Neural Model for the Prediction of Volume Losses in the Aging Process of Rums

Beatriz García-Castellanos; Osney Pérez-Ones; Lourdes Zumalacárregui-de-Cárdenas; Idania Blanco-Carvajal; Luis-Eduardo López-de-la-Maza

Beatriz García-Castellanos; Osney Pérez-Ones; Lourdes Zumalacárregui-de-Cárdenas; Idania Blanco-Carvajal; Luis-Eduardo López-de-la-Maza

Neural Model for the Prediction of Volume Losses in the Aging Process of Rums

Beatriz García-Castellanos[1]

Osney Pérez-Ones[2]

Lourdes Zumalacárregui-de-Cárdenas[3]

Idania Blanco-Carvajal[4]

Luis-Eduardo López-de-la-Maza[5]

Abstract

The rum aging process shows volume losses, called wastage. The numerical operation variables: product, boardwalk, horizontal and vertical positions, date, volume, alcoholic degree, temperature, humidity and aging time, recorded in databases, contain valuable information to study the process. MATLAB 2017 software was used to estimate volume losses. In the modeling of the rum aging process, the multilayer perceptron neuronal network with one and two hidden layers was used, varying the number of neurons in these between 4 and 10. The Levenberg-Marquadt (LM) and Bayesian training algorithms were compared (Bay) The increase in 6 consecutive iterations of the validation error and 1,000 as the maximum number of training cycles were the criteria used to stop the training. The input variables to the network were: numerical month, volume, temperature, humidity, initial alcoholic degree and aging time, while the output variable was

---

[1] Instituto Cubano de Investigaciones de los Derivados de la Caña de Azúcar (La Habana, Cuba). beatriz.garcia@icidca.azcuba.cu. ORCID: 0000-0001-8101-0638

[2] Ph. D. Universidad Tecnológica de La Habana "José Antonio Echeverría" (La Habana, Cuba). osney@quimica.cujae.edu.cu. ORCID: 0000-0002-0366-0317.

[3] Ph. D. Universidad Tecnológica de La Habana "José Antonio Echeverría" (La Habana, Cuba). lourdes@quimica.cujae.edu.cu. ORCID: 0000-0001-6921-737X.

[4] M. Sc. Instituto Cubano de Investigaciones de los Derivados de la Caña de Azúcar (La Habana, Cuba). idania.blanco@icidca.azcuba.cu. ORCID: 0000-0003-1281-3722.

[5] Universidad Tecnológica de La Habana "José Antonio Echeverría" (La Habana, Cuba). llopezm@quimica.cujae.edu.cu. ORCID: 0000-0002-7009-4415.

wastage. 546 pairs of input/output data were processed. The statistical Friedman and Wilcoxon tests were performed to select the best neural architecture according to the mean square error (MSE) criteria. The selected topology has a 6-4-4-1 structure, with an MSE of $2.1 \cdot 10^{-3}$ and a correlation factor (R) with experimental data of 0.9898. The neural network obtained was used to simulate thirteen initial aging conditions that were not used for training and validation, detecting a coefficient of determination (R^2) of 0.9961.

Keywords: aging; artificial neural networks; modeling; rums; volume losses.

Modelo neuronal para la predicción de mermas en el proceso de añejamiento de rones

Resumen

El proceso de añejamiento de ron experimenta pérdidas de volumen, denominadas mermas. Las variables numéricas de operación: producto, rambla, posiciones horizontal y vertical, fecha, volumen, grado alcohólico, temperatura, humedad y tiempo de añejamiento, registradas en bases de datos, contienen información valiosa para estudiar el proceso. Se utilizó el software MATLAB 2017 para estimar las pérdidas en volumen. En la modelación del proceso de añejamiento de ron se utilizó la red neuronal perceptrón multicapa con una y dos capas ocultas, variándose el número de neuronas en estas entre 4 y 10. Se compararon los algoritmos de entrenamiento Levenberg-Marquadt (L-M) y Bayesiano (Bay). El incremento en 6 iteraciones consecutivas del error de validación y 1 000 como número máximo de ciclo de entrenamiento fueron los criterios utilizados para detener el entrenamiento. Las variables de entrada a la red fueron: mes numérico, volumen, temperatura, humedad, grado alcohólico inicial y tiempo de añejamiento, mientras que la variable de salida fue mermas. Se procesaron 546 pares de datos de entrada/salida. Se realizaron las pruebas estadísticas de Friedman y Wilcoxon para la selección de la arquitectura neuronal de mejor comportamiento de acuerdo al criterio del error cuadrático medio (MSE). La topología seleccionada presenta la estructura 6-4-4-1, con un MSE de $2.1 \cdot 10^{-3}$ y un factor de correlación (R) con los datos experimentales de 0.9981. La red neuronal obtenida se empleó para la simulación de trece

