How to Optimize Gower Distance Weights for the k-Medoids Clustering Algorithm to Obtain Mobility Profiles of the Swiss Population

Alperen Bektas[1] and René Schumann[2]

Abstract—This piece of research aims to obtain mobility profiles of the Swiss population. To that end, a survey of the Swiss Statistical Office (FSO) called Mobility and Transport Micro-census (MTMC) is utilized. Along with a qualitative method *clustering*, the respondents in the survey are clustered based on their mobility characteristics to obtain their profiles. The clustering, in particular acquiring medoids (centrotypes or exemplars), helps us then to generate a synthetic population of Switzerland. To gain medoids of each cluster, the k-Medoids clustering algorithm is utilized which partitions instances based on their positions in a latent space (symmetric distance matrix). Distances that shape this space can be generated by various metrics e.g. Euclidean, Gower, Manhattan. Since in this study features are *mixed-type* (e.g. numeric, categorical, etc.), the Gower distance metric is preferred. In this study, the default weights of the Gower distance are optimized to obtain a higher Average Silhouette Width (ASW) value of the clustering results. ASW can be used to measure the quality of clustering results in which high value leads to higher intra-cluster homogeneity and inter-cluster dissimilarity. So, maximizing the ASW value improves the quality of the clusters which is the goal of the optimization. At the end, this process helps us to obtain more accurate mobility profiles of the Swiss population.

I. INTRODUCTION

Knowing mobility dynamics of a society is always advantageous for policy makers in terms of different aspects such as future investments. As an environmental aspect, nowadays there is a huge effort on decarbonizing mobility sector that constitutes around one-fifth of the global carbon-dioxide emissions [1], [2]. Several hard-measures, such as diesel bans or increasing taxes on fossil fuels, are under consideration to reduce total emissions from the mobility sector. In addition, there are also people-centric ways to handle the problem. One of them is changing mobility behaviors of individuals e.g. encouraging them to use more soft-mobility modes. For this purpose, it is helpful to learn mobility profiles (or mobility patterns) of individuals. Knowing which people use private car, bike, train, etc. frequently and their characteristics are important to create strategies to deal with decarbonization targets. To this end, in this piece of research, we aim obtain mobility profiles of the Swiss population based on their mobility characteristics. Clustering, as a qualitative approach, is employed to obtain the mobility patterns. Individuals who have similar characteristics are grouped in the same clusters. Afterwards, clusters, their members and a representative (medoid) of each cluster are obtained. Apart from exploring mobility patterns of the society, the clustering results can be then used to generate a synthetic population of Switzerland. Synthetic population generation is not the specific goal of this paper but the obtained results can be utilized as a milestone for future work. Since the medoids and the intra-cluster feature distributions are known, crafting a synthetic population stays one step ahead.

The Swiss Statistical Office (FSO) conducts quinquennial surveys to observe characteristics of the Swiss population including their mobility practices. The surveys are released with the name 'Mobility and Transport Micro-census (MTMC)' [3]. The last release refers to 2015, is employed as the main data set in this research. The data contain a large number of features, separated in several sub-tables, related to activities (e.g. trips), properties (e.g. car, accommodation) and characteristic of the respondents (e.g. age, education level). This research aims to cluster the individuals (respondents), not the activities. Hence, the features, which describe individuals' mobility characteristics such as their modal choices, are selected in a final data set that is employed to place the respondents in a latent space. The features qualify the activities e.g. origins and destinations of trips, are excluded. In the same vein, the features depict properties of the respondents, e.g. engine size of respondents' cars are not considered for the final data set. Basically, an ex-ante feature selection takes place before running the clustering algorithm. The selected features are illustrated in Section III.

