SYNTHESIZING NATURALISTIC DRIVING DATA: A FURTHER REVIEW

K VENTER and K MURONGA

CSIR Built Environment, Transport Systems and Operations, PO Box 321, Pretoria, 0001, Tel: 012 841 3856, kventer@csir.co.za, *Tel: 012 841 2337, kmuronga@csir.co.za.

ABSTRACT

The Naturalistic Driving Study (NDS) methodology has in the past decade proven extremely valuable in providing rich contextual information about the driver, the vehicle and driving environment. Internationally the uptake of the methodology is growing and especially more developed countries are employing NDS on larger and grander scales. As the methodology evolves new data challenges necessitate the development of novel approaches to manage and analyse the data.

The NDS methodology has been applied twice within the South African context. Large amounts of quantitative and qualitative data were collected and different application software products are currently used to transcribe and analyse the data. The process is extremely resource intensive and working with the data remains a learning curve. Recommendations were put forward in an earlier study toward the management and integration of these Very Large Databases in order to simplify the analyses of the data. This paper provides feedback in terms of the progress made with the implementation of the recommendations as applied in two new investigations that made use of the previously collected material. The findings though indicate that the battle is far from over and concludes with a review of additional strategies and further recommendations for developing an approach to work with these data and databases.

1. INTRODUCTION

1.1. Background

Naturalistic Driving Studies (NDS) have since the original Strategic Highway Research Programme (SHRP) or 100-car study in 2005 been successfully employed to investigate various types of driver behaviour within the context of the vehicle and environment. Countries such as the United States (Stutts et al., 2005) European Union (Sagberg et al., 2011) as well as Australia (Regan et al., 2013) are undertaking increasingly larger and diverse NDS studies. The United States Second Strategic Highway Programme (SHRP2) is the largest study of its kind ever undertaken with approximately 2360 participants in 2012 (Campbell, 2012). The European Naturalistic Driving Study (UDRIVE) currently collects data from two hundred participants (Horizon 2020, 2016) in seven European Union (EU) member

states while in Australia and New Zealand four-hundred vehicles have been fitted with these data collection systems (Regan et al., 2013).

Countries conducting these studies on a large scale are investing considerable resources in terms of equipment, skills and time into the projects (Regan et al., 2013). However, the benefits for these countries in terms of making inputs into interventions that reduces the number of serious and fatal crashes seems to outweigh the significant costs associated with the layout and implementation of these projects (Sagberg et al., 2011; Regan et al., 2013). Bärgman (2015) states that Naturalistic Driving Data (NDD) is essential for informing the design of measures such as in-vehicle technology or safer infrastructure that has the potential to reduce crashes and associated casualties significantly.

The worth and the effectiveness of the methodology are accentuated when seeing the range and variety of topics under investigation. The methodology has been expanded from studying light motor vehicle drivers to include heavy vehicle drivers (Barr et al., 2003), commercial vehicle drivers (Wiegand et al., 2009) as well as driver behaviour from special road user groups such as novice drivers (Lee et al., 2011), older drivers (Blanchard et al., 2010) and motorcyclists (Uchida et al., 2010). More recently the methodology has been extended to investigate into the interaction between motorised and non-motorised transport (NMT) users (Twisk et al., 2012).

1.2. Overview of NDS approach

The methodology is technically possible due to advances in information and communication technologies, improvements in storage capacities, data-mining, image processing and increasingly low-cost camera technology (Eenink et al., 2014). Typically, a vehicle is equipped with a Data Acquisition System (DAS) device that continuously capture information about vehicle movements (acceleration, deceleration, position on the road, driving speed), the driver (position, eye, head and hand movements, and other in-vehicle activities) as well as environmental information from the surrounding area (traffic densities, time headway, following distance, road and weather conditions). The methodology offers insight and detail that traditional methodologies cannot capture (Williamson et al., 2015). In this lies the significance because all behaviour is observed, including normal driver behaviour as well as behaviour preceding a crash or near-crash (Stutts et al., 2005; Regan et al., 2013).

In the local studies NDD was collected through an in-vehicle DAS comprising of inward and outward facing cameras, Global Positioning System (GPS) and an on-board computer that records date, time, speed, acceleration and deceleration (Muronga and Venter, 2014).

