Mohammad CHEGINI, Jürgen BERNARD, Jian CUI, Fatemeh CHEGINI, Alexei SOURIN, Keith ANDREWS, Tobias SCHRECK. Interactive visual labelling versus active learning: an experimental comparison. *Frontiers of Information Technology & Electronic Engineering,* 21(4):524-535. https://doi.org/10.1631/FITEE.1900549

Interactive visual labelling versus active learning: an experimental comparison

Key words: Interactive visual labelling; Active learning; Visual analytics

Corresponding author: Mohammad CHEGINI
E-mail: m.chegini@cgv.tugraz.at
ORCID: https://orcid.org/0000-0002-3516-8685

Introduction

- Labelling is assigning a class from the label alphabet to an instance (a record) in a multivariate dataset.
- Supervised machine learning algorithms must be trained on a labelled dataset.
- Active learning algorithms can help the analyst by suggesting instances to label.
- Interactive visual labelling tools build explorable visual overviews on top of active learning algorithms and can outperform classic active learning techniques regarding accuracy.

Motivation

- Since there are multiple visualisation and interaction techniques, the following research question arises: **How do characteristics of these techniques and datasets affect performance and user experience for visual interactive labelling tasks?**
- To answer the questions, a comparative user study of three well-known interactive visualisation techniques for labelling tasks is performed: similarity map, scatterplot matrix (SPLOM), and parallel coordinates.

Methods

- Using mVis, the performances of three different visualisation techniques for labelling a multivariate dataset were compared.
- mVis allows the analyst to select one or multiple instances for labelling.
- In mVis, every time a set of instances is labelled, random forest runs in the background and suggests potential labels for all currently unlabelled instances.
- For the experiment, the user was restricted to selecting a single instance at each step.

