Trajectory User Linking in C-ITS Data Analysis

Juliet Chebet Moso[1,2], Stéphane Cormier[1], Hacène Fouchal[1]
[1]CReSTIC EA 3804, 51097
[1]Université de Reims Champagne-Ardenne
Reims, France
juliet-chebet.moso@etudiant.univ-reims.fr, {stephane.cormier, hacene.fouchal}@univ-reims.fr

Cyril de Runz
BDTLN, LIFAT
University of Tours
Tours, France
cyril.derunz@univ-tours.fr

John Wandeto[2]
[2]Computer Science
[2]Dedan Kimathi University of Technology
Nyeri, Kenya
john.wandeto@dkut.ac.ke

***Abstract*—Vehicles in an Intelligent Transport Network exchange a lot of messages. Every message sent is generated with an identifier of the transmitting vehicle. To respect the user privacy, an identifier is kept only over a specified time interval. The need that arises is, given that multiple identifiers are assigned to a vehicle, are we able to group the identifiers and detect those which belong to the same vehicle? We solved this Trajectory-User Linking problem by chaining anonymous trajectories to potential vehicles by considering similarity in movement patterns. Our method managed to link trajectory segments to their common vehicles which we validated through map matching of the trajectories using QGIS.**

***Index Terms*—Trajectory-User Linking, Moving objects, Similarity measure, Semantic trajectory**

I. Introduction

The development of wireless communication technology, geographical information systems, embedded positioning devices and ubiquitous devices has facilitated collection and storage of vast quantities of mobility data. The collection of mobility data is done by online or offline means through devices attached/carried by the moving objects, road side units among other techniques. These data typically contains information describing the movement of people, goods, vehicles, air crafts, animals, natural phenomena (hurricanes, tornadoes, and ocean currents), etc. Each trace of a moving entity is a multi-attribute, time-ordered sequence of locations.

According to Liu, Wang and Qu cited in [1], trajectory mining can be viewed as a process of analyzing mobility traces with the aim of discovering spatial, spatial-temporal and behavioral patterns through clustering, classification, anomaly detection, and interesting location detection. Trajectory data mining can also be categorized into the following phases [2]: (a) pre-processing (trajectory compression, stay-point detection, trajectory segmentation and map matching), (b) data management (indexing and storing the data for efficient retrieval) and (c) pattern mining (clustering, classifying, and detecting outliers). The key driving force in trajectory data analysis can be "economic (logistical optimization, customer behavior analysis, targeted advertising), scientific (animal behavior analysis, healthcare), administrative (urban planning, criminal investigation), or private" [3]. The present challenge is how to exploit these data to extract useful knowledge and information for improvement of mobility levels [4].

When considering moving objects especially on road networks, the paths taken are linked to the prevailing traffic environment and conditions. When analyzing these trajectories it is important to incorporate the environmental information so as to gain a better understanding on the movement patterns [5]. Daily trajectories reveal aspects of lifestyle and behavior of moving objects which when analyzed can be able to show a connection or relationship between them. Trajectory pattern analysis is invaluable in applications such as: recommender systems, public security systems, and path planning in emergency evacuations [6].

In order to gain useful knowledge from trajectories, the raw points need to be enriched with semantic features, which is essentially a challenging task. One technique for solving this problem is to use experts to annotate the semantic features on the raw trajectories or to let the users attach semantic labels to their trajectories. Another approach is to associate points of interest (POIs) with the location information such that the POIs become the semantic labels [2], [7]. Semantically enriching a trajectory with background information makes querying and analysis simpler and enhances pattern identification [8]. This in turn facilitates behavior analysis of moving entities. Semantic trajectories can be applied in context-aware computing, trip recommender systems and life experience sharing [2].

In extraction of semantic patterns, the purpose of visits to a location and the time when the particular pattern occurred are important aspects to consider. However, it is challenging to identify the reason for visiting a region due to the fact that the region can cover multiple POIs and most of the time the POIs are not captured as attributes of the trajectory. Also, moving objects generate a lot of redundant highly sampled data over a long period of time, as a result of the low cost of storage and advances in battery technology. The high sampling-rates over a long period of time is an effective method to increase the probability of capturing more patterns during pattern mining. However, making one-by-one comparisons of the un-simplified raw trajectories is practically impossible and computational intensive. To mitigate this challenge, compression and pruning techniques can be employed during trajectory data mining [6].

