Artificial Neural Networks for Diagnostics of Water-Ethanol Solutions by Raman Spectra

Igor Isaev[1,2(✉)], Sergey Burikov[1,2], Tatiana Dolenko[1,2], Kirill Laptinskiy[1,2], and Sergey Dolenko[1]

[1] D.V. Skobeltsyn Institute of Nuclear Physics, M.V. Lomonosov Moscow State University, Moscow, Russia
isaev_igor@mail.ru, dolenko@sinp.msu.ru
[2] Physical Department, M.V. Lomonosov Moscow State University, Moscow, Russia

Abstract. The present paper is devoted to an elaboration of a method of diagnosis of alcoholic beverages using artificial neural networks: the inverse problem of spectroscopy – determination of concentrations of ethanol, methanol, fusel oil, ethyl acetate in water-ethanol solutions – was solved using Raman spectra. We obtained the following accuracies of concentration determination: 0.25% vol. for ethanol, 0.19% vol. for fusel oil, 0.35% vol. for methanol, and 0.29% vol. for ethyl acetate. The obtained results demonstrate the prospects of using Raman spectroscopy in combination with modern data processing methods (artificial neural networks) for the elaboration of an express non-contact method of detection of harmful and dangerous impurities in alcoholic beverages, as well as for the detection of counterfeit and low-quality beverages.

Keywords: Neural networks · Inverse problems · Raman spectroscopy
Water-ethanol solutions

1 Introduction

At present, when there is a significant amount of danger to health low-quality products on the market, it is necessary to control the quality of alcoholic beverages. Consumption of counterfeit alcoholic drinks is dangerous for life because even a small amount of toxic impurities (methyl alcohol, fusel oils, etc.) can cause intoxication of the human organism – allergenic, immunomodulatory, genotoxic actions. One of the most famous strong alcoholic beverages is vodka produced by fermentation and distillation of grain, potatoes, sugar beet, grape, etc. [1]. Vodka is a water solution of ethanol with a concentration of 37.5–56% vol. (Russia GOST standard 12712-2013); therefore, the development of express and non-contact methods of diagnostics of water-ethanol solutions is an extremely urgent task.

Study performed at the expense of Russian Science Foundation, project no.14-11-00579.

B. Kryzhanovsky et al. (Eds.): NEUROINFORMATICS 2018, SCI 799, pp. 167–175, 2019.
https://doi.org/10.1007/978-3-030-01328-8_18

Control of quality of strong alcoholic beverages assumes the solution of two separate problems: (1) determination of content (concentration) of ethyl alcohol; (2) determination of potentially dangerous impurities and their concentrations. There are many methods of determination of the content of ethyl alcohol in water-ethanol solutions: by measuring the density of the sample [2], refractive index [3], boiling point of the solutions [4], chromatographic method [5], method of nuclear magnetic resonance (NMR) [6, 7] etc. Non-contact means of vibrational spectroscopy – IR absorption spectroscopy [8] and Raman spectroscopy [9] - are also widely used. Most of them are also used to solve the second problem – determination of the concentrations of impurities dangerous to a human in alcoholic beverages. First of all, it applies to chromatography, NMR, methods of vibrational spectroscopy.

It should be noted that the methods based on the measurement of the density of the solution or its refractive index, in the presence of a large number of impurities in the solution are not precise enough. Chromatographic analysis, chemical methods, and NMR method provide high accuracy of determination of substances in solutions, but they are time-consuming and expensive. Besides, these methods are contacting, i.e., they require opening the container and extracting a certain amount of sample from it.

For practical implementation, the method of Raman spectroscopy favorably differs from the other mentioned methods. It is non-contact, express, we can apply it in real time mode, it does not require complicated sample preparation and expensive reagents, so it is widely used for the diagnostics of various drinks. Thus, Raman spectroscopy is used for determination of the content of glucose, sucrose and fructose in food drinks [10], for studying the structure of water-ethanol solutions [11, 12], for determination of the content of ethanol and methanol in alcoholic beverages [13], for identification of various types of whiskey [14].

