Engineered Decision Tree for Judging Spacecraft Collision Risks

Sven Rebhan, Nils Einecke, Viktor Losing, Steffen Limmer, Sebastian Schmitt

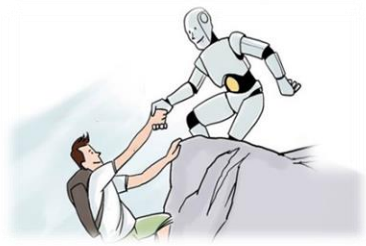
Honda Research Institute Europe GmbH, Offenbach, Germany

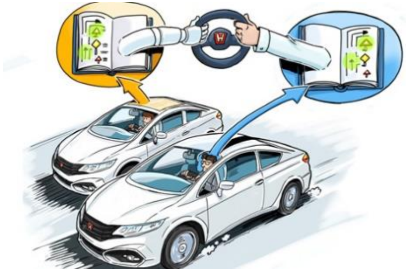
Side-Workshop -- 8th European Conference on Space Debris

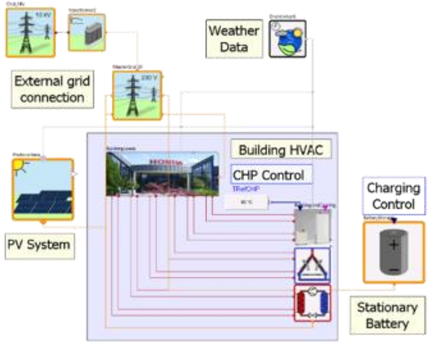
23.04.2020

Honda Research Institute Europe

- Founded in 2003
- Fundamental research for Honda
- Three sites:
  Tokyo (Japan), San Jose (US), Offenbach (Germany)
- Research Fields:
  Cooperative Intelligence, Learning, Personalization,
  Data Analytics, System Optimization, Cooperative Engineering, **Energy Management**, System Architecture,
  Risk and Planning, Perception and Prediction



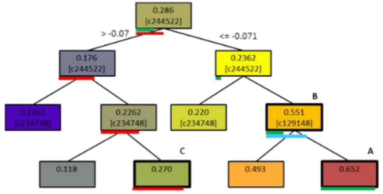




Starting Point

- Test HRI algorithms and expertise in different domains
  - Check generality
  - Learn about approaches in other domains

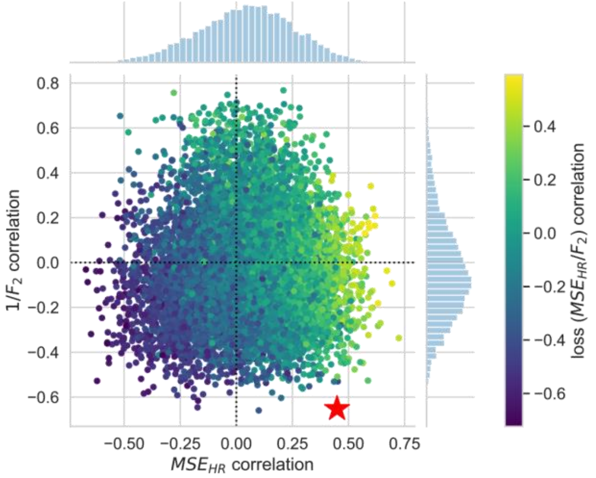
- Started to apply wide range of approaches to the problem
  - Extraction of time series features, machine learning, random forests, statistical analysis and manual data engineering
  - Side note: approaches for risk analysis in traffic scenarios could not be applied → based on assumption like known street layout
  - All approaches managed to beat the fixed-value baseline
  - Best ML approach achieved overall score of 0.83 (final: 0.555)

