

PROPERTY INVESTIGATIONS WITH AN AUTOMAT CORRELATION ROUTINE AND APPLICATIONS FOR A SET OF ALLOYS

L. Jäntschi

Faculty of Science and Engineering of Materials, Technical University of Cluj-Napoca, Romania

ABSTRACT: The present paper is focused on modeling of statistical data processing with applications in field of material science and engineering. A new method of data processing is presented and applied on a set of 10 Ni–Mn–Ga ferromagnetic ordered shape memory alloys that are known to exhibit phonon softening and soft mode condensation into a premartensitic phase prior to the martensitic transformation itself. The method allows to identify the correlations between data sets and to exploit them later in statistical study of alloys. An algorithm for computing data was implemented in preprocessed hypertext language (PHP) and a hypertext markup language interface for them was also realized and putted onto comp.east.utcluj.ro educational web server and is accessible via http protocol at the address <http://comp.east.utcluj.ro/~lori/research/regression/linear/v1.3/>. The program running for the set of alloys allow to identify groups of alloys properties and give qualitative measure of correlations between properties. Surfaces of property dependencies are also fitted.

KEYWORDS: linear regression, automat processing of data, web applications, set of alloys.

1. INTRODUCTION

Many statistical procedures for processing data are now available [1]. Most of them offer a voluble set of possibilities and variants, but which one to consider them? That is not a easy question and the frequent answer is: that is choice of analyst [2,3]. Data mining technology offer in this area of knowledge some answers, but not a complete answer [4].

By other hand, to interpret experiment results, data need to be well processed [5]. Structure investigations are frequently combined with statistical processing [6]. In most of cases, best results are obtained with specific procedures in contrast to general numeric algorithms [7,8].

Modeling of structure is benefit to property predictions [9,10]. Nonstandard statistical evaluation procedures then are helpful [11].

The presented model make data preprocessing to a set of 10 Ni–Mn–Ga ferromagnetic ordered shape memory alloys that are known to exhibit phonon softening and soft mode condensation into a premartensitic phase prior to the martensitic transformation itself and is a extension added to the model presented in book [12].

2. METHOD

The logic scheme of data preprocessing is presented in figure 1.

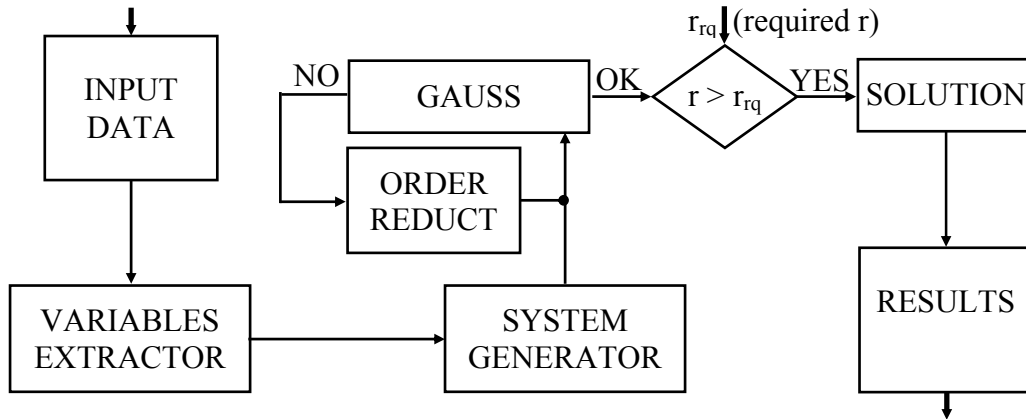


Figure 1. Data automat processing algorithm

The INPUT module read a text format data, process input data, split it into rows and columns and computes average means. If name n_rows it assigned to number of rows, n_cols to number of columns, $data$ to array of data, the output of module INPUT is computed by formulas:

$$M[i,j] = \sum_{k=1}^{n_rows} data[k][i] \cdot data[k][j] / n_rows; \quad M[0,j] = \sum_{k=1}^{n_rows} data[k][i] / n_rows, \quad 1 \leq i, j \leq n_cols \quad (1)$$

Linear regression and PLS (partial least squares) are most used methods in statistical processing of data. Presented method uses them. The output of INPUT module is used as input in GAUSS and RESULTS modules. GAUSS module solves a linear system of equations in form (2a):

$$\sum_{j=1}^{n_cols} M[i,j] \cdot x_j = 1, \quad 1 \leq i \leq n_cols \quad \text{OR} \quad \sum_{j=1}^{n_cols-1} M[i,j] \cdot x_j = M[0,j], \quad 1 \leq i < n_cols \quad (2)$$

