Use of Machine Learning for Expanding Realistic and Usable Routes for Data Analysis on Sustainable Mobility

Fabian Schirmer[a], Andreas Freymann[b], Anamaria Cristescu[c] and Niklas Geisinger[d]

Fraunhofer IAO, KEIM, Esslingen, Germany

Keywords: Neuronal Networks, Cloaking, Data Protection, Location Data, Mobility, Machine Learning.

Abstract: The current mobility or the transition to more sustainable alternatives are constantly changing. For Promoting a sustainable mobility and for investing in a proper infrastructure, we need accurate data regarding the mobility behavior. Gathering location information such as GPS can help to improve the charging infrastructure and the bicycle or pedestrian paths. This motivates the citizens to use sustainable means of transportation such as bicycles or electric cars. However, using personal information via GPS data can cause some challenges: preserving data privacy while keeping data quality to get useful analysis results. This paper presents an advanced approach of processing GPS data based on machine learning and spatial cloaking in contrast to current approaches focusing on common algorithms only. The evaluation has been conducted by generating simulated GPS trips. As a result, the presented approach provides an algorithm that prevents a complete loss of useful data while protecting the privacy of each user in cases where cloaking areas are close together.

1 INTRODUCTION

One of the biggest challenges that governments must face worldwide, is to achieve the Paris Agreement goals, which demand to keep the temperature below 2°C and reduce emissions (Pan et al., 2017). The transport sector is a crucial intervention field since it plays a main role in global carbon emission (Ortmeyer and Pillay, 2001). In fact, almost 85% of transport emissions can be assigned to road travel (Ritchie, 2020). In Germany, sustainability of the transport sector has become a major concern of the government (Merkel, 2018). Germany is one of the most motorized countries globally, where the private usage of the car is up to 43% due to the total number of trips (Bundesministerium für Verkehr und digitale Infrastruktur, 2020). As an example, Chemnitz, a German city, is no exception, where the private usage of a car is being a preferred mode of transport (Chemnitz, Stadt der Moderne, 2020).

Promoting sustainable development in the transport sector is, therefore, an essential topic for science as well. According to a recent research on transportation planning measures, a better infrastructure can motivate citizens to swift to more sustainable modes of transport which comprises, e.g., separated bicycle paths increase passengers' readiness to use the bicycle up to 55% (Wardman, 2007). (Martens, 2007) demonstrates that high-quality cycling infrastructure increases bike-and-ride mobility behavior. This implies that satisfaction with the urban mobility infrastructure is crucial for shifting to sustainable modes of transport, such as public transport, walking, or cycling. Therefore, this requires further studies on passengers 'mobility behavior and needs, if we want to design the infrastructure that facilitates the transport mode shift effectively.

The city of Chemnitz recently launched the project New Urban Awareness of Mobility in Chemnitz (NUMIC) to address the issue of mobility shift (Chemnitz, Stadt der Moderne, 2020). Within this project, a smartphone application has been developed to track trips of the citizens of Chemnitz via GPS. By sharing the GPS information of their trips, the citizens can contribute to the understanding of the mobility behavior patterns, which would help to increase the sustainability of the mobility infrastructure. GPS data collection has become a popular method for the evaluation of mobility behavior (Niu et al., 2014). Such approach can be used to connect the information about the infrastructure with specific location data. Another

[a] https://orcid.org/0000-0002-7032-8242
[b] https://orcid.org/0000-0002-3735-4545
[c] https://orcid.org/0000-0002-5299-1972
[d] https://orcid.org/0000-0003-0224-8213

