A Framework for Cognitive Bias Detection and Feedback in a Visual Analytics Environment

Alexander Nussbaumer*, Katrien Verbert‡, Eva-C. Hillemann*, Michael A. Bedek* and Dietrich Albert*
* Cognitive Science Section, Knowledge Technologies Institute
Graz University of Technology, Austria
Email: {alexander.nussbaumer, eva.hillemann, michael.bedek, dietrich.albert}@tugraz.at
‡ HCI Research Group, Department of Computer Sciecne
Katholieke Universiteit Leuven, Belgium
Email: katrien.verbert@kuleuven.be

***Abstract*—This paper presents a framework that supports the detection and mitigation of cognitive biases in visual analytics environments for criminal analysis. Criminal analysts often use visual analytics environments for their analysis of large data sets, for gaining insights on criminal events and patterns of criminal events, and for drawing conclusions and making decisions. However, due to the nature of human cognition, these cognitive processes may lead to systematic errors, so-called cognitive biases. The most prominent and relevant cognitive bias in the intelligence field is the confirmation bias, in which an analyst disproportionally considers and selects information that supports the initial expectation and hypothesis. The framework presented in this paper describes a model, how the possible occurence of the confirmation bias can be detected automatically, while the analyst makes use of the visual environment. Moreover, based on this information, different feedback methods are employed that support and encourage the mitigation of the confirmation bias. This framework is in a work-in-progress state and contains research objectives and directions, the framework design, initial implementations, plans for further development and integration, as well as user-centric evaluation.**

I. Introduction

The development and application of new knowledge and information technologies have enormous influence on the way people live, work and learn. In the intelligence sector criminal analysts receive huge amounts of data, countless unsystematic dots of information, of which they are constantly required to understand and make sense of. Sense-making in this context means that analysts have to find and interpret relevant facts by actively constructing a meaningful and functional representation of some aspects of the whole picture. Visual Analytics technologies can help to make this task easier for the analyst by giving support to the human reasoning and sense-making processes.

Even though the support for sense-making with visual analytics technologies is helpful and valuable, there is still a well known problem of systematic errors, so-called cognitive biases, that might hinder analysts from doing accurate analysis by drawing sound conclusions. Biases occur when uncertainty, complexity, and time constraints prohibit people from making optimal decisions. In such situations they unconsciously apply heuristics, which can be thought as "rules of thumb" when assessing the value, importance, and meaning of information. Though these heuristics are useful in general, they can lead to severe and systematic errors in judgement [1] or decision biases. In the context of intelligence analysis, these "systematic errors" or cognitive biases can occur in every phase of the intelligence cycle causing errors in judgement, such as discounting, misinterpreting, ignoring, rejection or overlooking information [2]. One of the most well-known cognitive biases is the confirmation bias, in which an analyst disproportionally considers and selects information that supports the initial expectation and hypothesis. In this paper, we focus specifically on confirmation bias as defined by Nickerson [3], who describes it as the the seeking or interpreting of evidence in ways that suit existing beliefs, expectations, or a hypothesis in hand. For example, in a prominent case (NSU case in Germany) the invistigators assumed that several murders were motivated by a conflict within the ethnical group of the victims. As consequence they investigated only within this ethical group. However, it turned out that there was a xenophobic background and that the offenders were not part of the investigated people.

Some approaches have outbeen described in literature how the confirmation bias can be mitigated. For example, Heuer [2] proposes a simple methodology called the Analysis of Competing Hypotheses (ACH) to mitigate the confirmation biases. ACH consists of a manual approach that guides users in identifying possible hypothesis and collecting a list of significant evidence for and against each hypothesis. A hypothesis-evidence matrix visualisation is used to support the user in the decision making process. The approach guides users to manually distribute attention more evenly across all hypotheses and evidence as a way to mitigate confirmation bias. Other methods for mitigating the confirmation bias are the use of different visualisation techniques and the indication where data is lacking or uncertain [4], as well as the provision of computerised critique questions and the support for decision making in groups [3].

