Visual Analytics: How Much Visualization and How Much Analytics?

Daniel A. Keim
University of Konstanz
78457 Konstanz
Germany
Daniel.Keim@uni-konstanz.de

Florian Mansmann
University of Konstanz
78457 Konstanz
Germany
Florian.Mansmann@uni-konstanz.de

Jim Thomas
Pacific Northwest National Laboratory
Richland, WA
U.S.A.
Jim.Thomas@pnl.gov

ABSTRACT

The term *Visual Analytics* has been around for almost five years by now, but still there are on-going discussions about what it actually is and in particular what is new about it. The core of our view on Visual Analytics is the new enabling and accessible analytic reasoning interactions supported by the combination of automated and visual analysis. In this paper, we outline the scope of Visual Analytics using two problem and three methodological classes in order to work out the need for and purpose of Visual Analytics. Thereby, the respective advantages and disadvantages of automated and visual analysis methods are explained plus examples of analytic reasoning interaction leading to a glimpse into the future of how Visual Analytics methods will enable us to go beyond what is possible when separately using the two methods.

1. INTRODUCTION

Visual Analytics is the science of analytical reasoning supported by interactive visual interfaces [10]. Over the last decades data was produced at an incredible rate. However, the ability to collect and store this data is increasing at a faster rate than the ability to analyze it. While purely automatic or purely visual analysis methods were developed in the last decades, the complex nature of many problems makes it indispensable to include humans at an early stage in the data analysis process. Visual Analytics methods allow decision makers to combine their flexibility, creativity, and background knowledge with the enormous storage and processing capacities of today's computers to gain insight into complex problems. The goal of visual analytics research is thus to turn the information overload into an opportunity by enabling decision-makers to examine this massive information stream to take effective actions in real-time situations.

Automatic analysis techniques such as statistics and data mining developed independently from visualization and interaction techniques. However, some key thoughts extended the scope of the fields into what is today called Visual Analytics research. One of the most important steps in this direction was the need to move from confirmatory data analysis to exploratory data analysis, which was first stated in the statistics research community by John W. Tukey in his book "Exploratory data analysis" [12].

Later, with the availability of graphical user interfaces and proper interaction devices, a whole research community devoted their efforts to information visualization [2; 3; 9; 13]. At some stage, this community recognized the potential of integrating the user in the KDD process through effective and efficient visualization techniques, interaction capabilities and knowledge transfer leading to *visual data exploration* or *visual data mining* [4]. This integration considerably widened both the information visualization and the data mining fields, resulting in new techniques and plenty of interesting and important research opportunities.

The term *Visual Analytics* was coined by Jim Thomas in the research and development agenda "Illuminating the Path" [10], which had a strong focus on Homeland Security in the United States. Meanwhile, the term is used in a wider context, describing a new multidisciplinary field that combines various research areas including visualization, human-computer interaction, data analysis, data management, geo-spatial and temporal data processing and statistics [6; 5; 7].

In general, Visual Analytics has the capability to transform many of our daily work processes and make them both more effective and efficient. With regards to effectiveness, we can clearly see that information visualization technology is often applied to help users to obtain and maintain an overview in various situations. Combining it with automated analysis results and reasoning analytics in Visual Analytics systems, enables scaling to larger and more challenging problems. Regarding the efficiency, it is noticeable that automated analysis often speeds up analysis tasks considerably whereas the visual representation is then used to efficiently communicate the outcome to the user or to disseminate research results to a wider audience.

While Visual Analytics was originally introduced for solving challenging problems that were unsolvable using automatic or visual analysis alone, meanwhile many Visual Analytics applications have demonstrated the broadening of this technology and its applicability to a much wider area.

In this paper, we first explain automatic analysis and visualization as well as their combination and the applicability of these three method classes to analytical problems as well as to general application areas of IT. Afterwards, advantages and shortcomings of visual and automated analysis methods are assessed in order to work out the benefits of combining them in Visual Analytics applications. The last section then summarizes our view on Visual Analytics.

