Scaling, Granulation, and Fuzzy Attributes in Formal Concept Analysis

Radim Belohlavek, *Member IEEE*, and Jan Konecny

***Abstract*— The present paper deals with scaling within the framework of formal concept analysis (FCA) of data with fuzzy attributes. In ordinary FCA, the input is a data table with yes/no attributes. Scaling is a process of transformation of data tables with general attributes, e.g. nominal, ordinal, etc., to data tables with yes/no attributes. This way, data tables with general attributes can be analyzed by means of FCA.**

We propose a new way of scaling, namely, scaling of general attributes to fuzzy attributes. After such a scaling, the data can be analyzed by means of FCA developed for data with fuzzy attributes. Compared to ordinary scaling to yes/no attributes, our scaling procedure is less sensitive to how a user defines a scale which eliminates the arbitrariness of user's definition of a scale. This is the main advantage of our approach. In addition, scaling to fuzzy attributes is appealing from the point of view of knowledge representation and is connected to Zadeh's concept of linguistic variable. We present a general definition of scaling, examples comparing our approach to ordinary scaling, and theorems which answer some naturally arising questions regarding sensitivity of FCA to the definition of a scale.

I. Problem setting

Tabular data describing objects and their attributes represents a basic form of data. Among various methods for analysis of object-attribute data, formal concept analysis (FCA) is becoming increasingly popular, see [9], [7]. The main aim in FCA is to extract interesting clusters (called formal concepts) from tabular data. A partially ordered collection of all formal concept is called a concept lattice.

In the basic setting, the input data to FCA is organized in a table such as the one in Tab. I. The table represents

TABLE I

Table describing objects x_1, x_2, x_3 and their yes/no attributes y_1, y_2, y_3, y_4.

	y_1	y_2	y_3	y_4
x_1	1	1	0	0
x_2	0	1	1	0
x_3	0	0	1	1

information about objects x_1, x_2, x_3 and their yes/no attributes (binary attributes, crisp attributes) y_1, y_2, y_3, y_4. According to the table, object x_1 has attribute y_1 (table entry contains 1), object x_2 does not have attribute y_1 (table entry contains 0), etc. In order for FCA to have the capability to analyze data with general attributes, like the one in Tab. II, FCA uses so-called (conceptual) scaling. Scaling, basically, represents a transformation of a table with general attributes to a table with yes/no attributes. For instance (see later for an exact definition of conceptual scaling), attribute age could be replaced by three yes/no attributes a_y, a_m, a_o, corresponding to age intervals $[0, 30]$, $[31, 50]$, $[51, \infty]$, which represent "young", "middle", "old". That is, a person with age, say, 18, has attribute a_y but has neither attribute a_m nor a_o. In a similar way, one can introduce attributes h_s, h_m, h_t corresponding to height intervals $[0, 160]$, $[161, 180]$, and $[181, \infty]$. This way, Tab. II can be transformed into a table with yes/no attributes like the one in Tab. III. After scaling, the data can be processed by means of FCA.

TABLE II

Table describing persons Alice, ..., George (objects) and their attributes (age in years, height in cm, presence of symptom).

	age	height	symptom
Alice	23	165	1
Boris	30	180	0
Cyril	31	167	1
David	43	159	0
Ellen	24	155	1
Fred	64	170	0
George	30	190	0

TABLE III

Table describing persons Alice, ..., George (objects) and their attributes after scaling to crisp attributes.

	a_y	a_m	a_o	h_s	h_m	h_t	symptom
Alice	1	0	0	0	1	0	1
Boris	1	0	0	0	1	0	0
Cyril	0	1	0	0	1	0	1
David	0	1	0	1	0	0	0
Ellen	1	0	0	1	0	0	1
Fred	0	0	1	0	1	0	0
George	1	0	0	0	0	1	0

Scaling is a particular form of information granulation [17]. Namely, new attributes, such as a_y, represent granules. For instance, a_y represents a granule consisting of age values $[0, 30]$. As convincingly argued by Zadeh, see e.g. [17], granules involved in human reasoning are vague rather than sharply delineated. Typical examples are granules corresponding to linguistic expressions like "young", "old", "tall", etc. Therefore, these granules should be represented by fuzzy sets rather than ordinary sets. From this point of view, scaling to crisp (binary, yes/no) attributes is not appropriate. For instance, if scaling is performed according to Tab. III, attribute

R. Belohlavek is with the Binghamton University – SUNY, U. S. A. (e-mail: rbelohla@binghamton.edu) and with the Palacky University, Olomouc, Czech Republic.

J. Konecny is with the Palacky University, Olomouc, Czech Republic (email: konecny@inf.upol.cz).

a_y ("young") fully applies to a person of age 30, but does not apply at all to a person of age 31, which is counterintuitive.

There is an obvious way to overcome this problem. Namely, instead of crisp attributes, one can use fuzzy attributes. After a scaling to fuzzy attributes using a set $\{0, 0.5, 1\}$ of truth degrees, the transformed table might look like the one in Tab. IV.

