A Decision Support Tool for Evaluating Big Data Investment in Transport: Empirical evidence from European use cases of NOESIS project

Christos Katrakazas\textsuperscript{a}, Natalia Sobrino\textsuperscript{b,c}, Ilias Trochidis\textsuperscript{d}, Jose Manuel Vassallo\textsuperscript{b}, Stratos Arampatzis\textsuperscript{d}, Constantinos Antoniou\textsuperscript{e}

\textsuperscript{a}National Technical University of Athens, Department of Transportation Planning and Engineering, 15773, Athens, Greece
\textsuperscript{b}Transportation Research Centre – TRANSyT, Universidad Politecnica de Madrid, Madrid 28040, Spain
\textsuperscript{c}Civil Engineering: Construction, Infrastructure and Transport Department, Universidad Politecnica de Madrid, Madrid 28014, Spain
\textsuperscript{d}Ortelio LTD, Coventry University Technology Park, Puma Way, Coventry, CV1 2TT, United Kingdom
\textsuperscript{e}Chair of Transportation Systems Engineering, Technical University of Munich, Munich 80333, Germany

Abstract

Big Data technologies have become extremely popular in transportation applications worldwide. The European Union (EU) Horizon 2020 (H2020) NOESIS project (http://www.noesis-project.eu) aims at improving the understanding about the impact of big data by creating a Decision Support Tool for evaluating big data investment in transport. Towards this aim, key challenges of Big Data utilization in the transportation domain are identified, use cases of applications are recorded and a benefit evaluation of Big Data applications is designed, in order to support future decisions. These tasks form the scope of this research. Initially, 13 areas of focus for Big Data in Transport are presented, followed by the first library of recorded Big Data use cases, along with a Data Benefit Analysis methodology and a decision support tool. As a result, a first holistic approach towards exploiting the socioeconomic impact of transportation investments using Big Data is formed.

Keywords: Big data, Transport, Decision support tool, Data Benefit Analysis
1. Introduction - The NOESIS Project

The NOESIS project aims at identifying the critical factors/features, which lead to successful implementation of Big Data technologies and services in the field of transport and logistics with significant value generation from a socioeconomic viewpoint. This is achieved through the examination of areas and contexts throughout Europe, in which ICT investments and exploitation of data should be implemented. The impacts of Big Data are evaluated in a series of transportation use cases (Big Data in Transport Library) by developing and applying a “Learning framework” (Decision Support Tool) and a Value Capture mechanism, which will estimate the expected benefits and costs (Data Benefit Analysis).

The main outcome of NOESIS is a Decision Support tool, which will be able to assess the value generated (i.e. socioeconomic impact) from Big Data investments, taking as input the specific characteristics and contextual information of the transport system under evaluation and associating it with a predefined set of use cases with similar characteristics by employing a number of big data (machine learning) techniques.

2. Big Data and Emerging Transport Challenges

In the last years, many Big Data technologies have been applied to the transportation sector all over the world, which has created a wide range of challenges and opportunities in the field of transport. The definition of major transportation areas and sub-problems that could benefit by Big Data are of interest for the transportation community. For this reason, one of the main tasks of the NOESIS project was to identify the major transport-related challenges, with regards to Big Data use (Katrakazas et al. 2019). The literature review that was conducted was useful to categorize each big data in transport use case of the Big Data in Transport Library (BDTL), based on the challenges the solution faced to. The literature review led to the identification 10 major challenges: i) Environment, ii) Connected and Autonomous Vehicles (CAVs), iii) Road Safety, iv) Traffic Management, v) Transport planning, vi) Freight and Logistics, vii) Aviation, viii) Railways, ix) Cost-Effectiveness and x) Data-related issues. After the initial derivation of the challenges along with the corresponding sub-problems within each challenge from the state-of-the-art in the literature, the project partners, as well as external experts on Big Data, provided feedback and reviewed the list of challenges. After the final validation, the updated challenges list contained 13 challenges, which covered the majority of the transportation domain spectrum with regards to Big Data usage:

- Environment and health
- Automation
- Safety and security
- Transport management and operation
- Transport policy and planning
- Freight and logistics
- Integration (MaaS)
- Funding, Financing, Cost efficiency
- Social / Psychological aspects
- Data related
- Quality of service
- Resilience
- Maintenance

3. The Big Data in Transport Library

One of the main outcomes of the NOESIS project is the development of the first collection of Big Data use cases in Transport, which form the BDTL. The BDTL constitutes a reference point as for the first time Transport Challenges have been associated with datasets, applications, along with their potential value. The BDTL has been developed as an open and easily accessible website, which serves as a knowledge hub where use cases and relevant findings are stored and can be accessed by all users.

In its final form, the NOESIS Big Data in Transport Library will consist of three different elements:
- Standardized descriptions of a collection of transport and logistics systems’ use cases including all contextual information about the transport system under study (area, population, transport mode, etc.)
- A list of Big Data products, which are matched with specific transport and logistics use cases. It is noteworthy that each NOESIS use case can be matched with many Data products, which correspond to alternative potential combinations of Big Data technologies, which are or could be applied in the use case.
- The expected value generated (“label”) for each pair of transport use case and associated Big Data product, the so-called “Transport Use Case-Big Data product” pair. The socio-economic value generated is the outcome of the Data Benefit Analysis methodology.

Currently, the BDTL consists of 85 validated use cases of Big Data usage for transportation problems, covering the majority of transport modes, as well as both the freight and passenger sectors. For each of the use cases, information is provided with regards to the challenges, the data characteristics, operating and investment cost details, as well as the reported socioeconomical benefits. A screenshot of the BDTL current form and the viewable information of each use case is displayed in Figure 1.

Figure 1. The NOESIS BDTL

4. Data Benefits Analysis (DBA)

Transportation projects often require large initial investments, which result in benefits for the society. Driven by the information revolution, a recent category of transport project is that of big data analytics and technologies applied in transport systems and network. Similar to other traditional transport project, big data innovations in
transport make sense if they improve the quality of life for society, the so-called benefits for society. In this line, the Data Benefit Analysis (DBA) aims at identifying and quantifying the value that Big Data applications may provide to the company that promoted it and to society as a whole. This methodology identifies a set of indicators which constitute a measure (in qualitative terms) of how large or small the impacts of a Big Data application will be in a given transport use case. The proposed DBA is applied to the NOESIS BDTL use cases. The results of the DBA in terms of value generation constitute the core of the Decision Support Tool, which is described in Section 5. Therefore, the DBA constitutes an identification and assessment of benefits of using big data solutions in transport.

4.1. Identification of big data solutions in transport benefits

The first step of the DBA is to identify a set of indicators, which constitute a measure (in qualitative terms) of how large or small the impacts of a big data solutions will be in a given transport use case.

On the one hand, one of the impacts of big data solutions in transport are the costs that the organization should be faced to: investment costs or capital costs and operational and maintenance costs of the big data solution. On the other hand, Table 1 provides an overview of key transport-specific benefits deriving from Big Data solutions from the two viewpoints: (i) the organization and (ii) the user/society as a whole. For instance, gathering insights on customer needs and feedback (through, for example, social media, or MaaS application) allows for service improvements while satisfying the increasing demand of transport users (Hill et al. 2017).

| Benefits or profits for the organization which promotes the big data solution | Benefits for users and society |
|---|---|
| Life-cycle cost savings | Travel time savings |
| To generate more revenue | To improve information provision to clients/users |
| To better known clients | To improve quality of service |
| To improve Know-How: scalability and generalization to other solutions | To improve planning and coordination across stakeholders |
| To improve supervision and surveillance | To improve reliability |
| | To improve safety and security |
| | To reduce environmental impacts |
| | Job creation |

The measurement of these costs and benefits for each of the big data in transport use cases are based on qualitative terms of: nothing, low, medium, and high. The proposed DBA is applied to the NOESIS use cases of the Big Data in Transport Library. The results of the DBA in terms of value generation constitutes the core of the Decision Support Tool. Results from DST will help decisions makers of the application of big data solution in transport.

