Estimation of bearing wear based on consistence-oriented modelling

László Pitlik, Marcell Pitlik, László Pitlik (jun) – MY-X team

Abstract

There is data asset about bearing wear where the exposure dynamic of the bearings is arbitrary different in case-to-case-situations like in the real life. The noises of the bearings have been measured (Xi). The realisation time (Y) of noise-level of 7 starting from a given timestamp should be modelled based on optimized staircase functions with a doubled attribute-set. The correlation between facts and estimations: 0.98. The simple attribute-set contains 9 variables (5 variable of the absolute noises: like sum, max, min, std. deviation, variance + 4 variables of the noise-differences between two neighboured time unit: sum, max, std. deviation, variance) – in each case with a ranking transformation where the higher level has the better ranking number. The doubled set integrate the inverse ranking values of the 9 variables (c.f. based on the-lower-the-better-principle. There are 10 bearings with full time measurements and a 11th bearing with only 300 time-units. The objects of the learning patterns are time layers of 50 time-units (it means 6\*50 for each full-time-measured bearing – see ABCDEF-layers). Therefore, the number of the objects: 60+6 and the Y-values have a special pattern for the 6 layers of a particular bearing: the difference of the values of consequence variables for one bearing is always 50 time-units. The above-mentioned basic model using 60 learning objects and with the correlation value of 0.98 leads to a correlation between the estimated pattern and the ideal pattern of monotonous 50 units to a correlation of 0.97 in cases of the 6 estimation objects. This is a kind of massive consistence where the basic model with 10 bearing could not be tested before. The estimation for the 11th bearing is 500 for realisation of the noise level of 7. The average (557) and the median (530) of the 10 bearings are higher than this basic estimation. The statistical and anti-discriminative derivation of the best estimation layers (see ABCDEF) let identify the E-layer as the best one (estimation value lower than 500 – 446+21). The naïve and second-best layer is layer C (estimation value 514+5). The last correlation layer had to support the decision between lower or higher than 500. In order to derive a new and independent approach, each (66) objects have been integrated into a learning model-series where the Y-values of the last 6 objects (of the 11th bearing) were simulated values with the special pattern of 50-unit-differences from 250 to 750 (step 50). All these learning phases presented estimation values where a new anti-discriminative analysis could be executed. The 11 objects (simulated Y-value-sets) have been compared to each other based on 6 attributes derived from the estimated values of the partial simulated Y-patterns: error^2 for the last 6 objects compared to the ideal pattern, error^2 for the first 60 objects, max, min, std. deviation and variance for the 66 objects. The ranking of the simulated situation (c.f. sensitivity analysis) let interpret two objects as identical good: the set of 500 and the set of 550. This last argumentation in the consistence chain leads to the conclusion, the 11th bearing should be capable of working further 200-250 time-units after the already realized 300 time-units. This method of the estimation (forecasting) of bearing wear does not use any testing methods. Instead of testing, the quality assurance of the modelling can be ensured based on parallel argumentation with high fitting to each other. Therefore, it is possible to model without testing based on consistence-effects!

Keywords

similarity analysis, modelling without testing, multi-layered consistence checking, modelling of time (-changes), AI-based term creation, anti-discriminative modelling

# INTRODUCTION

## The targeted tasks

There is a database containing real noise measurements regarding the wear of a bearing in an experimental setting. The load applied during the experiments is considered to be random. Using the aforementioned database, the following modelling problem is to be solved: what is the time (Y) left before reaching the noise level that indicates the need for replacing the bearing? The independent variables (X) are descriptive statistics ordered / ranked both ascending and descending to facilitate domesticated polynomization features of staircase functions. This method aims to generate consistence layers that may be able to validate and support the final conclusion without testing.

## Aims

The main goal of this paper is to demonstrate how consistence-based modelling (without testing) can be executed. Also, the estimation of critical wear (when a replacement is imminent) of bearings based on noise detection data is a practical task to be solved.

## Target groups

This paper is a (context-free) case study for all who are interested in novel modelling opportunities. The owner of the original dataset may gain a forecast (regarding the need for placement of the bearings) in higher quality and reliability than previously learn-and-test modelling approaches were able to grant. The similarity analyses used in the modelling process provide special layers of quality: symmetrical functions stand for the validity of the results and double attribute sets used in the modelling process allow staircase functions to exhibit polynomial behaviour and better represent optimum-like relations. The anti-discriminative optimization process bypasses all subjective influences of scoring and weighing systems.

