Modern Chemistry
2014; 2(2): 15-18
Published online May 30, 2014 (http://www.sciencepublishinggroup.com/j/mc)
doi: 10.11648/j.mc.20140202.12

Predicting the melting point of organic compounds consist of carbon, Hydrogen, nitrogen and oxygen using multi layer Perceptron artificial neural networks

Yahya Hassanzadeh-Nazarabadi[1], S. Majed Modaresi[2], S. Bahram Jafari[3], Sanaz Taheri-Boshrooyeh[4]

[1]Mobile Robot Department, Parse Lab of Robotic, Mashhad, Iran
[2]Chemistry Department, Ferdowsi University, Park Sq, Mashad, Iran
[3]Chemistry Department, Tabriz Uneversity, Abresan, Tabriz, Iran
[4]Mobile Robot Department, Parse Lab of Robotic, Mashhad, Iran

Email address:
ya_ha_na@ieee.org (Y. Hassanzadeh-Nazarabadi), ma_modr@hotmail.com (S. M. Modaresi), bjamm66@yahoo.com (S. B. Jafari)

To cite this article:
Yahya Hassanzadeh-Nazarabadi, S. Majed Modaresi, S. Bahram Jafari, Sanaz Taheri-Boshrooyeh. Predicting the Melting Point of Organic Compounds Consist of Carbon, Hydrogen, Nitrogen and Oxygen Using Multi Layer Perceptron Artificial Neural Networks. *Modern Chemistry.* Vol. 2, No. 2, 2014, pp. 15-18. doi: 10.11648/j.mc.20140202.12

Abstract: So far the methods used to predict or calculate the melting point of organic compunds do not focus on the compound nature, they mostly use microscopic physio-chemical properties of materials. In this paper the disadvantage of such traditional methods will be defined. Then a new method is introduced. This method uses the nature properties of compounds to estimate their melting point based on an artificial neural network and offsets the disadvantges of pervious ones.

Keywords: Artificial Neural Networks, Neurons, Matlab 2013, Fitnet Function, Levenberg-Marquart Algorithm

1. Introduction

The temperature at which a phase state of a solid would be changed and becomes a liquid is the melting point.This is one of the most useful properties between chemical compounds which is frequently measured and used. It is one of the most important factor in industry, such as controlling solubility of a compound, formulation of medicines and in pharmacology science and it is widely used in chemical engineering design and experiments[1, 2]. The melting point of solid compounds largly depend on the molecular shape and intermolecular forces between molecules such as london dispersion forces, hydrogen bonding forces and other intermolecular interactions[2].

However a large amount of data are available in handbooks, the thermal instability or impurity of some compounds may deviate measurements of experiment[3].

Since it is not practical to measure the melting point of all synthesis compounds readily, as the need arises, estimation methods are generally coming to be in use. Under such conditions, automized methods and models could be used to provide a logical prediction[1]. So far, most of the models which have been introduced are physio-chemical models based on complicated mathematical equations. Each of these models focus on a series of specefic physio-chemial features, but not the bulk properties of matter. Although it is essential to have knowledge about physio-chemical properties of chemical compounds to predict their behaviour especially melting point[3], it could be predicted by training of a neural network in an easier manner.

In this paper, training neural networks for estimation of melting point is discussed. First previous works will be studied and their advantages and disadvantages will be explained. In the next part design of an artificial neural network for this prediction is introduced.

2. Study of Previous Work

In this part we divide the models of determining melting point into two categories: The classic and the modern one. Classic model was first invented and it is used in laboratory such as DSC and Thiele Tube.

DSC is a Differential Scanning Calorimeter which measures the transferred energy as heat to or from a sample at constant pressure during a physical or chemical change. Originally it would give us the entalpy, but it is possible to

calculate the melting point by a series of calculations[4].
In fact when this apparatus is used during the determination the enthalpy of fusion, it is possible to find the relation between enthalpy of fusion and melting point of mentioned matter. As it is illustrated in equation number 1.
It is provide in an ideal solution; mole fraction of solute (X) at saturation is a function of heat of fusion (ΔH) and in addition melting point of the solid (T_{fus}) and the temperature of solution (T)[4].

$$\ln X = - \Delta H_{fus}/R \left(1/T - 1/T_{fus} \right) \quad (1)$$

Although it is a new method, but it is not reachable for everyone. It is because of the expensivity of the device used in this method.