However, in the presence of a large number of various impurities in drinks, the solution of the inverse problem (IP) of Raman spectroscopy – determination of the type and concentrations of contaminants in multicomponent water-ethanol solutions – is significantly complicated. Therefore, instead of conventional methods of solution of such multi-parameter ill-posed IPs, adaptive methods of data analysis are used. They provide acceptable solutions [15, 16]; e.g., artificial neural networks (ANN) are used to solve such IPs [17]. ANN are actively used in solution of a wide range of problems associated with pattern recognition, forecasting, classification, etc. In particular, ANNs are rather widely used in spectroscopy. So, ANNs were used for determination of the type and concentration of salts dissolved in water by Raman spectra [15, 16, 18], for express determination of components of wine by the absorption spectra [19], for determination of blood glucose by the spectra of IR absorption [20].

In this study, the problem of diagnostics of water-ethanol solutions by Raman spectra included (1) determination of the concentration of ethyl alcohol in the solution and (2) determination of the concentration of harmful impurities in vodka. The method of ANN was used to solve these problems.

2 Problem Statement and Data Preparation

The impurities most commonly occurring in counterfeit and low-quality alcohol – methanol (the main cause of poisoning by counterfeit alcohol), fusel oil, and ethyl acetate, – were used as harmful substances. Since the main components of the fusel oil are isoamyl and isopropyl alcohols, a mixture of 70/30 (by volume) of isoamyl and isopropyl alcohols was used as a model of fusel oil. Methanol, fusel oil, ethyl acetate were dissolved in the mixture of ethanol and water (concentrations 35, 38, 40, 42, 45, 49, 53, 57%). In this way, vodkas of various strengths were simulated. The following impurity concentrations were used: methanol – 0, 0.05, 0.14, 0.4, 1.1, 3.1, 8.6, 24%; fusel oil - 0, 0.025, 0.07, 0.22, 0.66, 2, 6, 18%; ethyl acetate – 0, 0.17, 0.35, 0.7, 1.4, 2.8, 5.6, 11.2%. For the preparation of the solutions, ethyl alcohol category "alpha," isoamyl alcohol category "pure for analysis," isopropyl alcohol category "chemically pure," ethyl acetate category "pure," methyl alcohol category HPLC (suitable for high-performance liquid chromatography) were used.

The Raman spectra were obtained using a laser spectrometer which included an argon laser (488 nm wavelength, 200 mW power) and a registration system consisting of a monochromator (grating 900 grooves/mm, 500 mm focal length) and a cooled CCD camera. Edge-filter was used to suppress the elastic scattering. Spectra were recorded in two ranges with centers at 520 nm (low-frequency region of the spectrum or the region of so-called "fingerprints") and 573 nm (the region of stretching vibrations). As a result, we obtained the full Raman spectrum in the range of 200–3800 cm^{-1}. The useful spectral resolution was 2 cm^{-1}. For each of the ranges, 10 spectra were recorded (two cycles of measurements, 5 spectra each). The acquisition time for one spectrum was 2 s. The obtained 10 spectra were averaged.

The principle possibility of determining the concentration of ethanol and harmful impurities is caused by the fact that ethyl, methyl, isoamyl, isopropyl alcohols and ethyl acetate have their specific lines in the "fingerprint" spectral region (200–1600 cm^{-1}) and in the area of valence vibrations of CH- and OH-groups (2600–3800 cm^{-1}) (Fig. 1). In the case of a single-component solution, one or several Raman bands specific to the component can be selected, and its concentration can be determined by the intensity of this band [13]. In the case of multicomponent solutions, the situation is complicated by the fact that the vibrational bands of different substances overlap.

Among the negative factors that reduce the accuracy of the solution of the problem in real conditions (in the diagnostics of real drinks) are also detector noise, fluorescent pedestal caused by impurities, and fluctuations in the intensity of the Raman signal, produced by the instability of the laser radiation power. The ANN method can successfully overcome these difficulties due to the ability of the ANN to learn, to summarize the provided information and to identify hidden patterns.

The dataset prepared to implement the ANN solution of the problem contained 4046 patterns. We divided it into training, validation and test sets in the ratio of 70:20:10, respectively. Thus, the training set contained 2800 patterns, validation set – 800 patterns, test set – 446 patterns. Training of the ANN was performed on the training set, stop at the minimum error on the validation set was used to prevent overtraining, and independent evaluation of the results was made on the test set.