Beatriz García-Castellanos; Osney Pérez-Ones; Lourdes Zumalacárregui-de-Cárdenas; Idania Blanco-Carvajal; Luis-Eduardo López-de-la-Maza

condiciones iniciales de añejamiento que no fueron empleadas para el entrenamiento y la validación, detectándose un coeficiente de determinación (R2) de 0.9961.

Palabras clave: añejamiento; mermas; modelación; redes neuronales artificiales; rones.

Modelo neuronal para a predição de mermas no processo de envelhecimento de runs

Resumo

O processo de envelhecimento de rum experimenta perdas de volume, denominadas mermas. As variáveis numéricas de operação: produto, rambla, posições horizontal e vertical, data, volume, grau alcoólico, temperatura, humidade e tempo de envelhecimento, registradas em bases de dados, contêm informação valiosa para estudar o processo. Utilizou-se o software MATLAB 2017 para estimar as perdas em volume. Na modelação do processo de envelhecimento de rum utilizou-se a rede neuronal perceptron multicamada com uma e duas camadas ocultas, variando-se o número de neurônios nestas entre 4 e 10. Compararam-se os algoritmos de treinamento Levenberg-Marquadt (L-M) e Bayesiano (Bay). O incremento em 6 iterações consecutivas do erro de validação e 1 000 como número máximo de ciclo de treinamento foram os critérios utilizados para deter o treinamento. As variáveis de entrada à rede foram: mês numérico, volume, temperatura, humidade, grau alcoólico inicial e tempo de envelhecimento, enquanto que a variável de saída foi mermas. Processaram-se 546 pares de dados de entrada/saída. Realizaram-se as provas estatísticas de Friedman e Wilcoxon para a seleção da arquitetura neuronal de melhor comportamento de acordo ao critério do erro quadrático médio (ECM). A topologia selecionada apresenta a estrutura 6-4-4-1, com um ECM de $2.1 \cdot 10^{-3}$ e um fator de correlação (R) com os dados experimentais de 0.9981. A rede neuronal obtida empregou-se para a simulação de treze condições iniciais de envelhecimento que não foram empregadas para o treinamento e a validação, detectando-se um coeficiente de determinação (R^2) de 0.9961.

Palavras chave: envelhecimento; mermas; modelação; redes neuronais artificiais; runs.

Beatriz García-Castellanos; Osney Pérez-Ones; Lourdes Zumalacárregui-de-Cárdenas; Idania Blanco-Carvajal; Luis-Eduardo López-de-la-Maza

I. INTRODUCTION

In fresh rum, as in most distilled alcoholic beverages, the aroma reminds of the used raw material. Aroma varies when the fresh rum rests in oak containers for a certain time, commonly known as "aging" or aging time. During this time, reactions that cause a transformation of the original organoleptic properties of the distillates occur naturally. [1]

The technological production process of the Aged rum that is carried out at the Alcohol and Beverage Reference Center (CERALBE), belonging to the Cuban Research Institute of Sugarcane Byproducts (ICIDCA), comprises several stages. During the rum aging process, product losses, popularly known as "the Angel portion", occur. The aging of rum does not change or transform the drink but develops and sublimates its latent qualities [2]. That is why, in the context of excellence in which these drinks compete, there is an interest in studying the decrease in the volume of rum during aging concerning the environmental conditions. These losses have not been updated in CERALBE recently, although the volume of losses is known to be high.

The existing technology in the aging cellars allowed the study of the wastages during 13 months; measuring the liquid level of the barrels, alcoholic strength, temperature and humidity. All this stored memory constitutes a valuable source of information that can be useful in understanding the present and predicting the future.