2. VISUAL ANALYTICS PROBLEMS

Visual Analytics can be best explained using the two problem classes (1) Analytical Problems and (2) General Application Areas of IT as illustrated in Figure 1. In order to solve problems in these two classes, there are three methodological classes: a) Automatic Analysis, b) Visualization, and c) Visual Analytics.

Let us first consider the problem classes. *Analytical problems* have in common that there is an inherent logic to each problem, which makes it possible to assess them in a rational way. However, this does not imply that all these analytical problems are solvable. Very often, we do not have sufficient computational or human resources to solve the problems within our lifetime using the presently available methods. While analytical problems are within the class *General Application Areas of IT*, problems of the latter class are not necessarily hard problems. They could involve tasks as simple as sending or receiving an e-mail.

Automatic Analysis methods can be used to solve some of these analytical problems. In particular, these methods apply when we have means for measuring and comparing the quality of candidate solutions to the problem at hand. While there is a broad range of problems where these methods apply, they fail when algorithms are trapped in local optima, which are unrelated to the globally best solution. *Visualization* methods on the other hand, use human background knowledge, creativity and intuition to solve the problem at hand. While these approaches often give very good results for small data sets, they fail when the available data for solving the problem is too large to be captured by a human analyst. *Visual Analytics* combines the strengths of both worlds: On the one hand they take advantage of intelligent algorithms and vast computational power of modern computers and on the other hand they integrate human background knowledge and intuition to find a good solution.

Traditionally, most Visual Analytics research has focused on mastering these problems that are only solvable through a combination of visual and automatic analysis. While the importance of some of these problems in natural sciences, business or government have justified investments in Visual Analytics during the last five years, the acceptability of users to work with such systems has broadened the original scope considerably and will have a strong influence on future developments. Such Visual Analytics systems can for example be used to solve simpler problems, that are also solvable with means. However, effectiveness and efficiency of Visual Analytics systems justify their usage even in non-classical Visual Analytics applications. For example, a visual tool to support people in archiving their e-mails in several folder taking into account textual similarity and a visual interface displaying a ranking of the most relevant folders solves a task, which can also be solved with traditional methods. However, the Visual Analytics application speeds up the performance of the user in completing the task and is therefore a valid reason for the use of Visual Analytics. Likewise, improving the effectiveness of users also motivates the use of Visual Analytics in general application areas of IT.

Another aspect to consider is the quality of the solution to a problem. Depending on the available means for finding good solutions, users might agree to accept suboptimal solutions. Nevertheless, these suboptimal solutions can also be completely unacceptable in some application scenarios.

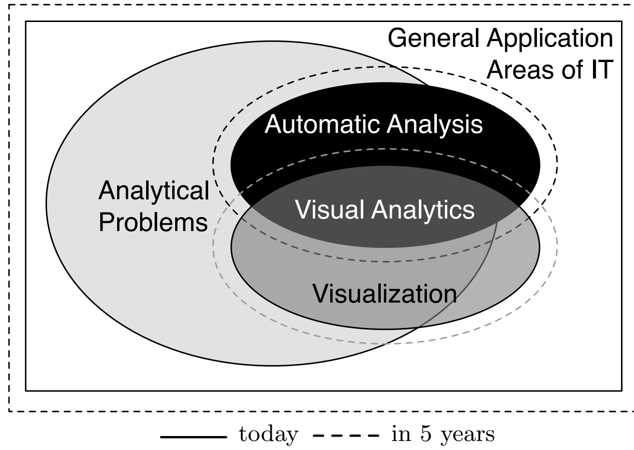

Figure 1: The two problem classes General Application Areas of IT (white) and Analytical Problems (light gray) can be solved using Automatic Analysis (black), Visualization (gray), and Visual Analytics (dark gray). Note that not every automatic or visual analysis problem is a Visual Analytics problem if other effective and efficient ways of solving the problem exist.