The earlier method introduced before DSC is thiele tube. Thiele tube is a glass tube designed to contain heating oil and a thermometer to which a small tube containing the sample is attached. In this model a small amount of the substance to be tested is placed in small test tube which is in turn tied to a thermometer and placed in heating bath. According to a series of lab observations as a result melting point would be determined.

There is another method which was smartly invented and was frequently used in the last century in order to determine the melting point of the materials which have melting point interval between 50°C to 260°C, developed by the Austrian pharmacognosist "Ludwig Kofler" and is also named as Kofler bench or Kofler hot-stage microscope. This apparatus is formed of a metalic band which is made of an alloy of steel. It has 36 cm length and 4 cm of width and is heated by an electrical source. This apparatus will show the melting point value by a mechanical pointer[5]. Although this apparatus could be helpful but every time before using, calibration is needed and consequently this shows the weakness of this old method.

It deserves to introduce a faster and newer method for determining the melting point of chemical compounds which acts directly and is specially used for organic compounds, these are named as electrothermal apparatuses which can delineate the melting point of compounds by following procedure:

For determining the melting point a beam with a constant wavelength (λ) releases from a source throw the sample. During the time that the sample is still in its solid state, the beam doesn't receive to the adjusted photocell which is toward it but as it is became as a liquid the beam recieves to the photo cell. By this way it become possible to determine the needed heat and also the given heat by a sensor which has been adjusted inside the apparatus.

Example for these apparatuses are found variously such as "IA9000 Series" or "1101D and 1102D Mel-Temp" that are able to determine the mentioned value with a tolerance of 0.1°C and 1°C ,respectively.

Such as these devices could be reachable for some laboratories but not mostly, because these devices are relatively expensive and it cannot be possible for all chemists who are working all around the world to have one of them. And in addition it is time wasting because some time this period of determining the melting point can spend too much time.

By the progess of the science, new methods were invented. Such as methods of quantitative structure-property relationship (QSPR)[2, 6]. QSPR could be used to estimate melting point. This method is specificly for organic compounds, such as alkanes, amines, ketones, and aldehydes[3]. In fact QSPR models are empirical equations which are in use for estimating thermodynamical and physical properties of molecules. Equation number 2 shows the QSPR formula for estimation of the physical properties.

$$P = a + b.\ D_1 + c\ .\ D_2 + d\ .\ D_3 + \ldots \quad (2)$$

In the equation number 2, P is a physical property of interest and a,b,c, … are regression coefficients and finally D_1 , D_2 , D_3 , … are parameters derived from the molecular structure[7]. Although this model could be useful for estimation the melting point of organic compounds, It has a very important and complicated part named descriptors. Descriptors are electrostatic and quantum chemical descriptions. The electrostatic descriptors are parameters which depend on the charge distribution within the molecule, including the dipole moment[7].

This need a chain of parctical and experimental activities to calculate the parameters. Beside of caculating the parameters, they should be inserted to the equation to reach the final relation. It will cost a large calculations processing loads.

Moreover, these estimational methods are not supposed to use bulk properties, but they mostly use microscopic physio-chemical properties of materials. In order to reach a more natural method that will use the nature and bulk properties of molecules a one year experimental study done by the authors of this paper. The results will be explained below.

3. Predicting Melting Point Usnig Neural Network

The first step was finding a valid refrence in order to extract the databases.[8] 289 samples were extracted. These samples were compounds consist of Carbon, Hydrogen, Nitrogen and Oxygen. Figure 1 shows the method of coding the compounds as the input for the neural network. As seen in this figure, each compund was coded in the column of table as combination of 4 numbers. The first row shows the number of Oxygen atoms of compund, the second one shows the number of Nitrogen atoms, the third one shows the number of Carbon atoms, the forth one shows the number of Hydrogen atoms and finally the last row shows the melting point of the selected compound. For instance, the second column stands for C_8H_8O with the melting point of 21 degree of Celcius.