Data mining (DM) is the process of extracting useful and understandable knowledge, previously unknown, from large amounts of data stored in different formats [3]. It allows prediction, classification, association, grouping and correlation tasks based on statistical techniques such as the analysis of principal and computational components such as artificial neural networks [4]. Currently, the DM has become popular due to the increase in the computing capacity of the computers, combined with the increase in the data storage capacity and its quality [5].

Artificial neural networks constitute a computational tool that mimics the functioning of the human brain because it can learn patterns or behaviors from a database [6-7]. Obtaining predictive models from training, which is developed by presenting an

input matrix and its corresponding output, has allowed the modeling of different processes.

The predictive models obtained by data mining techniques constitute an alternative to the mathematical models and at the same time, a tool to analyze the information stored in the rum aging processes to predict the percentage of volume losses based on the variables that are registered.

II. MATERIALS AND METHODS

A. Creation and Training of the Neural Network

The multilayer perceptron neuronal network with one and two hidden layers, feed-forward network, backpropagation training algorithm was used for the modeling of the rum aging process. This type of neural network is easy to use and allows the modeling of complex functions [7-8].

The number of neurons in the hidden layers was varied from 4 to 10 with each of the training algorithms used: Levenberg-Marquadt (L-M) and Bayesian (Bay). The input variables to the network were: numerical month, volume, temperature, humidity, initial alcoholic grade and aging time, while the output variable was volume losses. Table 1 shows the minimum and maximum values for each variable. 546 pairs of input / output data were processed. The original data were normalized between 0.0 and 1.0, given the differences between their magnitudes. For the partition of training data, the "dividerand" function was used, with the default division of 70% for training, 15% for testing and 15%, for validation. The increase in 6 consecutive iterations of the validation error and 1.000 as the maximum number of training cycles were the criteria used to stop the training. Both the creation of the neural network and its training were carried out in Matlab 2017.

Table 1. Value range for each variable.

Variable	V (L)	GA (ºGL)	T (ºC)	H (%)	t (years)	m (L)
Minimum	154.33	55.72	23.82	43.75	3.46	0.59
Maximum	172.51	56.74	37.00	74.00	4.69	2.11
Range	18.18	1.03	13.18	30.25	1.22	1.52

*V: volume; GA: alcoholic grade; T: temperature, H: humidity; t: aging time, m: volume losses.

Beatriz García-Castellanos; Osney Pérez-Ones; Lourdes Zumalacárregui-de-Cárdenas; Idania Blanco-Carvajal; Luis-Eduardo López-de-la-Maza

For the selection of the optimal number of neurons in the hidden layer, three criteria were taken into account: the mean square error in the validation of the model (MSE), the average absolute error (MAE) and the correlation coefficient (R) between the values of losses estimated by the neuronal model and the real values. [9-10].

1) Friedman Test. The Friedman test allows the comparison of several samples, is used for the selection of the best behavioral neuronal architecture according to the criteria of the mean square error, in the case that the results were very similar and the decision becomes difficult. This test has two hypotheses: h_0 (null hypothesis), which raises the equality between all the medians in a group, and h_1 (alternative hypothesis), which raises the non-equality among all the medians in the group. It is a non-parametric test. The selection of the neural architecture is based on the P value; if this is less than 0.05 the null hypothesis is rejected.

2) Wilcoxon Test. When the Friedman test reveals that there are statistically significant differences between the group medians, it is necessary to perform the Wilcoxon test to select the best behavioral neuronal architecture. The Wilcoxon test (non-parametric) allows a comparison of pairs of samples. It has two hypotheses: h0 (null hypothesis), which remarks the equality between two medians, and h1 (alternative hypothesis), which mentions the non-equality between two medians. The selection of the neural architecture is based on the P-value; if it is less than 0.05 then the null hypothesis is rejected [10].

III. Results and Discussion

A. Determination of Noise, Cleaning and Selection of the Data to be Used

A data with 900 instances and ten variables were obtained, five of them qualitative: product, track, date and horizontal and vertical positions; while the remaining: aging time, volume, temperature, humidity and initial alcoholic degree are quantitative. The initial data matrix was reduced to 546 values as it presented incomplete measurements for three barrels (19, 21 and 159) and the period from February to September 2014.