3. COMBINING AUTOMATIC AND VISUAL ANALYSIS

In many cases automated analytics is favored towards interactive visual analysis since getting the user involved in the analysis process can be an unpredictable and cost-intensive undertaking. However, many real-world problems are not well-defined from the very beginning and can thus not be analyzed by an automated algorithm. Especially when these algorithms are applied on vaguely defined problems, the relationship between the input and the output of the algorithms often remains unclear to the analyst. Therefore, the question arises whether the analyst can trust the system or not. Furthermore, some problems require dynamic adaptation of the analysis solution, which is very difficult to be handled by an automated algorithm.

The major drawback of visual data analysis is probably the fact that it is a cost-intensive activity to pay highly specialized experts who need to be trained on the software for many days. One of the goals of visual analytics new systems is to provide "walk-up usable" interfaces, which greatly reduces training and learning time on new technologies. In case of a real-time data analysis, these experts sometimes need to be available 24/7. However, the advantages of visual analysis often outweigh the drawbacks: Using interactive analysis, the users are given means to steer the analysis process in an intelligent way, which results in a trusted and more meaningful end result.

Visual Analytics aims at combining the best of both worlds. Taking efficient automated analysis methods wherever appropriate while allowing the user to combine their output with his or her background knowledge and intuition. On the one hand, this approach establishes a higher level of trust into the end result for problems that can be solved using automated or visual analysis. On the other hand, it is only through Visual Analytics that some of today's most pressing data analysis problems become solvable since neither automated analysis nor visualization alone can provide

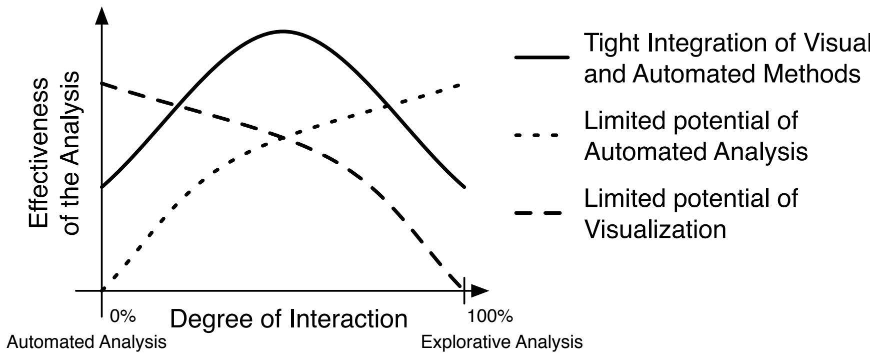

Figure 2: Potential of Visual Analytics

solutions to these large-scale and complex problems.

In this context, we have to think about what it means to solve a problem. Since not all data analysis problems can be precisely formulated in words not to mention in mathematical formulas, we have to face fuzzy solutions as output of the analysis. Sometimes, we do not even know whether a given solution is the best possible solution or just a good one. Depending on the analysis task, the available data and the evaluation measure, we have to settle for sub-optimal results. In the end, it boils down to a trade-off between the resources needed to come up with a better solution or the potential damage that a sub-optimal solution will cause.

From a conceptual perspective, the traditional analysis processes as described in the KDD pipeline for automated analysis or in the Information Seeking Mantra [8] for visual analysis are merged in the *Visual Analytics Mantra*: "Analyze first, show the important, zoom, filter and analyze further, details on demand". Furthermore, we have to think about the levels of integration between automatic and visual analysis (cf. [1]) in order to properly design future Visual Analytics applications.

In a recent article [11] many applications were discussed that will take advantage of the new visual analytics technologies. In addition 5 systems were analyzed to characterize the specific visual analytics reasoning techniques that enabled rapid insightful knowledge discovery. Some of these were whole-part relationship; relationship discovery; combined exploratory and confirmatory interaction; supporting multiple data types; temporal, geospatial and linked views; groupings and outlier identification; labeling, analytic reporting, and several more. These analytic interactions are key to successful visual analytics technologies that are walkup usable and sometimes scale independent.