O	1	3	4	4
N	0	0	0	1
C	8	8	8	7
H	8	8	8	11
MP	21	205.3	109	132

Figure 1. *The format of coding the components.*

4. Implementation of the Artificial Neural Network

In order to use the artificial neural networks(ANN) for the function aproximation, it will be efficient to design a two layers ANN. In fact as a practical experience, in the function approximation application of the ANN, the maximum layers number will be 3 and the efficient layers number will be 2. However in some cases as an exception, the ANN for function approximation would have more than 3 layers. Therefore as is shown in the figure number 2, the designed ANN for this research is consist of two layers. Considering the element types that each compound includes, the first layer of the ANN would have four inputs. One input for the number of Oxygen atoms of the compund, one for the Hydrogen atoms number, one for Nitrogen and the last one for Carbon. As the desirable result would be only the melting point of compound, the artificial neural network would have only one outpout as the melting point. The activation function for the first layer is tanh(x). This function shown a better performance comparing with the others for this research. The second layer activation function is pureline(x). This function will be used as the output layer activation function in the all function approximation applications of ANN[9].

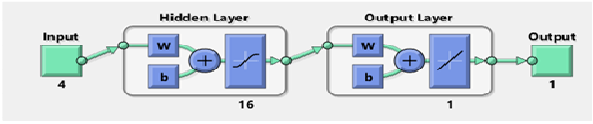

Figure 2. *The two layers ANN designed for melting points approximation.*

The traning algorithm used in this research is Levenberg-Marquart.[9] Considering Figure number 2, this by means of "fitnet" function was produced by MATLAB 2013.[9, 10] The samples for train, validation and test purposes were choosed by a random algorithm from the statistical society(consist of 289 sampels). As it is shown in the figure number 2, the neurons number of the first layer of ANN is 16. This number was obtained from the equation number 3. This equation defines the first layer's maximum neurons number of a two layer ANN that is used for a function approximation purpose.[9]

$$N1 \leq \frac{k(Ni+No)-No}{Ni+No+1} \quad (3)$$

In the equation number 3, k is the total number of the samples for training(in this experiment k = 289), *Ni is* the number of the ANN inputs (here Ni = 4), N_1 *is* the neuron first layer's neuron number and finally *No* the number of the outputs (here No = 1).[9]

By replacement of the parameters of the ANN is designed in figure number 2 in the equation number 3, it is determined that $N_1 \leq 239$. This shows the maximum possible neurons of the first layer of ANN is designed in figure number 2. If $N_1 > 239$, the ANN would be overtrained. Overtraining of the network means to memorize the answer rather than learning the relation between the whole sampels.[9] For instence, it is as the same as forcing a student to memorize the solution of a mathematical problem instead of learning how to solve it and this causes reduction of accuracy. Also the relation between samples will not be determined.

In Order to choose the best possible number for N_1, the ANN was tranied with all the N_1 integer amounts between 8 and 239. The test accuracy of the ANN was measured in each time of train. According to the figure number 3, it was found that the ANN is desigend for predicting the melting points will have the maximum accuracy(about 60%) of test data if its first layer neurons number will be 16.

5. Results

The tarining results of the ANN designed in the figure number 2 for predicting the melting points of organic compunds consist of Oxygen, Nitrogen, Hydrogen and Carbon is shown in the figure number 4 for traning, validation, test data and also the whole ANN itself. The train datas are used just for training. The Validation datas are for reforming the train of the ANN, they used for testing the ANN but the result of test will affect the whole ANN traning process. The test datas are only for examinating the accuracy of ANN and they do not have any effect on the training process.[9] Therefore the test data is the best way to find out how the ANN responses to the new and unknown sampels.

As the figure number 4 illustrates, the accuracy of the neural network for traning due to the sampels is 64.285%. In the validation procedure the accurcy of 60.544% was measured. The most important result was reported from the traning procedure and the shiny point of this research was the accuracy of 65.408% for testing new sampels. This accuracy shows that the designed ANN for melting point prediction will predict the melting point of new infamiliar materials with the accuracy of about 65%. Althought the accuracy of the whole ANN for this procedure is 63.733%.