5. NOESIS Decision Support Tool

As discussed earlier, the impact of Big Data is evaluated within the NOESIS project by analysing a series of transportation use cases. Based on the data created during the implementation of the BDTL, the NOESIS Decision Support Tool (DST) has been developed (Figure 2). The DST is able to predict the socio-economic value generated from Big Data investments, taking as input the specific characteristics of the transport system under evaluation and associating it with a predefined set of use cases with similar characteristics by employing a number of machine learning techniques.

The first step towards developing the Decision Support Tool was to analyze the 85 use cases and to identify the most critical factors/features. The question was which features should be used to create a predictive model. This is a difficult question, which requires deep knowledge of the problem domain. These features have been selected by applying machine learning techniques and by talking with experts in the field. The features that have been selected are: Transport mode, Transport sector, Type of data collected, Sample size of data, Operating costs, Investment costs, Transport challenges.
After the selection of the key features, the DST identifies the underlying patterns of “successful” applications of Big Data in transport. This is a “learning” process directly related to pattern recognition and machine learning. So far, we have trained the NOESIS DST with 85 use cases. The user inputs the following selected critical features: Transport mode, Transport sector, Type of data collected, Sample size of data, OPEX, CAPEX, Transport challenges. After completing the above information, the user receives as output the potential benefits for the organization and for the society, as defined in Table 2. Clustering with PCA and K-Means is then used to identify clusters in NOESIS BDTL and to identify the cluster that the potential use case belongs to.

![NOESIS Decision Support Tool](http://noesis-project.eu/toolbox/decision.php)

6. Results

6.1. Description of the BDTL

Figure 3 gives an overview of the mode and sector distribution of the 85 NOESIS use cases.

![NOESIS BDTL cases distribution](image-url)
As it can be observed, the vast majority of the cases are concerned with road transportation, while the combination of road and rail comes second in population.

With regards to their impact, tables 2 and 3, give an overview of the effect of the cases 11 factors (Revenue, Business Opportunities, Improving KnowHow, Better relationship with Clients, enhanced Supervision and Surveillance, Reduced Environmental Impacts, Quality of Service, better Information Provision, Travel Time Savings, Reliability and Job creation) per mode and sector (table 2) and per operating and investment costs (table 3).

Table 2. Most frequent level of effect of mode and sector on the impacts of a use case

