

mVis in the Wild: Pre-Study of an Interactive Visual Machine Learning System for Labelling

Mohammad Chegini*
Graz University of Technology, Austria
NTU, Singapore

Jürgen Bernard†
UBC, Canada

Lin Shao‡
Graz University of Technology, Austria

Alexei Sourin§
NTU, Singapore

Keith Andrews¶
Graz University of Technology, Austria

Tobias Schreck||
Graz University of Technology, Austria

ABSTRACT

Many machine learning algorithms require a labelled training dataset. The task of labelling a multivariate dataset can be tedious, but can be supported by systems combining interactive visualisation and machine learning techniques into a single interface. mVis is such a system, providing a unified ecosystem to explore multivariate datasets and execute machine learning algorithms to build labelled datasets.

This paper describes a pre-study evaluation of the mVis system, comprising case studies in two different domains: collaborative intelligence and daily activities. In each case study, a volunteer researcher was asked to use mVis to explore, analyse, and label their own dataset in their own environment, while thinking out loud. The case studies provided valuable leanings in terms of the usability of the system, understanding how different analysts work, and identifying important missing features.

Index Terms: Human-centered computing—Visualization—Visualization design and evaluation methods

1 INTRODUCTION

A multivariate dataset has two or more dimensions and usually comes with a large number of records. A common task for analysts and researchers from various fields is to explore, partition, and label multivariate datasets. Automated algorithms often used in Machine Learning (ML), such as clustering and classification, can help label datasets and partition them into multiple classes. While automated ML algorithms create robust models for a dataset, choosing an appropriate algorithm and its parameters for a particular dataset is not always straightforward. Moreover, ML algorithms are blind to the context of dimensions and some of them, including supervised techniques, work only for datasets for which a sufficiently large training set of labelled records is available.

On the other hand, user-centred approaches such as visual clustering delegate labelling tasks to the analyst. Although labelling a small dataset with a limited number of records can be done manually by the analyst, labelling a large multivariate dataset can quickly become time-consuming and in some cases an impossible task. Furthermore, manually labelling a dataset may decrease precision. A combination of user-centred and automated approaches can potentially build robust models which are controlled and verified by an expert.

* e-mail: m.chegini@cgv.tugraz.at

† e-mail: jubernar@cs.ubc.ca

‡ e-mail: l.shao@cgv.tugraz.at

§ e-mail: assourin@ntu.edu.sg

¶ e-mail: kandrews@tugraz.at

|| e-mail: t.schreck@cgv.tugraz.at



Figure 1: *The facilitator (left) and researcher (right) conducting a case study with mVis [7].*

mVis [7] is an interactive visual machine learning system, which helps an analyst either partition an unlabeled dataset or explore and verify the partitions of a labelled one. While it is challenging to develop interactive ML systems which do not overwhelm the analyst with parameters and options, at least one study has shown the usefulness of such interactive visual systems over pure ML algorithms (e.g. active learning) for labelling tasks [3].

There are many ways to evaluate interactive systems for visual analysis [1, 2] and many motivations behind such evaluations [10]. However, as previous researchers have noted, it can be challenging to evaluate such systems [5, 8, 13]. Datasets can vary wildly and tasks are often dependent on the kind of data being explored. In many applications, domain experts are recruited for evaluation. However, a domain expert is not always available or willing. It is also hard to measure and compare the “insights” which such systems are designed to discover [11]. The difficulty of running controlled experiments has led to the increasing use of qualitative evaluation methods involving case studies and observation of individual users [12, 16].