TABLE IV
TABLE DESCRIBING PERSONS ALICE, ..., GEORGE (OBJECTS) AND THEIR ATTRIBUTES AFTER SCALING TO FUZZY ATTRIBUTES USING SET $\{0, 0.5, 1\}$ OF TRUTH DEGREES .

	a_y	a_m	a_o	h_s	h_m	h_t	symptom
Alice	1	0.5	0	0.5	1	0	1
Boris	1	0.5	0	0	0.5	0.5	0
Cyril	0.5	1	0	0.5	1	0	1
David	0	1	0.5	1	0.5	0	0
Ellen	1	0.5	0	1	0.5	0	1
Fred	0	0	1	0.5	1	0	0
George	1	0.5	0	0	0	1	0

Scaling, however, is a kind of data preprocessing. After scaling, data is processed by means of FCA. That is, one can extract all formal concepts from the data, a non-redundant basis of attribute implications from the data, etc. In the basic setting, FCA analyzes data with yes/no attributes. Therefore, scaling to fuzzy attributes results in data to which ordinary FCA cannot be applied. Fortunately, in a series of papers, FCA was extended to be applicable to data with fuzzy attributes, see e.g. [2], [3], [5], [6], [15]. Scaling to fuzzy attributes therefore nicely fits algorithms and theoretical foundations of existing methods.

The main aim of this paper is to propose a general definition of scaling to fuzzy attributes, look in detail at the basic properties of this scaling, and compare by means of illustrative examples to ordinary scaling. A point to emphasize is that scaling to fuzzy attributes behaves naturally w.r.t. similarity of attribute values and lends themselves to sensitivity analysis for which we present results in this paper.

Note that [15] and [16] are related papers. We will comment on them in the full version of this paper.

II. PRELIMINARIES

A. Fuzzy Sets and Fuzzy Logic

We use sets of truth degrees equipped with operations (logical connectives) so that is becomes a complete residuated lattice [10], i.e., an algebra $\mathbf{L} = \langle L, \wedge, \vee, \otimes, \rightarrow, 0, 1\rangle$ such that $\langle L, \wedge, \vee, 0, 1\rangle$ is a complete lattice with 0 and 1 being the least and greatest element of L, respectively; $\langle L, \otimes, 1\rangle$ is a commutative monoid (i.e. $\otimes$ is commutative, associative, and $a \otimes 1 = 1 \otimes a = a$ for each $a \in L$); $\otimes$ and $\rightarrow$ satisfy so-called adjointness property, i.e., $a \otimes b \leq c$ iff $a \leq b \rightarrow c$ for each $a, b, c \in L$. A (truth-stressing) hedge on $\mathbf{L}$ [10], [11] is a particular unary function on L. We will use hedges satisfying the following properties: $1^* = 1$, $a^* \leq a$, $(a \rightarrow b)^* \leq a^* \rightarrow b^*$, $a^{**} = a^*$, for each $a, b \in L$. Elements a of L are called truth degrees. $\otimes$ and $\rightarrow$ are (truth functions of) "fuzzy conjunction" and "fuzzy implication". Hedge * is a (truth function of) logical connective "very true", see [10], [11]. Properties of hedges have natural interpretations, e.g. "$a^* \leq a$" can be read: "if formula φ is very true, then φ is true", etc. Note that hedges other than truth-stressing ones like "at least a little bit true" have different properties and are not considered in our paper.

A common choice of $\mathbf{L}$ is a structure with $L = [0, 1]$ (unit interval), $\wedge$ and $\vee$ being minimum and maximum, $\otimes$ being a left-continuous t-norm with the corresponding $\rightarrow$. Three most important pairs of adjoint operations $\otimes$ and $\rightarrow$ on the unit interval are the well-known Łukasiewicz, Goguen (product), and Gödel (minimum) operations, see, e.g. [2], [10], [12]. In applications, we often need a finite linearly ordered $\mathbf{L}$. For instance, one can put $L = \{a_0 = 0, a_1, \ldots, a_n = 1\} \subseteq [0, 1]$ ($a_0 < \cdots < a_n$) with $\otimes$ given by $a_k \otimes a_l = a_{\max(k+l-n,0)}$ and the corresponding $\rightarrow$ given by $a_k \rightarrow a_l = a_{\min(n-k+l,n)}$. Such an $\mathbf{L}$ is called a finite Łukasiewicz chain. Another possibility is a finite Gödel chain which consists of L and restrictions of Gödel operations on $[0, 1]$ to L.

Two boundary cases of (truth-stressing) hedges are (i) identity, i.e. $a^* = a$ ($a \in L$); (ii) globalization:

$$a^* = \begin{cases} 1 & \text{if } a = 1, \\ 0 & \text{otherwise.} \end{cases} \qquad (1)$$

A special case of a complete residuated lattice with hedge is a two-element Boolean algebra $\langle\{0, 1\}, \wedge, \vee, \otimes, \rightarrow, ^*, 0, 1\rangle$, denoted by $\mathbf{2}$, which is the structure of truth degrees of the classical logic. That is, operations $\wedge, \vee, \otimes, \rightarrow$ of $\mathbf{2}$ are truth functions (interpretations) of the corresponding logical connectives of classical logic and $0^* = 0$, $1^* = 1$.