The final data set consists of mixed-type features. It contains both numerical (continuous) and categorical features. Type of features are important when a selection of a distance metric takes place because some distance metrics (e.g. Euclidean) are able to handle only continuous features [4], [5] while a distance matrix is being created. For categorical features, a data transformation is required beforehand. The Gower distance [6] is capable of handling mixed-type data sets, which might contain all types of data, without any external transformation [7], [8]. There are some other approaches to handle mixed-type features such as cluster ensemble approach. But empirical research shows that the Gower distance provides a more effective way of tackling the issue of mixed-type features [9]. To that end, it is preferred in this study to create a distance matrix which contains pairwise distances (dissimilarities) among instances (respondents). These distances determine positions

[1]A. Bektas is with the SI-Lab, Ph.D. student at HES-SO Valais/Wallis, Rue de Technopole 10, 3960 Sierre, Valais, Switzerland `alperen.bektas at hevs.ch`
[2]R. Schumann is the head of the SI-Lab, Professor at HES-SO Valais/Wallis, Rue de Technopole 10, 3960 Sierre, Valais, Switzerland `rene.schumann at hevs.ch`

of instances in a multi-dimensional latent space. Therefore, the distance matrix can be utilized as a latent space which is then employed by the clustering algorithm for separation based on positions. A brief description of the Gower distance is given in Section II.

One of the potential extensions of this research (future work) is generating a synthetic population of Switzerland. A synthetic population can be generated for each cluster separately with narrowed down intra-cluster attribute space. Then generated cluster level synthetic data can be merged to obtain a full synthetic population. The idea is, first, obtaining medoids (together with their clusters), cluster sizes and intra-cluster feature distributions (e.g. marginal frequencies, conditional probabilities). Then, adding some noises to features of medoids according to intra-cluster feature distributions. At the end, a realistic synthetic population is gained. Having regard to this potential extension, in this research, one of the partitioning algorithms the k-Medoids (partitioning around medoids, PAM) is chosen to cluster instances. As the name suggest, it clusters instances around medoids [10], [11]. At the end of the clustering, it presents medoids of each cluster (together with other cluster members). In addition to generating a synthetic population, obtaining medoids can help to figure out the characteristics of clusters. They have the lowest cumulative distance towards other instances in the same cluster. The k-Medoids requires a latent space (distance matrix). It is compatible with the matrices that are created by the Gower distance metric [9]. Therefore, in this piece of research, the Gower distance and the k-Medoids are employed together to perform the clustering.

The k-Medoids algorithm gets the number of clusters information (k) as an input parameter. It is an unsupervised approach. Hence, the optimal number of clusters should be pre-specified. This ex-ante situation is one of the handicaps of it [12]. But there are some cluster validation methods which can help to get an idea about the optimal number of clusters (e.g. average Silhouette Width (ASW), Calninski and Harabasz Index (CH) and Pearson version of Hubert's Γ (PH) [13], [14]). ASW is one of the most widely used approaches that measures how well an instance is matched with its own cluster [14]. It is basically a fitness measure that reflects how maximized intra-cluster homogeneity and inter-cluster dissimilarity [15]. The idea for pre-specifying the optimal number of clusters is, trying different k-values in an interval for partitioning and appointing one of them that has the highest ASW value as the optimal number of clusters. Detailed information about how ASW values are used to pre-specify the optimal number of clusters is given in Section II.

In this study, while the profiles are being created, an optimization process takes place to improve the clustering results. The optimization targets obtaining clusters with higher quality. In other words, it takes aim at appointing the instances in the right clusters as much as possible. In general, the quality of clusters is measured by;

- Compactness: It reflects the intra-cluster cohesion (how homogeneous are the instances within the same cluster).

- A lower intra-cluster variation is an indicator of a good clustering.
- Separation: It measures how well-separated a cluster is from other clusters. A higher inter-cluster dissimilarity is an indicator of a good clustering.

ASW reflects both compactness and separation of clusters (see Eq. 4). So it can be used as a qualitative measurement. In this study, the ASW value of clusters is utilized as a metric that should be maximized. Firstly, the optimal number of clusters that has the highest ASW value is found. Then, its ASW value is improved. To this end, the default weights (all equal to 1 [6]) of the Gower distance metric are optimized while a distance matrix is being created. This optimization brings a higher ASW value of the clusters that leads to obtaining higher-quality. In Section IV, the AWS values with the default weights and with the optimized weights are compared. In Section II, detailed information about how optimization take place in the overall clustering concept is illustrated.

II. OVERALL ARCHITECTURE

The main goal of this research is obtaining mobility profiles of the Swiss population through clustering. While they are being obtained, an optimization process takes place to improve the quality of the clusters. The idea hinges on splitting the instances into the clusters more precisely. To this end, an overall architecture (concept) is created that draws all the procedure. It is split into two steps. The steps (illustrated in the Fig. 1) are briefly explained as following;

- 1st step: In this step the optimal number of clusters is obtained.
- 2nd step: The ASW value of the optimal number of clusters (obtained in the first step) is improved through optimizing the default Gower weights.