6. Conclusions

In this research first the importance of melting point measurment and it's applications were discussed. Then the researches were done so far to measure the melting point of organic compunds were reviewed. It was cleared that this methods have some disadvantages such as heavy load of calculations. Also they just focused on the bulk physio-chemical properties of materials and not their microscopic properties. Then an experimental method of

predicting the melting point of organic compunds was introduced. This method uses a two layer artificial neural networks and models the nature of the compunds. In the contrary of the pervious methods that need heavy loads of calculations or special instruments, this method could perform on a personal computer using Matlab software. Also it could predict the melting point of new invented or unfamiliar organic compunds with the accuracy of about 65%.

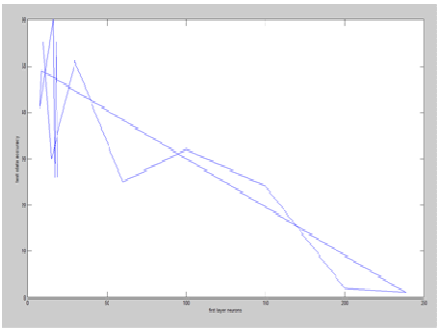

Figure 3.*Comparing the accuracy of traning with the number of neuruns, Accuracy (Y axis) vs number of neuruns(X axis) per training in ANN*

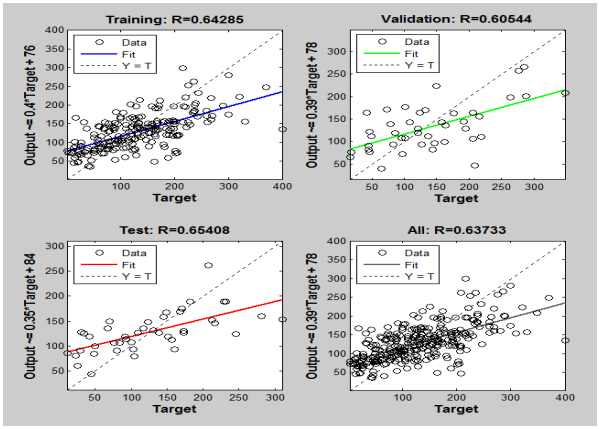

Figure 4. *The training results of ANN.*

References

[1] Gani, J.M.a.R., Group-Contribution Based Estimation of Pure Component Properties. Fluid Phase Equilibria, 2001: p. 183-208.

[2] Rahman, J.C.d.a.M.H., QSAR Approach To The Prediction Of Melting Points Of Substituted Anilines. 6th Int. Conf. on Mathematical Modlling, 1988. 11: p. 843-846.

[3] WANG Qiang, M.P.a.N.S., Position Group Contribution Method for Estimation of Melting Point of Organic Compounds. Chinese Journal of Chemical Engineering, 2009. 17: p. 468-472.

[4] Paula, P.A.a.J.d., Physical Chemistry. 2010: Freeman , NewYork.

[5] Aalae Alkhalil , J.B.N., Clive J. Roberts , Jonathan W. Aylott , and Jonathan C. Burley, Confocal Raman Microscope Mapping of a Kofler Melt. Crystal Growth and Design - ACS Publications, 2011: p. 422-430.

[6] Fangyou Yana, S.X., Qiang Wangb, Zhen Yanga, Peisheng Maa, Predicting the melting points of ionic liquids by the Quantitative Structure Property Relationship method using a topological index. The Journal of Chemical Thermodynamics, 2013. 62: p. 196-200.

[7] Jonsdottir, J.D.D.a.S.O., QSPR Models For Various Physical Properties of Carbohydrates Based on Molecular Mechanics and Quantum Chemical Calculations. Carbohydrate Research, 2004. 339: p. 269-280.

[8] Handbook of Chemistry and Physics. 84 ed. 2004: CRC Press.

[9] Menhaj, M.B., Fundamentals of Neural Networks. Vol. Computational Intelligence (vol.1). 2012.

[10] Yekta, A.F., Guidance of MATLAB7. 2005.