This paper describes a pre-study evaluation of the mVis system, comprising two case studies each in a different domain (collaborative intelligence and daily activities). In each case study, a volunteer researcher with no previous experience of mVis was observed as they used mVis to explore, label, and verify a dataset from their own domain. The researchers were asked to talk out loud as they worked to provide greater insight into their thought process [9]. Afterwards, the researchers participated in a semi-structured interview. This type of evaluation is useful to (1) test the usability of the system, (2) understand how the current implementation helps the analyst with their tasks, and (3) identify important missing features. The results will be used to inform and plan future evaluations. Figure 1 shows the setup of a case study. Figure 2 shows a screenshot of

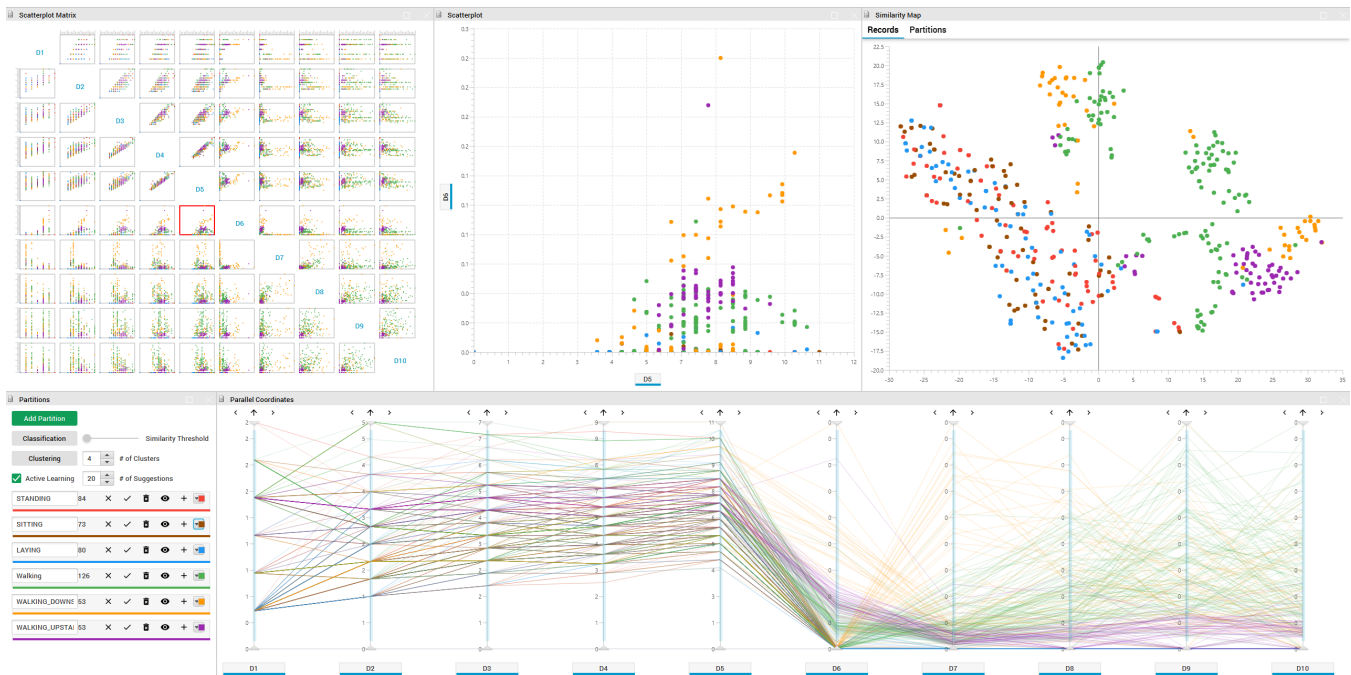


Figure 2: A screenshot of mVis taken during the second case study.

mVis during the second case study.

2 RELATED WORK

A number of systems have sought to combine Visual Analytics (VA) with Machine Learning (ML) techniques to create advanced interactive visual machine learning systems which help an analyst explore and build complex models for large multivariate datasets. For example, Brown et al. [4] introduced Dis-Function, a system which allows the analyst to explore a dataset and actively define a distance function instead of directly manipulating parameters in the model. Another example of this combination is Regression Lens [15], an interactive visual tool to explore a dataset through scatterplots and perform local regression analysis. Later, Chegini et al. [6] extended this work to support local pattern search in scatterplots using model and shape-based distance functions. In a comparative study, Bernard et al. [3] demonstrated the usefulness of interactive visual systems over active learning for labelling tasks.

