Unsupervised Clustering of Eye Tracking Data

Fabian Göbel
Institute of Cartography and Geoinformation, ETH Zurich, Zurich, Switzerland
goebelf@ethz.ch

Henry Martin
Institute of Cartography and Geoinformation, ETH Zurich, Zurich, Switzerland
martinhe@ethz.ch

Abstract

The reading behavior on maps can strongly vary with factors such as background knowledge, mental model, task or the visual design of a map. Therefore, in cartography, eye tracking experiments have a long tradition to foster the visual attention. In this work-in-progress, we use an unsupervised machine learning pipeline for clustering eye tracking data. In particular, we focus on methods that help to validate and evaluate the clustering results since this is a common issue of unsupervised machine learning. First results indicate that validation using the silhouette score alone is a poor choice and should, for example, be accompanied by a visual validation using t-distributed stochastic neighbor embedding (t-SNE).

2012 ACM Subject Classification Computing methodologies → Cluster analysis

Keywords and phrases eye tracking; unsupervised machine learning; clustering; map task

Funding Swiss National Science Foundation Grant No.: 200021_162886.

1 Introduction

Human factors such as background knowledge or mental model as well as the cartographic design of a map have a strong impact on how people read and understand maps. Therefore, eye tracking experiments have a long history in cartographic research (refer to [5] for an overview). Answering where and how map readers spent their visual attention on a map is crucial to evaluate its design and usability. However, due to the high data rates of up to 1500 Hz that are needed to track the quickest movements of the human eye, these experiments usually result in large log files, which requires using automated data analysis.

There are many examples of the successful application of supervised machine learning methods on eye tracking data. For instance, Kiefer et al. recognize six predefined map activities based on gaze patterns [4]. However, the application of supervised learning methods requires ground truth information. In cases where these are not available, unsupervised learning techniques are an option to explore the data. An example is the work by Jonietz et al. who used GPS trajectory data to search for user groups with similar changes in mobility behavior [3].

This work evaluates the suitability of unsupervised machine learning methods for clustering eye tracking data. In the following, we first propose a clustering framework and discuss each of its steps with the help of real data from a prior eye tracking experiment. This allows to exemplify the framework and test a wide-range of different hyperparameters. Finally, the results are validated and evaluated by comparing the clustering results with t-SNE visualization, before we draw a conclusion and give an outlook on future work.

Spatial big data and machine learning in GIScience, Workshop at GIScience 2018.
Editor: Martin Raubal, Shaowen Wang, Mengyu Guo, David Jonietz, Peter Kiefer;
LIPICS Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

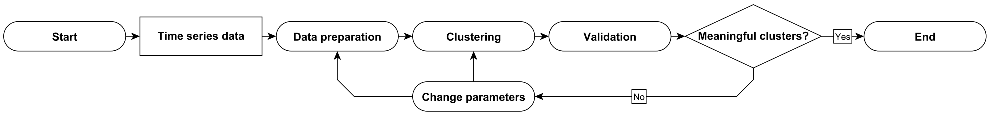


Figure 1 Overview of the clustering pipeline.

Table 1 Features used for clustering. Columns are interpreted as {statistic x indicator x type}. For instance, the first column reads: "total duration of trial".

statistics	total	mean, variance, median, 95th percentile, 5th percentile, total	total	rate of	mean, variance, weighted mean	covariance of	variance of	total
indicator	duration of	distance between succeeding, duration, angle between succeeding	number of, changes of curvature of, changes of direction of		dispersion x of, dispersion y of	dispersion of	duration on	dwells on
type	trial	fixations					City A, City B, Legend, Map	

2 Framework for clustering Eye Tracking Data

The proposed framework provides a structured approach to test different options of data preparation methods and clustering algorithms together with a wide-range of hyperparameters. Figure 1 gives a general overview of the iterative pipeline of the framework.

In order to explore patterns in gaze behavior, we use the data from an experiment published previously [2]. Participants performed a common comparison task on a map, which involved interacting with the legend. Three different maps with varying symbol density and three legend types were tested. In total, 18 samples were extracted for the clustering process.

