Optimizing Processes in Visual Data Analysis through Progressive Computations

Cagatay Turkay
giCentre, City University London
Cagatay.Turkay.1@city.ac.uk

Helwig Hauser
University of Bergen
Helwig.Hauser@ii.uib.no

I. Introduction

The seminal work by Card et al. [1] investigates different aspects of human capabilities in our communication with the outer world, be it another human or a computer. In their paper, they describe the human computer interaction process as a dialogue between the user and the computer. Based on their investigation of psychology literature, they present three *human time constants* that characterize the temporal characteristics of our related human capabilities. These constants are reported to be highly important to achieve an optimal communication between the user and the computer.

Visual analytics (VA) is, in particular, an interactive and iterative dialogue between the human and the computer. The interactive analysis process is a sequence of actions by the user and responses by the computer. As such, it is of vital importance to think of visual analytics as a dialogue and to properly address the perceptual and cognitive capabilities of humans in this dialogue in the light of the already mentioned three time constants.

As a result of recent research activities, powerful processes in visual analysis involve the tight integration of computational tools and interactive methods [2]. Our solution moderates the temporal aspects of the interactive visual steering of computational analysis tools (Figure 1). Instead of forcing the user to wait for an interactive computation to finish, our methodology aims to present a best possible result within an acceptable time frame. And depending on the interpretation of these first approximate results, the user might either wait for more accurate results to compute or continue to explore the data by updating his/her interactive inputs. This approach is inline with the suggestion by Card et al. [1] that reads *"…When the cycle time becomes too high, cooperating rendering processes reduce the quality of rendering …"*.

In order to maintain the temporal limitations set forth by the human time constants, we develop a mechanism where the computational tool guarantees to respond within a fixed time, i.e., 1 second. To achieve this, we make use of *online algorithms*, which are capable of processing the data piece-by-piece sequentially [3]. These algorithms do not need the whole data to operate and can update the results as new data becomes available. One common method to use online algorithms is to pass the data items row by row, so that each update cycle of the algorithm is performed in a limited time [3]. To be able to utilize this incremental computing nature of online algorithms, we use them in combination with a sampling method.

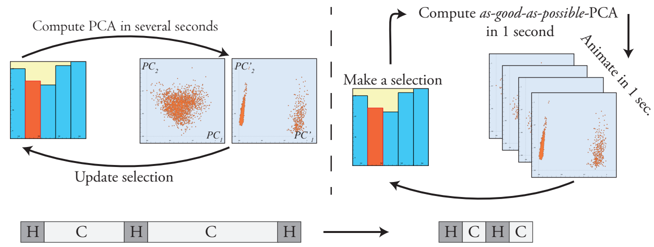

Figure 1. Optimizing an analytical process against perceptual constraints. In a conventional setup (left), the user first requests a (re-)computation of PCA results (with a selection of variables), then he/she waits a certain time for the results. This waiting could potentially interrupt the dialogue between the user and the computer. In order to address this, our suggested optimization (right) computes PCA results *as good as possible* within 1 sec. in response to a selection by the user. And whenever new selections are made, the results are re-computed in no more than 1 sec. and the visualizations animate in 1 sec. to display the new results.

When working with massive datasets, our methods can enable the user to get an understanding of the data quickly. And after finding interesting relations in the data, e.g., which dimension subset to use for clustering, the user can refer to a sophisticated offline algorithm (which potentially takes a long time to compute) to get more accurate results. This amounts to a more efficient pipeline compared to using the costly algorithms without any prior investigation of the data.

References

[1] S. K. Card, G. G. Robertson, and J. D. Mackinlay, "The information visualizer, an information workspace," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. New York, NY, USA: ACM, 1991, pp. 181–186.

[2] C. Turkay, A. Lundervold, A. Lundervold, and H. Hauser, "Representative factor generation for the interactive visual analysis of high-dimensional data," *Visualization and Computer Graphics, IEEE Transactions on*, vol. 18, no. 12, pp. 2621–2630, 2012.

[3] S. Albers, "Online algorithms: a survey," *Mathematical Programming*, vol. 97, no. 1, pp. 3–26, 2003.