There are many ways to evaluate interactive systems for visual analysis [1, 2]. Lam et al. [10] systematically reviewed over 800 visualisation publications and identified seven scenarios (motivations) for the evaluation of information visualisations: three for understanding data analysis processes and four for evaluating the visualisations themselves. Sedlmair et al. [14] identified nine stages of design when designing visualisations for domain experts. According to these scenarios and design stages, Wong et al. [18] suggested appropriate evaluation methods for each stage.

It can, however, be challenging to evaluate such systems [5, 8, 13]. Running controlled experiments on interactive visual systems can be particularly challenging. Datasets can vary wildly and tasks are often dependent on the kind of data being explored. Domain experts can be hard to find or unwilling to participate [18]. It is also hard to measure and compare the “insights” which such systems are designed to discover [11].

The difficulty of running controlled experiments has led to the increasing use of qualitative evaluation methods involving case studies and (longer term) observation of individual users. Shneiderman and Plaisant [16] introduced the idea of the Multi-dimensional In-depth

Long-term Case (MILC) study, a structured process to evaluate a VA system by observing a small number of domain experts using the system with their own datasets over a longer period of time. The MILC method has been shown to give a comprehensive understanding and high-quality results [12, 17]. This paper describes two stripped-down case studies of the mVis system with domain experts (in the spirit of MILC), which will be used to guide and inform future development and evaluation.

3 PRE-STUDY OF MVIS

mVis [7] is an interactive visualisation tool for exploring and labelling multivariate datasets, implemented in Java. The system consists of four linked visualisation views and one panel to manipulate partitions, as shown in Figure 2. The analyst can explore a multivariate dataset using a scatterplot matrix (SPLOM), scatterplot, similarity map (projection by PCA, MDS, or t-SNE), and parallel coordinates plot. The analyst can freely select records in any view and create new partitions or add to existing ones. Moreover, several machine learning operations, including clustering, classification, and active learning, help the analyst build robust models interactively. The integration of multiple-linked view exploration, with visual clustering and other state-of-the-art ML techniques support a domain expert in labelling a multivariate dataset, which can then be used as a training set for other supervised ML algorithms.

Since the focus is not on a specific domain, gathering requirements and evaluation feedback is complex. It is necessary to conduct several domain expert studies to identify common requirements and improve usability of the system. Since the system is still evolving, qualitative studies involving observation and thinking-aloud, followed by semi-structured interviews are preferred over other types of user study.

The pre-study evaluation of the mVis system comprised two case studies, each with a domain expert. In each case study, the domain expert was a volunteer researcher with no previous experience of mVis. The facilitator first provided the researcher with a five-minute introduction to mVis. Then, the researcher was observed as they used mVis in their own office environment to explore, label, and

verify a familiar dataset from their own domain, while thinking out loud [9]. The facilitator sat next to the participant and took notes, and provided assistance with mVis when asked. Afterwards, the participant was interviewed in a semi-structured way to discover (1) their general impression of mVis (2) any missing features, and (3) in which stages of analysis mVis proved useful.

3.1 Case Study 1: Collaborative Intelligence Dataset

The volunteer researcher in the first case study was Monika. She is working on a collaborative intelligence platform dataset consisting of 718 records and ten dimensions. Each record represents a user and the dimensions are quantitative numbers associated with activities of the user on the platform (for example, number of *comments* and number of *reports*). Monika has been working with this dataset for over a year.