2.1 Data Preparation

Because most clustering algorithms require tabular data, the pipeline start with a feature extraction step. We manually create features, that are specially tailored to reflect the scan pattern of a user's gaze behavior. Table 1 gives an overview of the features which quantify the spatio-temporal characteristics of the eye tracking data.

Feature extraction from time series data quickly leads to a high number of features: In our case, we create 37 features, which is a high value regarding 18 samples. Therefore, we include a feature selection step in our framework which filters features based on their dispersion. We use the interquartile ratio as it is robust to outliers [3].

2.2 Clustering

The next step is the clustering itself. Our framework applies three common clustering algorithms: *KMeans*, *spectral clustering*, and *DBscan*. KMeans is a simple and well-known clustering algorithm, however, it can only represent convex clusters. To address the case of non-convex clusters, we also apply spectral clustering. Furthermore, in this step, we test DBscan, which is often used for GPS trajectory analysis and does not require the definition of the number of clusters.

2.3 Validation

In general, unsupervised learning problems, such as clustering, are very hard to evaluate, because of the missing ground truth information. Although, in our scenario, the clustering algorithms produce a valid result for most hyperparameter configurations, the challenge to identify meaningful results remains. We propose a two-step approach to evaluate resulting clustering configurations: At first, we apply internal clustering validation measures, namely the silhouette score (SIL)[7] and the distribution of users per class (e.g. uniform distribution of users over all classes vs. all but one user in the same class) to exclude meaningless or trivial results. One disadvantage of the SIL is that it favors convex clusters, which might not reflect the structure of the data. Therefore, we do not use the SIL[1] for ranking but as a threshold to filter the results.

In a second step, we evaluate the clustering results by visual inspection similar to [1]. A dimension reduction method projects high-dimensional data to a human readable two-dimensional space. We chose t-distributed stochastic neighbor embedding (t-SNE), which has been proven to be a powerful tool for the visualization of high-dimensional data [6, 8]. Although, t-SNE can unveil structure in the data the method can not be used as a quantitative clustering algorithm, as it does not preserve distances [8].

In our framework, we use t-SNE to visualize structure in the data by projecting the unlabeled, high-dimensional data into two dimensions. We then color the data points in the resulting two-dimensional plot according to the clustering results. This allows to visually evaluate if the clustering result corresponds to the result of the t-SNE projection. If both results are corresponding, the clustering result is likely to be meaningful. If the results are not in line, it is still possible that both algorithms have uncovered different meaningful patterns but we can not validate our clustering result.

2.4 Example

We calculated all possible clustering results of the example dataset, by running the different algorithms with a wide range of hyperparameters (comparable to an extensive grid search). 29700 parameter combinations were tested, which results in 9952 valid clusterings (a result is valid if it has more than 1 cluster and if the algorithm converged). Following the framework, results with a SIL below 0 were excluded and only the ones with an interesting distribution of users among the different classes were manually picked. We then visualize the promising picks using the t-SNE algorithm and color the data points according to the label of the clustering result.

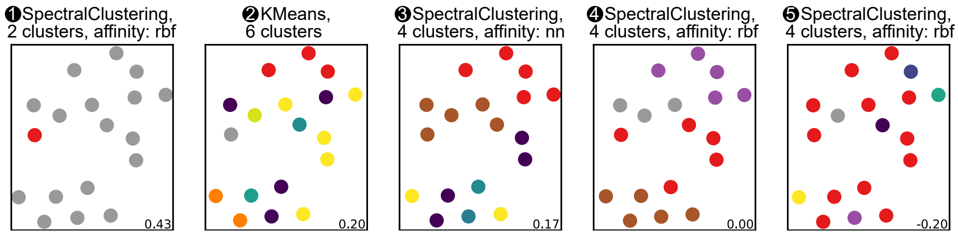
Figure 2a shows one single t-SNE projection colored with five different clustering results (result with negative SIL is included for comparison). The graphs show that the solution with the highest SIL (0.43) is a trivial solution where all points but one are in the same cluster. Although the forth clustering has a rather low silhouette score, it looks very promising. This might be an example of the inability of the SIL to acknowledge non-convex clusters.

