Visualization and Decision-Making using Structural Information[1]

Boris Kovalerchuk
Department of Computer Science
Central Washington University
Ellensburg, WA, 98926-7520, USA

Abstract

Visualization and computational intelligence methods such as data mining and knowledge discovery are important tools for decision support in imaging science. The goal of this paper is to discuss the blending of these areas to improve decision making for a variety of applications. The paper offers a conceptual model for integrating decision-making models, models for discovery of relations (computational intelligence models) and methods of visual correlation. Analysts use visual correlation for discovering relations and decision-makers use them for visualizing the decision under consideration. The model reflects this difference. The blending of these models is illustrated in the example of stopping cholera in London in the 19th century. The last part of the paper is devoted to systematization of visual correlation (VC) methods and their efficiency.

1. Introduction

Computational intelligence methods such as data mining, "expert" mining (extracting patterns by interacting with experts), machine learning, inductive reasoning strategies, expert systems, fuzzy logic and neural networks discover hidden patterns. Visualization methods present them to domain experts in an appropriate form. On the other hand, visualization helps domain experts to discover hidden patterns (correlations) directly augmenting computational intelligence methods. For domain experts (analysts and decision-makers) discovering and visualization of hidden patterns is important, but it is only a part of their decision making process. The paper addresses problems of decision support by discovering and visualizing patterns based on structural information. We start from two examples on events in Aden in 2000 and in London in 1854.

Every day mass media shows many impressive visualizations of events in the World. For instance, "Time" magazine [1] provided a rich multilevel visualization of information about the attack on the USS Cole in Aden in October 2000. This visualization starts from the World and ends with an individual injured sailor, presenting six levels of visualization: (1) World, (2) region, (3) port, (4) ship, (5) hole, and (6) sailor. The visualization shows many details about USS Cole's equipment including armament, but it is not very helpful for decision making -- how to prevent such deadly attacks. There are two reasons for that: (1) decision making is not a mass media goal and (2) "rich" information is actually scarce for decision making.

Another example was described by E. Tufte [2, pp. 27-37] on cholera epidemic in London in 1854 using an original work by Dr. J. Snow [3]. This example includes several visualizations. Some of them just show growing death toll day by day in September 1854 ("within two hundred and fifty yards of the spot where Cambridge Street joins Broad Street, there were upwards of five hundred fatal attacks of cholera in ten days"). These visualizations do not help to make decisions how to stop the epidemic. Another one matched/correlated the death tolls with locations of water pumps-wells. It was extremely useful for decision-making (DM). It prompted the authority to shut down a specific pump located in the area with a high death toll. Dr. Show marked deaths from cholera on the map, along with locations of the area's 11 community water pump-wells.

2. Modeling approach

An analysis of these examples shows that the first example (Aden) lacks a *discovered relation* between the attack and attributes useful for decision making to prevent such attacks. The second example obviously has this relation discovered by Dr. Show on September 7, 1854, which allowed the Board of Guardians of St. James's Parish to make a decision to prevent a further spread of the cholera by shutting down the pump-well on September 8, 1854. The epidemic ended during two weeks.

These two examples help us to make a point on the concept of a Visualization Useful for Decision Making (*VUDM*) based on:

- *discovered relations/patterns (DRP) and*
- *a decision making model (DMM).*

The first example (Aden) does not include components DRP and DMM. The second example (London) includes both of them:

(i) Discovered relation – people used water from well d *(death)* on Broad St. more often died from cholera than people used any other wells i;

$"\ i\ (i^1d)\ D(d) > D(i),$ where $D(\ )$ is the number of dead after drinking water from a well. "There were only ten deaths in houses situated decidedly nearer to another street <Broad St.> pump ".

(ii) Decision making model – shut down a well d if the death toll of people who used this well is higher than for people who used other wells i;

$"\ i\ ((i^1d)\ D(d) > D(i) \Rightarrow Shutdown(d).$

The last DMM is very simple and people often even do not notice that the model is there. However this simple model is a result of very non-trivial Dr. Show's discovery of the relation between use of water wells and death toll [2,3]. Next we argue that two categories of the DDM models are needed:

(a) A model for a decision-maker (e.g., city managers, the board of guardians of the parish) and

(b) A model for an analyst (e.g., Dr. Show) who discovers relations for a decision-maker.

