

OBAVUM: An Ontology-based Approach to Visualizing Urban Mobility Data

Thiago Sobral, Vera Costa, José Borges

Department of Industrial Engineering and Management
Faculty of Engineering of the University of Porto
Porto, Portugal
thiago@fe.up.pt, veracosta@fe.up.pt, jlborges@fe.up.pt

Tânia Fontes, Teresa Galvão

INESC TEC - INESC Technology and Science and FEUP
Faculty of Engineering of the University of Porto
Porto, Portugal
trfontes@fe.up.pt, tgalvao@fe.up.pt

Abstract—This paper proposes an ontology-based approach to visualizing urban mobility data. Our approach, which is in ongoing development, is centered in a visualization-oriented urban mobility ontology that is used to semantically characterize data and visualization techniques. We present a practical application to a public transportation network of the city of Porto, Portugal. We address how semantics can empower and facilitate tasks like automatic recommendation of visualization techniques, and definition of a data filter based on passengers' journey patterns.

Keywords- *urban mobility; information visualization; ontology; semantic web technologies*

I. INTRODUCTION

Related literature shows how visualization techniques are relevant for urban mobility themes like congestion analysis [1]; optimization of public transportation networks based on mobile network data [2]; detection of abnormal fluctuations in passenger ridership [3]; and modeling and visual representation of time in visualizations [4]. Nevertheless, a literature review study [5] explains that most of these techniques seems to focus more on helping to understand the system usage and to improve urban planning, rather than to improve the usability and accessibility of the system for the travelers. In fact, those techniques are generally restricted to a dataset with a particular schema. Such schema-dependency may limit the reproducibility and scalability of techniques in similar contexts by others.

The use of ontologies and Semantic Web technologies (SWT) can be used to solve this problem and increase the power of visualizations. Nonetheless, few works have developed semantic visualizations for urban mobility data. Some studies mostly address the development of tools for supporting passengers in route planning and tourist information retrieval [6,7,8]. Others have developed semantic visualization for 3D urban environments: Pina et al. [9] illustrates a scenario in which the user decides which bus line to take using the visualization tool. Gomes et al. [10] developed an ontology-based visualization tool for urban planning, notably on classification of land use.

Urban mobility data are generally not provided with semantics, implying that the knowledge extraction process depends solely on human inputs. Therefore, the OBAVUM approach comprises the development of semantically rich urban mobility visualizations. Such approach considers the

integration and characterization of data through a visualization-oriented urban mobility ontology, which is also used to semantically characterize visualizations, according to their intrinsic characteristics as well as human perception factors, so as to enhance the adequacy and utility of visualizations in practical contexts by transportation analysts, our target users.

The objectives of this paper are:

- To describe the main components of the visualization-oriented urban mobility ontology;
- To describe and exemplify how data and visualization techniques can be characterized with this ontology.

We present a practical application with real data retrieved from the public transportation network of Porto, Portugal. The example explores the semantic specification of a visualization technique and shows how semantics can empower tasks like (1) automatic recommendation of visualization techniques for target users; (2) and definition of a data filter based on passengers' journey patterns.

II. AN ONTOLOGY-BASED APPROACH

The visualization-oriented urban mobility ontology has a number of classes and properties related to infrastructure and events of an urban mobility system (e.g. *BusStop* and *TicketValidation*). Elements are derived after analyzing datasets from complex transportation networks of cities like Porto, London, and Rio de Janeiro. In this section, we present the main ontology elements related to data characterization (see section II.A). After, we present in further detail the elements related to the semantic characterization of visualizations (see section II.B). Fig. 1 presents all the ontology elements addressed in this paper.

A. Characterization of Urban Mobility Data

The results of our previous studies [11,12] allowed characterizing urban mobility events in three main classes:

- *TravelIntention*: an expectation of traveling in a transportation network, generally derived from data related to route planner or schedule consultation (in case of public transportation);
- *TravelEvent*: a travel made by a passenger;
- *TrafficFlow*: detection of moving entities (vehicles or people).

