Issue 3 | September 2019

MASTER

Multiple ASpects TrajEctoRy management and analysis

Inside this issue

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant No 777695 Duration: March 1 2018 | February 28 2022

Content

Editorial

Welcome to the third issue of the MASTER newsletter!
We are very proud that **Prof. Yannnis Theodoridis from University of Piraeus Research Center** kindly contributed with an invited article with title: "*Learning from our movements: The mobility data analytics era*" discussing his vision on big data and mobility analytics.
In this issue we also meet the **SoBigData project**, the European Research Infrastructure for Big Data and Social Mining. Similarly to MASTER, SoBigData is coordinated by CNR Pisa and both projects have several synergic points in common: from the mobility topics, to the quality of the resources shared: people/researchers, data, methods and infrastructures.
We have selected two very recent papers developed thanks to the partners collaboration in MASTER. The first article entitled "*Towards Semantic-Aware Multiple-Aspect Trajectory Similarity Measuring*" reports on the result of the collaboration between the National Research Council of Italy and the Federal University of Santa Catarina. The contribution is a similarity measure taking into account the multiple semantic aspects of trajectories.
The second article entitled "*A network abstraction model for vessel trajectory analysis and anomaly detection*" presents the result of the collaboration between Harokopio University and Dalhousie University proposing a network based model for better understanding of the behavior of vessels in the sea, specifically focussing on anomaly detection.
As past events, we report two workshops supported by MASTER partners: The Second Workshop on Big Mobility Data Analysis, held in Lisbon in conjunction to "Extended Database Technology" conference on March 26th (http://bmda19. datastories.org) and the First Workshop on Fairness, Accountability, Transparency, Ethics, and Society on the Web (http://fates19. isti.cnr.it), in conjunction with "The Web Conference" in San Francisco, last May 14th, 2019.
As forthcoming events, we highlight the first project workshop called MASTER to be held in conjunction with ECML-PKDD conference in Wurzburg, Germany on September 16th, 2019.

You can download this and previous issues of the newsletter from the MASTER web site: http://www.master-project-h2020.eu
The next issue will be published at the beginning of March 2020, reporting more secondments experiences, research results and events.

Stay tuned and happy reading!

Chiara Renso, Project Coordinator

Learning from our movements

The mobility data analytics era

Yannis Theodoridis, University of Piraeus, PIraeus, Greece

Once upon a time, it was the ChoroChronos EU research project[1]. The challenge at that time ('90s) was to bring spatial and temporal database aspects together in a, then emerging, integrated spatio-termporal domain. As time was passing, new challenges appeared and addressed by the researchers of the field: efficient system architectures, knowledge discovery from mobility data, privacy aspects, etc. Nowadays, in the era of Data Science and Big Data, mobility data analytics aims at learning from objects' movements, covering a range of methods and solutions, from de-noising of location information and integrating with multiple heterogeneous related sources to predictive analytics, either offline and online.

Let's catch the thread from the beginning … Back in '90s, research in spatial and temporal databases, separately, resulted in pretty mature results to contribute in real-world DBMSs. In the spatial database field, Oracle introduced Spatial Data Option in 1996, OGC released its first specifications in 1997, and PostGIS was launched in 2001. On the other hand, TSQL2 language specification was developed in 1993. The 'marriage' of the two fields was led by research projects in both US and Europe; focusing on the latter, the notable ChoroChronos EU project (1996-2000) aimed at bringing together the two communities and integrate their ideas in the so-called **spatio-temporal databases**, where time would be considered a first-class citizen [Frank et al. 1999; Güting et al. 2000; Koubarakis et al. 2003].

What followed was the focus of research in point objects due to the popularity of related applications (tracking of moving objects via GPS technology), which led to the **"moving object trajectory"** concept[2] and, as expected, raised challenges on knowledge discovery from this new type of data as well as on personal data privacy. This brought the 'dialogue' with other than database management domains, including machine learning / data mining as well as data privacy and security. For instance, the GeoPKDD (2005-2009), MODAP (2009-2012) and MOVE (2009-2013) EU projects aimed at **devising knowledge discovery and (privacy-preserving) analysis methods for trajectories of moving objects**, bringing together ICT researchers and domain specialists [Giannoti & Pedreschi, 2008; Renso et al. 2013]. The advances in social networks and linked open data in '00s also resulted in the so-called location-based social networking and relevant applications. Thus, a new trajectory variation was born, the **semantic trajectories** [Parent et al. 2013], studied by e.g. the SEEK EU project (2012-2015), the objective of which was to envisage a new semantic enriched knowledge discovery process where the semantic aspect (in the sense of the meaning of the movement) would be embedded at each step.

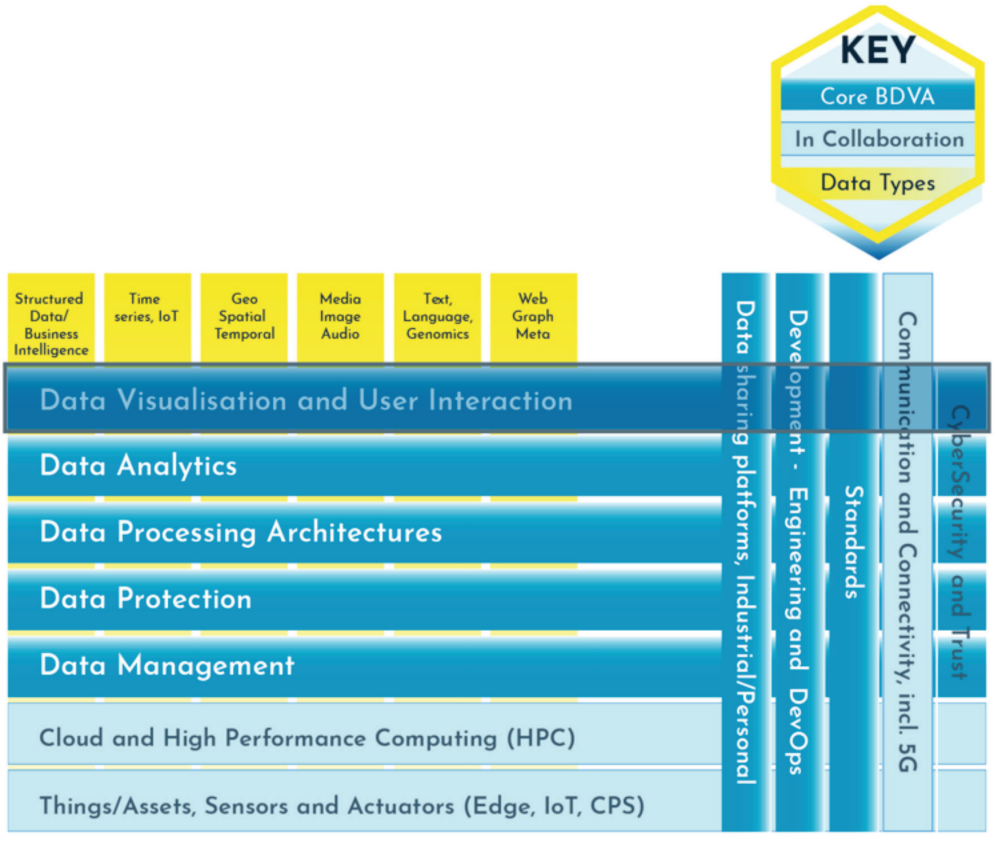


Fig. 1: The BDVA reference model (source: bdva.eu)

Nowadays, one can find plenty of sensor data and open sources of related information, mature Big Data technologies, plethora of Data Science methods and tools. In this environment, a hot research topic is that of **Mobility Data Analytics** (MDA). The range of processes covered under this term includes data acquisition and processing (typically from multiple and varying data sources), data management (storage and indexing, as usually…), data mining, data privacy, data visualization and user interaction [Pelekis and Theodoridis, 2014].

