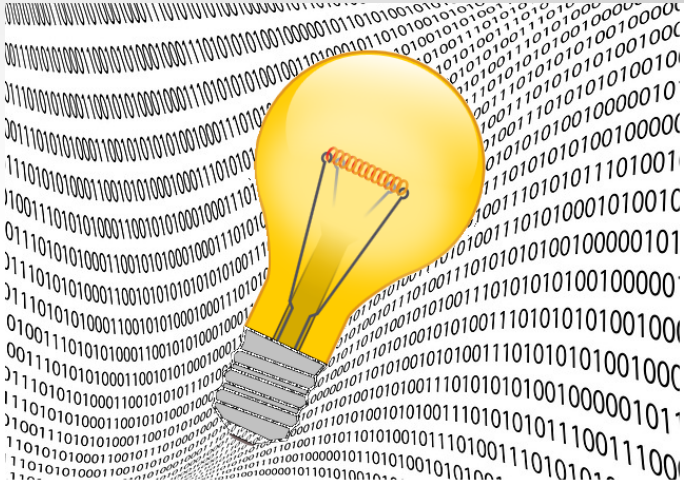
Data Mining on Crash Simulation Data



JONAS OORLYNCK & RICKY NATHVANI

Fraunhofer SCAI

GEORG-AUGUST-UNIVERSITÄT GÖTTINGEN

UNIVERSITEIT GENT

UNIVERSITY OF OXFORD

Overview

Using “Data Mining” to analyse car crash simulation data; what do we want?

What is Data mining?

Analysing car parts from the BMW manufacturing process and correlating them to the crash simulation data; will it work?

Data sets

Data Preparation

Data sets

Correlations

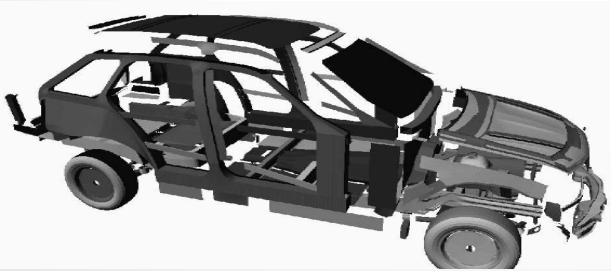
Analysis

Patterns

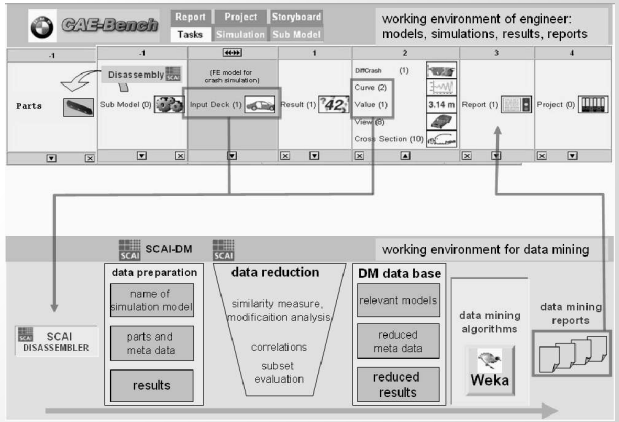
Setup

The question(s) ?

The data



The environment



Data Preparation

What?: convert raw data in data minable data and cleaning it up

Why?: data mining algorithms can't process data in its original format

⇨ data preparation = ***most time consuming process*** of data mining task

Example: Preparing crash simulation data

1) Export raw data from CAE-Bench
2) Computing meta data
3) Data cleaning and sorting
4) Similarity analysis and data reduction
5) Evaluation and final cleaning

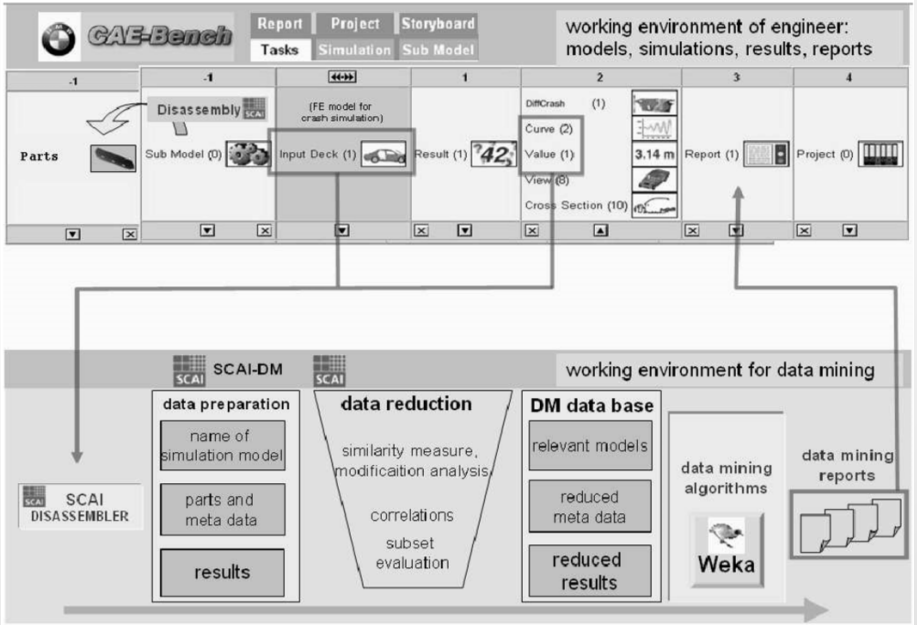
⇨ Results stored in tables for data mining!