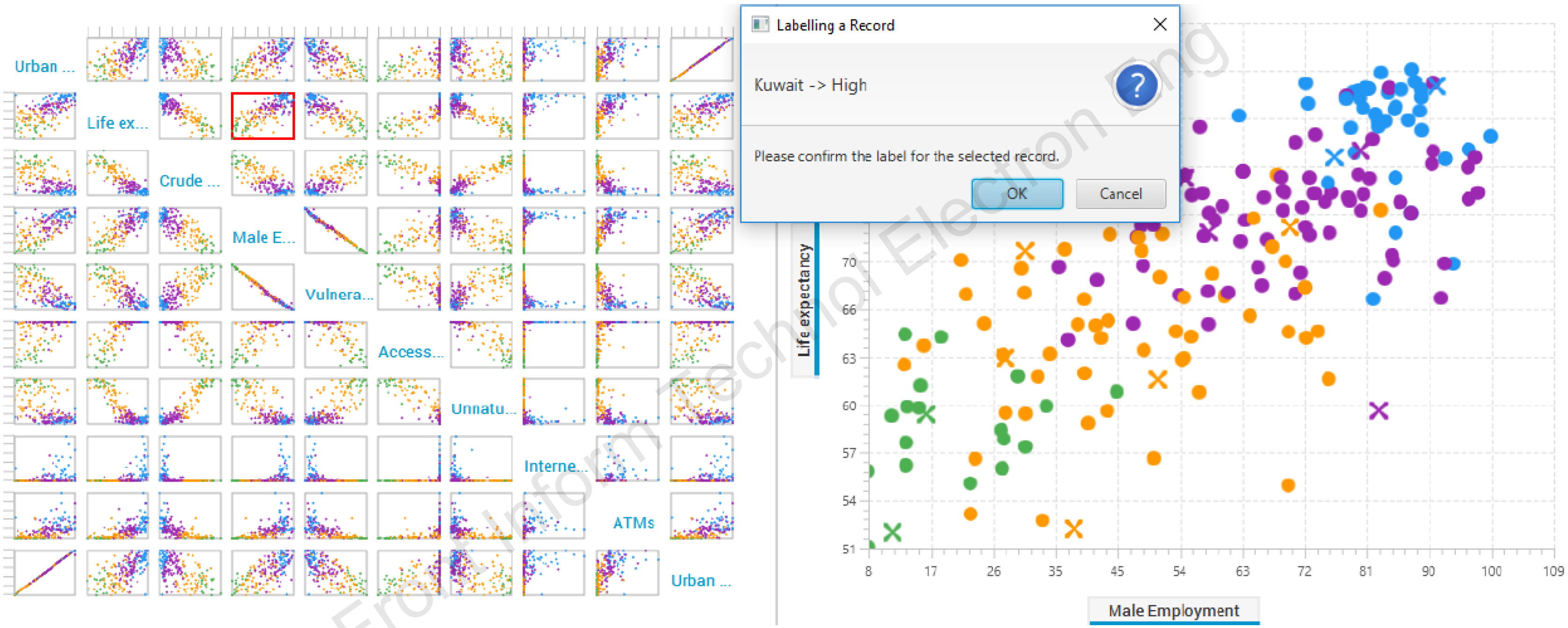
Methods

- The performances of visualisation techniques are compared with those of active learning methods in terms of accuracy.
- Three active learning methods were used: smallest margin, entropy-based sampling, and least significant confidence. The average accuracy in each step was used to compare the results.
- Greedy selection represents the best possible labelling result, and is the theoretical upper limit of what could be achieved.

Screenshot of mVis

The mVis tool, showing the SPLOM at top left, detailed scatterplot at top middle, similarity map at top right, and parallel coordinates at bottom right, for the MNIST2 dataset

Labelling in mVis



The user has selected the scatterplot of life expectancy vs. male employment in the SPLOM on the left and has selected the instance of Kuwait for labelling in the detailed scatterplot view on the right. The dialogue on the upper middle of the screen asks the user to confirm the label for that instance.

Research questions

- ***RQ1*** How do three individual visualisation techniques (similarity map, SPLOM, and parallel coordinates) compare in terms of accuracy of the resulting classifier?
- ***RQ2*** How does interactive visual labelling (IVL) with the three visualisation techniques compare to non-interactive labelling based on active learning (AL) selection?
- ***RQ3*** Which of the three visualisation techniques are rated higher by users in terms of user experience and confidence during selection of records to label?
- ***RQ4*** Do users adopt a different labelling strategy depending on the visualisation being used?

Study design

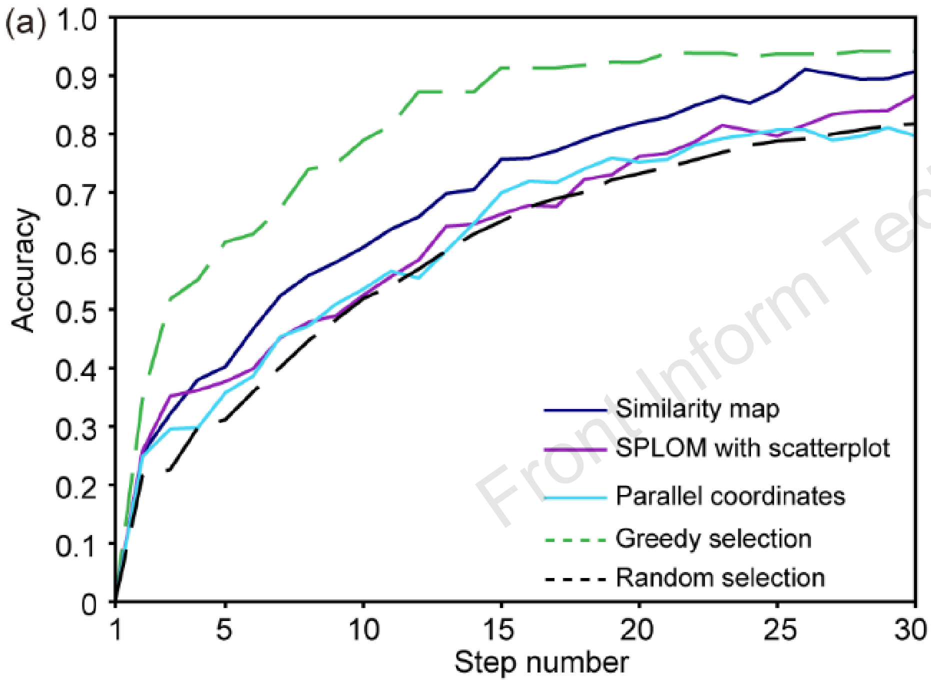
- A comparative experiment was conducted to evaluate the effectiveness of three individual visualization techniques for interactive labelling, based on which records were selected by test users for labelling.
- MNIST and the World Bank (WB) datasets were used for the experiment.
- The study was carried out in a quiet lab and 10 participants were recruited.
- During their test session, participants were asked to think aloud, and to ask questions. All sessions were captured by screen recording, and three sessions were additionally recorded with an external video camera for later analysis.

