CODATA Prague Workshop
Information Visualization, Presentation, and Design
29-31 March 2004

Goals of Analysis for Visualization and Visual Data Mining Tasks

Thomas Nocke and Heidrun Schumann
University of Rostock
D-18059 Rostock, Germany
nocke@informatik.uni-rostock.de
schumann@informatik.uni-rostock.de

Abstract

Extensive work has been done in the area of visual data mining in the last few years. This includes the development of visual and non-visual mining methods for large, heterogeneous and uncertain data. An appropriate selection and parameterization of such mining techniques are sensitive processes, especially if we have to consider highly complex analysis environments. One of the most important factors are goals of analysis, which specify the user's intentions in the analysis process. This paper describes a new, general specification for goals of analysis. It extends existing approaches by aspects of importance for real-world application scenarios.

Introduction

The rising complexity of current data sets requires new techniques to support users to handle their visual analysis. North, Conklin and Saini [4] stated that modern relational database technologies allow efficient and flexible data management, but today's visualization techniques are developed by only one programmer and are specific to one certain problem. Although this statement does not apply for innovative, flexible and adaptive visualization environments, it is true for many current visualizations, which do not distinguish between different contexts and user's goals. Thus, users are provided with a lot of information they might not be interested in. Goals of analysis are a suitable approach to specify the user's intentions for analysis processes in visual data mining and visualization environments. Thus, the gap between internal technique implementation, visualization expert-knowledge and the user's problem context knowledge can be bridgeded. Based on such goals, appropriate visual and non-visual mining techniques can be selected, parameterized and managed (semi-) automatically.

There are a variety of approaches to formalize and use goals of analysis in visualization systems. Roth and Mattis [6] introduce **information seeking goals** (e.g. value lookup, value comparison, identify distributions, find correlations) and characterize visualization expressivity by **metadata** for relational data (e.g. by scale type, cardinality of data records). Wehrend and Lewis [9] separate goals into **objects** (e.g. scalars, positions) and **operations** (e.g. identify, compare). Keller and Keller [2] present expressive visualization examples for various visualization goals combining **data types, operations, objects** and **application context**. Robertson [5] introduces the **location** factor to be analyzed (whether a user wants global, local or punctual information out of the data). In its Task by Data Type Taxonomy, Shneiderman [8] introduces several **tasks** (e.g. overview, zoom, and filter) and seven **data types** (e.g. 1D, 2D, Network). Fujishiro et al. [1] introduce **analysis targets** (scalar, nominal, direction, ...) and separate them from **analysis actions**. This separation is a suitable approach, because it allows specifying arbitrary targets independently from the kind of action to be performed. Furthermore, Fujishiro et al. [1] combine these two categories with Sneiderman's tasks [8].

Altogether, the application of goals of analysis has only partially been established in visualization and visual data mining systems. This may be due to the various interpretations of goals in the user minds of varying application backgrounds and for this reason, resulting ambiguites in terms. To overcome this problem, different views on goals have to be established, that fill the gap between internal processing, visualization expert vocabulary and application domain language. Therefore, the following aspects have to be considered. They integrate isolated aspects of the approaches mentioned above. Moreover, some further important aspects such as the level of analysis are added:

1. User driven aspects:
    a. **Task** (general goals) - Overview, Extract, Details-on-Demand, … (Shneiderman [8]),
    b. **Action** (special goals) - Associate, Classify, Compare, Identify, Reveal, ... (Wehrend and Lewis [9], Roth and Mattis [6], Keller and Keller [2], Fujishiro et al. [1]),
    c. **Location** - global, local and punctual (Robertson [5]),
2. Data driven aspects:
    a. **Data features and data types of interest (target)** - Scalar, Nominal, Direction, ... (Wehrend and Lewis [9], Roth and Mattis [6], Keller and Keller [2], Fujishiro et al. [1]), but as well e.g. Outlier and Typical value,
    b. **Data classes** - 1D, 2D, 3D, ND, Temporal, Tree, Network, and Map (Shneiderman [8]),
    c. **Dependence** – dependent and independent variables (Keller and Keller [2], dos Santos and Brodlie [7]),
    d. **Specialization** – general or data class specific goals,
    e. **Description** - abstract, verbal, XML, or internal,
3. Context aspects:
    a. **Application context** – neutral and application dependent (Keller and Keller [2]),
    b. **Level of analysis**, i.e. the visual analysis is provided by
        i. operator graphs consisting of visual and non-visual-mining operators,
        ii. operator graphs only consisting of visual operators,
        iii. single visualization techniques visualizing a static data set by one picture only,
4. **Complexity** aspect: elementary and composed goals.

Considering all these aspects offers a comprehensive view on the specification of analysis goals. This also includes for instance the treatment of goals concerning spatial or temporal analysis tasks (see e.g. the location aspect or the temporal data class). In the following, we will describe a new specification of goals of analysis based on the aspects mentioned above in more detail (section "Goal specification"). Afterwards, a framework will be presented which supports the gathering and handling of goals (section "Goal gathering and storage"). Finally, we will conclude with remarks on our approach and further work planned to apply goals of analysis in climate research context (section "Conclusions and future work").