Having $\mathbf{L}$, we define usual notions: an $\mathbf{L}$-set (fuzzy set) A in universe U is a mapping $A\colon U \rightarrow L$, $A(u)$ being interpreted as "the degree to which u belongs to A". If $U = \{u_1, \ldots, u_n\}$ then A can be denoted by $A = \{^{a_1}/u_1, \ldots, ^{a_n}/u_n\}$ meaning that $A(u_i)$ equals a_i for each $i = 1, \ldots, n$. Let $\mathbf{L}^U$ denote the collection of all $\mathbf{L}$-sets in U. The operations with $\mathbf{L}$-sets are defined componentwise. For instance, the intersection of $\mathbf{L}$-sets $A, B \in \mathbf{L}^U$ is an $\mathbf{L}$-set $A \cap B$ in U such that $(A \cap B)(u) = A(u) \wedge B(u)$ for each $u \in U$, etc. $\mathbf{2}$-sets (operations with $\mathbf{2}$-sets) can be identified with the ordinary (crisp) sets (operations with ordinary sets) of the naive set theory. Binary $\mathbf{L}$-relations (binary fuzzy relations) between X and Y can be thought of as $\mathbf{L}$-sets in the universe $X \times Y$.

Given $A, B \in \mathbf{L}^U$, we define a subsethood degree $S(A, B)$ and a degree $A \approx B$ of equality by

$$S(A, B) = \textstyle\bigwedge_{u \in U}\big(A(u) \rightarrow B(u)\big), \qquad (2)$$

$$A \approx B = \textstyle\bigwedge_{u \in U}\big(A(u) \leftrightarrow B(u)\big), \qquad (3)$$

which generalize the classical subsethood relation $\subseteq$ and equality relation $=$. Note that $a \leftrightarrow b = (a \rightarrow b) \wedge (b \rightarrow a)$. Put verbally, $S(A, B)$ and $A \approx B$ are degrees to which A is a subset of B and to which A equals B, respectively. In particular, we write $A \subseteq B$ iff $S(A, B) = 1$. As a consequence, $A \subseteq B$ iff $A(u) \leq B(u)$ for each $u \in U$.

B. Formal Concept Analysis of Data With Fuzzy Attributes

This section summarizes basic notions of formal concept analysis of data with fuzzy attributes. A data table with fuzzy attributes, which represents the input table, is represented by a triplet $\langle X, Y, I\rangle$ (formal fuzzy context, in terms of FCA) where X is a set of objects, Y is a finite set of attributes, and $I \in \mathbf{L}^{X\times Y}$ is a binary fuzzy relation between X and Y assigning to each object $x \in X$ and each attribute $y \in Y$ a degree $I(x, y) \in L$ to which x has y. $\langle X, Y, I\rangle$ can be thought of as a table with rows and columns corresponding to objects $x \in X$ and attributes $y \in Y$, respectively, and table entries containing degrees $I(x, y)$.

For $A \in \mathbf{L}^X$, $B \in \mathbf{L}^Y$ (i.e. A is a fuzzy set of objects, B is a fuzzy set of attributes), we define fuzzy sets $A^\uparrow \in \mathbf{L}^Y$ (fuzzy set of attributes), $B^\downarrow \in \mathbf{L}^X$ (fuzzy set of objects) by

$$A^\uparrow(y) = \textstyle\bigwedge_{x\in X}\big(A(x)^{*_X} \to I(x, y)\big),$$
$$B^\downarrow(x) = \textstyle\bigwedge_{y\in Y}\big(B(y)^{*_Y} \to I(x, y)\big).$$

^{*_X} and ^{*_Y} are hedges; they control the number of extracted formal concepts, see [5]. Put verbally, $A^\uparrow$ is the fuzzy set of all attributes from Y shared by all objects from A (and similarly for $B^\downarrow$). A formal (fuzzy) concept in $\langle X, Y, I\rangle$ is a pair $\langle A, B\rangle$ of $A \in \mathbf{L}^X$ and $B \in \mathbf{L}^Y$ satisfying $A^\uparrow = B$ and $B^\downarrow = A$. That is, a fuzzy concept consists of a fuzzy set A (so-called extent) of objects which fall under the concept and a fuzzy set B (so-called intent) of attributes which fall under the concept such that A is the fuzzy set of all objects sharing all attributes from B and, conversely, B is the fuzzy set of all attributes from Y shared by all objects from A.

A collection $\mathcal{B}(X, Y, I) = \{\langle A, B\rangle \mid A^\uparrow = B, B^\downarrow = A\}$ of all formal concepts in $\langle X, Y, I\rangle$ can be equipped with a partial order $\leq$ modeling the subconcept-superconcept hierarchy (e.g., *dog* $\leq$ *mammal*) defined by

$$\langle A_1, B_1\rangle \leq \langle A_2, B_2\rangle \text{ iff } A_1 \subseteq A_2 \text{ (iff } B_2 \subseteq B_1). \quad (4)$$

Note that $^\uparrow$ and $^\downarrow$ form a fuzzy Galois connection [2] and that $\mathcal{B}(X, Y, I)$ is in fact a set of all fixed points of this Galois connection. Under $\leq$, $\mathcal{B}(X, Y, I)$ happens to be a complete lattice, called a fuzzy concept lattice of $\langle X, Y, I\rangle$. The basic structure of fuzzy concept lattices is described by so-called main theorem of concept lattices [2], [3], [5]. Note that FCA of data with fuzzy attributes generalizes ordinary FCA. Namely, if $\mathbf{L}$ is a two-element Boolean algebra $\mathbf{2}$, the above notions (formal concept, concept lattice) become the notions of ordinary FCA provided we identify ordinary sets with their characteristic functions. FCA with binary attributes is thoroughly covered in [7], [9] where one can find theoretical foundations, methods and algorithms, and applications in various areas.