1.3. South African topics and applications

The NDS methodology has previously been employed to conduct two experiments during the years 2011 and 2013. In 2011, NDD was collected for a pilot project from a company driver and the purpose of the experiment was to explore the usefulness of the methodology in a South African context (Venter, 2012). The results were favourable leading to a second study in 2013 where four primary drivers participated

in an experiment to investigate the differences between novice and experience drivers in South Africa. This experiment collected data over a six month period. The focus of the research was to investigate the differences in scanning behaviour at selected pre-identified locations (stop streets, traffic circles and traffic lights) in the road environment (Venter and Sinclair, 2014).

More recently these data were again reanalysed to include an investigation as to whether or not drivers perceive non-motorised transport (NMT) users as hazards in the road environment. A fourth study reviewed new material not previously analysed, considering driver distraction and inattentiveness (Road Traffic Management Corporation, 2015).

1.4. The data challenge

Although this approach is considered one of the most promising methodologies to provide simultaneous insight into the driver, vehicle and environment interactions, the methodology is costly and resource intensive (Backer-Grondähl, 2009; Regan et al., 2013; Williamson, 2015). The challenges experienced with the South African data collection process, preparation, storage and analysis have previously been highlighted (Muronga et al., 2014).

Large amounts of qualitative and quantitative data were collected and a key challenge was the integration of the different formats as well as the management and analysis of the data (Muronga et al., 2014). These challenges facilitated a review of international practices in an attempt to make recommendations for the development of a NDD management strategy for the South African context that addresses:

- The processes for collecting, storing and managing the large volumes of different types of data;
- Address the integration of different data formats which currently makes use of multiple programmes and processes for analysis;
- Automation in order to reduce the amount of time it took to standardise the data and to prepare it for eventual analysis.

2. PURPOSE OF THIS PAPER

The mining of the collected data is ongoing and new research topics have been explored using the existing dataset. The purpose of this paper is to provide feedback as to how, in the context of the new research topics, the recommendations from the previous 2014 study have been implemented and applied.

3. TOWARDS A NDD MANAGEMENT STRATEGY

3.1. Overview

The previous recommendations pertaining to the management of Very Large Databases (VLDB) included hardware upgrades and approaches for data integration. Data integration strategies included data transformation for more efficient analysis, typology development where the classification scheme can be reapplied to new research or data consolidations to create new variables for use in further analysis (Muronga et al., 2014). In terms of automating the process of “merging” and “matching” the corresponding quantitative and qualitative data suggestions included algorithms for organising and indexing, searching and highlighting unique characteristics with pattern and object recognition was the most promising (Muronga et al., 2014).

3.2. Hardware upgrades

In accordance with the recommendations the hardware was upgraded and attempts were made to automate the data management process (Muronga et al., 2014). Hardware upgrades included improving the computers processing speed as well as making available more internal and external hard drive capacity.

3.3. Attempt to automate the process

Previously the analysis used the qualitative data (image material) as a departure point where videos were watched and coded in-vivo (Venter and Sinclair, 2014). The qualitative data set was then matched with the corresponding quantitative dataset. This was, however, time consuming and resource intensive (Muronga et al., 2014). In an attempt to speed the process up, the thought was to tackle the problem from a quantitative perspective by developing a search strategy for interrogating the quantitative information. The quantitative information comprise of hundreds of thousands of entries logged for vehicle movements. These values include GPS coordinates, date, time, speed, acceleration and deceleration.

In the driver distraction study, the literature review highlighted pre-identified vehicle parameters that are internationally accepted and associated with distracted and inattentive driving. These parameters included, for example, deviations from the posted speed limit, mostly driving slower (Leung et al., 2012; Burns et al., 2002) as well as deviations in lateral acceleration (greater or equal to 0.7 g) or longitudinal acceleration deviations equal or greater to 0.6g (Klauer et al., 2006). Use was made of these parameters to interrogate the quantitative dataset in an attempt to isolate these specific situations with the endeavour to pin pointing and narrowing the search for behaviour in the qualitative data more rapidly and accurately (Perez and Hankey, 2013). This attempt to automate the process was, however, not successful as the predefined values could not be identified in the datasets.

As a result of this problem the researchers again had to revert to the use of the qualitative datasets with the exception that this time a predefined coding and classification scheme was applied instead of in-vivo coding. The image material (videos) was transcribed (using the BX4000 system into .avi files at eight frames per

second) and matched with the log files containing vehicle information from each corresponding video into Microsoft Excel files for quantitative analysis.

Despite the vehicle parameter search not providing the expected evidence of vehicle deviations associated with particular driver behaviour, the qualitative analysis proved that there were (a) indications of distracted and inattentive driving and (b) that although drivers recognise NMT users in the road environment, recognition was inadequate and infrequent.