## Usefulness

Modelling without testing is able to process (learn) all provided data without loss of information. The high quality of the evaluation is maintained by the automated and objective way it can be done: when the consistence and validity of an estimated answer is questionable, the "robot" is able to use the "i do not know" answer to differentiate between the potential output scenarios.

## Motivation

The default challenge in artificial intelligence research is to fulfil the KNUTH-principle: cf. [https://miau.my-x.hu/miau2009/index\_tki.php3?\_filterText0=\*knuth](https://miau.my-x.hu/miau2009/index_tki.php3?_filterText0=*knuth). Also, as genetic algorithms indicate, novel modelling (search optimization) approaches are needed regarding both the representation of knowledge (staircase functions exhibiting domesticated polinomization features) and the hermeneutics of the results (consistence-based results).

# LITERATURE BACKGROUND

This time, the chapter is rather symbolic: it is important to emphasize that similar problems were addressed using neural network modelling, but the consistence-based approach was not yet reported. This later method is, however, covered extensively in other publications by members of the My-X Team.

## Bearing degradation

* Prognosis of Bearing and Gear Wears Using Convolutional Neural Network with Hybrid Loss Function: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7349655/>

## Consistence-based modelling

* Consistence-layers of similarity analyses: [https://miau.my-x.hu/myx-free/index.php3?\_filterText2=\*konziszten](https://miau.my-x.hu/myx-free/index.php3?_filterText2=*konziszten)
* Global and partial aspects of consistence regarding time series: <http://miau.gau.hu/miau/53/autotrend.doc>
* Space/time in consistence-based modelling: <http://miau.my-x.hu/miau/82/kjm_hu_ecology_full.doc>
* Consistence of multi layered modelling: <http://miau.my-x.hu/miau/111/chf30.doc>
* Modelling without testing: <https://miau.my-x.hu/miau2009/index.php3?x=e159>
* Further reading: <https://miau.my-x.hu/miau2009/index.php3?x=e0&string=konziszten> & <https://www.google.com/search?q=%22tesztel%C3%A9s+n%C3%A9lk%C3%BCl%22+modellez%C3%A9s+site%3Amiau.my-x.hu>

# DATA AND METHODS

Details (source-XLS): <https://miau.my-x.hu/miau/280/degradation_path_01.xlsx>

The following figures tables and diagrams are all exported from the above-mentioned spreadsheet file. The detailed descriptions given in this document are to facilitate the reproducibility of various layers of the XLS file.

## Original dataset



FIGURE NO.1 Characteristics of the 10+1 cases. X-axis is time, Y-axis is noise levels, 7 is the threshold value. (source: Worksheet OAM\_nyrs)



FIGURE NO.2 Differential view of the 10+1 cases. X-axis is time, Y-axis is noise level differences. (source: Worksheet OAM\_nyrs)

Based on Figures 1 and 2:

* If only one "case" (reaching noise level 7) is studied, the evaluation must be able to give an estimate regarding the next occurrence, cf. JOKER and the similarity of similarities.

See: <https://miau.my-x.hu/miau/280/joker_min_2.xlsx>, also <https://miau.my-x.hu/miau/280/joker_min.xlsx>, also <https://miau.my-x.hu/miau2009/index.php3?x=e0&string=joker> This also means that modelling without testing is not an irrational challenge -- but one where the expected accuracy must be carefully examined. Possible tools:

* + direct / inverse proportionalities can always generate estimations
	+ when evaluating more cases, mean and expected values can be obtained
	+ using similarities is a more complex approach than descriptive statistics
* Further questions, instead of the estimation when the noise level 7 will be exceeded:
	+ will the next reach to noise level 7 happen sooner or later than the average so far? (median?)
	+ which previous experiment has the strongest resemblance to the 11th (not finished) noise levels?
	+ is the original noise level graph or the differential view more suitable to estimate the expected values?

## Preparing the OAM

The object-attribute matrix (OAM) can be generated from the original data -- to do so, in many cases, it requires the experience and best impulses of the person who is controlling the evaluation. On the other hand, the principles and methods in the background must always be as objective and systematic as possible.