Procedure and tasks

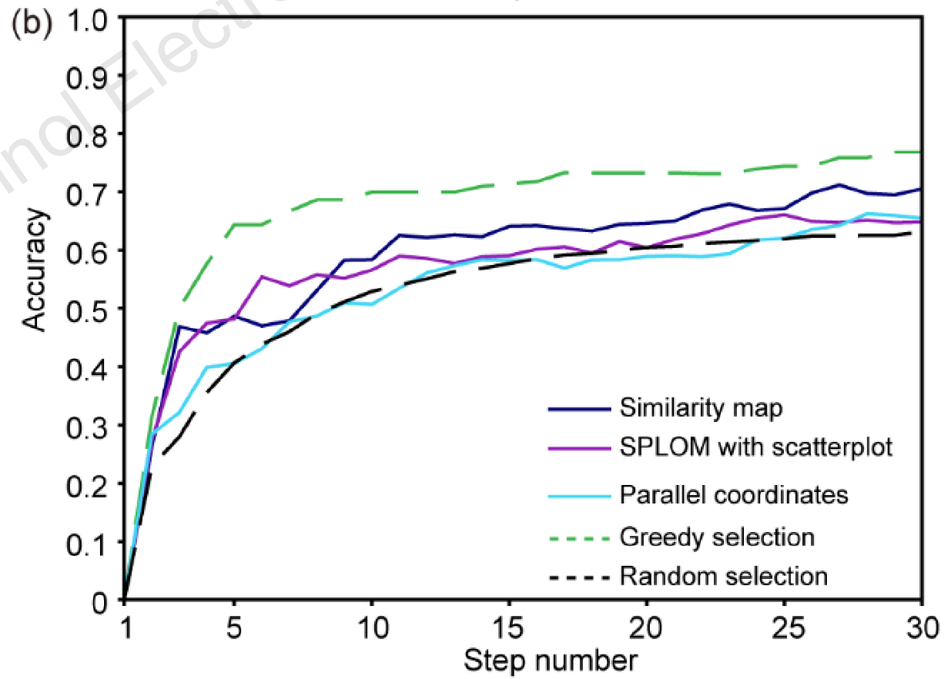
1. Opening: introduction and background questionnaire.
2. Tutorial: demonstration of mVis and practice with the MNIST2 dataset.
3. Test session: six experimental conditions, labelling each of the two datasets with each of the three visualisations.
4. Closing: interview with the participant.