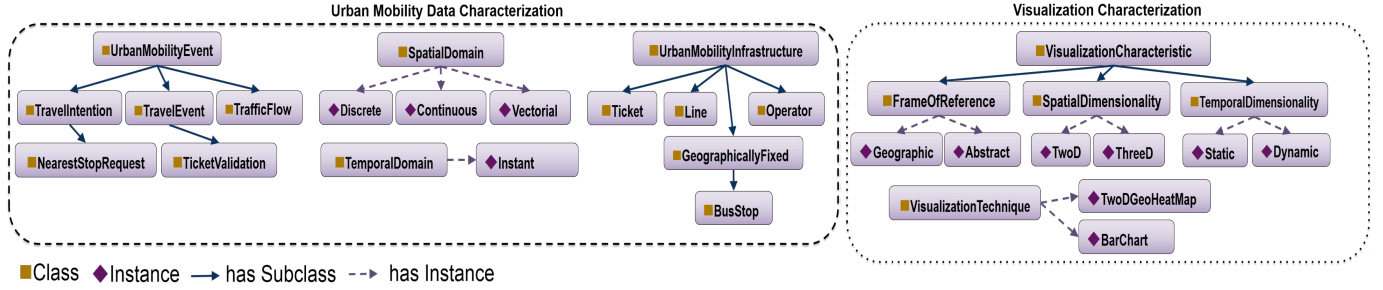


Figure 1. An excerpt of the OBAVUM ontology presenting the elements addressed in this paper.

Urban mobility data usually have different spatiotemporal types, with each data unit belonging to one of the following spatial instances of the class *SpatialDomain*:

- *Discrete*: a finite, “well defined” set of spatial coordinates (e.g. location of bus stops). By “well defined” we mean that locations are likely to be previously known or associated to fixed locations;
- *Continuous*: a finite but much broader set of spatial coordinates, resembling the notion of continuity (e.g. all possible latitude-longitude pairs in a city);
- *Vectorial*: a set containing two or more nodes of a transportation system (e.g. origin-destination pairs).

One advantage of ontologies and semantic data over common approaches (e.g. relational databases) is the support of logical reasoning, which allows the computer to extract knowledge based on pre-determined logic rules. We specified a set of rules for automatic inference of the spatial domain of data, given herein in human-readable Semantic Web Rule Language (SWRL) syntax.

We defined that the spatial domain of a certain instance (a semantic data unit) of urban mobility event is *Discrete* whenever it occurs at a location that belongs to the class *GeographicallyFixed*, which comprises all physical elements that have a fixed position. The rule is given by the following expression:

$$\text{occursAt}(?x, ?y) \wedge \text{GeographicallyFixed}(?y) \Rightarrow \text{hasSpatialDomain}(?x, \text{Discrete}) \quad (1)$$

A *Continuous* spatial domain is defined when an event occurs at a location that is not geographically fixed:

$$\text{occursAt}(?x, ?y) \wedge \neg \text{GeographicallyFixed}(?y) \Rightarrow \text{hasSpatialDomain}(?x, \text{Continuous}) \quad (2)$$

Similarly, we defined that a *Vectorial* spatial domain applies whenever an instance has defined both properties, *hasOrigin* and *hasDestination*:

$$\text{hasOrigin}(?x, ?y) \wedge \text{hasDestination}(?x, ?z) \Rightarrow \text{hasSpatialDomain}(?x, \text{Vectorial}) \quad (3)$$

On regards to time domain, we only considered data units that can be associated to instants in time by the means of a timestamp. Hence, data units can belong to an *Instant* temporal domain. Data units related to more complex time

representations (e.g. time intervals, multiple time instants) are still not supported.

B. Characterization of Visualization Techniques

In order to develop semantically rich visualization techniques, the ontology contains a number of elements that allow for their characterization, which includes essential properties stated on visualization-related literature such as (1) *name*; (2) *frame of reference* (geographic or abstract); (3) *time representation* (static or dynamic); and (4) *dimensionality of presentation space* (2D or 3D) [13].

An important feature of the OBAVUM approach is the ability to semantically characterize human perception factors. As we acknowledge that visualization techniques should be user-centered, the specification of those factors are expected to occur after the evaluation with target users.

After carrying an extensive user-centered design process with 10 transportation analysts, we gathered important inputs related to the adequacy of visualization techniques on regards to the type of data involved, and the specific information that analysts want to retrieve [11,12]. The following factors were defined as properties in the ontology:

- Prefix: *hasRecommended*

- *Subject*: urban mobility concepts that were considered adequate to be visualized with a given technique (e.g. *TicketValidation*, *NearestStopRequest*);
- *SpatialDomain*: the recommended spatial domains of a dataset for a given technique (e.g. *Discrete*, *Continuous* or *Vectorial*);
- *TemporalDomain*: the recommended temporal domains of a dataset for a given technique (e.g. *Instant*).