The first *model (a)* can and should be simple, similar to model (ii) above. The second model (b) can be complex enough to cover a wide range of possible decision alternatives. In the London example the model (ii) produced a single decision alternative – to shut down the pump-well d. A *model (b)* can include much more alternatives to be explored by an analyst:

(1) Restrict access of new people to the city,

(2) Restrict contacts between people in the city limit,

(3) Restrict consumption of some food,

(4) Use some medication,

(5) Restrict contacts of population with some animals,

(6) Restrict consumption of some drinks.

Actually research of Dr. Show resulted in the last alternative (specifically to restrict/prohibit consumption of water from the well d). We have no historic evidence that Dr. Show really considered all alternatives (1) -(6). It is most likely that he came to the well water alternative without a formalized model such as the model (b). Our goal is to show that if his decision-making and visualization process would be driven by a DMM with alternatives (1)-(6) then the water alternative (6) could surface naturally and be investigated. This alternative can guide an investigation (including exploratory visualizations) instead of relying on insight of such extraordinary people as Dr. Show. We illustrate the concept of model-based approach in Figure 1.
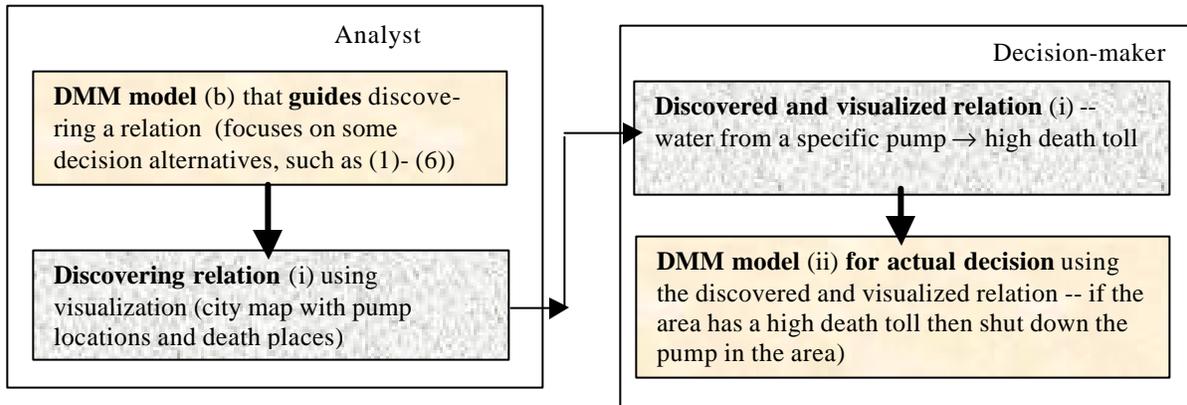
Analyst	Decision-maker
DMM model (b) that **guides** discove-ring a relation (focuses on some decision alternatives, such as (1)- (6))	**Discovered and visualized relation** (i) -- water from a specific pump → high death toll
Discovering relation (i) using visualization (city map with pump locations and death places)	**DMM model** (ii) **for actual decision** using the discovered and visualized relation -- if the area has a high death toll then shut down the pump in the area)

