



International Journal of ChemTech Research

CODEN (USA): IJCRGG ISSN: 0974-4290 Vol.7, No.5, pp 2203-2207, 2014-2015

Classifying the Suitability of River Cauvery Water for Drinking using Data Mining Classifiers and Authenticating Using Fuzzy logic

Durga Karthik¹*, K.Vijayarekha² and S. Raga Subha³

¹CSE, SRC,SASTRA University, Kumbakonam. India.
² School of EEE, SASTRA University, Thanjavur, India.
³ SRC,SASTRA University, Kumbakonam. India.

Abstract: River has been the cradle for all Civilizations and one among them is the Dravidian civilization nurtured along river Cauvery. Originating at Talakaveri in western ghats in Karnataka, flows for nearly 416 km in Tamilnadu before draining at Bay of Bengal.Cauvery delta in Tamilnadu is the fertile region and Nagapattinum district is gifted with various tributaries flowing through the district.Various parameters such as Ph, Ec, HC03,Cl,Mg,Na,Ca etc were tested from the river water for analysing the suitability of water for drinking purpose. The parameters were classified using Data Mining classifiers and authentication was carried using fuzzy logic.

Keywords: Cauvery water, Classification, Fuzzy logic, Drinking, Data mining.

Introduction

Nagapattinam is a southern district of Tamilnadu, with 10.7700° N, 79.8300° Eas geographical coordinates. The major water resource of this district is dependent on Cauvery water not only for irrigation but for house hold purpose. The water quality identification is required for the region as many water borne disease is frequent in this district. Due to industrialization, population growth and disposal of waste has led to pollution of the natural resources.

River water quality for drinking can be identified by analyzing the various parameters such as Ph, Ec, HCo3, Cl, Mg,Ca,Na,K, Residual Sodium Carbonate (RSC),Sodium Absorption Ration (SAR). Profiling water quality for drinking purpose has been carried out at various locations including ground water¹, river water ^{2,3} from sub station.

Data mining⁴ has emerged with various techniques that can be applied for classification or prediction from a large volume of data. Water potablity⁵ assessment requires analysis of multivariate⁶, so identifying the principal components using PCA⁷ has emerged as a better method in dimensionality reduction and also for better classification or prediction.

Experimental Methods

Water samples from river Cauvery were collected in a clean plastic cans during (November 27/2014 – Dec 2/2014) and were given for analysis at Aaduthurai Rice Research Institute, Aduthurai, India.The parameters are as follows.

Physhows the water acidity based on hydrogen ion concentration, EC indicates the dissolved ions in water, Chlorides indicates the presence of Cl,HCo3 bicarbonates, Mg, Ca, Na ,K indicates the Magnesium , calcium, Sodium and potassium present in the water sample. The other two parameters are residual Sodium

carbonate (RSC) and Sodium absorption ratio(SAR). River water samples from 35 locations of Nagapattinum District and 3 samples from adjacent locations from Thanjavur district were also collected for analysis.

Classification

Our aim is to classify the water into four groups such as highly suitable, fairly suitable, moderate,poor,bad. Data mining techniques provide various methods for classification and can give very accurate results on using them appropriately. As 11 parameters are present, Principal component Analysis (PCA) can be employed to identify the major components and is generally employed for dimensionality reduction. Classifying the data after dimension reduction results in accurate classification. Hence PCA using IBM –SPSS tool was employed on theriver data given in Table 1.