A typical MDA application tracks in real-time a population of humans, vehicles, vessels, aircrafts and handles the resulted trajectories, enriched with heterogeneous context, in order, for instance, to be able to assess traffic situation or drivers' behavior, forecast anticipated movements in short-term or schedule traffic in long-term, react as soon as an 'anomaly' is detected, and so on [Vouros et al. 2018]. In this framework, challenges touch almost each architectural layer of an MDA system, compatible with the EU BDVA reference model (illustrated in Fig. 1):

-**data sources** of interest include, on the one hand, streaming information, such as the GPS signals transmitted by the objects themselves, the objects' locations tracked by an external device (e.g. radar), and live weather information, and, on the other hand, archive collections of census, meteorological, etc. data; all of this information to be correctly and smoothly integrated;

-**data processing** requires cleansing (de-noising, smoothing) and semantic enrichment of incoming data as already mentioned, segmentation into trajectories that make sense (e.g. from/to predefined places of interest), and storage in appropriate (relational or NoSQL) stores;

-**data management** includes efficient querying and retrieval, which assume implementation of query algorithms and maintenance of indexing mechanisms suitable for these purposes;

-**data analytics** requires both offline and online algorithms: cluster and frequent pattern analysis for detecting typical movement patterns as well as interesting outliers can run offline on the historical data whereas movement prediction and anomaly detection over the incoming stream should run online; in order for the online methods to be effective, they should take into consideration the results of the offline methods (a now tracked object may be considered to have anomalous behavior if it is quite dissimilar to past typical patterns or quite similar to past outliers) and this makes things even more challenging;

-**data visualization and user interaction** is an essential component in order for the mobility analyst to get familiar with the data he/she is requested to analyze and interact with the above methods and tools; visual analytics (VA) is challenging, especially in mobility data [Andrienko et al. 2013]

This is, more or less, the context of a number of recent EU projects, including datAcron (http://datacron-project.eu; 2016-2018), where the use cases are on maritime and aviation, Track&Know (https://trackandknowproject.eu; 2018-2020) on drivers in urban environment, and MASTER (http://www.master-project-h2020.eu; 2018-2022) on land transportation, sea monitoring, and tourism.

Fortunately, research is an everlasting story. So, what's next? In the near future, we expect to see advances, for instance, in self-organizing and self-cleansing information integration tools (can we claim that we know what information is relevant to enrich a location or assess its accuracy better than a crawler searching the Semantic Web?), close to storage-less architectures following the IoT paradigm (if data resides in its original source, what is the need to replicate it in our storage? what if data is unaffordable huge and only meta-data as well as patterns extracted from data are stored locally?), and Explainable AI (why this and not that technique in order to analyze location data? our right to explanation). We expect these and other even more exciting research outcomes to appear in the years to come.

NOTE

[1] Disclaimer: the flashback in past is biased since it only refers by name to projects where the author has participated. The author's intention is by no means to provide an exhaustive survey of research activities related to the topic of the article.

[2] According to DBLP (dblp.uni-trier.de), the first papers with this term in their title appeared in 2000.

REFERENCES

Andrienko GL, et al. (2013): Visual Analytics of Movement. Springer.

Frank AU, et al. (1999): Chorochronos: A Research Network for Spatiotemporal Database Systems. SIGMOD Record 28(3): 12-21.

Giannotti F & Pedreschi D (2008): Mobility, Data Mining and Privacy - Geographic Knowledge Discovery. Springer.

Güting RH, et al. (2000): A foundation for representing and querying moving objects. ACM Trans. Database Syst. 25(1): 1-42.

Koubarakis M, et al. (2003): Spatio-Temporal Databases: The CHOROCHRONOS Approach. LNCS 2520, Springer.

Parent C, et al. (2013): Semantic trajectories modeling and analysis. ACM Comput. Surv. 45(4): 42:1-42:32.

Pelekis N & Theodoridis Y (2014): Mobility Data Management and Exploration. Springer.

Renso C, et al. (2013): Mobility Data: Modeling, Management, and Understanding. Cambridge University Press.

Vouros GA, et al. (2018): Big Data Analytics for Time Critical Mobility Forecasting: Recent Progress and Research Challenges. EDBT 2018: 612-623.

Meeting the SoBigData project

Cristina Muntean, ISTI-CNR, Pisa, Italy

SoBigData is the European Research Infrastructure for Big Data and Social Mining. Similarly to MASTER, SoBigData is coordinated by CNR Pisa and both projects have several points in common. The synergy between these two projects highlight both the relevance of the projects and the quality of the resources shared: people/researchers, data, methods and infrastructure.

SoBigData has integrated resources and has created an infrastructure where sharing data and methods among text miners, visual analytics researchers, socio-economic scientists, network scientists, political scientists, humanities researchers can indeed occur. Furthermore the research infrastructure serves a large community of social sensing and social mining researchers and it has reduced the gap between existing research centres present at European level.
SoBigData is enabling easy comparison, re-use and integration of state-of-the-art big social data, methods, and services, into new research. As an open research infrastructure, SoBigData promotes repeatable and open science. Although SoBigData is primarily aimed at serving the needs of researchers, the openly available datasets and open source methods and services provided by the new research infrastructure will also impact industrial and other stakeholders (e.g. government bodies, non-profit organisations, funders, policy makers).

SoBigData is not only strengthening the existing clusters of excellence in social data mining research, but also creating a pan-European, inter-disciplinary community of social data scientists, fostered by extensive training, networking, and innovation activities. In fact, SoBigData promotes networking, dissemination and innovation actions by means of workshops, summer schools,datathons, training courses and knowledge transfer.

SoBigData covers different thematic research environments, called exploratories. In SoBigData, researchers do analyses and use innovative methodologies in the exploratories:

- **City of Citizens**, covering smart cities, human mobility behavior analysis
- **Well-being & Economy**, covering poverty indicators, spatial analysis
- **Societal Debates**, covering text and social network analysis
- **Migration Studies**, covering macroscopic human flows, social behavior analysis
- **Sports Data Science**, covering training indicators, performance predictors
- **Explainable Machine Learning**, covering artificial intelligence, machine learning, explanation, ethics

From the brief description of the SoBigData projects it's immediately evident that there are several common goals and interests between MASTER and SoBigData.