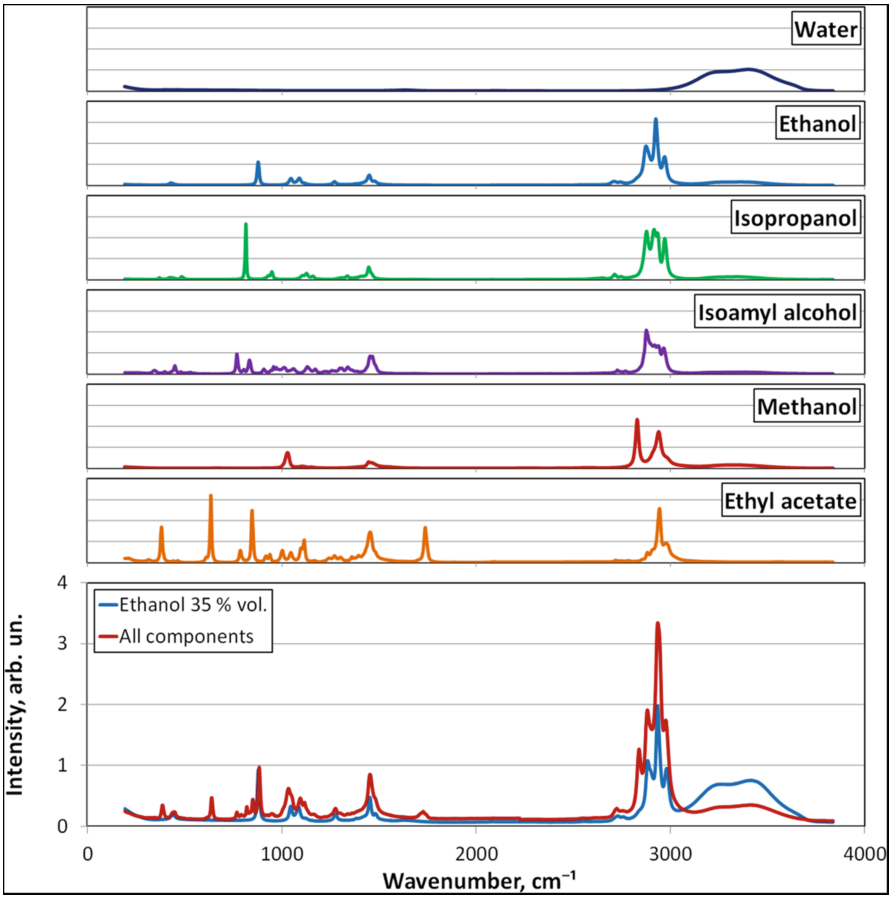

Fig. 1. Top – Raman spectra of pure substances. Bottom – examples of Raman spectra of the studied water-ethanol solutions.

Thus, the regression problem stated in this study was to determine the values of 4 parameters by 2048 input characteristics. The sought-for parameters were the concentrations of fusel oil, ethyl acetate, methanol, and ethanol.

3 Selection of the Optimal MLP Architecture

The ANN architecture used to solve this problem was the multi-layer perceptron (MLP). This study included a search for the optimal numbers of layers and neurons, as well as search for the optimal method of selection of significant input features for the MLP. For each configuration of the parameters, 5 MLPs were trained with various weights initialization; the statistical indicators of their application were averaged.

The multi-parameter IP was solved with two ways of parameter determination. For autonomous determination (ADP), a separate single-output ANN was trained for each parameter. For simultaneous determination (SDP), a single 4-output ANN was built.

At the first stage, the results for MLPs with various numbers of hidden layers (HL) and neurons in them were compared. The results are presented in Fig. 2.

