# Visual Analytics Methods

Summary of chapter 3 and 4 from

Mastering the information age - Solving problems with visual analytics.

Eurographics Association, 2010.

László Kiss

CQI6IY

Univerity of Debrecen

2018

# Summary of Chapter 3

## Data management

### Motivation

According to the Data Management International Association: “Data Resource Management is the development and execution of architectures, policies, practices and procedures that properly manage the full data lifecycle needs for an enterprise”. VA applications need a reliable and efficient data management system layer to manage their data lifecycle.

There are some issues which need to be solved during data management:

- Heterogeneity of data sources: it is necessary to integrate the various data sources
- Different data types: it is a big challange to analyse and convert the different input data sets
- Data streams:handling this type of dataform could be a big and bigger challange due related to a modern IT infrastructure
- Working under pressure: The informations provided by VA applications could be a base some important decisions.
- Time consuming activities: It is a difficult process to managing different data format and type.

The data management layer has to be ensured the continuity between the classical methods and the new ways of data proccesses. The classical data management rely on RDBMS (Relational Database Management System) and SQL (Structured Query Language) and the modern ways come from the Big Data World.

### State of the Art

The market of the data management systems is domainated by RDBMS based solutions and this domination still growing. Due this fact the Visual Analytics data management has to integrate this. One of the most promising tools is the semantic integration which is based on exchange of metadatas.

There are some other techniques related to RDBMS and they share many goals with Visual Analytics. We talk about Data Warehouses, OLAP (On-Line Analytical Processing) and Data Mining.

Main shared goal:

- Data warehouses: support of decisions
- OLAP: provide a different points of view to the datasets.
- Data mining: discover the knowledgement

Data reductuion and abstraction: it is vital because of the optimatization of the physical and logical layer.

Data quality: all of the data management system has a commom problem about the input data. data could be incomplete, inconsistent, or contain measurement errors.

The most important tasks about data quality:

- Linking different views of the same data
- Restoring missing data
- Polishing the data

There are a lot of examples the represent the involvation of Visual Analytics to the classical data management task like OLAP, DW and data mining.

Visual OLAP: https://www.kyubit.com/OLAP-Analysis

Visual data mining: https://www.sas.com/sk_sk/news/press-releases/2016/september/visual-data-mining-machine-ax-2016.html

Visual data reduction: https://eprints.cs.univie.ac.at/4794/

## Challenges and opportunities

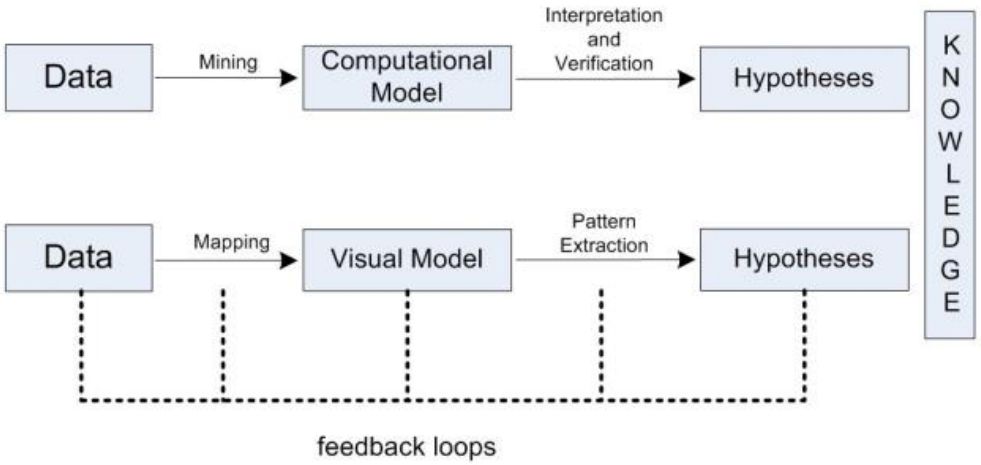
There are a lot of challenges for both data management and visual analytics such as missing data problem, data uncertainty and issues related to size of this datasets.In one hand Visual Analytics has to solve this, but ont he other hand VA can support with new solution about this problems.

# Summary of Chapter 4

## Data mining

### Motivation

Since the beginning of digital world humans were required in the data analysis process. It is necessary because complexity of the problem space. (For example Human Genome Project). See the next figure easy to understand the importance of common work between humans and digital system. We can see ther difference between the classical data mining procces and new way of data mining supported by VA.




It is obvious that new tools and new methodologies are developed to help the experts to extract the relevant information. Visual Analytics is a successful combination of automated and visual analysis. There are some research fields which can prove the useful of VA.

Bioinformatics: Due to the increase of complexity of research, the rate of this methods is getting higher and higher.

Climate change: Visual Analytics support the prediction of environmental processes.

Pattern identification: It is one of the main goal in KDD (knowledge discovery from data) and the combination of automated proccesses and VA could be a next step in this area.

Spatio temporal data mining: the Visual Analytics opens a new possibilities, because we can put the raw data to a map and other special plots and we can discover new relations between datas.

## Benefits in industry and „real-world”

Basicly the industry received well the new methods in data management, the companies recognized early the all of the new possibilities.

Some important application area:

Marketing data: Visual Analytics can help to deal with the flood of information and understand the „meaning” of large datasets

Process industry: VA can provide a way of making sense the large datasets generated by factories related to manufacturing.

Software industry: The size and complexity of this projects is currently growing rapidly, so the VA methods can support the analysis of large datasets with new methods.

Pharmaceutical industry: This is one of the most complex area in industry, so involve the new techniques to this processes is vital.

## State of Art

As we have seen, the objective of knowledge discovery and data mining is to extract information from large datasets.

Possible support activities in variable fields in data mining

Statistical and mathematical tools: The data visualization is a basic part of this area since beginning, so the connection is obvious.

Specific algorithmic tools: There are some current tools – Graphviz or Pajek – which are more focused on on the analysis of social and complex network data by taking advantage of network/graph visualisation.

Visual analytics libraries: It is a very good new possibility to create a new support tools to realize VA.

Visual data mining tools: It is a new type of softwares, which can creates visualisations to reveal hidden patterns from datasets.

Web tools and packages: With these tools, users can create visualisations using their own data.

Scientific visualisation tools: It is a representation of data graphically as a means to gain understanding the contexts of datasets.

Combined methods: There have been some attempts to combine data mining methods and visualisation. It will be a success story, but it requires a lot of efforts.

## Challenges and opportunities

The most important problem in this section could be a question of human-computer integration. Considering with this problem we can divide the actual topics to five groups.

- Analytical reasoning:
- Visual representations and interaction techniques:
- Data representations and transformations
- Production, presentation and dissemination
- Moving research into practice

In terms of KDD, the first (analytical reasoning) and third (data representations and transformations) categories are highly relevant.

Visual Analytics has the same issues about the data like classic data mining such as:

- (qualitative) textual data
- data stored in (distributed) databases
- data received from sensors
- spatial data such as satellite imagery
- audio and video

To handling these problems data standards and system of metadata required.

There are some special software and tools to realize the visual data mining. There were developed to a specific task and requires special experties.This methods can lead a Visual Controlled Data Mining which can be a next step in this topic.

The evaluation of this methods is very difficult, it requires special measurement and criteria. We have to analyze the opportunities about the collaboration with other techniques. I think if we use this opportunities in proper way, the Visual Analytics support our research and development activities.