The session was conducted on a laptop with a 12-inch display in Monika's office. The facilitator imported the dataset into mVis after cleaning it. The facilitator then explained the user interface of mVis for around 5 minutes and asked Monika to freely explore the dataset while verbalising her thoughts. She started by investigating patterns in pairwise dimensions. For example, she realised that users who participate in *chats* do not often give *comments*. She also identified a relationship between *comments* and *reports*. Later, she performed several brushing interactions, using either the parallel coordinates plot or scatterplot. After the initial exploration phase, she created two partitions using first k-means and then hierarchical clustering. The new partitions mainly separated active and inactive users. She deleted all partitions and then performed another k-means clustering with k equal to four. Monika removed *votes* from the participating dimensions, since it is not a good indicator for clustering based on her prior experience. The result was better this time. For example, she identified that one of the clusters are users having a low number of *reports*. In summary, Monika used mVis primarily to (1) find new patterns among users and dimensions, (2) verify her previous observations about the dataset, and (3) create meaningful partitions.

A semi-structured interview was conducted shortly after the observation phase. The facilitator started by asking Monika's impression of mVis. She said she could confirm many previous observations in her explorations and even find new ones. She stated "Understanding this dataset would have been so much easier, if I had had this tool a year ago". The facilitator asked which features mVis is missing. She noted a lack of interactive help, lack of an easy-to-use dataset importer, lack of a feature to save previous partitions, and no means to visually compare old and new partitions. Monika later added that mVis is especially useful for people who do not have enough knowledge to use other tools such as R and Python. Finally, the facilitator asked in which stages of analysis mVis can be useful. Monika mentioned mVis is useful in the initial phase of data exploration to understand dimensions and records before creating advanced models. The duration of the session after importing the dataset was 55 minutes.

3.2 Case Study 2: Daily Activities Dataset

The participant in the second case study is Jian, a researcher working on a dataset about daily activities. The dataset has 25 dimensions and 412 records collected by feature extraction from time-series signals. The signals are gathered by placing various sensors in users' pockets to observe their daily activities. Each record belongs to a specific user and has already been manually assigned to (labelled with) exactly one of six states (classes) denoting daily activities: *walking*, *walking upstairs*, *walking downstairs*, *sitting*, *standing*, and *laying*. Figure 2 shows this dataset in mVis.

The session was conducted in Jian's office on a PC with a 24-inch display. The PC screen and user's voice were recorded on video. First, Jian explained the dataset to the facilitator and remarked that he has been exploring the dataset using conventional data science

tools such as R, Python, and SQL. He mentioned that he never tried to visualise the dataset using standard visualisation techniques. After introducing mVis to him, the facilitator imported the dataset into mVis. Jian reacted to the visualisation by stating "this looks great". Since the dataset was already labelled, he focused on finding relationships between dimensions and partitions. He first imported all 25 dimensions, and later decided to include only the first 10 dimensions into mVis. He started by identifying dimensions and their relationships with partitions. For example, he realised a dimension called *D8* can separate records labelled as *laying* from the others. By looking at the SPLOM, he discovered that many patterns recur among pairs of dimensions. He remarked that mVis does an excellent job in grouping dimensions and partitions. He later performed clustering to observe the differences between manual labelling and automatic partitioning using ML algorithms. At this point, he wished there was a visual comparison tool to compare previous and current partitions. In summary, Jian used mVis to (1) identify correlations between dimensions and partitions, (2) find relationships between pairs of dimensions, and (3) verify the manually labelled model in the dataset.

After the observation phase, the facilitator conducted a semi-structured interview. He first asked Jian about his impression of mVis. Jian mentioned that mVis might be useful for finding patterns in a dataset and he would use it for data exploration and model validation. Regarding missing features, he made two suggestions. First, a comparison tool to compare various partitions and secondly a guidance module to explain patterns in scatterplots. Lastly, he answered the question about which stage of data analysis mVis is useful by mentioning initial exploration. He added mVis could also play a crucial role in the validation phase. The duration of the session, including the introduction to mVis was 42 minutes.

4 DISCUSSION

The pre-study with two case studies involved a combination of thinking-aloud, observation, and interview. It demonstrated the general utility of mVis and illuminated future directions. mVis is a general purpose system and is not designed for a specific domain, therefore it is crucial to work with a variety of analysts and domain experts to define common analysis approaches and goals and address these in the system.

