

Control of coffee grinding with Artificial Neural Networks

Luca Mesin, Diego Alberto, Eros Pasero

Electronics Department
Politecnico di Torino
Turin, Italy
eros.pasero@polito.it

Alberto Cabilli

R&D
LAVAZZA
Turin, Italy
a.cabilli@lavazza.it

Abstract— Quality assessment and standardization of the property of the final product is fundamental in food industry. Coffee particle granulometry and density are continuously monitored during coffee beans grinding and grinders are controlled by operators in order to keep coffee particle granulometry within specific thresholds. In this work, a neural system is used to learn how to control two grinders used for coffee production at LAVAZZA factory, obtaining average control error of the order of a few μm . The results appear promising for the future development of an automatic decision support system.

Keywords—component; Artificial Neural Network; Partial mutual information; coffee; grinding.

I. INTRODUCTION

Nowadays in the food industry continuous controls and sophisticated evaluations are essential in the assessment of the quality of food, as required by European regulations on this subject [1,2]. Furthermore, in every factory the quality of products needs also to be maintained at a high level during the complete production chain. Various external factors can influence the quality of the final product, as material, mashing, fermentation, maturation and blending conditions. For this reason, automated controls are needed in order to guarantee high and stable production quality [3,4]. Moreover, it is often necessary to deal with an adaptive control, since different food varieties may require similar treatments [5,6].

An example of an effective and promising adaptive processing applied to food industry is represented by the Artificial Neural Networks (ANN) [7,8].

In this paper, we present the analyses carried out on several features of interest related to coffee production in the LAVAZZA [9] factory. Taking into account the rich bibliography on this subject [10 – 12], our aim is to study coffee feature time series and their mutual degree of influence using ANN [14]. In this way, the behavior of the principal variables can be controlled and forecasted during the coffee production to help and provide human operators with the most secure and probable regulation estimates. Our aim is to avoid unwanted breakdown of the complete production chain that

may occur if some of the most important parameters go out of range.

II. METHODS

A. Data acquisition

Datasets of some coffee production variables are recorded along the production line. Granulometry and densitometry are performed on coffee particles obtained after grinding, to check the quality of the product. Coffee granulometry is obtained using a laser diffraction analyser (Sympatec RODOS/M – HELOS [15]), providing a histogram of size with 17 bins from which 4 indexes are computed, as detailed below. Moreover, coffee bulk density is measured using a densitometer. Then, data are loaded by the operators and managed by the LAVAZZA Quality Control System which, by means of a Statistical Process Control software (ASI DATAMYTE [16]), allows data storing, searching and simple analysis, like identification of linear trends.

The size distribution is about bimodal. Fig. 1 shows the four parameters that are extracted to describe this probability density function (PDF): x_{10} , x_{50} , x_{90} , and *coffee dust* (or simply *dust*). The first three values represent the upper limits of the interval in the PDF that contain the 10%, 50%, and 90%, respectively, of all coffee particles dimension, they are expressed in μm . The fourth parameter (*dust*) estimates the particles percentage with a size lower than 100 μm . The production line recorded only four variables extracted from the granulometry PDF and the density of the product. Such variables were estimated as the most significant in characterizing the final product. Operators were instructed to manually regulate the grinders when these variables went out of specific ranges. Our full dataset consisted of 2570 patterns. On the basis of these measures, the operator decides if a change in the regulation of burr grinders is required. Sometimes, the granulometry variables are close to threshold values which requires the product to be monitored with a high frequency. If such threshold are undergone, production is stopped.