Schirmer, F., Freymann, A., Cristescu, A. and Geisinger, N.
Use of Machine Learning for Expanding Realistic and Usable Routes for Data Analysis on Sustainable Mobility.
DOI: 10.5220/0010395201560163
In *Proceedings of the 6th International Conference on Internet of Things, Big Data and Security (IoTBDS 2021)*, pages 156-163
ISBN: 978-989-758-504-3

feature of the application is that the user can make suggestions for the infrastructure improvements for an exact GPS location. The tracking of the entire trip and the location-based notifications can then be used to analyze the mobility behavior and needs. Based on the results, planners can improve and develop the existing mobility infrastructure. However, the main challenge in this context is to ensure the data privacy because tracking of GPS trips usually comprises important personal information. A possible data protection solution is to reduce the precision of the GPS trips, which can be done by cloaking the start point and end point of each trip. This is done by standard cloaking algorithms that replace GPS information of the trips within a specific area. This solution prevents the possibility of extracting personal information, and as a result, it does not allow the identification of a tracked person. However, even if the use of this standard cloaking algorithm protects the privacy, it inevitably leads to the loss of the most relevant GPS data. This paper presents an advanced approach that preserves relevant information of such GPS trips while simultaneously ensures the data protection and privacy of each tracked person. This approach suggests an algorithm, which expands GPS data points of a trip into another free defined cloaking areas to blur the start and the end of a trip. We call this algorithm Expanded Cloaking Algorithm (EXCL-Algorithm). We applied this algorithm to two specific scenarios to demonstrate that this algorithm maintains relevant data while ensuring data privacy. In the upcoming chapters, we describe these evaluation scenarios. The paper is structured as follows. Section 2 focuses on the related publications and provides insights into relevant topics, including cloaking algorithms and data privacy. Section 3 describes the problem scenarios including the use of a standard cloaking algorithm. The privacy conditions are explained in section 4. Section 5 presents the solution to cloak the GPS tracks based on EXCL-Algorithm. Section 6 summarizes the results and outcomes of the suggested solution and, finally, Section 7 closes this paper by providing the conclusion and an outlook on future work.

2 RELATED WORK

Tracking has become a prominent IT field, which comprises various technologies and methods developed to date. Especially widespread is the usage of location-based data in our daily life (Zhu et al., 2013). However, the data protection associated with tracking has become a controversial topic since GPS data can reveal highly personal information(Trujillo-Rasua and Domingo-Ferrer, 2015; Feng and Timmermans, 2017). This poses a challenge for the developers of tracking services because the data needed for the analysis has, in many cases, to be protected. In the case of the NUMIC project, the mobility tracking application requires sensitive data on commuters' daily routes to analyze the mobility patterns and infrastructure deficiencies; some parts of this data must be protected to avoid identifying the persons who participate. This section discusses several common approaches to data privacy protection, namely *data aggregation*, *k-anonymity*, *cloaking* and *extension*.

Data aggregation brings some information to a higher level by removing specific identifiers. However, a well-known weakness of this approach is that specific information can be re-identified, e.g. by enriching aggregations with other data (Sweeney, 2012).

This approach moves the users' identity away from the original position (Zhu et al., 2013). Claudio et al. (2005) address this issue in the context of location-based services (i.e. navigation services for weather forecasting or location-based marketing), which represent sensitive private information. Their research contribution regards a framework that evaluates the risk of using pseudonyms to protect sensitive data and presents using *k-anonymity* as a more secure approach. Niu et al. (2014) use the k-anonymity for location-based services to increase privacy by offering an algorithm based on dummy locations. The result also shows an increase in protecting the privacy level. However, the probability that all persons have similar sensitive data increases if data from many persons are collected. In that case, the k-anonymity lacks in anonymity (Claudio et al., 2005). In the specific case described in this paper, k-anonymity is not an appropriate approach since it is not determinable how many users participate. Since the number of app users in the NUMIC project is not estimable, this approach is for the evaluation of the collected data not suitable.

Another technique to protect data privacy is *spatial cloaking* (Jeansoulin et al., 2010; Chow and Mokbel, 2011), which is also acknowledged in the context of location-based services (Wang and Wang, 2010). Chow et al. (2011) describe a common approach that blurs sensitive information into a cloaking area. They contribute to it by offering a spatial cloaking algorithm designed for mobile peer-2-peer environments, which can increase the scalability, privacy protection, and effectiveness of the algorithm. Ghinita et al. (Ghinita et al., 2007) go one step further and present a distributed architecture for anonymous location-based queries, which should fill in the gap of existing solutions. They claim, for example, that centralized solutions cause a bottleneck related to the amount of location-based queries. Furthermore, Wang and Wang (2010) also claim that traditional spatial cloaking approaches lack a central service where all information comes together. Thus, they address an in-device spatial cloaking and propose using cloud service to get relevant information on the cloak region

The literature review leads us to a central dilemma: the necessary data protection automatically causes a greater inaccuracy of the GPS data, which in turn decreases the informative value of the data. Hohet al. addressed this issue by presenting an uncertainty-aware path cloaking algorithm using a novel time-to-confusion metric in the context of GPS trips of vehicles (Hoh et al., 2007).

