Using decision tree algorithms for modelling driver distraction situations.

Tariq Mouhtadi, Utrecht University

May 16, 2017

Contents

Abstract

Algorithms based on decision trees are widely used in artificial intelligence research. In this thesis we apply two different of these algorithms to data gathered in a driver distraction study with 25 participants. The majority of studies in cognitive modelling only start gathering data after formulating a hypothesis, based on theory in the domain. We starts with the data however and analyse it to try and find interesting features. The idea behind this approach is to try and discover less obvious patterns. First we extracted 7 features from the original data to get a dataset that we could apply our algorithms to. Our algorithms were able to correctly predict 54% and 57% respectively of the users either focussing on the road or on their phone. We found that inherent properties such as age and gender are not strong predictors of performance in driving. On the other hand looking at the actual driving performance can be used to see if a participant is focussing on the road or their phone. It is important to note for further research that the original data assumed that participants were following instructions to focus on the road or the phone, but did not verify it. To gain better insights from decision tree algorithms it is important that the data is checked to ensure that the data is consistent. For improving the accuracy of the algorithms it would also greatly help to have a larger number of participants. This task had participants look at their phone to fill in a number. For further research it would be useful to look at tasks with distractions in other modalities with the same algorithms.

1 Introduction

Computers are quite efficient at processing large amounts of data. We should use this by this inherent property should be leveraged by applying machine learning to large data sets, and letting experts interpret the results with their irreplaceable domain knowledge.
This paper explores the use of machine learning for cognitive modeling by modeling on a dual task of driving and dialing a phone number, using the data from Janssen and Brumby (2010).
We will distinguish between ”white-box” and ”black-box” machine learning, and explain why only ”white-box” is suitable for feature selection.

In most cognitive modelling research a hypothesis is formed based on the theory, and subsequently data is gathered to test this hypothesis. We present a different approach by starting with the data. We will determine which features from this data are interesting, and which are not. This leads to recommendations on which type of features should be expanded upon for further research.

The large amount of data that can be obtained nowadays bringing together the fields of cognitive modeling and machine learning. These fields are very compatible, as from the perspective of cognitive modeling there is a situation where wide sets of data can be gathered from experiments, and there is a need for algorithms and models to process this data (Janssen, Gould, Li, Brumby, & Cox,2015). From the other side in the field of machine learning, one of the necessities for successful work are large data sets. Previously it was asserted that at least 50 data points per participant were needed (Webb, Pazzani, & Billsus,2001). But nowadays we have data sets with thousands or even millions of points per user.
It is also said that by 2021 machine learning should be a core technology in the information economy (Webb et al.,2001). We hope to spread the use of these tools in cognitive modeling research.

To demonstrate the effectiveness of the right machine learning tool for this problem we show an application of decision tree algorithms to the data. This should aid in model selection by finding an interesting subset of variables.

Decision trees are a well proven way of working with data that is both powerful and accessible, and the use of the scikit-learn python library yields the potential of using state-of-the-art processes in the field of artificial intelligence, with a well tested library. This decreases the chance of statistically incorrect results.

2 Related Work

We will start by explaining the theory of decision tree models, and then expand on the domain surrounding the dual task.

2.1 Machine Learning

Machine learning is the use of advanced learning algorithms to create models that reliably fit a data set. The models are created by training them on a 'training set' and scoring them with a 'test set'. The differences between various machine learning algorithms lie in their method of improving their models, or learning, and the structure of the resulting model.

A variety of different ways of obtaining the training data and test data can also be observed. With common usage of taking a split from the total data, e.g. 80:20, with a randomly chosen 20% of the data going into the test set, and 80% going to the training set. Or another method of folding the data where different slices of the data set are taken as test data every run.
The goal of this paper is to demonstrate a simple and elegant use of machine learning on the domain of driver distraction settings. We will use the classic 80:20 split, as it is the most accessible way of creating the training and test data sets. The 80:20 split is called the "Pareto Principle" and is well established in artificial intelligence.

Carbonell, Michalski, and Mitchell (1983) introduce the three research foci that machine learning is based around:

- "Task-Oriented Studies - the development and analysis of learning systems to improve performance in a predetermined set of tasks (also known as the "engineering approach")."
- "Cognitive Simulation - the investigation and computer simulation of human learning processes."
- "Theoretical Analysis - the theoretical exploration of the space of possible learning methods and algorithms independent of application domain."

This paper is in the second of these foci the 'Cognitive Simulation', and uses insights from research into the 'Theoretical Analysis' of machine learning to base the models on.

Within the field of machine learning there is a divide that can be made between two types of algorithms for learning.

Black box is an umbrella term for all algorithms that do not reveal how they learn a task.
You can give a black box model an input and it will return an output, without explaining how this output is calculated. These black box algorithms work great for learning to recognize incredibly complex patterns in large sets of data. The biggest issue with the techniques however is that it is quite troublesome to predict the appearance of any bugs in the calculation. The only way to reliably test the results from a black box algorithm is to do tests with large data sets annotated by humans.