Moving objects may have different strategies when they need to report their locations to a central repository, such as time-based, distance-based, and prediction-based strategies. They may also suspend the communication with a central server for a while and resume later. The overall result is that the lengths and time stamps of the trajectories will be different and the trajectories may also be segmented with gaps (missing readings). Each trip in a trajectory dataset includes an identification (ID) of the device it was recorded from. Device IDs enables chaining of consecutive trips of the same vehicle to rebuild movement over a longer period of time, which provides a better insight into mobility patterns. However, device IDs may periodically change for privacy or some other reasons, which clearly limits the analysis. Thus, to be able to gain knowledge from trajectories a method for chaining anonymous trajectories and filling missing gaps is required.

We propose to solve the Trajectory-User Linking (TUL) problem by chaining anonymous trajectories to potential vehicles by considering similarity in movement patterns. This will be performed as a pre-processing step for the characterization and semantic analysis of moving objects though behavior analysis. We make the following contributions: (a) we present a detailed state of the art on trajectory linking, trajectory classification and identify the open research issues; (b) we investigate trajectory linking problem using a real dataset of messages generated in Cooperative Intelligent Transportation System (C-ITS); (c) we validate our results using map matching.

The rest of this paper is structured as follows: Section II presents the state of the art investigation on Trajectory-User Linking and trajectory classification. Section III presents the problem statement, methodology and description of the dataset. Section IV presents the experiments and results, and Section V presents the conclusion and future work.

II. Related Works

This section introduces works on Trajectory-User Linking and trajectory classification.

A. Trajectory-User Linking (TUL)

Trajectory-User Linking is a recent area of research in location based social network applications (LBSNs) [9]. It is motivated by the fact that LBSN applications generate a lot of data which are usually stripped of the user identifiers as a way of anonymizing the data and preserving privacy. On the other hand, linking these trajectories to the users who generated them can provide invaluable information for recommendation systems and identification of criminals through phone signals and check-ins among other applications. Solving TUL is a challenging task due to the large number of user classes and the sparsity of data. A Recurrent Neural Networks (RNN) based semi-supervised learning model, called TULER (TUL via Embedding and RNN) is proposed in [9] which learns the semantic mobility patterns of spatio-temporal data by correlating trajectories to the users who generated them. TULER is designed to identify the dependencies inherent in check-in data and infer hidden patterns of users.

Another semi-supervised learning framework, TULVAE (TUL via Variational AutoEncoder) is proposed in [10] which learns human mobility in a neural generative architecture with stochastic latent variables that span hidden states in RNN. It considers the fact that human trajectories especially in geo-tagged social media are sparse with high-dimensionality and may contain embedded hierarchical semantic structures. TULVAE handles the data sparsity problem by analysis large volumes of unlabeled data which is a source of useful knowledge and unique individual mobility patterns.

While considering the heterogeneity of mobility data due to the growing number of location based services and the need for a deep understanding of user behavior across multiple services, DPLink [11] is proposed. DPLink is an end-to-end deep learning based framework for performing user identity linkage task on heterogeneous mobility data collected from different services with different properties. It is made up of a feature extractor including a location encoder and a trajectory encoder to extract representative features from the trajectory and a comparator to compare and decide whether to link two trajectories as the same user. A multi-modal embedding network and a co-attention mechanism in DPLink handle the low-quality problem of mobility data.

B. Trajectory classification

Trajectory classification is a process of identifying the class of a moving object based on its movement path. The goal can be to identify a type of vessel, the transportation mode, type of animal or a specific user based on their movement patterns [12]. The key input to a trajectory classification task is a sequence of spatio-temporal points. The main classification process follows three stages [2]: (a) Trajectory segmentation, (b) Feature extraction from the segments, and (c) Building of the classification model (e.g. Dynamic Bayesian Network (DBN), Hidden Markov Model (HMM), and Conditional Random Field (CRF) which consider information from local points/segments and the sequential patterns between contiguous points/segments).