III. Scaling

A. Many-Valued Contexts (Data Tables With General Attributes)

We start by introducing a formal counterpart to data tables with general attributes. To comply with FCA, we use the notion of a many-valued context, see [9].

Definition 1: A many-valued context (data table with general attributes) is a tuple $\mathcal{D} = \langle X, Y, W, I\rangle$ where X is a nonempty finite set of objects, Y is a finite set of (many-valued) attributes, W is a set of values, and I is a ternary relation between X, Y, and W, i.e., $I \subseteq X \times Y \times W$, such that

$$\langle x, y, w\rangle \in I \text{ and } \langle x, y, v\rangle \in I \text{ imply } w = v.$$

Remark 1: (1) A many-valued context can be thought of as representing a table with rows corresponding to $x \in X$, columns corresponding to $y \in Y$, and table entries at the intersection of row x and column y containing values $w \in W$ provided $\langle x, y, w\rangle \in I$ and containing blanks if there is no $w \in W$ with $\langle x, y, w\rangle \in I$.

(2) One can see that each $y \in Y$ can be considered a partial function from X to W. Therefore, we often write

$$y(x) = w \;\text{ instead of } \langle x, y, w\rangle \in I.$$

A set

$$\mathrm{dom}(y) = \{x \in X \mid \langle x, y, w\rangle \in I \text{ for some } w \in W\}$$

is called a domain of y. Attribute $y \in Y$ is called complete if $\mathrm{dom}(y) = X$, i.e. if the table contains some value in every row in the column corresponding to y. A many-valued context is called complete if each of its attributes is complete.

(3) From the point of view of theory of relational databases, a complete many-valued context is essentially a relation over a relation scheme Y, see [14]. Namely, each $y \in Y$ can be considered an attribute in the sense of relational databases and putting

$$D_y = \{w \mid \langle x, y, w\rangle \in I \text{ for some } x \in X\},$$

D_y is a domain for y. In this paper, we consider only complete many-valued contexts.

(4) In what follows, we prefer term data table (with general attributes) over the term many-valued context.

B. Scales and Plain Scaling With Fuzzy Attributes

In this section, we introduce scaling using fuzzy attributes.

Definition 2: A scale (or, $\mathbf{L}$-scale) for attribute $y \in Y$ is a data table $\mathbb{S}_y = \langle X_y, Y_y, I_y\rangle$ with fuzzy attributes (formal fuzzy context) such that $D_y \subseteq X_y$. Objects $w \in X_y$ are called scale values, attributes of Y_y are called scale attributes.

Remark 2: The concept of a scale can be seen a particular case of Zadeh's concept of a linguistic variable. Zadeh's linguistic variable is defined as quintuple $\langle \chi, T, U, G, \sigma\rangle$ in which χ is a name of the linguistic variable, T denotes a set of terms of χ (syntactic values), U is a universe, G is a syntactic rule (usually a grammar) which generates terms of χ, and σ is a semantic rule associating with each term X its meaning $\sigma(X)$, which is an $\mathbf{L}$-set over a universe U. We can simplify the notion of a linguistic variable by removing the syntactic rule. What remains is a quadruple $\langle \chi, T, U, \sigma\rangle$.

A scale $\mathbb{S}_y = \langle X_y, Y_y, I_y\rangle$ can then be considered as (a simplified) linguistic variable $\langle y, Y_y, X_y, \sigma\rangle$, where $(\sigma(z))(w) = I_y(w, z)$ for $z \in Y_y$ and $w \in X_y$. That is, scale attributes are considered as terms and scale values are considered as elements of the universe of the linguistic variable.

Example 1: Consider the data table with general attributes depicted in Tab. II. A set X consists of Alice, ..., George, a set Y consists of age, height, and symptom, W is a set containing all of the table entries, and we have $\langle$ Alice, age, 23$\rangle \in I$, etc.

An example of a scale for age is depicted in Tab. V. Here, $X_{\text{age}} = \{0, 1, \ldots, 150\}$, i.e., scale objects are numbers $0, \ldots, 150$, $Y_{\text{age}} = \{\mathrm{a}_y, \mathrm{a}_m, \mathrm{a}_o\}$, i.e., scale attributes are fuzzy attributes corresponding to linguistic terms "young", "middle", "old", and we have $I_{\text{age}}(5, \mathrm{a}_y) = 1$, etc. For simplicity, all rows corresponding to $0, 1, \ldots, 20$ of the scale are represented by a single row labeled 0–20 and the same for 21–30, etc.

An example of a scale for height is depicted in Tab. VI. Here, $X_{\text{height}} = [120, 200]$, $Y_{\text{height}} = \{\mathrm{h}_s, \mathrm{h}_m, \mathrm{h}_t\}$, i.e., scale attributes are fuzzy attributes corresponding to linguistic terms "short", "medium", "tall", and we have $I_{\text{height}}(165, \mathrm{h}_s) = 0.5$, etc. Again, rows corresponding to values from $[120, 150]$ are represented by a single row and the like for other groups of values.