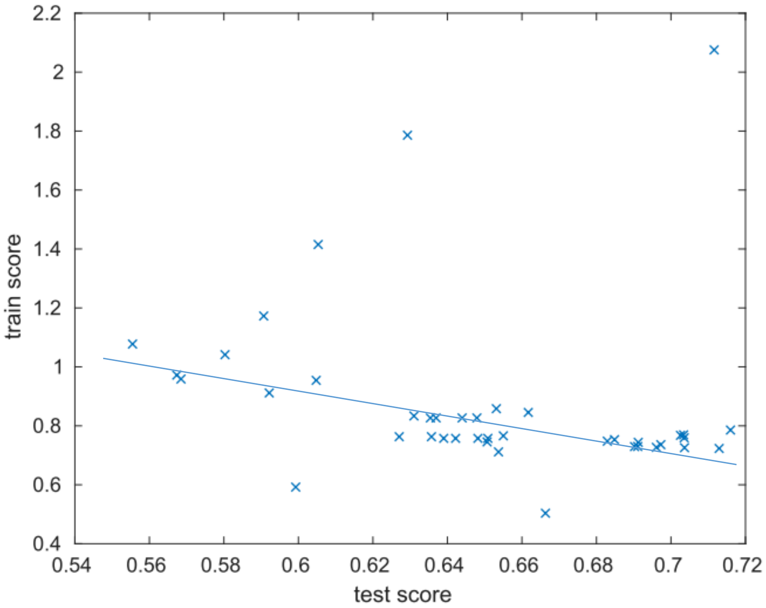


- Pushing ML approaches further on train set led to performance drop on tests set

Anti-Correlation train-test for F_2 score

- After getting below 1.0 score further advancements led to a suspicious anti-correlation between train and test results
- Further advances with ML approaches were very difficult
- Post-analysis of challenge team revealed:
  - Manual selection of high-risk events changed set characteristics
  - Unluckily a very unfavorable test selection was done



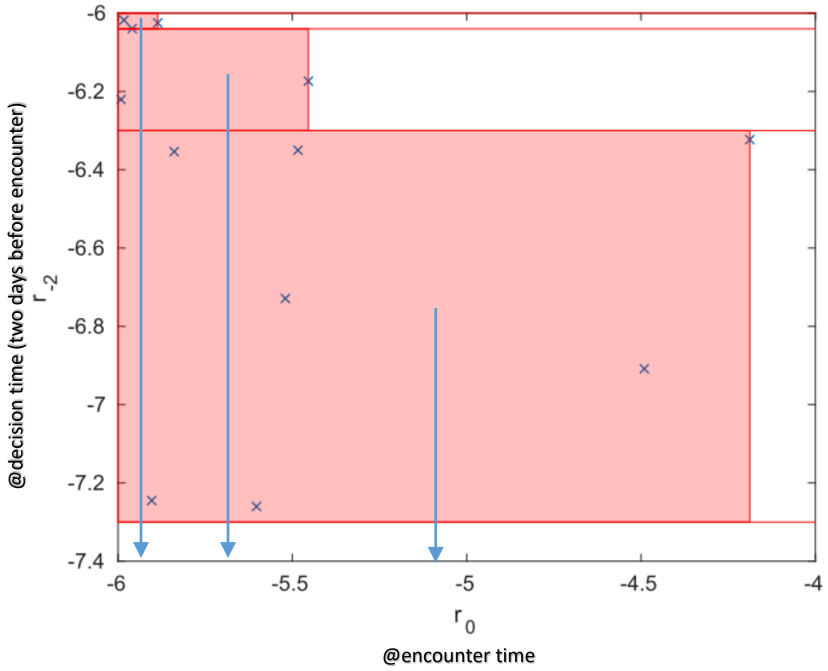


Major Findings in Train Data

1. Events slightly above high-low threshold @decision time tend to move to high risk at encounter
→ set close-threshold risks to high risk in a cascade

$$r_0 = \begin{cases} -5.10, & \text{if } -7.30 \leq r_{-2} < -6.40 \\ -5.60, & \text{if } -6.40 \leq r_{-2} < -6.04 \\ -5.95, & \text{if } -6.04 \leq r_{-2} < -6.00 \end{cases}$$