In contrast to bias mitigation, there are almost no methods known for the automatic detection of confirmation biases. Endert et al. [5] describes an approach of using possible trends derived from the temporal history of keyword weighting. Converging trends in the weighting of entities might

indicate confirmation bias, whereas diverging weights might represent an analysis involving multiple hypotheses. Fisher [6] describes a method for the detection of the confirmation bias by presenting a questionnaire before and after working on a task. The confirmation bias can be calculated from the change of the answer patterns. This is in our view the only available method for the operationalisation of the detection of the confirmation bias. Apart from these methods and to the best of our knowledge there is no method to automatically detect the confirmation from interaction data without additional questionnaires or specific activities to be done by the user.

In this paper we present a framework for detecting and mitigating confirmation biases in a visual analytics environment. New concepts for the bias detection are presented and combined with existing mitigation strategies. These concepts are integrated in a visual analytics environment where data of the users' interactions with information visualisation tools are used as input for the bias detection, and feedback for bias mitigation is presented to the user. Incorporating knowledge of cognitive processes into adaptive visualisations has been identified as a key challenge for the visual analytics community by Kristin Cook and James Thomas [7][8]. The visual analytics process combines automatic and visual analysis methods with a tight coupling through human interaction in order to gain knowledge from data [9]. In this sense the analyst is kept in the feedback loop by using interaction data to give meaningful support for bias mitigation.

The proposed framework described in Section 3 is our approach and contribution to and integration of the research fields of intelligence analysis, cognitive biases, and visual analytics. It consists of both research approaches and software components integrated to an adaptive visual analytics environment. Section 2 describes the objective of this framework in detail.

II. Objectives

The overall goal is to research and develop adaptive visual analytics systems that can prevent the user from seeking supporting evidence for original beliefs by challenging the users beliefs, if the user is suffering from a confirmation bias during an investigation work. This aim leads to a framework that includes (a) methods for bias detection, (b) several feedback methods to mitigate a potential confirmation bias, (c) an approach to evaluate the soundness of the developed detection and mitigation strategies, and (d) a way to qualify the data visualisation tools regarding their tendency to induce or prevent biases.

The figure below (Figure 1) depicts the overall research activities and software components of the framework for bias detection and bias mitigation. The core part is the visual analytics environment that the analysts use for their work. The visual analytics platform is connected to data sets about committed crimes. The use of interactive visualisation tools produces log data that are used to analyse the behaviour of the analysts and to detect potential biases. Based on such results, various forms of feedback are given to the analysts using the visual analytics platform. In addition to this feedback, the results from the bias detection are also used to assess the visualisation tools in the environment regarding their tendency to induce or reduce the confirmation bias.

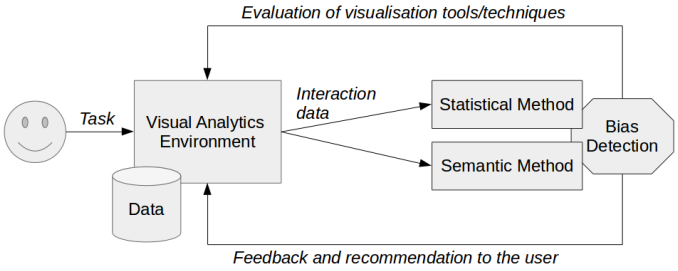

Fig. 1. Framework for bias detection and feedback

From a research point of view the main activities and challenges are the development of the detection methods, the provision of meaningful feedback for bias mitigation, and the evaluation of the effectiveness and soundness of these methods. From a development point of view the main activities are the development and integration of the individual components, which are the visual analytics environment including visual tools and data storages, the bias detection comportments, and the feedback methods integrated in the visual analytics environment. The next section describes the individual research and development activities in the context of the overall framework.