COMMON RESEARCH THEME: MOBILITY

Both MASTER and SoBigData cover aspects related to mobility. If for MASTER we are more interested in holistic trajectories at a more general level, in SoBigData this interest is focused mainly on urban mobility and services for the citizens.

The City of Citizens exploratory tells stories about cities and people living in it. Data scientists describe those territories by means of data, statistics and models. This allows citizens and local administrator to better understand cities and how to improve them. How do people move into the city? How does the traffic change during the day? And how does it vary during the week? How does the tourism presence affect the traffic? As a case-study, the traffic in the Italian cities of Pisa and Florence was analyzed by exploiting Big Data sources such as mobile phone traces, vehicular GPS and social media data as a proxy of human behaviour. The results can be useful for both local administrators and citizens. The local administrators could have a tool to quantify accurately city's traffic and understand the city's usage, so they could take better decisions to manage mobility. Citizens could take information to know the traffic situation in real time and they could choose the best and fastest way. This could be useful in carpooling, too. Indeed, Big Data analysis can suggest to citizens who can share the travel with them.

Both Master and SoBigData work with trajectories, although of different type and granularity. A fruitful interaction between two projects could come from SoBigData datasets and infrastructure upon which to deploy the methods and techniques designed by the MASTER project.

THE HUMAN RESOURCE

Both projects invest in training and enhancing the research experience of students, PhDs and researchers through research visits. SoBigData offers funding for Transnational Access, while MASTER funds secondments, both serving the same purpose. Researchers have the opportunity to interact with the local experts, discuss research questions, run experiments on non-public datasets and methods, present results at workshops/seminars.

The SoBigData project invites researchers and professionals to apply to participate in Short-Term Scientific Missions (STSMs) to carry forward their own big data projects. These opportunities are offered as part of SoBigData's Transnational Access (TNA) activities and calls for applications. SoBigData welcomes

applications from individuals with a scientific interest, professionals, startups and innovators that may benefit from training in data science and social media analytics. In order to apply you have to fill the Project Application Form. Funding for a short-term scientific mission (2 weeks to 2 months) is available up to 4500 euros per participant (to cover the cost of daily subsistence, accommodation, and European flights). STSM bursaries are awarded on a competitive basis, according to the procedure described in the application pack and eligibility criteria, and based upon the quality of the applicant, the scientific merit of the proposed project, and their personal statement.

PRIVACY AND ETHICS

Privacy and ethics are important issues when it comes to data and data analysis. Trajectories and location data of users need to preserve privacy and to be used in ethical ways. SoBigData has created a MOOC on these topics and MASTER researchers could benefit from taking it.

Both MASTER and SoBigData promote an ethical approach to data science. Digital records may contain potentially sensitive information on personal activities or content protected by intellectual property rights, so it is necessary to have special policies for the collection and analysis of large-scale datasets. Projects need to collect and analyse data in an ethical manner by-design, implement privacy enhancing tools and work towards ensuring its analyses are also correct, fair and non-discriminatory.

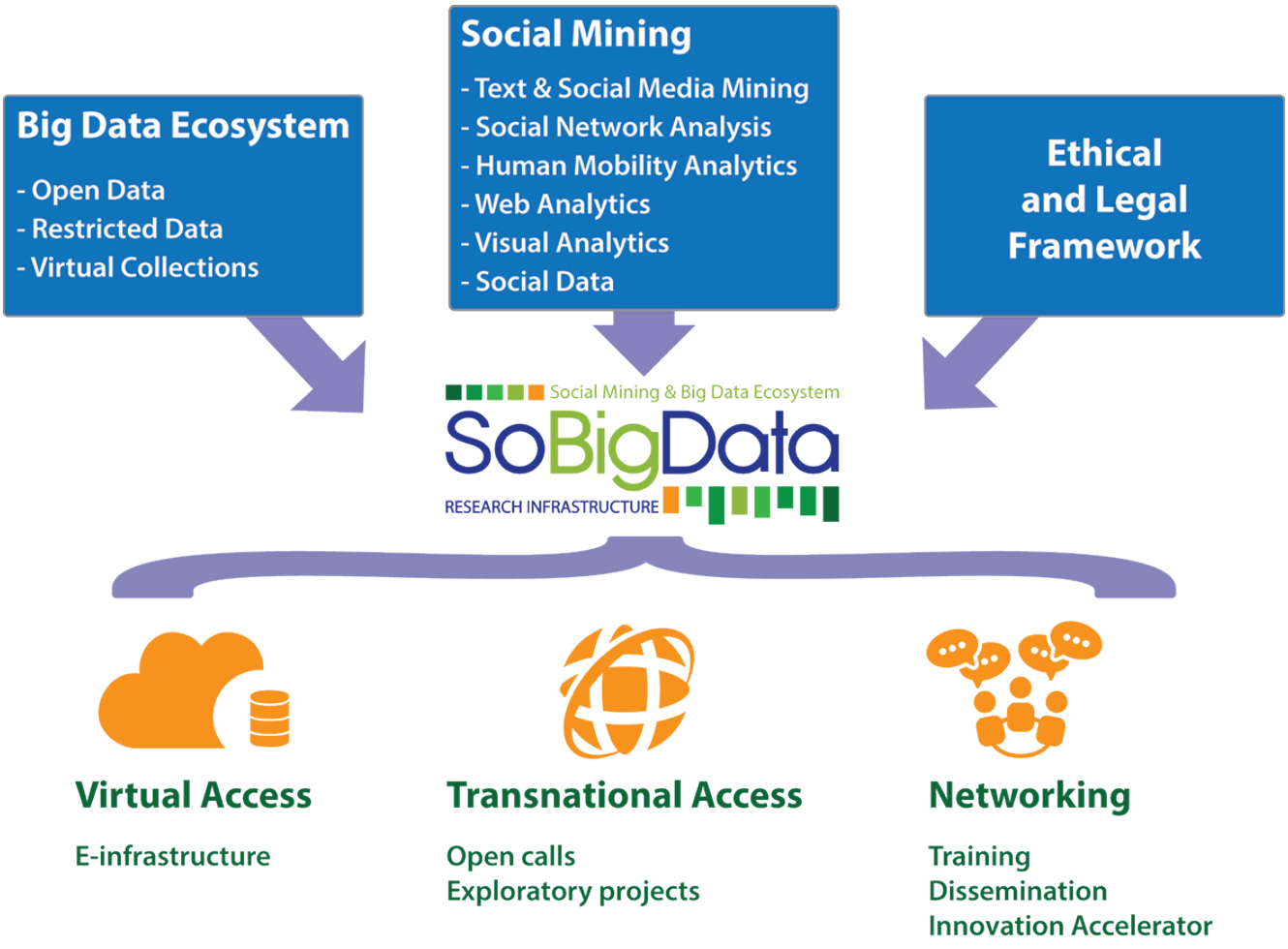
SoBigData has developed an online course in order to share with the community these basic elements about ethics, data protection, and intellectual property law. Visit the website of the FAIR (First Aid for Responsible Data Scientists by SoBigData) and enjoy their MOOC aimed at:

- Researchers, who want to become more aware of ethical issues
- Companies, which want to contribute to our community
- Students, who are curious about ethics, privacy and law

http://www.sobigdata.eu/ethics

SoBigData is now moving toward **SoBigData++** as advanced community. **The collaboration between SoBigData++ and MASTER** will continue with new transnational access experiences available for MASTER researchers and other opportunities.