Similarly, Scheider et al. (2020) displayed and examined a strategy to obfuscate GPS trips by *extending* the trip. In addition, they introduced a method for simulated crowding. These methods have numerous advantages, but also a few disadvantages. Firstly, the length of a trip increases by 50%. Secondly, it requires more than a doubled runtime.

Thus, this paper builds on the research findings of Scheider et al. (2020) and adds the spatial cloaking (Wang and Wang, 2010) to eliminate disadvantages of each algorithm. The spatial cloaking by itself deletes significant GPS data needed for further analysis, whereas the simulated crowding in the study of Scheider et al. produces too much unrealistic data (**?**). However, the combination of these two methods outweighs their disadvantages.

3 PROBLEM DEFINITION

Figure 1 shows a running example of how a common spatial cloaking can be applied. It illustrates a GPS trip between a start and a target area in the city of Chemnitz, Germany. Additionally, Feedback messages from the user can be added to the trip. While the path between these two areas is visible, the exact GPS points within the start area and target area have been blurred out (i.e., GPS points have been deleted) into cloak regions. This deletion generally enables the privacy protection of a user who uses, for example, a smartphone application to record a trip. The start and target points of the trip are not determinable anymore. This makes it impossible for an outsider to map the trip to a home, a working place, or other visited places of a specific person. However, the main trip between the start and the target area has been maintained, which can be used to connect a text message to a specific location on the path.

The size of each cloaking area depends on the population density. Large cloaking areas represent the urban areas with a lower population density, whereas small cloaking areas represent the urban areas with a higher population density. The described scenario shows a common way to use cloaking. However, in some specific scenarios that can occur in this context, the start and target points might be too close to each other. Based on the use of the standard cloaking algorithm, this results in a vast loss of GPS data or even in a deletion of the entire trip. Two scenarios can be identified, which are affected by the described problem of being to close to each other. They will be described in the following.

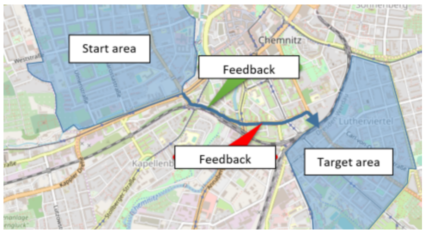

Figure 1: Spatial cloaking.

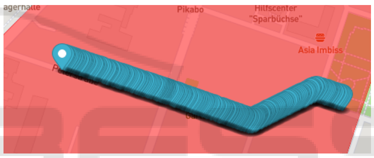

Figure 2: Scenario one: One area.

Scenario One - One Area: In some cases, especially for short trips, there is a high probability that the start and target points of a trip lie within a single cloaking area. More precisely, the start point and the target point of the trip would be the same and are respectively congruent (see Figure 2). If we use the common cloaking algorithm and the user remains in this cloaking area, all data points would be deleted, and no GPS data points would lie between these two areas. The only information left is the red area.

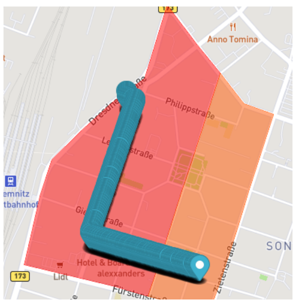

Figure 3: Scenario two: bordering areas.

Scenario Two - Areas Adjoin Each Other: Another scenario is when the start point and the target point lie in different cloaking areas, but these areas are directly adjoined to each other as it is shown in Figure 3. Similar to Scenario one, there would be no GPS data points between these two areas since all the data points would be deleted by the common cloaking algorithm. As a result, the important analysis could not be performed. Since two areas replace the tracked GPS data, more data is lost.

