><td></td></tr>
<tr><td>B</td><td>30</td><td></td><td>×</td></tr>
<tr><td>AB</td><td>10</td><td>×</td><td>×</td></tr>
</table>
</td>
<td>
<table>
<tr><th></th><th>mult</th><th>a</th><th>b</th></tr>
<tr><td>T</td><td>20</td><td></td><td></td></tr>
<tr><td>A</td><td>20</td><td>×</td><td></td></tr>
<tr><td>B</td><td>20</td><td></td><td>×</td></tr>
<tr><td>AB</td><td>20</td><td>×</td><td>×</td></tr>
</table>
</td>
<td>
<table>
<tr><th></th><th>mult</th><th>a</th><th>b</th></tr>
<tr><td>T</td><td>30</td><td></td><td></td></tr>
<tr><td>A</td><td>10</td><td>×</td><td></td></tr>
<tr><td>B</td><td>10</td><td></td><td>×</td></tr>
<tr><td>AB</td><td>30</td><td>×</td><td>×</td></tr>
</table>
</td>
</tr>
<tr>
<td>$\mathit{lift}(\{a\} \rightarrow \{b\}) = \frac{1/8}{1/2 \cdot 1/2} = 1/2 < 1.0$</td>
<td>$\mathit{lift}(\{a\} \rightarrow \{b\}) = \frac{1/4}{1/2 \cdot 1/2} = 1.0$</td>
<td>$\mathit{lift}(\{a\} \rightarrow \{b\}) = \frac{3/8}{1/2 \cdot 1/2} = 3/2 > 1.0$</td>
</tr>
<tr>
<td>
</td>
<td>
</td>
<td>
</td>
</tr>
<tr><td>stretched</td><td>well-formed</td><td>compressed</td></tr>
</table>

By scaling the unscaled position with respect to the y-coordinate, the final position of the node is determined:

$$pos(A, B) := \frac{\log_2(supp(B))}{upos_y(A, B)} \cdot upos(A, B) \tag{27}$$

To illustrate this presentation method in more detail, we use the context shown in Table 5. This context is characterized by the independence of $\{k\}$ and $\{m\}$, which is proved mathematically in Formula (28) by calculating the lift of $\{k\} \to \{m\}$.

$$lift(\{k\} \to \{m\}) := \frac{supp(\{k\} \to \{m\})}{supp(\{k\}) \cdot supp(\{m\})} = \frac{0.05}{0.25 \cdot 0.2} = 1.0 \tag{28}$$

Table 5. Example context.

	mult	g	h	k	m
G	25	×			
H	30		×		
GH	5	×	×		
GHK	20	×	×	×	
GHM	15	×	×		×
GHKM	5	×	×	×	×

Figure 11 displays the corresponding concept lattice as a cascading additive line diagram. It can be seen that the top-element forms a parallelogram with the nodes labeled with "GHK", "GHM" and "GHKM", thus displaying the independence of $\{k\}$ and $\{m\}$. If the node labeled "GHKM" were higher, meaning it had a larger y-coordinate, then $\{k\}$ and $\{m\}$ would be positively correlated. If the y-coordinate were smaller, the parallelogram would be stretched on the lower side. This would correspond to a negative correlation.

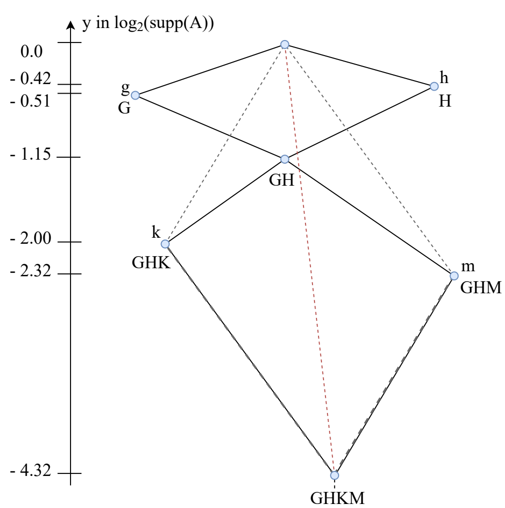


Fig. 11. Cascading line diagram of the context shown in Table 5.

