Chapter #

VISUAL DATA MINING WITH SIMULTANEOUS RESCALING

Evgeni Vityaev[1] and Boris Kovalerchuk[2]
[1] *Institute of Mathematics, Russian Academy of Sciences, Russia,*
[2]*Central Washington University,USA*

Abstract: Visualization is used in data mining for the visual presentation of already discovered patterns and for discovering new patterns visually. Success in both tasks depends on the ability of presenting abstract patterns as simple visual patterns. Getting simple visualizations for complex abstract patterns is an especially challenging problem. A new approach called inverse visualization (IV) is suggested for addressing the problem of visualizing complex patterns. The approach is based on specially designed data preprocessing. Preprocessing based on a transformation theorem is proved in this chapter. A mathematical formalism is derived from the Representative Measurement Theory. The possibility of solving inverse visualization tasks is illustrated on functional non-linear additive dependencies. The approach is called inverse visualization because it does not use data "as is" and does not follow the traditional sequence: discover pattern → visualize pattern. The new sequence is: convert data to a visualizable form → discover patterns with predefined visualization.

Key words: Visual data mining, simultaneous scaling, non-linear dependency, data pre processing, reverse visualization.

1. INTRODUCTION

Visual data mining is a growing area of research and applications [Keim, 2001; Fayyad, Grinstein & Wierse, 2001; Spence, 2001; Ware, 2000; Mille, 2001; Soukup & Davidson, 2002]. It includes two visually related tasks: the visual discovery of patterns and the visualization of discovered patterns in a specific form. This visual form should be *perceivable*, *understandable* and

interpretable in a domain. It is often formulated that the visualization should be simple - that is judged afterwards informally.

The most promising visual discovery approach is a *heterogeneous approach* that combines (a) analytical manipulation (AM) with data to transform them and (b) pure visual discovery (VD) by interactively observing transformed data.

In this chapter we attempt to formalize the concept of a *simple visualization* for data mining using ideas from classical physics, specifically from the theory of physical structures [Kulakov, 1971]. The next goal is to develop an AM technique that can provide a simple visualization. The chapter concludes with a simulation example showing the application of the AM technique on data.

Traditional *visualization* generally follows the sequence:

<patterns> → <visualization>.

In contrast, *visual discovery* reverses the sequence:

<visualization> → <patterns>.

That is, in visualization, we start from patterns and produce a visualization while in visual discovery, we start from visualization and produce patterns.

Thus, we call **visual discovery** an **Inverse Visualization Task** (**IVT**) with the goal to *find data transformations that permit the generation of a simple, clear visualization of data and patterns*. Success in this endeavor depends on the properties of the data transformations and the data mining methods involved. Many data mining practitioners share the opinion that practically any data mining method will discover meaningful patterns for "good" data while few if any will produce meaningful patterns for "poor" data.

One of the goals of IVT is to transform "poor" data into "good" data thus permitting a wide variety of data mining methods to be used for the successful discovery of hidden patterns.

For now, we will not attempt to define formally "poor" and "good" data. Rather, we will show that in classical physics such transformations have been used successfully for a long time to discover patterns, which are now classical (fundamental) physical laws, without formal definitions of "good" data. We note that the laws of classical physics are simple so the problem of their visualization is not difficult.

However, the lessons learned from classical physics can help in other domains where patterns do not appear to be simple, but first we need to understand the reasons for the simplicity of laws in classical physics. Are these

reasons specific to physics or can they be exploited for domains such as finance, medicine, remote sensing, and image analysis?

An explanation of simplicity in physics follows from two theories: the Representative Measurement Theory [Krantz, Luce, Suppes & Tversky 1990] and the Physical Structures Theory [Kulakov, 1971; Mikhailichenko, 1985]. Measurement theory [Krantz et al., 1990, v.1] demonstrates that a system of physical quantities and fundamental laws will have a simple representation because they are obtained through a procedure that *simultaneously scales* the variables involved in the laws.

Traditionally data mining *does not involve simultaneous scaling.* Note that simultaneous scaling is different from the data *normalization* procedures used in neural networks to speed up search, see for example [Rao & Rao, 1995]. The typical normalization in neural networks transforms the scale of each variable *independently* and non-linearly to some interval, such as [-1,1]. On the other hand, simultaneous scaling of variables x, y and z might transform these variables into new scales x', y' and z' so that the law has the simple linear form, perhaps $y'= x'+z'$. In general, *laws of classical physics show that if all variables included in a law are scaled simultaneously then the law can assume a relatively simple form.*

The problem of finding efficient, simultaneous scaling transformations was not posed and solved by Representative Measurement Theory. This theory explains the simplicity effect but lacks a constructive way to achieve it. On the other hand, Representative Measurement Theory has wider area of application than physics only. For instance, psychology has benefited significantly from it [Krantz et al., 1971]. This observation raises a hope that simultaneous scaling will be beneficial in other areas too. This, of course, requires designing simultaneous scaling transformations.