When classifying trajectories, clustering can be performed by assigning similar trajectories to groups (clusters) such that the inter-class similarity is low and the intra-class similarity is high. Clustering facilitates the extraction of collective movement characteristics of objects resulting in behavior prediction which is used for decision support in location recommendation, destination prediction, weather forecast, urban planning and market research [13]. The current focus of trajectory clustering research is finding appropriate features for trajectory representation, similarity measures and development of algorithms for spatial data clustering [14]. The main challenge is how to identify relevant features that distinguish the class of a single point, trajectory segment or the whole trajectory and how to select the most discriminate features to be used in building the classification model[15]. A common discriminant feature is the distance between two trajectories or sub-trajectory segments which is computed using a distance measure or metric based on the type of application.

Trajectory similarity encompasses the geometric patterns of moving objects as well as the semantic generalizations derived from the raw trajectories. Several works in literature have considered the geometric or sequential features of trajectories when analyzing user similarity. Similarity among trajectories is often measured in terms of the co-location frequency (feature-based representations), which is the number of times two moving objects appear spatially close to one another. Other approaches for measuring similarity include subsequence similarity metrics such as the length of the Longest common subsequence(LCSS) [16], Edit Distance on Real Sequences (EDR) [17], Common Visit Time Interval (CVTI) [18], Maximal Semantic Trajectory Pattern (MSTP) [19], Multidimensional Similarity Measure (MSM) [20], and Stops and Moves Similarity Measure (SMSM) [21].

LCSS reduces the impact of noisy data by defining distance and matching thresholds.Two points match when their distance is less than a given threshold in all dimensions. However, LCSS ignores possible gaps in sequences, which, for certain problems, results in the same similarity value for different pairs of trajectories. EDR uses an edit distance measure to compute similarity between elements where a match considers all dimensions. Penalties are assigned according to the length of the gaps between two matched sub-sequences resulting in more accurate results than LCSS. CVTI integrates the semantic dimension of stops with temporal dimension. It does not allow heterogeneous data such as stops and moves to be modeled and measured together.

MSTP measures the similarity between two semantic trajectory patterns by considering the frequency at which stops are visited. However, it does not handle multiple data dimensions and does not consider moves between stops. In MSM the similarity score is built upon the matching scores of all pairs of elements that have at least one matching dimension. Partial similarity is assigned according to the number of dimensions in which elements match. It allows definition of different weights for every dimension.It however, ignores the order of stops and does not consider moves. It may assign a high similarity score even if two trajectories are only similar for a small portion of their length. SMSM considers both stops and moves within the trajectory and performs partial dimension matching and partial ordering of stops through assignment of weights. However, estimation of weights may be challenging for users.

Trajectory data have diverse formats which are unique to application requirements; therefore, different mining techniques and similarity measures are applicable based on the scenario being modeled. When looking at the applicability of similarity measures based on trajectory dimensions, LCSS and EDR require all elements to match across all dimensions, while MSM considers matching pairs in a single dimension. In scenarios where the trajectory data contains outliers LCSS, EDR, MSM and SMSM can be applied since they are robust to noise. When dealing with semantic trajectories MSM and SMSM are good options though, LCSS and EDR can be extended for semantic trajectory mining. When considering applications that use GPS trajectories annotated with stops only or trajectories extracted from social media, the best measure is MSM since it handles sparse data. MSM is particularly useful when one wants to find users who visited the same place at similar times without considering the order of visits. When order of visits is important, SMSM is the most appropriate since it considers the order of the stops. SMSM is also applicable in situations where one wants to extract the most similar paths or most popular routes between stops.

III. Problem Statement and Methodology

A. Problem Statement

The vehicles of an Intelligent Transport Network (ITS) exchange a lot of messages. Every message sent is generated with an identifier of the transmitting vehicle. To respect the user privacy, the identifiers of each vehicle are changed regularly. An identifier is kept only over a time interval. The issue we want to study is, given the multiple identifiers assigned to a vehicle are we able to group the identifiers and detect those which belong to the same vehicle? We adopt the definition of [9] for Trajectory User-Linkability problem:

Let $T_{vi} = m_{i1}, m_{i2}, \dots, m_{in}$ denote a trajectory generated by the vehicle v_i during a time interval, where $m_{ij}(j \in [1, n])$ is a message sent from a specific location at time t_j. Given that the identifier is changed after a time period, trajectory $T_x = m_1, m_2, \dots, m_y$ generated by the same vehicle in the next time interval with a different identifier is considered unlinked. TUL can thus be defined as:

Suppose we have a number of unlinked trajectories $T = t_1, \dots, t_m$ generated by a set of vehicles $V = v_1, \dots, v_n (m \gg n)$, TUL learns a function that links unlinked trajectories to the vehicles: $T \rightarrow V$

B. Methodology

In a C-ITS environment cooperative awareness is achieved through exchange of CAMs which contain position information. This can act as a privacy threat to the drivers since an eavesdropper can be able to create a detailed mobility pattern of the driver. To mitigate this, Pseudonym schemes are used to provide anonymous communication. A pseudonym generally provides authentication for vehicles which can use multiple pseudonyms in order to guarantee unlinkability of actions[22]. This involves the change of pseudonyms after a preset time period so as to prevent linkability of one pseudonym to another which can in turn result in the identity of a vehicle and consequently that of the driver being revealed if one is able to identify the home address

A first work consists in grouping as much as possible the different identifiers which represent sub-trajectories of one vehicle. A complete grouping with all the identifiers of each vehicle may be difficult to obtain but grouping some identifiers can be obtained. For example, if the last message of an identifier is spatially and temporally close to the first message obtained with another identifier and the change in attributes like speed and heading angle is consistent, then the change of identifier from the last message to the first one is obtained for the same vehicle. Thus the two identifiers are linked and belong to the same vehicle. In this example, the work consists in defining a reliable link between two messages with different identifiers.

Then we detect the contradictions between messages. For instance, if two messages give the same localization at the same time, then their identifiers cannot belong to the same vehicle. These contradictions help to define the group of identifiers for each vehicle by rejecting the identifiers leading to a contradiction. The framework to be followed in the analysis is shown in Fig.1.

Definition: *Trajectory*: A raw trajectory consists on a sequence of n points $T = [p_1, p_2, \ldots, p_n]$, in which $p_i = x, y, z, t, A$, where *x, y ,z* represent the position of the moving object in space, t is the timestamp and A represents other attributes associated with the point (i.e. speed, heading angle and drive direction)

In this study a trajectory is considered as the consolidation of messages uniquely identified by a single identifier.

C. Dataset Description

In our study we used a real dataset of Cooperative Awareness Messages (CAM) collected in France between September 2018 and August 2019 under a C-ITS project [23]. The purpose of CAMs is to give dynamic information about the vehicle (i.e. speed, position, heading (direction of motion with regard to true north) etc.). A vehicle sends CAMs to its neighbourhood using Vehicle-to-Vehicle (V2V) or Vehicle-to-Infrastructure (V2I) communications. The frequency of CAM message generation varies from 10Hz to 1Hz (100 milliseconds to 1000 milliseconds). Each CAM is uniquely defined by a stationid (Pseudonym) and timestamp. In this dataset each vehicle was assigned unique stationids which were changed periodically for privacy reasons resulting in a total number of 3866 unique IDs and a total number of 10,174,437 CAM messages sent.

In this study, each message has an identifier (id) associated with the transmitting vehicle but this vehicle is unknown. The message also includes a time stamp (time), the localization of the vehicle with latitude (lat), longitude (long) and altitude (alt), the speed (speed), the heading angle (angle) of the vehicle and the drive direction(direction). Thus the message is a data defined with 8 variables: id, time, lat, long, alt, speed, angle, direction. The variables lat, long, alt are the three position variables. Speed, angle and direction variables are used as variables of the behavior of the transmitting vehicle.

IV. Experimental Evaluation and Results

TUL problem is currently an active research area in location based social networks where the aim is to identify the users who generate check-in trajectory data. In this study we look at this problem in relation to trajectories generated by vehicles on a constraint road network. Our aim is to generate continuous trajectories using anonymous data while ensuring that privacy is preserved. In order to link the trajectories we consider the following conditions for triggering CAM generation as specified in ETSI EN 302 637-2 standard [24]:

- If the absolute difference between the current heading value of the vehicle and the heading value included in the last transmitted CAM by the same vehicle exceeds 4°;
- If the distance between the current position of the vehicle and the position included in the last transmitted CAM by the same vehicle exceeds 4 metres;
- If the absolute difference between the current speed of the vehicle and the speed included in the last transmitted CAM by the same vehicle exceeds 0.5 m/s.