The cascading additive line diagram allows to read at each node that has exactly two predecessors, the (in)dependence of these predecessors using the parallelogram method. Furthermore, in all cases, it is possible to read the dependence by means of the y-coordinates, as discussed in Section 3.1.

4 Prototype Implementation

To demonstrate the operation of the cascading additive line diagram, we implemented a visualisation prototype. The source code is freely available online [1]. The tool creates concept lattice diagrams for given formal contexts. It is able to generate both additive and cascading line diagrams.

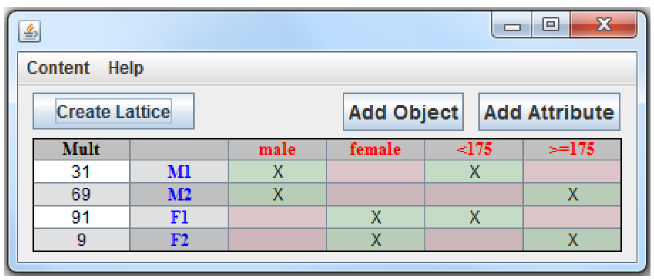


Fig. 12. Display of a cross table in the program.

Figure 12 shows the user interface for entering the formal context. In the screenshot, the example from Table 3 has been entered. Clicking on 'Create Lattice' triggers the creation of the concept lattice diagram. The additive line diagram is created as described in Section 2.1. The cascading line diagram is created as described in Section 3.4.

Upon displaying the diagram, the irreducible nodes are marked by green squares. In the cascading view, the independence of two attributes is expressed by the parallelogram formed with the top element. By left-clicking on a reducible node with two predecessors, the parallelogram that would result if the predecessors were independent is shown in red. This is shown in Fig. 13, using the context from Table 3, which was considered in detail in Section 3.2. Figure 13 also shows an information window, which opens upon right-clicking on a node.

5 Conclusion

The goal of this paper was to develop a representation for concept lattice diagrams that – in addition to the attribute-logical relationships – allows statistical relationships between attributes to be read off the diagram. Toward a more succinct representation of the input data, the cross table defining the formal context was extended by a multiplier column, where the positive natural number in this

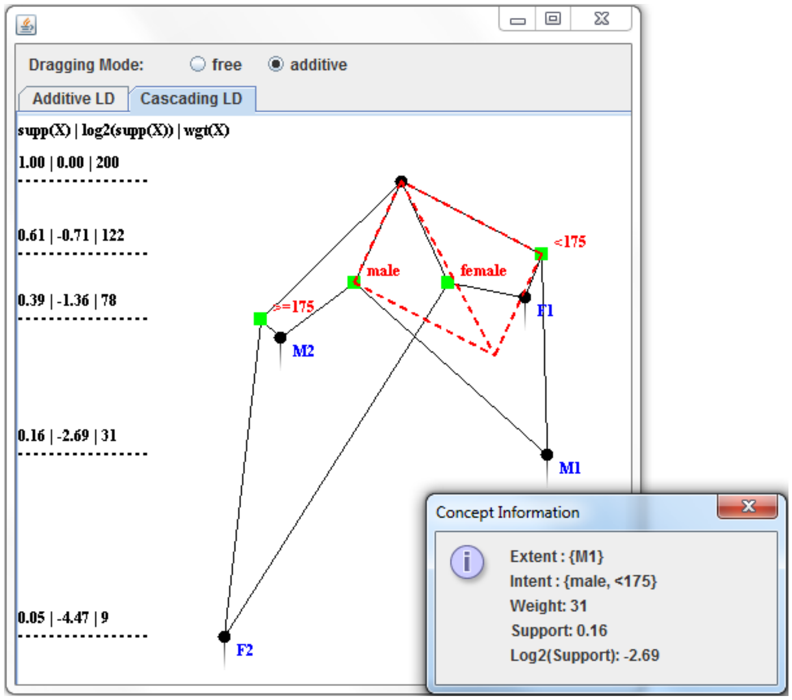


Fig. 13. Display of a cascading additive line diagram in the program with the parallelogram with the top element shown and the additional window opened.

cell indicates the weight or multiplicity of the associated object. This can be understood to represent an aggregation of several objects with the same intent from a normal cross table.