One of the key observations of the study is how the nature of the dataset can change interactions with the system. For example, in the case of the collaborative intelligence dataset, the focus was on selecting individual records and investigating each dimension thoroughly. The smaller nature of the dataset allowed the analyst to do this. In case of the daily activities dataset, it was not possible for the analyst to explore all records and dimensions. The analyst was more interested in typical pairwise patterns and relationships between dimensions. By conducting more case studies on a variety of datasets from a variety of domains, it is hoped that further behavioural patterns and typical tasks will be revealed.

The results of the pre-study verified that mVis is moving in the right direction, but still lacks some key features. The observations showed that both participants could work with mVis without the need for much assistance. It confirmed that mVis is easy-to-use and fast to learn. Nevertheless, both participants asked for an interactive help module, and one of the participants requested a guidance module. Using this module, the system will show the correlation between dimensions and partitions as extra information. Moreover, the system could generate an automatic description of visible patterns in the dataset and guide the user toward them.

It is planned to release mVis as standalone, cross-platform software. First, however, several critical features, including proper dataset import and export, history tracking, and selection of dimensions still have to be implemented. Moreover, further case studies need to be conducted, including possibly more detailed and longer-

term case studies using the full MILC methodology.

5 CONCLUDING REMARKS

The evaluation of interactive visual analytics systems is often challenging. Many evaluation methods are available, from controlled experiments and comparative studies, through heuristic evaluations and formative usability studies, to longer term case studies. Choosing the right evaluation method at the right stage of a project can be tricky. Controlled experiments require significant effort. At the early stages of system development, a focus on design case studies and formative usability evaluation can be more beneficial and require less effort.

Conducting even stripped-down case studies with mVis has been hugely beneficial in terms of discovering usability issues, understanding how analysts might use the system to help with their tasks, and identifying important missing features. The results will also be used to inform and plan future evaluations. The authors would appreciate feedback from the other workshop participants as to how best to approach the continuing evaluation of mVis.

ACKNOWLEDGMENTS

Many thanks to researchers at the IALab of Monash University and at Fraunhofer Singapore, specifically to Monika Schwarz and Jian Cui, for their participation in the pre-study. Thanks also to Luise Middel and Ajin Joseph from Fraunhofer Singapore for helping with the setup.

