AUTOMATED MODEL FITTING FOR METALLURGICAL AND OTHER INDUSTRIES

GAUDE-FUGAROLAS Daniel

dgaude Prime Innovation SLU, Vilassar de Mar (Spain), EU, dgaude@cantab.net

Abstract

One of the aspects defining the paradigm sometimes called Industry 4.0 is the availability of large amounts of data generated by modern processes. Nowadays, most process parameters can be collected automatically and therefore, the large volume of process data generated becomes close to impossible to analyse manually, or at least, to be able to extract from it all the useful information that it contains.

Advanced mathematical and computational tools to analyse data and create models automatically have existed for some time, being labeled at one time or another as Expert Systems, Artificial Intelligence, Machine Learning, Data Mining, Statistical Pattern Recognition etc. However, only recently the combination of automated data collection and fast computers have allowed to extend the benefits of such techniques beyond some niche applications and industries.

This work introduces a new project to develop a comprehensive set of such tools aimed to fit industrial needs. As starter, several advanced model fitting techniques are applied to a metallurgical example. As dataset, a publicly available database on creep rupture is used. Supervised learning algorithms (using gradient descent training, normal equation regression, artificial neural networks), when correctly applied, allow to fit models to any set of complex data and to make reliable predictions with them. Unsupervised learning methods on the other hand, may also be used to find structure in the data without any a priori knowledge of such underlying structure (anomaly detection, clustering).

Keywords: Automated Model Fitting, Regression, Artificial Neural Networks, Anomaly Detection, Cluster Analysis

1. INTRODUCTION

Although scientific research has progressed greatly in understanding the underlying phenomena on many material behaviour and processing, there remain many situations where accurate quantitative descriptions are still lacking.

The lack of progress in predicting the precise outcome of an industrial process, for instance, is related to its dependence on large numbers of variables, that although usually well understood individually, become untractable when the whole process is considered.

Moreover, many process parameters can be collected automatically and therefore, the large volume of process data generated becomes close to impossible to analyse manually, or at least, to be able to extract from it all the useful information that it contains.

Advanced mathematical and computational tools to analyse data and create models automatically have allowed to extend the benefits of such techniques into many industries. Supervised learning tools, like Regression and Classification to produce predictive models, as well as unsupervised learning tools like Anomaly Prediction and Clustering Analysis to extract information on the underlying structure in the data, all assist modern industry and research in taming many a complex phenomenon [1-5].

2. RESOURCES USED: SOURCES, DATABASE AND COMPUTING RESOURCES

2.1. MAP and Creep Rupture database

In order to provide some examples of the application of the mentioned modelling techniques, a set of empirical metallurgical data has been used. This database is provided by the Materials Algorithms Project, a Materials Science repository created by the Phase Transformations Group of the University of Cambridge and the National Physical Laboratory (UK), which can be openly accessed at http://www.phase-trans.msm.cam.ac.uk/map/mapmain.html.

In particular, the chosen database is MAP_DATA_CREEP_RUPTURE. This database was compiled by F. Brun and Dr. T. Yoshida from the published literature and consists of 2066 instances of creep rupture stress results, presented as function of 30 inputs including chemical composition, heat treatment conditions and test parameters [6].

The composition of each alloy is defined by C, Si, Mn, P, S, Cr, Mo, W, Ni, Cu, V, Nb, N, Al, B, Co, Ta, O, Re. Normalising, tempering and annealing heat treatments are defined by time and temperature treatment and cooling severity. The Creep tests are defined by temperature, stress and resulting time to rupture.

The data has been normalised and randomised and split in three smaller datasets. The first one, the training set, containing the 60% of the datapoints, and then two datasets with a 20% of datapoints each to be used as test and cross validation sets, i.e. used to compare the resulting models using unseen data.

For anomaly detection and clustering analysis however, the whole dataset has been used.

2.2. A note on computer used

The computer used to run these calculations has an Intel Core i7-4720HQ CPU, running at 2.60GHz with 4 physical cores (and, by hyperthreading, simulating 8 virtual cores) and 15.6 GiB of RAM. It runs 64-bit Ubuntu Linux 16.04 LTS, and the mathematical programming environtment, Octave [7].

3. SUPERVISED LEARNING & ADVANCED REGRESSION

Supervised learning refers to the construction of models with the purpose of making predictions (both quantitative or qualitative). It is characterised by a model training process in which a set of answers, or labels is provided. The model thus constructed should be able to correctly describe the phenomenon and therefore to predict the correct answers in new examples.