TABLE V
SCALE FOR ATTRIBUTE AGE FOR DATA TABLE FROM TAB. II.

	a_y	a_m	a_o
0–20	1	0	0
21–30	1	0.5	0
31–40	0.5	1	0
41–50	0	1	0.5
51–60	0	0.5	1
61–150	0	0	1

TABLE VI
SCALE FOR ATTRIBUTE HEIGHT FOR DATA TABLE FROM TAB. II.

	h_s	h_m	h_t
120–150	1	0	0
151–160	1	0.5	0
161–170	0.5	1	0
171–180	0	0.5	0.5
181–200	0	0	1

It is often the case that domains D_y (and the scale objects X_y) for some attributes come naturally equipped with similarity relations. That is to say, we naturally consider some values $v, w \in X_y$ similar to each other, some not. This pertains in particular to numerical domains such as age or height. In fact, similarity is a matter of degree and an appropriate approach to capture this intuition is to consider sets X_y of objects equipped with similarity relations $\approx_y$, i.e. binary fuzzy relations $\approx_y$: $X_y \times X_y \to L$ such that degree

$$v \approx_y w$$

from L is interpreted as a degree to which v is similar to w. One should require $\approx_y$ to be reflexive, i.e., satisfying $w \approx_y w = 1$ for each $w \in X_y$, symmetric, i.e., satisfying $v \approx_y w = w \approx_y v$ for each $v, w \in X_y$, and perhaps also transitive, i.e. satisfying $(u \approx_y v) \otimes (v \approx_y w) \leq (u \approx_y w)$ for each $u, v, w \in X_y$. A criterion for a scale to be reasonable can then be defined as follows.

Definition 3: Scale $\mathbb{S}_y$ is called admissible w.r.t. $\approx_y$ if for each values $w_1, w_2 \in X_y$ and scale attribute $z \in Y_y$ we have

$$(w_1 \approx_y w_2) \otimes I_y(w_1, z) \leq I_y(w_2, z).$$

Remark 3: Admissibility means: if w_1 and w_2 are similar and a scale attribute z applies to value w_1 then z applies to w_2 as well. In a sense, an admissible scale respects similarity relation $\approx_y$.

Example 2: Consider fuzzy relations $\approx_y$ defined on X_y for y being both age and height by rule

$$w_1 \approx_y w_2 = \begin{cases} 1 & \text{if } w_1 = w_2, \\ 0.5 & \text{if } 0 < |w_1 - w_2| < 5, \\ 0 & \text{otherwise} \end{cases}$$

One can check that both of the scales from Tab. V and Tab. VI are admissible w.r.t. $\approx_y$.

It is our contention that some ordinary crisp scales appear to be unnatural and problematic simply just because they ignore the underlying similarity on X_y (similarity is not considered in ordinary scaling). For instance, in case of the ordinary scale which is used to transform Tab. II to Tab. III, Boris has scale attribute young while Cyril who is only 1 year older is considered middle aged but not young. A series of questions arises like why to separate Boris and Cyril by their ages when their ages are very similar? This kind of arbitrariness is, in our opinion, an apparent disadvantage of the ordinary concept of scale and scaling.

Example 3: A practical consequence of what we just described is the following. For a formal concept $\langle A_1, B_1\rangle$ of the concept lattice corresponding to Tab. III which is generated by a_y, i.e., $A_1 = \{\mathrm{a}_y\}^{\downarrow}$, we have

$$A_1 = \{\text{Alice, Boris, Ellen, George}\},$$

i.e. the formal concept does not apply to Cyril at all. On the other hand, for a formal concept $\langle A_2, B_2\rangle$ which is generated by a_m, i.e., $A_2 = \{\mathrm{a}_m\}^{\downarrow}$, we have

$$A_2 = \{\text{Cyril, David}\},$$

i.e. the formal concept does not apply to Boris at all. That is, these formal concepts completely separate Boris and Cyril.

Therefore, admissible scales seem to capture the intuitive requirement to take the underlying similarities $\approx_y$ into account.

Given scales for a data table with general attributes, we can transform the data table into a table with fuzzy attributes. The following definition says how to do it.

Definition 4 (plain scaling): For a data table $\mathcal{D} = \langle X, Y, W, I\rangle$ (as above), scales $\mathbb{S}_y$ ($y \in Y$), the derived table with fuzzy attributes (w.r.t. plain scaling) is a table $\langle X, Z, J\rangle$ with fuzzy attributes defined by

– $Z = \bigcup_{y \in Y}\{y\} \times Y_y$,

– $J(x, \langle y, z\rangle) = I_y(w, z)$ for $y(x) = w$.

Denote by $\mathcal{B}(\mathcal{D}, \mathbb{S})$, where for $\mathcal{D} = \langle X, Y, W, I\rangle$ and $\mathbb{S} = \{\mathbb{S}_y \,|\, y \in Y\}$, the fuzzy concept lattice corresponding to the derived table, i.e.,

$$\mathcal{B}(\mathcal{D}, \mathbb{S}) = \mathcal{B}(X, Z, J),$$

where $\langle X, Z, J\rangle$ is the table with fuzzy attributes derived from $\mathcal{D}$ and $\mathbb{S}$.

Example 4: A derived table corresponding to table with general attributes from Tab. II, scales from Tab. V and Tab. VI, and a trivial scale for attribute symptom, is just the table with fuzzy attributes shown in Tab. IV.