Figure 1. Conceptual DM model structure and visualization.
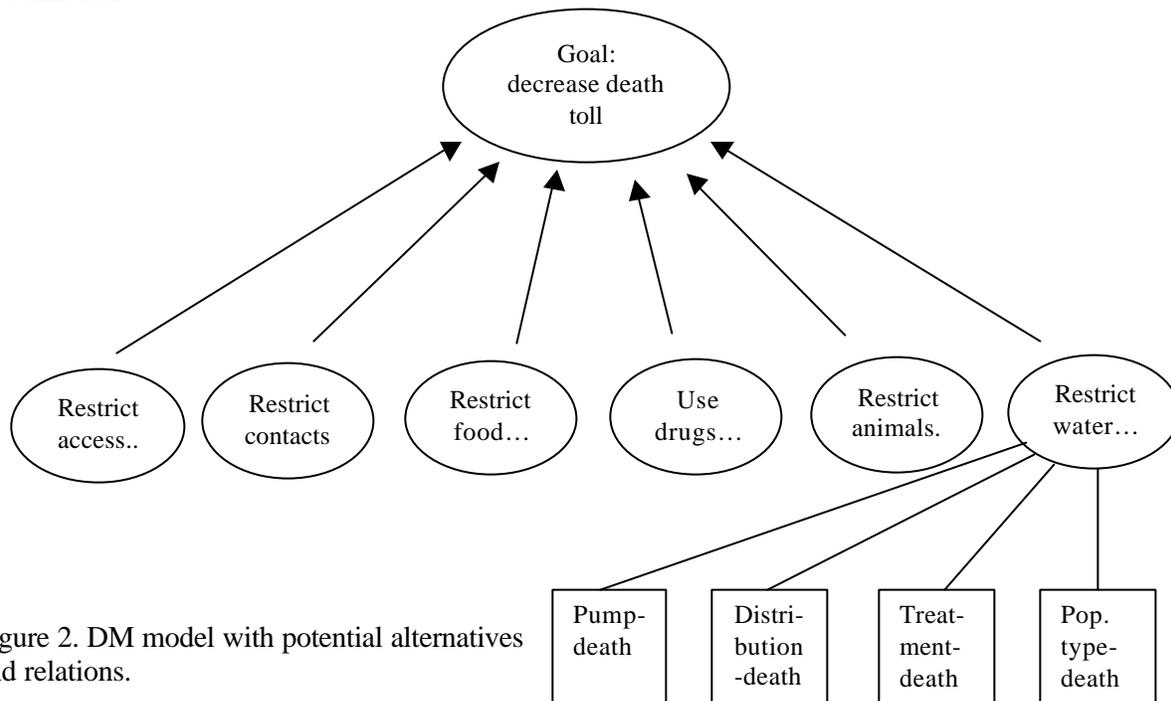

Figure 2. DM model with potential alternatives and relations.

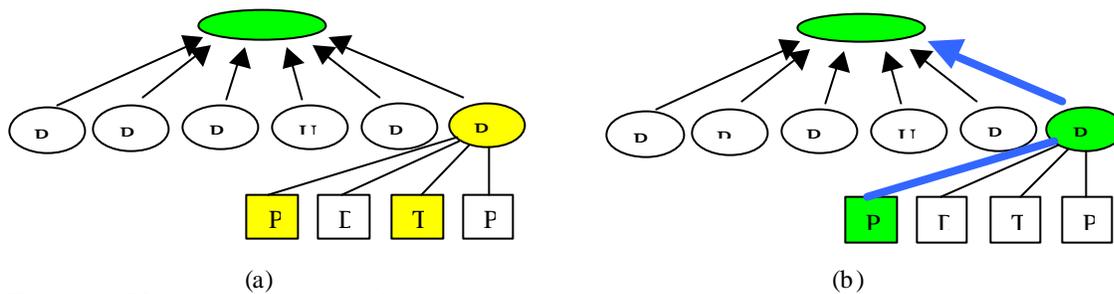(a)   (b)

Figure 3. Visual correlation in the process of discovery of relations.

3. DMM model and discovery of relations

Next we clarify connections between the model (b) with alternatives (1)-(6) and discovering relations such as (i). The DMM model (b) is not formulated in terms of any relations. It should be elaborated to include more objects and then investigate relations between them. For instance, alternative (6) can be developed to include such objects as water pumps, water distribution from the pumps, ways of water treatment, type of population. These objects may be suspected of being related to the high death toll. We call this *structured information* for DM. Thus, the DM model will grow as a tree (see Figure 2). The rectangles show relations to be investigated.

After providing such structural information an analyst can investigate relations between death toll and each of the components: pumps, distribution routes, ways of water treatment and type of population. Visualization is a natural element in this analytical process. For instance, after short investigation an analyst can come to the conclusion that water treatment such as boiling and the pump are the most related objects to the death toll. It is visualized by coloring these components (see Figure 3a). The next stage of investigation could discover a very important relation between the pump and the death toll.