3.1. One-step fit of Linear Regression

If we have a dataset matrix X and a vector of labels or answers Y, it is possible to build the following model:

$$X \cdot \theta = Y \tag{1}$$

It is possible to obtain the fitting parameters θ to construct a linear regression model in one single step, by using the Normal Equation method (Eq. 2).

$$\theta = X^t \cdot X^{-1} \cdot X^t \cdot Y \tag{2}$$

This method has several advantages. Obviously, one is that it is solved in a single step. Additionally this method does not require any optimisation parameter. The practical application of this method however, still has some caveats. On one hand, matrix X is not always invertible. For instance, in real systems, often some parameters are a linear combination of some other parameters in the system, making X non-invertible. The error involved in experimental measurement of the data also introduces difficulties in the stability of this

method. Finally, but not least important, the computational cost of obtaining the inverse of X grows rapidly with the size of the database. For these reasons, alternative iterative methods are often favoured.

3.2. Iterative fit of Linear Regression

A commonly used alternative to the Normal Equation method to obtain the fitting parameters in a Linear Regression is related to the Gradient Descent method. Gradient descent is an iterative method based on the obtention of the optimal set of θ parameters by the progressive reduction of the error function of the parametrisation of the Linear Regresion. This is obtained by varying the parameters θ following the first derivative of the error function, until the global minimum is reached.

The gradient descent method has the advantage of being less computationally demanding, specially for large databases, than the Normal Equation method. Additionally, it does not suffer from stability issues related to the invertibility of the matrix X. However, it requires the optimisation of a learning parameter α, that define the step used in each iteration.

Finally, the gradient descent method is also suitable to train Logistic Regression models (classification).

3.3. Precision Comparison

The results of training a model using the two methods presented above are presented in **Table 1**. The average error obtained over the training set is very similar, as well as that obtained when making predictions over unseen data (the test set). In both cases, and as it was expected, the result on unseen data is slightly worse but the increase in predictive error is small (<3%).

Table 1 Linear Regression average train and test errors for both Normal Equation and gradient descent methods. Gradient Descent method using α=0.3 and 5000 iterations

Method	Average Train Error	Average Test Error	Diff.(/%)
Normal Eq.	0.00448497	0.00460893	2.76
Grad. Desc.	0.00448732	0.00462049	2.97

3.4. Artificial Neural Network Regression

An Artificial Neural Network is a non-linear regression method remotely related to the operation of biological neurons. The inputs to a neural node are operated over an activation function and transferred as inputs to the next neural layer, and eventually produce the output Y.

Different architectures of ANN are possible, by varying the number of hidden layers and number of nodes in each hidden layer, as well as the activation functions used in each of them. Commonly used activation functions are hyperbolic tangent, sigmoid and linear functions. Because of the non-linear nature of its transference function a neural network can capture interactions between the inputs and output that would not be possible using a traditional linear regression model. ANN are used both in regression and classification models [1 - 3, 5].

3.5. Risk of Overfitting, Regularisation.

Unfortunatelly, there are some risks with automated model fitting of which we need to be aware. Precisely because the fitting of the model is performed automatically, there is the need to ensure that the model describes the phenomenon of interest, and not the noise in the experimental data. It is a common mistake to fit increasingly complex models to a dataset, obtaining very good error scores, to discover later on that the model is unable to descrive unseen examples of the data (overfitting, or high variance fit, **Figure 1c)**. The

oposite problem consists in having a overtly simple model (i.e. a model that does not include all the rellevant parameters) leading to a high bias fit (**Figure 1a**) [1 - 3, 5].

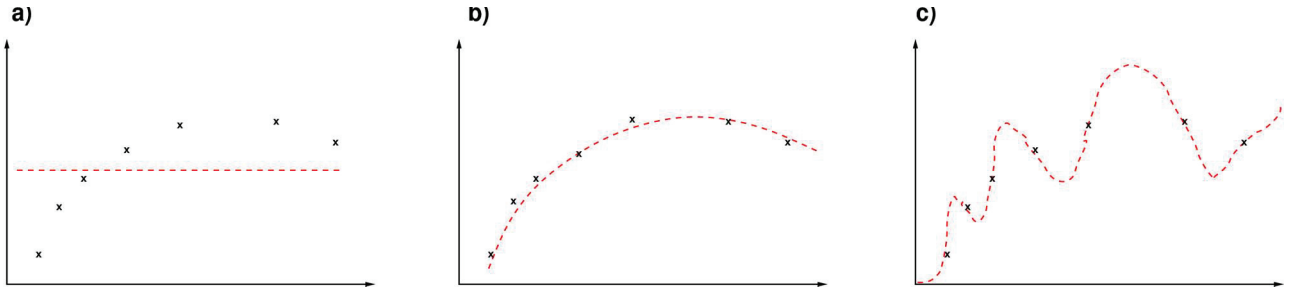

Figure 1 Finding the right balance during model fitting.
a) High bias model (underfit). b) Balanced fit. c) High variance model (overfit)

In the ANN examples produced, a degree of overfitting was detected (the error when making predictions on the test dataset was clearly larger than the error on the training set). Regularisation is a method to reduce high variance fit (overfit), by penalising the complexity of the model [1 - 3, 5].