For more info:
http://www.sobigdata.eu



Secondments

Experiences and results

Since the publication of previous newsletter issue MASTER partners performed 9 more secondments.

PONTIFICIAL UNIVERSITY OF RIO DE JANEIRO (PUC), BRAZIL

Three secondments has been executed at the Pontificial Catholic University of Rio de Janiero hosted by Prof. Casanova. Two researchers from CNR, Cristina Ioana Muntean and Vinicius Monteiro de Lira has been seconded in the period of March 2019 while researcher from UNIVE Andrea Marin has been seconded during the month of July 2019. Cristina collaborated with Prof. Casanova in developing ranking methods for RDF documents, while Vinicus discussed the transportation scenario and the available datasets. Andrea collaborated with the PUC group in comparing the transportation datasets available in Rio with the ones available in Venice to define a simulation model for the transportation scenario.

FEDERAL UNIVERSITY OF SANTA CATARINA (UFSC), BRAZIL

Two researchers from CNR visited the Federal University of Santa Catarina in Florianopolis in Brazil hosted by Prof. Vania Bogorny and her research group.

The main scientific outcomes of these secondments has been to start a collaboration on multiple aspects trajectory clustering, privacy for multiple aspects trajectories and similarity of multiple aspects trajectorie (see article at page 11).

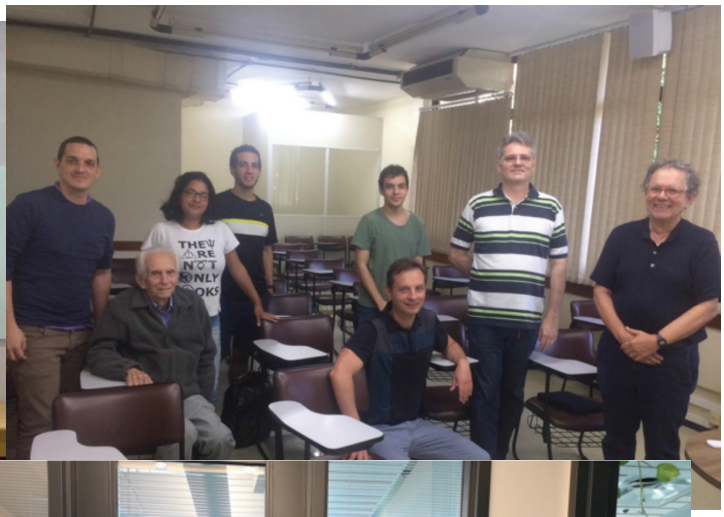
FEDERAL UNIVERSITY OF CEARA' (UFC), BRAZIL

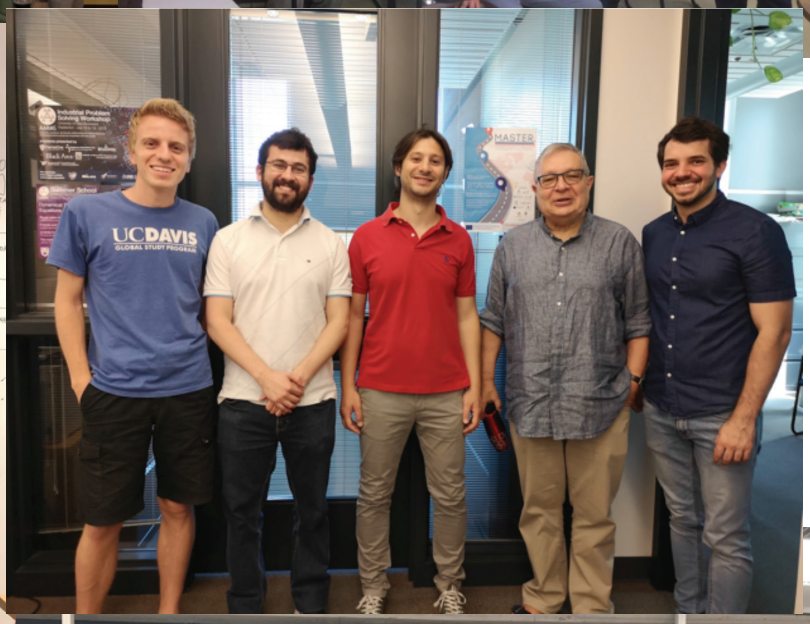
Two researchers from CNR, Andrea Michienzi and Chiara Renso started their secondments to Federal University of Ceara' in Fortaleza in Brazil. The work done is in the direction of developing analysis methods that cope with the multiple dimensions and complex aspects associated with holistic trajectories. Here, we considered as mining technique a graph analysis with 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 these kinds of data is to identify the nodes which have a maximum influence in the graph.

DALHOUSIE UNIVERSITY (DAL), CANADA

Two researchers from CNR, travelled to Halifax (Canada) for their secondments to Dalhousie University in the period June and July 2019: Vinicius Monteiro

de Lira and Emanuele Carlini.
Vinicius and Emanuele collaborated with Prof. Stan Matwin and his research group in developing a temporal graph-based approach that uses AIS data to build complex network representations of the vessel trajectories all over the world-wide ports. The idea is to answer data-driven questions in different context of analysis like Trends/Popularity (e.g, what are the trending ports along the year?), Prediction (e.g, what are the predictions of importation and exportation for the commercial blocks?) and Community detection (e.g. Is it possible to infer communities of vessels arriving to a port?).





Interviews with secondees

SECONDMENT EXPERIENCE IN THREE WORDS

Idea, collaboration, persons

Life-changing, challenging, meetings

Interesting, challenging, stimulating

Full of activities, amazing, fruitful

Networking, Collaboration and Exchange

Collaboration, creative research, tropical

Interesting, challenging, stimulating

Intense, productive, joyful

THE MOST APRECIATED ASPECT

The meeting with new researchers and novel methodologies

To share ideas and start a collaboration with a very receptive environment

The amazing people I had the possibility to meet

The very kind availability of the hosting unit manager and all his group, you feel “at home”

The research experience and the life/travel experience

The networking and exchange of knowledge with students, researchers, and professors that share common research topics

The availability of our hosts and their focus on collaboration for growing

The vibrant environment, many active research groups, a lot of enthusiasm and very high quality of the students, very warm welcome

THE BRING HOME MESSAGE

Enthusiasm brings new ideas

Research is on exchanging and growing ideas

To change work environment once in a while is very beneficial for a researcher

Sharing ideas with people with different expertise can yield very interesting future projects

Integration of knowledge derived from different datasets is a particularly challenging topic

Greats ideas can come up when putting together people with different expertise and domain background

The deep and detailed research discussions

Learn something new every day

Towards Semantic-Aware

Multiple-Aspect Trajectory Similarity Measuring

Lucas May Petry, Federal University of Santa Catarina, Brazil

We are living in the information age, generating data basically about everything we do, including when we move (and when we don't). Even though in the past mobility data was constrained to the data generated by GPS devices, nowadays we have available much richer movement data that can be extracted from social networks, such as Facebook, Foursquare, and Twitter, to name a few. By enriching mobility data with such information, GPS trajectories become what we call multiple-aspect trajectories.