S.No	Location	рН	EC	HCO3	CI	CSO4	Са	Mg	Na	К	RSC	SAR
1	Kollidam	5.2	0.3	6.4	1.1	0.1	0.5	1	3	0.5	0	1.2
2	Mayilduthurai	6.5	0.4	5.2	8.2	0.1	2	6.5	4	0.5	6.2	9.2
3	Muvallur	6.8	0.2	6.4	9.2	0.1	1.5	2.5	7	0.5	0.9	2.5
4	Kuthalam	7.2	0.3	4.2	3.4	0.1	6	2	6	0.5	1.6	4.36
5	Villiayanallur	7.5	0.4	4.8	16	0.1	2	8	8	1	0.7	5.41
6	Manajjeari	7.4	0.2	6.3	7.2	0.1	3	10	10.8	0.5	0	3.65
7	Anjarvarthalai	7.6	0.6	5.5	2.5	0.3	2.4	6	25	0.5	0	4.46
8	Komal	6.8	0.5	5.87	44	0.2	1.5	8	8	1	0	5.24
9	Komal	7.5	0.3	7	18	0.1	1.6	8	3	3.7	0	3.21
10	Malliyam	7.9	0.4	7.2	15	0.2	3	1	8.2	4	0	0.87
11	Kilayur	6.2	0.6	5.5	20	0.2	15	4	4.5	0.5	0	3.79
12	Karuvazhakarai	6.8	0.4	5.2	11	0.2	3.5	46.5	6	0.4	0	2.29
13	Malaiyur	7.2	0.2	6.3	1.9	0.1	1.2	5	6.5	5	0	4.16
14	Vazhur	7.5	0.5	805	5	0.1	6	7	7	4	0	1.47
15	Mangainallure	7.1	0.2	7.9	37	0.1	1.5	8.8	1	0.5	0	1.91
16	Mappadugai	7.3	0.6	5.5	2.5	0.1	2.8	3.3	5	0.4	2.9	4.97
17	Elandhangudi	7.4	0.5	2.1	4	0.1	1.5	4	13	0.4	0	5.96
18	Senthangudi	7.6	0.3	8	5	0.6	0.5	5	6	1	0	3.16
19	Komal	6.8	0.5	7	5	0.1	1.8	10	3	4.5	3.2	4.6
20	Mariyur	7.5	0.3	4.9	4.8	0	3	1.2	6	4	0.4	1.99
21	Kovangudi	7.9	0.4	4.2	1.8	0	1	7	8	1.2	0	0.54
22	Asikadu	7.2	0.6	7.5	24	1.5	2	1.5	6	4	0	2.54
23	Therazhandhur	7.5	0.4	6.5	2.3	0.1	1.6	1.7	6	8	0	0.73
24	Mangkudi	6.2	0.2	8.7	3.4	0.1	1.5	2.2	5	5	0.4	0.41
25	Tharmapuram	6.8	0.5	10.2	1.5	0.1	3	4.1	4.5	1	0	22.36
26	Sembanorkoil	7.2	0.3	6.4	29	0.1	7.8	8.1	4.5	3	0	6.76
27	Srinivasapuran	7.5	0.4	5.9	1.7	0.1	2	3.7	3	6	0	1.98
28	Kuthalam	7.1	0.6	5.9	3.5	0.1	1.4	3.1	1.2	5	0	5.11
29	sitharkadu	7.3	0.4	3.5	28	1	2.8	4	2.8	3	2.8	4.02
30	Seathrapalapu	7.6	0.4	6	4.2	0.1	0.5	14	5	1.5	0	8.08
31	Thiruvalankadu	6.8	0.2	5.1	-	2	2	4.4	2.9	1	8.5	12
32	Tharmapuram	7.5	0.5	5.9	7.2	0	2.5	6	2	1	0	1.9
33	Manjalvaikal	7.9	0.2	5.5	7	2	1.5	8	8	1.5	2.1	1.8
34	Thirruvidaikazh	6.2	0.6	8.4	10	1	1.8	10	6	1	0	2.4
35	Thirukkadaiyur	6.8	0.6	5.4	2.2	1	1.1	1.2	3	1	0.4	2.4
36	Thanjavur	6.2	0.5	8	3.6	3.1	1	7	10	4	9.3	17.89
37	Swamimalai	6.8	0.3	9	14	0.1	1	1.7	1.5	0.5	-0.8	3.09
38	Aaduthurai	7.2	0.5	5.4	5.2	0.1	12	22	1	0.5	-3.7	0.72

Table 1: Cauvery River Water dataFor Nagapattinum and Thanjavur District.

