

Formulation of Special Fats by Neural Networks: A Statistical Approach

J.M. Block^a, D. Barrera-Arellano^{a,*}, M.F. Figueiredo^b, F.C. Gomide^b, and L. Sauer^c

^aLaboratório de Óleos e Gorduras, FEA-UNICAMP, Campinas, São Paulo, ^bDepartamento de Engenharia de Computação e Automação Industrial, FEE-UNICAMP, Campinas, São Paulo, and ^cDepartamento de Computação e Estatística, CCET-UFMS, Campo Grande, Mato Grosso do Sul, Brazil

ABSTRACT: In the present work, a neural network able to formulate fats with three ingredients derived from soybean (one refined oil and two hydrogenated base stocks) was built and trained. The training of the network was accomplished with data on the solid fat content (SFC) of 112 products, association with the proportions of the raw material used in their formulation. After the training, the network furnished, from the requested solid profiles, the possible formulations for the desired product. According to the statistical analysis applied to the results obtained, larger mean errors were observed in products with very low SFC and the smallest errors were found in products with high SFC. Regarding different temperatures, the network performance was more accurate for 10, 20, and 25°C than for 30, 35, and 37.5°C, where the lower measurements resulted in larger relative errors. According to evaluation by industrial experts, all the responses furnished by the network after its training were considered within the acceptable variation limits. For these experts, the network knowledge generalization (accomplished with products not presented during the training) was considered highly efficient (nearly 100%).

Paper no. J9065 in *JAACS* 76, 1357–1361 (November 1999).

KEY WORDS: Blending, fat formulation, hydrogenated fats, neural networks, shortenings.

The fats and oils industry, following a worldwide trend, is using more and more computers in the automation of their facilities and processes. These procedures have as a goal the production of higher-quality and lower-cost products for an ever more competitive market. Besides the automation reached through computation, artificial intelligence has been proposed for process control in widely diverse fields (1). Neural networks, one of the branches of artificial intelligence, have already been used in the fats and oils field to classify different kinds of oils (2,3), to characterize and classify edible oil blends (4), and to detect olive oil and butter adulterations (5,6).

Despite developments in the computer science field and the incorporation of use computer in the 1970s and 1980s, the

production of special fats through blending still depends a great deal on formulators and makes use of many trial-and-error procedures (7–9). It was demonstrated in a previous work that neural networks could be used to formulate hydrogenated fats in a fast and efficient manner (10).

In the present work, a similar neural network was built (10), trained with 112 products that were formulated with different raw materials, also derived from soybean. The network performance was statistically evaluated and by experts from the fat-formulating industry.

EXPERIMENTAL PROCEDURES

Raw materials. Two hydrogenated bases derived from soybean oil and one refined soybean oil were furnished by Sadia S.A. (Paranagua, Brazil). Characteristics of these products are presented in Table 1.

Equipment. A Pentium II computer, 266 MHz, 32Mb, 3.0 Gb HD (Blaster); pulsed nuclear magnetic resonance spectrometer (Minispec PC 120; Bruker, Karlsruhe, Germany); and an infrared spectrophotometer (model 1600 Series FTIR; Perkin Elmer, Norwalk, CT).

Analytical methods. Iodine value, AOCS Cd 1b-87 (11); slip melting point, AOCS Cc 3-25 (11); Isolated *trans* isomers, AOAC (12); and solid fat content (SFC), AOCS Cd 16-81 (11), serial method, temperatures: 10, 20, 25, 30, 35, and 37.5°C.

Neural network. The topology of the network used and its main characteristics were described by Block *et al.* (10).