The other type of algorithms are called white box algorithms. These generate a model with a structure that can be visualized and analyzed by researchers or even audited by a panel of experts.
This ability to transparently reveal the inner structure of a model to test its validity is crucial in

achieving trustworthy predictions as mistakes in generating models can have severe consequences in for example medical applications. Therefore these models should be validated by an external data set and by surveying a panel of exports on the domain. Which is only possible for white box algorithms, and decision trees have the advantage that can easily be expressed as rules, and as such are white box models. (Dreiseitl & Ohno-Machado,2002).

2.1.1 Decision Tree Classifier

Decision tree models have been found as the most useful models in the domain of data mining for their reasonable accuracy and inexpensive computation (Du & Zhan,2002).
At the foundation of decision tree algorithms lie simple statistical models. A decision tree algorithm consists of a number of decisions that lead to leaf nodes in the tree. As the idea for decision trees is use a statistical model for each decision in the tree (Jordan,1994). This means that every path through the tree is a complex statistical model. The entirety of the tree can in effect be reduced to a classical statistical model. In the case of a classification decision tree algorithm, like we use, that model can then be used to classify data into disjoint sets.

When handling decision trees within a statistical framework both likelihood and Bayesian theory can be applied analyse and optimise the models (Jordan,1994).

The objective of classification is building models using the training data that can be used to classify new data whose class labels are unknown (Du & Zhan,2002).

In these experiments the class labels are 'focussed on dialing' and 'focussed on driving'. These come from the original task, by Janssen and Brumby (2010), where participants were instructed either to focus on staying in the middle of the road, 'focussed on driving', or to type in the phone number as fast as possible, 'focussed on dialing'.

2.1.2 Interpreting Decision Trees

To gather new insights from a trained decision tree model, researchers can use the model to classify new data to predict unseen data points. Decision trees were used in the field of psychiatry where classification and regression tree (CART) analysis generated decision trees that assess psychiatric patients to determine their suicide risks (Mann et al.,2008). They achieved 69% accuracy in identifying patients that had previously attempted suicide in the last 30 days from those who had never attempted suicide. Which can be very helpful identifying future attempts, as they note that these 30 days are the most relevant in assessing suicide risk.

A domain that has benefitted from using decision tree based models is spectrum analysis for chemical and biological research, here using 'gini importance' of features to eliminate irrelevant features helps creating more accurate models (Menze et al.,2009). This gini importance is a way of calculating the effect a feature has on a calculation by counting the amount of times a feature is used to split the data, and what effect each of these splits has on the classification. It is this gini importance that we will use to determine the effect our features will have on classifying the dual task.

2.2 Cognitive Modeling

Cognitive modeling is a domain in which researchers model human behavior by creating theory based models that mimic brain processes. Such as making a model for how attention is used in a driving task. Cognitive modeling is interesting from an artificial intelligence perspective because it also concerns itself with human performance from an integrated standpoint.

With the high risks accidents in centralized production units it is important to correctly predict human performance. To effectively use modern information technology in designing safe interfaces, it is needed to study human performance as a whole. Whereas it is normally studied in separate paradigms (Rasmussen,1983).

Rasmussen (1983) concludes that a systematic description of human performance is needed. Noting that these should cover both daily routine and stressful accidents. Machine learning can be used to create such systematic descriptions.

To create a machine learning model suitable for a cognitive modeling perspective, we first need to define some key characteristics of the domain. The features that shall later be used for the model will be analyzed with two cognitive modeling concepts, 'types of data' and 'types of knowledge'.

Types of data

Rasmussen (1983) defines three types of data: signals, signs, and symbols.
In a task where a participant has to remember a phone number and input it into a phone these three types of data can be identified and studied.
The input of a number into the phone is a signal, a time-space action that has an effect on the state of the total input.
The current state of the number in the phone is a sign, it is the result of the environment and various signals.
The end state of the number in the phone is represented as a symbol, this type of data is an abstract representation of a situation.

Studying these types of data can show us whether it is interesting to look at events during a task, or the situation surrounding it. This helps define the type of feature that is being looked at.

Types of knowledge

Three types of knowledge, referring to the types of data, are defined by Rasmussen (1983):

- **skills-based knowledge** information is perceived as signals.
- **rule-based knowledge** information is received as signs.
- **knowledge-based knowledge** information is represented as symbols.

In a task where a participant has to drive his car from point A to point B all three types of knowledge can be specified. Steering the car through a corner is skills-based knowledge, a driver has obtained the skill to steer his car and reacts to signals such as how the car reacts to his braking.
Reacting to the car drifting of from the center of the road is rule-based knowledge, there is no conscious thought involved, as the driver intuitively corrects themselves to adhere to a certain rule, in this case driving in their lane.
Knowing the route to drive from A to B is knowledge-based, the driver has to think about possible routes and plan his course, sometimes also adjusting to external symbols, such driving around a traffic jam.

Specifying the domain

Since the start of Artificial Intelligence there has been a focus on human intelligence, as McCarthy, Minsky, and Rochester (1955) noted in their proposal that started the field: "The study is to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can

in principle be so precisely described that a machine can be made to simulate it."
Modelling and simulating human intelligence is quite an accurate, albeit simplified, description of the field of Cognitive Modelling. It is thus not too far-fetched to join these fields in research.