Figure 2b shows the t-SNE label distribution of the forth clustering from Figure 2a with different perplexity values. Due to the stochastic nature and the dependency on the perplexity of the t-SNE, the position of a point and the distance between points varies greatly between the different results. The graphs show that the results by the spectral clustering are fairly robust and correspond well to the t-SNE projections with different hyperparameters.

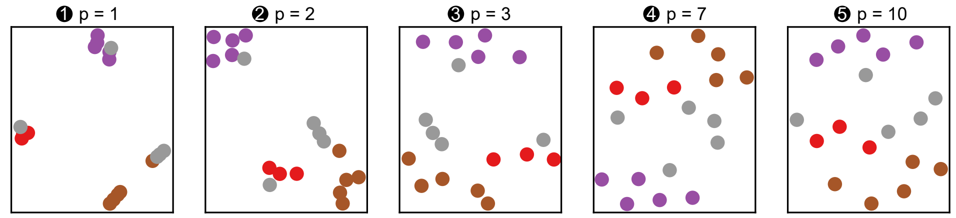
3 Discussion and Conclusion

In this work, we proposed a framework for unsupervised clustering with a special focus on the validation of the clustering results. The approach was exemplified with a data set of an eye tracking experiment on cartographic maps. We showed that a validation based on

[1] The SIL ranges between -1 and 1; values closer to 1 indicate well defined and separated clusters



(a) T-SNE graphs with a constant perplexity of 7. Colors indicate results of different clusterings. Headers state the respective algorithms. The SIL is shown at the bottom right of each plot.



(b) T-SNE graphs with varying perplexity (p). Coloring uses the same labels as in plot 4 from a).

the silhouette score alone can be misleading and should be accompanied by other validation methods. Promising results could be achieved by visual inspection based on t-SNE. Although we have demonstrated the framework with only very few data points, the approach is suited for large-scale datasets. The main contribution of this work-in-progress is the introduction of a technique to find a theoretically good solution in a structured way.

However, the question of how to finally choose the best clustering result is an ongoing topic of discussion. Although we could demonstrate the potential of t-SNE for evaluation of a clustering result, domain knowledge is crucial for both, choosing meaningful parameters and extracting more elaborate features that describe the interaction of a human with a map and the inspected map content in more detail.

References

1 K. Chen and L. Liu. Validating and refining clusters via visual rendering. In *Data Mining, 2003. ICDM 2003. Third IEEE International Conference on*, pages 501–504. IEEE, 2003.

2 F. Göbel, P. Kiefer, I. Giannopoulos, A. T. Duchowski, and M. Raubal. Improving Map Reading with Gaze-Adaptive Legends. In *ETRA '18: 2018 Symposium on Eye Tracking Research & Applications*. ACM, 2018.

3 D. Jonietz, D. Bucher, H. Martin, and M. Raubal. Identifying and Interpreting Clusters of Persons with Similar Mobility Behaviour Change Processes. In *The Annual International Conference on Geographic Information Science*, pages 291–307. Springer, 2018.

4 P. Kiefer, I. Giannopoulos, and M. Raubal. Using Eye Movements to Recognize Activities on Cartographic Maps. *Proceedings of the 21st SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pages 498–501, 2013.

5 P. Kiefer, I. Giannopoulos, M. Raubal, and A. Duchowski. Eye Tracking for Spatial Research: Cognition, Computation, Challenges. *Spatial Cognition & Computation*, 17(1-2):1–19, 2017.

6 L. v. d. Maaten and G. Hinton. Visualizing data using t-SNE. *Journal of machine learning research*, 9:2579–2605, Nov 2008.

7 P. J. Rousseeuw. Silhouettes: A Graphical Aid to the Interpretation and Validation of Cluster Analysis. *Journal of computational and applied mathematics*, 20:53–65, 1987.

8 M. Wattenberg, F. Viégas, and I. Johnson. How to Use t-SNE Effectively. *Distill*, 2016.