MASTER

Multiple ASpect TrajEctoRy management and analysis

Project Acronym	***MASTER***
Project Full Name	***M**ultiple **AS**pects **T**raj**E**cto**R**y management and analysis*
Project Number	***777695***
Deliverable Title	***Final version of similarity measures and analysis methods for holistic trajectories***
Deliverable No.	***D4.2***
Contract Delivery Date	***30/04/2023 (M62)***
Actual Delivery Date	***28/04/2023(M62)***
Responsible Authors	*Karine Zeitouni (Unit Manager of UVSQ)*
	This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie-Slodowska-Curie grant No 777695

DOCUMENT INFORMATION

GRANT AGREEMENT N.	777695
PROJECT ACRONYM	MASTER
PROJECT FULL NAME	Multiple Aspects Trajectory Management and Analysis
STARTING DATE (DUR.)	01/03/2018 (70 months)
ENDING DATE	31/12/2023
PROJECT WEBSITE	http://www.master-project-h2020.eu
COORDINATOR	Chiara Renso (CNR)
WORKPACKAGE N. | TITLE | START – END MONTH	WP4 | **Holistic Trajectories analysis methods** | M8 – M70
WORKPACKAGE LEADER	HUA
DELIVERABLE N. | TITLE	D4.2| **Final version of similarity measures and analysis methods for holistic trajectories**
RESPONSIBLE AUTHOR	Karine Zeitouni (Unit Manager of UVSQ)
DATE OF DELIVERY (CONTRACTUAL)	30 April 2023 (M62)
DATE OF DELIVERY (SUBMITTED)	28 April 2023 (M62)
VERSION | STATUS	1.0
NATURE	REPORT
DISSEMINATION LEVEL	PUBLIC
AUTHORS (PARTNER)	CNR, HUA, UPRC, UVSQ, DAL, PUC, UFC, UFSC

CONTRIBUTORS	Emanuele Carlini (HUA), Antonios Makris (HUA), Andrea Michienzi (CNR), Vinicius Monteiro de Lira (CNR), Raffaele Perego (CNR), Christos Sardianos (HUA), Konstantinos Tserpes (HUA), Iraklis Varlamis (HUA), Michela Rial (CNR), Chiara Renso (CNR), Francesco Lettich (CNR), Mohammad Abboud (UVSQ)

ACRONYM LIST

MASTER	Multiple Aspects Trajectory Management and Analysis
ICT	Information and Communication Technologies
ISTI	Istituto di Scienza e Tecnologie dell'Informazione "A. Faedo"
CNR	Consiglio Nazionale delle Ricerche
UNIVE	Ca' Foscari University of Venice
UVSQ	University of Versailles Saint-Quentin
UFC	Federal University of Ceara'
UPRC	University of Pireaus Research Center
HUA	Harokopio University of Athens
PUC	Pontificial University of Rio de Janeiro
DAL	Dalhousie University
THIRA	Municipality of Thira
ER	Experienced Researcher
ESR	Early Stage Researcher
AIS	Automatic Identification System (AIS)

TABLE OF CONTENTS

1. INTRODUCTION

The objective of this deliverable (D4.2) is to "*report about joint research work on similarity, machine learning and data mining methods for holistic trajectories*" as stated in the MASTER Grant Agreement Annex 1 Part A pages 15-16. According to the GA, this deliverable reports on the scientific activities of Task 4.1 and Task 4.2 at M24, which were accomplished during the secondments linked to WP4. It is linked to milestone MS3.

Contributors to this deliverable: CNR contributed to T4.1 with similarity methods and T4.2 with ML methods. HUA on T4.2 for recommendation systems. UFSC to T4.1 on similarity methods. UFC on T4.2 on graph analysis, DAL and PUC contributed to T4.2 on Machine learning, UVSQ to Task 4.2 on representation learning on trajectories.

2. WP4 OBJECTIVES AND TASKS

According to GA Annex I Part A , WP4 "encloses all the research activities related to the analysis of large repositories of holistic trajectories, the privacy and big data issues arising in this context, and to the shared discussion among MASTER partners of the concepts and vision around the future research challenges. The activities of the WP are split among 5 tasks, the first two having the foundational aim of providing proper similarity measures and analysis methods based on machine learning and data mining techniques. Training is embedded in each secondment with seminars and courses, including Summer School and Dagstuhl Seminar."