Within the domain of cognitive modelling there is a specific research field that is obviously suited for a data-driven approach, multitasking. Multitasking is studying the performance of subjects while performing a multitude of tasks. This creates difficulties in building consistent models, as it is hard to predict the effect one task will have on the others.

Multitasking in driving
Computational modelling allows researchers to simulate driver behavior, and explore its constraints and parameters (Salvucci,2006). According to Salvucci (2006) the research community has seen a growing push towards models that unify the many aspects of driving into a single computational model.
Thus we need algorithms that can take a large number of features and use it to create all encompassing models that predict driver behaviour. Machine Learning is perfectly suited to selecting relevant information from large structured and unstructured datasets, taking into account all dimensions available in this data.

Michon (1985) divides the task of the driver in three levels of skill:

- **strategical (planning)** planning a route to a destination.
- **tactical (maneuvering)** evading other cars and objects.
- **operational (control)** operating the steering wheel and gas pedal, basic unconscious control actions.

Safe driving typically involves all of these three levels (Salvucci,2006).
This makes driving a very interesting task for studying external effects on human behaviour, as these combined processes have the effect of taking up the full mental capacity of the driver.

So when driving is combined with other tasks the driver is forced to divide his attention between the tasks.
Driving can also involve secondary tasks such as tuning a radio or dialing a number on a cellphone (Salvucci,2006). It is this dialing that the original experiment used as a distraction task.

3 Experimental Design

The original study had participants perform four related tasks. We use the data gathered during these tasks to build our models and find interesting features related to multitasking.

3.1 Tasks

Typing practice task
In the practice task the participants had to correctly enter parts of the total phone number. This was meant to train them in remembering the number. It is interesting to see how long they took to complete these trials, as it shows how adept they are with a cellphone.

Typing pre-test task

This had participants correctly enter the entire phone number and is interesting for the same reason as the practice task.

Single task Driving
In the single driving task participants had to just keep the car centered on the road. This task reveals how adept they are at driving a car in general.

Dual task Driving
The dual-task was originally performed in Janssen and Brumby (2010) to research the interleaving of attention in multitasking, they explain their study: 'we investigate whether dual-task interleaving is confined to subtask boundaries. To address this question, we use a dialing while driving dual-task paradigm.'

Participants were instructed either to focus on keeping their car centered, or on being fast in dialing the number. This forces participants to adopt different strategies to achieve their goal.
Every trial in this task is annotated with one of these focus conditions. Which makes it a good target for classification algorithms.

3.2 Data

The data from the original experiment consisted of 2.114.150 rows of events during the four tasks. There were 23 variables in the dataset measuring different parts of the tasks, and indicating which trial was happening for which subject.

3.3 Original Model

The original model that was made in Janssen and Brumby (2010) focusses on the dual task trials. They use the state of the phone number and the deviation from the center of the road to look at strategies a driver uses to stay on the road.
This model had its focus on the dialing aspect of the dual task, whereas we focus on the driving part.

The model shows that strategic decisions are being made about the right time to switch back to focussing on the road.

4 Research Questions

We formulated three research questions to help us look at the data in a focussed way.
For this we look at symbols and signals from the theory (Rasmussen,1983).
RQ1 Are symbols useful in predicting driver distraction.
For this we look at gender, age, handedness, driving skill, and typing skill.

RQ2 Are signals useful in predicting driver distraction.
For this we look at deviation from the center of the road during the distraction task.

RQ3 Can we use decision trees to accurately predict driver behavior.
To look at this we implement two decision trees on the data, and check their accuracy.

5 Procedure

6 Measurements

In this section we will go over the different features that we will use in the decision trees later on. These features were measured by questionaire for the symbols. And by the four tasks in the original experiment for the signals.

6.1 Gender

The population consisted of 17 men and 8 women. This feature could be interesting to see if there are biological reasons underlying the differences between participants. A number of researchers have done studies into the differences between men and women in multitasking.

The hunter-gatherer hypothesis suggests that due to evolutionary reasons, women are better at multitasking than men (Silverman & Eals,1992;Eals & Silverman,1994). If this hypotheses holds true, women should be perform better in the dual-task than their male counterparts.

Research done to study this hypothesis concluded that female cognitive control is more robust against control demand compared to males, and their findings suggest that women have the potential to be better multitaskers than men (Ren, Zhou, & Fu,2009). This again point to women performing better than men.

On the other hand Mäntylä (2013) found that over the course of multiple experiments men were better at multitasking than females, and in research done specifically on the area of driving women were found to self-regulate more than men. This caused them to disconnect from the driving task due to handle the stress (Gwyther & Holland,2012). This could indicate that women perform worse at least on a driving task.

Even though these researchers seem to come to different conclusions, they all indicate that gender should be a significant feature in multitasking in driving. Therefore it should certainly be included in any analysis of driving tasks.