The components are generally grouped based on covariance exhibited by the data. The second component shows the vital parameters such as Ph,EC,HCo3,Mg,Ca,Na.Ph indicates if the water is acidic or basic, EC refers to the content of ions in water, Bicarbonate levels shows contamination, Magnesium and calcium refers to the hardness of water. Sodium corresponds to the EC level, hence the presence of sodium is not included. The above parameters without sodium were used for clustering using Weka Tool. EM clustering was employed to cluster the data. Three clusters were generated by EM clustering shown in the figure 2.

On applying PCA for the data using SPSS tool, it showed 3 components shown in Figure 1.







Figure 2: River Cauvery water clusters using EM clusters

Three clusters were identified and the Clustered Instances are given in table 2.

Table 2: River Cauvery Water Cluster instances Using EM Clustering

S.no	Cluster	Number of Instance	Percentage of Instance
1	0	9	26%
2	1	11	31%
3	2	15	43%

Classification Using Fuzzy Logic

The EM (Expectation Maximization) method is an unsupervised iterative algorithm that clusters the instance based on probability distribution. To authenticate the above clusters , a fuzzy based rules using WHO standards were derived.

 $Ph (water) = \begin{cases} 7 \le x \ge 8.5 & 1\\ else & 0 \end{cases}$



based on the cumulative value the river water is classified as given below

	(5	well suited
	4	fairly suited
Cumulative water =	3	moderate
	2	poor
	1	bad

On applying the above logic to the river cauverywater, all the data grouped cumulatively into 5,4 and 3 respectively given in the Figure 2. The categories were well suited, fairly suited and moderate, hence three categories were formed given in table 3. The categories formed from fuzzy and data mining algorithm correlated.



Figure 2: River Cauvery water classification using Fuzzy rules.

Table 3: Classification of River Cauvery water using Fuzzy rules.

S.	Classification&	Number of	Percentage of
No	Class	locations	Instance
1	Well Suited (5)	16	42%
2	Fairly Suited (4)	14	37%
3	Moderate (3)	8	21%

Conclusion

River Cauvery water on analysis shows that it is suitable for drinking water in Nagapattinum district. Data mining techniques prove to be faster and accurate on applying them to appropriate data. Fuzzy logic technique along with data mining has found to correlate in classification and fuzzy logic can be applied for authenticating the data mining classification. The above method is simpler and faster, hence can be applied for identifying the water suitability in a given region.

References

- 1. Kamakshaiah.Kolli, R.Seshadri, "Ground Water Quality Assessment Using Data Mining Techniques", International Journal of Computer Appliocations, 2013, Vol 76, No:15, 39-42.
- 2. Vikram Bhardwaj, Dhruvsensingh and A K Singh, "Water Quality of the Chhoti River Using Principal Component Analysis, Ganga Plain, India", J Earth Syst. Sci, 2010, 119, No: 1, 117-127.
- 3. Kunwar P.Singh, Amrita Malik, SaritaSinha," Water Quality and apportionment of pollution sources of Gomti river (India) using Multivariate statistical Techniques a case study", AnalyticaChemica Acta, 2005, 538, 355-374.
- 4. Shoba.G,Shobha G,"WaterQuality Prediction Using Data Mining Techniques: A surver",International Journal of Engineering and Computer Science",2014,Vol:3,No:6,6299-6306.
- 5. DurgaKarthik and K.Vijayarekha,"Multivariate Data Mining Techniques for Assessing water Potability,Rasayan J.Chem,2014,Vol:7, No:3,256-259.
- 6. DurgaKarthik, KaruppaiyanVijayarekha, SubramaniyanSekar," Profiling Water Quality Using the Multivariate Chemometric Method. *Polish Journal of Environmental Studies*", 2014, Vol. 23(2).1175-1178.
- DurgaKarthik and K. VijayaRekha, SATTVA-Statistical Affirmative Testing Tool for Various Adulterants. Research Journal of Applied Sciences, Engineering and Technology", 2012, 4(24).5357-5360.