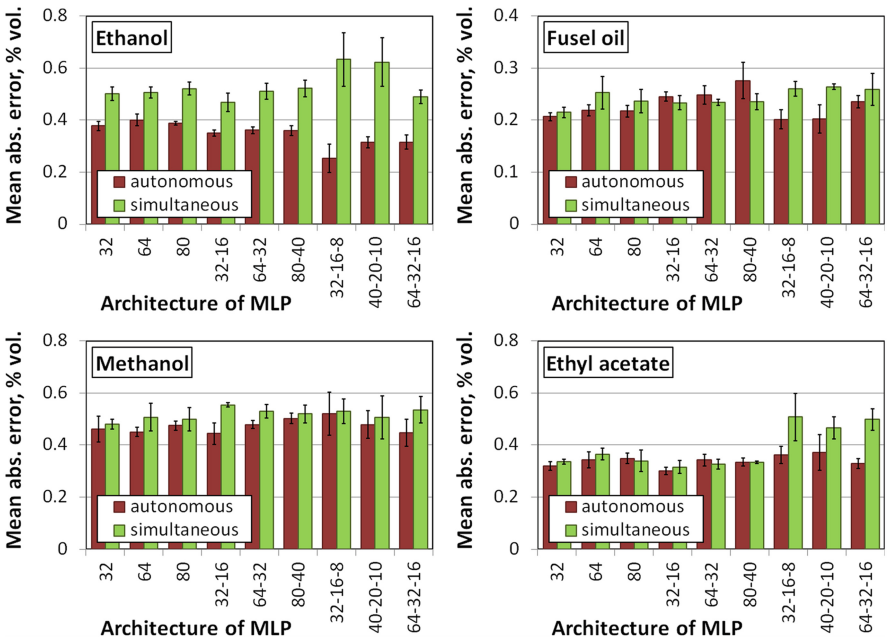

Fig. 2. Mean absolute error, % vol. for each determined component, for various MLP architectures and for different ways of parameter determination.

It can be seen that ADP gives a consistently better result for ethanol and slightly better results for other components of the solutions. All the considered architectures give comparable results except that MLP with three HL perform better for ethanol with ADP, and show some deterioration with SDP.

The obtained results may indirectly indicate that the complexity of the problem is not very high and that the data set has sufficient representativity.

4 Significant Feature Selection

At the second stage, we considered the selection of significant input features by various methods. The selected features were used to train an MLP with 32 neurons in the single HL. The training was also carried out in the modes of ADP and SDP.

Three of the considered methods were based on the calculation of the significance of interrelation between each of the input features and each of the determined