4. BACK TO THE DRAWING BOARD

4.1. Overview

The search to simplify the data analysis process is therefore on-going and this section provides an overview of new and emerging technologies that can be used for the management and interrogation of VLDBs. In terms of speeding up the interrogation process, machine learning is considered a key strategy for especially qualitative data interrogation. The convergence of qualitative and quantitative methodologies remains problematic but strategies to overcome this challenge include “chunking”, creation of narratives as well as working towards a harmonised set of definitions and standardised software for NDD analysis.

4.2. International experience with the data challenges

Similar data challenges are experienced by international research teams. In preparation for the EU UDRIVE Study, Backer-Grondähl et al. (2009:15) highlighted “the build-up of a large database that includes measures of an enormous amount of variables that are relevant for investigating various traffic safety issues” as a concern. Even after inception of the UDRIVE study, Saint Pierre (2014) emphasises that the “massive” heterogenous data, and the manual annotation of video material remains problematic. Williamson et al. (2015) highlight that the Australian NDS study (ANDS) data set take-up a lot of space differs in format and need different approaches and programmes to download, transcribe and analyse the data. In addition to the data challenges, the skills to “search for a needle in a haystack” need to be developed (Saint Pierre, 2014).

4.3. Emerging methodologies for managing NDD and VLDBS

4.3.1. Relational vs. Non-Relational databases

In the past, most organisations relied on relational databases as they are very good in transaction management and querying capabilities, but these databases are not able to store and process big data effectively (Abramova & Bernardino, 2013). To be able to store and process big data in VLDBs organisations are increasingly considering alternatives to legacy relational infrastructure. Organisation such as Facebook, Google and Amazon considered NoSQL databases for their needs to handle large volumes of data. NoSQL databases are non-relational databases designed to handle large scale of data storage and allow for data processing on large number of servers (Moniruzzaman & Hossain, 2013).

The difference between relational databases and NoSQL is that relational databases make use of structured query language “SQL” to access and manipulate data. This is possible as all tables on the databases are related and data is organised in columns and rows, whilst with NoSQL, meaning “Not Only SQL” databases, the data does not have to be related and does not make use of a SQL interface and databases are not built primarily on tables. NoSQL databases can manage large volumes of data for analytical processing and offer computational storage of applications such as those for big data analytics, business intelligence and social networking.

Making use of NoSQL databases to manage NDD can assist researchers in solving the problems experienced in terms of the volumes of data that are increasing at a very fast rate and not easy to analyse, especially video images. Dealing with the image analysis can be a demanding task not only in terms of computational complexity, but also in terms of the researcher’s knowledge (Lakovidis and Diamantis, 2014). It requires knowledge of application programming interfaces (APIs) and some theoretical knowledge of digital signal processing and data mining. NoSQL databases allows for multiple researchers with various skills to work on the same project, as it applies the agile methodology and not the legendary waterfall method.

A recommendation would be to employ software programmers on NDS projects to assist with customisation of the data processing tools to meet the requirements of the project, as well as developing tools such as dashboards, to view the data.

4.3.2. Search strategies

Manipulation of large and diverse datasets is not new and a pre-requisite is to understand the contents of the different data sets and how they compare to each other (El-Geresy and Abdelmoty, 2001). The idea put forward by El-Geresy et al. (2001) is in line with recent research into machine learning where spatial representations are coded within the image data sets using adjacency and orientation of the relationships between objects in the image sets.

Currently image material is selected and the whole image set is observed for the desired interaction or behaviour (Jovanis, 2013). If the NDD analysis continues to be more effective through the qualitative data analysis, consideration should be given to creating spatial representation between, for example the driver and objects such as NMT users outside of the vehicle to facilitate new search strategies within the qualitative data.

4.3.3. Convergence of methodologies

The current research approach combines different qualitative and quantitative methodologies to analyse the data. According to Angell (2014) the convergence of methods are difficult especially as with NDS, the research topic spans the full complexity of human behaviour and choice along with occurring in a larger complex setting such as the road and traffic environment. When using different methodologies to analyse and interpret the data, care should be taken to consider the similarities

and differences of these approaches, especially when developing new thinking frameworks (Angell, 2014).