6.2 Handedness

Of the participants 4 were left handed with the remaining 21 being right handed. This is not a lot of research being done to handedness in multitasking, however in research done into multitasking with interpreters and translation. Stachowiak (2015) found no effect related to handedness.

6.3 Age

The variations in age in the data set are shown in Figure 1. Age seems to have an influence on frontal lobe structure causing differences in age related impairments between individuals, and older adults do show consistent impairments in multitasking. But this seems to be mitigated by depending more on external cues instead of internal cues (Kievit et al.,2014).

In research done by Chaparro, Wood, and Carberry (2005) between two participant groups, where the young were 33 participants with an average age of 27, and the elder a group of 21 aged around 69, showed that the young subjects detect significantly more roadsigns. And although both groups were affected by a secondary task, the elder were significantly worse at multitasking. Which is consistent with reports that elderly drivers are more likely to cause accidents due to missing signs or traffic lights.

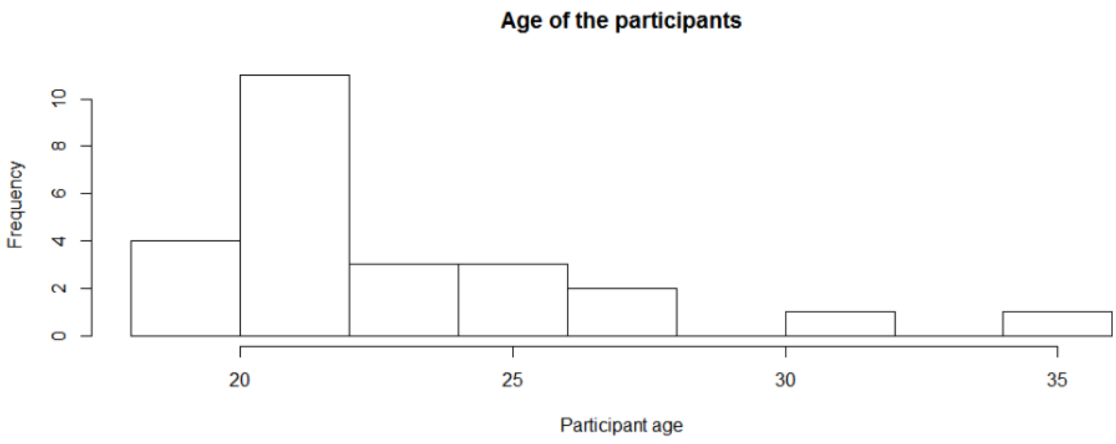


Figure 1: A histogram of participant age

6.4 Typing speed

The dialing task had participant enter a single UK-style phone number on a cellphone. To measure the typing the participants had to dial a number on a Nokia 6300 on the left of their steering wheel (Janssen & Brumby,2010). They also had to correct mistakes by typing "*", and start and finish the trial by typing a "#". The measure for typing speed in both typing tasks was the time until task completion.

6.5 Driving skill

In the driving task the participant has to control the steering in a vehicle that is going a constant speed of 88.5 km/h. We are interested in the maximum deviation to the left of the center of the road, and the maximum deviation to the right of the center of the road during each trial. A smaller range of road use is a better score in this case.

$$largest\_deviation\_to\_the\_right - largest\_deviation\_to\_the\_left = range$$

Where the largest deviation to the right is the highest value the lateral deviation from the center of the road had. And the largest deviation to the left is the lowest value of this lateral deviation, and is always negative as the minimum is 0. The difference between these gives us the range for that trial. Which we take as the measure for driving performance.

These values were measured in a simulation on a Logitech G25 with a 30-inch monitor. The simulation with had so the participants had to steer to keep in the center (Janssen & Brumby,2010).

7 Method

In this thesis we use two decision tree algorithms to create models based on the data previously described.
To be able to use these algorithms we must first pre-process the data into a structure that the algorithm can learn from. We call this a vector, and in this case the vectors contain seven variables.

After the data is structured it can be given to algorithms to create the two models. The data is first split into a training set and a test set. With the 80:20 split discussed earlier.

The algorithms will use the gini coefficient to select the best features for each decision step. We will then look at the scores for the fit of the model and to the gini importance of each feature in the models.

gini coefficient
The gini coefficient is an often used measure in the fields of artificial intelligence, economics, statistics and probability. It was devised as a measure of inequality in data, and in our decision trees it describes the balance between the classifications in each node.
A gini coefficient of 0.5 means that there are equal numbers of both classification in that node. On the other hand a lower measure like 0.1 means that there are almost only values in of the classes in that node.
This gini coefficient is most often the first measure taught in machine learning education, and is considered the default for choosing features in these algorithms.
The other measure is also commonly used for the same purpose is entropy.

score of the model
The score of the model is measured in the simplest way. We look at the amount of trials in the test set it correctly predicts and divide it by the total number of trials in that set.

gini importance
The mean decrease impurity is a measure used in machine learning to indicate how much an impact a feature has on the prediction. If a feature is good at splitting the data it has a high mean decrease impurity. When the splitter used is the gini coefficient the mean decrease impurity is also called gini importance (Louppe, Wehenkel, Sutera, & Geurts,2013).