| Mode, Sector | Sample Size | OPEX | CAPEX | Revenue | BusinessOp | KnowHow | Clients | Supervision | Surveillance | ReduceEnv | Impacts | QoS | Information provision | TravelTime | Reliability | JobCreation |
|--------------|-------------|------|-------|---------|------------|---------|---------|-------------|--------------|-----------|---------|-----|-----------------------|------------|-------------|-------------|
| Air, Both    | 1,000,000-... | R    | E     | M       | M         | L       | L       | H           | L            | M         | H       | No  | L          | L          | L           | L           |
| Air, Rail, Both | 10,001-100,000 | R    | E     | No     | L         | M       | L       | H           | L            | H         | No     | H    | L          | No         | L           | No          |
| Air, Rail, Road, Maritime Passenger | 100,001-1,000,000 | C    | C     | L       | M         | M       | M       | L           | M            | M         | H       | No  | L          | No         | L           | No          |
| Air, Rail, Road, Maritime, Inland waterway Passenger | 1,000,000-... | C    | R     | L       | H         | H       | H       | L           | No           | L         | H       | L    | No         | No         | No          | No          |
| Maritime Both | 100,001-1,000,000 | R    | E     | L       | L         | M       | M       | H           | M            | M         | M       | M    | L          | M          | M           | L           |
| Maritime, Freight | 100,001-1,000,000 | R    | R     | No     | L         | No      | L       | No          | L            | H         | M       | No  | L          | No         | No          | No          |
| Maritime, Inland waterways Both | 100,001-1,000,000 | R    | E     | No     | No        | H       | M       | H           | H            | M         | H       | M    | L          | No         | No          | L           |
| Rail Both | 100,001-1,000,000 | R    | E     | L       | No        | M       | M       | M           | M            | No        | M       | M    | L          | M          | M           | L           |
| Rail, Freight | 10,001-100,000 | R    | R     | M       | M         | H       | H       | M           | L            | H         | No     | L    | No         | No         | No          | No          |
| Rail, Passenger | 1,000,000-... | C    | R     | H       | H         | M       | H       | L           | H            | L         | M       | No  | L          | No         | No          | No          |
| Rail, Road Both | 1,000,000-... | R    | E     | No     | L         | M       | H       | L           | H            | M         | L       | L    | L          | L          | L           | L           |
| Rail, Road, Passenger | 1,000,000-... | R    | R     | L       | M         | H       | H       | L           | H            | M         | No     | L    | No         | No         | No          | L           |
| Rail, Road, Inland waterways Passenger | 0-1,000 | R    | R     | L       | No        | No      | H       | L           | H            | M         | M       | M    | M          | L          | L           | L           |
| Road Both | 1,000,000-... | R    | R     | M       | M         | H       | M       | H           | H            | H         | M       | L    | M          | M          | L           | L           |
| Road, Freight | 1,000,000-... | R    | R     | No     | H         | M       | H       | M           | H            | H         | L       | No  | L          | No         | No          | No          |
| Road, Passenger | 1,000,000-... | R    | R     | M       | H         | H       | H       | M           | H            | H         | L       | No  | L          | No         | L           | No          |
| Road, Maritime Freight | 10,001-100,000 | R    | E     | M       | M         | H       | L       | No          | H            | H         | No     | M    | L          | No         | No          | No          |
| Road, Rail Both | 100,001-1,000,000 | R    | R     | M       | L         | M       | H       | L           | M            | M         | M       | H    | L          | L          | L           | L           |
| Road, Rail Passenger | 1,000,000-... | R    | R     | M       | L         | No      | H       | H           | M            | H         | H       | H    | H          | L          | L           | L           |
Table 3. Most frequent level of effect of operating and investment cost on the impacts of a use case  
(N: None, L: Low, M: Medium, H: High)

| OP Costs   | Inv. Costs | Revenue | Business | Opportunity | KnowHow | Clients | Supervision | Environment Impacts | QoS | Information Provision | Travel Time | Reliability | JobCreation |
|------------|------------|---------|----------|-------------|---------|---------|--------------|----------------------|-----|----------------------|-------------|-------------|-------------|
| Cheap      | Cheap      | No      | No       | H           | M       | H       | L            | H         | H     | H       | H           | M            | No          |
| Cheap      | Reasonable | L       | H        | M           | H       | H       | L            | H         | H     | L       | H           | L            | L           |
| Expensive  | Expensive  | H       | H        | M           | L       | M       | M            | H         | H     | H       | H           | H            | H           |
| Expensive  | Reasonable | L       | M        | H           | H       | H       | L            | H         | L     | L       | L           | L            | L           |
| Reasonable | Cheap      | M       | H        | H           | M       | H       | M            | H         | H     | M       | L           | L            | L           |
| Reasonable | Expensive  | No      | L        | M           | H       | M       | M            | H         | H     | L       | H           | L            | L           |
| Reasonable | Reasonable | M       | H        | H           | H       | M       | H            | H         | H     | H       | H           | No           | L           |

From Table 2, it is demonstrated that the majority of the modes and sectors operate on cheap and reasonable costs with regards to their Big Data usage, with the exception being Air transportation, Maritime, Rail with a large sample size, the combination of Road and Maritime transportation for freight. With regards to the results according to the DST, it is observed that the highest positive impacts comes the enhancement of supervision and surveillance as well as the improvement of the Quality of Service and the provision of information to the customers. On the other hand, the lowest impact is provided to job creation, where among all modes and sectors the impact is negligible or very low.