These are objectives of WP4:

1. Conduct research activities related to the **analysis** of large repositories of holistic trajectories. This objective has been achieved regarding the first two years of activities as detailed later in Task 4.1 and Task 4.2 reported in the present deliverable.

2. Conduct research activities related to **privacy** and **big data** issues arising in the context of large repositories of holistic trajectories. This objective has been partially achieved with activities in Task 4.3 and Task 4.4 but are not reported in this deliverable. As stated in the grant agreement, these activities will be reported in the Deliverable D4.4 "Big Data and privacy issues in holistic trajectories analysis" at M70.

3. Facilitate the shared discussion among MASTER partners of the concepts and vision around the **future research challenges.** This objective is related to Task 4.5 whose activities will be reported in deliverable D4.3.

4. Contribute to the training activities embedded in each secondment with seminars and courses held at the hosting institution, as well as to the Summer School and the Dagstuhl Seminar organized within MASTER project.

WP4 consists in the following tasks:

Task 4.1 Computing Similarity of holistic trajectories (Leader: CNR) (M8-M60);

Task 4.2 Machine learning and data mining methods (Leader: UVSQ) (M8-M60);

Task 4.3 Privacy issues in holistic trajectories analysis (Leader: UPRC) (M8-M60);

Task 4.4 Big Data issues in holistic trajectories analysis (Leader: HUA) (M8-M60);

Task 4.5 Future challenges (Leader: CNR) (M36-M70).

Task 4.6 Training in holistic trajectories analysis methods (Leader: UNIVE) (M8-M70).

Table 1 below shows all the secondments linked to the whole WP4 at M60. The total number of secondments indicated in the table column "Second. N" is 24. The total PMs of WP4 up to M62 is 25,73.

This deliverable is based on activities carried out during the secondments associated to tasks T4.1 and T4.2 as stated in the GA and highlighted in bold in Table 1 for a total of 14,75PMs. The other listed secondments will be reported in deliverables D4.4 Privacy and Big Data issues for holistic trajectories due at M70 and D4.3 Future challenges in holistic trajectories due at M70 as specified in the Task column.

Research activity carried out for tasks T4.1 and T4.2 are reported in the present deliverable; this activity produced the publications listed in Section 4. This deliverable contributes to MS5.

Table 1: Secondments executed from M8 to M62 linked to WP4 "Holistic Trajectories analysis methods"

RD N.	Second. N	Fellow ID	Profile	Name	Last Name	Sending Institution	Hosting Institution	From	To	PM	Task
19	97	15	ESR	Antonios	Makris	HUA	UFSC	01/12/2018	31/12/2018	1	T4.4
20	**98,71**	**16**	**ER**	**Christos**	**Sardianos**	**HUA**	**UFSC**	**01/12/2018 06/11/2019**	**31/12/2018 05/12/2019**	**2**	**T4.2**
21	**9, 107**	**3**	**ER**	**Raffaele**	**Perego**	**CNR**	**UFSC**	**11/03/2019 26/10/2019**	**06/04/2019 30/11/2019**	**2.07**	**T4.1**
25	**89**	**19**	**ESR**	**Andrea**	**Michienzi**	**CNR**	**UFC**	**12/06/2019**	**13/07/2019**	**1.07**	**T4.2**
26	**5**	**7**	**ESR**	**Vinicius**	**Monteiro de Lira**	**CNR**	**DAL**	**29/05/2019 15/06/2019**	**08/06/2019 07/08/2019**	**2.17**	**T4.2**