Examples of new thinking frameworks for analysis of NDD include chunking, creation of narratives as well as applying both quantitative and qualitative coding schemes to the data sets (Bärgman, 2015). Chunking refers to creation of aggregate measures of continuous NDD across trips or entire driving conditions (Bärgman, 2015). Previously analysis was done by selecting image data for a specific time period (e.g. Driving week 1) and the whole clip is reviewed and coded from beginning to end. The original reasoning was that by observing whole time periods it will be possible to observe behaviour change over time. Chunking however separates datasets in equivalent subsets which is useful to understand specific indicators such as prevalence of a specific behaviour across the whole data set instead of segments. Bärgman (2015) as part of a modified approach for the Driver Reliability and Error Analysis Method (DREAM), created narratives which is similar to previous approaches by the research team, where in-vivo coding were used to describe events in the data using grounded theory to develop new insights related to the driver behaviour (Venter et al., 2014).

5. CONCLUSION

As indicated earlier, the value of the methodology lies in the rich contextual information that can be mined and analysed for years to come. Despite it not yet being possible to significantly improve the data interrogation or analysis process, each project and process brings new insight and possibilities for working toward developing new methodologies to standardise the collation and analyses processes.

The convergence of methodologies for the South African NDD analysis remains problematic, but the process is evolving slowly allowing for a synthesis of approaches which could contribute to the development of new frameworks and conceptual structures. One approach currently being explored is the use of Grounded Theory to develop coding schemes applicable to the South African context in which the coding of driver behaviour takes place. Grounded Theory is a qualitative approach to synthesising the data and traditionally not associated with hard sciences such as engineering. However Grounded Theory might be a useful tool to construct and develop new thinking frameworks applicable to the South African driving context.

Despite the challenges that have been experienced in working and managing these VLDBs, it has not only provided new insight into topic specific South African driver behaviour but has contributed to new thinking regarding managing the data and methodologies for analysis of the NDD. Progress might be slow, but each new project is a learning curve which provides an opportunity to develop new thinking frameworks and methodologies that could potentially expedite the analysis and results.

This NDD set is small in comparison to international data sets. Due to limited resources, only a fraction of the collected data has been interrogated. This data still holds great potential for understanding and interpreting specific aspects of driver and other road user behaviour within the South African road and traffic environment.

Efforts to expand the use of the NDS methodology on a much larger scale is ongoing and plans for a large representative national study on the way.

The benefits of investing in the NDS methodology is undeniable not only in terms of the advantages that it could have towards informing topic specific road safety strategy and interventions in the country but also in terms of the fact that with such advances comes the opportunity for South Africa to contribute to a larger international body of research, putting South Africa on the international road safety map. However, teams with multiple skills and capacity to manage, search and analyse the data in order to extract intelligence from such a study is essential.

REFERENCES

Abramova, V. and Bernardino, J, 2013, NoSQL databases: MongoDB vs cassandra', Proceedings of the International C* Conference on Computer Science and Software Engineering, ACM, pp. 14-22.

Angell, L.S, 2014, 'An Opportunity for Convergence? Understanding the Prevalence and Risk of Distracted Driving Through the Use of Crash Databases, Crash Investigations, and Other Approaches', Engaged Driving Symposium, Annals of Advances in Automotive Medicine, pp. 40-59.

Backer-Grondähl, A., Phillips, R., Sagberg, F., Touliou, K., & Gatscha, M, 2009, 'Topics and applications of previous and current naturalistic driving studies. Deliverable D1.1. PROLOGUE project', European Commission, Brussels, Belgium.

Bärgman, J, 2015, 'On the analysis of naturalistic driving data: Development and Evaluation of Methods for Analysis of Naturalistic Driving Data from a Variety of Data Sources', Department of Applied Mechanics, Chalmers University, Gothenburg, Sweden.

Barr, L.C. Yang, C.Y.D. and Ranney, T.A, 2003, 'An exploratory analysis of truck driver distraction', Transportation Research Board. Washington, D.C., pp. 1-21.

Blanchard, R.A., Myers, A.M and Porter, M.M, 2010, 'Correspondence between self-reported and objective measures of driving exposure and patterns in older drivers', Accident Analysis and Prevention,42, pp. 523-529.

Campbell, K., 2012, 'The SHRP 2 Naturalistic Driving Study: Addressing Driver Performance and Behavior in Traffic Safety', Transport Research News 282, September–October, pp. 30-35.

Congui, M., Whelan, M., Oxley, J., Charlton, J., D'Elia, A and Muir, C, 2008, 'Child pedestrians: Factors associated with the ability to cross roads safety and development of a training package', Monash University Accident Research Centre, Victoria Australia.

Eenink, R., Barnard,Y., Baumann, M., Augros, X., Utesch, F, 2014 'UDRIVE: the European naturalistic driving study', Transport Research Arena, Paris, pp. 1-10.