Fortunately, the theory of Physical Structures provides an answer for this problem via the *constructive classification* of all functional expressions of all possible fundamental physical laws [Mikhailichenko, 1985]. Classes defined by this classification have an important property -- any other functional expressions of a physical law can be transformed to one of the given classes by a monotone transformation of all involved variables.

The procedure for deriving such transformation is the *simultaneous scaling* of these variables. This result shows, that every physical law can be described as class of expressions that can be converted to each other by monotone transformations of the variables contained in the law. This means that all laws can be enumerated in the classification from of all functional expressions of all possible fundamental physical laws [Mikhailichenko, 1985]. All laws of this classification have simple form and by extension, the problem of their visualization is simple too. All **complexity of visualization** of

these laws is thus converted into the design of a **monotone transformation** of the *n*-tuples of variables involved.

2. DEFINITIONS

Let us define a class of functions F, which can be transformed to the linear function $y = x + z$ by monotone transformations. There are many possible ways to define the class F. It is convenient to assume that F is given through a set of axioms. Suppose that a dataset D from a specific domain (e.g., finance) represents a set of triples x, y, and z where $y = f(x, z)$ and the function f is not known analytically. Suppose that the function f is known only through tabulated values from D and possibly some meaningful (for the domain) properties in the form of axioms. We assume that:

- real-world variables x, y and z are mapped to real numbers $\boldsymbol{R}$ by some measurement procedures,
- the order relation on $\boldsymbol{R}$ is not just a numeric relation but it has interpretation as a real-world relation for the variable y, and
- in the same way the equality relation on $\boldsymbol{R}$ is interpreted for variables x and z.

We define the class F of functions $f \in F$ on $X_f{\times}Z_f$, where $X_f \subset \boldsymbol{R}$, $X_f \neq \varnothing$, and $Z_f \subset \boldsymbol{R}$, $Z_f \neq \varnothing$. Functions from F satisfy the five properties of *additive conjoint structure* [Krantz et al., 1971, p.256]:

(1). $\forall\, z_1, z_2, \ \exists\, x \ \ (f(x, z_1) \geq f(x, z_2) \Rightarrow \forall\, x' \ (f(x', z_1) \geq f(x', z_2)) \ )$

(2). $\forall\, x_1, x_2, x_3, z_1, z_2, z_3$
$(f(x_1, z_2) = f(x_2, z_1))\ \&\ (f(x_1, z_3) = f(x_3, z_1)) \ \Rightarrow (f(x_2, z_3) = f(x_3, z_2))$

(3). For any three of x_1, x_2, z_1, z_2 the fourth of them exists such that
$f(x_1, z_2) = f(x_2, z_1)$

(4). $\exists\, x_1, x_2, z \ (f(x_1, z) \neq f(x_2, z))$

(5). For any z_1, z_2 : $z_1 \neq z_2$, if a sequence $x_1, x_2, \ldots, x_i, \ldots$ of elements of X_f is determined and satisfies the following properties: $\forall i,\ x_i < x_{max}$
$f(x_1, z_1) = f(x_2, z_2),\ f(x_2, z_1) = f(x_3, z_2),\ f(x_3, z_1) = f(x_4, z_2), \ldots,$
$f(x_i, z_1) = f(x_{(i+1)}, z_2), \ldots$
then this sequence is finite.

In addition properties (2) and (3) should also take place with x replaced by z and vice versa.

3. THEOREM ON SILMULTANEOUS SCALING

The theorem below is based on axioms (1)-(5) and is used for design of a simultaneous scaling procedure.

Theorem [Krantz et al., 1971, p.257]**:**

1. For any function $f \in F$ there are one-to-one functions φ_x, φ_z and a monotone function φ such that
   $\varphi f(x, z) = \varphi_x(x) + \varphi_z(z)$, $<x, z> \in X_f \times Z_f$,.

2. If $\varphi`(x)$, $\varphi`(z)$ are two other functions with the same property, then there exist constants $\alpha > 0$, β_1, and β_2, such that
   $\varphi`_x(x) = \alpha\varphi_x(x) + \beta_1$, $\varphi_z(z) = \alpha\varphi`_z(z) + \beta_2$
   $f'(x', z') = \varphi f(\varphi_x(x'), \varphi_z(z'))$
   where φ is a strictly monotone function, and φ_x, φ_z are one-to-one functions from F.

Proof [Krantz et al., 1971, p.264-266]. The proof follows from the fact, that the set axioms (1) – (5) represents an additive conjoint structure.

Let function $f \in F$ on $X_f \times Z_f$, satisfy the axioms (1)-(5). By virtue of the axiom (4) there are points on the plane $<x_0, z_0>$, and $<x_1, z_0>$ such that $f(x_0, z_0) \neq f(x_1, z_0)$ (see Figure 1).