Training, operation, learning verification, and network efficiency. The methodology used for the training operation, learning verification, and knowledge generalization of the network are as follows: (i) Formulation of 112 products from three described raw materials; (ii) determination of the SFC in the formulated products; (iii) network training having as input data the determined solid profiles, and the proportion of raw material as the output data; (iv) inserting of the trained network in a computational system: interface network/user; (v) learning verification through the solicitation of the formulation of the products used in training, when the response of the minor error is experimentally verified; and (vi) efficiency verification in generalizing unknown data, asking the formu-

*To whom correspondence should be addressed at Laboratório de Óleos e Gorduras, Caixa Postal 6091, Faculdade de Engenharia de Alimentos, UNICAMP, CEP 13083-970, Campinas, SP, Brazil.
E-mail: daniel@fea.unicamp.br

TABLE 1
Characteristics of the Raw Material^a

Raw material	IV	SP	Trans (%)	SFC (%)					
				10°C	20°C	25°C	30°C	35°C	37.5°C
Soybean oil	124.9	ND	ND	ND	ND	ND	ND	ND	ND
Base 1	86.2	32.1	36.0	35.7	18.9	11.1	3.6	0.0	0.0
Base 2	64.7	41.9	46.8	66.9	47.9	39.8	30.8	14.2	7.5

^aIV, iodine value; Trans, trans fatty acids (%); SP, softening point, °C; SFC, solid fat content; ND, not determined.

TABLE 2
Classification of Products Used in Network Training According to the Solid Fat Content at 10°C

Solid fat contents at 10°C (%)	Number of products	%
Class 1 (0–9.9)	9	8.0
Class 2 (10–19.9)	19	17.0
Class 3 (20–29.9)	27	24.1
Class 4 (30–39.9)	26	23.2
Class 5 (40–49.9)	21	19.7
Class 6 (50–59.9)	9	8.0
Class 7 (>59.9)	1	0.8
Total	112	100

lation of 17 products, that, although belonging to the example universe, were not used for training.

Statistical analysis data. Variance analysis was carried out with Microsoft Excel v. 97 (Richmond, WA), based upon the absolute value of the relative error between the network suggested SFC values and the experimental values obtained with the network furnished formulation:

$$\text{relative error (\%)} = \left| \frac{(\text{suggested value} - \text{experimental value})}{\text{suggested value}} \times 100 \right| \quad [1]$$

TABLE 3
Some Examples^a of Formulated Products Used in Learning and Learning Verification of the Neural Network

Product	Formulation	Solid fat content (%)						
		10°C	20°C	25°C	30°C	35°C	37.5°C	
1	Solicited	(0:90:10)	57.35	39.18	30.73	23.93	12.34	7.15
	Determined	(5:87.3:7.7)	56.23	39.84	33.54	24.76	12.19	6.66
2	Solicited	(0:80:20)	48.51	32.66	25.47	19.65	9.48	5.35
	Determined	(10:74.7:15.3)	47.67	33.01	27.30	19.22	8.68	4.79
3	Solicited	(0:70:30)	40.22	26.73	20.23	14.63	7.04	3.60
	Determined	(0:69.6:30.3)	39.57	26.97	22.34	15.25	6.81	3.66
4	Solicited	(0:60:40)	33.66	21.32	15.37	11.19	4.78	2.15
	Determined	(4.9:57.2:37.8)	33.11	21.94	17.99	11.34	5.29	2.23
5	Solicited	(0:50:50)	26.71	15.73	11.40	7.35	2.69	0.96
	Determined	(14.9:40.6:44.5)	26.43	15.87	12.38	7.55	2.58	0.90
6	Solicited	(0:40:60)	20.52	11.99	8.26	7.80	1.62	0.37
	Determined	(0:41:59)	20.91	12.06	9.11	4.91	1.65	0.32
7	Solicited	(0:30:70)	14.25	7.31	5.12	2.76	0.42	0.00
	Determined	(0:30.1:69.9)	13.87	7.18	5.31	2.67	0.54	0.00
8	Solicited	(0:20:80)	9.55	4.49	2.17	1.10	0.00	0.00
	Determined	(25.1:5.7:69.2)	9.44	3.74	1.72	0.79	0.15	0.00
9	Solicited	(0:10:90)	4.16	1.52	0.72	0.00	0.00	0.00
	Determined	(0:8.3:91.6)	3.42	1.15	0.51	0.00	0.00	0.00
10	Solicited	(10:80:10)	53.01	35.89	28.41	21.26	10.48	5.69
	Determined	(20:73.6:6.4)	53.01	35.75	30.50	21.60	10.09	5.12

^aThe complete data (112 examples) can be found on the Fats & Oils Lab home page at: <http://www.fea.unicamp.br/dta/conteudo/oleos/table5.htm>.