Example 5: We are going to demonstrate that the undesirable effect of separation of Boris and Cyril by both formal concepts generated by scale attributes corresponding to "young" and "middle" is not present when we use scaling to fuzzy attributes, cf. Example 3. Namely, for a formal fuzzy concept $\langle A_1, B_1\rangle$ of the concept lattice corresponding to Tab. IV which is generated by a_y, i.e., $A_1 = \{{}^1/\mathrm{a}_y\}^{\downarrow}$, we have

$$A_1 = \{\text{Alice, Boris, } {}^{0.5}/\text{Cyril, Ellen, George}\},$$

i.e. the formal concept partially covers Cyril. On the other hand, for a formal concept $\langle A_2, B_2\rangle$ which is generated by a_m, i.e., $A_2 = \{{}^1/\mathrm{a}_m\}^{\downarrow}$, we have

$$A_2 = \{{}^{0.5}/\text{Alice, } {}^{0.5}/\text{Boris, Cyril, David, } {}^{0.5}/\text{Ellen, } {}^{0.5}/\text{George}\},$$

i.e. the formal concept partially covers Boris (and also Alice, Ellen, and George).

The next example shows that scaling using fuzzy attributes is beneficial from the point of attribute dependencies.

Example 6: Consider data table in Tab. IV and a fuzzy attribute implication

$$\{{}^{0.5}/\mathrm{a}_y, {}^{0.5}/\mathrm{a}_m, {}^{0.5}/\mathrm{h}_s, {}^{0.5}/\mathrm{h}_m\} \Rightarrow \{\mathrm{sym}\}. \quad (5)$$

We refer a reader to [6] for an overview on fuzzy attribute implications. Without going to details on semantics of fuzzy attribute implications, one can intuitively see that (5) is true in degree 1 in the data from Tab. IV (this is the case when we use globalization in the definition of truth degree of fuzzy attribute implications). This says that, in the data, a person who is young to degree at least 0.5, middle-aged to degree at least 0.5, short to degree at least 0.5, and medium-high to degree at least 0.5, has the symptom. That is, a person who is in between young and middle-aged and in between short and medium-high, has the symptom.

If we use scaling with crisp attributes, i.e., one gets the data from Tab. III, and use ordinary attribute implications, the situation is different. Namely, the following ordinary attribute implications which are related to (5) can be considered:

$$\{\mathrm{a}_m, \mathrm{h}_m\} \Rightarrow \{\text{symptom}\}, \quad (6)$$

$$\{\mathrm{a}_y, \mathrm{h}_s\} \Rightarrow \{\text{symptom}\}, \quad (7)$$

$$\{\mathrm{a}_y, \mathrm{h}_m\} \Rightarrow \{\text{symptom}\}, \quad (8)$$

$$\{\mathrm{a}_m, \mathrm{h}_s\} \Rightarrow \{\text{symptom}\}. \quad (9)$$

One can easily see that both (6) and (7) are true, but (8) and (9) are not true in Tab. III. One difference w.r.t. (5) is that in scaling with fuzzy attributes, one has a single fuzzy attribute implication, while with ordinary attributes we have two attribute implications describing the dependency of symptom on age and height. Note also that both (6) and (7) are supported just by a single row in Tab. III and are, therefore, too specific. More importantly, however, if one considers the original data from Tab. II, our contention is that (5) naturally captures the dependency of symptom on age and height.

Note that due to restricted scope of this paper, we do not list all formal concepts of the fuzzy concept lattice corresponding to Tab. IV, nor do we present its Hasse diagram. Likewise, we do not list all fuzzy attribute implications. More detailed examples will be present in upcoming papers.

IV. Sensitivity issues in scaling: a theoretical insight

One of the above-mentioned disadvantages of ordinary scaling is that it is very sensitive to user's selection of scale attributes. A very small difference in the definition of scale attributes may lead to a large difference in the resulting concept lattices. For instance, if we define attribute a_y ("young") as applying to ages from $[0, 31]$ instead of $[0, 30]$, the concepts of the resulting concept lattice will sharply change in that some objects disappear from some concepts and will appear in other concepts (e.g., Cyril disappears from a formal concept generated by "middle" and will appear in a formal concept generated by "young"). This is not desirable. Where shall a user draw a line between "young" and "middle"? This is a question we can not get rid of when using ordinary scaling.

We argued above that this effect can be mitigated when scaling with fuzzy attributes is used. However, a problem regarding the arbitrariness of boundaries, which a user defines for scale attributes, remains. The boundaries are now membership functions of scale attributes which are now fuzzy attributes. The basic question is: What happens if instead of scale $\mathbb{S}^1_y = \langle X_y, Y_y, I^1_y\rangle$, a user selects scale $\mathbb{S}^2_y = \langle X_y, Y_y, I^2_y\rangle$ which has a similar membership function, i.e., when $I^1_y(x, y)$ is close to $I^2_y(w, z)$ for any $w \in X_y$ and $z \in Y_y$? Suppose we have two sets of scales, say $\mathbb{S}^1 = \{\mathbb{S}^1_y \,|\, y \in Y\}$ and $\mathbb{S}^2 = \{\mathbb{S}^2_y \,|\, y \in Y\}$, such that $\mathbb{S}^1_y$ is similar to $\mathbb{S}^2_y$ for each $y \in Y$. Is it true, then, that the resulting concept lattices $\mathcal{B}(\mathcal{D}, \mathbb{S}^1)$ and $\mathcal{B}(\mathcal{D}, \mathbb{S}^2)$ are similar in some natural way of measuring similarity of concept lattices? In what follows, we are going to provide a positive answer to this question.

