A Human Factors Approach to Enhanced Machine Learning in Cars

Joseph SMYTH, Stewart BIRRELL, Lech BIREK, Kris KOBYLINSKI, Alex MOUZAKITIS and Paul JENNINGS

WMG – The University of Warwick, UK
Jaguar Land Rover, Coventry, UK

Abstract. Using machine learning techniques, it is possible to learn and subsequently automate certain driver-focused features in consumer vehicles. A human factors approach is taken to review current machine learning systems. Subsequently, it is found that current methods used for machine learning involve long learning times and might not be sufficient to grasp a true understanding of interaction, routine and feature use - a new method is proposed. Issues surrounding trust and acceptance in automation are also explored and recommendations made.

Keywords. Machine Learning, Automotive, Automation, User routine

1. Introduction

The basic concept of machine learning is for a system to learn about a certain task based on information provided to it and information that it can gather itself - without being explicitly programmed by the user. Originating as a sub-area of artificial intelligence, some applications of machine learning have been evidenced in recent products such as a self-learning watering system (MyBlossom, 2015) and further products such as Smart Lightbulbs (Power-up your lights, 2015) are bringing self-learning features into the home. With all self-learning/machine learning and automated systems the acceptance, desirability and uptake by consumers is heavily dependent on the system's ability to complete a task to the same, or better standard than a user can. Considering a consumer-focused machine learning (ML) system, the way in which it can achieve this relies initially on its capability to access information and learn the desired task(s).

Many automotive manufactures (such as BMW, Jaguar Land Rover, Mercedes) are currently looking to apply ML and autonomous technology to cars each with varying scope. One application of such ML could be to learn a driver's pattern of engagement and use of certain vehicle features and then automate these tasks to mimic a driver's normal feature interaction routine. Eventually, this could take care of many of the non-essential driving tasks - allowing the driver to concentrate solely on the task of driving without having to expend cognitive effort on other aspects of the vehicle - bringing considerable safety and user experience (UX) benefits.

Some examples of current feature-sets found in the literature for such ML systems include (but are not limited to): predictive call list which suggests 'speed-dial' contacts, phone reminder to remind the driver of a forgotten phone, heated/cooled seats that warm/cool the vehicle's seats autonomously, timed climate that auto activates the climate, cabin temperature that autonomously adjusts the cabin temperature, seat massage that activates the in-built seat massage autonomously and media which selects the preferred in-car entertainment source. From reviewing the literature on various ML systems, and with an understanding of the process of learning and automation, three areas of interest for further research were highlighted which form the basis of this paper:

Question 1. (Q1) How can the accuracy of ML systems predictions be improved?

Question 2. (Q2) How can the time taken for the ML system to learn a routine be decreased?
Question 3. (Q3) How can users trust and acceptance of ML systems be managed?

2. Methods

To answer the three highlighted questions, it was necessary to fragment them into sub-questions to ensure effective focus of research. To begin this exploration, two models explaining an example of an ML function were assessed and critiqued:

Figure 1. Generic timeline of Machine learning.

The timeline in Figure 1 shows three stages, consisting of: A learning phase - whereby a system is recording data and 'learning' a user's routine. An activation stage - where a system has enough information to begin automation. Finally, the automated phase - where a system is automating the feature. The grey area highlighted shows the human factors interest in this project – the conclusion of the learning phase and the transition into automation. Where it is shown how the activation point is directly related to, and solely dependent on the effectiveness of the learning phase and therefore depends entirely on the system's ability to understand the user's routine (i.e. learning). The learning phase in typical machine learning systems can be invisible to the user, where some current systems have no communication with the user until activation is reached. Considering Figure 1 further questions were found:

- What can be done in the 'learning phase' to bring the 'activation' point forward? (Q2)
- How can actions in the 'learning phase' ensure accuracy in the 'automated phase'? (Q1)
- Would communication between system and user benefit the ML efficiency? (Q1 Q2 + Q3)
- How does the user feel about the time taken to automate considering their **perceived** level of routine? (Q2 + Q3)