As more semantic information about movement becomes available, more sophisticated mobility analysis techniques become necessary. One technique that is very useful and needed by many algorithms are similarity measures, as they are the basis for nearest neighbor search, nearest neighbor classifiers, and clustering. Existing similarity measures either consider all trajectory dimensions together, enforcing a full match in all dimensions, or consider all dimensions as independent, not allowing a relationship (dependency) between a subset of dimensions. In view of this, in this work we propose a new similarity measure for multiple-aspect trajectories, namely **MUltIple-aspect TrAjectory Similarity (MUITAS)**. Differently from existing trajectory similarity measures, MUITAS allows the definition of attribute relationships in the similarity assessment.

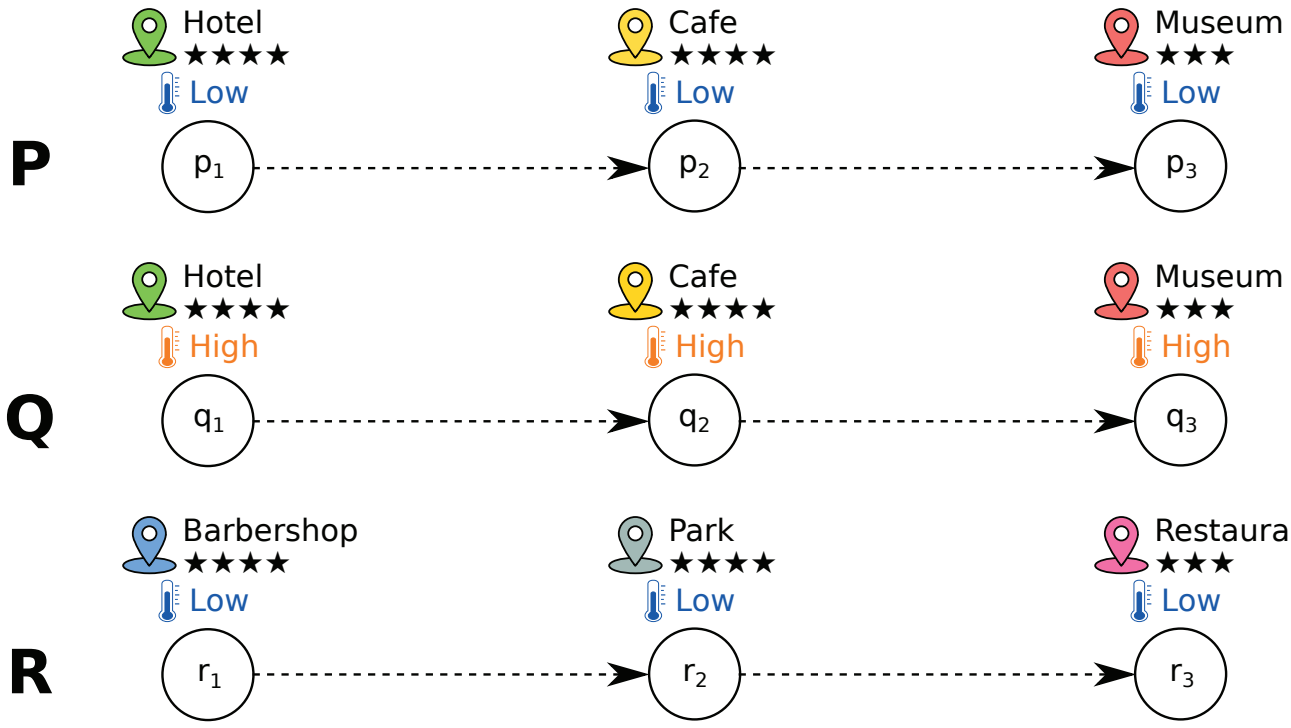


Figure 1. Example of multiple aspect trajectories P, Q, and R.

Let us consider the example shown in *Figure 1*, with trajectories P, Q, and R. For the sake of simplicity we consider only three attributes in the example: the category of place visited and its rating, representing the POI aspect; and the temperature representing the weather aspect. Trajectories P and Q visit the same categories of places (hotel, cafe, and museum) and with the same rating. The main difference between P and Q is that P occurs where the temperature is low, while for trajectory Q the temperature is high. Trajectory R, on the other hand, goes to different POIs (barbershop, park, and restaurant), but their ratings are the same as for trajectories P and Q, and the temperature is always low.

Now let us suppose that we want to find the trajectory that is the most similar to P. In our example, trajectory Q is the most similar to P, because both trajectories visit places of the same category (museum, cafe, and hotel) and with the same rating, only differing in the temperature. However, without considering the semantics of the attributes and their relationships, trajectories P and Q and trajectories P and R would have the same similarity score by state-of-the-art methods, because they all share two common attributes: P and Q share the POI category and rating, while P and R share the rating and temperature. We

claim that when analyzing the similarity of multiple-aspect trajectories, the semantics of the attributes and their relationships is more important than simply counting the number of attribute values that match or do not match.

We believe that even though trajectory P shares two attribute values with R (rating and temperature), the trajectories are semantically different, because they visit completely different categories of places. In our example the attribute rating is associated to the place visited, so its semantics relies on the aspect POI, and its meaning is lost without the POI category, so these attributes should not be disassociated or, in other words, are dependent.

Therefore, we take these characteristics into account to develop MUITAS, the new similarity measure that supports both independent and dependent attributes, allowing the definition of attributes that have a semantic relationship. The measure also supports the use of a different distance function for each attribute and the definition of a weight that represents the importance degree of each attribute.

We evaluate MUITAS with two real-world trajectory datasets. The first one contains user check-ins on Foursquare, which we enriched with weather information and other characteristics about the places visited by the users, since such aspects may potentially enhance the similarity analysis of trajectories. The second dataset was collected and annotated by volunteers in Pisa, Italy, in the context of the "TagMyDay" experiment and made available to MASTER researchers from the **SoBigData** project (http://www.sobigdata.eu) research infrastructure. This dataset contains user trajectories annotated with the means of transportation, the activity performed by the user, the weather conditions, the distance travelled, etc. The experimental results show significant improvement of MUITAS when compared to existing works.

Even though we focused on multiple-aspect trajectories, the proposed similarity measure can be applied to any type of trajectory or sequenced data in a variety of applications. Future works include considering heterogeneous trajectories, in the sense that they may not be annotated with the same attributes or characteristics. Moreover, we want to investigate machine learning approaches for defining the relationships between attributes required by MUITAS. This is useful when we have none or little knowledge about the application domain, or even when trajectories have too many attributes.