B. Data preprocessing

The resulting datasets were constituted by time series sampled at variable time instants. Data needed to be windowed

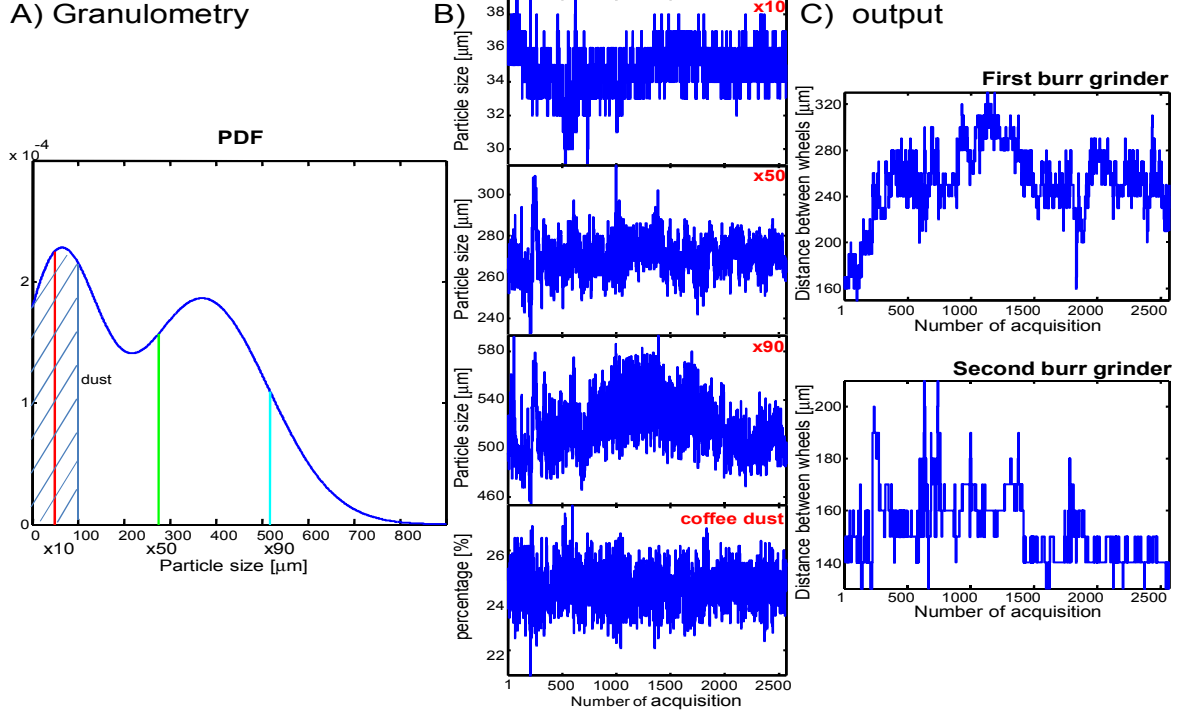


Figure 1. A) Granulometry (PDF) obtained fitting real data with the sum of two Gaussians. Data extracted from the PDF are indicated: $x10$, $x50$ and $x90$ are the 10th, the 50th and the 90th percentile; *dust* is the integral of the PDF from 0 to 100 μm . B) Data extracted from granulometry ($x10$, $x50$, $x90$, and *dust*), which, together with the density of the product, constitute the input time series for the neural control systems. C) First and second burr grinding output values (output to be learned by the neural control systems).

and resampled (at a constant sampling time of 15 minutes) after cubic spline interpolation to provide an efficient and effective training for the ANN approach we intended to develop. Furthermore, being interested in the one sample step ahead prediction, we took two input training patterns at a time to train the net.

C. Feature selection

The characteristics of the product are described by the variables extracted from granulometry (coffee dust, $x10$, $x50$, $x90$) and by the density (see Sect. II.A). On the basis of these variables, the operator decides how to control burr grinders. As a support to operator decision, a neural control system was considered. Two supervised artificial neural networks (ANN) were used to control the first and the second burr grinder, respectively. Thus, the output was the distance between the wheels of the first and second burr grinder, respectively. The possible input variables were the above mentioned data (coffee dust, $x10$, $x50$, $x90$ and density) measured at the present time or with a delay up to two sampling intervals (sampling interval equal to 15 minutes, see Sect. II.B) and the delayed output up to two sampling intervals.