27	**10**	**20**	**ER**	**Emanuele**	**Carlini**	**CNR**	**DAL**	**02/07/2019**	**06/08/2019**	**1.17**	**T4.2**
34	**70**	**9**	**ER**	**Irakils**	**Varlamis**	**HUA**	**UFSC**	**06/11/2019**	**05/12/2019**	**1**	**T4.1**
35	**72**	**16**	**ER**	**Christos**	**Sardianos**	**HUA**	**PUC**	**06/12/2019**	**05/01/2020**	**1**	**T4.2**
36	68	8	ER	Konstantinos	Tserpes	HUA	PUC	06/12/2019	05/01/2020	1	T4.4
37	99	9	ER	Iraklis	Varlamis	HUA	PUC	06/12/2019	05/01/2020	1	T4.4
38	**105**	**7**	**ESR**	**Vinicius Cesar**	**Monteiro de Lira**	**CNR**	**UFC**	**30/11/2019**	**02/01/2020**	**1.13**	**T4.2**
40	76	27	ESR	Tarlis	Torelli Portela	UFSC	CNR	02/01/2022	02/02/2022	1.03	T4.4
42	**14**	**30**	**ER**	**Francesco**	**Lettich**	**CNR**	**UFC**	**29/03/2022**	**30/04/2022**	**1.07**	**T4.2**
46	**13**	**3**	**ER**	**Raffaele**	**Perego**	**CNR**	**PUC**	**10/08/2022**	**09/09/2022**	**1**	**T4.2**
48	**57**	**31**	**ESR**	**Mohammad**	**Abboud**	**UVSQ**	**UFC**	**17/06/2022**	**18/07/2022**	**1.07**	**T4.2**
47	96	5	ER	Karine	Zeitouni	UVSQ	UFC	12/07/2022	12/08/2022	1.03	T4.4
49	73	15	ESR	Antonios	Makris	HUA	DAL	21/06/2022	20/07/2022	1	T4.4
4	18	4	ER	Chiara	Renso	CNR	UFSC	12/03/2023	04/04/2023	0.7	T4.5
22	113	22	TECH	Beatrice	Rapisarda	CNR	UFSC	12/03/2023	04/04/2023	0.7	T4.5
54	116	33	ESR	Chiara	Pugliese	CNR	UFSC	12/03/2023	11/04/2023	1	T4.5
56	111	35	ESR	Guido	Rocchietti	CNR	UFSC	23/03/2023	06/05/2023(*)	1,47 (1,26)	T4.5
57	91	34	ESR	Maddalena	Amendola	CNR	UFSC	23/03/2023	06/05/2023(*)	1,47 (1,26)	T4.5

(*) secondment to be completed at the time the deliverable is submitted. In parenthesis the PMs up to M62.

TASK 4.1 COMPUTING SIMILARITY OF HOLISTIC TRAJECTORIES (LEADER: CNR) (M8-M60).

According to the GA Annex 1 Part A this task "addresses the definition and implementation of metrics to estimate the similarity between pairs of holistic trajectories."

This task has been executed during secondments #9 and #107 to UFSC by the researcher Raffaele Perego (CNR), and #70 to UFSC by the researchers Iraklis Varlamis (HUA), for a total of 3,07 person months.

During the secondments #9 and #107 to UFSC, Raffaele Perego (CNR) focused on the study of the computational cost of the similarity measures proposed by the research group of Prof. Vania Bogorny (UFSC). The scalability of the methods is an open issue and they identified some alternative approaches to extract signatures from trajectories and compute effective similarities among these signatures in an efficient way. In continuation of this work, Iraklis Varlamis (HUA) who was seconded to UFSC (Secondment #70), contributed to the proposal of a novel similarity metric that leverages machine learning in computing complex similarity functions. It is based on the creation of binary clustering trees on each aspect and then combines all similarities using a "forest" that allows multi-aspect comparisons.

TASK 4.2 MACHINE LEARNING AND DATA MINING METHODS (LEADER: UVSQ) (M8-M60).

According to the GA Annex 1 Part A this task "focuses on the analysis methods, which in the context of MASTER, must cope with the multiple dimensions and complex aspects associated with holistic trajectories. The mining methods considered are: clustering, prediction with ML, graph analysis with TDG, recommenders."
This task has been addressed in secondments #98 and #71 to UFSC by researcher Christos Sardianos (HUA); in secondments #89 and #105 to UFC by researchers Andrea Michienzi (CNR) and Vinicius Monteiro de Lira (CNR); in secondments #5 and #10 to DAL by Vinicius Monteiro de Lira (CNR) and Emanuele Carlini (CNR); and in secondment #72 to PUC by Christos Sardianos (HUA). This effort represented 8,44 person months in total.