Statistical analysis was also accomplished by excluding outliers (relative mean error greater than 100%), which could result in dubious data analyses. To facilitate the analysis of the results, the universe of the products used in the training step of the network was classified according to the SFC at 10°C (Table 2).

Network expert evaluation. Two experts with considerable experience in fat formulation currently working in Brazilian fat products industries have analyzed the results based on acceptable variation ranges for the product solid profiles. The variation range was of 2.5, 2.0, 2.0, 1.5, 1.0, and 0.5% for temperatures of 10, 20, 25, 30, 35, and 37.5°C, respectively.

RESULTS AND DISCUSSION

Some examples of formulated products used in learning and learning verification of the neural network are presented in Table 3; complete data (112 examples) can be found at: <http://www.fea.unicamp.br/dta/conteudo/oleos/table5.htm>. The term “Solicited” refers to the SFC blend profiles used in the network training and for which the network will furnish a response of formulation (response of minor mean error) corresponding to the proportion in percentage of each raw mate-

rial. The proposed formulation by the network was elaborated and the solid profiles experimentally determined (“Experimental” term). The network was built to furnish, besides the proportion of each raw material, the expected solid profiles for each product (theoretical data, not presented).

According to the results, among the 112 products formulated by the network, 85 formulations (75.9%) were different from those presented during the training step, indicating that the network goes through its own way to solve the problem that is presented to it.

According to the variation analysis applied to the 112 products of the network, a significant difference was observed ($P = 0.005$) among the products 95, 98, 44, 112, 49 and the others.

Products 49, 112, and 44 presented as a common characteristic, low SFC at 10°C (11.8; 6.1, and 10.6%, respectively), typical values of fluid fats. Products 98 and 95 presented typical solid profiles of soft spreads and soft margarines, with SFC around 20% at 10°C; product 95, despite the high mean error, presented a characteristic graph of what formulators have agreed to call “perfect melting.” This type of graph presents the following profile: 5°C = 27%; 10°C = 22%; 20°C = 10%; 30°C = 2%; 35°C = 0%, and 40°C = 0%.

The mean errors presented by these products, which can vary from 455.6% (blend 95) to 1083.0% (blend 49), were higher at temperatures between 30 and 37.5°C. At these temperatures, the solid contents of these products were very close or equal to zero, resulting in very high relative errors, not necessarily meaning an exaggerated or unacceptable difference among the results.

The smaller mean error was observed for product 66 (0.8%), characterized as a hard product with high solids content (49.5% at 10°C and 4% at 37.5°C).

For the experts, all the responses furnished by the network were considered within the defined specifications for the formulated products. In some cases (35 products), the experimental response was better than that suggested by the network in all or in most of the temperatures (products 6, 7, 9, 10, 12, 18, 20, 24, 31, 39, 43–51, 53, 55, 63, 65, 66, 86, 88, 90, 91, 98, 101, 103, 104, 108, 110, 112).

Among the obtained results, the following products were considered outliers: 45, 8, 39, 54, 95, 98, 44, 112 and 49, with mean errors between 124.6 (45) and 1083.0% (49). These products presented very varied solid profiles, with SFC between 6.11 and 30.12% at 10°C. Nevertheless, all of them presented high relative errors implied by the SFC equal to zero at the temperatures of 30, 35, and 37.5°C, where, obtained through formulation with the network, responses were obtained between 0.13 and 0.33%. Although these products were classified as atypical, all the results were considered adequate, no response having unacceptable values. According to the statistical analysis accomplished after excluding these results, significant differences among the responses were not obtained for the products.