III. Framework Design

A. Research and visual analytics environment

The first step in the creation of this framework is the preparation and integration of an environment that puts together all software components and that allows to perform research work and user studies. It provides a platform for visualisation tools that allow dynamic interaction with data sets and help in the sense-making process, in order to fulfill an investigation task. The development of such an environment will include the use of existing interactive visualisation tools and libraries that will be adapted to the needs of analysts and our studies. These visualisation and interaction tools are taken from previous work, including the MUVA project on Managing Uncertainty in Visual Analytics [10] and the VALCRI project on visual analytics for criminal intelligence [11].

In order to work with realistic and complex situations, data sets will be used from three different sources. First, real crime data from open data archives will be used, for example form the US City Open Data Census[1] or from the UK Police[2]. A second data source are data sets provided by the Visual Analytics Community[3]. These data sets are created to enable the creation of and experiments with visual analytics

[1] http://us-city.census.okfn.org/dataset/crimestats
[2] https://data.police.uk/
[3] http://www.vacommunity.org/VAST+Challenge

environments. Third, data sets with real (but anonymised) crime events from the VALCRI project will be used.

In conjunction with elaborating the research environment, typical user tasks will be defined that are meaningful for intelligence analysts and represent their daily work (e.g. finding similar and related crimes or tracing the activities of a criminal). The development of the research environment and the selection of the data sets are closely related to the user tasks.

In order to capture, store, and nalysse user interaction data, the Equalia service [12] is being used. Equalia was originally developed to support the whole evaluation process of adaptive systems. It is a service-based component that supports the systematic capture and management of interaction and log data. Based on a process model, the data are automatically analysed and made available for further processing.

B. Bias detection

In general two different types of methods are researched and tried out that allow for the detection of confirmation bias. The first method makes use of a statistical procedure and the second one of a semantic procedure. In both cases we expect likelihoods rather than certainty that a user suffers from a confirmation bias.

Both methods are based on the selective exposure experiment for the detection of the confirmation bias [6]. Unlike our objective to automatically detect the confirmation bias, this experiment employs a questionnaire before and after the user performs a task. Based on the answer patterns in the pre-and post-questionnaire, it can be detected whether a user is biased.

The statistical method compares interaction behaviour of biased and non-biased users. In order to classify users according to their state according to their confirmation bias, they will participate in the selective exposure experiment. Then they complete some tasks and their interaction data is classified according to the bias state of the respective user and can be seen as training data for the detection algorithm. An appropriate detection algorithm has to be selected and adapted and could be based on an existing one from the machine learning field (e.g. Vector Space Model for similarity measures between documents (e.g. [13]).

The semantic method is based on a model of cognitive processes and how they are related to the confirmation bias. By doing thinking-aloud tests, relationships between observable user interactions and non-observable cognitive processes will be established. Using the selective exposure experiment it can then be detected which and how intensive cognitive processes are involved in a biased and unbiased investigation task. Similarity measures as described above are applied to do the automatic detection. Another alternative for detecting biases is the random walk model that has recently been used in the context of (non-cognitive) biases (e.g. [14]).

To the best of our knowledge the above described selective exposure experiment is the only way to operationalise the measurement of the confirmation bias. Our two method types described above are hypotheses for the automatic detection and their elaboration constitutes part of the core research.

C. Feedback

If a likelihood for a confirmation bias has been identified through the detection methods described above, feedback to the user will be provided to make the analyst aware of a potential bias. Following the ideas presented in [15], several methods are available to provide feedback to the analysts and change their thinking process. The first method is the change of perspective and view of the data. For example, a different visualisation technique can be used to present the data to the analyst in a different form. Further, multiple views can present the same data at the same time in different ways. The second method deals with the use of uncertainty of the available data. If data is presented, it will be shown how certain the data is in terms of its evidence. Indicating that data is uncertain is supposed to make the user aware of the thinking process. Third, computerised critique questions will explicitly make the user aware to rethink the current hypothesis. Fourth, explicit prompts to rethink the own hypothesis is another method. Fifth, a prompt to discuss the current hypothesis in groups or with peers might change the thinking direction. Another approach may rely on the Analysis of Competing Hypotheses method presented by Heuer [2] and visualise multiple hypotheses and evidences in a matrix to support awareness.