If answer of algorithm solving is *undetermined system* and null variable is x_{n_cols} then GAUSS module solve determined system of n_cols order given by equation (2b). If answer of algorithm solving is *undetermined system* and null variable is different form x_{n_cols} then GAUSS module pass extended system matrix to REDUCT ORDER module. If is input in REDUCT ORDER module then is an *undetermined system* and this it extract null row and column corresponding to the null variable and the resulting matrix of $(n_cols-1) \times n_cols$ dimension is passed again to GAUSS module. When system is solved a unique solution is found. Then, System extended matrix contain at column n_cols the coefficients of regression equation:

$$a_1 \cdot x_1 + \dots + a_i \cdot x_i + \dots + a_{n_cols} \cdot x_{n_cols} + a_{n_cols+1} = 0 \quad (3)$$

where the coefficients a_{n_cols+1} and a_{n_cols+1} are resulted regression coefficients. Note that equation (3) is in implicit form; to obtain a explicit form is necessary to extract dependent variable from (3). The last coefficient is assigned to -1: $a_{n_cols+1} = -1$.

At end of module SOLUTION it result an implicit linear regression equation between given variables through his values in columns (equation 3). Equation 3 can be exploited to obtain explicit linear regression equations for each variable that has no null coefficient a_i :

$$\hat{x}_i = \left(\frac{a_1}{-a_i} \right) \cdot x_1 + \dots + \left(\frac{a_{i-1}}{-a_i} \right) \cdot x_{i-1} + \left(\frac{a_{i+1}}{-a_i} \right) \cdot x_{i+1} + \dots + \left(\frac{a_{n_cols}}{-a_i} \right) \cdot x_{n_cols} + \left(\frac{a_{n_cols+1}}{-a_i} \right) \quad (4)$$

Sum of residues S_i can be now evaluated. To compare one equation to another, a order value is required. Let to explicit this. If x_1 values (data[1] from input) are percents expressed in values from 0 to 100 and x_2 is premartensitic temperature transformation expressed in K with values from 100 to 600, then also sum of residues are expressed square of same measurement units. To make independence of measurement unit and measure order, values S_i are divided with own sum of squares of variable measurements ($M[i,i]$ from INPUT module, equation 1). Final equation, with substitution $x_i = \text{data}[k,i]$, $1 \leq k \leq n_rows$ and summing Q_i is:

$$S_i = \left(\frac{a_{n_cols+1}}{a_i} + \sum_{j=1}^{n_cols} \frac{a_j}{a_i} \cdot x_i \right)^2, Q_i = \sum_{k=1}^{n_rows} \left(\frac{a_{n_cols+1}}{a_i} + \sum_{j=1}^{n_cols} \frac{a_j}{a_i} \cdot \text{data}[k,i] \right)^2 / M[i,i] \quad (5)$$

and express relative residues of variable x_i when variable x_i is assumed to be dependent of independent variables $x_1, \dots, x_{i-1}, x_{i+1}, x_{n_cols}$. Note that the dependence and independence statistical concept is hard to prove in practical situations, but will see later, can be decelerated. For a good correlation, Q_i should be smallest possible value.

Another quantitative measure for a good correlation is correlation coefficient between measured data x_i and estimated values \hat{x}_i from equation (4). Assuming that M is mean operator, for r is given by:

$$r(x_i, \hat{x}_i) = \frac{M(x_i \cdot \hat{x}_i) - M(x_i) \cdot M(\hat{x}_i)}{(M(x_i \cdot x_i) - M(x_i) \cdot M(x_i))^{1/2} \cdot (M(\hat{x}_i \cdot \hat{x}_i) - M(\hat{x}_i) \cdot M(\hat{x}_i))^{1/2}} \quad (6)$$

The absolute value of r must be high for a good correlation. More additionally tests are also available in other programs such as Microsoft Excell or Statsoft Statistica.

3. ALGORITHM AND IMPLEMENTATION

The implementation of algorithm is relative simple, if are used a flexible language processing. In terms of programming, portability of resulted program can be a problem. As example, if we are chose to implement the algorithm in Visual Basic, the execution of the program is restricted to Windows machines. If Perl is our choice, a Unix-based machine is necessary to run program. Even if we chose to implement the program in C language, we will have serious difficulties to compile the programs on machines running with different operating systems. Other questions require an answer: We want a server based application or client based application? We want a server side or a client side application?