The main difference between the algorithms lies in the maximum depth of the model it is allowed to create. For one algorithm the limit is set to three layers, and the other has no limit on the depth. These models should not yield drastically different results. But the unlimited model has the potential to reveal smaller effects. This comes at the risk of overfitting on training data however.

7.1 pre-processing

The original research noted the subjects age, handedness and gender. Before doing the dual task participants were tasked to do 3 single tasks first. A phone practice task, typing in smaller pieces of the phone number. A phone pre test task, typing in the whole phone number. And a single driving task, just driving in the middle of the road.

For both phone related tasks we took the mean of the maximum time a participant took to do a trial, and standardized it. For the driving task we calculated a range per trial for deviation from the middle of the road, and took the mean of this single driving score. The dual task has the same range as the single task, but now an individual value per trial instead of the mean.

7.2 3-layer model

7.2.1 Plan

The main model of this paper is a decision tree model with a limit on the amount of layers the model can have. This is set to 3 layers which gives us a quite simple model, as shown in the appendix, that avoids overfitting on the data.

This model will use the gini coefficient to split the data to the best features.

7.2.2 Implementation

For generating the model we use sci-kit learn, a python library specifically made to answer the growing need for statistical data analysis by non-specialists in all fields of science (Pedregosa et al.,2011).
For the three layer model we use the DecisionTreeClassifier from sci-kit learn on the training data. There are three parameters we have to consider here. Criterion, splitter and max_depth.
The criterion in this case is gini as we want to use the gini coefficient to make our decisions.
For splitter we have the options "best" or "random", where best uses the splitter to make a decision, and random chooses a random feature. We choose "best" as we are interested in the most interesting features.
Lastly this model will have a max depth of 3 layers, so we set max_depth to 3.

7.2.3 Score

The simple accuracy score for this model was 57%.

The gini importances for this model are:

range	gender	handedness	age	phonepracticetime	phonepretesttime	singledrivingscore
0.63	0.00	0.00	0.00	0.06	0.31	0.00

7.3 unlimited model

7.3.1 Plan

The unlimited model is to be generated exactly like the three layer model, with the difference that is has no limit on depth.
This allows us to see if more information could be extracted by making a more complex model.

7.3.2 Implementation

We again use the DecisionTreeClassifier for this model, with the same criterion("gini") and splitter("best"). By setting the max_depth to None we allow the algorithm to create a model without thinking about the amount of layers, only stopping when it cannot make any meaningful splits anymore.

7.3.3 Score

The simple accuracy score for this model was 54%.

The gini importances for this model are:

range	gender	handedness	age	phonepracticetime	phonepretesttime	singledrivingscore
0.74	0.03	0.01	0.04	0.04	0.10	0.05

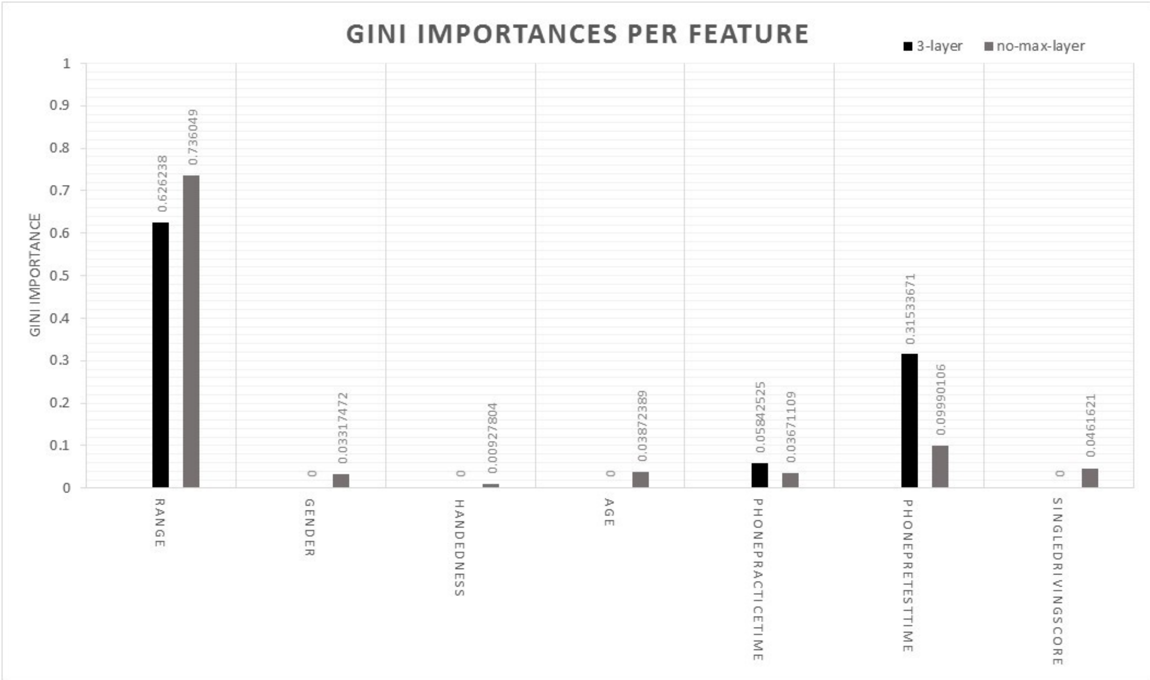


Figure 2: An overview of gini importances in the two models

8 Results

In figure 2 the gini importances of both models are depicted. Both models had very similar results. They also were almost equals in accuracy with scores of 57% and 54%.