All these methods have different levels of interventions and might disturb the workflow and thinking process of the user. On the one hand, this is a wanted effect to de-bias thinking, on the other hand this can also be annoying. Finding a good balance is part of the research. In general, discovering and understanding effective feedback methods is the second core research question.

D. User Studies

Empirical user studies are used for two reasons and in two situations. First, studies are needed in the research process to get training data for the bias detection methods. Participants doing investigation tasks are evaluated if they are biased or not. The interaction data of the biased and unbiased participants are used for pattern analysis.

The second type of user studies targets the evaluation of the overall approach. User studies will be conducted by asking them to perform several tasks with a visual analytics environment with different methods to support bias mitigation (experimental conditions) and a baseline VA environment without such support (baseline condition). Data will be collected by recording user behaviour, such as the time taken to perform the task, the number of steps, the number of recommended items that are accepted etc. As in our previous work, the data encoding scheme of Brown et al. [16] will be adopted to record interactions of subjects with the system. This data is then analysed in a second step to identify which recommendation and visualisation techniques are most effective to mitigate biases. Commonly-used precision and recall metrics will be used

to measure recommendation accuracy. In addition, subjective user data will be collected with post-study questionnaires. The ResQue framework [17] will be used to collect data about quality of recommended items, perceived ease-of-use, and user satisfaction. In addition the framework of Kijnenburg [18] will be applied to collect additional information about personal and situational characteristics that may have an influence on the effectiveness of feedback and recommendations.

E. Available Components

In the current state we have available several components of the framework needed for the realisation of the overall bias detection and feedback approach:

- Visualisation tools and software libraries for visualisation
- Data sets of criminal events from open data repositories
- Logging and analysis service (Equalia)
- Confirmation bias experiment as online study

Current work focuses on the bias detection algorithms and concrete feedback mechanisms based on these results. Furthermore the integration of these components requires development work and the validation the overall approach requires user studies.

IV. Conclusion

This paper describes our current work towards automatic bias detection in criminal intelligence. This work is a combination of research of new methods and development and integration of system components. The presented framework outlines how the research work is put into practice. Next steps include the implementation of the missing components (mainly the bias detection methods), the integration of the several components, and the conducting of user studies.

To the best of our knowledge, no successful research work has been conducted on automatic detection of confirmation biases. The idea to detect such biases from interaction data has been introduced by Endert et al. [19]. The authors proposed the use of so called semantic interactions to detect the analytic reasoning process of the user, but so far no research has been done to evaluate whether the approach can be used successfully to detect confirmation biases, and which algorithms are suitable for detecting such biases.

Available research is limited to classification and description of biases (e.g. [20]), models of cognitive processes related to biases [21], bias mitigation strategies [22], and experiments to detect biases (Selective Exposure experiment [6]). However, there are no methods known to detect biases automatically. In this work, we build on this existing research to develop a methodology for detecting biases in real time.

Acknowledgment

The research leading to these results in project VALCRI has received funding from the European Union 7th Framework Programme (FP7/2007-2013) under grant agreement no FP7-IP-123456.

References

[1] A. Tversky and D. Kahneman, "Judgment under uncertainty: Heuristics and biases," *Science*, vol. 185, no. 4157, pp. 1124–1131, 1974.

[2] R. J. Heuer, *Psychology of Intelligence Analysis*. Central Intelligence Agency, 1999.

[3] R. S. Nickerson, "Confirmation bias: A ubiquitous phenomenon in many guises," *Review of general psychology*, vol. 2, no. 2, p. 175, 1998.

[4] M. Cook and H. Smallman, "Human factors of the confirmation bias in intelligence analysis: Decision support from graphical evidence landscapes," *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 50, no. 5, pp. 745–754, 2008.

[5] A. Endert, L. Bradel, and C. North, "Beyond control panels: Direct manipulation for visual analytics," *EEE Computer Graphics and Applications*, vol. 33, no. 4, pp. 6–13, 2013.