As example, a client side application can have disadvantage of execution on client, and dependence of processing speed by power of client machine. If we prefer this variant, a java script or visual basic script is our programming language. Most benefit to portability and execution speed seems to be a PHP (post processed hypertext) variant of implementation. A PHP script can be putted on any server or client with PHP processor and executed from them trough http server (Apache, Squid, ...) and client (Internet Explorer, Netscape).

Another advantage of PHP using is the possibility to link our algorithm with a materials database (d-Base, Interbase, MySQL, PostGRES format) and input data can be then loaded from them.

As conclusion, a PHP implementation is our choice. A graphical interface was built in html with a TEXTAREA for input data and an INPUT SUBMIT button for submitting data to the server. The server is a Free BSD Unix based server with an Apache web server running on. The server is hosted in educational network of Technical University of Cluj-Napoca with address 193.226.7.211 and name comp.east.utcluj.ro.

With PHP technology, was build a routine for pseudo domain names, that redirect client to different pages, depending on his input of domain name in client http browser. The program build have 21 subroutines and a main program, specified below:

- function af_ec(\$n,\$coef,&\$t) // display a equation;
- function af_mt(\$titlu,&\$tabel,\$n_r,\$n_c) //display any matrix with an title;
- function af_rez(\$n_r,\$n_c,&\$d,&\$m,&\$c,&\$t,\$n_o,\$spr) //list a table with best equations founded;
- function ch_ln(\$l1,\$l2,&\$cc,\$r) // Gauss linear algebra method, change two lines in system extended matrix;
- function cnk(\$k,\$n_r,\$n_c,&\$data,&\$tab,\$spr,&\$dep,&\$inv) //make recursive all possible combination c(n,k);
- function data_copy(\$n_r,\$n_c,&\$d,&\$t,&\$d_t,&\$n_t) //extract an subset of data from entire set;
- function do_means(&\$data,&\$mean,\$n_rows,\$n_cols) //make all (xi, xi*xj) means;
- function ec_by(\$n,&\$coef,\$by,&\$c_by) //calculate coefficients for explicit equation from implicit equation coefficients;
- function ec_val(\$n,&\$valori,&\$coef) //compute value of implicit equation in given point
- function estimare(\$n_r,\$n_c,&\$d,&\$c,\$x,&\$x_est) //compute value for explicite equation in given point for given dependent variable;
- function im_ln(\$nr,\$rw,&\$cc,\$r) //Gauss linear algebra method, make unitary element in system extended matrix;
- function mx_rw(\$cl,&\$cc,\$r) //Gauss linear algebra method, find the best line for zeroes in system extended matrix;
- function n_to_s(\$nr) //format and display a real number;
- function r_stat(\$n,\$k,&\$d1,&\$d2) //compute correlation coefficient r;
- function rd_gs(\$cc,\$r,&\$cf) //Gauss linear algebra iterative algorithm;
- function reg_lin_1(\$n_r,\$n_c,&\$d,&\$t,\$spr) //make linear regresion if possible; return answer;
- function res(&\$t,\$n) //reset counter for recursive c(n,k);
- function sum_r(\$n_rows,\$n_cols,&\$data,&\$coef,\$cor) //calculate sum of residues;
- function ze_pd(&\$cc,\$r) //Gauss linear algebra method, make supdiag. zeroes in system matrix;
- function ze_sd(\$e,&\$cc,\$r) //Gauss linear algebra method, make subdiag. zeroes in system matrix;
- main program //input data and requested minimal correlation coefficient and display founded equations;

4. RESULTS AND DISCUSSION

A set of Ni–Mn–Ga ferromagnetic ordered shape memory alloys are used for investigation [13].

The properties are: 1 - Alloy State (Poly- or Single-crystalline alloy), with values 1, -1 (PC, SC respectively); 2 - e/a, electron/atom ratio; 3 - Concentration of Ni, in %; 4 - Concentration of Mn, in %; 5 - Concentration of Ga, in %; 6 - T₁ (rows 1-7), T_M' (rows 8-10), in K, temperatures: T₁= martensitic transformation; T_M'=intermartensitic transformation in Group III alloys; 7 - T_M, premartensitic temperature transformation, in K.

Table 1. Input data values (outputted by PHP program)

0	1	2	3	4	5	6	7	0	1	2	3	4	5	6	7
1	1	7.35	49.6	21.9	28.5	4.2	178	6	1	7.56	47.7	30.5	21.8	227	240
2	1	7.36	47.6	25.7	26.7	4.2	152	7	1	7.57	51.1	24.9	24	197	248
3	-1	7.45	49.7	24.3	26	183	218	8	-1	7.78	53.1	26.6	20.3	417	379
4	1	7.5	50.9	23.4	25.7	113	224	9	-1	7.83	51.2	31.1	17.7	443	415
5	-1	7.51	49.2	26.6	24.2	184	240	10	-1	7.91	59	19.4	21.6	633	517

Program computes and output the regression equations. With an $r_{rq} = 0.9$ the program found over 60 different implicit equations of linear regression with $r > r_{rq}$, almost impossible to obtain by hand or in some program with statistics kernel. If value of r_{rq} is increased to $r_{rq} = 0.99$, number of implicit equation founded is reduced at 29. For $r_{rq} = 0.999$ number of implicit equation founded is 12.