RQ1 *Are symbols useful in predicting driver distraction.*
We did not find any strong effects from any of the symbols.
The symbol with the highest gini importance was the phonepretesttime.
The effect of this variable appears greater when the model is less complex.
Most of the symbols seemed to have no effect on driver distraction.

RQ2 *Are signals useful in predicting driver distraction.*
The range in the dual task was the best feature for predicting driving behavior. The result of the model are not significant, but it does point to signals being effective in predicting driver distraction.

RQ3 *Can we use decision trees to accurately predict driver behavior.*
Both models had a decent accuracy in predicting driver behavior. They were very simple models, and still managed to get above change results.

9 General Discussion

Summary of results
Both models got modest scores for predictions on the focus condition, but these results are not statistically significant. The models do provide insight into which variables are interesting for further research. Furthermore in other research it was noted that other algorithms such as logistic regression can yield better results, when used on the variables that were found to be interesting with decision trees (Menze et al.,2009).

Our main result is that how one drives is the dual-task is most indicative of the focus condition. With gini importances of 0.63 and 0.74 it is by far the most important feature we measured. This also shows that there is a definite effect measurable from focussing on a phone while driving.

For the typing features we did not get such a strong result. The pre-test time seemed to have some promise with one gini importance of 0.32, but we cannot draw any conclusions on the effect of typing proficiency on dual task performance.

According to the theory age should have an effect on multitasking performance. In our models we did not find such an effect, as the gini importances were just 0.00 and 0.03. A probable reason why we did not find any effect on age is due to the small range of ages of our participants. With the age feature varying between 19 and 35 which would be classified as the same age group in the theory (Chaparro et al.,2005).

Handedness has no effect on multitasking according to the theory (Stachowiak,2015). We found gini importances of 0.00 for handedness on both models, also suggesting there is no effect.

Contrary to the effect of the driving performance during the dual-task, we found no effect with the skill of the driver. With gini importances of 0.00 and 0.05 inherent driving skill does not seem to be very interesting.

In the current theory there is research to suggest men are better at multitasking (Mäntylä,2013;Gwyther & Holland,2012) and there is other research to point to women being better at multitasking (Silverman & Eals,1992;Eals & Silverman,1994;Ren et al.,2009). With low gini importances of 0.00 and 0.03 gender seems to have no effect on driving performance.

Although the resulting scores for the fit of the models were only 54% and 57%. It is hard to say in this case if this means the fit is not good. It is important to note that there was no verification in the original data to see if participants were following instructions. This stops us from making any strong conclusions about the errors in the models. For further research it is thus very important to check the validity of the labels used in experiments.

Implications for theory

The model originally used with the data (Janssen & Brumby,2010) used driving and typing performance to investigate how good people are at multitasking. We found that at least driving, and probably typing performance is a good measure for determining the focus condition. This confirms that the original research was targeting the right features.

This research again confirms that there is a measurable effect on driving when one is distracted by their phone.

Using decision trees was useful because it allowed for a larger number of features to be compared to eachother. The gini importances show us that signals are more interesting for studying multitasking than signs. The direct implication of this finding for the theory is that researchers should focus on gathering data related to the task that is performed, and focus less on the inherent properties of the participants.

Implications for design

According to both decision tree models monitoring the deviation from the center of the road can indicate if a driver is focussing on the road, or on a secondary task. This could be used to warn drivers, and pull their attention back to the road. There are already car manufacturers applying techniques to warn drivers when they become fatigued. For example, Mercedes-Benz claims to use 70 measured variables to notice a driver losing focus and warn them to take a break. (Mercedes-Benz,2017)

It is certainly possible for car manufacturers to implement systems that warn drivers when they are being distracted. This could prevent accidents from happening by warning drivers when they become distracted.

For researchers the use of decision trees can be a useful tool for feature selection. By looking at the tasks starting from the data interesting features can be separated from features without any effect.

Limitations and Future work
The greatest limitation of this thesis is the use of just one measurement for driving performance. For deeper investigations into dual task driving performance it would be interesting to look into more variables related to driving and typing. Measuring braking reaction speed, or a drivers reaction to drifting from the center of the road for example.

There are more features that can be extracted from the data that we used. However we believe that it is more useful to start a new study by gathering new data. There are a few things that are important in gathering such data. Conditions from the data should be verified. A larger number of participants is needed, as 25 is not enough to be able to generalize the models. More features of the task being performed should be measured.

Ideally future work should be conducted with car manufacturers on closed tracks. Using the precise measurements Mercedes-Benz claims to do would help greatly in studying driver performance. Thus we recommend that research should be done in collaboration with the industry.