The selection of optimal input features for the ANN is of great importance, in order to reduce the measurement noise, to counteract the difficulties of facing a problem with large dimension and to improve performance. We used a method based on the selection of the input data providing maximal information on the output. In order to avoid redundancy among

input features, the algorithm proposed in [17, 18, 19] was applied. It determines the interdependencies between candidate variables computing the Partial Mutual Information (PMI), which represents the information between a considered variable and the output that is not contained in the already selected features. Using the PMI selection method, the candidate features were ordered for decreasing values of mutual information with the output, as reported in Tab. I, for each burr grinder.

D. Application of MLP

Then, to map the input to each desired output, a set of multilayer perceptrons (MLP) was used. A single hidden layer was used, which is sufficient to approximate any nonlinear function (universal approximation property, [14]). Different MLPs had a different number of inputs (chosen following the decreasing order of PMI with the desired output), a number of neurons in the hidden layer in the range 4 – 20 (with sigmoidal activation function) and a single output neuron (with linear activation function). Different MLP topologies were trained by modifying iteratively the weights and the bias in order to reduce the error in fitting the desired output, applying the quasi-Newton algorithm [20] for a number of iterations in the range 50-200. We used one separate and independent MLP per grinder because, training a single MLP for both I and II burr grinders, we appreciated the same behavior in the error distribution for the II burr grinder, while a worse estimation of the control of the first grinder. Furthermore, the quasi-Newton

optimization algorithm was used since it provided lower errors on training, validation, and test sets, if compared with backpropagation (the decrease of the average error was nearly 2 μm for the training set and 1 μm for validation and test sets). On the other hand, the quasi-Newton algorithm requires a longer processing time (on the order of the 30-32 % for both the burr grinders) with respect to backpropagation. This increment could be due to the different implementation or to the simpler backpropagation algorithm, which requires only the estimate of the gradient of the error, whereas the quasi-Newton algorithm requires estimates of derivatives of the error up to the second order.

The data were divided into training (70% of data), validation (15% of data) and test sets (15% of data), choosing randomly the samples associated to each set, for 10 times. The network with highest average generalization capability (i.e., the minimum error) on the validation set was selected, and the error on the test set was further considered. All the analyses have been carried out using a MATLAB platform [21].

III. RESULTS

Fig. 1 shows the raw data considered in this work. The granulometry of a sample of the product (Fig 1A) is used to extract the variables (x10, x50, x90, and dust, Fig. 1B), which, together with the density of the product, constitute the input features for the neural control systems. The output to be learned by the neural control systems is the distance of the wheels of the two burr grinders (Fig. 1C).

The PMI criterion indicated the importance of the features in predicting the control of the grinders, as indicated in Tab. I. Different MLPs were trained. The ones obtaining better generalization on the validation set have 7 and 10 inputs, 14 hidden neurons and their weights were updated for 150 and 200 iterations of the optimization algorithm, for the I and II burr grinder, respectively. On the other hand, designing one single MLP for both grinders, the optimal network requires 5 inputs, 20 hidden neurons, and the weights were updated for 200 iterations of the quasi-Newton optimization algorithm. The first five most important inputs, selected also in this case by means of the PMI criterion, are the followings (written in order of decreasing PMI): II burr grinder delayed of 1 time step, I burr

grinder delayed of 1 time step, II burr grinder delayed of 2 time steps, I burr grinder delayed of 2 time steps, and x50.

The control outputs estimated using two MLPs, one for each grinder, are shown both for simulated (Fig. 2) and experimental (Fig. 3) data. Simulated data were obtained by changing in time the amplitude of the first Gaussian function modeling a bi-modal PDF (Fig. 2A) and computing the granulometry parameters and imposing a constant density (Fig. 2B). Applying on simulated input time series the MLPs trained on real data, the controls of the two grinder follow the same periodicity of the variation of the input variables. Moreover, we notice that, as the particle size is decreasing, the distance of the wheels is increased, in order to compensate for such a decrease, in line with our expectations. When applied to real data, Fig. 3 indicates that the MLPs follow reasonably well the output, with good generalization performances on the test data set.

The distribution of control error on the test set is shown in details in Fig. 4. The second grinder undergoes few small corrections with respect to an average position and can be followed with lower error with respect to the first.