We performed trajectory mining using PostgreSQL database with the spatial extension PostGIS used for storing and processing spatial data. We also used Quantum GIS (QGIS) an open-source cross-platform desktop geographic information system application that supports viewing, editing, and analysis of geospatial data. QGIS was majorly used for visualization and map matching of the trajectories as a validation step. Data pre-processing was done by removing noise from the data. The distribution of all trajectories is shown in Fig.2. We then extracted origin-destination pairs from the trajectories whereby an origin is the first message of each trajectory and a destination is the last message of the trajectory. Fig.3 shows the distribution of the origin (green colour)-destination (red colour) pairs.

Considering the fact that each vehicle was assigned multiple identifiers, we sort out to link identifiers which occurred on the same day. Taking the destination points, we extracted the

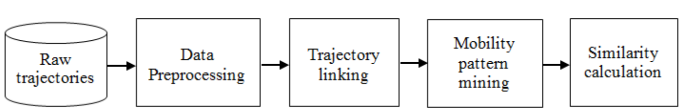

Fig. 1. Trajectory mining framework.

Fig. 2. Distribution of all trajectories.

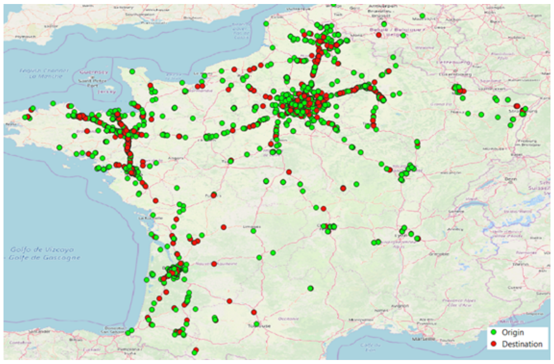

Fig. 3. Distribution of origin-destination pairs.

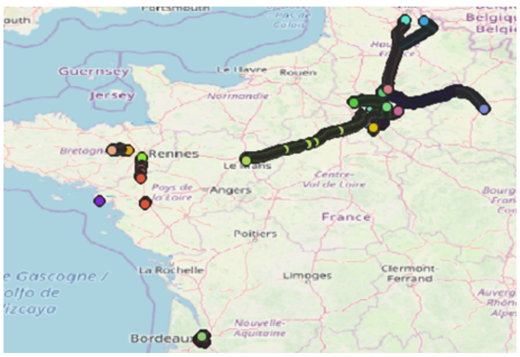

Fig. 4. Distribution of trajectories for the 5th and 6th of April 2019.

nearest origin point within 170 meters (since the highest speed recorded in the dataset was 163m/s) and also filtered out the results by implementing the CAM generation trigger conditions as additional constraints. The distance computation was done using the *ST_DistanceSpheroid* function in PostgreSQL which gives the linear distance between two longitude/latitude points. We also used the CAM generating frequency of 100 – 1000 milliseconds as a constraint in order to get exact matches in time and space.

During matching we were specifically targeting the matches which occurred on the same date within a few seconds difference so as to get trajectories which are continuous in space and time. Also, as a test for continuity, the matched trajectories had to be traveling in the same direction during the change of identifiers.After processing all the trajectories we were able to get 867 matching/linked ID pairs with the month of April having the highest matches at 124 IDs. We selected 45 trajectories generated on the 5th and 6th April 2019 (as shown in Fig.4) and after processing 10 trajectories linked with others to generate a total of 35 trajectories. The highest number of linkages per trajectory was four trajectories where ID 1 linked to 2 then 3 and finally 4 both in time and space as shown in Fig.5, thus generating one continuous trajectory as shown in Fig.6. To validate the linkage/matching of trajectories, we performed map-matching to ensure that the trajectories are on the same road and moving in the same direction.

Fig. 5. Continuity validation of linked trajectories.

Fig. 6. Continuous trajectory after linking four trajectories.

The fact that the trajectories are constrained by a road network increases the probability of linking trajectory segments to the generating user given background knowledge and behavioral aspects of movement like speed, heading and drive direction. However, complete linkage of all segments to the generating users is a difficult task and might not be possible. This is proven by the fact that out of 3866 trajectories, we were only able to link 867 pairs which is 22.43% of the total number of trajectories. The monthly analysis of linked trajectories is shown in Fig.7 which indicates the total number of IDs per month, the total number of IDs linked per month and the linkage percentage.