Blue flow-lines in the Fig. 1 contribute just to the first step. The red ones are for only the second step. The black flow-lines are step-neutral. They contribute to both.

A. The first step

In this step, first, the instances are placed in a latent space that is represented by a symmetric Gower distance matrix. All pairwise distances are created feature by feature. Distances on categorical features are calculated according to Eq. 1 while it is according to Eq. 2 for numeric ones.

$$D_{i,j,k} = \left\{ \begin{array}{ll} 1, & X_{i,k} \neq X_{j,k} \\ 0, & X_{i,k} = X_{j,k} \end{array} \right\} \qquad (1)$$

In the equation above, X_i and X_j indicate instances i and j whilst k represents a categorical feature. This dichotomous output of the equation can be adjusted also for ordinal features. The scale (between 0 and 1) is divided the number of ordinal values. For instance, if an ordinal feature contains 4 levels (e.g. 1, 2, 3, 4), the distance between 4 and 3 becomes 0.25 [16]. For numeric features, in the equation Eq. 2, r_c indicates the scale (range). It is simply $\max X_c - \min X_c$. The equation brings a normalization for values.
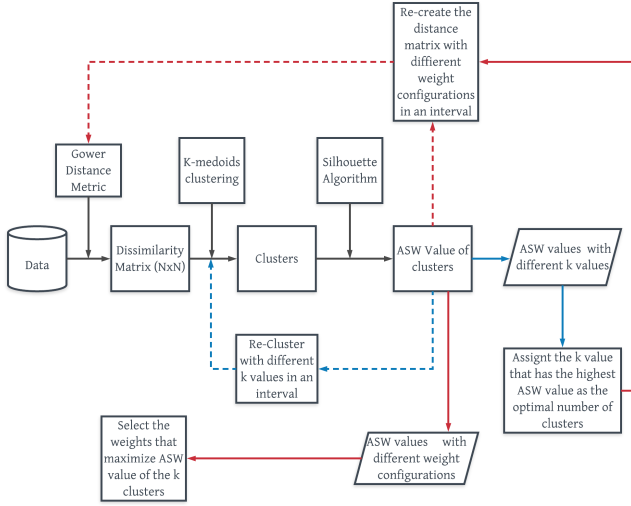
Fig. 1. The Overall Concept/Architecture (contribution of the colors to the steps; blue: the first, red: the second, black: contributes to both)

$$D_{i,j,k} = \frac{|X_{i,c} - X_{j,c}|}{r_c} \qquad (2)$$

Based on each feature, pairwise distances are generated and scaled between 0 and 1. After that, calculated n distances (n=number of features) are combined to obtain a final distance [6] (see Eq. 3). Each feature might have a different weight in aggregation. In the equation, *w* indicates the weight of a feature. By default, each weight equals to 1 [6]. They can be tuned (that is what this study focus on). After all pairwise distances are calculated, they are represented in a symmetric distance matrix. The matrix acts as an latent space whereby the k-Medoids partitions the instances into a pre-specified number of clusters.

$$D_{i,j} = \frac{\sum_{k=1}^{N} w_k D_{i,j,k}}{\sum_{k=1}^{N} w_k} \qquad (3)$$

When clusters are obtained, in the post-clustering phase, the ASW value of the k-clusters are calculated according to Eq. 4 which depicts the Silhouette value of instance *i*. The feature a_i indicates average dissimilarity of *i* to all other objects in the cluster *a* (the smaller the value, the better the assignment). Another feature b_i reflects the minimum dissimilarity of the instance *i* to all objects in any other cluster (the closest cluster to i except its own cluster). The Eq. 4 returns values between -1 and 1. Values close to 1 indicates that instance *i* is assigned to the proper cluster. For -1, it is vice-versa. Average *Sil* values of all instances (ASW) gives an idea about the quality of the clustering [15].

$$Sil = \frac{b_i - a_i}{\max\{a_i, b_i\}} \qquad (4)$$

As it has been mentioned in the introduction section, the k-Medoids requires pre-specifying the optimal number of

clusters k (see the pseudo-code in the Algorithm 1). To that end, as we see in the overall concept (see Fig. 1), there is a loop indicated with dashed blue lines. In the loop, different k-values in an interval are tested. Basically, the k-Medoids employs the same latent space (matrix) with different k-values. ASW values of each k-value are recorded. Then, when the loop ends, the k-value that has the highest ASW value is assigned as the optimal number of clusters. At the end of the first step, the optimal number of clusters -k- and their ASW value are obtained.