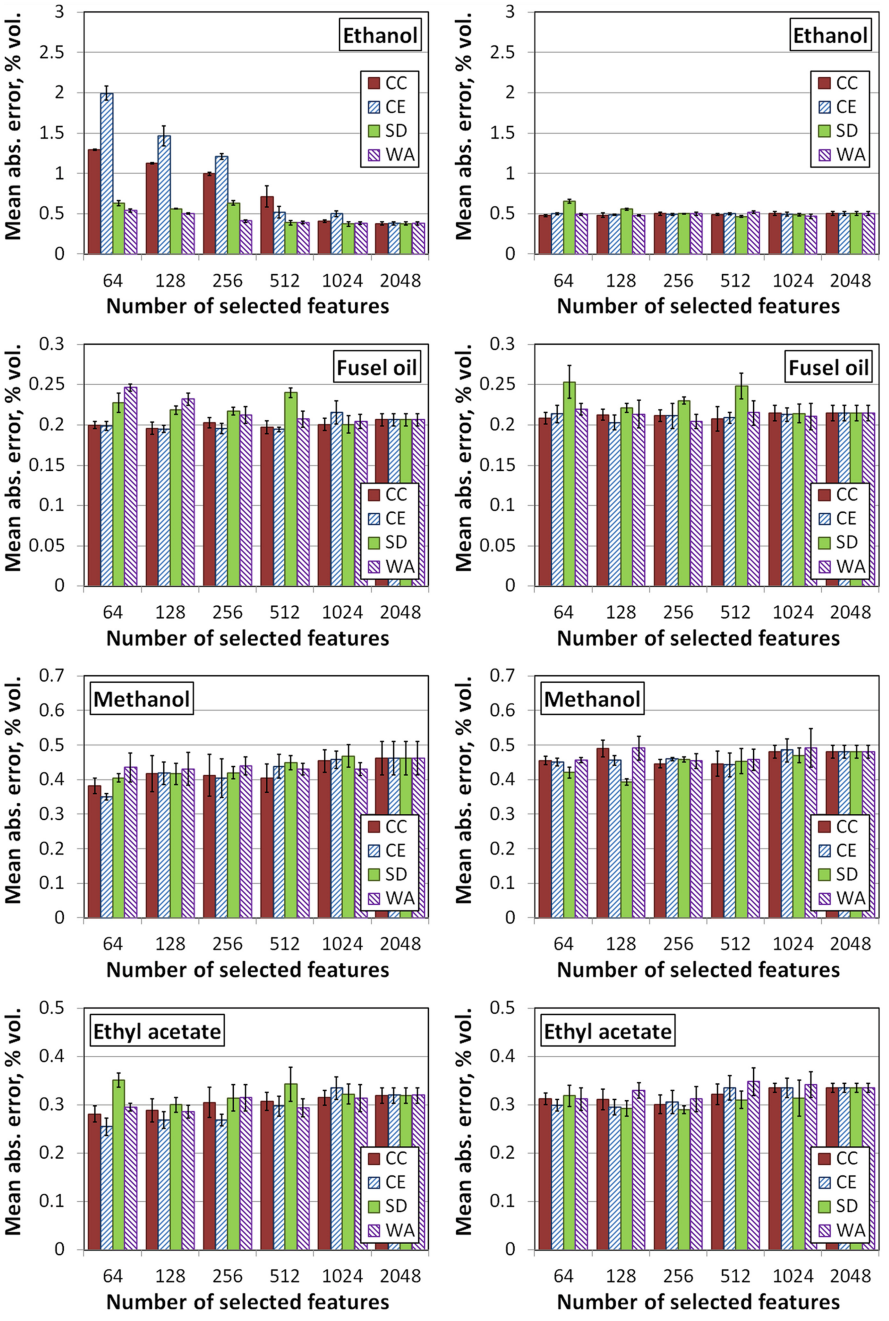

Fig. 3. Mean absolute error, % vol. for each determined component, for various methods of significant feature selection, for various numbers of features. Left – autonomous determination of parameters, right – simultaneous determination of parameters.

parameters. The values of cross-correlation (CC), cross-entropy (CE), and the importance of inputs obtained through the weights analysis of a neural network (WA) were used as the significance value. Those features, the significance value of which exceeded a certain threshold, were selected to be used. The threshold values were set in such a way that a certain pre-defined number of features were selected. For SDP, the network input was fed with all the features selected for at least one of the determined parameters. So, in this case, the number of features was, in fact, higher than for ADP.

The selection of features by the standard deviation (SD) value is based on the assumption that the amount of information carried by the feature is proportional to the entropy, which is, in turn, proportional to SD. Here also the threshold was set in such a way that a certain number of features were selected. As this method uses no information about the outputs, the numbers of features for ADP and SDP were the same.

The results of solving the problem using feature selection are shown in Fig. 3. For fusel oil, methanol and ethyl acetate, feature selection in some cases can slightly reduce the error; the best results are obtained with CE. For ethanol, there is a sharp drop in the quality of the solution when using feature selection through CC and CE in the case of ADP. This fact is since all characteristic lines of the ethanol spectrum overlap with the lines of other components. As a result, without information about the content of other components, it is impossible to accurately determine the concentration of ethanol with a reduced number of features. If such information is received (in the case of SDP), no fall in the quality of the solution is observed.

5 Conclusion

According to the results of the study, the following conclusions can be drawn

- Raman spectroscopy with MLP processing of spectra can be successfully used to detect hazardous impurities in aqueous ethanol solutions.
- Autonomous determination gives a consistently better result than simultaneous for ethanol and slightly better results for fusel oil, methanol, and ethyl acetate.
- All the considered architectures give comparable results except that MLP with three hidden layers perform better for ethanol with autonomous determination, and show some deterioration with simultaneous determination of the parameters.
- Significant feature selection provides a small improvement in the quality of the solution; the best results we obtained with cross-entropy.

The best-obtained accuracy of the determination of component concentrations was: 0.25% vol. for ethanol (enough for practical purposes); 0.19% vol. for fusel oil (less than the maximum permissible concentration (MPC)); 0.35% vol. for methanol and 0.29% vol. for ethyl acetate (several times greater than MPC but more than an order of magnitude less than the lethal concentration for 100 ml of drink).