* The objects are the cases (10+1), all reduced to 304 timeframes (the amount of data available for the 11th scenario)
	+ In other perspective, objects could be the "slices" of the original experiments per 50 timeframes each
	+ also, the cumulative data, in each iteration a 50-frame longer case is evaluated
	+ (and, of course, the limit does not have to 50 timeframe, or even a fixed interval)
* The attributes can be the descriptive statistics of the previously described objects. Also, aggregated data can be used. When object "slicing" is used, the data can be absolute or differential, e.g.
	+ mean
	+ max
	+ min
	+ standard deviation
	+ variance, (etc.)

The original OAM can be found on Worksheet OAM\_nyrs. The final set of attributes may seem an arbitrary (intuitive) choice, but another possibility would be to use a combinatorically generated much larger set (cf. brute force techniques) where even the sums and products of various attribute pairs (triplets etc.) are included. It is a common misunderstanding that attributes that show a high correlation value with the Y values are worth more (cf. <https://miau.my-x.hu/miau/274/real_values_of_attributes.docx>). Similarly, the more data does not necessarily and always lead to the better results (cf. <https://miau.my-x.hu/miau/phd/Barta_Gergo_Doktori_ertekezes_tezisei_0604.pdf>).

When the attribute set (Xi values) is fixed, the potential values of Y are also limited to some extent. In this particular case, the ABCDEF layers are used to form the objects and thus, the Y values are the time differences between the last timeframe used and the noise level 7 reached. For the 10 original "full" experiments, it also gives a constraint for the values of Y: they are "shifted" by 50 timeframes each. When the AI learns the pattern, it will used as an important source for improving consistence, cf. Figure 3.



FIGURE NO.3 Steps and results of the preparation of the OAM. Everything is measured on the "noise scale" except Y which is the timeframes left to reach noise level 7. (Source: Worksheet OAM\_nyrs)

## Modelling

First (cf. Figures 3 and 4), a "complete" OAM is to be shaped: superfluous rows and columns must be excluded.



FIGURE NO.4 The raw OAM containing the original data. Some rows are excluded. The X variables have the original noise units, the Y (dependent) variable gives the number of time frames before reaching the critical (7) noise level. (source: Worksheet OAM\_dpl)



FIGURE NO.5 The OAM with the doubled set of attributes -- and with results. Some rows are excluded. The X variables are ranked, the Y (dependent) variable uses a linear transformation of its original value. (source: Worksheets OAM\_all\_cases and OAM\_dpl, from Row 73)

The linear transformation of the Y values is a reaction to the characteristics of the LP engine. Since there are only positive values permitted, the additive part of the transformation allows the model to under-estimate the original values. The multiplicative part facilitates the finer "resolution" of the information in the situation. The optimal transformation factors were determined empirically since the desired way to operate the robot / AI is when it faces no mathematical restraints producing the estimated values. An estimation is considered optimal if the sum of the original Y values and the sum of the estimated values are the same. Also, the high correlation between the previous two datasets is expected along with a high validity.

Regarding the online AI engine used in this task, the high validity means minimal symmetry disturbance in the staircase functions. Under standard modelling conditions, this means that the object with the "best" results should "lose" a contest when the ranking methods of the raw data are set upside-down. When a doubled set of attributes is used, both the direct and the inverse layers are parts of the same analysis and thus, this interpretation of the validity is disregarded.

The use of a doubled attribute set is an answer to the following problem: neural networks tend to give estimations in higher-order polynomial expressions which do not show the experimentally expectable monotonous behaviour. Using the direct and inverted attributes in the same model allows a flexible mathematical representation while reducing the polynomial patterns (cf. interpretable ANN).

There are two models with doubled attribute sets:

* one with 10x6 = 60 objects (rows), the "complete" cases;
* another one with 11x6 = 66 objects. The latter uses all the complete cases and the six known layers of the 11th experiment.

Thus, ranking goes from 1 to 60 or 66, depending on the models -- for the 11th experiment, the question is, how many values are greater (among the original 60 values) than the one at hand. It is worth mentioning that the above test-ranking automatically allows the complete 2^n combinatorial possibilities (cf. <https://miau.my-x.hu/miau/256/bme_mnb_pm_pl.docx>) where ’*n*’ is the number of attributes used.

The OAM and the results are in the XLS file (see sources). Worksheet modell1 contains the results where the correlation coefficient value is above .98 (cf. Figure 5). Worksheet “OAM-dpl” contains (starting from Row 140) the estimates for the 11th experiment based on the previous 10x6 objects used in the "learning" phase.