1) Topology of the neural network. The number of neurons in the hidden layer was varied from 4 to 10 with each of the training algorithms used (Levenberg-Marquadt and Bayesian) and the behavior of all topologies were compared. The coefficient of determination (R^2) of the losses for each of the topologies, as well as the mean square error (MSE) and the mean absolute error (MAE, are shown in Table 2.

Table 2. Topology comparison with a hidden layer for each algorithm.

Neurons in the hidden layer	Levenberg-Marquadt			Bayesian		
	R^2	MSE	MAE	R^2	MSE	MAE
4	0.9271	0.0978	0.2734	0.9926	0.0092	0.0702
5	0.9325	0.0660	0.1864	0.9933	0.0230	0.1295
6	0.9890	0.0177	0.1080	0.9870	0.0361	0.1604
7	0.9885	0.0336	0.1415	0.9966	0.0080	0.0712
8	0.9567	0.0968	0.2478	0.9651	0.0824	0.2452
9	0.9779	0.0375	0.1611	0.9856	0.0277	0.1196
10	0.9684	0.0465	0.1664	0.9797	0.0403	0.1619

From Table 2 it is shown that the coefficient of determination values are in the range of 0.9271 to 0.9966; the mean square error values are between 0.0080 and 0.0978; while the absolute error is in the range of 0.0702 to 0.2734. By following the criteria of the highest coefficient of determination and the lowest error value, the best-performing neural networks have the structure 6-6-1 and 6-7-1 using the algorithm LM and 6-7-1 and 6-4-1 using the Bayesian algorithm.

According to the previous result and the similarity in the order of the mean square error for each topology, it was concluded that it was necessary to perform the Friedman test to determine if there were statistically significant differences for each configuration.

P-value for the Friedman test was 0.0926 for the L-M training algorithm, while in the case of the Bayesian it was 0.0580 (both greater than 0.05). Therefore, there was equality among all medians and there were no statistically significant differences between the behaviors of the different topologies for each training algorithm used.

Subsequently, networks were modeled with two hidden layers and with the best configurations of neurons obtained, alternating in the same way between the training algorithms mentioned. The statisticians, for the selection of the best model with two hidden layers, are shown in Tables 3 and 4 for each algorithm.

Beatriz García-Castellanos; Osney Pérez-Ones; Lourdes Zumalacárregui-de-Cárdenas; Idania Blanco-Carvajal; Luis-Eduardo López-de-la-Maza

Table 3. Topology comparison with two hidden layers for the L-M algorithm.

Neurons in the hidden layer	R^2	MSE	MAE
7-6	0.9651	0.0357	0.1516
6-7	0.9771	0.0250	0.1167
10-7	0.9401	0.2852	0.3099
7-10	0.9625	0.0577	0.1994
10-6	0.9755	0.0896	0.2479
6-10	0.9464	0.0833	0.2544
6-6	0.9428	0.0553	0.1874
7-7	0.9773	0.0370	0.1622
6-4	0.9753	0.0345	0.1420
4-6	0.8516	0.2213	0.3652
4-7	0.9833	0.0190	0.0933
7-4	0.9570	0.0782	0.2295

Table 4. Topology comparison with two hidden layers for the Bay algorithm.

Neurons in the hidden layer	R^2	MSE	MAE
7-4	0.9630	0.2167	0.3704
4-7	0.9957	0.0118	0.0936
10-7	0.9757	0.0936	0.2333
7-10	0.9887	0.0323	0.1410
10-4	0.9888	0.0284	0.1318
4-10	0.9955	0.0046	0.0432
7-7	0.9905	0.0525	0.1860
4-4	0.9981	0.0021	0.0389

The topologies shown in Tables 3 and 4 were tested for Friedman to determine if there were statistically significant differences with respect to each configuration. In the first case (L-M algorithm) the result of the P-value for the Friedman test was 0.6959, greater than 0.05, so there are no significant differences between the behaviors of the different topologies. In the second case (Bayesian algorithm), there was a difference between the medians of the topologies, presenting a P-value for the Friedman test of 0.0005, so the Wilcoxon test was performed to define the best neuronal topology according to the criterion of the mean square error. This test was performed between the best topology of this group (6-4-4-1) and the remaining ones; according to the criteria: mean square error, correlation coefficient and absolute error. The results of the Wilcoxon test are shown in Table 5.