Rescaling algorithm. Let's simultaneously scale X, Z, and Y ($y = f(x, z)$) as follows:

- assign value 0 to x_0 of the scale X; record it as $x_0 = 0$;
- assign value 1 to x_1;
- assign value 0 to z_0 of scale Z;
- assign values $f(x_0, z_0) = 0$ and $f(x_1, z_0) = 1$ for function f.

By virtue of the axiom (3) for three elements x_0, z_0, x_1 there exists fourth z_1, such that

$$f(x_0, z_1) = f(x_1, z_0).$$

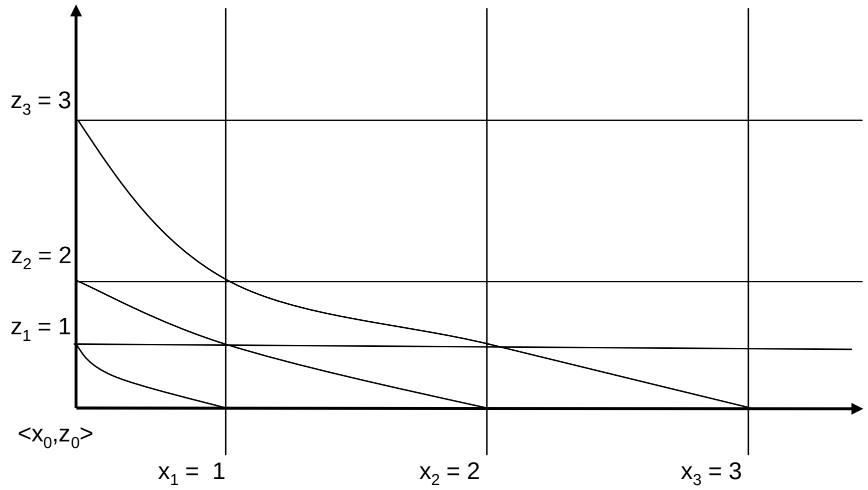


Figure 1. Simultaneous rescaling process

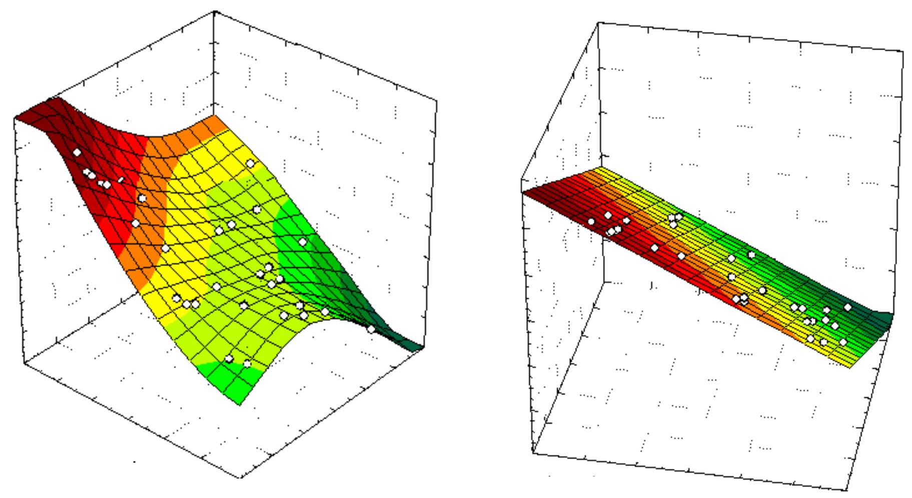
Let us link the points $<x_0, z_1>$, $<x_1, z_0>$ as shown in Figure 1. Along this line the function has identical values. These values are the values of *Y* scale (which is not shown on the picture). It is easy to see, that these values of *x*, *z*, and *y* satisfy the function $x + z = y$. We take a point $<x_1, z_1>$ and assign value $y = f(x_1, z_1) = 2$ for this point.

Next we again apply the axiom (3). At first we apply it to values x_1, x_0, z_1 and receive x_2 such that $f(x_1, z_1) = f(x_2, z_0)$ and then we apply it to values x_1, x_0, z_1 and receive z_2, such that $f(x_0, z_2) = f(x_1, z_1)$. After that we assign value $y = f(x_0, z_2) = f(x_1, z_1) = f(x_2, z_0) = 2$. Now we consider new points $<x_2, z_1>$ and $<x_1, z_2>$.

To make the given construction possible for all new points $<x_0, z_3>$, $<x_3, z_0>$ it is necessary, that the values of the function would be identical $f(x_2, z_1) = f(x_1, z_2)$ for points $<x_2, z_1>$ and $<x_1, z_2>$. The equality $f(x_2, z_1) = f(x_1, z_2)$ follows from the axiom 2.

Figures 2 and 3 present such transformation. The surface in Figure 2 is transformed to the surface in Figure 3 by the simultaneous rescaling of variables *x*, *z*, and *y*. It follows from the theorem, that if properties (1)-(5) take place for some variables *x*, *y*, *z*, then the function $f \in$ F can be converted to function $y = x + z$ by rescaling variables. After this the visualization of rescaled data and function $y = x + z$ is obvious (see Figure 3).