Algorithm 1 Pseudo-code of the k-Medoids (PAM) Clustering Algorithm

---

Input: m (a dissimilarity matrix), k (the number of clusters)
Output: k clusters

    Initialization: select arbitrary k of the n cases as the medoids
1: **while** As long as the Mean Square Error (MSE) decreases **do**
2:     Distribute n cases to k medoids
3:     Update medoids from the cases added to the cluster
4:     Calculate MSE for each update
5:     **if** If MSE does not decrease anymore **then**
6:         **break**
7:     **end if**
8: **end while**
9: **return** $k\ clusters$

B. The second step

In the second step, the ASW value of the k-clusters, that is gained in the first step, is improved. To that end, the default Gower weights are optimized in a loop that is indicated with the dashed red line in Fig. 1. An optimization function `optim` [17] in `R` language is utilized which is detailedly explained in section III. According to each weight configuration, the instances are placed in a new latent space. In other words, the pairwise distances change for each weight configuration. Then, the k-Medoids partitions the instances into k clusters based on the new matrix. At the end, the ASW values of k clusters according to different weight configurations, are compared. The weight configuration that brings the highest ASW value for the k-clusters is chosen as the optimized weights. Thus, at the end of the second step, the optimized weights and the clusters that are created through them are obtained.

III. DETAILING CONFIGURATION

In this section, the configuration settings of the overall concept are explained. The flow in the Fig. 1 begins with an ex-ante feature selection. Some mobility related features of the respondents in the Micro-census survey are chosen heuristically in a final data set. They are used when the instances are placed in a latent space as the active features. The final data set consists of 6 features which of five are categorical (including ordinals) and only one feature is numeric. They are listed below.

1) Number of cars in the household
2) Number of daily trips
3) Modal-choice
4) Has-HT
5) Multimodality
6) Daily Distance

The first feature is the number of cars in the household. It is an ordinal variable. Individuals of a household which has no car available, choose probably either public transportation facilities or soft mobility modes with respect to the distance of trips. Or if a household has more than 2 cars, its members might not behave reluctantly to use them. Therefore the number of cars in households can influence daily mobility practices. The majority of the Swiss households have one car (around 50%). It is followed by 2 and 3 cars per household. Around 8% of the Swiss households has no car.

The second variable is the number of trips per day. It is an ordinal variable. A trip might be long like going to another city or might be just visiting a bakery/grocery. The majority of the population perform 2 trips per day. They consist one-fourth of the population. It can be interpreted in this way that these people are mostly commuters. They go from home to work in mornings and turn back at night (2 trips). Around 4% of the population perform just one trip. It means that they do not turn back to their domiciles. There are some outliers who perform up to 21 trips. But overall, the bulk of the population perform between 2 and 6 trips.

One of the important features is modal-choice. It is a categorical variable. In the survey (Micro-census) people are asked how they perform their mobility activities (trips) on a particular day. A respondent may use more than one mode to perform his trips (multimodality). In this case, his most frequently used mode is assigned as the main modal choice. It is called as primary means of transport.

The fourth feature is Has-HT. It is a dichotomous variable (yes or no). Half-fare (HT) travel-cards allow 50% reduction for the public transportation modes in Switzerland. So, ownership of it can be used to identify the mobility patterns. Owners of that travel-card use probably more public transportation than non-owners. Around 40% of the population has a HT travel-card.

Multimodality is another dichotomous feature. It is also a binary categorical variable (yes or no). According to the survey, if a respondent uses more than one mode e.g. bike and train, he is appointed as a multi-modal otherwise mono-modal. 70% of the population prefer monomodalilty. Just around 30% prefers multimodality. In general, mono-modals use their own properties e.g. private-car, bike, walking whilst multi-modals use at least one public transportation mode.

The last feature is the only numeric feature, daily kilometers. It is used to identify long-distance and short-distance performers. In average, the Swiss population performs around 37 kilometers per day. These kilometers are higher for train users whilst they are very low for walkers or bike users.