TABLE I. FORECASTED BURR GRINDINGS AND RELATED PMI SORTED VARIABLES (10 MOST IMPORTANT FEATURES)

I burr grind.	PMI values	II burr grind.	PMI values
I burr grind. d1 ^a	1.360	II burr grind. d1 ^a	1.547
density d1 ^a	0.152	II burr grind. d2 ^a	0.141
x10 d1 ^a	0.134	density	0.042
dust d2 ^a	0.122	dust d2 ^a	0.041
x90 d1 ^a	0.120	x50 d1 ^a	0.040
x90	0.118	dust d1 ^a	0.038
II burr grind.	0.113	I burr grind. d1 ^a	0.037
x10 d2 ^a	0.106	x10	0.035
x10	0.101	x90 d1 ^a	0.034
x50 d2 ^a	0.085	I burr grind.	0.032

a. d1 (d2) means a sample delayed of one (two) sample time(s).

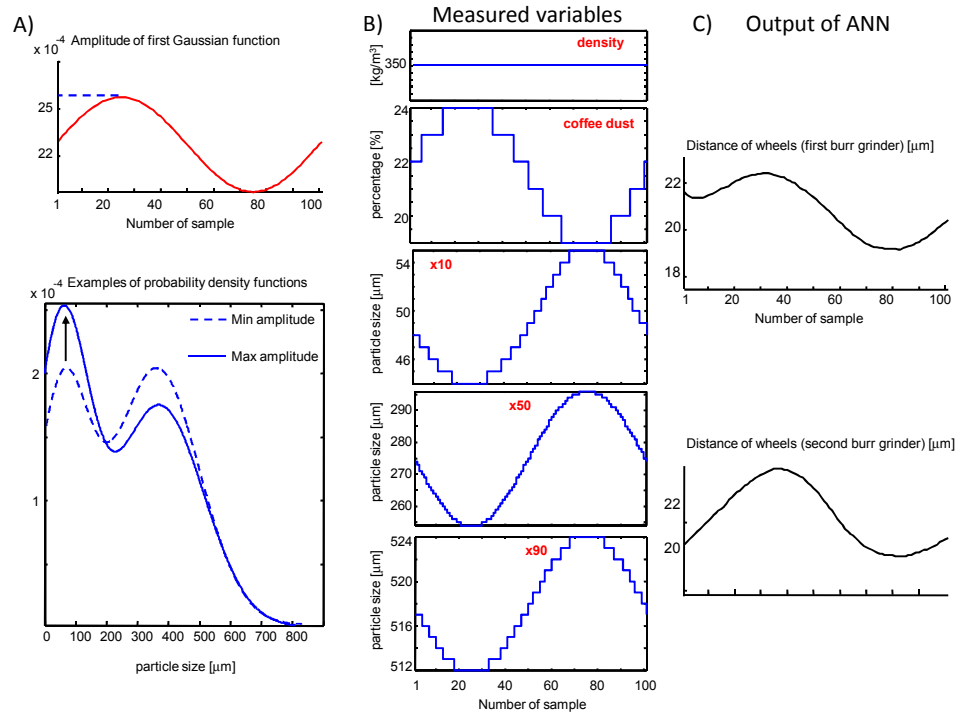


Figure 2. I and II burr grinder forecasted outputs (right) using MLPs trained on real data, applied to simulated input time series (shown in the center) estimated from the simulated PDF, which is obtained modifying in time the PDF estimated fitting real data shown in Fig 1. The simulated variation is a periodic change of the amplitude of the left Gaussian function used to represent the PDF.