Now the analyst can report the discovery to decision-makers (city managers). If they want to be sure that the analyst did not overlook some important alternatives then graphs (a) and (b) from Figure 3 provide them this information. If the decision-makers just want to consider a course of action based on the discovered relation then only a simple colored part of Figure 3(b) is needed. This part is presented on Figure 4.

Next we consider how visualization can help to discover the relation between pumps and the death toll. This is Dr. Show's classical work [2,3] (see also Figure 5 presenting the idea of Show's visualization). The idea of visualization is to bring together a city map with pumps (circles) and death locations (squares). Figure 5 shows a higher death toll around one of the

pumps. Now the analyst can report the discovery to decision-makers (city managers). If they want to be sure that the analyst did not overlook some important alternatives then graphs (a) and (b) from Figure 3 provide them this information. If the decision-makers just want to consider a course of action based on the discovered relation then only a simple colored part of Figure 3(b) is needed. This part is presented on Figure 4.
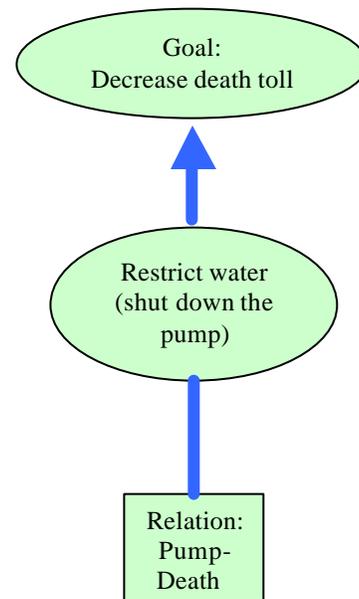

Figure 4. Final decision making model with the discovered relation: visual correlation approach.
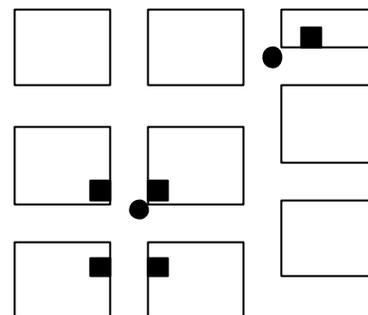

Figure 5. Mapping pumps and death toll.

This visualization was critical for discovering the relation between the death toll and pumps, but once it has been discovered many *other visualizations* can serve as well as this one for convincing decision makers that the pump has to be shut down to save lives. Figure 5 shows a simple alternative visualization. To get the death toll in Figure 6 we need to use a specific area

around each pump. Figure 6 shows this plot for distance 250 yards from pumps, other similar figures can be for 500, 1000, and 1500 yards. Obviously Figure 6 is simpler than Figure 5 especially if all 11 pumps studied by Dr. Show and all city areas with these pumps would be presented on the map. The map contains a lot of information irrelevant to the decision-making on the cholera epidemic.

Next, visualization on Figure 6 is not new for decision-makers. They are familiar with this type of plot. This is a standard plot of the relation used for checking correlation. Thus, decision-makers can concentrate on making decisions instead of studying a new way of data presentation. We did not find evidences that for final decision-making the simple visual correlation (Figure 6) has any disadvantages in comparison with maps such as presented on Figure 5 when a relation is already discovered. It looks that a simple visual correlation can serve pretty well. Moreover, it is possible that new developments of visual correlation methods for the final decision-making are not needed. Consistent use of known visual correlation methods and their combinations has obvious advantage – decision-makers know and trust them. However, the situation can be significantly different when visual correlation is needed as a tool for discovery of *unknown relations*. There are plenty of examples when only new tools made possible to discover the
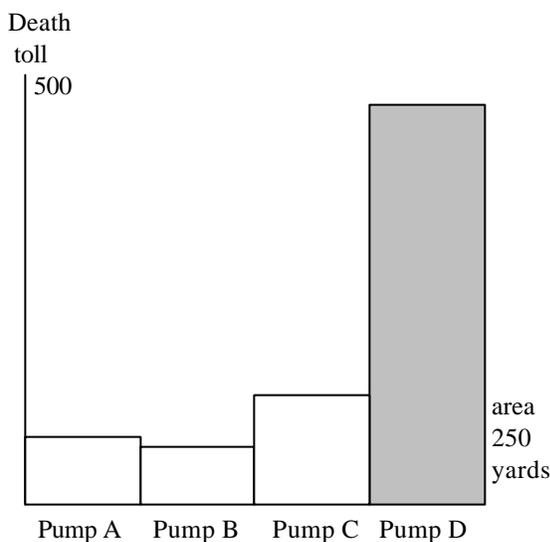