The mean errors obtained by the network for the different classes of products are presented in Figure 1. The majority of

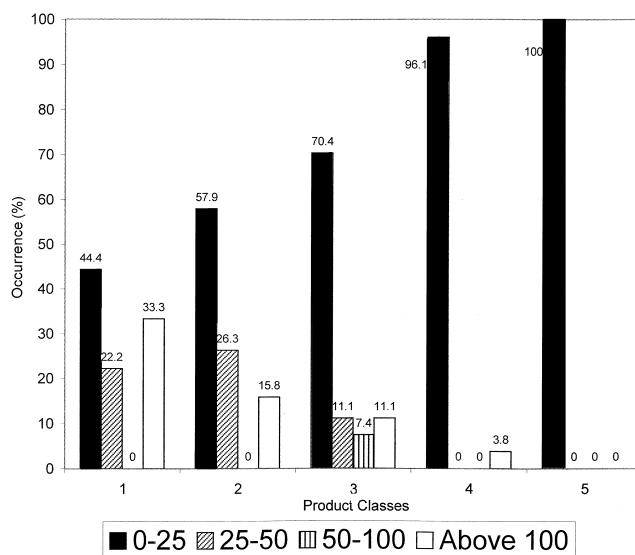


FIG. 1. Mean errors presented according to the different classes of products.

the responses presented mean errors of $\leq 25\%$, representing 100% of the responses in the classes 5 and 6. Mean errors above 100% occurred in the classes 1, 2, 3 and 4.

The performance of the network related to temperatures for (i) all the results and (ii) the atypical data can be observed in Table 4.

In the statistical analysis for all the results, a significant difference was observed ($P = 0.0003$) between 35°C and 10, 20, and 25°C. The mean errors found varied between 3.4 (10°C) and 164.9% (35°C). A significant difference was detected in the analysis after excluding the atypical data ($P = 7.2 \times 10^{-11}$) between the temperatures of 35 and 37.5°C and the other temperatures.

The performance of the network at the different temperatures, with and without the atypical data, was similar, although after excluding the atypical data, the errors at 30, 35 and 37.5°C were much smaller, indicating a less accurate performance for these products at these temperatures.

The results obtained by the network in relation to its efficiency in generalizing knowledge, can be observed in Table 5.

According to statistical analysis, no difference was observed among the products ($P = 0.46$). The smaller absolute mean error (AME) (2.28%) was observed in product 10, of hard consistency [AE = absolute error = (solicited – experi-

TABLE 4
Performance of the Network at the Temperatures Studied (with and without the outliers)^a

Mean error	Temperatures (°C)					
	10	20	25	30	35	37.5
For all data (%)	3.4 ^a	3.8 ^a	8.3 ^a	44.5 ^{a,b}	164.9 ^{a,b}	101.9 ^b
Without outliers (%)	3.3 ^a	3.8 ^a	8.1 ^a	9.1 ^a	28.3 ^b	24.6 ^b

^aDifferent superscript roman letters mean significant differences among the samples.

TABLE 5
Network Efficiency in Generalizing Data from Unknowns (i.e., blends not used in training)