For both mentioned reasons, we created a unique cloaking algorithm (EXCL-Algorithm) to handle such situations by preserving GPS data as well as protecting the privacy.

4 PRIVACY CONDITIONS

For training and testing the EXCL-Algorithm, GPS trips were generated through the method of simulation by considering the data protection regulations. We used the open source software tool SUMO (Simulation of Urban MObility) (Sumo, 2020) since it enables to create realistic GPS trips. This includes for example, waiting times due to traffic lights or multimodal traffic (using a combination of several different modes of transportation). Finally, this tool enables to test different modes of transportation as well as to test the plausibility of such generated and expanded trips.

For the simulation of GPS trips, two random adjacent areas within the city of Chemnitz have been chosen. The map data is retrieved from OpenStreetMap (OSM) and imported into the SUMO model. For scenario one, one single area is used, whereas in scenario two, two cloaking areas lie directly next to each other. For each of those two scenarios, 20 trips simulated by SUMO are used. Several irregular trips have also been included in the scenarios (e.g., trips that start or end in dead-end roads) in order to verify the algorithm even for such complex cases. Furthermore, each scenario comprises ten trips simulated by vehicles and ten trips simulated by pedestrians as they pass parks and other pedestrian-only areas. The routes of the trips have various lengths, ranging from 24 meters in scenario one to 973 meters in scenario two.

With specifying the trips as input for SUMO, the simulation results, finally, comprises the position of each vehicle or pedestrian. To obtain realistic GPS trips, the exported data is then thinned out to only include data points with a random distance between 0.5 and 3 meters to each other, which results in having between 15 to 632 data points per trip. These different distances reflect the stop-and-go behavior of each user. Table 1 shows the average distance between two consecutive data points (column: "Avg. dist."), the range of total length (column: "Length"), and the number of data points (column: "Points") of the trips for the two scenarios, each further divided into car (denoted as C) and pedestrian routes (denoted as P).

Table 1: Overview of the simulated data.

Scenario	Length	Points	Avg. dist.
One (C)	303 - 831 m	177 - 460	1.73 m
One (P)	24 - 692 m	15 - 365	1.54 m
Two (C)	135 - 870 m	79 - 367	1.72 m
Two (P)	52 - 973 m	36 - 632	1.53 m

5 SOLUTION APPROACH

As mentioned before, using the common cloaking approach for both presented scenarios, all data points within the cloaking areas would be deleted. One the one hand, this ensures data protection as it makes it impossible to retrieve any information from the trips anymore. On the other hand, detailed data analysis is impossible. This section presents the Expanded Cloaking Algorithm (EXCL-Algorithm) that preserves the single GPS data points while simultaneously preserving the data privacy.

One important characteristic of the EXCL-Algorithm is the expansion of the available GPS trip (red bubbles), which is shown in 4. For that, the EXCL-Algorithm creates new cloaking areas (blue rectangles) where new artificial start and target points will be located. These artificial points are essential, because they expanse the original start and target points with artificial start and target points (green bubbles) in order to blur the original start and target points into just GPS points without having a starting or ending character. This results in protecting private information such as private addresses. In comparison to the common cloaking algorithm, which would delete all original GPS data points (red bubbles), the EXCL-Algorithm creates due to the additional artificial GPS data points (green bubbles) new artificial cloaking areas (blue rectangles): start and target area. As a result, the data points between these new two areas can be maintained and used for further analysis and studies. The protection of private data is also granted, since the trip cannot be linked to a specific person. Furthermore, the EXCL-Algorithm enables that it is not apparent which data points are artificially generated and which are real. The only information about each of the GPS trips is that it was manipulated by an expanded algorithm.

The cloaking of each GPS using the EXCL-Algorithm is done in two directions: from the start as well as from the target point of the trip. For that

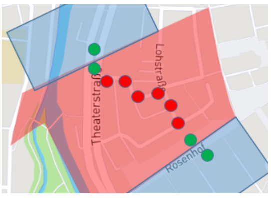

Figure 4: Scenario Two: Two areas.

a new artificial starting point as well as a target point are generated by adding and expanding a new starting and target cloaking area. Finally, both sides of each GPS trip are then expanded until the origin cloaking area (red area) is left. The new GPS track consists of a new artificial start point and a new artificial target point which both lie within a new and differing cloaking area (blue rectangles). All GPS points within the origin cloaking area are preserved.