It appears the method for learning as explained in Figure 1 is solely dependent on the user exhibiting a routine (as that is often the only data being collected), which opens an interesting area for further research. Considering this, Figure 2 was next examined which details an example of a method used by ML systems to learn routine and reach automation:

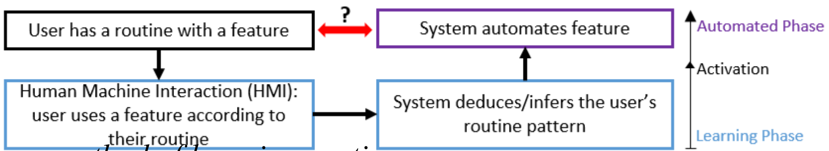

Figure 2. Common method of learning routine

Figure 2 describes the process of user interaction, the deducing of routine and subsequent automation - all of which is based on the user's HMI interaction. The red arrow shows a discrepancy between the system's automation and the actual routine of the user. Two further sub-questions can be drawn from Figure 2:

- What other forms of information can a ML system use during the 'learning phase'? (Q1+Q2)
- What are the differences (perceived and actual) between the features' outputs during the learning stage and the features' outputs during the automated phase? (i.e. What are the differences between the user's actual routine and the system's prediction of routine)

Considering both Figure 1 and Figure 2 it was also interesting to consider the impact of trust and mistrust of automation (Q3). To address this, an initial literature search was first conducted to give a background for further research. This revealed how trust is 'an important component in technology acceptance and adoption' (Bahmanziari, Pearson, & Crosby, 2003) and in order for potential customers to trust technology (such as ML and autonomous systems), it is important for OEMs (Original Equipment Manufacturers) to build up a level of trust between themselves and their potential customers. It has also been previously shown how ‘brand trust is rooted in the result of past experience with the brand and it is also positively associated with brand loyalty’ (Meschtscherjakov, Wilfinger, Scherndl, & Tscheligi, 2009). These findings are important for new autonomy-based features such as in ML applications as they show how such projects will be key in building trust between consumers and OEMs. This trust could later be ‘exploited’ to ensure the smooth introduction of future automated systems where trust is key (such as self-driving cars).

Trust also has a vital role to play in the time taken in the ‘learning phase’, as it was learnt that mistrust in the learning algorithms can increase perceived time to automate; whereas trust in the system and understanding of system status / progress could decrease perceived times and allow users to be more accepting of the ‘learning phase’ (see (Karaca, Erbil, & Ozment , 2011). Considering the understanding of system status, it has also been shown how ‘trusting smart systems depends on those systems sharing the user’s goals and systems are deemed more ‘trustworthy and acceptable when they also provide information’ (Verberne, Ham, & Midden, 2012).

Following the critique of current models and an initial literature search concerned with trust and acceptance, the scope for a focused literature review was then set. The focus of this review covered:

• Understanding routine and interaction - to increase prediction accuracy (addressing Q1)
• Measuring patterns and engagement - to speed up the learning time (addressing Q2)
• Communication requirements between a ML system and user (addressing Q1, Q2 and Q3)
• The effects and affects of acceptance, trust and mistrust on automation (as required from the gaps in the research and as raised in Q3)

Using Warwick University’s Library catalogue and online journal subscriptions a literature search was conducted using key terms from the highlighted questions. Literature was assessed for accuracy, reliability and validity considering data collection techniques, appropriateness of assessments/ statistical tests consistency of findings with previous work, references to support statements and conclusions and applicability of research to the questions highlighted.

It is apparent that there has been a significant focus recently on increasing automation and adding ML ability to homes – commonly known as ‘smart homes’. However, little work has been published about ML and routine in an automotive context. For this reason, the literature review began by looking for extractable lessons learnt from smart home research. Later, healthcare research was analysed with particular focus on information about decreasing perceived waiting time in doctors’ surgeries. Further to this, routine was examined from a psychological perspective to understand its true nature. Other models of machine learning were also examined particularly one from stock market trading where information about other methods were reviewed and critiqued. Considering trust and acceptance - research from healthcare, cyber security and automotive specific sectors were looked at and applicability to automotive ML was examined.