Table 5. Comparison between pairs of topologies by Wilcoxon test.

Compared topologies	P-value	Significant differences
6-4-4-1 y 6-7-4-1	0.0025	*
6-4-4-1 y 6-4-7-1	0.0138	*
6-4-4-1 y 6-10-7-1	0.0240	*
6-4-4-1 y 6-7-10-1	0.0103	*
6-4-4-1 y 6-10-4-1	0.0274	*
6-4-4-1 y 6-4-10-1	0.3300	
6-4-4-1 y 6-7-7-1	0.0004	*

There are statistically significant differences with respect to the mean square error between the topology (6-4-4-1) and all the remaining ones, except for the configuration (6-4-10-1). Therefore, only those two could be considered selectable in this group, taking into account the criterion of the lowest mean square error. Table 6 shows a selection of the best topologies according to the mentioned criteria, in addition to the determination coefficient calculated for each of them.

Table 6. Selection of the best topologies.

Algorithm / Topologies	R	MSE	MAE	R^2
L-M/6-6-1	0.9890	0.0177	0.1080	0.9780
L-M/6-7-1	0.9885	0.0336	0.1415	0.9772
Bay/6-4-1	0.9926	0.0092	0.0702	0.9853
Bay/6-7-1	0.9966	0.0080	0.0712	0.9931
L-M/6-4-7-1	0.9833	0.0190	0.0933	0.9669
Bay/6-4-10-1	0.9955	0.0046	0.0432	0.9909
Bay/6-4-4-1	0.9981	0.0021	0.0389	0.9961

The topologies underlined in Table 6 constitute the best combinations of statisticians. The four selected correspond to the Bayesian training algorithm, two of them have a hidden layer and the other two. These were subjected to the Friedman test to determine if there were statistically significant differences. The P-value obtained was 0.0153, less than 0.05, showing that there were differences with the topology (6-7-1). The Wilcoxon test was performed to define if there were differences between the best topology (6-4-4-1) with the (6-4-1) since it had been determined, previously, that there was no difference with (6-4-10-1). The P-value obtained was 0.0513, so there are no differences. From these last four selected topologies, anyone expect (6-7-1) can be chosen.

Beatriz García-Castellanos; Osney Pérez-Ones; Lourdes Zumalacárregui-de-Cárdenas; Idania Blanco-Carvajal; Luis-Eduardo López-de-la-Maza

Based on the results of the best model according to the topology of the neural network, according to the criteria of the mean square error and the correlation coefficient separately, it was decided that the neuronal model that best predicts the losses in the rum aging process is (6-4-4-1). The higher correlation coefficient among all topologies, a moderate structural complexity that allows savings in Matlab software calculations and low mean square error values, are reasons that justify the previous decision.

2) Simulation. In order to verify the predictive capacity of the neural network, already trained and validated, the topology (6-4-4-1) was used to simulate 13 initial aging conditions. These values were determined as the average of the thirteen months of research for each variable analyzed, to ensure the interpolator character of the network. The quality of the model can be seen in Figure 1, where the actual values and those estimated by the neuronal model for the different initial conditions are shown. The average error in the estimate is 3.03%, the maximum error being 7.4%.

Fig. 1. Comparison chart between actual and predicted values for the neural network model (6-4-4-1).

The regression line between the values estimated by the neuronal model and the values of actual losses, as well as the coefficient of determination are shown in Figure 2.
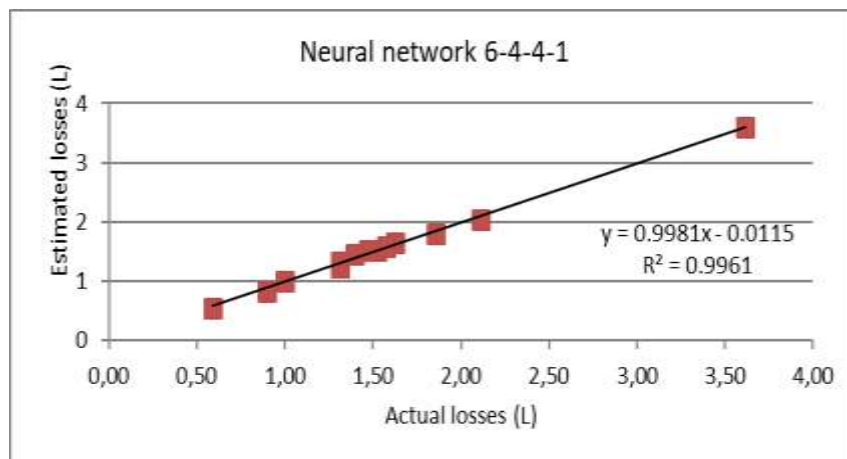
Fig. 2 Regression line between predicted and real values for the neural network model (6-4-4-1).