Table 2 Artificial Neural Network with minimal regulatisation (λ=1) and with regularisation(λ=75). In all cases, one hidden layer and twice as many hidden units as input parameters

Act. Funct.	Reg. param.	Train Error Function	Test Error Function	Diff.(/%)
Lin/Lin	λ=1	0.005596	0.0131382	134.78
Tgh/Lin	λ=1	0.00566931	0.0136643	141.02
Tgh/Tgh	λ=1	0.00568072	0.0138385	143.60
Tgh/Tgh	λ=75	0.0139584	0.0140113	0.38

4. UNSUPERVISED LEARNING

Unsupervised learning refers to the analysis of the data without providing any a priori target results. The analysis performed by an unsupervised model training process provides a new insight of the underlying structure of the data [1-3,5].

4.1. Anomaly Prediction

In any industry it would be extremely useful to detect any deviation from the normal operations, before they produce any detrimental result, a faulty product or mechanical failure. Anomaly prediction tools are able to provide that warning in industrial applications [5,8-9].

This is possible by training a multivariate probability distribution function out of the relevant process parameters and using it to detect any outliers in the behaviour of the system. By setting a tolerance threshold inferred from previously detected anomalies it is possible to infer if a current instance is beyond the normal functioning conditions. This method however, only gives warning of the existence of an outlier, but without informing what is unusual with it. Nevertheless, it allows to single out that speciffic instance for further quality tests, for instance. Or, it could be combined with a classification algorithm to predict the nature of the fault [3, 5].

Within the purpose of this article, the system studied (creep rupture time data) is not really amenable to perform an anomaly prediction in the industrial sense. However, the multivariate probability distribution function has nevertheless been determined, and it allowed to rank the different datapoints in terms of likeliness, and for instance infer the difference in plausibility for each datum. Just as an example, say that the difference between most likely result to least likely one has been found to be of 58 orders of magnitude ($1.98 \cdot 10^{-2}$ and

$5.37 \cdot 10^{-61}$). That suggests a very skewed distribution with some outliers that would probably be worth studying in more detail.

4.2. Cluster Analysis

The information contained in a database presents a degree of variance (i.e. diversity). By grouping the data in a reduced number of clusters, each one containing only those datapoints with similar characteristics, the overall variance is reduced. Subsequent analysis of the clustering scheme may provide better understanding of the phenomena described by the data, as underlying independent cases will be isolated. However, although it is always possible to reduce further the variance of the whole dataset by subdividing it into smaller clusters, ultimately a clustering scheme assigning just one or a few datapoints to each cluster would not provide any additional information, so a trade-off needs to be achieved.

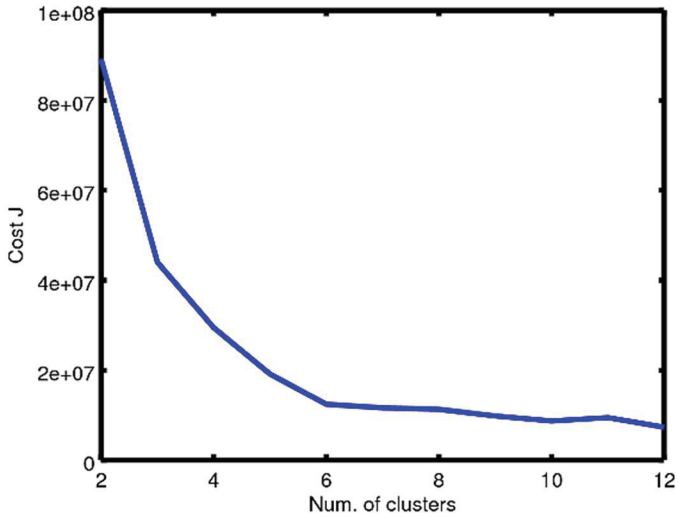

Figure 2 Using Cost J as proxy for overall variance, this is plotted against different clustering schemes. In this case, a trade-off between variance reduction and reduced number of clusters suggest a clustering scheme including around 5 clusters