Goal specification

On the one hand, a specification of goals has to consider the mentioned aspects. On the other hand, they must be provided in a user interface that is intuitive and easily manageable. To avoid user overstrain as well as loss of important aspects, we targeted on a well-balanced trade-off in our specification. Because of this, we developed a three-stage procedure: Main aspects which should be provided by a suitable user interface are all user driven aspects as well as the main data properties as matter of visualization. Further aspects such as **specialization** and **application context** are added by a suitable attribution. Finally, the **complexity** is added by combining attributed goals.

Thus, we define **analysis task**, **analysis action**, **analysis location** and **analysis target** as main aspects. Then, we define AL as the set of all possible values specifying **analysis location,** and respectively AT for **analysis target,** AC for **analysis action** and AK for **analysis task.** Now we define an **elementary goal** (eg) as one element of one of these sets.

Then, we enrich every elementary goal by the 2 aspects **specialization** and **application context,** with SP as set of all **specialization** parameter values and respectively AP for **application context**. In doing so, we define an **attributed elementary goal** (aeg) as follows: *aeg = f (eg, sp, ap)*, with *sp Î SP* and *ap Î AP* and with the attribution function *f: EG x SP x AP ® AEG*, with EG the set of all **elementary goals** and AEG as the set of all **attributed elementary goals**. **Data classes** and **data class specific goals** are integrated into the **specialization** attribute. This allows for instance to add to the analysis target “surface” the **specialization** information “3D volume data”. An example for an added **application context** could be the “evaluate” **action** for a “simulation” background.

Based on these definitions, we define a **composed goal** (cg) as a subset of the union of all **attributed elementary goals** (aeg): *cg Í AEG*. Now, these **composed goals** can be labeled using a naming function *n: CG ® String*, with CG being the set of all **composed goals** and String being the set of possible goal names. This naming function for goals in combination with the attributes **specialization** and **application context** allows deriving different views on goals in different application scenarios.

The question which values of the specified goal sets we need, is still focus of our research. In a first step, we use the assignments from literature, and enrich them with application dependent values. Moreover, the **description** (and labeling) aspect is of high relevance for understandability and interpretability by the users. Specifying **composed goals**, hiding their internal structure to non-expert users and name them application background dependent, creates easily understandable views to the goals. Thus, internal and abstract details can be hidden, and vocabulary can be customized for the user’s background language.

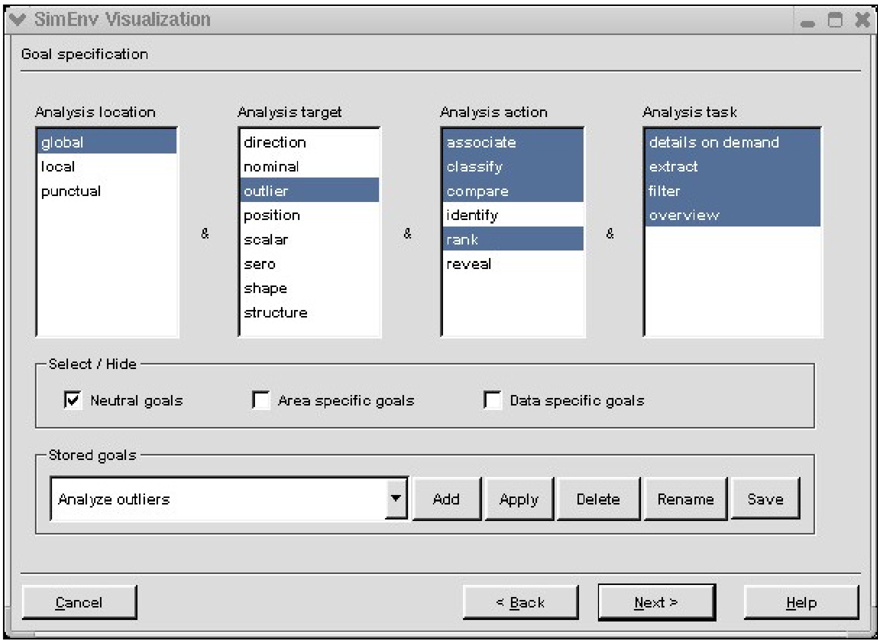
In summary, we have provided a flexible taxonomy for specifying user goals in analysis processes, that comprises general goals for general-purpuse application as well as for special-purpose applications using the naming function. All the introduced approaches from the state of the art can be encoded with this taxonomy (adding further aspects by attribution if necessary). For instance, the examples of Keller and Keller [2] with the aspects **analysis action**, **analysis target**, **application background** and **specialization** (including data class specific information) can be expressed by our taxonomy.

Goal gathering and storage

We have developed a special component to handle goals of analysis and integrated it into a general data mining framework for the investigation of climate data. The component consists of a data structure to store **elementary** and **composed goals**, a GUI part and I/O functionality to store and reload defined goals. Figure 1 shows the GUI part of the component and an excerpt of the XML-files storing **elementary** and **composed goals**.