IV. CONCLUSIONS

The neural network obtained with the topology 6-4-4-1 was used to model the aging process and demonstrated its ability to estimate the losses in this process satisfactorily. The high value of the coefficient of determination (0.9961) between the simulated values and the real values, and the low mean square error in the model validation indicate the convenience of neural networks in the modeling of the aging process. On average, the error that the trained network commits in the estimate is 3.03%.

AUTHOR'S CONTRIBUTION

The collection and analysis of the information, as well as the writing of the document, were carried out by Beatriz García Castellanos. The coordination and organization of the work were done by Osney Pérez Ones and Lourdes Zumalacárregui de Cárdenas. Luis López de la Maza supported the analysis of the information. Idania Blanco Carvajal supported the interpretation of the aging process. Likewise, all authors made significant contributions to the document and agree with its publication and state that there are no conflicts of interest in this study.

REFERENCES

[1]   Ministerio de la industria alimentaria. Resolución No. 12/19. La Habana, Cuba, 2019.

[2]   J. Marcano, "El ron. Tradición en Las Antillas," 2019. https://mipais.jmarcano.com/economia/ron.html

Beatriz García-Castellanos; Osney Pérez-Ones; Lourdes Zumalacárregui-de-Cárdenas; Idania Blanco-Carvajal; Luis-Eduardo López-de-la-Maza

[3]   I. Witten, and E. Frank. *Data Mining*. Amsterdam: Morgan Kaufmann Publishers, 2005.

[4]   M. Servente, "Algoritmos TDIDT aplicados a la minería de datos inteligente". Tesis de grado, Universidad de Buenos Aires, Buenos Aires, Feb. 2002.

[5]   L. Zumalacárregui, O. Pérez, F. Hernández, G. Cruz, G. "Modelación del equilibrio líquido-vapor a presión constante de mezclas etanol-agua utilizando redes neuronales artificiales," *Tecnología Química*, vol. 38 (3), pp. 527-548, Jul. 2018.

[6]   H. Abdi, and L. J. Williams "Principal component analysis," *Wiley Interdisciplinary Reviews: Computational Statistics*, vol. 2 (4), pp. 433-59, Jul.-Aug.2010. https://doi.org/10.1002/wics.101

[7]   E. F. Caicedo, and J. A. López, *Una aproximación práctica a las redes neuronales artificiales* Santiago de Cali: Programa Editorial Universidad del Valle, Edición digital, Jul. 2017.

[8]   J. D. Terán-Villanueva, S. Ibarra-Martínez, J. Laria-Menchaca, J. A. Castán-Rocha, M. G. Treviño-Berrones, A. H. García-Ruiz, and J. E. Martínez-Infante, "Estudio de redes neuronales para el pronóstico de la demanda de asignaturas," *Revista Facultad de Ingeniería*, vol. 28 (50), pp. 30-39, Ene. 2019. https://doi.org/10.19053/01211129.v28.n50.2019.8783

[9]   M. A Korany, H. Mahgoub, O. Fahmy, and H. Maher, "Application of artificial neural networks for response surface modelling in HPLC method development," *Journal of Advanced Research*, vol. 3 (1), pp. 53-63, Jan. 2012. https://doi.org/10.1016/j.jare.2011.04.001

[10] L. E. López, L. Zumalacárregui, O. Pérez, and O. Llanes, "Obtención de un modelo neuronal para la estimación de la concentración de etanol en la destilería Héctor Molina," *Tecnología Química*, vol. 38 (2), pp. 315-325, May. 2018.