In the end, we can only assess how much visualization and how much analytics are required for a particular problem by assessing the users' capabilities, the analysis task and the available data. Larger data sets and well-defined problems might be better solvable through analytics whereas visualization establishes a stronger confidence in the end results while using and extending background knowledge about the problem at hand. Figure 2 shows the potential of Visual Analytics on three different kind of problems. The dashed line represents the effectiveness of analyzing problems, such as automatic electric switching, customer scoring or credit card approval, that can be solved more effectively through automated analysis. In contrast to this, problems such as the search for the airplane of Steve Fossett in huge amounts of high resolution satellite images, are still better solvable through humans. Combining the best of both worlds through visual analytics applications is a very promising solution for problems that can neither be effectively solvable through automated analysis nor explorative analysis as shown by the solid curve. A very interesting research question is therefore to develop methods for determining the optimal combination of visualization and automated analysis methods to solve different classes of problems by taking into consideration the user, the task and the characteristics of the data sets.

4. CONCLUSIONS

In this paper we defined Visual Analytics as new enabling and accessible analytic reasoning interactions supported by the combination of automated and visual analysis. After a short outline of the field's history we explain the broadening of this technology into more general application areas of IT. While it was originally meant to solve some of the most difficult problems in government, business and science we foresee its applicability to everyday processes due to both the efficiency and effectiveness of Visual Analytics applications.

As a further contribution, this paper discusses advantages and disadvantages of visualization and automatic analysis in order to work out the joint potential of Visual Analytics applications. Thereby, we consider (a) the effect of sub-optimal solutions, (b) the need for adapting existing analysis processes using the Visual Analytics Mantra by integrating both visualization and automatic analysis and (c) a brief assessment of how much of each is needed.

5. ACKNOWLEDGEMENTS

This work was funded by the EU VisMaster Coordination Action in the Future and Emerging Technologies (FET) programme under FET-Open grant number 225429 and by the Pacific Northwest National Laboratory (PNNL) in Richland, WA, U.S.A.

6. REFERENCES

[1] E. Bertini and D. Lalanne. Surveying the complementary role of automatic data analysis and visualization in knowledge discovery. In *VAKD '09: Proceedings of*

the ACM SIGKDD Workshop on Visual Analytics and Knowledge Discovery, pages 12–20, New York, NY, USA, 2009. ACM.

[2] S. W. Card, J. D. Mackinlay, and B. Shneiderman, editors. *Readings in information visualization: using vision to think*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1999.

[3] C. Chen. *Information Visualization - Beyond the Horizon*. Springer, 2nd edition, 2004.

[4] D. A. Keim. Visual exploration of large data sets. *Communications of the ACM (CACM)*, 44(8):38–44, 2001.

[5] D. A. Keim, F. Mansmann, A. Stoffel, and H. Ziegler. *Visual Analytics*. Springer, 2009. Encyclopedia of Database Systems.

[6] D. A. Keim and J. Thomas. Scope and challenges of visual analytics, 2007. Tutorial at IEEE Visualization, http://kops.ub.uni-konstanz.de/volltexte/2008/6899/.

[7] J. Kielman, J. Thomas, and R. May. The future of visual analytics. *Information Visulalization*, 2009.

[8] B. Shneiderman. The eyes have it: A task by data type taxonomy for information visualizations. In *IEEE Symposium on Visual Languages*, pages 336–343, 1996.

[9] R. Spence. *Information Visualization - Design for Interaction*. Pearson Education Limited, 2nd edition, 2006.

[10] J. Thomas and K. Cook. *Illuminating the Path: Research and Development Agenda for Visual Analytics*. IEEE-Press, 2005.

[11] J. Thomas and J. Kielman. Challenges for visual analytics. *Information Visualization Journal*, 11:309–314, 2009. Special Issue: Foundations and Frontiers of Visual Analytics.

[12] J. W. Tukey. *Exploratory Data Analysis*. Addison-Wesley, Reading MA, 1977.

[13] C. Ware. *Information Visualization - Perception for Design*. Morgan Kaufmann Publishers, 2nd edition, 2004.