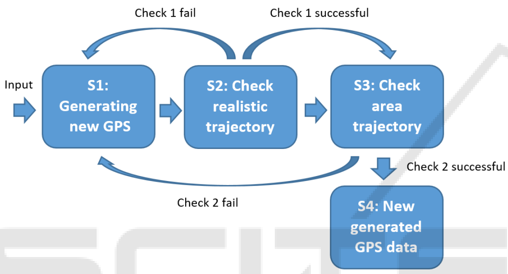

Figure 5: Steps of the EXCL-Algorithm.

Figure 5 gives an overview of the main steps of the EXCL-Algorithm explained in the respective subsection. Generally, the outcome of these steps are the artificially generated GPS points of a trip. The first step (S1) is responsible for generating new artificial GPS data for the new cloaking areas with the help of machine learning (see subsection 5.1). The result of the first step will then be checked afterwards in step two (S2) which checks the realistic trajectory and step three (S3) which is responsible for checking the area trajectory. If S1 or S2 fails, the previous step is active again. The checking of S2 and S3 includes the checking of the distance between each of the trajectories and whether the original area is deleted (see subsection 5.2 and 5.3). The results are then added to the start and to the target of the existing trip in the last step 4 (S4), which enables the possibility to use the whole trip for further data analysis.

5.1 Generating New GPS Data (S1)

Step S1 is responsible for generating new artificial GPS data, which is essential to expand the original trip into new defined cloaking areas. Generating new GPS data in addition to the existing ones is done with one of the machine learning algorithms of section 5.1.2. Thus, the following section deals, firstly, with structuring of the input data, which is needed for the machine learning algorithms and, secondly, with the comparison of different machine learning algorithms.

5.1.1 Input Data

The input data (i.e., GPS data points from the original GPS trip) is always different due to the different length of each trip. Large areas may allow more GPS data than in smaller areas. Moreover, it is important to consider the distances between the GPS points. Due to the different modes of transportation, distances from GPS points vary significantly. Regarding the training data, it is structured in a way that the data point *d(t)* is used to estimate the next data point *d(t+1)*. This training method is used for the whole GPS data. Figure 6 shows the procedure how the input data is processed by the EXCL-Algorithm. The black arrow on the upper side stands for the course of the training.

Figure 6: Structuring the input data.

The last GPS point, which is represented by the blue rectangle in Figure 6, is used for the first check of the realistic trajectory to control the distance between the GPS points in S2. Using this structure enables the machine learning algorithm to get a meaning of the overall trajectory.

5.1.2 Comparison of Different Machine Learning Algorithms

Based on the above mentioned input data, different machine learning algorithms (e.g., Neural Networks, Linear Regression, or Long Short-Term Memory) need to be trained and evaluated. The comparison is done based on accuracy and realistic values (S2 and S3). In addition, the amount of training data is required to generate the GPS data points. For this reason, different lengths are simulated, as shown in Table 1. The machine learning algorithms are based on supervised learning, where the training data represents the existing GPS data points. The decision of appropriate output data of the algorithm is based on checks of the approximately same distance between the GPS data and the leaving of the original area.

For our EXCL-Algorithm we compare three machine learning algorithms for generating new GPS data: Linear Regression (LR), neural network (NN) and Long Short-Term Memory(LSTM). Table 2 shows these machine learning algorithms in terms of their accuracy and duration of training in epochs. The values in the table are based on the lowest amount of training data. Increasing the training data will produce slightly better values for model accuracy and the number of epochs needed for good results. Finally, we picked the best training accuracy for the first try to expand the existing trip.

Table 2: Overview machine learning.

Criterion	LR	NN	LSTM
Model accuracy	10^{-8}	10^{-8}	10^{-9}
Epochs	-	40	20

Each machine learning algorithm is equipped with Early Stopping, which stops the training if the accuracy starts to increase.