The rescaling algorithm requires that values of a function *f* on specific pairs of values $<x, z>$ satisfy properties (1)-(5) of the theorem. These properties are true for preference relations used in Decision Theory [Keeney & Raiffa, 1976], but this is not a universally true condition for other tasks.



(a) (b)

Figure 2. Data visualization: (a) original data, (b) simultaneously rescaled data

4. A TEST EXAMPLE

A test example must satisfy several requirements to be really convincing:

(1) It should contain regularities (patterns) known in advance;

(2) These regularities should have at least a hypothetically meaningful interpretation in the domain;

(3) These regularities should not be obvious when data is prescreened and visualized prior to rescaling as in Figure 2 (a).

Table 1, shown in section 5 below, contains data to meet these requirements. It is produced in the way described below:

- Attributes a_1, a_2, a_3 and a_5, a_6, a_7, a_8, a_9 are created by using a random number generator. For instance attributes a_1, a_2, a_3 could model some basic independent indicators of product manufacturing.
- Attribute a_4 is a sum of the first two attributes, $a_4 = a_1 + a_2$.
- Attribute a_{10} is a target attribute, it is equal to some random monotone transformation F of the difference $a_4 - a_3$, i.e.,
  $a_{10} = F(a_4 - a_3) = F(a_1 + a_2 - a_3)$.

These attributes may have different *interpretations*. In one of them attributes a_5, a_6, a_7, a_8 and a_9 represent noise. They are random and unrelated to the target attribute a_{10}. A hypothetical interpretation of the regularity $F(a_1 + a_2 - a_3)$ could be productivity or production efficiency or revenue. Attribute a_1 may indicate initial capital (scaled from 0 to 10), attribute a_2 may indicate

quality of management (also scaled from 0 to 10) and attribute a_3 could be a tax level (scaled from 0 to 10). Attributes a_1 and a_2 contribute positively to revenue while attribute a_3 contributes negatively.
A relatively complex monotone transformation is motivated by an intention:

- to solve a **realistic task**. In real-world tasks, if there are any hidden patterns (regularities), they are usually disguised and significantly corrupted. Experience shows that monotone regularities are common for many data mining tasks.
- to show **unique capabilities** of the simultaneous scaling method. Traditional methods that do not use simultaneous scaling can not discover a regularity corrupted by a random monotone transformation. The only way to do this is to analyze all interpretable order relations $\leq_1, \leq_2, \leq_3, \ldots, \leq_{10}$ for all attributes. These regularities are revealed by the simultaneous scaling method.
- to find **meaningful patterns**, regularities, and functions. For instance, typically regression analysis produces functions that just interpolate data without meaningful interpretation in the domain. In contrast the simultaneous scaling method produces meaningful regularities such as

$$\forall\, a, b\ (a \leq_1 b \ \&\ a \leq_2 b \Rightarrow a \leq_{10} b)$$

The data in Table 1 encodes the following regularities by design:

$$\begin{aligned}
&\forall\, a, b\ (a >_3 b \ \&\ a \leq_4 b \Rightarrow a \leq_{10} b)\\
&\forall\, a, b\ (a \leq_3 b \ \&\ a >_4 b \Rightarrow a >_{10} b)\\
&\forall\, a, b\ (a \leq_1 b \ \&\ a \leq_2 b \ \&\ a >_3 b \Rightarrow a \leq_{10} b)\\
&\forall\, a, b\ (a >_1 b \ \&\ a >_2 b \ \&\ a \leq_3 b \Rightarrow a >_{10} b)
\end{aligned} \tag{1}$$

5. DISCOVERING SIMULTANEOUS SCALING

The Discovery System [Kovalerchuk & Vityaev, 2000] can discover all monotone regularities including those shown in (1) above and are actually encoded in Table 1 along with random noise. When regularities (1) are discovered, a simultaneous monotone rescaling of the data can be arranged and the straightforward and simple visualization presented in Figure 3 below will be generated.

Thus the major challenge is discovering the monotone regularities. The Discovery System searches sequentially for monotone regularities starting from simplest ones:

$$\forall\, a, b\ (a \leq_i b \Rightarrow a >_{10} b),\ i = 1, \ldots, 9; \tag{2}$$