3. Results

The key points to emerge from the literature review are summarised below:

Learning, routine and user engagement (Addressing Questions 1 and 2)

• An ML system needs to be able to differentiate between a wrong prediction, a one-off change in driver interaction and a change in routine (Rashidi & Cook, 2009)
• An ML system should look for a variety of ways of collecting data from the user to aid in the learning of their routine and to understand the hierarchy and impact of the data collected (Das, Cooke, Bhattacharya, Heierman III, & Lin, 2002)
• Consider start-up triggers for activation of features - what might these be and how can they

be controlled to ensure accuracy and learning of routines? (Rashidi & Cook, 2009)
• Consider methods for allowing users to input and edit patterns and automation timetables to add to satisfaction, predictability and accuracy (Rashidi & Cook, 2009) (Domingos, 2012)
• Be aware of the non-rigid and emotionally dependant variation in routine and patterns to improve accuracy of automation and improve learning ability (Cohen, 2007)
• The extent to which different routine influencers affect each routine characteristic and feature engagement would be a beneficial study for an ML project. (Cohen, 2007)
• A multidisciplinary approach to improving the models depicted in Figure 1 and Figure 2 would be beneficial and including influences from computer science would give a better understanding of what information/data is useful to collect. Ranking information importance in a hierarchy would benefit information processing (Das, Cooke, Bhattacharya, Heierman III, & Lin, 2002)
• Testing ML/automation within a vehicle simulator will improve safety and limit effects on trust and acceptance (Karel, Cornelie, Tineke, Bart, & Marika, 2008) (Brookhuis & De Waard, 2005).

Trust and acceptance (Addressing Question 3)

• Ensure transparency in the learning times and system status to reduce perceived learning times (Karaca, Erbil, & Ozment , 2011) and (Pitrou, et al., 2009)
• Mistrust in the learning ability of a system increases perceived time to automate (Karaca, Erbil, & Ozment , 2011) (Weyner, Fink, & Adelt, 2015)
• Look for ways to improve privacy and handling of personal data, including being transparent with users about data processing within an ML system (Jacobsson, Boldt, & Carlsson, 2016)
• Workload and cognitive strain could be useful indicators of acceptance, through understanding the measurements before/after automation (Karel, Cornelie, Tineke, Bart, & Marika, 2008)
• People are becoming more expectant of automation in vehicles (Weyner, Fink, & Adelt, 2015)
• Increasing number of malfunctions and errors increased the participants' perception of lower control (Weyner, Fink, & Adelt, 2015)
• Lack of system status information and therefore 'mode confusion' decreases perception of control and therefore acceptance (Weyner, Fink, & Adelt, 2015)
• Lack of control should be monitored to ensure increased automation in an ML system is not affecting acceptance as predicted in the 'Lack of Control Theory' (Weyner, Fink, & Adelt, 2015)
• To measure trust it is essential to consider extraneous attributes such as opinions of, and predisposition to trust technology (Weyner, Fink, & Adelt, 2015) (Lee & Katrina, 2004)
• User errors due to poorly designed HMI can result in lower trust that the system will achieve a desired task (De Vries, Midden, & Bouwhuis, 2003)
• Reliability of reported trust in a system can be influenced by the users' bias (De Vries, Midden, & Bouwhuis, 2003) (Dzindolet, Pierce, Beck, & Dawe, 2002)
• Communication of errors and full disclosure can increase trust in the system (Mazor, et al., 2006) (Mazor, et al., 2004) (Schwappach & Koeck, 2004)
• Trust should be assessed according to the following criteria – ability, integrity and benevolence (Mayer, Davis, & Schoorman, 1995)