REFERENCES

- [1] Keith Andrews. Evaluating information visualisations. In *Proc. AVI 2006 Workshop on BEyond time and errors: novel evaluation methods for Information Visualization (BELIV'06)*, pages 1–5. ACM Press, 2006. ISBN 1595935622. doi: 10.1145/1168149.1168151. URL <http://ftp.isds.tugraz.at/pub/papers/andrews-beliv2006.pdf>.
- [2] Keith Andrews. Evaluation comes in many guises. In *CHI 2008 Workshop on BEyond time and errors: novel evaluation methods for Information Visualization (BELIV'08)*, 2008. URL <http://ftp.isds.tugraz.at/pub/papers/andrews-beliv08.pdf>.
- [3] Jürgen Bernard, Marco Hutter, Matthias Zeppelzauer, Dieter Fellner, and Michael Sedlmair. Comparing visual-interactive labeling with active learning: An experimental study. *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, 24(1):298–308, 2018. ISSN 1077-2626. doi: 10.1109/TVCG.2017.2744818.
- [4] Eli T. Brown, Jingjing Liu, Carla E. Brodley, and Remco Chang. Dis-function: Learning distance functions interactively. In *Proc. IEEE Conference on Visual Analytics Science and Technology (VAST 2012)*, pages 83–92, 2012. doi: 10.1109/VAST.2012.6400486.
- [5] Sheelagh Carpendale. Evaluating information visualizations. In Andreas Kerren, John T. Stasko, Jean-Daniel Fekete, and Chris North, editors, *Information Visualization: Human-Centered Issues and Perspectives*, number 4950 in LNCS, pages 19–45. Springer, 2008. ISBN 354070955X. doi: 10.1007/978-3-540-70956-5_2.
- [6] Mohammad Chegini, Lin Shao, Robert Gregor, Dirk J. Lehmann, Keith Andrews, and Tobias Schreck. Interactive visual exploration of local patterns in large scatterplot spaces. *Computer Graphics Forum (CGF)*, 37(3):99–109, 2018. doi: 10.1111/cgf.13404.
- [7] Mohammad Chegini, Jürgen Bernard, Philip Berger, Alexei Sourin, Keith Andrews, and Tobias Schreck. Interactive labelling of a multivariate dataset for supervised machine learning using linked visualisations, clustering, and active learning. *Visual Informatics*, 3(1):9–17, 2019. doi: 10.1016/j.visinf.2019.03.002.
- [8] Anamaria Crisan and Madison Elliott. How to evaluate an evaluation study? comparing and contrasting practices in vis with those of other disciplines. In *Proc. Vis 2018 Workshop on Evaluation and Beyond – Methodological Approaches for Visualization (BELIV 2018)*, pages 28–36. IEEE, 2018. doi: 10.1109/BELIV.2018.8634420.
- [9] K. Anders Ericsson and Herbert A. Simon. *Protocol Analysis: Verbal Reports as Data*. MIT Press, 1984.
- [10] Heidi Lam, Enrico Bertini, Petra Isenberg, Catherine Plaisant, and Sheelagh Carpendale. Empirical studies in information visualization: Seven scenarios. *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, 18(9):1520–1536, 2012. ISSN 1077-2626. doi: 10.1109/TVCG.2011.279.
- [11] Chris North. Toward measuring visualization insight. *IEEE Computer Graphics and Applications*, 26(3):6–9, 2006. doi: 10.1109/MCG.2006.70.
- [12] Adam Perer and Ben Shneiderman. Integrating statistics and visualization for exploratory power: From long-term case studies to design guidelines. *IEEE Computer Graphics and Applications*, 29(3):39–51, 2009. doi: 10.1109/MCG.2009.44.
- [13] Catherine Plaisant. The challenge of information visualization evaluation. In *Proc. Advanced Visual Interfaces (AVI 2004)*, pages 109–116, Gallipoli, Italy, 2004. ACM. ISBN 1581138679. doi: 10.1145/989863.989880.
- [14] Michael Sedlmair, Miriah Meyer, and Tamara Munzner. Design study methodology: Reflections from the trenches and the stacks. *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, 18(12):2431–2440, 2012. ISSN 1077-2626. doi: 10.1109/TVCG.2012.213.
- [15] Lin Shao, Aishwarya Mahajan, Tobias Schreck, and Dirk J. Lehmann. Interactive regression lens for exploring scatter plots. *Computer Graphics Forum*, 36(3):157–166, 2017. doi: 10.1111/cgf.13176.
- [16] Ben Shneiderman and Catherine Plaisant. Strategies for evaluating information visualization tools: Multi-dimensional in-depth long-term case studies. In *Proc. AVI 2006 Workshop on BEyond time and errors: novel evaluation methods for Information Visualization (BELIV '06)*, pages 1–7, Venice, Italy, 2006. ACM Press. ISBN 1595935622. doi: 10.1145/1168149.1168158.
- [17] Eliane R. A. Valiati, Carla M. D. S. Freitas, and Marcelo S. Pimenta. Using multi-dimensional in-depth long-term case studies for information visualization evaluation. In *Proc. CHI 2008 Workshop on BEyond time and errors: novel evaluation methods for Information Visualization (BELIV 2008)*, pages 1–7, Florence, Italy, 2008. ACM. doi: 10.1145/1377966.1377978.
- [18] Yuet Ling Wong, Krishna Madhavan, and Niklas Elmqvist. Towards characterizing domain experts as a user group. In *Proc. IEEE Evaluation and Beyond - Methodological Approaches for Visualization (BELIV 2018)*, pages 1–10. IEEE, 2018. doi: 10.1109/BELIV.2018.8634026.