Table 1. Test data

#	1	2	3	4	5	6	7	8	9	10	#	1	2	3	4	5	6	7	8	9	10
1	8	2	8	10	5	9	0	4	1	53	*41*	1	4	3	5	7	9	6	6	7	53
2	6	4	0	10	1	1	8	1	5	79	*42*	3	8	9	11	5	6	2	4	2	53
3	4	1	3	5	0	3	8	7	1	53	*43*	9	1	5	10	1	8	0	6	3	61
4	8	0	9	8	9	1	5	0	0	30	*44*	8	6	6	14	6	2	6	7	2	72
5	7	8	4	15	5	0	2	8	6	83	*45*	0	8	9	8	9	7	2	1	1	30
6	9	1	8	10	9	7	0	3	4	53	*46*	4	4	3	8	9	2	8	2	5	61
7	9	5	5	14	3	3	9	7	7	75	*47*	9	7	8	16	4	4	5	0	9	72
8	5	8	2	13	7	1	6	6	8	83	*48*	6	4	7	10	4	7	4	9	6	56
9	1	9	4	10	0	1	4	7	1	64	*49*	5	2	4	7	7	7	7	7	3	56
10	1	2	3	3	4	0	3	2	7	31	*50*	2	2	9	4	3	6	2	5	2	12
11	7	5	5	12	6	8	0	2	4	66	*51*	1	7	8	8	5	0	6	5	6	31
12	2	1	6	3	3	8	1	7	2	17	*52*	9	1	8	10	3	6	5	5	4	53
13	4	6	5	10	6	6	7	5	4	61	*53*	8	1	5	9	5	2	7	3	1	58
14	1	5	8	6	4	9	0	9	5	24	*54*	2	7	7	9	8	0	4	8	9	53
15	7	3	9	10	1	4	5	8	2	51	*55*	3	5	6	8	5	6	2	4	2	53
16	2	6	9	8	9	5	3	5	2	30	*56*	5	5	1	10	8	7	9	1	3	75
17	1	8	9	9	6	6	3	8	2	31	*57*	3	5	7	8	8	3	3	6	7	51
18	9	3	7	12	3	9	1	2	1	61	*58*	4	5	4	9	8	6	4	8	1	61
19	0	4	4	4	4	2	0	3	4	31	*59*	2	9	4	11	4	7	4	0	1	66
20	9	1	5	10	5	4	2	6	1	61	*60*	4	1	9	5	9	0	2	8	5	14
21	8	7	6	15	1	2	7	1	1	75	*61*	9	1	5	10	4	0	7	9	2	61
22	7	9	6	16	9	6	7	3	1	79	*62*	5	8	4	13	8	6	2	5	8	75
23	8	9	2	17	1	5	6	3	9	96	*63*	5	8	6	13	6	0	5	9	9	66
24	8	2	7	10	9	4	1	1	6	56	*64*	9	2	5	11	2	1	9	7	2	64
25	0	1	7	1	6	9	0	6	6	9	*65*	5	8	8	13	3	6	2	1	7	61
26	7	7	1	14	5	0	5	2	5	91	*66*	3	7	9	10	2	1	6	4	7	51
27	2	9	7	11	2	1	1	5	3	58	*67*	8	4	8	12	8	2	4	9	0	58
28	5	5	1	10	2	4	5	8	8	75	*68*	7	0	8	7	6	0	8	1	5	30
29	1	7	0	8	6	3	6	9	7	72	*69*	4	1	6	5	9	2	8	9	7	30
30	4	6	0	10	0	9	8	2	7	79	*70*	8	6	8	16	4	8	8	2	8	72
31	0	6	1	6	3	7	8	6	1	61	*71*	1	5	5	6	2	3	7	6	8	51
32	8	4	4	12	2	6	6	4	6	72	*72*	1	3	9	4	2	4	5	6	4	12
33	3	4	6	7	7	2	1	7	6	51	*73*	7	3	0	10	6	3	7	6	6	79
34	2	7	8	9	3	3	8	4	4	51	*74*	9	0	4	9	3	8	0	0	5	61
35	6	5	9	11	7	2	4	5	5	53	*75*	5	7	4	12	1	6	8	6	8	72
36	9	3	9	12	8	5	8	6	0	56	*76*	5	1	9	6	1	4	8	8	4	17
37	1	6	2	7	0	7	6	0	8	61	*77*	3	0	5	3	7	6	5	4	1	24
38	6	0	0	6	0	2	3	7	5	64	*78*	7	7	7	14	7	9	7	5	7	66
39	2	7	9	9	0	1	6	9	8	31	*79*	0	2	9	2	5	8	5	1	8	5
40	4	8	2	12	0	9	5	1	5	79	*80*	3	4	4	7	0	9	9	7	8	56

After testing them we discover regularity

$$\forall\ a, b\ (a \leq_4 b \Rightarrow a \leq_{10} b).$$

with a statistical confidence level equal to 0.0001. This regularity is not in the list of regularities (1) above, although it is true for data from Table 1. Next the systems tests more complex regularities:

$$\forall\ a, b\ (a >_i b\ \&\ a \leq_j b \Rightarrow a \leq_{10} b) \quad i, j = 1, ..., 9$$

and finds a regularity

$$\forall\ a, b\ (a >_3 b\ \&\ a \leq_4 b \Rightarrow a \leq_{10} b)$$

with a statistical confidence level equal to 0.025.