The automatic inference of recommended visualization techniques is achieved by building a semantic query that associates the property *hasRecommendedVisualization()* to an instance whenever its properties match the ones of the visualization technique. Fig. 2 presents an example regarding a ticket validation instance and two visualization techniques that were inferred as recommended for it.

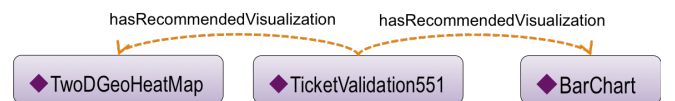


Figure 2. Example of recommended visualizations that were automatically inferred for the ticket validation instance 551.

III. APPLICATION TO A PUBLIC TRANSPORTATION NETWORK

We present the semantic specification of a visualization technique, preceded by an example of data integration by mapping the entries of the datasets onto the ontology.

A. Study domain

In the Metropolitan Area of Porto, the public transportation network covers an area of 1,575 km² and serves 1.75 million of inhabitants [14]. The network is composed of 126 buses lines (urban and regional), 6 metro lines, 3 tram lines, 3 train lines and 1 cable line [15]. This network is operated by 11 transport providers, of which Metro do Porto and STCP are the largest.

The Porto network is based on an intermodal and flexible ticket system: called Andante. Andante is an open zonal system, based on smart cards, that requires validation only when boarding. A validated occasional ticket allows for unlimited travel within a specified area and time period: 1 hour for the minimum 2-zone ticket, and longer as the number of zones increases. Andante holders can use different lines and transport modes in a single ticket.

B. Material and Methods

Two datasets were used as reference:

- Dataset 1 (D1): Ticket validation information collected from bus (STCP company) and metro;
- Dataset 2 (D2): Users' locations at the time they asked Move-me mobile app [16] to search for the nearest transportation stops. Move-me provides real time information about Porto's public transportation system.

The following entries present part of the schemes of D1 and D2 that were considered for this example:

D1(operator,cardID,stop,line,direction,validationDateTime)
D2(requestDateTime,latitude,longitude)

Entries of D1 and D2 were mapped onto the ontology elements, resulting in triples in Resource Description Framework (RDF) form (subject + predicate + object), which is the standard data model for semantic data.

An example of the resulting instances and their respective triples follows:

```
:TicketValidation100 rdf:type :TicketValidation ;
:hasOperator :OperatorSTCP ;
:hasTicket :Ticket590 ;
:occursAt :BusStop1137 ;
:hasLine :Line305 ;
:hasDirection "2"^^xsd:Integer ;
:hasValidationDateTime "2013-08-08T07:59:56"^^xsd:DateTime .

:NearestStopRequest301 rdf:type :NearestStopRequest ;
:hasRequestDateTime "2010-05-05T19:53:12"^^xsd:DateTime ;
:occursAt [ :latitude "41.145251"^^xsd:Double ;
:longitude "-8.616015"^^xsd:Double . ] .
```

Our example assumes that other supporting datasets containing information such as operators, bus stops and lines were already mapped onto the ontology. Examples of their

resulting instances found above are :OperatorSTCP, BusStop1137: and :Line305. Those instances also have associated properties and values in the same way we exemplified for a ticket validation and a nearest stop request.

We selected the 2D Geographic Heat Map visualization technique as an example (see Fig. 3), due to its intuitive nature and acceptability among the transportation analysts we consulted. The density of data in a certain geographic area gives the concept of heat. Higher data density is depicted as higher color temperatures. Transportation analysts evaluated it in our previous studies [11,12] through semi-structured interviews and exploratory usability tests.

By nature, ontologies offer an advantage: other individuals can extend them by defining new classes and properties, in case the existing ones are too generic or do not suffice for using in particular contexts. As an example, we defined a data filter based on passengers' journey patterns particularly for the city of Porto.