Conclusion
There is potential for decision tree and other white box algorithms to aid scientists from other disciplines in their research. The strongest contribution that these algorithms can offer is in feature selection. These algorithms can take in a large number of features and select those that have the strongest correlations.

In this thesis we found that driving performance can be used to determine if a driver is focussing on the road. More accurate predictions can be made by measuring more aspects of driving.
Our results suggests that it does not matter so much who is driving, but how they are driving.

References

Carbonell, J. G., Michalski, R. S., & Mitchell, T. M. (1983). An overview of machine learning. In *Machine learning* (pp. 3–23). Springer.

Chaparro, A., Wood, J. M., & Carberry, T. (2005). Effects of age and auditory and visual dual tasks on closed-road driving performance. *Optometry & Vision Science*, *82*(8), 747–754.

Dreiseitl, S., & Ohno-Machado, L. (2002). Logistic regression and artificial neural network classification models: a methodology review. *Journal of biomedical informatics*, *35*(5), 352–359.

Du, W., & Zhan, Z. (2002). Building decision tree classifier on private data. In *Proceedings of the ieee international conference on privacy, security and data mining-volume 14* (pp. 1–8).

Eals, M., & Silverman, I. (1994). The hunter-gatherer theory of spatial sex differences: Proximate factors mediating the female advantage in recall of object arrays. *Ethology and Sociobiology*, *15*(2), 95–105.

Gwyther, H., & Holland, C. (2012). The effect of age, gender and attitudes on self-regulation in driving. *Accident Analysis & Prevention*, *45*, 19–28.

Janssen, C. P., & Brumby, D. P. (2010). Strategic adaptation to performance objectives in a dual-task setting. *Cognitive science*, *34*(8), 1548–1560.

Janssen, C. P., Gould, S. J., Li, S. Y., Brumby, D. P., & Cox, A. L. (2015). Integrating knowledge of multitasking and interruptions across different perspectives and research methods. *International Journal of Human-Computer Studies*, *79*, 1–5.

Jordan, M. I. (1994). A statistical approach to decision tree modeling. In *Proceedings of the seventh annual conference on computational learning theory* (pp. 13–20).

Kievit, R. A., Davis, S. W., Mitchell, D. J., Taylor, J. R., Duncan, J., Henson, R. N., et al. (2014). Distinct aspects of frontal lobe structure mediate age-related differences in fluid intelligence and multitasking. *Nature communications*, *5*.

Louppe, G., Wehenkel, L., Sutera, A., & Geurts, P. (2013). Understanding variable importances in forests of randomized trees. In *Advances in neural information processing systems* (pp. 431–439).

Mann, J. J., Ellis, S. P., Waternaux, C. M., Liu, X., Oquendo, M. A., Malone, K. M., et al. (2008). Classification trees distinguish suicide attempters in major psychiatric disorders: a model of clinical decision making. *The Journal of clinical psychiatry*, *69*(1), 23.

Mäntylä, T. (2013). Gender differences in multitasking reflect spatial ability.

McCarthy, J., Minsky, M., & Rochester, N. (1955). A proposal for the dartmouth summer research project on artificial intelligence.

Menze, B. H., Kelm, B. M., Masuch, R., Himmelreich, U., Bachert, P., Petrich, W., et al. (2009). A comparison of random forest and its gini importance with standard chemometric methods for the feature selection and classification of spectral data. *BMC bioinformatics*, *10*(1), 213.

Mercedes-Benz. (2017). *ATTENTION ASSIST - Mercedes-Benz.* `http://techcenter.mercedes-benz.com/en_CA/attention-assist/detail.html`. (Accessed: 2017-05-15)

Michon, J. A. (1985). A critical view of driver behavior models: what do we know, what should we do? In *Human behavior and traffic safety* (pp. 485–524). Springer.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: Machine learning in python. *Journal of Machine Learning Research*, *12*(Oct), 2825–2830.

Rasmussen, J. (1983). Skills, rules, and knowledge; signals, signs, and symbols, and other distinctions in human performance models. *IEEE transactions on systems, man, and cybernetics*(3), 257–266.