Linear Regression. Linear Regression always produces well results for expanding each trip. Linear Regression uses the method of least-squares, which is the method of finding the best-fitting line for the training data by minimizing the sum of the squares of the vertical deviations from each training point to the line. (Seber and Lee, 2012, p. 35 ff.)

Neuronal Network. The general neural network consists of an input-layer, one or more hidden layers, and an output layer. Each of the layers, in turn, consists of neurons. (Géron, 2018, p. 253 ff.)

$$\sum_{i=1}^{n} x_i \cdot w_i \quad (1)$$

Calculation of the neurons (Géron, 2018, p. 253 ff.)

The output depends on the inputs x, weights w (see equation 1), and the activation function for each neuron. The activation function is linear. The optimizer is Adam (a method of stochastic optimization) and the loss function is mean squared error. This specific choice is based on the most accurate result for the output of the model.

Long Short-Term Memory. One strength of the Long Short-Term Memory (LSTM) is extracting useful information about historical records. These cells learn over a long-range dependency to forecast the data. Most of the other algorithms still have problems with such a behavior. (Hua et al., 2019) We used it in the following case to get a meaning of the whole trajectory and predict realistic values. The optimizer chosen is Adam and the loss function is mean squared error.

5.2 Check Realistic Trajectory (S2)

After the artificially generating GPS points in S1, we conducted the feasibility. The purpose of the first check (S2) is to ensure that the generated GPS data points are as realistic as possible (the distances between each of the GPS points are nearly the same). If the first validation is successfully done the next step S3 is checked. If not, the generation of new GPS data points will generate a new GPS point. The GPS data points should be generated in a realistic manner, so that the generated GPS points should be close to each other in the same way as the data given. This means that the distances (d) between each GPS point are nearly identically to the other points follow afterwards (see equation 2).

$$d(t+1) - d(t) \approx d(t+2) - d(t+1)... \quad (2)$$

5.3 Check Area Trajectory (S3)

If the generated GPS data points for the start and target area are added and the first check (S2) was successful, the next check (S3) will be performed. The second check (S3) examines whether the origin area is left or not. This enables preserving the whole trip and tries at the time to minimise the artificially generated GPS data points as much as possible without losing its data. The expansion of each GPS trip must be as short as possible. The automated generation of new GPS points is exactly specified. If a new area is entered, the generation of new GPS data is finished. However, if the newly entered areas at the beginning of the trip and at the end of the trip lie directly next to each other (scenario two, see figure 3), the generation of the GPS points is done again. Equally if the beginning and the end of the trip lie in the same zone (scenario one, see figure 2), the generation is done again, too.

```
while(GPSPointWithinOriginArea){
    generateNewGPSPoints();
}
if (startArea==targetArea){
    startNewGenerationGPSPoints();
}
```

Above listing shows the logic of the second check (S3). As long as the current generated GPS points stay within the original area, the generation of new GPS points will continue. If the origin area is left, the generation of new GPS points is over. If the start and target area are the same, the generation of new GPS points will start from the beginning to get different areas for the start and target area.

5.4 Generated New GPS Data (S4)

S4 does not only create a new artificial GPS data. To make it more difficult to differentiate between the original GPS data and the artificially generated GPS data, S4 also adds a small noise, as simulated data always seems identical. Bringing in a small deviation creates a more realistic data and complicates it to distinguish from the original data. Scenarios like waiting at traffic lights or bus stops produce more GPS data at one specific point than running at a higher pace. These both scenarios are also taken into account.

6 RESULTS

The results show that the EXCL-Algorithm can be performed with one of the presented machine learning algorithms of section 5.1.2. They produce similar results and provide realistic new trips indistinguishable from real trips. Unlike the common cloaking algorithm, the EXCL-Algorithm preserves the data of both presented scenarios that can be used for further analysis and studies and simultaneously protects the privacy as the trip cannot be mapped to a specific user. This also means that additional data such as the feedback assigned to each trip are not deleted in the scenario one and scenario two. In particular, trips with smaller distances can now be used for subsequent analysis. In the following the results and benefits of the EXCL-Algorithm will be presented for each of the two scenarios described in Section 3.