Inspired by the display form of additive line diagrams, a new display form was then developed, which allows for reading statistical and associative measures such as support, confidence, and lift from the concept lattice diagram. Furthermore, the statistical (in)dependence of two attributes can be derived from the diagram. This was achieved by a vertical positioning following the logarithmized support of the associated concept. Through additive scaling, it was also possible to develop a form of representation, called cascading additive line diagram, in which the dependence or independence of two attributes can be intuitively grasped from their formed parallelograms (or, rather, the deviation from the parallelogram shape).

We presented a prototype implementation that allows to input a formal context as a cross table and, from this, to derive, display, and adjust both additive and a cascading line diagrams.

It has been shown in this paper that a formal context's concept lattice can be represented by a line diagram from which, beyond the classical logical dependencies, statistical relationships between attributes can be read rather directly and intuitively.

As one avenue of future work, the proposed representation paradigm could be coupled with existing optimization approaches toward a beneficial choice of the vector assignment *vec* to the irreducible attributes, with the goal of avoiding (near-)overlap of nodes, nodes being positioned on (or close to) edges they are not incident with, and reducing the overall number of edge crossings.

Finally, our novel visualization approach will have to be tested empirically: determining if this way of presenting statistical information is indeed useful to human users and how well this paradigm works for larger concept lattices can only be found out by means of comprehensive user studies.

References

1. https://github.com/Klimpke/Cascading-Line-Diagrams-Visualizer
2. Körpergröße in Deutschland nach Geschlecht 2006 [Body height in Germany by gender 2006]. https://de.statista.com/statistik/daten/studie/1825/umfrage/koerpergroesse-nach-geschlecht/ (2006), accessed: 2020-09-23
3. Abdullah, Z., Saman, M.Y.M., Karim, B., Herawan, T., Deris, M.M., Hamdan, A.R.: FCA-ARMM: A model for mining association rules from formal concept analysis. In: Herawan, T., Ghazali, R., Nawi, N.M., Deris, M.M. (eds.) Proc. 2nd Int. Conf. on Soft Computing and Data Mining (SCDM). AISC, vol. 549, pp. 213–223. Springer (2016)
4. Agrawal, R., Imieliński, T., Swami, A.: Mining association rules between sets of items in large databases. In: Proc. Int. Conf. on Management of Data (SIGMOD). p. 207–216. ACM (1993)
5. Brin, S., Motwani, R., Ullman, J.D., Tsur, S.: Dynamic itemset counting and implication rules for market basket data. SIGMOD Rec. **26**(2), 255–264 (1997)
6. Ganter, B., Wille, R.: Formal Concept Analysis – Mathematical Foundations. Springer-Verlag (1999)
7. Hannan, T., Pogel, A.: Spring-based lattice drawing highlighting conceptual similarity. In: Missaoui, R., Schmid, J. (eds.) Proc. 4th Int. Conf. on Formal Concept Analysis (ICFCA). LNCS, vol. 3874, pp. 264–279 (2006)
8. Lakhal, L., Stumme, G.: Efficient mining of association rules based on formal concept analysis. In: Ganter, B., Stumme, G., Wille, R. (eds.) Formal Concept Analysis, Foundations and Applications. LNCS, vol. 3626, pp. 180–195. Springer (2005)
9. Mondal, K.C., Pasquier, N., Mukhopadhyay, A., Maulik, U., Bandhopadyay, S.: A new approach for association rule mining and bi-clustering using formal concept analysis. In: Perner, P. (ed.) Machine Learning and Data Mining in Pattern Recognition. pp. 86–101. Springer (2012)
10. Pogel, A., Hannan, T., Miller, L.: Visualization of concept lattices using weight functions. In: Wolff, K.E., Pfeiffer, H.D., Delugach, H.S. (eds.) Supp. Proc. 12th Int. Conf. on Conceptual Structures (ICCS). pp. 1–14. Shaker (2004)
11. Stumme, G., Taouil, R., Bastide, Y., Pasquier, N., Lakhal, L.: Intelligent structuring and reducing of association rules with formal concept analysis. In: Baader, F., Brewka, G., Eiter, T. (eds.) Proc. KI. pp. 335–350. Springer (2001)