C. Results and discussion

After analyzing all TicketValidation and NearestStopRequest instances that resulted from D1 and D2, a logic reasoner deduced their spatial domain. Since BusStop was defined as a subclass of GeographicallyFixed, all instances of BusStop are also geographically fixed. Hence, expression (1) allowed to infer that :TicketValidation100 has a discrete spatial domain, according to the tautology:

$$\begin{aligned} & \text{occursAt}(\text{TicketValidation100}, \text{BusStop1137}) \wedge \\ & \text{GeographicallyFixed}(\text{BusStop1137}) \Rightarrow \\ & \text{hasSpatialDomain}(\text{TicketValidation100}, \text{Discrete}) \end{aligned}$$

On the other hand, instance :NearestStopRequest301 had a continuous spatial domain, since the value of the property occursAt() does not belong to GeographicallyFixed. The latitude and longitude triples were omitted from the following tautology for the sake of readability. They were depicted as "*". Thus, expression (2) allowed to perform the following inference:

$$\begin{aligned} & \text{occursAt}(\text{NearestStopRequest301}, *) \wedge \neg \text{GeographicallyFixed}(*) \Rightarrow \\ & \text{hasSpatialDomain}(\text{NearestStopRequest301}, \text{Continuous}) \end{aligned}$$

We argue that inferring the spatial domain for each instance is more appropriate and flexible than stating, *a priori*, that all instances of a certain class belong to a particular one (e.g. "All TicketValidation should have Discrete spatial domain"). This prevents data units with missing spatial information from being associated to a spatial domain, causing issues during the visualization process.

Transport analysts considered the 2D Geographic Heat Map appropriate for analyzing travel events and intentions, particularly ticket validations and nearest stop requests. Analysts stated that data belonging to a continuous spatial domain would be preferred than discrete, although the visualization can technically depict both. At the current development stage, the ontology does not support the quantification of users' preference between types of data, thus both were equally defined as recommended.

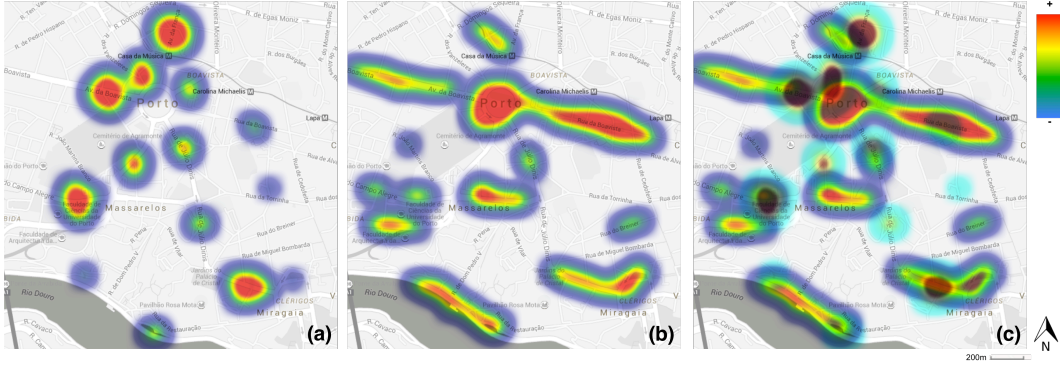


Figure 3. The 2D Geographic Heat Map visualization technique coupled with data from (a) ticket validations, (b) nearest stop requests and (c) both.

This visualization is only compatible with static representations of time (i.e. it shows all data units that have timestamps inside a time interval specified by the user).

Based on the analysts' inputs and the intrinsic characteristics of this visualization technique, we provided the following semantic specification:

```
:TwoDGeoHeatMap rdf:type :VisualizationTechnique ;
# Intrinsic characteristics
  rdf:label "2D Geographic Heat Map" ;
  :hasFrameofReference :Geographic ;
  :hasTemporalDimensionality :Static ;
  :hasSpatialDimensionality :TwoD ;
# Human perception factors
  :hasRecommendedSubject :TicketValidation ;
  :hasRecommendedSubject :NearestStopRequest ;
  :hasRecommendedSpatialDomain :Discrete ;
  :hasRecommendedSpatialDomain :Continuous ;
  :hasRecommendedTemporalDomain :Instant .
```

Note that our semantic specification for this particular visualization technique does not intend to be general. The rationale behind the OBAVUM approach is the way that the elements of the ontology can be used for providing such specification.

In order to identify which instances were compatible with the 2D Geographic Heat Map, we built a semantic query that constructed a graph of all instances that matched the recommendations of the visualization technique:

```
CONSTRUCT ( ?x :hasRecommendedVisualization :TwoDGeoHeatMap )
WHERE {
  ?x rdf:type ?subject .
  ?x :hasSpatialDomain ?spatialdomain .
  ?x :hasTemporalDomain ?temporaldomain .
  ?subject IN {
    SELECT ?sbj
    WHERE {
      :TwoDGeoHeatMap :hasRecommendedSubject ?sbj .
    }
  }
  ?spatialdomain IN {
    SELECT ?sd
    WHERE {
      :TwoDGeoHeatMap :hasRecommendedSpatialDomain ?sd .
    }
  }
  ?temporaldomain IN {
    SELECT ?td
    WHERE {
      :TwoDGeoHeatMap :hasRecommendedTemporalDomain ?td .
    }
  }
}
```

Executing this query produced instances similar to the ones depicted in Fig. 2. Ideally, the process of query

definition and execution should be internally performed in real time by the visualization tool, in order to not require users to execute it manually.