Accuracy of visualisation techniques for labelling

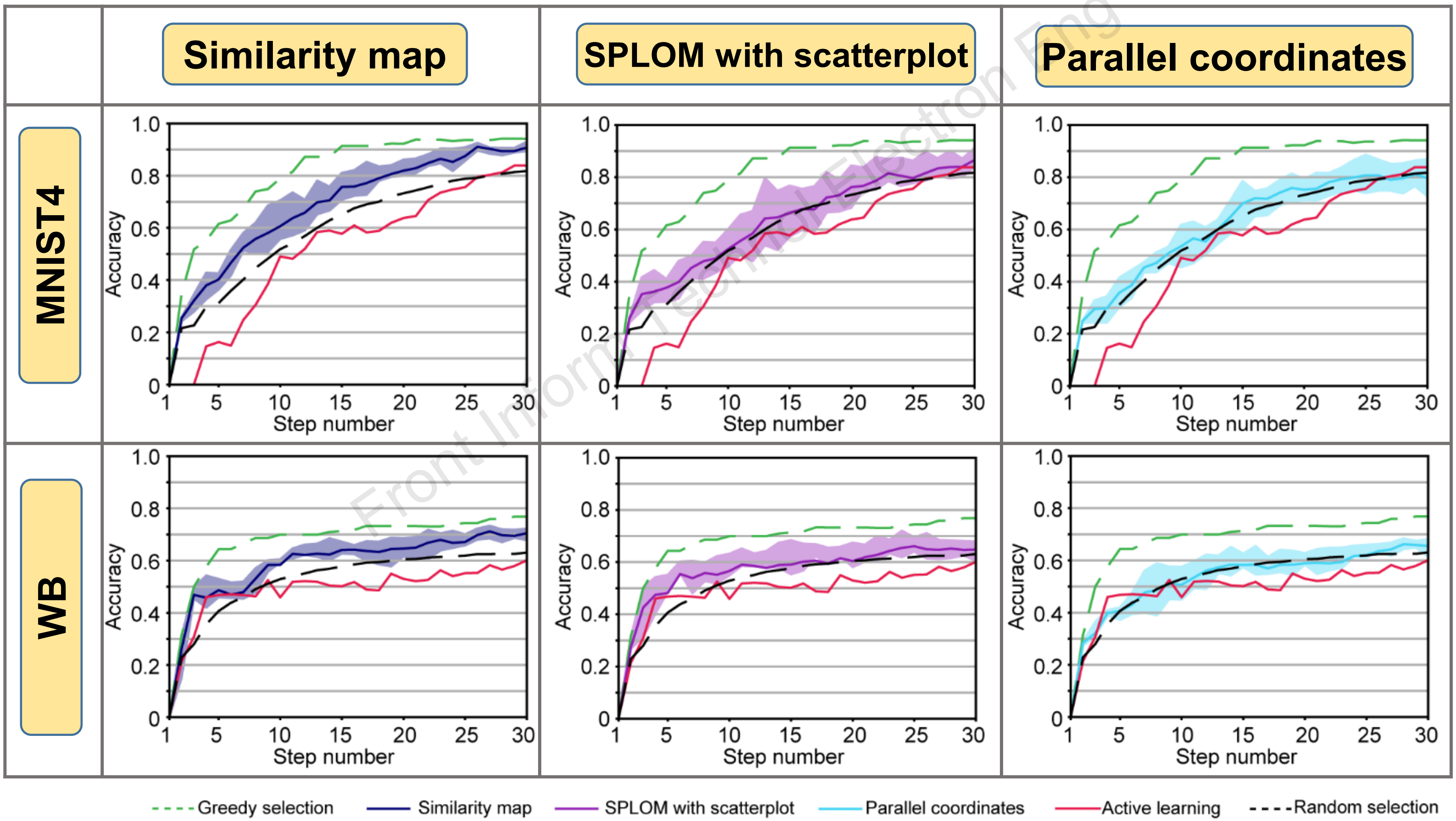
MNIST4



WB

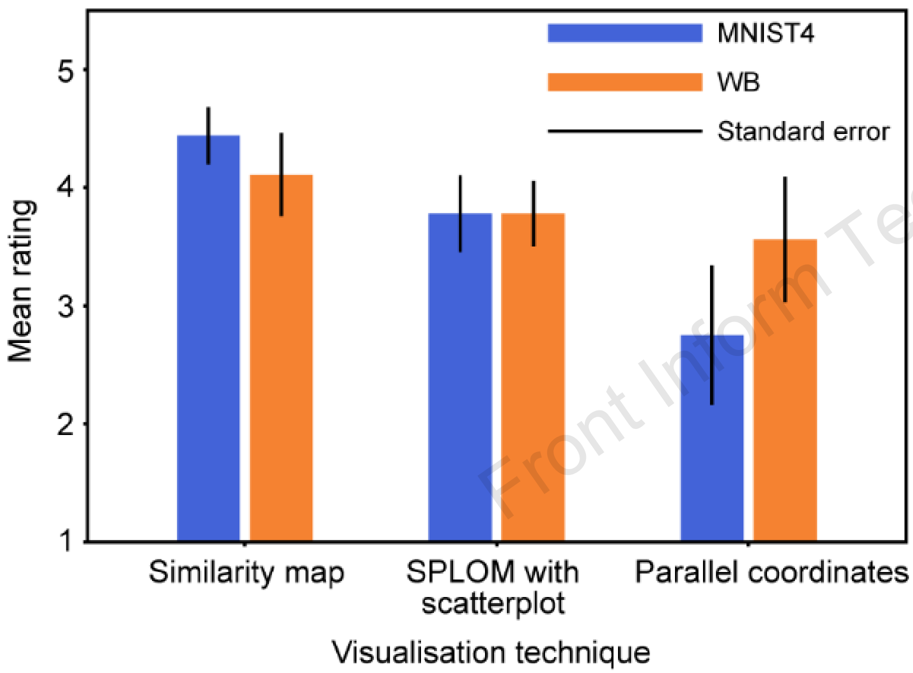


Accuracy of visualisation techniques for labelling 2

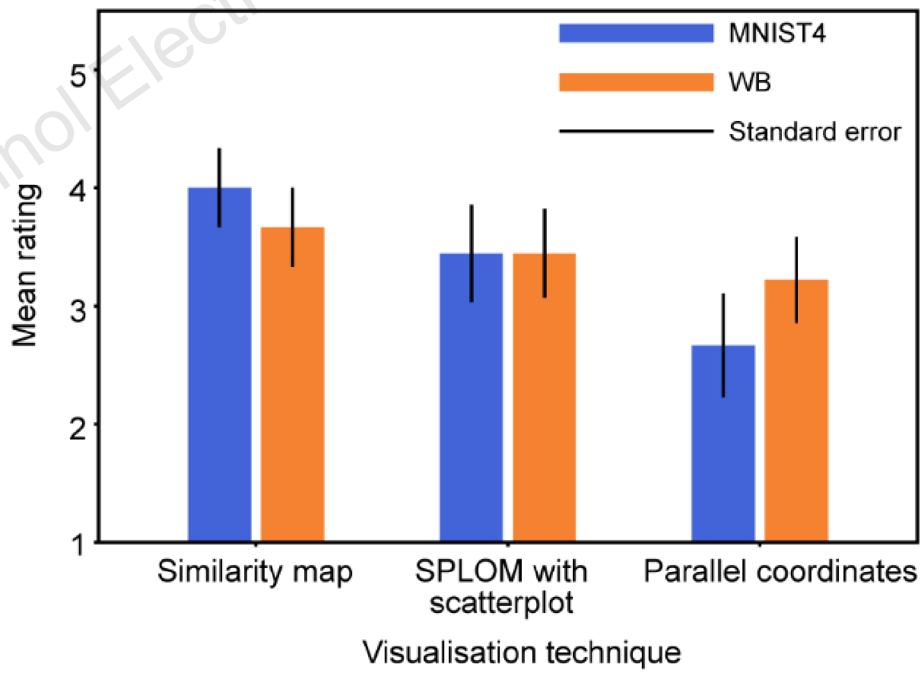


Mean rating given by the users

Labelling experience



Selection confidence



Discussion

- The results of the study are promising as they show that the classification performance of interactive visual labelling techniques can outperform those of active learning selection strategies.
- The similarity map seems to be the preferred view for labelling.
- When some example labelling is already available, some users prefer to use SPLOM with scatterplot for a more detailed insight.
- Parallel coordinates and SPLOM are suitable for finding relationships between dimensions, identifying clusters, and exploring data to make sense of it.

Conclusions

- We have compared three interactive visualisations with each other and with active learning for the purpose of labelling a multivariate dataset.
- All three interactive visualisations performed better than active learning algorithms, in terms of classification accuracy.
- The similarity map performed better than both SPLOM with scatterplot and parallel coordinates in both the MNIST4 and WB datasets.
- The results support the view that a user-in-the-loop approach is beneficial for creating training datasets.