4. Discussion

Previously in ML systems, recent user interaction alone was thought to be the sole effector and descriptor of routine, where Figure 2 is based on a static routine. However, the literature review has highlighted how 'routine' is far from being a static entity and it is constantly being changed and impacted (Cohen, 2007). To understand the extent of the flexibility and non-static nature of routine a diagram has been created highlighting the most relevant impactors:

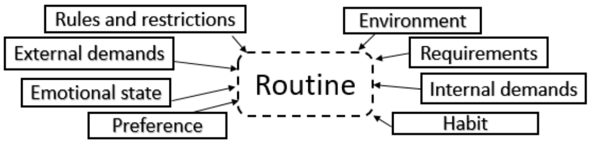

Figure 3 – Impactors on routine

With the information gleaned from the literature review (as summarised in the previous bullet points) it is possible to revisit and develop Figure 1 and Figure 2. In line with the literature and considering the multi-dimensional impactors of routine (shown in Figure 3) it appears that the goal of reducing learning time and increasing accuracy may not achievable through the linear method adopted in Figure 2. This paper adopts and describes a human factors approach to ML to enhance the previous models, including taking reference from a multi-disciplinary selection of sources which consider human-centred design. Where commonly ML technology is focused heavily on just the system capabilities and algorithm development, this approach looks to understand how the human can play a part in development of such systems. Many ML projects have previously overlooked the human, as an effector and resource for system development and as stakeholder in the development of ML and autonomous technology – where ML has previously been a computer science dominated and lead area of study.

Previously, Figure 2 showed a linear process of interaction, learning and 'mimicking' routine through automation. It is now clear that this is a single track approach to learning about routine and although such methods may have sufficed in previous ML projects, in abstract domains, applying the same approach to an automotive application is sub-optimal. Considering a human factors approach in implementing ML in a car one must consider a human-focused system which can improve and increase the facilitation of knowledge transfer between the user and system. This information transfer can take place though multiple streams such as user interfaces, self-programming, observation of interaction, contextual information gathering, connected cars and information sharing etc. The literature has shown how this has been achieved in other sectors and as such there is good evidence for its utility and a good starting point for further automotive focused research. In conclusion of the literature review and critique of the two models a new framework is proposed for a ML system's 'learning phase' (Figure 4):

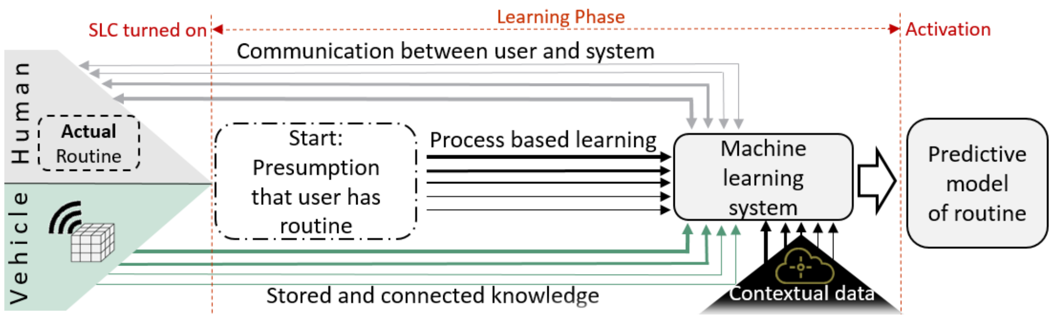

Figure 4 –Proposed framework for a machine learning system

The top left of the framework shows the 'Human' side where **actual** routine (a condensed version of that which is shown in Figure 3) is held. The grey arrows describe the possibility for communication of this routine between the human and the system. Examples in the literature (see (Das, Cooke, Bhattacharya, Heierman III, & Lin, 2002)) on how to achieve this pointed towards user interfaces, notifications and self-programming. Communication can be utilised to gain information about the user's routine, confirm system status and also query learnt information to check accuracy.