Let us first introduce a degree $\mathbb{S}^1_y \approx \mathbb{S}^2_y$ to which the scales $\mathbb{S}^1_y$ and $\mathbb{S}^2_y$ are similar.

Definition 5: For scales $\mathbb{S}^1_y = \langle D_y, Y_y, I^1_y\rangle$, $\mathbb{S}^2_y = \langle D_y, Y_y, I^2_y\rangle$ put

$$(\mathbb{S}^1_y \approx \mathbb{S}^2_y) = \textstyle\bigwedge_{w \in D_y, z \in Y_y} I^1_y(w, z) \leftrightarrow I^2_y(w, z).$$

That is, $\mathbb{S}^1_y \approx \mathbb{S}^2_y$ is a degree to which the membership functions of scale attributes z in $\mathbb{S}^1_y$ and $\mathbb{S}^2_y$ are similar if truth function $\leftrightarrow$ fuzzy equivalence is chosen to assess similarity of membership degrees.

For $\mathbb{S}^1_y = \langle D_y, Y_y, I^1_y \rangle$, $\mathbb{S}^2_y = \langle D_y, Y_y, I^2_y \rangle$ ($y \in Y$), and $\mathbb{S}^1 = \{\mathbb{S}^1_y \,|\, y \in Y\}$, $\mathbb{S}^2 = \{\mathbb{S}^2_y \,|\, y \in Y\}$, put

$$(\mathbb{S}^1 \approx \mathbb{S}^2) = \textstyle\bigwedge_{y \in Y} (\mathbb{S}^1_y \approx \mathbb{S}^2_y).$$

That is, $\mathbb{S}^1 \approx \mathbb{S}^2$ is a degree to which the corresponding scales from $\mathbb{S}^1$ and $\mathbb{S}^2$ are similar.

Now, we recall a definition of a degree of similarity between two fuzzy concept lattices. We proceed just for the case of similarity defined over extents (the case of intents is dual), see [1], [2].

$$\begin{aligned} &(\mathcal{B}(\mathcal{D}, \mathbb{S}^1) \approx \mathcal{B}(\mathcal{D}, \mathbb{S}^2)) = \\ = \quad &\textstyle\bigwedge_{\langle A_1, B_1 \rangle \in \mathcal{B}(\mathcal{D}, \mathbb{S}^1)} \bigvee_{\langle A_2, B_2 \rangle \in \mathcal{B}(\mathcal{D}, \mathbb{S}^2)} (A_1 \approx A_2) \wedge \\ &\textstyle\bigwedge_{\langle A_2, B_2 \rangle \in \mathcal{B}(\mathcal{D}, \mathbb{S}^2)} \bigvee_{\langle A_1, B_1 \rangle \in \mathcal{B}(\mathcal{D}, \mathbb{S}^1)} (A_1 \approx A_2). \end{aligned}$$

That is, $(\mathcal{B}(\mathcal{D}, \mathbb{S}^1) \approx \mathcal{B}(\mathcal{D}, \mathbb{S}^2))$ is a degree to which for each formal concept $\langle A_1, B_1 \rangle$ from $(\mathcal{B}(\mathcal{D}, \mathbb{S}^1)$ there exists a similar formal concept $\langle A_2, B_2 \rangle$ from $\mathcal{B}(\mathcal{D}, \mathbb{S}^2))$ and *vice versa*, when similarity of $\langle A_1, B_1 \rangle$ to $\langle A_2, B_2 \rangle$ is measured by similarity of their extents A_1 and A_2.

Then we get the following theorem.

Theorem 6: For a data table $\mathcal{D} = \langle X, Y, W, I \rangle$ (as above), scales $\mathbb{S}^1_y = \langle D_y, Y_y, I^1_y \rangle$, $\mathbb{S}^2_y = \langle D_y, Y_y, I^2_y \rangle$ ($y \in Y$), and $\mathbb{S}^1 = \{\mathbb{S}^1_y \,|\, y \in Y\}$, $\mathbb{S}^2 = \{\mathbb{S}^2_y \,|\, y \in Y\}$ we have

$$(\mathbb{S}^1 \approx \mathbb{S}^2) \leq (\mathcal{B}(\mathcal{D}, \mathbb{S}^1) \approx \mathcal{B}(\mathcal{D}, \mathbb{S}^2)).$$

Proof: Due to the limited scope, we omit the proof. ■

Theorem 6 says: "if scales $\mathbb{S}_1$ and $\mathbb{S}_2$ are similar then the resulting concept lattices $\mathcal{B}(\mathcal{D}, \mathbb{S}^1)$ and $\mathcal{B}(\mathcal{D}, \mathbb{S}^2))$ are similar too". Note that this is the exact meaning of Theorem 6 (one can see this by invoking basic rules of semantics of fuzzy logic). Therefore, small changes in definition of a scale lead to small changes in the resulting concept lattices which is a desirable property.

