Using Random Forest to estimate risk profiles and probability of breakdowns

\-

Lara Andrea Neira Gonzalez

Introduction

- Industrial machines face different types of risks and are usually highly complex assets
- Difficulty to assess risk often results from incomplete knowledge about operating conditions and risk influencing factors
- Are there any forward looking risk indicator that can be used for the estimation of riskiness of the machine?

Why Machine Learning?

- We are in the age of “Big Data”
- How can we build models explaining more variation and allow for better interpretation?
  - ➢ Machine learning is one way to efficiently examine the search space

Random Forest

- Supervised machine learning technique which estimates variable importance measures (VIMs) associated with outcome

- Random Forest is able to detect single and interaction effects because of its own architecture.

Data and study design

- 32 variables from a Yankee dryer in a paper plant ("paper machine")
- 6 variables from a evaporation station from a sugar plant ("sugar machine")
- The study design for both studies was to use a training set with 2/3 of cases/controls, and independent test sample with a 1/3

Methods

1. We performed RF based on the VIM AUC
2. Obtain empirical p-values for each variable
3. Take the empirical significant variables
4. Test for single effects from the empirical significant variables on our independent test dataset using Likelihood ratio tests from nested models based on general linear regressions.

Methods

5. Fit a model with all significant validated single effects
6. Test the AUC in the independent test

Results

Sugar machine

- All variables from the sugar machine were empirically significant

- After testing for the single effects and looking for significance after Bonferroni correction (90 %, 95% and 99% confidence levels) :

  - 30 variables were significant
    - 29 at 99% confidence level
    - 1 at 90% confidence level

Results

Sugar machine

Variable	Description	Significance	Explanatory impact	Way of impact	Mean
TC51_05	Juice temperature (1st evaporator inlet)	***	13.64	Risky	128.9
P57_04	P2 - juice pressure (valve outlet)	***	7.38	Protective	165.9
P74_01	P2 - water pressure (valve outlet)	***	7.19	Risky	2420.1
T51_08	Juice temperature (1st evaporator outlet)	***	7.11	Risky	133.1
T57_03	T - juice temperature (valve inlet)	***	6.98	Risky	94.2

- After fitting the model with all 30 significant validated variables, we got in an independent database an:

AUC = 0.83

Results

Paper machine

- All variables from the sugar machine were empirically significant

- After testing for the single effects and looking for significance after Bonferroni correction (90 %, 95% and 99% confidence levels) :

  - 5 variables were significant
    - 3 at 99% confidence level
    - 1 at 95% confidence level
    - 1 at 90% confidence level

Results

Paper machine

Variable	Description	Significance	Explanatory impact	Way of impact	Mean
throttle.position	throttle.position	Unessential	1.1	Protective	179.93
gas.mixer.position	gas.mixer.position	***	85.93	Risky	10309.91
turbo.position	turbo.position	***	3.95	Risky	289.94
turbo.setpoint	turbo.setpoint	***	3.62	Risky	296.37
U.gen.L1L2	U.gen.L1L2	**	2.42	Protective	304.89
speed	speed	*	2.24	Risky	13934.12

- After fitting the model with only the most significant variable, which have a $R^2 = 85.93$, the model reached in an independent database an:

AUC = 1

Discussion

- Both our studies found significant factors helping us to better understand the “mechanics of the failure” of the two machines and enable better prevention
- Our model predicts breakdowns better than any human and thus can help in decision making

Discussion

- Our findings can be used for creating risk profiles as our models could detect different significant factors which are playing an important role in the process of each machine

- The models can be used as a basis for pricing of insurances and managing of guarantees and warrantees for machine producer

Thank you

Our Team

Lara A. Neira Gonzalez

Alfredo D. Egídio dos Reis

Martin Kreer

José-María Guerra

Inrobin

LISBON SCHOOL OF ECONOMICS & MANAGEMENT
UNIVERSIDADE DE LISBOA

THE UNIVERSITY of EDINBURGH