Similarly, another parametric set of hypothetical regularities is generated and tested to discover the second regularity in (1). Next we can discover a regularity with all three variables in the antecedent if we substitute given attributes with parameters *i, j* and *k* that are the indexes of attributes. For instance, for discovering

$$\forall\ a, b\ (a \leq_1 b\ \&\ a \leq_2 b\ \&\ a >_3 b \Rightarrow a \leq_{10} b),$$

we generate a parametric set

$$\forall\ a, b\ (a \leq_i b\ \&\ a \leq_j b\ \&\ a >_k b \Rightarrow a \leq_{10} b)\ i, j, k = 1, ..., 9$$

and test it. The test reveals the needed regularity with a confidence level equal to 0.1.

6. ADDITIVE STRUCTURES IN DECISION MAKING

In section 2, we assumed an *additive conjoint structure* that permitted us to build a simple linear visualization. In this section, we discuss the motivation of using an additive structure from a decision-making viewpoint invoking an approach presented in [Keeney & Raiffa, 1976].

A decision-making problem is considered as a tradeoff between contradictory goals such as maximizing profit and minimizing risk. The tradeoff means that we try to substitute a chunk Δ_1 of goal G_1 that we cannot satisfy with a chunk Δ_2 of another goal G_2 that we can satisfy. The tradeoff assumption must be such that substitution makes sense. This is, in essence, leading

us to an additive conjoint structure assumption. Under this widely accepted assumption, the practical issue is finding the chunks Δ_1 and Δ_2.

One of the options that can be used to solve this problem is the *explicit way* where a SME (subject matter expert) declares, say, that Δ_1=3 and Δ_2=5 are equivalent for substitution purposes, that is SME formalizes preferences as a model. This is typically a very difficult task. Another approach is the *implicit* approach. In this approach, we just ask a SME to define preferences for, say, about 100 pairs of multi-criteria decisions.

A SME can say that a decision with attributes $(a_1, a_2, \ldots, a_n) = (1, 5, \ldots, 7)$ is better than a decision with attributes $(a_1, a_2, \ldots, a_n) = (3, 2, \ldots, 5)$. Alternatively, we may ask a SME to assign a priority to each $(a_1, a_2, \ldots, a_n)$ alterative using a 0 to 100 percentage scale. Table 1 can be interpreted in this way, where a_{10} can be viewed as a priority.

Both implicit alternatives provide us with a partially defined a scalar priority function, v:

$$v(x_1, \ldots, x_n) \geq v(y_1, \ldots, y_n) \Leftrightarrow (x_1, \ldots, x_n) \geq_{\mathrm{SME}} (y_1, \ldots, y_n)$$

Function $v(x_1, \ldots, x_n)$ is *additive* if $v(x_1, \ldots, x_n) = v_1(x_1) + v_2(x_2) + \ldots + v_n(x_n)$. There are several sets of axioms known for the relation $\geq_{\mathrm{SME}}$. If a set of such axioms is assumed to be true, then theorems can be proved that a numeric additive function v exists. One of these sets of axioms is presented below.

To be able to benefit from such mathematical results, we need to be able to test that the axioms are satisfied for an individual task and a dataset.

Two options are available: (1) to test axioms on a tabulated function v, such as that presented in Table 1, where $v(a_1, a_2, \ldots, a_9) = a_{10}$, and (2) to test axioms on tabulated on SME preferences $\{p_{ij}\}$, that record relations between different pairs of alternatives:

$$p_{ij} = 1 \Leftrightarrow x_i \geq_{\mathrm{SME}} x_j,$$

where $x_i = (x_{i1}, \ldots, x_{in})$ and $x_j = (x_{j1}, \ldots, x_{jn})$.

However, testing axioms provides only a conclusion that an additive function v exists, but functions $v_1(x_1), v_2(x_2), \ldots, v_n(x_n)$ would still need to be built. To build v, we need to apply a simultaneous scaling procedure to $x_1, x_2, \ldots, x_n$. Then when the functions $v_1(x_1), v_2(x_2), \ldots, v_n(x_n)$ are found we need to analyze what they mean. To do this, we can view a dataset using these functions with new axis (scales) as shown in figure 2(b). It is a linear surface. The SME can analyze how well this sum corresponds to the estimates of $v(x_1, \ldots, x_n)$. Thus visualization provides a simple representation of SME knowledge that can be communicated to another SME and tested independ-

ently. If a regularity v is multiplicative, $v(x_1, \ldots, x_n) = v_1(x_1)v_2(x_2)\ldots v_n(x_n)$ then

$$\ln v(x_1, \ldots, x_n) = \ln v_1(x_1) + \ln v_2(x_2) + \ldots + \ln v_n(x_n),$$

that is the logarithm of v is an additive function. Thus, we can use the same technique for multiplicative regularities.

7. PHYSICAL STRUCTURES

The simultaneous scaling linearization and visualization procedure described in section 5 is fully applicable to physical structures. It is based on the theorem that is a described below. In addition to this theorem, another theorem was proved by [Vityaev, 1985], which states that every physical structure of the rank (2, 2) satisfies all axioms (1)-(5) of the additive conjoint structure described in section 2.

The most general characteristic of all physical laws is that they are *equally* applicable to all objects. This fundamental property permits the derivation of the structure of physical laws by formulating fuctional equations of a special type and solving them.