This research work was a collaboration between the Federal University of Santa Catarina in Brazil and the National Research Council of Pisa in Italy, in the frame of the research project MASTER (Marie Skłodowska-Curie research grant from the EC and the H2020 framework, under the agreement No 777695).
Extracted from:
Lucas May Petry, Carlos Andres Ferrero, Luis Otavio Alvares, Chiara Renso, Vania Bogorny
Towards Semantic-Aware Multiple-Aspect Trajectory Similarity Measuring. In Transactions in GIS, Wiley
First published: 18 June 2019 https://doi.org/10.1111/tgis.12542

A network abstraction model

for vessel trajectory analysis and anomaly detection

Konstantinos Tserpes and Iraklis Varlamis, Harokopio University of Athens

In pursue of a better analysis and understanding of the behavior of vessels in the sea, the scientific community and industry alike have resorted to maritime surveillance systems. The wealth of information that such systems provide, with the most prominent being AIS, benefits a multitude of applications including anomalous behavior detection. Abnormal vessel behavior can be indicative for a set of note-worthy events, such as a vessel being in distress or a vessel performing illegal activities. The impact of those events is severe and has a multifaceted effect on the environment, society, economy, etc. It is, therefore, crucial to use the technology to enable the early detection of such-like events. The opportunity is now more relevant than ever, with distributed data sensors tracking and reporting vessel movements around the globe.

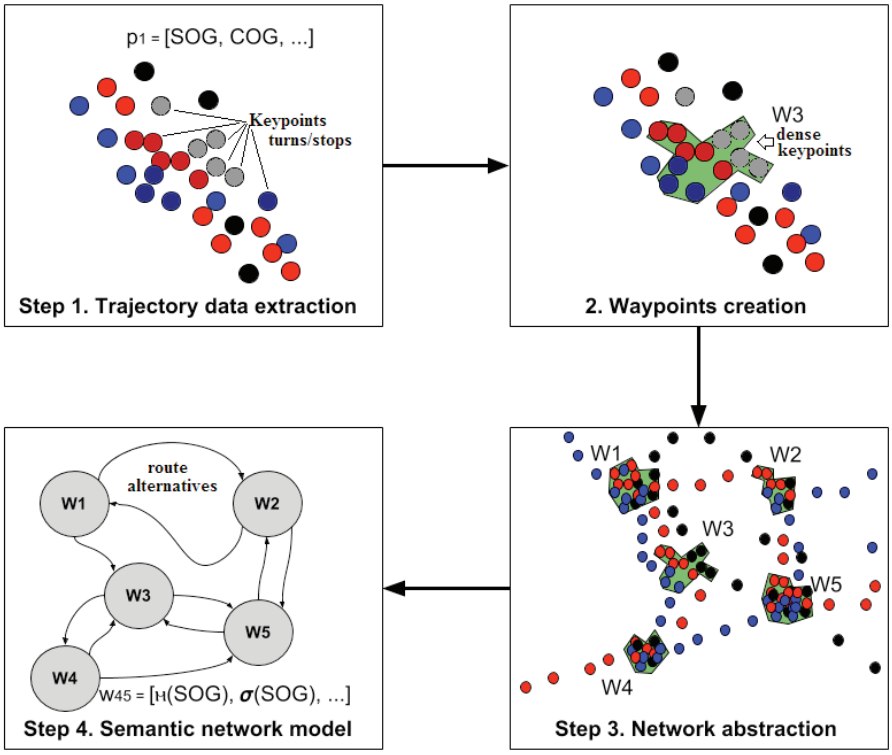


Figure 1: The main steps of the proposed model.

The original motivation behind our work was to tackle the problem of anomaly detection in the sea using a model of normal behavior as a basis, which will be extracted from existing vessel trajectory data. The intuition was that a better solution requires the attribution of context-based knowledge to vessel trajectory data, such as i) the waypoints that define vessel operations and the sort of movement patterns that they follow in relation to those waypoints (i.e., a region of interest for a given application) over time, ii) the sub-trajectories that compose the trajectory of a vessel and the features that can be extracted for them. The main idea is to use AIS data from multiple vessels to identify the spatial waypoints according to frequently observed vessels pattern, such as being stationary or making significant changes in their courses. Then, to understand the frequency and transition patterns of vessels moving from one waypoint to another using data from multiple vessels, and finally to generate a network that captures all this information. Given this network abstraction model, trajectory analysis can be performed to detect unexpected vessel behaviors.

This last step was of particular importance: handling the volume of AIS data, which comprises a vast data stream, constitutes a major challenge for traditional data analysis methods and machine learning algorithms. So, it is essential, before any further analysis, to simplify vessel trajectories and if possible to abstract the transactional model of AIS streams to a model that fits data mining and analytics.

A large part of the literature on vessel position prediction and anomaly detection focuses on the analysis of momentary GPS coordinates and vessel movement features (e.g., velocity, bearing). This point-based examination of the trajectories, however, implies the attri-

bution of the same value in the analysis to each GPS point and this may result in weak results in understanding vessel behavior. In reality, there are states in the vessels' trajectories, which are of high importance in the context of the real vessel operation that include the traversal of spatiotemporally defined waypoints (e.g. ports, off-shore platforms, capes, route deviations, vessel towing etc.) and a large number of states of low or no contribution to the solution of the problem at hand.

Early experiments with AIS data from the Mediterranean, gracefully provided by MarineTraffic.com, showed that the proposed network model is quite abstract and achieves a good compression of vast amounts of data collected from thousands of vessels that operate in an area. At the same time, it is very comprehensive in the information it keeps for vessels' trajectories and allows more complex analysis to be performed, such as clustering or classification of movement patterns. The network abstraction of vessel trajectories for a region, can be used for processing new AIS data that come as a stream for this region, and quickly detect vessels that move from one waypoint to another or deviate from the predefined routes.

The preliminary results in 5,782 distinct trips performed by 1,716 vessels during a one-month period, identified a number of potential outlier cases which were empirically evaluated as true positives in practically all cases.