Suppose a use case of an analyst willing to visualize data from nearest stop requests with a visualization tool compliant with the OBAVUM approach, with a library of semantically rich visualizations. After integrating data through the ontology, the analyst wants to know which visualization techniques could be of use with that data. The tool would then provide a list of recommended techniques by internally executing a straightforward semantic query similar to:

```
SELECT ?visualization
WHERE {
  ?x rdf:type :NearestStopRequest .
  ?x :hasRecommendedVisualization ?visualization .
}
```

Since the property *hasRecommendedVisualization()* was applied directly to instances, the list of recommended visualizations shown to the user mitigates any possibility of coupling them with non-conformant data. Moreover, the task of identifying recommended visualizations is reduced to a simple query.

Data integration brings the advantage of jointly visualizing multiple types of data, regardless of their initial differences in terms of schema or format. In that sense, the aforementioned query remained valid. Since semantic data is processed as a graph, only the techniques that belong to the intersection of recommended techniques would be shown in the query results. Therefore, in our example, if one wanted to analyze travel intentions (nearest stop requests) and travel events (ticket validations) simultaneously (see Fig. 3c), the 2D Geographic Heat Map would figure on the list of recommended visualizations.

Finally, as an example of ontology extension, we classified the travel behavior of Andante cards in D1 in terms of journey patterns, based on another study [17]. Such classification considered the concepts of spatial and temporal regularity. For instance, we defined that a passenger with a *Rigid* journey pattern has temporal regularity (the departure time deviation is within 1 hour), and a *Broad* spatial regularity (two primary journeys accounting for at least 74% of all his/her journeys, on average). Examples of frequent journeys made by passengers that follow this pattern are home/work/home or home/school/home.

We then defined the following rule for assigning Andante cards to a *Rigid* journey pattern:

$$\text{hasSpatialRegularity}(?x, \text{Broad}) \wedge \text{hasTemporalRegularity}(?x, \text{true}) \Rightarrow \text{hasJourneyPattern}(?x, \text{Rigid})$$

The evaluation of spatial and temporal regularity patterns (using *hasSpatialRegularity()* and *hasTemporalRegularity()*) required the specification of semantic queries, in the same rationale we used for constructing the graph of all instances that were suitable for the visualization technique. Fig. 4 schematizes the automatic inference of the journey pattern rule applied to a specific ticket.

Given that all properties defined in the ontology can be assigned to a domain (in the mathematical sense), it is possible for a visualization technique to show to the user which properties can be used as a data filter, instead of restricting the specification of filters into visualizations themselves. In the case of *hasJourneyPattern()*, we asserted that the domain of this property is *TicketValidation*. Hence, a visualization tool could show to the user all properties that can act as a filter by listing the ones that have *TicketValidation* as domain:

```
SELECT ?propertyName
WHERE { ?propertyName rdf:type owl:ObjectProperty
        ?propertyName rdfs:domain :TicketValidation . }
```

IV. CONCLUSIONS AND FUTURE PERSPECTIVES

This paper proposed the use of ontologies and SWT as an approach to visualizing urban mobility data, such as public transportation ridership, vehicle flows and the like. The results showed that our approach has considerable advantages, such as enhanced knowledge extraction capabilities due to the provision of semantics to data and visualizations, and better support for data integration from multiple sources, which can eliminate schema dependency and facilitate reproducibility.

Since this work is in ongoing development, future perspectives include extending the ontology with other urban mobility elements for data characterization and the refinement of existing ones, as well as a deeper specification of elements used for characterizing visualization techniques. Furthermore, we expect to formalize our approach by constructing a methodology oriented to the development of semantically rich urban mobility visualizations. We seek to involve analysts continuously for evaluation purposes.

The ontology is expected to be publicly available and documented. We also envisage the development of a proof-of-concept prototype with a set of visualization techniques based on the OBAVUM approach.