In the four columns at the top (see left part of figure 1), **elementary goals** from the four categories **location**, **target**, **action** and **task** can be selected. Below, the display of neutral, **application specific** and **data specific goals** can be filtered in and out. Later on, a selection of several classes of **application dependent** and **data specific goals** is planned in this GUI area. At the bottom, the so-defined sets of **elementary goals** can be managed as **composed goals**. This includes adding, changing, deleting, renaming and storing them. Thus, our framework provides an efficient base for interactively realizing the naming function and generating **application specific goals**.

In figure 1 a complex goal for analyzing outliers is in focus. The selected **elementary goals** can be interpreted in the following way: “globally”, “outliers” should be “associated”, “compared”, “classified” and “ranked”. Furthermore, the visual mining technique to be selected should support “overview”, “extract”, “filter” and “extract” on “outliers”. This **composed goal** is currently denoted as “analyze outliers” and could be renamed for instance, in “global outlier analysis”.




```
<GoalItem>
  <Value text = "local"/>
  <Type text = "AO"/>
  <Specialization text =
     "general"/>
  <Context text =
     "neutral"/>
</GoalItem> ...
<GoalItem>
  <Value text="surface"/>
  <Type text= "TA"/>
  <Specialization text  =
     "volume data"/>
  <Context text =
     "neutral"/>
</GoalItem>
```

```
<StoredGoals>
 <Goal name = "Analyze
outliers">
  <AO v0 = "global"/>
  <AC v0 = "associate"
      v1 = "classify"
      v2 = "compare"
      v3 = "rank"/>
  <TA v0 = "outlier"/>
  <AT v0 = "details on
           demand"
      v1 = "extract"
      v2 = "filter"
      v3 = "overview"/>
 </Goal> ...
</StoredGoals>
```

Figure 1. Screenshot of the **goal of analysis dialog** from our climate investigation framework (left) – it contains the selection of elementary goals in four listboxes for each of the four main goal sets (top left), a section to filter goals (center left) and a management area for the **composed goals** to be stored (bottom right); **elementary goal** XML definition file (top right); XML file to store **composed goals** (bottom right).

In the right part of figure the storage of **attributed elementary goals** (top right) and of stored **composed goals** is presented.

In a first version of our framework, a couple of goals are linked statically with a set of mining techniques. For more complex goals as described, a combination of several visual presentations is provided, each of them supporting a special goal. The described component for specifying goals of analysis is part of a more complex architecture for analyzing climate data. This architecture includes, besides the goal specification, a tool for extracting metadata (see Nocke and Schumman [3]). As a further component, a visualization design tool is currently under development. It uses the output of the metadata and the goal specification component to select and parameterize appropriate data mining operator networks combining automated mining and visualization techniques basing on the visualization system OpenDX.

Conclusions and further work

In this paper, we introduced a systematic specification to define comprehensive goals for different purpose visual data mining tasks and outlined its application opportunities. The specification is designed to support single visualization technique scenarios as well as the selection and generation of complex visual data mining operator networks.

Nevertheless, there are still challenges for future work: On the one hand, we will test the specification of complex goals and their adaptability for the climate research context. Although first experiments with real users using the goals in our visualization design framework (in a climate simulation environment scenario) were successful, we have to continue testing and evaluating its usability. On the other hand, extensive work has to be done to apply them to simplify the visualization process, i.e. to use it in the visualization design tool.

References

[1] Fujishiro, I., Furuhata, R., Ichikawa, Y., and Takeshima, Y., GADGET/IV: A Taxonomic Approach to Semi-Automatic Design of Information Visualization Applications Using Modular Visualization Environment, IEEE Information Visualization, 2000, 77ff.

[2] Keller, P.R., and Keller, M.M, Visual Cues. Practical Data Visualization. IEEE Computer Society Press, Los Alamitos, 1993

[3] Nocke, T., and Schumann, H., Meta Data for Visual Data Mining, Proceedings Computer Graphics and Imaging, CGIM 2002, Kaua'i, Hawaii, USA, August, 2002

[4] North, C., Conklin, N. and Saini, V., Visualization Schemas for flexible Information Visualization, Proceedings of Symposium on Information Visualization’02, Boston, 2002

[5] Robertson, P. K., A Methodology for Scientific Data Visualisation: Chosing Representations Based on a Natural Scene Paradigm, Proceedings Visualisation '90, IEEE Computer Society Press, Los Alimatos, 1990.

[6] Roth, S.F., and Mattis, J., Data Characterization for Intelligent Graphics Presentation, Proceedings of the Computer-Human Interaction Conference (CHI '90), 1990.

[7] dos Santos, S., and Brodlie, K., Gaining understanding of multivariate and multidimensional Data through Visualization. To be published in: Computers & Graphics, Vol. 18. No. 3, 2004

[8] Shneiderman, B., The Eyes Have It: A Task by Data Type Taxonomy for Information Visualization, Proceedings IEEE Symposium on Visual Languages'96, Los Alamos 1996.

[9] Wehrend, S., and Lewis, C., A problem-oriented classification of visualization techniques, IEEE Visualization'90, 1990, 139-143.