6.1 Extrapolation Scenario One

Figure 8 shows the trip expanded with the EXCL-Algorithm. In comparison to Figure 7, the entire trip will be preserved. Normally, all data points lying in the red area will be deleted. So for further analysis only the red area is available.

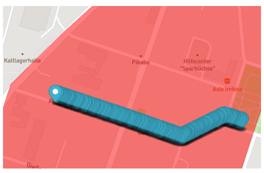

Figure 7: Initial situation: scenario one.

Figure 8: Result of scenario one using EXCL-Algorithm.

Due to the use of the EXCL-Algorithm the trip still can be used because the EXCL-Algorithm defines a new start and target area (yellow areas in Figure 8) of the trip. For an outsider, it is impossible to distinguish between the artificially generated GPS data and the original tracked GPS data from the users.

6.2 Extrapolation Scenario Two

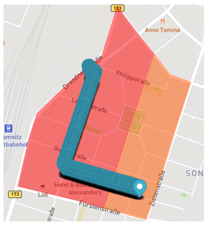

Figure 9: Initial situation: scenario two

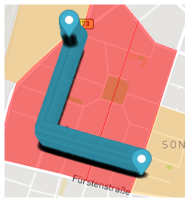

Figure 10: Result of scenario two using EXCL-Algorithm

As well as scenario one, scenario two can also be optimized through our EXCL-Algorithm. Figure 10 demonstrates the result of the EXCL-Algorithm. Compared to the initial consideration (see Figure 9), two completely new areas (yellow territories see figure 10) are entered, which are the new starting area (the top left corner of Figure 10) and the target area (the lower right corner of Figure 10). All data points within the red area and orange area (Figure 9) are preserved. Normally the whole data points in Figure 9 cannot be used for the mobility analysis. Due to the use of our EXCL-Algorithm the complete number of data points can now be used for further work.

7 CONCLUSION

In this paper, we presented an expanded cloaking algorithm, called EXCL-Algorithm, based on supervised learning to expand real GPS data points of a trip by adding new artificial start and target points within new defined cloaking areas. The EXCL-Algorithm enables to ensure the privacy of each user as well as preserves the main trip data in reference to two specific scenarios where the original start points and target points are close together. The essential advantage of this method is to protect trip data against deletion in compliance with data protection legislation.

The core of the EXCL-Algorithm consists of four steps. They comprise, for example, the examination of different machine learning algorithms in order to generate suitable artificial GPS data points. The verification of the artificially generated GPS points is done by focusing on the distance between the GPS data points as well as by ensuring that new areas are entered to prevent that scenario one (only one area without any GPS data) or scenario two (two areas directly lie to each other but no GPS data) occur again. These verification steps, finally, enable to evaluate our EXCL-Algorithm based on two specific scenarios.

The content of this research refers to a generic view to all modes of transports. However, this means that the generation of new GPS data is not only based on the existing car or bicycle routes. The generated data can for example vary from roads or enter green areas as well. So, for an outsider, it is still impossible to distinguish between artificially generated data or real data because all modes of transportation can be combined for tracking the routes in the application.

Finally, the EXCL-Algorithm preserves data that can be used for further studies and protects the privacy of each user who contributes by tracking the daily trips. Further studies can focus to generate only those GPS data points lying on roads for cars or pavements for pedestrians depending on the provided data. This studies would precise the EXCL-Algorithm and its data generation and would focus on only one mode of transportation.

ACKNOWLEDGEMENTS

This paper is based on the research and development joint project NUMIC - New Urban Awareness of Mobility in Chemnitz. It is founded by the Federal Ministry of Education and Research (BMBF) and the European Social Fund under the grant number 01UR1804A. The responsibility for this publication lies with the authors.

REFERENCES

Bundesministerium für Verkehr und digitale Infrastruktur (2020). Mobilität in Deutschland (MiD). ttps: //www.bmvi.de/SaredDocs/DE/Artikel/G/ mobilitaet-in-deutschland.html. [Online, accessed 2-November-2020].