Figure 6. Alternative visualization.

hidden relations. Discovery of relations is an area of computational intelligence with such areas as data mining, machine learning and knowledge discovery [Kovalerchuk, Vityaev et al, 2000, 2001] and their visual aspects. Many visual correlation methods presented below are suitable for both purposes.

4. Review of current studies on visual correlation

Current studies on visual correlation range from formally defined classical linear correlation in statistics to very informally defined correlation between statements in a natural language. There are two major concepts of visual correlation:
(1) visual correlation as a *result of visualization* of objects/events (O/E) correlation and
(2) visual correlation as a *process of correlating* O/E visually.

Table 1 includes examples from variety of fields, which belong to both categories. The goal of this review is to structure and classify practices and methods used in many fields. We are interested in their generalization and in finding common ground for new applications. It is important to note that examples vary in their level of exact definition and presentation of correlation to the user:

- *High level.* A system makes exact and clear correlation with the design stage.
- *Medium level.* A system does not correlate objects in advance, but provides a user with interactive tools such as a curve-matching cursor.
- *Low level.* A system mostly relies on a human perceptual mechanism, providing similar graphical or multimedia presentation of correlated entities and some pointing mechanisms.

Table 1 codifies a variety of visual correlation methods. Table 2 presents examples of visual correlation software.

5. Criteria for visual correlation efficiency

Visual correlation shares many efficiency criteria with visualization in general. Table 3

contains a list of criteria for visual correlation efficiency. The problem of measuring information density (IDS) for visual information is highly nontrivial. For textual information, this problem has been well known for a long time. C. Shannon offered a measure of information -- information entropy -- applicable for transmission of information using communication channels. In visual correlation, we have a specific communication channel transmitting information from a computer to a human.

6. Conclusion

Visual correlation of objects and events is not yet shaped as a separate area of the decision-making process in general. A variety of methods have been developed independently in different fields with little or no communication and without common terminology. A significant amount generalization work should be done.

References

[1] Ratnesar, R., Sneak Attack, *Time*, Oct. 23, 2000, vol. 156 no. 17
[2] Tufte, E., *Visual Explanations: images and quantities, evidence and narrative*, Graphics Press, Connecticut, 1997.
[3] Snow, J., *On the Mode of Communication of Cholera*, London, 1855.
[4] Shaffer E., Reed D., Whitmore S., Shaffer B., Virtue: Performance visualization of parallel and distributed applications, *Computer*, v. 12, 1999, pp. 44-51.
[5] Loe Feijs, Roel de Jong, 3D visualization of software architectures, *Communications of ACM*, vol. 41, n. 12, pp. 73-78., 2000.
[6] Kovalerchuk, B., Vityaev, E., *Data Mining in Finance*, Kluwer, 2000.
[7] Kovalerchuk, B., Vityaev, E., Ruiz, J, Consistent and Complete Data and "Expert" Mining in Medicine, In: *Medical data Mining and Knowledge Discovery*, Springer, 2001, pp. 237-280.

Table 1. Visual correlation methods.