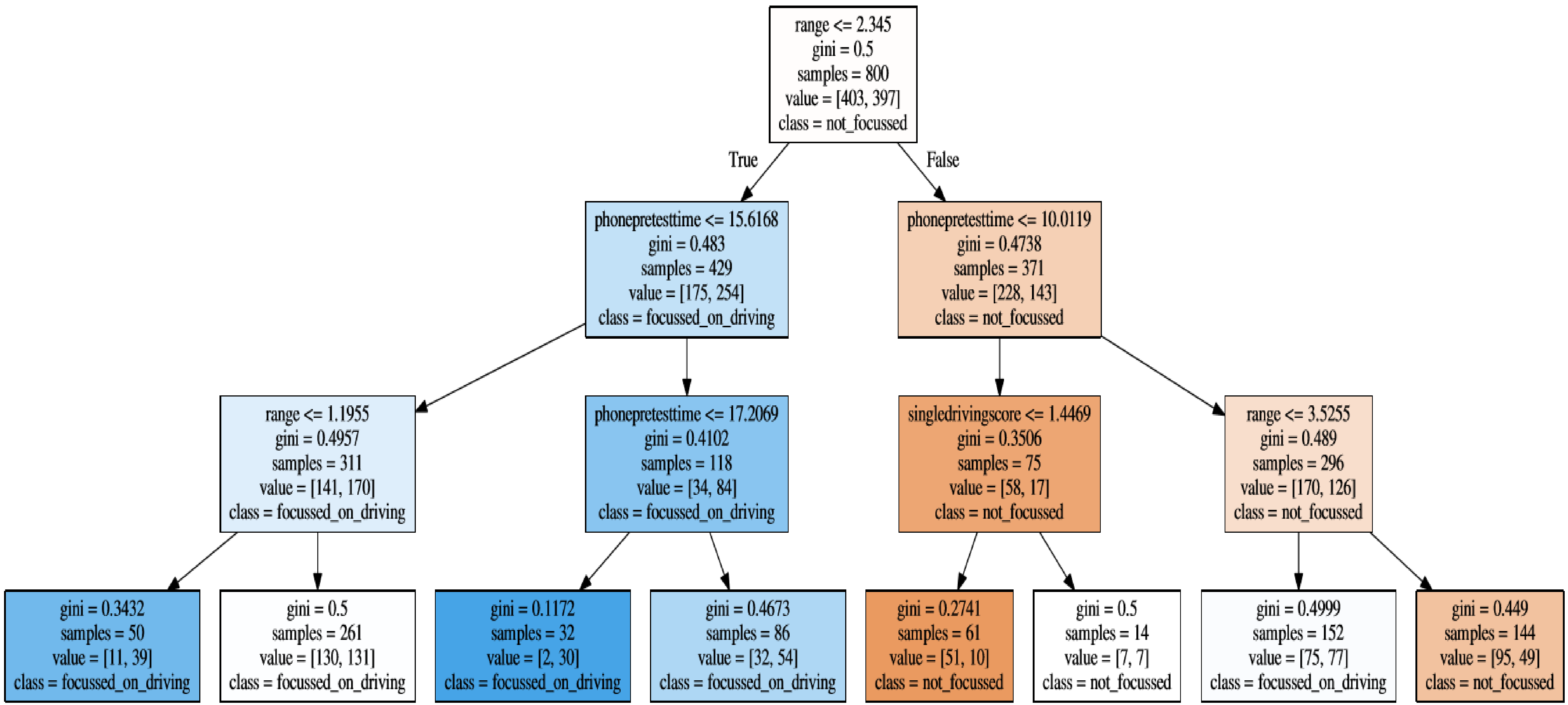
Ren, D., Zhou, H., & Fu, X. (2009). A deeper look at gender difference in multitasking: Gender-specific mechanism of cognitive control. In *2009 fifth international conference on natural computation* (Vol. 5, pp. 13–17).

Salvucci, D. D. (2006). Modeling driver behavior in a cognitive architecture. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *48*(2), 362–380.

Silverman, I., & Eals, M. (1992). Sex differences in spatial abilities: Evolutionary theory and data. In *Portions of this paper were presented at the meetings of the international society for human ethology in binghamton, ny, jun 1990, the human behavior and evolution society in los angeles, ca, aug 1990, and the european sociobiological society in prague, czechoslovakia, aug 1991.*

Stachowiak, K. (2015). On multitasking in interpreters and translators.

Webb, G. I., Pazzani, M. J., & Billsus, D. (2001). Machine learning for user modeling. *User modeling and user-adapted interaction*, *11*(1-2), 19–29.


range <= 2.345
gini = 0.5
samples = 800
value = [403, 397]
class = not_focussed
True
False
phonepretesttime <= 15.6168
gini = 0.483
samples = 429
value = [175, 254]
class = focussed_on_driving
phonepretesttime <= 10.0119
gini = 0.4738
samples = 371
value = [228, 143]
class = not_focussed
range <= 1.1955
gini = 0.4957
samples = 311
value = [141, 170]
class = focussed_on_driving
phonepretesttime <= 17.2069
gini = 0.4102
samples = 118
value = [34, 84]
class = focussed_on_driving
singledrivingscore <= 1.4469
gini = 0.3506
samples = 75
value = [58, 17]
class = not_focussed
range <= 3.5255
gini = 0.489
samples = 296
value = [170, 126]
class = not_focussed
gini = 0.3432
samples = 50
value = [11, 39]
class = focussed_on_driving
gini = 0.5
samples = 261
value = [130, 131]
class = focussed_on_driving
gini = 0.1172
samples = 32
value = [2, 30]
class = focussed_on_driving
gini = 0.4673
samples = 86
value = [32, 54]
class = focussed_on_driving
gini = 0.2741
samples = 61
value = [51, 10]
class = not_focussed
gini = 0.5
samples = 14
value = [7, 7]
class = not_focussed
gini = 0.4999
samples = 152
value = [75, 77]
class = focussed_on_driving
gini = 0.449
samples = 144
value = [95, 49]
class = not_focussed