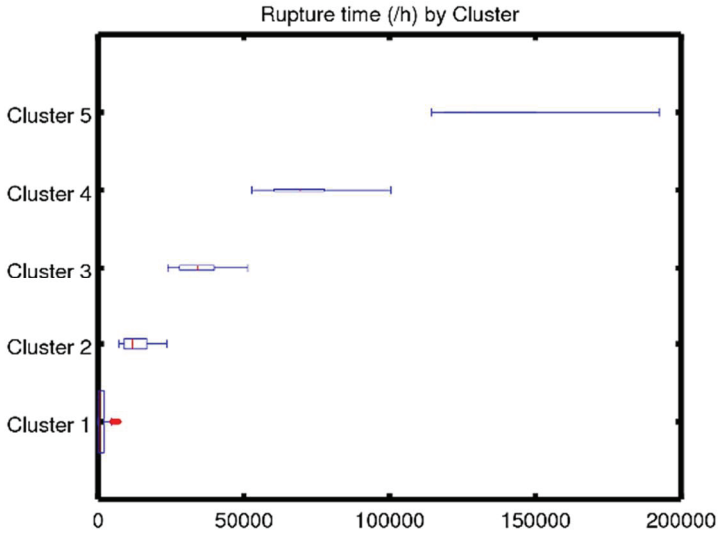

Figure 3 Using the chosen 5 clusters some differences in the data become evident. In the case of Time to Rupture, each cluster includes only tests with a similar range of rupture time

The method groups the data in different combinations of groups or clusters, and determines the overall resulting variance for each case. This variance is compared for different clustering schemes in order to find a trade-off between variance reduction and reduced number of clusters.

Applying this method to the MAP_DATA_CREEP_RUPTURE database, a clustering scheme including around 5 clusters seems to be optimal. Using the chosen 5 clusters some differences in the data become evident. For instance, in the case of Time to Rupture, each cluster includes only tests with a similar range of rupture time.

5. EXTRA EXAMPLES

All of these techniques can be used independently, but their capabilities may also be combined to build more powerful tools. The possibilities are almost endless.

For instance, in a prototype development factory, an Anomaly Prediction model may be used to detect outliers in the manufacturing process (and therefore, possible faulty components), and its output combined with a Logistic Regression (or Classification model) that classifies each of those possible anomalies into one of the several possible fault modes, significantly facilitating quality control procedures.

Another example, involves the description (and prediction) of the outcome of a complex process or phenomenon. Cluster Analysis may be performed first, to identify a number of characteristic 'cases' in the process. Then, when a Linear Regression or Artificial Neural Network is used to describe and make predictions on the outcome of the process, the clustering information may be used as an input parametre, reducing the variability in the data. If necesary, independent models for each cluster could be developed even.

6. CONCLUSION

A new project on Automated Model Fitting aimed to develop tools useful for industry and research has been presented. Many are the techniques and methods that can be included in labels like Expert Systems, Artificial Intelligence, Machine Learning, and others. In the end, many of those consist on automated methods to develop models describing some complex phenomenon, either with the aim of predicting some behaviour or to find the underlying structure in the data, and therefore reducing its variance.

Two methods to train a Linear Regression have been presented, as well as an Artificial Neural Network model, and some of their advantages and disadvantages discussed. Unsupervised training has been exemplified with an Anomaly Prediction method and Clustering Analysis.

The examples have been developed using a publicly available database on Creep Rupture test data, although some additional industrial project applications have also been suggested.

REFERENCES

[1] BHADESHIA, H. K. D. H. Neural networks in materials science. *ISIJ International*, 1999, vol. 39 pp. 966-979.

[2] GAUDE-FUGAROLAS, D. *Modelling of transformations during induction hardening and tempering*. PhD thesis, University of Cambridge. Department of Materials Science and Metallurgy, Cambridge, 2002.

[3] MACKAY, D. *Information Theory, Inference, and Learning Algorithms*, 1st Ed. Cambridge University Press, Cambridge (UK), 2003.

[4] GAUDE-FUGAROLAS, D., YARDLEY, V., LARDON, J-M, MONTAGNON, J. DE CARLAN. Y. Advanced alloy design tools applied to the development of vanadium nitride strengthened high-temperature steels. *In Proc. 8th Liege Conference Materials for Advanced Power Engineering*. Liege, Belgium, 2006, vol. 3, pp 1319-1328.

[5] NG. A. Machine learning. https://www.coursera.org/learn/machine-learning, 2016. Accessed 26/05/2016.

[6] BRUN, F., YOSHIDA, T., ROBSON, J. D., NARAYAN, V., BHADESHIA, H. K. D. H., MACKAY, D. J. C. Theoretical design of ferritic creep resistant steels using neural network, kinetic and thermodynamic models. *Materials Science and Technology*, 1999, vol. 15, pp.547-554.

[7] EATON, J. W., BATEMAN, D., HAUBERG, S., WEHBRING. R. GNU Octave version 4.0.0 manual: a high-level interactive language for numerical computations. 2015.

[8] GILMORE, C., HAYDAMAN, J. Anomaly Detection and Machine Learning Methods for Network Intrusion, *Detection: an Industrially Focused Literature Review*, 2016, pp. 292-298.

[9] MARTÍ, L., SANCHEZ-PI, N., MOLINA, J.-M., BICHARRA GARCIA, A.C. Anomaly Detection based on Sensor Data in Petroleum Industry Applications. *Sensors*, 2015, vol. 15, pp. 2774-2797.