The bottom left side of the framework represents the 'Vehicle'. As found in the literature - particularly in stock market modelling (see (I Know First, 2016)), it is possible to store a 'knowledge base' of known entities about, or in relation to patterns, interactions and/or features. This information is multi-dimensional, can be learnt over time and is based on facts, knowledge and external/historic data. Figure 4 shows also how the knowledge base is transferable and shareable between different systems/vehicles. This knowledge base can be used to both inform the system of likely routines/patterns (based on historical data, shared data from other users, information about the surroundings, statistics and likelihood correlations etc.) and also help determine whether observed interactions are likely to be representational of a routine, or a one-off event that is not useful. This information can feed into the machine learning algorithm (as shown by the green arrows) to help inform the predictive model based on stored, connected and shared knowledge.

The red dashed vertical lines represent the beginning and end of the learning phase. The first stage of this learning phase must be - 'presumption that user has routine'. The system cannot know, at the beginning of the process, what the routine of the user is with each feature. Two further information streams are used to inform the 'machine learning algorithm'. These information streams are drawn as black arrows and represent 'processed based learning' and 'contextual data' (or context based learning):

Process based learning is a method of collecting data over time to build an understanding of a routine. Figure 2 shows one such method of process based learning through the monitoring of user interaction. The literature review showed how ML systems should utilize multiple streams of information/data collection. The literature pointed to a few options on how to do this including collecting data through an HMI, inferring usage through monitoring of a user's schedule and engagement with other devices and seeking information about the user's intent or reasons for activating features (see (Rashidi & Cook, 2009) and (Domingos, 2012)).

Contextual data is a level of learning focused on the contextual factors that influence a person's routine. These aspects have been previously discussed through terms such as 'start up triggers' in the literature (Rashidi & Cook, 2009) and account for independent factors that will (possibly be) unique for each feature. Through collecting contextual data about a user's routine the system can both identify possible start up triggers for routine interaction and also learn information about the user's requirements for a feature in relation to contextual information (considering time, location, vehicle occupants, purpose of driving, work day vs holiday etc. and many other factors).

In accordance with the literature, some streams of information will be more important than others. This 'hierarchy' of importance is represented in Figure 4 by thickness of arrows and shows how some information sources are more useful or more reliable than others.

It is important to note that the proposed framework in Figure 4 only accounts for the learning phase and does not consider the later activation stage or automated stages, and provides an explanation of how the learning phase may be shortened and activation accuracy could be increased. To this end, further research would be required in order to extrapolate this further into the automation stage. It is however recommended and expected that learning continues throughout the automated phase to account for changes in routine over time.

If the framework depicted in Figure 4 is properly implemented it is believed that the timeline shown in Figure 1 could be affected as the time taken to reach activation could be sooner and the strength of the information gathered in the learning phase can ensure accuracy in activation.

5. Conclusion

Through a review of current literature and the development of a modified ML framework; methods of improvement for current systems for machine learning have been highlighted. The success of the method proposed in this paper (see figure 4) for decreasing learning time, improving accuracy and managing trust/acceptance is untested and as such, no conclusion can be made yet as to its effectiveness. However, the literature suggests the key to speeding up learning time and improving accuracy does lie in the facilitation of information transfer between user and vehicle. What's more, through implementing the recommendations for increasing trust and acceptance also, it is hoped future ML projects can ensure a positive impact on its current users and gain longer term benefits for the OEM and their future features.

References

Bahmanziari, T., Pearson, M. J., & Crosby, L. (2003). Is Trust Important in Technology Adoption? A Policy Capturing Approach. *Journal of Computer Information Systems, 43*, 46-54.

Brookhuis, K., & De Waard, D. (2005). ADAS Acceptance and Effects of Behaviour: The Consequences of Automation. *Traffic and Transport Psychology. Theory and Application*, 273-278.

Cohen, M. D. (2007). Reading Dewey: Reflections on the Study of Routine. *Organization Studies, 25* (5), 773-786.

Das, S. K., Cooke, D. J., Bhattacharya, A., Heierman III, E. O., & Lin, T.-Y. (2002, December). The Role of Prediction Algorithms in the MavHome Smart Home Architecture. *IEEE Wireless Communications, 9*(6), 77-84.