V. Conclusion

In this work we considered the trajectory-linking problem and applied it to messages generated by vehicles in C-ITS. Based on our analysis, it is possible to link trajectories to the generating users if other distinguishing attributes (like speed,

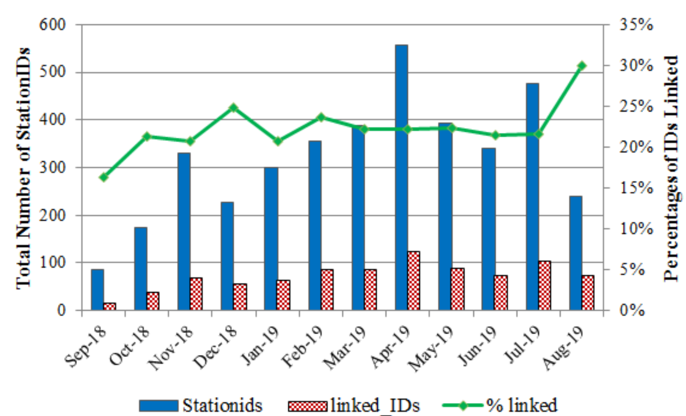

Fig. 7. Monthly analysis of linked trajectories.

heading angle, altitude and drive direction) and background knowledge on generation of the messages are considered when performing similarity analysis. It is also worth noting that the use of pseudonyms as a privacy and security measure has been proven to be a viable approach since we were not able to break the unlinkability requirement. As future work, we plan to semantically enrich the trajectories and perform frequent pattern mining and behavior analysis on the data.

References

[1] S. Shekhar, H. Xiong, and X. Zhou, Eds., *Encyclopedia of GIS*. Cham: Springer International Publishing, 2017.

[2] Y. Zheng, 'Trajectory Data Mining: An Overview', *ACM Trans. Intell. Syst. Technol.*, vol. 6, no. 3, pp. 1–41, May 2015, doi: 10.1145/2743025.

[3] F. Valdés and R. H. Güting, 'A framework for efficient multi-attribute movement data analysis', *The VLDB Journal*, vol. 28, no. 4, pp. 427–449, Aug. 2019, doi: 10.1007/s00778-018-0525-6.

[4] F. Alesiani, L. Moreira-Matias, and M. Faizrahnemoon, 'On Learning From Inaccurate and Incomplete Traffic Flow Data', *IEEE Trans. Intell. Transport. Syst.*, vol. 19, no. 11, pp. 3698–3708, Nov. 2018, doi: 10.1109/TITS.2018.2857622.

[5] T. Wu, J. Qin, and Y. Wan, 'TOST: A Topological Semantic Model for GPS Trajectories Inside Road Networks', *IJGI*, vol. 8, no. 9, p. 410, Sep. 2019, doi: 10.3390/ijgi8090410.

[6] Y. Cao et al., 'Effective spatio-temporal semantic trajectory generation for similar pattern group identification', *Int. J. Mach. Learn. & Cyber.*, Jul. 2019, doi: 10.1007/s13042-019-00973-y.

[7] Z. Yan, D. Chakraborty, C. Parent, S. Spaccapietra, and K. Aberer, 'Semantic trajectories: Mobility data computation and annotation', *ACM Trans. Intell. Syst. Technol.*, vol. 4, no. 3, p. 1, Jun. 2013, doi: 10.1145/2483669.2483682.

[8] A. Nishad and S. Abraham, 'SemTraClus: an algorithm for clustering and prioritizing semantic regions of spatio-temporal trajectories', *International Journal of Computers and Applications*, pp. 1–10, Sep. 2019, doi: 10.1080/1206212X.2019.1655853.

[9] Q. Gao, F. Zhou, K. Zhang, G. Trajcevski, X. Luo, and F. Zhang, 'Identifying Human Mobility via Trajectory Embeddings', in *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence*, Melbourne, Australia, Aug. 2017, pp. 1689–1695, doi: 10.24963/ijcai.2017/234.

[10] F. Zhou, Q. Gao, G. Trajcevski, K. Zhang, T. Zhong, and F. Zhang, 'Trajectory-User Linking via Variational AutoEncoder', in *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence*, Stockholm, Sweden, Jul. 2018, pp. 3212–3218, doi: 10.24963/ijcai.2018/446.