References

1. Marianski, S., Marianski, A.: Home Production of Vodkas. Infusions & Liqueurs. Bookmagic, LLC (2012)
2. Fuerst, A.: Determination of alcohol by pycnometer. Ind. Eng. Chem. Anal. Ed. **2**(1), 30–31 (1930)
3. Owuama, C., Ododo, J.: Refractometric determination of ethanol concentration. Food Chem. **48**(4), 415–417 (1993)
4. Iland, P., Ewart, A., et al.: Techniques for Chemical Analysis and Quality Monitoring during Winemaking. Patrick Iland Wine Promotions, Campbelltown (2000)
5. Leary, J.: A quantitative gas chromatographic ethanol determination. J. Chem. Educ. **60**(8), 675 (1983)
6. Isaac-Lam, M.: Determination of Alcohol Content in Alcoholic Beverages Using 45 MHz Benchtop NMR Spectrometer. Intern. J. of Spectroscopy 2016, Art.ID 2526946, 8 pp. (2016)
7. Zuriarrain, A., Zuriarrain, J., Villar, M., Berregi, I.: Quantitative determination of ethanol in cider by 1H NMR spectrometry. Food Control **50**, 758–762 (2015)
8. Lachenmeier, D., et al.: Rapid and mobile determination of alcoholic strength in wine, beer and spirits using a flow-through infrared sensor. Chem. Cent. J. **4**, 5 (2010)
9. Burikov, S., Dolenko, T., et al.: Diagnostics of aqueous ethanol solutions using Raman spectroscopy. Atmos. Ocean. Opt. **22**(11), 1082–1088 (2009)
10. Ilaslan, K., Boyaci, I., Topcu, A.: Rapid analysis of glucose, fructose and sucrose contents of commercial soft drinks using Raman spectroscopy. Food Control **48**, 56–61 (2015)
11. Hu, N., Wu, D., et al.: Structurability: a collective measure of the structural differences in vodkas. J. Agric. Food Chem. **58**(12), 7394–7401 (2010)
12. Burikov, S., Dolenko, S., et al.: Decomposition of water Raman stretching band with a combination of optimization methods. Mol. Phys. **108**(6), 739–747 (2010)
13. Boyaci, I., Genis, H., et al.: A novel method for quantification of ethanol and methanol in distilled alcoholic beverages using Raman spectroscopy. J. Raman Spectrosc. **43**(8), 1171–1176 (2012)
14. Kiefer, J., Cromwell, A.: Analysis of single malt Scotch whisky using Raman spectroscopy. Anal. Methods **9**, 511–518 (2017)
15. Burikov, S., Dolenko, S., et al.: Neural network solution of the inverse problem of identification and determination of partial concentrations of inorganic salts in multicomponent aqueous solution. In: Proceedings of XII Russian Scientific-Technical Conference "Neuroinformatics-2010", part 2, pp. 100–110. MEPhI, Moscow (2010)
16. Dolenko, S., Burikov, S., et al.: Adaptive methods for solving inverse problems in laser raman spectroscopy of multi-component solutions. Pattern Recognit. Image Anal. **22**(4), 551–558 (2012)
17. Hassoun, M.: Fundamentals of Artificial Neural Networks, 1st edn. MIT Press, Cambridge (1995)
18. Dolenko, S., Burikov, S., et al.: Neural network approaches to solution of the inverse problem of identification and determination of partial concentrations of salts in multicomponent water solutions. In: Wermter, S., Weber, C., et al. (eds.) ICANN 2014, LNCS, vol. 8681, pp. 805–812. Springer, Heidelberg (2014)
19. Martelo-Vidal, M., Vázquez, M.: Application of artificial neural networks coupled to UV–VIS–NIR spectroscopy for the rapid quantification of wine compounds in aqueous mixtures. CyTA J. Food **13**(1), 32–39 (2015)

20. Liu, W., Yang, W., Liu, L., Yu, Q.: Use of artificial neural networks in near-infrared spectroscopy calibrations for predicting glucose concentration in Urine. In: Huang, D.-S., Wunsch, D.C., Levine, D.S., Jo, K.-H. (eds.) ICIC 2008. LNCS, vol. 5226, pp. 1040–1046. Springer, Heidelberg (2008)
21. International Programme on Chemical Safety (IPCS). Environmental Health Criteria 196: Methanol, pp. 1–9 (1997)
22. Hazardous substances in industry. Handbook for chemists, engineers, and physicians, 7th edn. Volume 1. Organic substances. Lazarev, N., Levina, E. (eds.) Khimija Publ., Leningrad (1977). (in Russian)
23. Gosselin, R., Smith, R., Hodge, H.: Clinical Toxicology of Commercial Products, 5th edn. Williams & Wilkins, Baltimore (1984)
24. Golikov, S. (ed.): Emergency treatment of acute poisoning. Meditsina Publishers, Moscow, 312 pp. (1978). (In Russian)