Note that a more general theorem can be obtained for the case when the corresponding scales do not have the same numbers of scale attributes (to be presented in the full version of this paper).

Another useful result, which we are now going to present concerns the concept of admissibility of a scale w.r.t. similarity relations $\approx_y$. Suppose we have $\approx_y$ for each $y \in Y$. Note that if we do not want to consider a similarity on domain D_y, we may take the identity relation for $\approx_y$. Using $\approx_y$'s, one can define a degree $x_1 \approx_X x_2$ of similarity of objects $x_1, x_2 \in X$ for a given table $\langle X, Y, W, I \rangle$ by

$$(x_1 \approx_X x_2) = \textstyle\bigwedge_{y \in Y} y(x_1) \approx_y y(x_2).$$

That is, $x_1 \approx_X x_2$ is a degree to which for every attribute $y \in Y$, the value of x_1 on attribute y is similar to the value of x_2 on attribute y.

Then, one can prove the following theorem.

Theorem 7: Consider a data table $\mathcal{D} = \langle X, Y, W, I \rangle$ (as above), fuzzy relations $\approx_y$ on X_y ($y \in Y$) which are reflexive, symmetric, and transitive, and a collection of scales $\mathbb{S} = \{\mathbb{S}_y \,|\, y \in Y\}$ which are admissible w.r.t. $\approx_y$'s. Then for each formal concept $\langle A, B \rangle$ from $\mathcal{B}(\mathcal{D}, \mathbb{S})$ and any objects $x_1, x_2 \in X$ we have

$$A(x_1) \otimes (x_1 \approx_X x_2) \leq A(x_2).$$

Proof: Due to the limited scope, we omit the proof. ■

Note that Theorem 7 can be read as follows: If two objects are similar, they will not get separated by any formal concept of the derived table.

V. Conclusions and Future Research

We presented an approach to scaling using fuzzy attributes and argued that such scaling overcomes some problematic aspects of scaling using ordinary attributes. Our future research will focus on the following problems:

- Further study of scaling using fuzzy attributes (definition of standard scales, further study of similarity issues).
- Further study of scales in the sense of Pollandt [15]. Note that Pollandt developed a radically different approach to scaling within the framework of fuzzy concept lattices.
- Study of scaling using fuzzy attributes and its connections to fuzzy attribute implications.

Acknowledgment

Supported by grant No. 1ET101370417 of GA AV ČR, grant No. 201/02/P076 of the Czech Science Foundation, and by institutional support, research plan MSM 6198959214.

References

[1] Belohlavek R.: Similarity relations in concept lattices. *J. Logic and Computation* Vol. **10** No. **6**(2000), 823–845.
[2] Belohlavek R.: *Fuzzy Relational Systems: Foundations and Principles*. Kluwer, Academic/Plenum Publishers, New York, 2002.
[3] Belohlavek R.: Concept lattices and order in fuzzy logic. *Ann. Pure Appl. Logic* **128**(2004), 277–298.
[4] Belohlavek R.: Algorithms for fuzzy concept lattices. *Proc. Fourth Int. Conf. on Recent Advances in Soft Computing*. Nottingham, United Kingdom, 12–13 December, 2002, pp. 200–205.
[5] Belohlavek R., Vychodil V.: Reducing the size of fuzzy concept lattices by hedges. In: FUZZ-IEEE 2005, The IEEE International Conference on Fuzzy Systems, May 22–25, 2005, Reno (Nevada, USA), pp. 663–668.
[6] Belohlavek R., Vychodil V.: Attribute implications in a fuzzy setting. In: Missaoui R., Schmid J. (Eds.): ICFCA 2006, Lecture Notes in Artificial Intelligence **3874**, pp. 45-60, Springer-Verlag, Berlin/Heidelberg, 2006.
[7] Carpineto C., Romano G.: *Concept Data Analysis. Theory and Applications.* J. Wiley, 2004.
[8] Ganter B., Wille R.: Conceptual scaling. In: Roberts F. S. (Ed.): *Applications of combinatorics and graph theory to the biological and social sciences.* Springer–Verlag, Berlin–Heidelberg–New York, 1989, pp. 139–167.
[9] Ganter B., Wille R.: *Formal Concept Analysis. Mathematical Foundations.* Springer-Verlag, Berlin, 1999.
[10] Hájek P.: *Metamathematics of Fuzzy Logic.* Kluwer, Dordrecht, 1998.
[11] Hájek P.: On very true. *Fuzzy sets and systems* **124**(2001), 329–333.
[12] Klir G. J., Yuan B.: *Fuzzy Sets and Fuzzy Logic. Theory and Applications.* Prentice Hall, 1995.
[13] Krajci S.: A generalized concept lattice. *Logic Journal of IGPL* **13**(2005), 543–550.
[14] Maier D.: *The Theory of Relational Databases.* Computer Science Press, Rockville, 1983.
[15] Pollandt S.: *Fuzzy Begriffe.* Springer-Verlag, Berlin/Heidelberg, 1997.
[16] Wolff K. E.: Concepts in fuzzy scaling theory: order and granularity. *Fuzzy Sets and Systems* **132**(2002), 63–75.
[17] Zadeh, L. A.: Towards the theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic. *Fuzzy Sets and Systems* **90**(2)(1997), 111–127.