After the final data set is obtained with the features, in the first step of the overall concept, the goal is obtaining the optimal number of clusters. To this end, as it is indicated by the dashed blue line in the Fig. 1, the clustering process is repeated with different k-values in an interval (loop). The latent space (distance matrix) which is created through the default Gower weights, stays steady. But k-values change in the loop. Based on the same matrix, individuals are split into different k-clusters. The lower and upper bounds of the k-values are chosen 2 and 15 respectively ($k \in [2, 3, 4..15]$). Due to the computational complexity (high computational time), the upper bound of the interval is set to 15. Thus, instances in the matrix are split into 2 to 15 clusters. For each k-value, a corresponding ASW value of the clustering results is calculated and compared as it is explained in the previous section. The k-value that has the highest ASW value is appointed as the optimal number of clusters. The results of this step is illustrated in section IV.

The second step aims to optimize the default Gower distance weights to obtain a higher ASW value of the k-Medoids clustering which is employed to gain mobility profiles of the Swiss population. The optimization targets to improve the ASW value of k-clusters (k is obtained in the first step). Improving the ASW value leads to a higher intra-cluster cohesion and inter-cluster separation. This improves the quality of the clusters. To this end, optim function in R language is employed (see Algorithm 3). Basically, the optim function takes the initial parameters (the default weights in this case), and it tunes them to minimize the return of a target function (function A in this case). The function B in below optimizes the default weights to minimize the return of function A. The function A places the instances into a latent space that is shaped by the input weights which come from the function B. Then, it partitions the instances into k-clusters (k is the optimal number of clusters found in the first step). It returns the negative ASW value of the clustering results. Since optim function aims to minimize the return of a target function, the function A returns negative of the ASW values to maximize it. Upper and lower bounds of the weights (see Algorithm 3) are set as 1 and 3 respectively ($w \in [1, 3]$). Due to comparability with the default weights, the lower bound is set to 1. The upper bound is set to 3 to prevent dominance of one variable to the clustering results. All selected features qualitatively important and aimed to contribute to the clustering process.

Algorithm 2 Function A to find the ASW value of the clustering with the input weights

---

Input: weights
Output: ASW value
1: Calculate a Gower distance matrix with the weights
   $Gower\_dist \leftarrow daisy(data, weights = weights)$
2: Partition the cases into k clusters
   $clusters \leftarrow pam(Gower\_dist, k)$
3: Calculate the ASW value of the cases in the k clusters
   $ASW\ value \leftarrow clusters\$silinfo\$avg.width$
4: **return** $-ASW\ value$

Algorithm 3 Function B to optimize the default weights

Input: default weights,upper bound(u), lower bound(l)
Output: optimized weights
  1: Calculate the optimized weights through the optim function

$optimized\_weights \leftarrow optim(par = weights, fn = function\ A, lower = 1, upper = u, method =' L - BFGS - B')$

  2: **return** $optimized\_weights$

Features	Optimized Weights
N-of-cars	1.000000
Has-HT	2.469693
Trips	1.000000
Daily Distance	1.000000
Modal-Choice	3.000000
Multimodality	2.640402

Cluster Number	Absolute size of the clusters
1	3890
2	1645
3	706
4	7973
5	941
6	503
7	1053
8	3138
9	461
10	2519
11	3952
12	493
13	1217
Total	28491

IV. FINDINGS

In the first step, the optimal number of clusters is obtained. Different k-values in an interval [2-15] are tested and the one that has the highest ASW value is appointed as the optimal number of clusters. The k-values and the corresponding ASW values are illustrated in the Fig. 2 below.
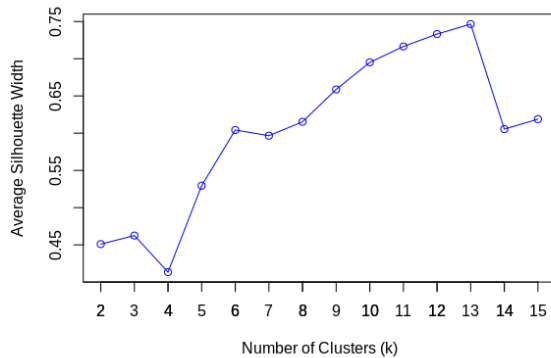