During HUA's secondment to UFSC (Secondment #71), the group discussed topics on semantic trajectory classification as a kind of ML (machine learning) method for prediction with ML. In particular existing results from UFSC were discussed on the machine learning task of semantic trajectory classification and the approaches and algorithms they proposed and implemented so far (e.g. Movelets). Those were compared to the trajectory classification techniques that HUA's team has applied in the maritime scenario (see K*apadais, K., Varlamis, I., Sardianos, C., Tserpes, K. (2019). A Framework for the Detection of Search and Rescue Patterns Using Shapelet Classification. Future Internet, 11(9), 192*). The research effort was also focused on the data mining methods that cluster single-aspect trajectories and designed methods that can extend them to multi-aspect trajectories with categorical, ordinal or numeric values, with implied or optional values. The current results can be summarised in a new clustering algorithm that combines a tree-based similarity measure and spectral clustering techniques which gave rise to a research paper accepted in The Big Mobility Data Analysis

workshop. During secondments #98 and #72 to UFSC and PUC, Christos Sardianos (HUA) focused on machine learning techniques for trajectory classification in the maritime domain.
During secondments #5 and #10 to DAL, Vinicius Monteiro de Lira (CNR) and Emanuele Carlini (CNR) adopted a temporal graph-based approach (with time dependent graph -TDG), that uses AIS data to build complex network representations of the vessel trajectories all over the world-wide ports. They consider the ports as nodes, and the vessel trips between the ports as edges. From these temporal graphs, they derive the time series of topological (such as betweenness centrality, degree centrality, etc) and semantic (average speed, average acceleration, etc) properties that allow to answer data-driven questions in different context of analysis: (1) Trends/Popularity (e.g, what are the trending ports along the year?); (2) Prediction (e.g, what are the predictions of importation and exportation for the commercial blocks?); (3) Community detection (e.g. Is it possible to infer communities of vessels arriving in a port?). A paper has been submitted to the Big Mobility Data Analysis (BMDA) workshop (see [2] in References).

During secondments #89 and #105 to UFC, by Andrea Michienzi (CNR) first, and continued later by Vinicius Monteiro de Lira (CNR), secondees considered as mining technique a graph analysis with the concepts time dependent graph (TDG), where the graph is obtained by the enriched trajectories considering the moving objects as nodes and the meeting points as edges which can be weighted with the strength of the relationship. One important problem in analysing this kind of data is to identify the nodes which have a maximum influence in the graph. This is of primary importance in many trajectory-based systems. For instance, if the trajectories model the movement of vehicles in a city, this can be used to understand which are the fastest routes to cross the city, if the trajectories model the movement of tourists, the detection of influential nodes, together with clustering techniques, may help in suggesting popular touristic routes patterns. The novelty of the new studied techniques with respect to the state of the art is that the graph obtained from holistic trajectories contains both temporal and spatial information. The temporal information helps us understand when the two moving objects meet and how much time they spend together, and the spatial information models the place where the two objects meet. This work has been conducted in collaboration with Professors Jose Antonio Macedo (UFC) and Ticiana Linhares (UFC), and the master student Hinessa Caminha (UFC).
As an update to the deliverable D4.1, this task has also been addressed in secondments #14, #13, #57.
During secondment #14 to UFC, Francesco Lettich (CNR), together with the hosting unit research group, zeroed in on potential lines of future research concerning external sensor trajectories (ESTs), i.e. sparse trajectories collected through the observations of a sensor network deployed within some urban area. In particular, with the ESTs the hosting unit has in the past dedicated a substantial amount of research efforts to tackle the next location prediction problem, which has been then addressed with machine learning and deep learning techniques. We report that these techniques have successively found use in real-world settings concerning urban policing and urban security in the city of Fortaleza (Brazil). One particular issue that the hosting unit has raised during the secondment is the need to enrich ESTs with further features, as this has the potential to substantially improve the

accuracy of predictive techniques. Said features can also come under the form of semantic dimensions, or aspects, such as weather, traffic, social media posts, events, points of interest of relevance, and more. In this context the CNR unit can offer its expertise in the semantic enrichment domain, where the integration of existing techniques or the introduction of novel ones can help attain the aforementioned goal.
Another problem that has been raised by the hosting unit is the need to represent external sensor trajectories, possibly enriched with many different features and aspects, with embeddings that are suitable for indexing, clustering, similarity search, or for use within some neural network architecture of choice. The hosting unit highlighted how the literature concerning this problem is missing and how this has the potential to open up further research collaborations with the CNR partner.
During secondment #13 to PUC, Raffaele Perego (CNR), conducted research focused on machine learning techniques able to cope with the scale and the multiple dimensions associated with holistic trajectories. Specifically, the researcher investigated scalable techniques for clustering and measuring similarities of items represented in latent spaces by highly dimensional vectors.