Product	Blend ^a	Solid fat content (%)						
		10	20	25	30	35	37.5°C	
1	Solicited	(5:5:90)	3.74	1.15	0.36	0.00	0.00	0.00
	Suggested	(0.07:7.29:92.64)	4.70	1.90	1.10	0.50	0.30	0.10
	Experimental	(0.07:7.29:92.64)	3.07	1.07	0.45	0.14	0.00	0.00
2	Solicited	(10:15:75)	10.19	3.97	2.74	1.14	0.00	0.00
	Suggested	(25.11:6.02:68.87)	9.70	4.20	2.50	1.20	0.40	0.20
	Experimental	(25.11:6.02:68.87)	9.93	3.50	1.45	0.62	0.00	0.00
3	Solicited	(15:25:60)	17.24	8.64	6.11	3.35	1.01	0.29
	Suggested	(20.37:21.93:57.70)	17.30	8.80	5.80	3.10	1.00	0.40
	Experimental	(20.37:21.93:57.70)	17.29	8.65	5.99	3.04	0.60	0.00
4	Solicited	(15:35:50)	23.60	13.77	10.16	6.05	2.14	0.59
	Suggested	(4.94:42.19:52.88)	23.50	13.80	10.10	6.20	2.10	1.00
	Experimental	(4.94:42.19:52.88)	24.26	14.50	11.17	6.64	2.14	0.94
5	Solicited	(15:55:30)	36.13	23.48	18.92	12.12	4.89	2.44
	Suggested	(9.97:57.89:32.14)	36.00	23.30	18.30	12.40	4.80	2.30
	Experimental	(9.97:57.89:32.14)	36.46	23.49	18.86	12.43	4.89	2.26
6	Solicited	(20:5:75)	7.96	3.13	1.25	0.23	0.02	0.00
	Suggested	(14.98:8.29:76.73)	7.90	3.40	2.00	0.90	0.40	0.20
	Experimental	(14.98:8.29:76.73)	7.97	3.01	1.33	0.72	0.07	0.00
7	Solicited	(20:35:45)	25.32	14.47	11.05	6.39	1.81	0.63
	Suggested	(19.84:35.57:44.59)	25.40	14.60	10.50	6.20	2.10	0.90
	Experimental	(19.84:35.57:44.59)	26.10	14.95	11.32	6.51	1.97	0.84
8	Solicited	(25:15:60)	15.34	7.14	4.57	2.05	0.06	0.00
	Suggested	(40.43:6.25:53.32)	15.20	7.00	4.30	2.00	0.60	0.30
	Experimental	(40.43:6.25:53.32)	15.52	6.98	3.89	1.20	0.05	0.00
9	Solicited	(25:60:15)	45.38	29.80	24.66	16.21	6.88	3.82
	Suggested	(30.08:58.87:11.05)	46.70	30.60	24.40	16.60	6.80	3.40
	Experimental	(30.08:58.87:11.05)	48.98	31.29	25.32	17.21	6.99	3.75
10	Solicited	(25:65:10)	49.73	32.60	26.41	18.64	7.92	4.30
	Suggested	(29.94:62.80:7.26)	49.80	33.00	26.60	18.40	7.70	4.00
	Experimental	(29.94:62.80:7.26)	50.64	32.94	26.80	18.92	7.50	4.41
11	Solicited	(35:30:35)	28.73	16.50	12.53	6.96	2.28	0.86
	Suggested	(39.63:28.67:31.70)	29.20	16.70	11.90	6.70	2.10	0.90
	Experimental	(39.63:28.67:31.70)	29.44	16.77	12.60	6.94	1.84	0.22
12	Solicited	(40:15:45)	20.64	10.73	7.29	3.40	0.36	0.00
	Suggested	(40.60:14.73:44.67)	20.70	10.50	6.80	3.40	0.36	0.00
	Experimental	(40.60:14.73:44.67)	20.71	9.94	6.62	3.35	0.39	0.00
13	Solicited	(40:35:25)	33.71	19.97	15.40	8.77	2.61	0.95
	Suggested	(39.60:35.14:25.26)	33.50	20.00	14.70	8.80	2.90	1.30
	Experimental	(39.60:35.14:25.26)	33.03	19.39	14.58	8.56	2.64	0.57
14	Solicited	(45:15:40)	23.42	12.23	8.23	4.04	1.01	0.11
	Suggested	(40.29:19.03:40.68)	23.30	12.30	8.20	4.30	1.30	0.60
	Experimental	(40.29:19.03:40.68)	23.83	12.35	8.72	4.36	1.28	0.00
15	Solicited	(45:40:15)	40.57	25.16	19.81	12.18	4.50	2.09
	Suggested	(40.19:44.58:15.22)	40.50	25.50	19.60	12.50	4.50	2.10
	Experimental	(40.19:44.58:15.22)	42.44	26.72	21.35	13.55	4.92	2.83
16	Solicited	(60:15:25)	28.87	15.23	11.16	5.33	1.10	0.40
	Suggested	(67.42:10.06:22.52)	28.80	15.20	10.00	4.90	1.40	0.60
	Experimental	(67.42:10.06:22.52)	29.15	15.44	10.51	5.05	0.90	0.00
17	Solicited	(65:25:10)	38.72	22.63	17.23	9.78	3.13	1.27
	Suggested	(60.42:29.59:9.99)	39.00	23.40	17.20	10.00	3.30	1.50
	Experimental	(60.42:29.59:9.99)	39.96	23.24	17.57	10.50	3.14	1.11

^aBlend = (base 1: base 2: oil).

mental)/solicited; AME = $(AE_{10} + AE_{20} \dots)/6$. The larger mean error observed was for product 1 (225%), which is liquid at room temperature. Among the 17 products only product 1 presented an error greater than 100%.