Fig. 2. Number of clusters and the corresponding ASW values

As can be seen in the figure, 13 clusters are the optimal (ASW value=0.7465). The corresponding ASW values have an upward-trend until 13. They peak at 13 and then decrease. For some k-values such as 4, the corresponding ASW value is lower than the previous and the next (3 and 5). The reason can be weak separation. The instances are not well-equilibrated into the clusters. Similar instances appear in different clusters. Therefore, adding a new cluster or removing one cluster can improve the ASW values in these cases. According to the Fig. 2, if the instances are split into 13 clusters, a better intra-cluster cohesion and inter-cluster separation is obtained. Thus, at the end of the first step, the optimal number of clusters is pre-specified as 13. In the second step, the goal is improving the ASW value (0.7465) of the 13 clusters. To this end, the function B is run. It employs the function A as a target function and aims to tune the initial parameters (the default weights) to minimize the return of the target function. After its run, the function B returns the optimal weights that are illustrated in the Table I.

According to the optimization results, some features (e.g. modal-choice and multimodality) have higher weights than others. Weights of there features (n-of-cars, trips and daily distance) do not change with the optimization. They are not as important as others to split the instances. Increasing their weights does not change the quality of the clustering. But they still play a role for distributing instances in the latent space. In other words, not only high-weighted features but also low-weighted ones contribute to the separation collectively. Based on the optimized weights, the instances are placed in a new latent space. Since the latent space is represented by distance matrices, a new Gower distance matrix is generated along with the optimized weights. Then, the instances are split into 13 clusters based on the distances. The ASW value of the clusters that are created through the optimized weights is observed as 0.8458. It was 0.7465 with the default weights. There is a significant improvement. As a control, the second highest ASW value is also checked. The second most optimal number of clusters is obtained when the k-value equals to 12 (see Fig. 2). After the optimization, its ASW value is improved to 0.8349 which was around 0.73 before. The optimization improves the ASW value but it is still less than 13 clusters'. In brief, the same instances are split into 13 clusters with the default and the optimized weights. It is observed that, with the optimized weights, the ASW value is improved, which leads to obtaining the clusters with higher quality.

The absolute sizes of the obtained clusters are illustrated in the Table II. As we can see sizes are quite heterogeneous. When we look at the medoids of the clusters, it is observed that, 4 of 13 medoids are private-car users. Their cluster sizes are significantly higher than others e.g. cluster 4. In total, they constitute around 60% of the population. These

four clusters differ from each other through the features like e.g. multimodality, daily kilometers. Walker medoids and their clusters come second place in terms of cluster size. There are two medoids who prefer on foot as their primary mode of transport (cluster 1 and 10). As might be expected, their daily kilometers are significantly lower than others' and than the average, since people generally do not walk for long distances. Members of the one walker cluster have a HT travel-card whilst the member of the other one don't. These clusters differ from each other through HT travel-card ownership. There are two train user medoids. Both of them make significantly more kilometers than other medoids. One of them has a HT travel-card whilst other one does not. Both prefer multimodality. There are also two bike/e-bike user medoids. They make slightly higher kilometers than walkers. One of them prefers multimodality who has also a HT travel-card. The second one has no HT travel-card and he prefers monomodalilty. There is only one cluster whose medoid prefers tram as a primary mode of transport. Its cluster size is one of the lowest. The medoid lives in a city center where tram exists. He makes more or less the same kilometers as the bike/e-bike users that is significantly less than the average. There are two bus user medoids whose cluster sizes are the lowest. One of them is multi-modal. They make more or less the same kilometers with the tram user medoid. One of them has a HT-travel-card whilst other one does not. In summary, mobility profiles hierarchically organized the active features so that the primary mode of transport of the medoids can be shared by different clusters which are mainly distinguished by has-HT and multimodality features.

V. LIMITATIONS

In the first step of the experiment, different k-values in an interval are tested (in a loop) and the k-value that brings the highest ASW value is appointed as the optimal number of clusters. In the loop, k-values vary between 2 and 15. This is the first limitation. The peak is found as 13; after which, the ASW values decrease. But there might be potential k-values out of the interval that improve the ASW value more than 13. Due to computational complexity, the first peak is chosen as the optimal. We wanted to look the trend out of the boundaries (2-15) with 10 step size (e.g. k = 20, 30, 40 ... 100). But the computation took days. Therefore we could not reflect that result in this paper.