That is a large set! If we are interested to study dependence between two variables form set, then we select the proper table from output of the program. The displayed equation resulted between x_0 and x_3 variables is: $+x_0*0.16+x_3*1.00=+2.58*10^1$ with $Q = 0.42$ and correlation coefficient $r = 0.99972$.

If we are looking for totally dependent variables (and here exists, sum of concentrations is 100%), the program find it and also eliminate one of them from set. The program response for correlating variables x_2 , x_3 and x_4 is: $-x_2*1.00*10^{-2}-x_3*1.00*10^{-2}-x_4*1.00*10^{-2}=-1.00$; $S=0$; $r=1.00$.

If we are looking for dependences between e/a and concentrations, simply select the founded equations from program output. In table 2 is showed the dependence of e/a by concentration of Ni and Mn:

Table 2. Dependence of e/a by Ni(x_2) and Mn(x_3) expressed by explicit and implicit equations

x_0	x_1	x_2	x_3	x_4	x_5	x_6	Equation	Residue	Correlation
0	1	1	1	0	0	0	$+x_1*1.00-x_2*7.02*10^{-2}-x_3*3.99*10^{-2}=+2.98$	$7.35*10^{-4}$	0.99995
0	1	1	1	0	0	0	$-x_1*1.42*10^1+x_2*1.00+x_3*0.56=-4.25*10^1$	$1.55*10^{-3}$	0.99997
0	1	1	1	0	0	0	$-x_1*2.50*10^1+x_2*1.75+x_3*1.00=-7.47*10^1$	$5.43*10^{-3}$	0.99992
0	1	1	1	0	0	0	$-x_1*0.33+x_2*2.34*10^{-2}+x_3*1.33*10^{-2}=-1.00$	$1.86*10^{-3}$	0.99995

If we are looking to express one of the temperatures depending by concentrations, then the equations from table 3 are useful. The equations that contain maximum of independent terms (without one concentration) is given at the end of output file of the program (see table 4).

Table 3. Dependence of T_M (premartensitic temperature) by Mn(x_3) and Ga(x_4) concentrations

x_0	x_1	x_2	x_3	x_4	x_5	x_6	Equation	Residue	Correlation
0	0	0	1	1	0	1	$+x_3*0.56+x_4*1.00+x_6*2.36*10^{-2}=+4.45*10^1$	$4.88*10^{-2}$	0.99321
0	0	0	1	1	0	1	$+x_3*2.37*10^1+x_4*4.23*10^1+x_6*1.00=+1.88*10^3$	0.16	0.99023

Table 4. Most comprehensive multi-linear dependence in data set

Equation	Residue	Correlation
$-x_0*6.77*10^{-4}+x_1*1.00+x_3*2.79*10^{-2}+x_4*6.61*10^{-2}+x_5*5.27*10^{-6}-x_6*1.09*10^{-4}=+9.82$	$4.12*10^{-4}$	0.99999
$-x_0*2.42*10^{-2}+x_1*3.57*10^1+x_3*1.00+x_4*2.36+x_5*1.88*10^{-4}-x_6*3.92*10^{-3}=+3.51*10^2$	$4.36*10^{-3}$	0.99995
$-x_0*1.02*10^{-2}+x_1*1.51*10^1+x_3*0.42+x_4*1.00+x_5*7.97*10^{-5}-x_6*1.65*10^{-3}=+1.48*10^2$	$1.98*10^{-3}$	0.99999