Single panel	Single dataset -single correlation plot	Linear correlation plot
		Curvilinear correlation plot (e.g., Fig.6)
	Multiple Datasets - Single Correlation Plot	*n* visualized entities in the single panel (e.g., Fig. 5)
	Glyphs	2-D Glyphs correlation
		3-D Glyph correlation
Line of panels side-by-side	Static pointers	Linked panels
	Dynamic interactive pointers	User sets up links interactively
	n panels side-by-side	*n* abstract visualizations side-by-side.
		n X-ray films side-by-side
		n real-world pictures side-by-side
Tree of panels	Vertical tree of panels	
	Horizontal tree of panels	
	Tree with the root in the center)	
Grid of panels	Table of n x n panels.	Grid of correlation and distribution plots
Network of panels	Static pointers	
	Dynamic interactive pointers	
	Side-by-side panels	
Nested panels	Nested panels for hierarchical views	Nested geographic maps and events
Panels in 3-D	Mountain panel	
	Fish eye	
	Room, gallery	
	Cone or disk tree	
Zooming and popping	Standard zooming	Geographic map zooming (2D or 3D)
	Zooming with changing metaphors and layouts	Magic Lens (2D or 3D)
Several monitors	Combination of all above listed methods	

Table 2 Visual correlation software: generic and field-specific.

Name/source	Capabilities
Generic statistical software (SAS, SPSS and so on)	Full range of single panel linear and curvilinear visual correlations of numeric datasets
Generic data mining software	Variety of visualizations of multidimensional data and discovered patterns
Correlation builder for wells *http://www.oilfield-systems.com/*	Rapid curve correlation using the curve-matching cursor. Displays curves and saves curves and correlation information. Supports use of predefined correlation templates for different users. Support mixture of templates in the same panel. Supports interactive set up of curve correlation parameters (width, gap)
Software application performance visualization [4]	The user selects an area for magnification and the system reveals magnified objects and their text descriptions using query and text modes. The system supports nested graphs for hierarchical views of geographic computations. The user selects a site and "drills down" and may see a different display metaphor and layout.
Software architecture visualization [5]	Uses VRML. Supports LEGO type 3-D glyph metaphor for software elements. General data types are brick with foundation. Special data types are brown bricks with four cylinders above. A data collector is a violet brick with foundation. Tables are cylinders.
Information visualization for arid lands *http://ialcworld.org/ Projects/95info.html*	The visual correlation of various geo-spatial and temporal data. Supports VC for satellite imagery, photographs, GIS layers, databases, spreadsheets, custom displays, etc. A visual interface to data through hypertext and hyper-area links. Supports access to a database from a map with "hotspot" which links to data related to that location. Supports the mouse to click on the "hotspot" for displaying the data related to that position.

Table 3. Criteria of efficiency of visual correlation.

Criterion	Description	Comment
TT -- Text time	Time for catching correlation of objects and events in a text.	
TS -- Text speed	Speed for catching correlation of objects and events in a text.	
VCT-- Visual correlation time	Time for catching up correlation visually	If VC time is relatively small then VC can be used in time-critical and/or information overloaded applications
IDS -- Information density	Amount of information presented visually. IDS is measured for a separate panel and as a sum of them (integral measure).	If IDS is relatively small then VC can be used in time-critical and/or information overloaded applications. If IDS is relatively large then VC can handle large applications. Visualization and visual correlation can be viewed as a specific type of data compression
VCS -- Speed of VC	IDS/VCT -- amount of information consumed per time unit.	If VCS is relatively high then VC can be used in time-critical and information overloaded applications
RVCT--Relative VC time	Time of visual correlation (VC1) relative to another visual correlation (VC2) or a text (TT1): VCT1/VCT2; VCT1/TT1.	If RVCT is relatively low then VC can be used in time-critical and information overloaded applications. Relative time can be measured in experiments without explicitly measuring information density of the VC.
RVCS -- Relative VC speed	Speed of visual correlation (VC1) relative to another visual correlation (VC2) or text TT1 VCS1/VCS2; VCS1/TS1	If RVCS is relatively high then VC can be used in time-critical and information overloaded applications. Relative time can be measured in experiments without explicitly measuring information density of the VC.
PIP -- parallel information processed	Amount of information processed in parallel in contrast with textual or other VC.	Image processing is a parallel process in contrast with sequential textual and audio information.
PIPS -- speed of parallel information processing	Speed of parallel information processing as a result of visual correlation in contrast with textual or other VC.	