

Detection of human alertness using supervised learning

01666 Fagprojekt – 05-03-2011 – Anders Hørsted (s082382)

Vejledere: Morten Mørup og Ole Winther

Resumé

Giver et hurtigt overblik over indholdet og resultaterne af rapporten. Højst en tredjedel A4 side.

- Den behandlede problemstilling¹
- Anvendte metoder
- Hovedresultater

¹Sej fodnote

Forord

Forordets funktion er at give læseren de oplysninger, som har betydning for at forstå rapporten og dens kontekst, men som ikke har direkte med rapportens substans at gøre.

- Hensigten og formålet med rapporten
- Identifikation af målgruppe
- Projektets forhistorie og kontekst
- Anerkendelse af hjælp
- (Dato for aflevering – placeres normalt her)
- (Underskrift af alle forfattere – placeres normalt her)

Notation

Indholdsfortegnelse

1 Indledning	1
1.1 Problemformulering	1
2 Kort om kaggle.com og The Ford Challenge	2
3 Om machine learning, pattern recognition og alt det andet jazz	3
4 Dataanalyse	3
5 Resultater	3
6 Definitioner	4

1 Indledning

Igennem de sidste 30 år er computerens ydeevne vokset markant. Den forbedrede ydeevne har givet mulighed for at udføre statistisk dataanalyse, i et omfang der ikke tidligere har været muligt. En af de ting der er blevet mulighed for, er at programmere såkaldte "classifiers". Et godt eksempel på en classifier, er de programmer bankerne bruger til at opdage evt. snyd med kreditkort. Baseret på alle tidligere korttransaktioner, og viden om hvilke der var "falske" transaktioner, kan bankerne programmere en classifier, der med høj præcision kan forudsige om en ny transaktion er snyd eller ej.

Denne opgave tager udgangspunkt i en konkurrence på hjemmesiden kaggle.com. Konkurrencen er udbudt af Ford Motors og handler om at udvikle en classifier, der kan forudsige om en bilist er ved at blive ukoncentreret, mens han/hun kører bil. Til at udvikle denne classifier har Ford foretaget en række målinger på bilister, mens de kørte bil, og til hver måling er det blevet noteret om bilisten var opmærksom eller ej. Med udgangspunkt i disse målinger udarbejdes – ved brug af de mest gængse metoder – en række classifiers, og deres evne til at klassificere ny data sammenlignes.

1.1 Problemformulering

I denne opgave vil jeg...

- ... bruge de mest gængse klassifikationsmodeller (nearest neighbour, logistic regression, neural networks og SVM) til at lave en classifier der (forhåbentlig) kan forudsige om en bilist er ved at falde i søvn.
- ... undersøge hvor stor indflydelse den indledende databehandling (feature selection og outlier removal) har på det endelige resultat.
- ... undersøge om classifieren kan forbedres ved at implementere en Hidden Markov Model, der tager hensyn til det temporale aspekt af data.
- ... implementere en ensemble classifier, der kombinerer resultatet af flere classifiers i én classifier.

2 Kort om kaggle.com og The Ford Challenge

$$a^2 + b^2 = c^2 \tag{1}$$

3 Om machine learning, pattern recognition og alt det andet jazz

4 Dataanalyse

5 Resultater

6 Definitioner