De Vries, P., Midden, C., & Bouwhuis, D. (2003). The Effects Of Errors On System Trust, Self-Confidence, And The Allocation of Control In Route Planning. *International Journal of Human-Computer Studies, 58*(6), 719-735.

Domingos, P. (2012). *A Few Useful Things to Know about Machine Learning.* Department of Computer Science and Engineering. Seattle, WA 98195-2350, U.S.A: University of Washington.

Dzindolet, M. T., Pierce, L. G., Beck, H. P., & Dawe, A. L. (2002). The Perceived Utility of Human and Automated Aids in a Visual Detection Task. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 44*(1), 79-94.

I Know First. (2016). *About I Know First*. (I Know First) Retrieved 09 02, 2016, from http://iknowfirst.com/about-i-know-first

Jacobsson, A., Boldt, M., & Carlsson, B. (2016). A Risk Analysis Of A Smart Home Automation System. *Future Generation Computer Systems, 56*, 719-773.

Karaca, M. A., Erbil, B., & Ozment , M. (2011). Waiting in the Emergency Room: Patient and Attendant Satisfaction and Perception. *European Journal of Surgical Sciences, 2*(1), 1-4.

Karel, A. B., Cornelie, J. V., Tineke, H., Bart, v., & Marika, H. (2008). Driving With a Congestion Assistant; Mental Workload And Acceptance. *Applied Ergonomics, 40*, 1019-1025.

Lee, J. D., & Katrina, S. A. (2004). Trust In Automation: Designing For Appropriate Reliance. *Human Factors, 46*(1), 50-80.

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An Integrative Model of Organisational Trust. *Academy of Management Review, 20*, 473-490.

Mazor, K. M., Simon, S. R., Yood, R. A., Martinson, B. C., Gunter, M. J., Reed, G. W., & Gurwitz, J. H. (2004). Health Plan Members' Views About Disclosure of Medical Errors. *Annals of Internal Medicine, 140*(6), 409-418.

Mazor, M. K., Reed, G. W., Yood, R. A., Fischer, M. A., Baril, J., & Gurwitz, J. H. (2006). Disclosure of Medical Errors What Factors Influence How Patients Respond? *Journal of General Internal Medicine, 21*(7), 704-710.

Meschtscherjakov, A., Wilfinger, D., Scherndl, T., & Tscheligi, M. (2009). Acceptance of Future Persuasive In-Car Interfaces Towards A More Economic Driving Behaviour. *1st International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (pp. 81-88). New York: AMC.

MyBlossom. (2015). *Blossom,The Smart Watering Controller*. Retrieved 07 20, 2016, from Blossom: http://myblossom.com/

Pitrou, L., Lecourt, A., Bailly, L., Brousse, B., Dauchet, L., & Ladner, J. (2009). Waiting Time And Assessment Of Patient Satisfaction. *European Journal of Emergency Medicine, 16*, 117-182.

Power-up your lights. (2015). Retrieved from BeOnHome: http://beonhome.com/

Rashidi, P., & Cook, D. J. (2009). Keeping the Resident in the Loop: Adapting the Smart Home to the User. *Institute of Electronic and Electrical Engineering, 39*(5), 949-959.

Schwappach, D. L., & Koeck, C. M. (2004). What Makes An Error Unacceptable? A Factorial Survey On The Disclosure Of Medical Errors. *International Journal for Quality in Healthcare, 16*(4), 317-326.

Verberne, F. M., Ham, J., & Midden, C. J. (2012, 10). Trust in Smart Systems, Sharing Driving Goals and Giving Information to Increase Trustworthiness and Acceptability of Smart Systems in Cars. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 54*(5), 799-810.

Weyner, J., Fink, R. D., & Adelt, F. (2015). Human-Machine Cooperation In Smart Cars. An Empirical Investigation To The Loss-Of-Control Thesis. *Safety Science, 72*, 199-208.