Let us consider two arbitrary sets of objects: a set M with elements $i, k, \ldots$ and a set N with elements $\alpha, \beta, \ldots$. Let us further suppose that for each pair $i \in$ M, $\alpha \in$ N is mapped to a real number $a_{i\alpha} \in \mathbf{R}$ by some experiment; that is, the set M×N is mapped to the matrix A = $\| a_{i\alpha} \|$ of such numbers.

If sets M and N are two sets of physical objects of different type, then the matrix $\| a_{i\alpha} \|$ is the result of experiments that describe the relationship between objects $i \in$ M and $\alpha \in$ N.

We will say that the **physical structure of the order (r, s)** is defined on sets M and N if a functional equation:

$$\Phi\begin{pmatrix} a_{i\alpha}, a_{i\beta}, \ldots, a_{i\gamma} \\ a_{k\alpha}, a_{k\beta}, \ldots, a_{k\gamma} \\ \ldots\ldots\ldots\ldots \\ a_{q\alpha}, a_{q\beta}, \ldots, a_{q\gamma} \end{pmatrix} = 0 \tag{3}$$

is satisfied for any r • s real numbers from matrix A = $\| a_{i\alpha} \|$,

$$\begin{matrix} a_{i\alpha}, a_{i\beta}, \ldots, a_{i\gamma} \\ a_{k\alpha}, a_{k\beta}, \ldots, a_{k\gamma} \\ \ldots\ldots\ldots\ldots \\ a_{q\alpha}, a_{q\beta}, \ldots, a_{q\gamma} \end{matrix}$$

that are located on the intersection of any r rows i,k,…, q and any s columns $\alpha, \beta, \ldots, \gamma$,

The function Φ must not depend on:

(1) the selection of r objects from the set M_r, where $M_r = \{ i, k, \ldots, q \} \subset M$, and

(2) the selection of the s objects from the set N, $N_s = \{ \alpha, \beta, \ldots, \gamma \} \subset N$

We assume that the function Φ is analytical and cannot be presented as superposition of other analytical functions with a smaller number of variables.

We will say that the functional equation (3) gives us a **physical law** of order (r, s) that is invariant relative to selection of finite sets M_r and N_s from sets M and N. The equation (3) is a symbolically written, infinite system of functional equations relative to an unknown function $\Phi(x_{11}, x_{12}, \ldots, x_{rs})$ of $r \cdot s$ variables and one unknown infinite matrix $A = \| a_{i\alpha} \|$, which represent one real valued function $a_{i\alpha}$ of two nonnumeric arguments *i* and α from M and N.

Mikhailichenko [Mikhailichenko, 1972] solved this system of equations and derived the analytical expressions for all possible physical laws that satisfy the equation (3). He proved a theorem stating that functions Φ and $a_{i\alpha}$ may have only one of the following forms:

1. for $r = s = 2$ $\quad a_{i\alpha} = \Psi^{1}(x_i+\xi_\alpha)$,

   $$\Psi(a_{i\alpha}) - \Psi(a_{i\beta}) - \Psi(a_{j\alpha}) + \Psi(a_{j\beta}) = 0;$$

2. for $r = 4, s = 2$

   $$a_{i\alpha} = \Psi^{1}[(x_i\xi_\alpha^{\ 1} + \xi_\alpha^{\ 2}) / (x_i + \xi_\alpha^{\ 3})],$$

   $$\begin{vmatrix} \Psi[a_{i\alpha}] & \Psi[a_{i\beta}] & \Psi[a_{i\alpha}] & \Psi[a_{i\beta}] & 1 \\ \Psi[a_{j\alpha}] & \Psi[a_{j\beta}] & \Psi[a_{j\alpha}] & \Psi[a_{j\beta}] & 1 \\ \Psi[a_{k\alpha}] & \Psi[a_{k\beta}] & \Psi[a_{k\alpha}] & \Psi[a_{k\beta}] & 1 \\ \Psi[a_{l\alpha}] & \Psi[a_{l\beta}] & \Psi[a_{l\alpha}] & \Psi[a_{l\beta}] & 1 \end{vmatrix} = 0;$$

3. for $r = s \geq 3$

   $$a_{i\alpha} = \Psi^{1}(x_i^{\ 1}\xi_\alpha^{\ 1} + \ldots + x_i^{\ m\text{-}2}\xi_\alpha^{\ m\text{-}2} + x_i^{\ m\text{-}1}\xi_\alpha^{\ m\text{-}1}),$$

$$\begin{vmatrix} \Psi[a_{i\alpha}] & \Psi[a_{i\beta}] & \dots & \Psi[a_{i\tau}] \\ \Psi[a_{i\alpha}] & \Psi[a_{i\beta}] & \dots & \Psi[a_{i\tau}] \\ \dots & \dots & \dots & \dots \\ \Psi[a_{v\alpha}] & \Psi[a_{v\beta}] & \dots & \Psi[a_{v\tau}] \end{vmatrix} = 0;$$