El-Geresy, B.A. and Abdelmoty, A.I, 2001, 'Qualitative Representations in Large Spatial Databases', Vision and pattern Recognition, IEEE, San Diego, United States of America, pp. 68-75.

Horizon 2020, 2016, 'All-natural driving study hits the road April 18, accessed on 20 April 2016 from https://ec.europa.eu/programmes/horizon2020/en/news/all-natural-driving-study-hits-road.

Iakovidis, D.K. and Diamantis, D, 2014, 'Open-Access Framework for Efficient Object-Oriented Development of Video Analysis Software', Journal of Software Engineering and Applications.

Klauer, S.G., Dingus, T. A., Neale, V. L., Sudweeks, J.D., and Ramsey, D.J, 2006, 'The Impact of Driver Inattention on Near-Crash/Crash Risk: An Analysis Using the 100-Car Naturalistic Driving Study Data', Final report, National Highway Traffic Safety Administration, Washington D.C.

Muronga, K and Venter, K, 2014, Naturalistic driving data: managing and working with large databases for road and traffic management research', Proceedings of the 33rd Southern African Transport Conference 7-11 July, Pretoria, South Africa, pp. 567-574.

Moniruzzaman, A. and Hossain, S.A, 2013, 'Nosql database: New era of databases for big data analytics-classification, characteristics and comparison', arXiv preprint arXiv: 1307.0191.

Perez, M.A. and Hankey, J.M, 2013, 'Distraction Index Framework', Final Report, National Surface Transportation Safety Centre for Excellence Blacksburg, Virginia.

Regan, M.A., Williamson, A., Grzebieta, R., Charlton, J., Lenne, M., Watson, B., Haworth, N., Rakotonirainy, A.,Woolley, J., Anderson, R., Senserrick, T., & Young, K, 2013, 'The Australian 400-car Naturalistic Driving Study: Innovation in road safety research and policy', Proceedings of the Australasian Road Safety Research, Policing & Education Conference 28th – 30th August,. Brisbane, Queensland, pp. 1-13.

Road Traffic Management Corporation, 2015', 'Inattentive and Distracted Driving', Road Traffic Management Corporation, Pretoria.

Sagberg, F., Eenink, R., Hoedemaeker., M., Lotan, T., Van Nes, N., Smokers, R., Welsch, R., and Winkelbauer, M, 2011, 'Recommendations for a large-scale European natralistic Driving Study Deliverable D4.1', TØI Institute of Transport Economics, Oslo, Norway.

Saint Pierre, G, 2014, 'From FoT to NDS: Recent developments in UDRIVE - The first large-scale European Naturalistic Driving Study', French Institute of Science and Technology for Transport, Development and Networks , 2014. 1-19.

Toledo, T and Lotan, T, 2006, 'In-Vehicle Data Recorder for Evaluation of Driving Behaviour and Safety', Transportation Record 1953, pp. 112-119.

Twisk, D.A.M., Van Nes, N., and Haupt, J, 2012, 'Understanding safety critical interactions between bicycles and motor vehicles in Europe by means of Naturalistic Driving techniques', Proceedings, International Cycling Safety Conference, 7-8 November, Helmond, The Netherlands, pp. 1-6 .

Uchida, N., Kawakoshi, M., Tagawa, T., and Mochida, T, 2010, 'An investigation of factors contributing to major crash types in Japan based on naturalistic data', IATSS 34, pp 22-30.

Venter K. and Sinclair, M, 2014, 'Exploratory study into South African novice driver behaviour', Proceedings of the 33th South African Transport Conference, Pretoria, pp. 1-10.

Venter, K, 2012, 'Eating, Drinking and uncontrolled steering: A South African exammple of distracted driving', South African Transport Conference, Pretoria. Pretoria, pp. 1-10.

Wiegand, D.M., Hanowski, R.J. and McDonald, S.E, 2009, 'Commercial Drivers' Health: A Naturalistic Study of Body Mass Index, Fatigue, and Involvement in Safety-in Safety critical events', Traffic Injury Prevention 10(6), pp. 573-579.

Williamson, A., Grzebieta, R., Eusebio, J., Zheng, W.Y., Wall, J., Charlton, J.D., Lenné, M., Haley, J., Barnes, B., Rakotonirainy, A., Woolley, J., Senserrick, J., Young, K., Haworth, N., Regan, M., Healy, S., Cavallo, A., Di Stefano, M., Wong, H.L, 2015 'The Australian Naturalistic Driving Study: from beginnings to launch', Proceedings of the Australasian Road Safety Conference 14-16 October, Gold Coast, Australia, pp. 1-7.