[11] J. Feng et al., 'DPLink: User Identity Linkage via Deep Neural Network From Heterogeneous Mobility Data', in *The World Wide Web Conference on - WWW '19*, San Francisco, CA, USA, 2019, pp. 459–469, doi: 10.1145/3308558.3313424.

[12] F. Vicenzi and L. M. Petry, 'Exploring Frequency-based Approaches for Efficient Trajectory Classification', In *Proceedings of the 35th Annual ACM Symposium on Applied Computing - SAC '20*, March 30-April 3, 2020, pp. 624-631, doi: 10.1145/3341105.3374045.

[13] Q. Yu, Y. Luo, C. Chen, and S. Chen, 'Trajectory similarity clustering based on multi-feature distance measurement', *Appl Intell*, vol. 49, no. 6, pp. 2315–2338, Jun. 2019, doi: 10.1007/s10489-018-1385-x.

[14] B. A. Sabarish, R. Karthi, and T. Gireeshkumar, 'Clustering of Trajectory Data Using Hierarchical Approaches', in *Computational Vision and Bio Inspired Computing*, vol. 28, D. J. Hemanth and S. Smys, Eds. Cham: Springer International Publishing, 2018, pp. 215–226.

[15] C. A. Ferrero, L. O. Alvares, W. Zalewski, and V. Bogorny, 'MOVELETS: exploring relevant subtrajectories for robust trajectory classification', In *Proceedings of the 33rd Annual ACM Symposium on Applied Computing -SAC'18*, Pau, France, 2018, pp. 849–856, doi: 10.1145/3167132.3167225.

[16] M. Vlachos, G. Kollios, and D. Gunopulos, 'Discovering similar multidimensional trajectories', in *Proceedings 18th International Conference on Data Engineering*, San Jose, CA, USA, 2002, pp. 673–684, doi: 10.1109/ICDE.2002.994784.

[17] L. Chen, M. T. Özsu, and V. Oria, 'Robust and fast similarity search for moving object trajectories', in *Proceedings of the 2005 ACM SIGMOD international conference on Management of data - SIGMOD '05*, Baltimore, Maryland, 2005, p. 491, doi: 10.1145/1066157.1066213.

[18] H.-Y. Kang, J.-S. Kim, and K.-J. Li, 'Similarity measures for trajectory of moving objects in cellular space', in *Proceedings of the 2009 ACM symposium on Applied Computing - SAC '09*, Honolulu, Hawaii, 2009, p. 1325, doi: 10.1145/1529282.1529580.

[19] J. J.-C. Ying, E. H.-C. Lu, W.-C. Lee, T.-C. Weng, and V. S. Tseng, 'Mining user similarity from semantic trajectories', in *Proceedings of the 2nd ACM SIGSPATIAL International Workshop on Location Based Social Networks - LBSN '10*, San Jose, California, 2010, p. 19, doi: 10.1145/1867699.1867703.

[20] A. S. Furtado, D. Kopanaki, L. O. Alvares, and V. Bogorny, 'Multidimensional Similarity Measuring for Semantic Trajectories: Multidimensional Similarity Measuring for Semantic Trajectories', *Trans. in GIS*, vol. 20, no. 2, pp. 280–298, Apr. 2016, doi: 10.1111/tgis.12156.

[21] A. L. Lehmann, L. O. Alvares, and V. Bogorny, 'SMSM: a similarity measure for trajectory stops and moves', *International Journal of Geographical Information Science*, vol. 33, no. 9, pp. 1847–1872, Sep. 2019, doi: 10.1080/13658816.2019.1605074.

[22] J. Petit, F. Schaub, M. Feiri, and F. Kargl, 'Pseudonym Schemes in Vehicular Networks: A Survey', *IEEE Commun. Surv. Tutorials*, vol. 17, no. 1, pp. 228–255, 2015, doi: 10.1109/COMST.2014.2345420.

[23] Scoop.developpement-durable.gouv.fr. 2020. *General Presentation - SCOOP Project : Connected Road And Vehicle.* [online] Available at: ¡http://www.scoop.developpement-durable.gouv.fr/en/general-presentation-a9.html¿ [Accessed 20 February 2020].

[24] ETSI E. 302 637-2 V1. 3.1-Intelligent Transport Systems (ITS); Vehicular Communications; Basic Set of Applications; Part 2: Specification of Cooperative Awareness Basic Service. ETSI, Sept. 2014.