Among the temperatures, no difference in performance was observed, whereas the mean errors varied from 3.2 to

98.1% representing the following sequence (in increasing order of errors): 10, 20, 25, 35, 37.5, and 30°C.

Confirming the high network generalization capacity, the experts considered that all the results for the formulated products not used in the training step presented solid profiles (SFC) within the required specifications.

Using 112 products in training seems not to have resulted in a superior performance of the network over a network trained with only 63 products (10). These results indicate that, more than a high number of samples used in the training, the choice of the raw materials is fundamental for its success.

Although statistical analysis has indicated quite high relative errors for some products, the analyses by the experts indicated a very high network efficiency, either for learning or for data generalization. The experts' analyses were based on a very restricted range of variation, most commonly used for table products. In the case of fats of industrial use the allowed range of variation also changes according to the temperature, depending on the use, but generally is much larger than that considered in the present work. This research demonstrates that neural networks can be a powerful tool in the formulation of special fats.

REFERENCES

1. Brito, L.A.L., Automated Total Quality Systems, *INFORM 4*: 829–832 (1993).
2. Romero, R.A.F., F.M. Lanças, S.J. Guizo, and S.R. Berton, Classification of Edible Oils Using Neural Networks, in *Proceedings of International Meeting on Fats and Oils Technology—Symposium and Exhibition*, Campinas, São Paulo, 9–11 July, edited by D. Barrera-Arellano and L.A.G. Golçalves, Campinas State University, Fats & Oils Lab, São Paulo, 1991, pp. 82–86.
3. Zupan, J., M. Novic, X. Li, and J. Gasteiger, Classification of Multi Component Analytical Data of Olive Oils Using Different Neural Networks, *Anal. Chim. Acta* 292:219–234 (1994).
4. Hussain, S., K.S. Devi, and P.J. Reddy, Characterization and Identification of Edible Oil Blends and Prediction of the Composition by Artificial Neural Networks—A Case Study, *Chemom. Intell. Lab. Syst.* 35:117–126 (1996).
5. Goodacre, R., D.B. Kell, and G. Bianchi, Rapid Assessment of the Adulteration of Virgin Olive Oils by Others Seed Oils Using Pyrolysis Mass Spectrometry and Artificial Neural Networks, *J. Sci. Food Agric.* 63:297–307 (1993).
6. Lipp, M., Determination of the Adulteration of Butter Fat by Its Triglyceride Composition Obtained by CG. A Comparison of the Suitability of PLS and Neural Networks, *Food Chem.* 55: 389–395 (1996).
7. Lefebvre, J., Finished Product Formulation, *J. Am. Oil Chem. Soc.* 60:295–300 (1983).
8. Smallwood, N.J. Using Computers for Oil Blending, *Ibid.* 66: 644–648 (1989).
9. Erickson, D.R., and M.D. Erickson, Hydrogenation and Base Stock Formulation Procedures, in *Practical Handbook of Soybean Processing and Utilization*, edited by D.R. Erickson, Champaign, AOCS Press, Champaign, 1995, pp. 218–238.
10. Block, J.M., D. Barrera-Arellano, M.F. Figueiredo, and F.A.C. Gomide, Blending Process Optimization into Special Fat Formulation by Neural Networks, *J. Am. Oil Chem. Soc.* 74: 1537–1541 (1997).
11. *Official Methods and Recommended Practices of the American Oil Chemists' Society*, 4th edn., edited by D. Firestone, American Oil Chemists' Society, Champaign, 1989.
12. *Official Methods of Analysis of the Association of Official Analytical Chemists*, 14th edn., edited by S. Williams, AOAC, Arlington, 1984.

[Received November 10, 1998; accepted June 24, 1999]