The second limitation is the upper bound of the weights. Due to preventing the dominance of a feature, it is set to 3. High computational time because of very small step size (seven digits), is another reason for that. But as we see in the Table I, the weight of model-choice is found at the border. It might increase if the upper bound is set more than 3. With a broader range of weights, a higher ASW value can be obtained.

VI. CONCLUSION

In methodological terms, this research gives a structured concept to improve the quality of clusters through the opti-

mization. As a distance metric, the Gower distance is utilized to create a latent space, whereby the k-Medoids partitions instances into the pre-specified number of clusters. Its default weights are optimized to improve Average Silhouette Width (ASW) value of the clusters. This improvement leads to a higher intra-cluster homogeneity and a higher inter-cluster separation. At the end, by the help of the optimization, more accurate clustering results are obtained. In an empirical point of view, this research provides a structured inside of the mobility dynamics of the Swiss population. As a next step, the results can be used to generate a synthetic population of Switzerland. To this end, the obtained medoids and the intra-cluster feature distributions are going to be utilized.

REFERENCES

[1] O. Edenhofer, R. Pichs-Madruga, Y. Sokona, E. Farahani, S. Kadner, K. Seyboth, A. Adler, I. Baum, S. Brunner, P. Eickemeier, *et al.*, "Summary for policymakers climate change 2014, mitigation of climate change," *IPCC 2014, Climate Change 2014: Contribution of Working Group III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, 2014.

[2] A. Bektas, K. Nguyen, V. Piana, and R. Schumann, "People-centric policies for decarbonization: Testing psycho-socio-economic approaches by an agent-based model of heterogeneous mobility demand," *CEF Conference*, june 2018.

[3] ARE/BfS, "Verkehrsverhalten der Bevlkerung Ergebnisse des Mikrozensus Mobilität und Verkehr 2015," *Federal Office for Spatial Development and Swiss Federal Statistical Office*, 2017.

[4] Z. Huang, "Extensions to the k-means algorithm for clustering large data sets with categorical values," *Data mining and knowledge discovery*, vol. 2, no. 3, pp. 283–304, 1998.

[5] C. Hennig and T. F. Liao, "How to find an appropriate clustering for mixed-type variables with application to socio-economic stratification," *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, vol. 62, no. 3, pp. 309–369, 2013.

[6] J. C. Gower, "A general coefficient of similarity and some of its properties," *Biometrics*, pp. 857–871, 1971.

[7] M. Ichino and H. Yaguchi, "Generalized minkowski metrics for mixed feature-type data analysis," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 24, no. 4, pp. 698–708, 1994.

[8] T. F. Cox and M. A. Cox, *Multidimensional scaling*. Chapman and hall/CRC, 2000.

[9] J. van den Hoven, "Clustering with optimised weights for gowers metric," *Netherlands: University of Amsterdam*, 2015.

[10] L. Kaufman and P. J. Rousseeuw, "Partitioning around medoids (program pam)," *Finding groups in data: an introduction to cluster analysis*, pp. 68–125, 1990.

[11] J. Han, M. Kamber, and A. K. Tung, "Spatial clustering methods in data mining," *Geographic data mining and knowledge discovery*, pp. 188–217, 2001.

[12] I.-S. Kang, T.-w. Kim, and K.-J. Li, "A spatial data mining method by delaunay triangulation," in *Proceedings of the 5th ACM international workshop on Advances in geographic information systems*, pp. 35–39, ACM, 1997.

[13] R. J. Campello and E. R. Hruschka, "A fuzzy extension of the silhouette width criterion for cluster analysis," *Fuzzy Sets and Systems*, vol. 157, no. 21, pp. 2858–2875, 2006.

[14] U. Maulik and S. Bandyopadhyay, "Performance evaluation of some clustering algorithms and validity indices," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 12, pp. 1650–1654, 2002.

[15] P. J. Rousseeuw, "Silhouettes: a graphical aid to the interpretation and validation of cluster analysis," *Journal of computational and applied mathematics*, vol. 20, pp. 53–65, 1987.

[16] J. Han, J. Pei, and M. Kamber, *Data mining: concepts and techniques*. Elsevier, 2011.

[17] E. Zurich, "General-purpose optimization (optim function)." https://stat.ethz.ch/R-manual/R-devel/library/stats/html/optim.html.