# RESULTS

## Layer No.1

Although further steps enhancing the consistence of the results may be introduced, the first result is given (cf. Worksheet OAM-dpl, bottom right corner) as after the 304 timeframes given in the dataset, another 200 is expected before reaching noise level 7 in the 11th experiment.

This estimation is supported by the 0.98 correlation value between the original and estimated data -- but not yet tested. This is an immediate usage of the model, and the exceptionally high correlation value may be an indicator of the risk of overlearning. Further consistence-calculations are needed in order to regain the possibility of the "i do not know" answers where the double attribute set staircase functions can not be validated using the symmetrical features.

## Layer No.2

Statistically speaking, the most robust estimation layer is "C" where the sum of the error values is minimal.



FIGURE NO.6 Modelling errors for each object and layer (ABCDEF). All units are time differences between estimated and original values of reaching noise level 7. (Source: Worksheets hiba1 and stat)

Figure 6 can be also interpreted differently:

* the "errors squared" criterion is best matched in layer "E" (significance?)
* an anti-discriminative similarity analysis arises when evaluating multiple criteria at once:
	+ minimal sum of errors
	+ minimal max of errors
	+ minimal security risk (i.e.: being closer to the noise level 7 -- since layer "A" is the first in a time series and this has some, even philosophical, implications on how the life of the bearing can be estimated at the very beginning of it)
	+ minimal difference between mean and median values
* The anti-discriminative similarity analysis favours layer "E"
* second place is granted to layer "C" (the "naive" winner)

Even a third layer of consistence must be determined since the estimations of layer "E" (446 and 467, after compensation) are contracting the final layer "F" estimation of 500. Also, layer "C" (514 and 519, after compensation) exhibits a difference in the opposite direction.

## Layer No.3

A complex series of assessments is used since the 10+1 experiments can be handled without the distinction between the complete and incomplete original scenarios -- the six layers of the 11th experiment can be computationally completed. Learning from this expanded dataset can be done in numerous simulated ways and the accuracy of the learning procedure can be optimized using multi-layered anti-discrimination similarity analyses.

- Worksheet “OAM\_all\_cases” shows a version where the 11th experiment is completed using the centred values of the previous estimates (final Y value of 500, and 50 subtracted in each layer to reach 450-400-350-300-250-200)

* Worksheet modell2 is itself a layer of consistence since it emphasizes that the last six objects are in the center of the above sixty.
* Worksheets “all\_i\_of\_n (\*)” are the inputs of the versions.
* Worksheets “modell\_i\_of\_n” show the corresponding results.
* Worksheet “becslesek” aggregates the modelling errors and introduces numerous final assessment attributes:
	+ 11negyzet: differences between the errors squared and the ideal results regarding the 11th experiment
	+ 1\_10negyzet: errors squared for the experiments 1-10 and their six layers each
	+ Max: highest error of the 66 cases
	+ Min: lowest error of the 66 cases
	+ Szórás: standard dev. of the 66 errors
	+ Variancia: variance of the 66 errors
* - The final results of the anti-discriminative modelling process is depicted in Figure 7.



FIGURE NO.7 Exploring the ideal intervals (source: OAM-simulation)

# Discussion

Consistence-based modelling may exclude controversial, indeterministic behaviors in an ideal situation (cf. <https://miau.my-x.hu/miau/279/tdk_zaj/TDK_PM_zaj_final.pdf> and <https://miau.my-x.hu/miau/277/szakdoga_PM_final.pdf>). Otherwise, the potential results may only be partially supporting each other. In some cases, the AI has to be able to express an "I do not know" hermeneutical decision.

Results are not necessarily black-and-white: some layers of the modelling indicate a 467 (time frames left) value which would be below the "central" 500. Also, cf. Figure 7, there are more aspects that align for an estimated value between 500 and 550, see the 519 marks.

# Conclusions

Consistence-based modelling may lead to context-dependent (e.g.: ABCDEF) and context-free (validity) layers of consistency. The aggregated result may be the "I do not know" answer of the AI / robot. There is also an "expected" value of the result, the one with the highest probability -- although this is rather an interval not a single value. Also, the method to this estimation is different from the typical probabilistic approach: a complex logical evaluation is used with no learning / testing phases in a situation where prompt results are expected -- not only a technical possibility but a "rational", the most rational conclusion available.