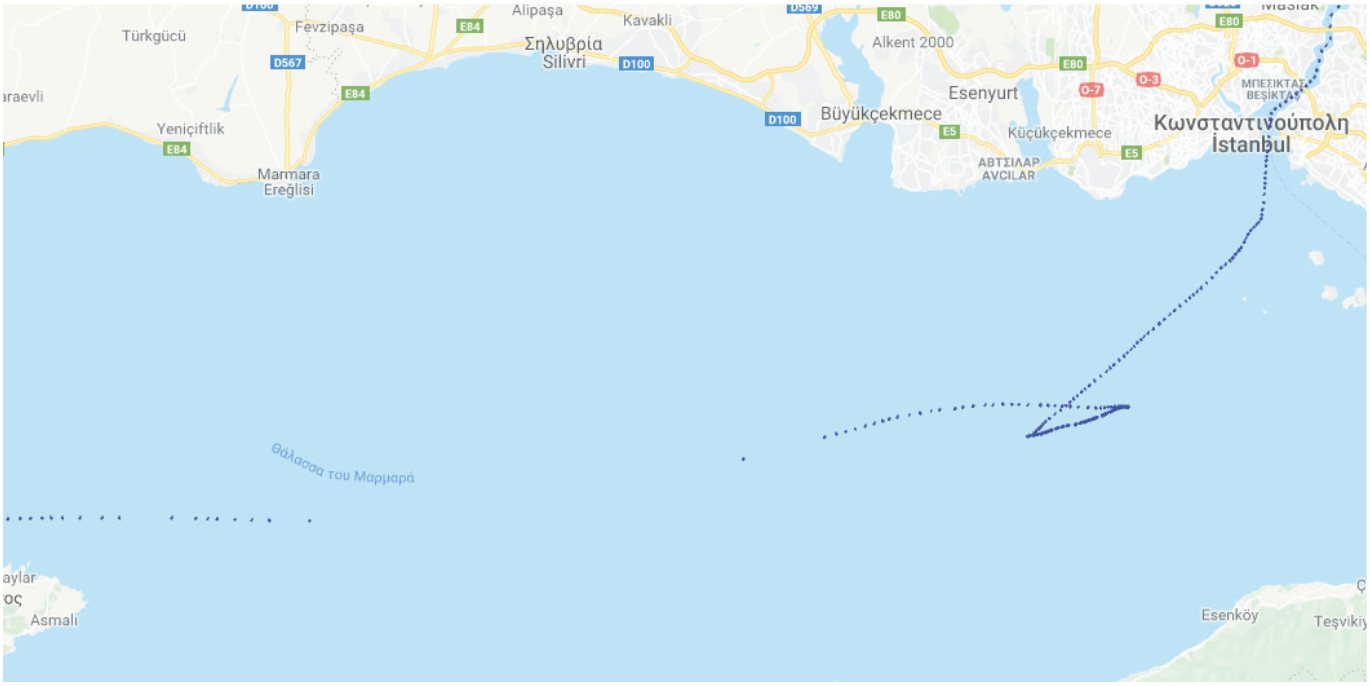


Figure 2: An extract of the trajectory of a test vessel in the dataset, which has been detected as outlier.

This research work was a collaboration between Harokopio and Dalhousie Universities in the frame of the research project MASTER (Marie Skłodowska-Curie research grant from the EC and the H2020 framework, under the agreement No 777695).
The first results were presented in the BMDA workshop of the EDBT/ICDT conference.

Varlamis, I., Tserpes, K., Etemad, M., Júnior, A. S., & Matwin, S. (2019). A Network Abstraction of Multi-vessel Trajectory Data for Detecting Anomalies. In EDBT/ICDT Workshops.

Past Events

MID TERM MEETING

The Mid Term meeting of MASTER has been held in Pisa, the coordinator location, last April 11th, 2019. The mid term meeting is a specific event of RISE projects to be held in the period between month 13 and month 18 with the presence of the Project Officer. The objective of the meeting is to discuss the implementation of the project and receive feedback from the EC.
The meeting has seen a broad participation of both European and Third Country partners. We are especially proud that three of our Third Country partners from Brazil and Canada were able to participate in person in the event. A total of 25 participants attended the meeting.
We had a total of 21 attendees from beneficiaries and 4 from Third Countries: 11 attendees from National Research Council of Italy, 2 from University Ca' Foscari of Venice (Italy), 2 from University of Piraeus Research Center (Greece) (1 by Skype), 2 from University of Versailles Saint-Quentin (France), 4 from Harokopio University of Athens (Greece), 2 from Federal University of Santa Catarina (Brasil), 1 from Federal University of Ceará (Brasil), 1 from Dalhouise University (Canada).

All the WPs has been presented by the respective WP leader and one secondee. We also included a "secondment experience" session presented by the Third Countries unit managers, together with a secondee, to briefly report their secondment experience. The Project Officer Simona Losmanova and the Independent Ethics Advisor Prof. Bettina Berendt participated in the event.
We also held a networking dinner the evening of April 10th, 2019 in the downtown of Pisa, where 21 attendees could participate.

Together with the Mid Term meeting we also hold a **RISE promotional event,** where external participants could attend a presentation of the RISE program and the forthcoming call, presented by the **Project Officer Simona Losmanova**. A total of 21 participants from outside the consortium, in addition to the 25 meeting participants, attended the event. Some of these attendees came from other institutes of local CNR, University of Pisa, Scuola Superiore San Anna, Scuola Normale Superiore, but also institutions from Florence and Venice. Mrs. Losmanova shared RISE promotional material with all participants.

BIG MOBILITY DATA ANALYTICS WORKSHOP

On March 26th the **Second International workshop on Big Mobility Data Analytics** has been held joint to EDBT conference in Lisbon (http://bmda19.datastories.eu). **MASTER** scientifically contributed to this workshop, being both UPRC and CNR researchers in the Organizing Committee which includes a total of 6 researchers: Alexander Artikis (University of Piraeus & NCSR Demokritos), Themis Palpanas (Paris Descartes University), Nikos Pelekis (University of Piraeus), Chiara Renso (ISTI-CNR Pisa), Yannis Theodoridis (University of Piraeus & Athena RIC), Dimitrios Zissis (University of the Aegean). The program included one keynote speaker, Prof. Dimitrios Gunopulos who presented his speech: "Urban data analysis: Travel time estimation in real time". The program included six papers and all of them are from MASTER partners (Harokopio, Piraeus, Versailles and Dalhousie Universities).

An audience of about 35 people attended the event.

The proceedings are freely accessible from the URL: http://ceur-ws.org/Vol-2322/

LIST OF ACCEPTED PAPERS:

+ **Integration of Mobility Data with Weather Information** | N. Koutroumanis, G. M. Santipantakis, A. Glenis, C. Doulkeridis, G. A. Vouros
+ **Towards A Semantic Indoor Trajectory Model** | A. Kontarinis, K. Zeitouni, C. Marinica, D. Vodislav, D. Kotzinos
+ **Performance Evaluation of MongoDB and PostgreSQL for Spatio-temporal Data** | A. Makris, K. Tserpes, G. Spiliopoulos, D. Anagnostopoulos
+ **A Trajectory Segmentation Algorithm Based on Interpolation-based Change Detection Strategies** | M. Etemad, A. Soares Junior, A. Hoseyni, J. Rose, S. Matwin
+ **A Network Abstraction of Multi-vessel Trajectory Data for Detecting Anomalies** | I. Varlamis, K. Tserpes, M. Etemad, A. Soares Junior, S. Matwin
+ **On Feature Selection and Evaluation of Transportation Mode Prediction Strategies** | M. Etemad, A. Soares, S. Matwin, L. Torgo

FATES ON THE WEB 2019

On May 14 2019, the **1st Workshop on Fairness, Accountability, Transparency, Ethics, and Society on the Web** has been held in conjunction with The Web Conference 2019, in San Francisco, CA, USA .
Can we build inclusive and representative machine-learning based-algorithms? Who is responsible for harm when algorithmic decision-making results in discriminatory outcomes? To whom should algorithms be transparent? What approaches to ethics might algorithms require?
The ***FATES on the Web 2019*** (Fairness, Accountability, Transparency, Ethics, and Society on the Web) tried to answer all these questions, bringing together researchers and enthusiasts concerned with the urgent challenges concerning algorithmic fairness and accountability, transparency, and ethics on data management and social interaction on the web.