Chemnitz, Stadt der Moderne (2020). Mobilitätsverhalten - Untersuchung "Mobilität in Städten". URL: https://www.chemnitz.de/chemnitz/de/unsere-stadt/verkehr/verkehrsplanung/mobilittsverhalten/index.html. [Online, accessed 30-October-2020].

Chow, C.-Y. and Mokbel, M. F. (2011). Privacy of spatial trajectories. In *Computing with spatial trajectories*, pages 109–141. Springer.

Claudio, B., Sean Wang, X., and Jajodia, S. (2005). Protecting privacy against location-based personal identification. *Jonker, Petković (Hg.) 2005 – Secure Data Management*, page 185–199.

Feng, T. and Timmermans, H. J. (2017). Using recurrent spatio-temporal profiles in gps panel data for enhancing imputation of activity type. *Big Data for Regional Science*, pages 121–130.

Ghinita, G., Kalnis, P., and Skiadopoulos, S. (2007). Proceedings of the 16th international conference on world wide web. page 371.

Géron, A. (2018). *Machine Learning mit Scikit-Learn und Tensorflow*. O'Reilly, Heidelberg.

Hoh, B., Gruteser, M., Xiong, H., and Alrabady, A. (2007). Preserving privacy in gps traces via uncertainty-aware path cloaking. page 161.

Hua, Y., Zhao, Z., Li, R., Chen, X., Liu, Z., and Zhang, H. (2019). Deep learning with long short-term memory for time series prediction. *IEEE Communications Magazine*, 57(6):114–119.

Jeansoulin, R., Papini, O., Prade, H., and Schockaert, S. (2010). *Methods for handling imperfect spatial information*, volume 256. Springer.

Martens, K. (2007). Promoting bike-and-ride: The dutch experience. *Transportation Research Part A: Policy and Practice*, 41(4):326–338.

Merkel, D. A. (2018). Rede von bundeskanzlerin dr. angela merkel beim ix. petersberger klimadialog am 19. juni 2018 in berlin. *Bulletin, die Bundesregierung*, 68.

Niu, B., Li, Q., Zhu, X., Cao, G., and Li, H. (2014). Achieving k-anonymity in privacy-aware location-based services. *IEEE INFOCOM 2014 2014*, page 754–762.

Ortmeyer, T. H. and Pillay, P. (2001). Trends in transportation sector technology energy use and greenhouse gas emissions. pages 1837–1847.

Pan, X., den Elzen, M., Höhne, N., Teng, F., and Wang, L. (2017). Exploring fair and ambitious mitigation contributions under the paris agreement. pages 49–56.

Ritchie, H. (2020). Cars, planes, trains: where do CO2 emissions from transport come from? URL: ttps://ourworldindata.org/co2-emissions-from-transport.\newblock[Online, accessed28-October-2020].\bibitem[Sceider et al., 2020]ExtendTrack Scheider, S., Wang, J., Mol, M., Schmitz, O., and Karssenberg, D. (2020). Obfuscating spatial point tracks with simulated crowding. *International Journal of Geographical Information Science*, 34(7):1398–1427.

Seber, G. and Lee, A. (2012). *Linear Regression Analysis*. Wiley Series in Probability and Statistics. Wiley.

Sumo (2020). Simulation of urban mobility. URL: eclipse.org/sumo/. [Online, accessed 07-September-2020].

Sweeney, L. (2012). k-anonymity: A model for protecting privacy: International journal of uncertainty, fuzziness and knowledge-based systems, 10(05), 557-570.

Trujillo-Rasua, R. and Domingo-Ferrer, J. (2015). Privacy in spatio-temporal databases: A microaggregation-based approach. 567:197–214.

Wang, S. and Wang, X. S. (2010). In-device spatial cloaking for mobile user privacy assisted by the cloud. pages 381–386.

Wardman, M., T. M. . P. M. (2007). Factors influencing the propensity to cycle to work. *Transportation Research Part A: Policy and Practice*, 41(4):339 – 350.

Zhu, X., Chi, H., Niu, B., Zhang, W., Li, Z., and Li, H. (2013). Mobicache: When k-anonymity meets cache. pages 820–825. IEEE.