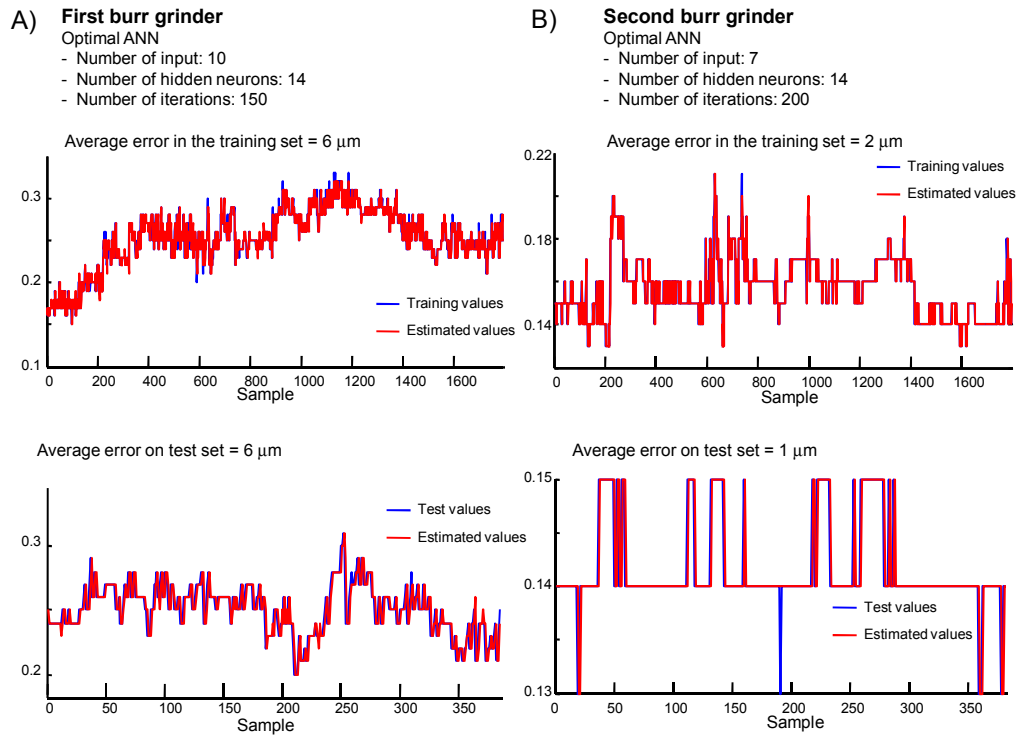
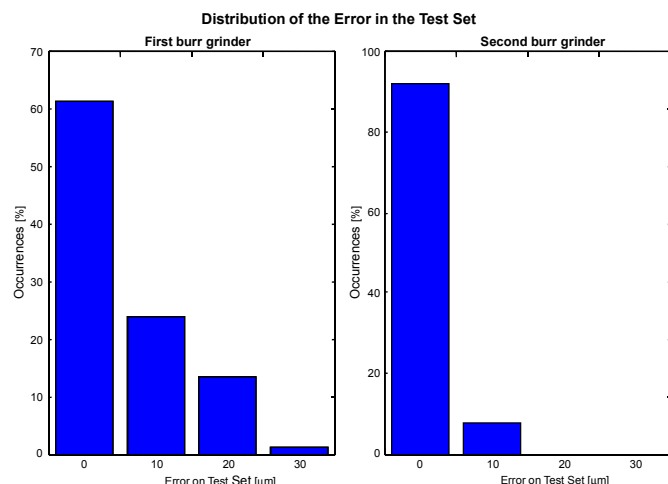


Figure 3. Optimal MLP topologies and estimated outputs on training and test sets for I burr grinding (A) and II burr grinding (B).



Test Set error histograms for predicted I burr grinding (left) and II grinding values (right).

DISCUSSION

A neural system was developed to learn how to control the two grinders used for coffee bean grinding in LAVAZZA. The input variables are obtained by measuring density and granulometry of samples. Such possible input features were ordered on the basis of PMI criterion and were selected considering the MLP topology guaranteeing optimal generalization performances. Control errors of a few μm were obtained (with a control step of 10 μm), both in the training and test set. Simulations indicate that the resulting control is in line with the choices that an experienced operator would take. Thus, the results are promising for a future integration of the control system in the production line, to support operators in controlling burr grinders. The performances so far obtained are sufficient for the considered industrial application, even if it is expected that they could be improved. Future works of our group goes in the direction of more complex and performing ANN designs.

