
PATTERNS CLASSIFICATION WITH APPLICATION

Sarah Ghanim Mahmood

Assistant lecturer

Collage of Administration Financial

Cihan University

ABSTRACT: *The present study tackles the topic of pattern classification, which is an essential issue for pattern recognition. It manifests a method, which is called ideal patterns that depends on Euclidean distance in order to make a classification for data related to well waters into two groups. The first is potable well waters, while the second group is that of not potable well waters.*

I. INTRODUCTION

The topic of pattern recognition is considered one of the modern subjects that polarize the attention of researchers in mathematics, statistics and computer sciences. The topic is also regarded as a hybrid field of knowledge in which mathematical, statistical and computing concepts are melted [Bishop, 2006 and Tou, 1977].

In brief, the pattern is a description of something, whereas recognition is a basic feature in humankind as well as other living creatures. In machine learning, pattern recognition can be identified as categorization of input data into diagnosable cases through distilling the moral features stated in the given data. Pattern recognition is a science employed in various fields of knowledge including medicine, psychology, psychiatry, flow of traffic, computer science in addition to security uses, [Theodoratos, 2010 and Trabelsi, 2011]

II. PATTERN CLASSIFICATION

Pattern classification is one of the basic issues of pattern recognition. Pattern classification is a process of categorizing data into certain groups or sections. If these groups are predefined, classification is considered supervised learning. Otherwise, they are regarded as unsupervised learning, [Koutroumbas, 2008]. Given database; i.e. $D = \{t_1, \dots, t_n\}$ of Tuples, $t_i = \{t_{i1}, \dots, t_{in}\}$ shall be number values, and a group of classes $C = \{C_1, \dots, C_m\}$ as $C_j = \{C_{j1}, \dots, C_{jm}\}$ is considered number values, the classification problem can be summed up by marking every t_i in class C_j so that:

$$\text{Sim}(t_i, C_j) \geq \text{Sim}(t_i, C_k), \quad \forall C_j, C_k \in C, \quad (1)$$

As $\text{Sim}(t_i, C_j)$ stands for similarity measure between t_i and C_j . it is largely known that there are many mathematical similarity measures; among which, Euclidean distance is most used. It is symbolized Dis [Tou, 1977].

III. CLASSIFICATION BASED UPON PATTERN IDEALS

Let's assume that there is a certain class (called m) $C_1, \dots, C_m$ and $Z_1, \dots, Z_m$ of pattern ideals. If $\underline{X}$ is an arbitrary vector, Euclidean Distance (as known) between vector $\underline{X}$ and pattern ideal Z_i will be defined as follow:

$$\begin{aligned} D_i &= \|\underline{X} - Z_i\| \\ &= \sqrt{(\underline{X} - Z_i)'(\underline{X} - Z_i)} \\ \therefore D_i^2 &= (\underline{X} - Z_i)'(\underline{X} - Z_i) \end{aligned}$$

i.e.

$D_i^2 = \underline{X}' \underline{X} - 2 \underline{X}' \underline{Z}_i + \underline{Z}_i' \underline{Z}_i, i = 1, \dots, m.$ (2) Therefore, if we want to classify the arbitrary vector $\underline{X}$ to one of the pattern ideals $Z_1, \dots, Z_m$, we can calculate D_i^2 for the values $i = 1, \dots, m$ then select classification $\underline{X}$ to exist in pattern Z_k that can implement:

$$D_k^2 = \min_{i=1, \dots, m} \{D_i^2\}. \quad (3)$$

As $\underline{X}' \underline{X}$ does not depend on i , we can reduce calculus in D_i^2 by defining the following decision function, [Theodoratos, 2008 and Theodoratos, 2010]:

$$d_i(\underline{X}) = \underline{X}' \underline{Z}_i - \frac{1}{2} \underline{Z}_i' \underline{Z}_i, i = 1, \dots, m. \quad (4)$$