[6] P. Fischer, E. Jonas, D. Frey, and S. SchulzHardt, "Selective exposure to information: The impact of information limits," *European Journal of Social Psychology*, vol. 35, no. 4, pp. 469–492, 2005.

[7] K. A. Cook and J. J. Thomas, Eds., *Illuminating the Path: The Research and Development Agenda for Visual Analytics*. Los Alamitos, CA, USA: IEEE Computer Society, 2005.

[8] J. J. Thomas and K. A. Cook, "A visual analytics agenda," *IEEE Computer Graphics and Applications*, vol. 26, no. 1, pp. 10–13, 2006.

[9] D. A. Keim, F. Mansmann, and J. Thomas, "Visual analytics: how much visualization and how much analytics?" *ACM SIGKDD Explorations Newsletter*, vol. 11, no. 2, pp. 5–8, 2010.

[10] K. Seipp, F. Gutirrez, X. Ochoa, and K. Verbert, "Visualising uncertainty in algorithm visualisations using techniques from geospatial visualisation," submitted to IEEE VAST 2016.

[11] J. Haider, M. Pohl, E.-C. Hillemann, A. Nussbaumer, S. Attfield, P. Passmore, and W. B. L. Wong, "Exploring the challenges of implementing guidelines for the design of visual analytics systems," *Proceedings of the Annual Meeting of Human Factors and Ergonomics Society*, vol. 59, no. 1, pp. 259–263, 2015.

[12] A. Nussbaumer, E.-C. Hillemann, C. M. Steiner, and D. Albert, "An evaluation system for digital libraries," in *Proceedings of the 2nd International Conference of Theory and Practice of Digital Libraries (TPDL 2012). Lecture Notes in Computer Science*, P. Zaphiris, G. Buchanan, E. Rasmussen, and F. Loizides, Eds., vol. 7489. Berlin, Heidelberg: Springer, 2012, pp. 414–419.

[13] G. Salton, A. Wong, and C. S. Yang, "A vector space model for automatic indexing," *Communications of the ACM*, vol. 18, no. 11, pp. 613–620, 1975.

[14] D. Volchenkov and P. Blanchard, *Fair and biased random walks on undirected graphs and related entropies*. Birkhuser, 2011.

[15] E.-C. Hillemann, A. Nussbaumer, and D. Albert, "The role of cognitive biases in criminal intelligence analysis and approaches for their mitigation," in *Proceedings of the European Intelligence and Security Informatics Conference (EISIC 2015)*, J. Brynielsson and M. H. Yap, Eds. New York, USA: IEEE, 2015, pp. 125–128.

[16] E. T. Brown, A. Ottley, H. Zhao, Q. Lin, R. Souvenir, A. Endert, and R. Chang, "Finding waldo: Learning about users from their interactions," *IEEE Transactions on Visualization and Computer Graphics*, vol. 20, no. 12, pp. 1663–1672, 2014.

[17] P. Pu, L. Chen, and R. Hu, "A user-centric evaluation framework for recommender systems," in *Proc. of the fifth ACM conference on Recommender systems*. ACM, 2011, pp. 157–164.

[18] B. P. Knijnenburg, M. C. Willemsen, Z. Gantner, H. Soncu, and C. Newell, "Explaining the user experience of recommender systems," *User Mod. and User-Adapted Int.*, vol. 22, no. 4-5, pp. 441–504, 2012.

[19] A. Endert, R. Chang, C. North, and M. Zhou, "Semantic interaction: Coupling cognition and computation through usable interactive analytics," *IEEE Computer Graphics and Applications*, vol. 35, no. 4, pp. 94–99, 2015.

[20] J. Baron, *Thinking and Deciding*, 4th ed. New York: Cambridge University Press, 2007.

[21] J. W. Payne, J. R. Bettman, and E. J. Johnson, *The adaptive decision maker*. Cambridge, England: Cambridge University Press, 1993.

[22] L. J. Sanna and N. Schwarz, "Metacognitive experiences and human judgment : The case of hindsight bias and its debiasing," *Current Directions in Psychological Science*, vol. 15, no. 4, pp. 172–176, 2006.