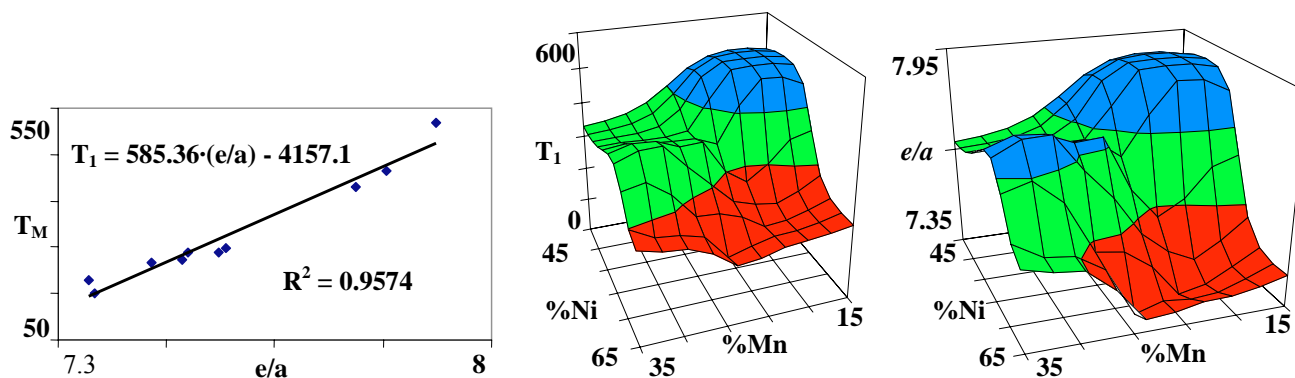


Figure 2. (a) Regression between T_1 and e/a , (b) surface plot of T_1 and (c) e/a by composition

In figure 2 are plotted some selected dependences from data set. Figure 2a show a mono-variable dependence between T_M and e/a , and figures 2b and 2c show surface dependences of T_1 and respectively e/a of concentrations (%Ni, %Mn).

5. CONCLUSIONS AND ACKNOWLEDGMENTS

Looking to the output sums of residues from tables, is easy to observe now that the properties type of alloy, and his martensitic, intermartensitic and premartenistic temperatures are interrelated having same order of sum residues in global equation, that is also expected conclusion. Very small same order of sum residues for concentrations suggest a strong interrelation between them, that is also true, because $\%Ni + \%Mn + \%Ga = 100$. This conclusion lead to consider the 3D plots fitted in figure 2 (b and c) of electron/atom ratio and T_1 temperature dependencies by concentration (%Ni,%Mn). The figure 3a prove good correlation between T_1 and e/a .

Author is grateful to the rector of Technical University of Cluj-Napoca, prof. dr. ing. Gheorghe Lazea for his policy on promoting information technology and also to the university staff for support in internet connection of lejpt web server. Useful support was also benefit from Romanian Ministry of Education for finance funding of MEC/CNCSIS contract 468/"A"/2002 and 281/"A"/2002.

REFERENCES

1. H. Nascu, L. Jäntschi, T. Hodisan, C. Cimpoi, G. Câmpan, Rev. Anal. Chem., **XVIII-6** (1999), 409.
2. L. Jäntschi, M. Unguresan, Int. Conf. Qual. Contr. Aut. Rob., Cluj-Napoca, Romania (2002), vol. 1, 194.
3. L. Jäntschi, H. I. Nascu, Int. Conf. Qual. Contr. Aut. Rob., Cluj-Napoca, Romania (2002), vol. 1, 259.
4. L. Jäntschi, M. Unguresan, Oradea Univ. Annals, Math. Fasc., **VIII** (2001), 105.
5. L. Marta, I. Deac, I. Fruth, M. Zaharescu, L. Jäntschi, Int. Conf. Sci. Mater., Cluj-Napoca, Romania (2000), vol. 2, 627.
6. L. Marta, I. Deac, I. Fruth, M. Zaharescu, L. Jäntschi, Int. Conf. Sci. Mater., Cluj-Napoca, Romania (2000), vol. 2, 633.
7. C. Cimpoi, L. Jäntschi, T. Hodisan, J. Plan. Chromatog. - Modern TLC, **11** (1998), 191.
8. C. Cimpoi, L. Jäntschi, T. Hodisan, J. Liq. Chromatogr. Rel. Techn., **10** (1999), 1429.
9. M. Diudea, L. Jäntschi, A. Graovac, Math/Chem/Comp'98 - Dubrovnik Int. Cour. & Conf. Interf. Math., Chem. Comp. Sci., (1998), vol.1, 3.
10. L. Jäntschi, G. Katona, M. Diudea, Commun. Math. Comput. Chem. (MATCH), **41** (2000), 151.
11. C. Sârbu, L. Jäntschi, Revista de Chimie, **1** (1998), 19.
12. M. Diudea, I. Gutman, L. Jäntschi, Molecular Topology, Nova Science, New York, (2001) 1-332.
13. V.A. Chernenko, J. Pons, C. Seguy', E. Cesari, Acta Mater., **50** (2002), 53.
14. V.A. Chernenko, J. Pons, C. Seguy', E. Cesari, Acta Mater., **50** (2002), 53.