With regards to the distinction of use cases per operating and investment costs as seen in Table 3, it is further distinguished that Supervision and Surveillance, Quality of Service and Information provision are significantly enhanced by utilizing Big Data regardless of the investment and operating costs. When both the investment and operating costs are reasonable the impact is usually higher, which is also evident when large amounts of money are invested (i.e. when operating costs and investment costs are both expensive).

6.2. Results from the Decision Support Tool

This subsection will present an example of using the NOESIS DST for making informed decisions on Big Data investments. The inputs for the DST are displayed in Table 4:

| Transport Mode | Road |
|----------------|------|
| Transport sector | Passenger |
| Type of Data | Location data |
| Sample size | 10,000 - 100,000 |
| OPEX | Reasonable |
| CAPEX | Reasonable |
| Transport Challenges | Transport Management and operations |
| Quality of service | |

These were sequentially changed for the following scenarios: i) increasing the sample size to 1.000,000 data points, ii) adding datasets containing new information, iii) adding social media data, iv) increasing the investment costs and v) increasing the operating costs. The base scenario in the DST and the potential impact on the use case are displayed in Figure 3.
Figure 4: Impact of characteristics on the impact of a use case using the DST (Scenarios from top left to bottom right: Base, i, ii, iii, iv, vi)

As can be seen from Figure 4, increasing the sample size of data has a positive effect on enhancing know-how and providing new business opportunities, while also improving reliability and saving travel time. However, new datasets do not affect the expected benefits of specific Big Data use case, except from the case of social media data, where new business opportunities arise and there is more information exchange between the business and the clients or users. A use case with more expensive investment costs, will lead to improved know-how and more business opportunities, but will lead to less job creation and less travel time savings. Similarly, increasing operating costs, also leads to less jobs and travel times, but also decreases business opportunities for the stakeholder.

7. Conclusions

The first version of the NOESIS Decision Support Tool has been trained based on a set of 85 use cases. We keep on training our algorithms with more datasets in order to enrich the learning experience of our tool. The preliminary results indicate that NOESIS Decision Support Tool can be used by policy makers, transport authorities, transport operators and businesses as a prescreening tool in order to understand the benefits that their solution would have in the society and in the organization. The final version of the NOESIS Decision Support Tool would also provide some further recommendations to experts in relation on which attributes they should change to increase specific benefits and examples of related use cases.

The BDTL, the DBA, and the decision support tool of NOESIS form the three pillars of the project. It is envisioned that these three pillars will consolidate knowledge on big data methodological and exploitational issues (i.e. with the BDTL), will assist understanding and predicting the potential value and benefits generated from big data applications in transport (i.e. with the Data Benefit Analysis), and will enhance evidence-based decision making
(i.e. with the Decision Support tool) for transport authorities, operators and businesses. Hence, a holistic evaluation framework in order to understand and assess the impact of Big Data will be available for the research and industrial community.

Acknowledgements

The research was funded by the EU H2020 NOESIS Project (Project Number: 769980).

References

Hill, N., Gibson, G., Guidorzi, E., Amaral, S., Kumar, P., & Jin, y. (2017) Scoping Study into Deriving Transport Benefits from Big Data and the Internet of Things in Smart Cities. Final Report for Department for Transport, Ricardo/ED10011/Issue Number 4. Retrieved from: https://www.gov.uk/government/publications/transport-benefits-from-big-data-and-the-internet-of-things-in-smart-cities

Katrakazas, C., Antoniou, C., Sobrino Vazquez, N., Trochidis, I., Arampatzis, S. (2019) Big Data and Emerging Transportation Challenges: Findings from the NOESIS project, IEEE 6th International Conference on Models and Technologies for Intelligent Transportation Systems, Krakow, Poland.