and also

$a_{i\alpha} = \Psi^{1}(x_i^{1}\xi_\alpha^{1} + \dots + x_i^{m-2}\xi_\alpha^{m-2} + x_i^{m-1} + \xi_\alpha^{m-1}),$

$$\begin{vmatrix} 0 & 1 & 1 & & 1 \\ 1 & \Psi[a_{i\alpha}] & \Psi[a_{i\beta}] & \dots & \Psi[a_{i\tau}] \\ 1 & \Psi[a_{i\alpha}] & \Psi[a_{i\beta}] & \dots & \Psi[a_{i\tau}] \\ \dots & \dots & \dots & \dots & \dots \\ 1 & \Psi[a_{v\alpha}] & \Psi[a_{v\beta}] & \dots & \Psi[a_{v\tau}] \end{vmatrix} = 0;$$

4. for $r = s + 1 \geq 3$

$a_{i\alpha} = \Psi^{1}(x_i^{1}\xi_\alpha^{1} + \dots + x_i^{m-2}\xi_\alpha^{m-2} + \xi_\alpha^{m-1}),$

$$\begin{vmatrix} 1 & \Psi[a_{i\alpha}] & \Psi[a_{i\beta}] & \dots & \Psi[a_{i\tau}] \\ 1 & \Psi[a_{i\alpha}] & \Psi[a_{i\beta}] & \dots & \Psi[a_{i\tau}] \\ \dots & \dots & \dots & \dots & \dots \\ 1 & \Psi[a_{v\alpha}] & \Psi[a_{v\beta}] & \dots & \Psi[a_{v\tau}] \end{vmatrix} = 0;$$

5. for $r - s \geq 2$, except the case $r = 4$, $s = 2$ physical structures are not exists.

   Ψ - is a strictly monotone function of one variable in some vicinity;

   Ψ^{1} - is an inverse function; x_i , ξ_α - are independent parametrs.

8. CONCLUSION

Evidence that an additive conjoint structure plays a fundamental role in two very different fields such as multi-criteria decision making and fundamental physical laws provides the basis for our belief that it also can play the same fundamental role in other fields. Thus, sumultaneous rescaling has a great potential as a major tool in visual data mining for the simplification of patterns to be visualized in a variety of fields. Further study will be directred at providing computationally efficient simultaneous rescaling algorithms for multi-dimensional data.

9. EXERCISES

1. Define an additive conjoint structure for $f \in F$ on $X_f \times W_f \times Z_f$ as a generalization of such structure for $f \in F$ on $X_f \times Z_f$ presented in section 2.
2. Formulate an analogue of the theorem presented in section 3 for 3-D function $f \in F$ on $X_f \times W_f \times Z_f$
3. Develop a rescaling algorithm for 3-D similar to presented in section 3 for 2-D.
4. Advanced. Solve problems in exercises 1-3 for n-dimensional case.

10. REFERENCES

Fayyad U., Grinstein G., and Wierse A., Eds., Information Visualization in Data Mining and Knowledge Discovery, Morgan-Kaufman, 2001.

Keim D., Visual Exploration of Large Data Sets, Communication of ACM, vol. 44, N.8, 2001, pp. 39-44.

Keeney R.L., Raiffa H. Decisions with Multiple Objectives: preferences and value Tradeoffs. John Wiley& Sons, 1976

Krantz DH, Luce RD, Suppes P, Tversky A: Foundations of Measurement v.1-3, Acad. Press, NY, London. 1971, 1989, 1990.

Kuakov Yu.I. The One Principal Underlying Classical Physics, Soviet Physics – Doclagy, V.15, #7, Jan., 1971, 666-668.

Kovalerchuk, B., Vityaev, E. Data Mining in Finance: Advances in Relational and Hybrid Methods, Kluwer Acad. Publ., Boston, 2000.

Mikhailichenko G.G. Phenomenological and Group Symmetry in the Geometry of two Sets (Theory of Physical Structures), Soviet Math. Docl. 32(2), 1985, 371-374.

Mikhailichenko G.G. Solution of functional equations in the theory of physical structures, Doklady, Soviet Academy of Sciences, 1972, v 206, N.5 1056-1058.

Mille, H. (Ed) Geographic Data Mining & Knowledge Discovery, Taylor and Francis, 2001

Rao H., Rao V., C++ Neural Networks and Fuzzy Logic, Hungry Minds, Inc, 1995.

Soukup T., Davidson I., Visual Data Mining: Techniques and Tools for Data Visualization and Mining, Wiley, 2002

Spence R., Information Visualization, Addison-Wesley, 2001.

Vityaev E.E. Numerical, algebraic, and constructive representation of one physical structure. In: Logical foundations MOZ (Methods of regularities discovery), Computer systems, #107, Novosibirsk, 1985, pp.40-51 (in Russian).

Ware C. Information Visualization: perception for Design, Acad. Press, 2000.