

DESTINATION PREDICTION WITH DECISION TREE

AN OPPOSITION BY MARTIN AGFJORD *

OVERALL

Presentation -

Report - Well written, Interesting problem and solution. I think you have solved the problem in an intuitive and elegant way by combining different services in order to give informative notifications to the user.

RECOMMENDATIONS

Layout

The thesis is divided into six chapters, Introduction, Related Work, Design and Implementation, Testing and Discussion, Conclusion and Future Work. A person with experience from technical reports like myself can easily find what it wants in the thesis because of the structure.

I do however have one recommendation about reorganizing the thesis. You introduce the concept of the connected car and describe what other companies have done previously. Then you describe what you are going to do in your thesis, but without describing the problem that you want to solve. I recommend you to include a problem description in the introduction chapter.

In fact, you already have a section about this in the chapter *Design and Implementation*. This section is better suited in the introduction in order to catch an interested reader early.

* Master thesis student at Department of Computer Science & Engineering, University of Gothenburg

Decision Tree

When I read the section about Decision Tree, I had problem understanding how it works. However, after looking at the visual example presented 8 pages later, it was perfectly clear to me. I recommend you to move this example to the section where you first describe Decision Tree. Or maybe you have a specific reason why you included the visual example later?

ID3

I also had some problems understanding why the ID3-algorithm works, I had to read the section multiple times. I recommend a more detailed description, preferably with an example. I often find it easier to understand an algorithm if I can compare the text description with the algorithm in pseudocode. You include an algorithm which generates a decision tree 6 pages later, but it is unclear whether this is the ID3-algorithm or not.

You describe *why* the ID3-algorithm works after you described *how* it works. This is the part I found most confusing, probably because I do not have much background within probability theory and machine learning. A general introduction to machine learning would be useful, where you describe how machine learning build up a model based on training data and you can describe each term by term like, training data, test data, classification, class, gain-function.

Here is a list of recommendations:

- The first thing you want to describe is that this is an algorithm which creates a decision tree. It feels like you assume the reader knows this.
- Include a figure with the pseudocode of ID3, and talk about how the pseudocode works, rather than describing the algorithm with only words.
- What is **entropy theory**? What is **disorder** in a data set? Elaborate on these two terms.
- What is an information-based method?

Training- and testdata

Overall, you describe your design and testing with data very carefully, but I miss a description of how the training and test data was decided. Without this description, one can make the very bad assumption that you tested on the same data as you trained on which would yield unreliable results.

QUESTIONS

Design and Implementation

You describe that if there is logs that a driver has been starting trips at 8:30, 9:01 and 16:59, then the possible times to push notification will be 8:00, 9:00 and 16:00. Why do you have this rule? Is it not possible to send the notification 30 minutes earlier or an hour? Rather than sending the notification to the closest previous hour.

Do Volvo see any future in using the results?

Notifications in real time

You have developed an app to give real time notifications. Is the application meant to run on phones? Or is it meant to be built in the Volvo Car computer?

In case of phones, it is easy to achieve real-time notifications by using the GPS on the phone.

What was the hardest part?

What would you change if you could do the project again?

DETAILS

Typos

Abstract

can achieve higher 75% -> can achieve higher than 75%

Introduction

Page 1. helps from Ericsson -> help from Ericsson

Related Work

Page 3. faster than human -> faster than a human

Page 3. more precisely and helping avoid -> more precise and helps to avoid

Design and Implementation

Page 15. recursive function recursively -> recursive function which recursively

Page 19. 9:01 and 16:59 -> 9:01 and 16:59

Misunderstandings / Unclear

Related Work

Page 3. However, there is **rare** study beyond car itself that can provide suggestions and guidance to drivers before they get into trouble.

What do you mean?

Page 5. Both of these two approaches had made **appreciate** results.

Wrong word?