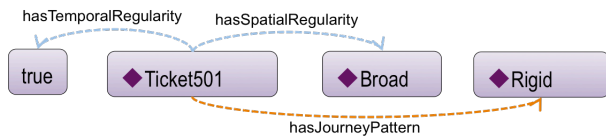


Figure 4. The asserted properties (in blue) for Ticket501 led to the automatic inference (in orange) of its journey pattern.

ACKNOWLEDGMENT

This work is fully supported by the doctoral grant PD/BD/105910/2014, provided by the Portuguese Science and Technology Foundation (FCT).

REFERENCES

- [1] T. Cheng, G. Tanaksaranond, C. Brunsdon, and J. Haworth, "Exploratory visualisation of congestion evolutions on urban transport networks," *Transportation Research Part C: Emerging Technologies*, vol. 36, pp. 296 – 306, 2013.
- [2] G. di Lorenzo et al., "Allaboard: Visual exploration of cellphone mobility data to optimise public transport," in *Proc. 19th Int. Conf. Intelligent User Interfaces (IUI '14)*, New York, NY, USA, pp. 335-340, ACM, 2014.
- [3] E. Polisciuc, A. O. Alves, C. Bento, and P. Machado, "Visualizing urban mobility," in *Special Interest Group Computer Graphics and Interactive Techniques Conf. (SIGGRAPH '13)*, Anaheim, CA, USA, 2013, pp. 115.
- [4] W. Zeng, C.-W. Fu, S. Müller Arisona, A. Erath, and H. Qu, "Visualizing Mobility of Public Transportation System," *IEEE Transactions Visualization and Computer Graphics (VAST'14)*, vol. 20, no. 12, pp. 1833-1842, 2014.
- [5] M. De Marchi, "Visualizing Public Transport Systems: State-of-the-Art and Future Challenges," unpublished. [Online]. Available: https://www.evl.uic.edu/creativecoding/cs524/RESEARCH/DeMarchi_Transit.pdf
- [6] E. D. Valle, I. Celino, and D. Dell'Aglia, "The experience of realizing a semantic web urban computing application," *Transactions in GIS*, vol. 14, no. 2, pp. 163-181, 2010.
- [7] K. M. de Oliveira et al., "Transportation ontology definition and application for the content personalization of user interfaces," *Expert Syst. Appl.*, vol. 40, pp. 3145-3159, 2013.
- [8] J. Wang, Z. Ding, and C. Jiang, "An ontology-based public transport query system," in *1st Int. Conf. Semantics, Knowledge and Grid (SKG '05)*, Beijing, China, pp. 62-64, 2005.
- [9] J. L. Pina, E. Cerezo, F. Seron, "Semantic visualization of 3D urban environments," *Multimedia Tools and Applications*, vol. 59, issue 2, pp. 505-521, 2012.
- [10] J.C. Gomes, N. Montenegro, P. Urbano, J. Duarte, "A Land Use Identification and Visualization Tool Driven by OWL Ontologies," in *12th OWL: Experiences and Directions Workshop*, Crete, Greece, 2012. [Online]. Available: http://ceur-ws.org/Vol-849/paper_5.pdf
- [11] T. Sobral, "Developing visualizations for urban mobility data: a user-centred approach," Master's thesis, Faculty of Engineering, University of Porto, 2014.
- [12] T. Sobral, T. G. Dias, and J. Borges, "Towards a conceptual framework for classifying visualisations of data from urban mobility services," in *Exploring Services Science (H. Nóvoa and M. Dragoicea, eds.)*, vol. 201 of *Lecture Notes in Business Information Processing*, pp. 228-242, Springer International Publishing, 2015.
- [13] W. Aigner, S. Miksch, H. Schumann, C. Tominski, "Visualization of Time-Oriented Data," Springer-Verlag London, 2011.
- [14] Instituto Nacional de Estatística I.P., Portugal – INE (2013). [Online]. Available: <http://www.ine.pt>
- [15] Transportes Intermodais do Porto – TIP (2015). [Online]. Available: <http://www.linhandante.com>
- [16] J.F. Cunha, T. G. Dias, "State of the Art and Future Perspectives for Smart Support Services for Public Transport," in *Service Orientation in Holonic and Multi-Agent Manufacturing and Robotics*, vol. 544, pp. 225-234. Springer International Publishing, 2014.
- [17] V. Costa, T. Fontes, P. M. Costa, T. G. Dias, "Prediction of Journey Destination in Urban Public Transport," in *17th Portuguese Conf. Artificial Intelligence (EPIA 2015)*, Coimbra, Portugal, pp. 169-180.