The workshop was co-located with The Web Conference 2019 (formerly known as WWW conference), hosted in San Francisco, California, celebrating this year the 30th anniversary of the Web.

The program included 19 high quality contributions in the form of long, short and discussion papers. Approximate number of attendees during the full day event was about 80 people.

More info here:
http://fates19.isti.cnr.it

LIST OF ACCEPTED PAPERS:

+ **Privacy and Transparency within the 4IR: Two faces of the same coin** | B. Teixeira; D. Schwabe; F. Santoro; F. Baião; M. L. Campos; L. Verona; C. Laufer; S. Barbosa; S. Lifschitz; R. Costa–
+ **Privacy-aware Linked Widgets** | J. D. Fernández; F.J. Ekaputra; P. Ruswono Aryan; A. Azzam; E. Kiesling
+ **Trust and trustworthiness in social recommender systems** | T. Hassan
+ **Black Hat Trolling, White Hat Trolling, and Hacking the Attention Landscape** | J. Matthews; M.Goerzen
+ **Fairness in the social influence maximization problem** | A. Stoica; A. Chaintreau
+ **Keynote talk by Frauke Kreuter: The Social Science of Privacy – Effects on Industry and Government Data Use**
+ **Can Location-Based Searches Create Exposure Bias?** | G. K. Patro; A. Banerjee; N. Ganguly; K. P. Gummadi; A. Chakraborty
+ **What's in a Name? The Need for Scalable External Audit Infrastructure** | A. Korolova; K. Shah
+ **In Defense of Synthetic Data** | L. Rodriguez; B. Howe
+ **Algorithms for Fair Team Formation in Online Labour Marketplaces** | G. Barnabò; A.Fazzone; C. Schwiegelshohn; S. Leonardi
+ **Quantifying the Impact of User Attention on Fair Group Representation in Ranked Lists** | P. Sapiezynski; W. Zeng; R. E. Robertson; A. Mislove; C. Wilson
+ **Unsupervised Topic Extraction from Privacy Policies** | D. Sarne; J. Schler; A. Singer; A. Sela; I. Bar Siman Tov
+ **Collaborative Explanation of Deep Models with Limited Interaction for Trade Secret and Privacy Preservation** | J. Domingo-Ferrer; C. Pérez; A.Blanco-Justicia
+ **Achieving Differential Privacy and Fairness in Logistic Regression** | D. Xu; S. Yuan; X. Wu
+ **On Preserving Sensitive Information of Multiple Aspect Trajectories In-House** | S. Giotakis; N. Pelekis
+ **Hegemony in Social Media and the effect of recommendations** | A. Stoica; A. Chaintreau
+ **Uncovering Social Media Bots: a Transparency-focused Approach** | E. Ferreira Dos Santos; D. Carvalho; L. Ruback; J. Oliveira
+ **Managing Bias in AI** | D. Roselli; J. Matthews; N. Talagala
+ **Nuanced Metrics for Measuring Unintended Bias with Real Data for Text Classification** | D. Borkan; L.Dixon; J. Sorensen; N. Thain; L. Vasserman
+ **Empirical analysis of bias in voice based personal assistants** | L. Lima; V. Furtado; E. Furtado; V.Almeida

Next Events

FIRST MASTER WORKSHOP

The first MASTER workshop will be held in conjunction with the ECML-PKDD conference in Wurzburg, Germany the day September 16th, 2019. The event includes the presentation of eight full papers, five of which are from MASTER partners presenting recent scientific results from project activities. The program also includes an invited keynote talk given by Prof. Yannis Theodoridis with title "**Learning from our movements – The mobility data analytics pipeline**".

LIST OF ACCEPTED PAPERS:

+ **Uncovering hidden concepts from AIS data: A network abstraction of maritime traffic for anomaly detection** | I. Kontopoulos, I. Varlamis, K. Tserpes
+ **Predicting Fishing Effort and Catch Using Semantic Trajectories and Machine Learning** | P. Adibi, F. Pranovi, A. Raffaetà, E. Russo, C. Silvestri, M. Simeoni, A. Soares, S. Matwin
+ **Online long-term trajectory prediction based on mined route patterns** | Petros Petrou, Panagiotis Tampakis, Harris Georgiou, Nikos Pelekis and Yannis Theodoridis (University of Piraeus, Greece)
+ **EvolvingClusters: Online Discovery of Group Patterns in Enriched Maritime Data** | G.S. Theodoropoulos, A. Tritsarolis, Y. Theodoridis
+ **A Neighborhood-augmented LSTM Model for Taxi-Passenger Demand Prediction** | T. Le Quy, M. Spiliopoulou, W. Nejdl, E. Ntoutsi
+ **Multi-Channel Convolutional Neural Networks for Handling Multi-Dimensional Semantic Trajectories and Predicting Future Semantic Locations** | A. Karatzoglou
+ **Nowcasting Unemployment Rates with Smartphone GPS data** | D. Moriwaki
+ **Prospective Data Model and Distributed Query Processing for Mobile Sensing Data Streams** | M. Brahem, K. Zeitouni, L. Yeh, H. El Hafyani

For more info see:
http://www.master-project-h2020.eu/workshop-master-2019/

BRIGHT - LA NOTTE DEI RICERCATORI 2019

MASTER is participating to the BRIGHT 2019, that will be held on September 27 in Pisa-Italy, with a seminar entitled
DIMMI DOVE VAI E TI DIRO' CHI SEI
(TELL ME WHERE YOU GO AND I WILL TELL YOU WHO YOU ARE)

The BRIGHT project aims at enhancing the visibility and perception of researchers among the general population in Tuscany, simultaneously with all the European Researchers' Night (ERN) initiatives.
More info here:
http://www.bright-toscana.it/seminari-2019/

MOBILE DATA MANAGEMENT CONFERENCE AND RELATED WORKSHOPS

MASTER is also supporting the organization of the 21th International Conference of Mobile Data Management 2020 to be held in Versailles, Paris, France, organized by our partner Prof. Karine Zeitouni from UVSQ.
MDM is seeking contributions for research papers and workshop proposals.
Tentative deadline for workshop proposal is December 8th 2019 and for research papers is January 17th, 2020.
More info here:
http://mdmconferences.org/mdm2020/

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant No 777695
Duration: March 1 2018 | February 28 2022

Copyright notice

Privacy statement
The personal data (names, email addresses…) and the other information entered in MASTER Newsletter will be treated according with the provision set out in Legislative Degree 196/2003 (known as Privacy Code) and subsequently integration and amendment.

Coordinator of the project: Chiara Renso | chiara.renso@isti.cnr.it

MASTER Newsletter is not for sale but it is distributed for purposes of study and research and published online at http://www.master-project-h2020.eu