1) Export of data from CAE-Bench

Export data from simulated crash test

Here: virtual car model after crash

CAE-Bench



2) Computing meta data

- Virtual car model = composition of FE-models of the individual car parts ⇨ ***Disassemble***
- Aim?: to analyse influence on crash behaviour due to changes in shape of car parts

What to use as meta data?

- FE-model of parts contains all *geometrical* data = hidden from data mining algorithms
- <u>Solution?:</u> define meta data such that it quantifies the geometry
- <u>Examples:</u> centre of gravity, moment of inertia, length of edges, surface size, etc.

⇨ All mesh independent, thus different meshing algorithms can be compared

3) Data cleaning and sorting

Cleaning: removing unnecessary data entrances

When?: During engineering process, parts are modified to test crash-worthiness. Unchanged parts can not be responsible for deviations in the simulation results (arrow) *Exclusion* from analysis!

<u>How?:</u> MD5-checksum for every mesh; filter out multiples of the same checksum (unchanged part)

⇨ Data mining tasks will then only include parts with multiple instances in the data base

4) Similarity analysis

- *Similarity measure* = geometric data of the different parts

- Parts are represented as a point in a multidimensional meta data space

⇨ Parts with similar meta data form clusters, which can be split up in sub-clusters through hierarchical clustering

- Here: weighted sum of meta data as single similarity measure

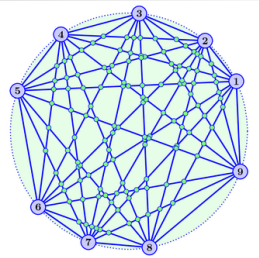
5) Evaluation

Last step:

A table is created containing 1 line of text per crash test containing the name of the model, similarity values of parts and result values of interest.

Data-mining on Similarity data

Attribute selection:

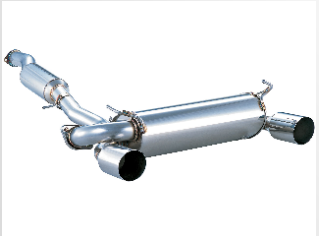


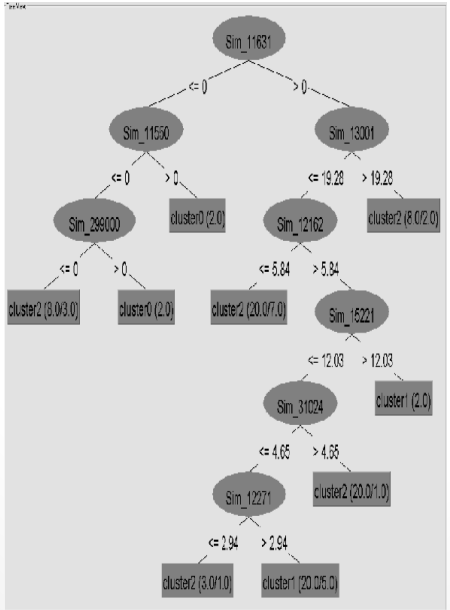
WEKA: Chi-Squared

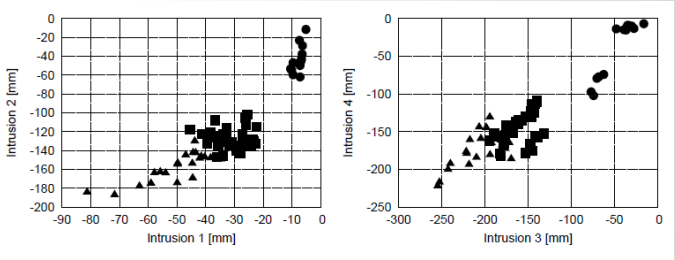
Model	Sim_8010	Sim_11011	Sim_11012	Sim_11013	Sim_11014	Sim_11021	Sim_11022	...	Intrusion 1	Intrusion 2
E6016VS01_fd01v02	0.13	3.11	3.11	0.91	0.91	0.53	0.53	...	-252.578	-217.411
E6016VS01_fd01v02a	0.13	3.11	3.11	0.91	0.91	0.53	0.53	...	-248.203	-217.2
e60vbg1+_fd02-1v06	0.0	1.13	1.0	4.08	4.02	0.01	0.01	...	-110.295	-114.728
e60vbg1+_fd02redv05	0.0	1.29	1.0	4.05	4.02	0.01	0.01	...	-191.374	-180.895
e60vbg1+_fd02v04	0.0	1.29	1.0	4.05	4.02	0.01	0.01	...	-118.985	-137.677
e60vbg1+_fd02v04b	0.0	1.29	1.0	4.05	4.02	0.01	0.01	...	-116.046	-126.97
e60vbg1+_fd02v07	0.0	1.13	1.0	4.08	4.02	0.01	0.01	...	-117.73	-123.492
e60vbg1+_fd03-1v09	0.0	1.37	1.37	3.53	3.54	0.35	0.35	...	-155.667	-169.425
e60vbg1+_fd03-2v10	0.0	1.37	1.37	3.53	3.54	0.35	0.35	...	-139.427	-145.34
e60vbg1+_fd03-2v10a	0.0	1.37	1.37	3.53	3.54	0.35	0.35	...	-148.988	-163.872
e60vbg1+_fd03-2v10b	0.0	1.37	1.37	3.53	3.54	0.35	0.35	...	-156.212	-170.325
e60vbg1+_fd03-3av12	0.0	1.37	1.37	3.53	3.54	0.35	0.35	...	-148.141	-146.928
e60vbg1+_fd03-3bv12b	0.0	1.37	1.37	3.53	3.54	0.35	0.35	...	-142.861	-162.149
e60vbg1+_fd03-3v11	0.0	1.37	1.37	3.53	3.54	0.35	0.35	...	-146.973	-153.957
e60vbg1+_fd03-4v13	0.0	1.37	1.37	3.53	3.54	0.35	0.35	...	-142.953	-145.71
e60vbg1+_fd03-5v02	0.0	1.37	1.37	3.53	3.54	0.35	0.35	...	-158.559	-166.717
e60vbg1+_fd03v08	0.0	1.37	1.37	3.53	3.54	0.35	0.35	...	-117.151	-126.494
e60vbg1+_fd04v14a	0.0	1.37	1.37	3.53	3.54	0.35	0.35	...	-148.246	-173.806

Data-mining on Similarity data

Decision Trees:

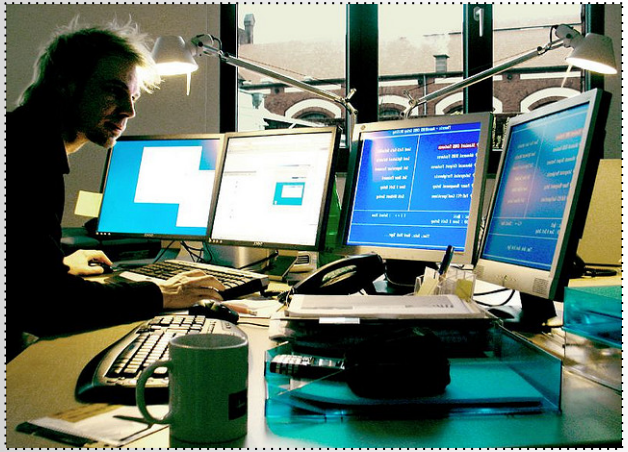


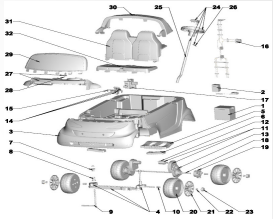


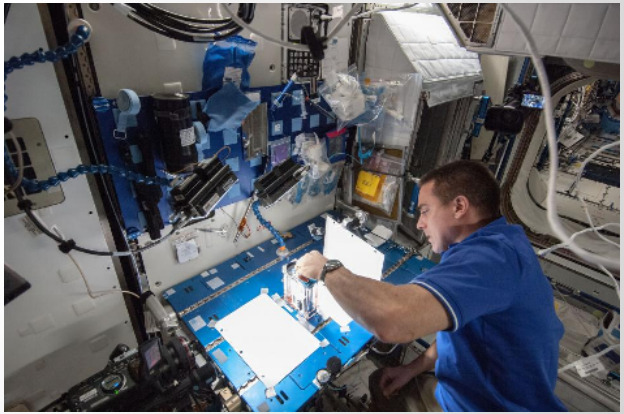


Data-mining on Similarity data

Data-mining Reports







Conclusions

Framework in place: the method has been demonstrated to be *applicable* to car crash data

Used in observations of model changes. Next step is to integrate into the CAE-Bench; the engineers working environment

Need more substantial data sets to check the effectiveness of the approach...the future is bright!

Thank you for your attention!

Any Questions?