2. Very high-risk events @decision time tend to get less risky at encounter
→ clip high risks



Final Decision Tree

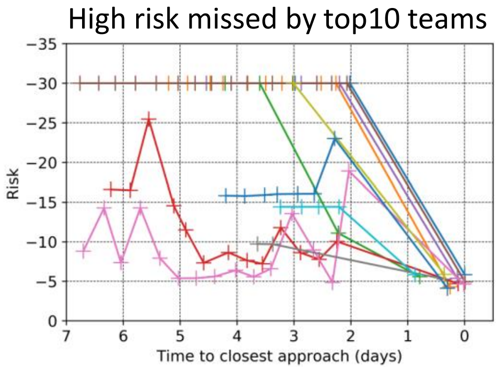
1. Naïve forecast (last risk prediction)
   - $r_0 = r_{-2}$
2. c_object_type
   - type==“payload”, “rocket body”, “tba” → r_0=-5.6
3. t_span
   - span < 0.5 → r_0=-6.00001
4. Clip high risks
   - r_{-2} > -3.5 → r_0=-3.5
5. Risk cascade
   - Slide 5
6. miss_distance
   - dist > 30,000m → r_0=-6.00001

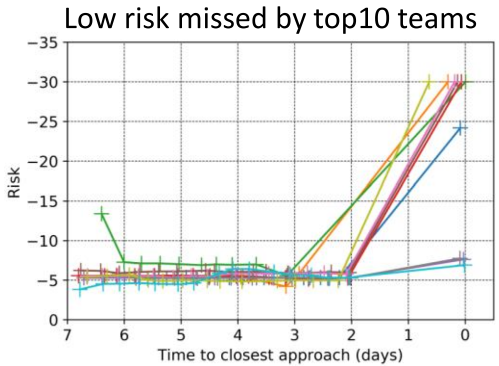
Main improvements achieved by <u>risk</u> feature itself!

	leaderboard	final		
technique	score LB	score GT	MSE	F_2
baseline	2.502	2.504	0.679	0.271
naive forecast	0.703	0.681	0.513	0.753
c_object_type	0.685	0.664	0.492	0.742
t_span	0.683	0.670	0.495	0.739
low to high risks	0.648	0.638	0.481	0.754
clip highest risks	0.636	0.635	0.479	0.754
low risk cascade	0.568	0.562	0.411	0.732
miss_distance	0.555	0.555	0.407	0.733

Lessons Learned

- ML on collision avoidance dataset is very difficult
- Risk (r_{-2}) value is most significant feature
- Other features seem to have almost no impact
- Time series seem to have very little effect
  → some encounter have strong risk jumps
  without prior indication





- Post-challenge analysis revealed: decision tree does not generalize well
  → reason I: statistical difference of test and train set
  → reason II: no train-validate splitting for manual engineering
- Insights from decision tree are less informative
- Single steps might improve prediction but require different parameters

Statistical analysis on random splits of all available data

approach	score	F2	MSR
Last risk prediction	0.762	0.60	0.46
Decision tree	0.745	0.51	0.38
Decision tree (tuned)	0.699	0.60	0.42

Challenge Insights

- Space collision avoidance shares dilemma with autonomous driving
  → there is almost not data of real impact events
- All events with maneuver actions were removed for challenge
  → only non-critical events are left
  (critical events would have an avoidance maneuver and are thus removed)

- Suggestions:
  - gather more data
  - find ways to include maneuver events
  - change approach to predict expert decision

- Multi-objective score might require different leaderboard
  → it might be favorable to use a ranking scheme instead
  - Algorithms are ranked for each objective
  - Algorithm performance is the mean rank

Middlebury stereo benchmark 2012

Algorithm	Avg. Rank	Tsukuba ground truth nonocc	Tsukuba all	Tsukuba disc	Venus ground truth nonocc	Venus all	Venus disc	Average percent of bad pixels (explanation)
PMF [119]	12.5	11.0 39	11.4 36	16.0 32	0.72 8	0.92 7	5.27 7	7.69
SegAggr [144]	15.8	12.4 60	12.9 54	17.3 51	**0.28** 1	**0.41** 1	**2.09** 1	7.49
LAMC-DSM [123]	17.1	9.34 29	10.1 30	13.5 10	1.48 17	2.10 16	8.19 23	9.20
PM-Forest [162]	18.8	11.1 41	11.8 43	17.3 52	3.11 30	3.14 23	4.57 4	7.80