Finally, the control solution proposed for the considered industrial case has the potential to be easily extended and applied to other processes related to food industry. Preliminary studies of our group are in progress on nuts selection and management.

- [1] <http://www.food.gov.uk/foodindustry/regulation/>.
- [2] R.B Leslie, J.C Oliveira, and A.G Medin, "Food Forum: a research forum for an innovative and globally competitive European Food Industry," Food Research International, Vol. 33, Issues 3–4, pp. 295–297, 2000.
- [3] S.Linko, "Expert systems—what can they do for the food industry?," Trends in Food Science & Technology, Vol. 9, Issue 1, , pp. 3–12, 1998.
- [4] S. V. Ilyukhin, T. A. Haley, and R. K. Singh, "A survey of automation practices in the food industry," Food Control, Vol. 12, Issue 5, pp. 285–296, 2001.
- [5] Julio R Banga, E. Bsa-Canto, C. G. Moles, and A. A. Alonso, "Improving food processing using modern optimization methods," Trends in Food Science & Technology, Vol. 14, Issue 4, pp. 131–144, 2003.
- [6] N.S. Bhuvaneswari, G. Uma, and T.R. Rangaswamy, "Adaptive and optimal control of a non-linear process using intelligent controllers," Applied Soft Computing, Vol. 9, Issue 1, pp. 182–190, 2009.
- [7] J.S. Torrecilla, L. Otero, and P.D. Sanz, "Artificial neural networks: a promising tool to design and optimize high-pressure food processes," Journal of Food Engineering, Vol. 69, Issue 3, pp. 299–306, 2005.
- [8] D. Guyer, X. Yang, "Use of genetic artificial neural networks and spectral imaging for defect detection on cherries," Computers and Electronics in Agriculture, Vol. 29, Issue 3, pp. 179–194, 2000.
- [9] LAVAZZA S.p.A., Strada Settimo 410, Torino.
- [10] O. Tominaga, F. Ito, T. Hanai, H. Honda, and T. Kobayashi, "Sensory Modeling of Coffee with a Fuzzy Neural Network", Journal of Food Science, Vol. 67, nr. 1, pp. 363–368, 2001.
- [11] J.A. Hernandez, B. Heyd, and G. Trystram, "Prediction of brightness and surface area kinetics during coffee roasting", Journal of Food Engineering, Vol. 89, Issue 2, pp. 156–163, 2008.
- [12] D.H.F. Yip and W.W.H. Yu, "Classification of Coffee using Artificial Neural Network", IEEE International Conference on Evolutionary Computation, Nagoya, Japan, pp. 655 –658, 1996.
- [13] S. Singh, E.L. Hines, and J.W. Gardner, "Fuzzy neural computing of coffee and tainted water data from an electronic nose", Sens. Actuators B, Vol. 30, nr. 3, pp. 190–195, 1996.
- [14] S. Haykin, "Neural Networks: A comprehensive Foundation", Ed. Prentice Hall, Upper Saddle River, US-NJ, 1999.
- [15] <http://www.asidatamye.com/>.
- [16] <http://www.sympatec.com/>.
- [17] A. Sharma, "Seasonal to interannual rainfall probabilistic forecasts for improved water supply management: Part 1 - A strategy for system predictor identification," Journal of Hydrology, Vol. 239, Issues 1–4, pp. 232–239, 2000.
- [18] L. Mesin, F. Orione, R. Taormina, and E. Pasero, "A feature selection method for air quality forecasting," Proceedings of the 20th International Conference on Artificial Neural Networks (ICANN), Thessaloniki, Greece, September 15–18, 2010.
- [19] R.J. May, H.R. Maier, G.C. Dandy, and T.M.K. Gayani Fernando, "Non-linear variable selection for artificial neural networks using partial mutual information," Envir. Mod. And Soft., Vol. 23, pp. 1312–1326, 2008.
- [20] E. Barnard, "Optimization for training neural nets," IEEE Transactions on Neural Networks, Vol. 3, Issue 2, pp. 232–240, 1992.
- [21] MATLAB, "The language of technical computing", <http://www.mathworks.com>.