In this case the decision made to select classification $\underline{X}$ to exist in pattern Z_k if:

$$d_k(\underline{X}) > d_i(\underline{X}), \forall i = 1, \dots, m.$$

IV. THE PRACTICAL SIDE

Since ancient times, groundwater has been used as the main source for human consumption, especially in places that surface water whether as rivers or lakes did not exist near or around them. This water was widely used even when surface water was available, but this river water was contaminated or its quality was worse than that of the groundwater. Groundwater is known as the water coming out of the ground whether in the forms of springs, fountains or sunk wells. Random samples were taken from old and newly sunk wells in different areas of Nineveh Province north of Iraq at different depths. Those samples were subject to analysis in the laboratories of the University of Mosul, according to the modalities [APHA, AWWA, WEF, 1998] standard tests for the World Health Organization in 1998. Several quality indicators were followed up for the taken samples.

The following table represents the data for research that represents twelve indicators (which are the basic features of the quality of ground water) taken from wells in twenty-one different area in the province of Nineveh, which were analyzed in the practical side of the research, [Mahmood, F.Y., 1994].

Since the percentages and amounts are varying with each other; and in order to have objective accounts, each variable was re-standardized by dividing the value to the largest value in it. It was conducted through computer processing MATLAB R2011a. Through cluster analysis (factorization) on the research () made on the same data, it was evident that we could choose two types ideal for water wells as the basic contents: the first ideal pattern represented by (Wayne Village), and second ideal pattern represented by (Baaj- Khoytla).[Muthana Sobhei, 2012]

Table (1) Distribution of Wells as per Ideal Patterns

Potable Water (Owaynat Village)	Non-drinkable Water (Baaj- Khoytla)
East Rabia 0.4644	Um el Kahaf 1.4401
Kharmal Village 0.5681	Al-Ayadhya 1.8354
Owaynat Village 0.6681	Baaj- Khoytla 2.4941
Rabia- Saada Village 0.5696	Mosul- Ugab Valley 0.8827
Zummar 0.5767	Bowayza 2.3474

Al- Khazar 0.6813	Bartalla 0.9063
Al-Shaquly 0.3786	Bartalla- Poultry Field 1.9243
Shallalat 0.4663	Nimrud 2.4397
Makhmour 0.2149	Hammam el Alil Intersection 1.1725
Kokchely 0.5621	
Nimrud 2.4397	

REFERENCES

- 1- APHA, AWWA, WPCF (1985), "Standard Methods for Water Examination and Tests" New-York.
- 2- Bishop, Christopher (2006). "Pattern Recognition and Machine Learning", Berlin: Springer.
- 3- Jain, Anil.K. Duin, Robert.P.W. Mao, Jianchang (2000). "Statistical pattern recognition: a review", IEEE Transactions on Pattern Analysis and Machine Intelligence 22 (1).
- 4- Koutroumbas, Konstantinos; Theodoridis, Sergios (2008). "Pattern Recognition", (4th ed.). Boston: Academic Press.
- 5- Tou ,Julius T. and R. C. Gonzalez (1977) "Pattern Recognition Principles ", Addison-Wesley.
- 6- Theodoridis, Sergios, Koutroumbas, Konstantinos,(2008), "Pattern Recognition", (4th ed.). Boston: Academic Press.
- 7- Theodoridis Sergios, Konstantinos Koutroumbas, "An Introduction to Pattern Recognition AMATLAB Approach", Elsevier Inc., 2010.
- 8- Tou ,Julius T. and R. C. Gonzalez (1977) "Pattern Recognition Principles ", Addison-Wesley.
- 9- Trabelsi, Salsabil; Elouedi, Zied; and Lingras, Pawan; (2011) , " Belief Rough Set Classification for Web Mining based on Dynamic Core ", 10th International Conference on Intelligent Systems Design and Applications; pp. 403-408.
- 10- Muthana Sobhei & Reiad Mahmood (2012) " Employment of Cluster Analysis and Nearest Neighbor Method in Pattern Recognition with the Application on Groundwater Quality in the Province of Nineveh". Iraqi Journal of Statistical Sciences.Vol 21No [141-165].
- 11- Mahmood, F.Y. (1994), "Physio-chemical evaluation of groundwater of some wells in Nineveh district used for drinking and domestic purpose" . Al-Rafidain engineering journal. Vol. 2, No.2.