During UVSQ's secondment #57 UFC the secondee Mohammad Abboud investigated on different topics related to advanced analysis and mining of multi-aspect trajectories. First, he discussed with the hosting group about their ongoing work on representation learning of trajectories defined by external sensors (ESTs). This approach is inspired by word embedding and transformers, popular in today's language models in Natural Language Processing (NLP) domain. The objective is to evaluate their capacity to learn a vector space of trajectories (trajectory embedding) and to capture their similarities. Another work presented by the secondee to UFC group was about activity recognition from trajectories defined by sensors or wearables. It was applied to micro-environment recognition in the context of environmental Mobile Crowd Sensing (MCS). In this use case, air quality measurements such as particulate matter (PM), NO2, Black Carbon, etc. were collected from the crowd using mobile sensor boxes and GPS trackers. A multi-view approach was proposed to learn from the time series of each type of pollutant and from the speed (derived from the GPS data): a 1st level learner is trained separately for each measure, then a meta-learner is trained on the predictions of the first level-learners and provides the final prediction of the micro-environment. UFC is also starting a project related to sensor-based pollution analysis in the city of Fortaleza. We discussed the perspectives of collaboration.
The third subject of the collaboration revolved around our ongoing work on fine-grained tweet geolocation. In this work, we are interested in enriching the qualitative text by adding the location information of the tweets at fine granularity. Our approach is based on two views: a spatial view handling the location keywords in the tweets, and a textual view handling the correlation between geotagged and non-geotagged tweets. Then a multi-view model is implemented on top of the two mentioned views.

3. CONCLUSIONS

In summary, main contributions related to this deliverable are as follows:

- Investigation of alternative approaches to extract signatures from trajectories for similarity measures (Secondments #9 and #107)
- Machine learning for complex similarity functions (secondment #70)
- Trajectory classification techniques using machine learning (Secondments #71)
- Graph based approaches with TDG for AIS data and traffic data (secondments #5, #10, #89, #105)
- Embeddings technique for sparse trajectories datasets (Secondment #14)
- High dimensional vectors for multiple aspects trajectories (Secondments #13,#57)

These scientific results are reported also in the 5 publications listed in Section 4.
The research activities carried out during secondments also included training activities of ESRs including participation to seminars and courses, as specified in the GA. These activities have been executed during the secondments of WP4. This deliverable corresponds to MS5.

4. PUBLICATIONS

[1] Petry, L. M., Ferrero, C. A., Alvares, L. O., Renso, C., & Bogorny, V. (2019). Towards semantic-aware multiple-aspect trajectory similarity measuring. Transactions in GIS, 23(5), 960-975. This is publication N. 25 SyGMA as article in journal. Link to repository is https://zenodo.org/record/3685667#.XlPHRi2ZNUM

[2] Emanuele Carlini, Vinicius Monteiro de Lira, Amilcar Soares, Mohammad Etemad, Bruno Brandoli Machado and Stan Matwin, Uncovering vessel movement patterns from AIS data with graph evolution analysis, BMDA workshop at EDBT 2020 to appear. This is publication N. 28 in SyGMA as workshop proceedings. Link to repository https://zenodo.org/record/3685892#.XlPZvC2ZNUM

[3] Kapadais, K., Varlamis, I., Sardianos, C., Tserpes, K. (2019). A Framework for the Detection of Search and Rescue Patterns Using Shapelet Classification. Future Internet, 11(9), 192. This is publication N.27 in SyGMA as article in journal. Link to repository https://zenodo.org/record/3685813#.XlPJ2S2ZNUM

[4] Varlamis, I., Tserpes, K., Etemad, M., Júnior, A. S., & Matwin, S. (2019). A network abstraction of multi-vessel trajectory data for detecting anomalies. In Proceedings of the 2nd International Workshop on "Big Mobility Data Analytics" at EDBT 2019, Lisbon, Portugal. This is publication N. 15 in SyGMA as workshop proceedings. Link to repository https://zenodo.org/record/2649606#.XlPIEi2ZNUM.

[5] Ioannis Kontopoulos, Iraklis Varlamis and Konstantinos Tserpes, Uncovering hidden concepts from AIS data: A network abstraction of maritime traffic for anomaly detection. First International workshop on Multiple Aspects Management and Analysis MASTER 2019 @ ECML-PKDD, Wurzburg, Germany September 16 2019. This is publication N. 23 in SyGMA as chapter in a book (workshop proceedings). Link to repository https://zenodo.org/record/3678139#.XlPK-C2ZNUM