ISSN: 2278 – 1323
International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)
Volume 2, No 5, May 2013

CLASSIFICATION BY K-MEANS CLUSTERING

MUKESH KUMAR CHOUDHAR[1], MANDEEP SINGH SAINI[2], PALVEE[3]

[1]M.Tech (E.C.E), I I T T Engineering College, Pojewal, P. T.U, (Punjab)

[2,3]M.Tech (E.C.E), Guru Nanak Dev Engineering College, Ludhiana, P. T. U. Regional Centre, (Punjab)

***ABSTRACT* — Clustering is an important task for machine learning which gives best discriminability among different subsets of features. It is usually a Classification problem with unsupervised learning paradigm. Recently unsupervised learning paradigms have gained tremendous attention, especially in the field of electrochemistry, bioinformatics. A novel impedance Tongue (i-Tongue) employing non specific multi-electrode electrochemical impedance spectroscopy is used for classification of Indian black tea. Impedance response at logarithmic frequency interval (features) ranging from 15 MHz (high frequency range) to 20 Hz (low frequency range) of three different type of electrodes were measured by using standard electrochemical workstation, which is used as our features dataset. Further the dimensions of these feature dataset containing impedances at particular frequency intervals are reduced by using Principal Component Analysis (PCA). Our proposed algorithm uses features similarity to distinguish between different tea samples by using a K-Means Clustering as a classifier to find the optimal data locations to have the best discriminability with minimum intra-cluster distance and maximum inter-cluster distance among different tea classes.**

***Index Terms*— Taste sensor, Multichannel taste sensor Amino acids, Food.**

[1]. INTRODUCTION

Tea is one of the widely consumed beverage in the world and India being the second largest producer, has its tremendous agro-commercial importance. Teas are majorly classified as black tea (fermented), green tea (unfermented), oolong tea (semi-fermented) apart from local variations in manufacturing processes. Orthodox and Cut–Tear–Curl (CTC) are the principal categories of black tea; their manufacturing techniques differ considerably and have a pronounced impact on the formative and degradative patterns of various cellular components [11]. Consumer acceptability of tea depends upon its flavour and taste, on the other hand flavour and taste depends upon spatiotemporal variability of the crop and manufacturing processes, which in turn highly influence its chemical composition, and are very critical in determining its quality [12]. Traditionally tea quality is assessed by tea tasters who have their own jargons to describe various quality attributes of a tea infusion. These jargons are sometimes not only difficult to comprehend by consumers but also highly subjective. It is therefore important to develop precise chemical or physical methods for the objective estimation of tea quality. Many attempts have been made by various researchers to correlate tea quality with its chemical composition. A number of efforts have been made to classify different beverages including tea using sensor array and electrochemical techniques such as Cyclic Voltammetry, Potentiometry and Conductivity. However, in comparison to potentiometry, especially with voltammetry, the impedance measurements are advantageous because of the potential experimental simplicity and the reduction of the response times [1].

One of the interesting areas of research in a multi-sensor system is signal pre processing.

Raw sensor signals, depending upon the mode of use of sensors, like steady-state phase, transient phase or both, carry information in different domains like time, frequency, physical parameters. This calls for computing techniques for feature selection and extraction. Evolutionary techniques like Genetic Algorithms, Particle Swarm Optimization are majorly used for feature/feature subset selection [12].

Instrumental evaluation of black tea is quite complex because of presence of many compounds and therefore it is being distinguished by tea tasters on their scores [8]. Impedance tongue are sensor array for qualitative and quantitative analysis and it is used to differentiate basic standard taste. The classification models can be created by using supervised and unsupervised techniques and artificial neural network [16].

This device can lead to lower calibration cost dataset and easy adaptability. This i-Tongue can reduces human sensory test panels, precise measurement of taste, formulation development time and cost and have many advantages and benefits [19].

Amol P. Bhondekar and Mopsy Dhiman [15] presented a novel iTongue for Indian black tea discrimination based on multi-electrode Electrochemical Impedance Spectroscopy.

Impedance response of platinum, gold, silver, glassy carbon, polyaniline and polypyrrole working electrodes in tea infusions for a sinusoidal excitation in the frequency range of 1Hz to 100 kHz has been measured. Also, the percentage of major chemical constituents responsible for the tea quality has been determined by HPLC and UV–vis spectrophotometer. The correlations between the frequency specific impedance response of working electrodes and the chemical concentrations which depend on sample variability, such as harvest interval, manufacturing process and brand type have been established. These correlations were further used for dimensionality reduction and discriminability enhancement.

A. IMPEDANCE TONGUE

The electronic tongue technology has been successfully employed for recognition and quality analysis of various food and agro products, viz., wine, cola, meat, fish, coffee, etc. Also, it has successfully established its capability for other applications like medical diagnostics and environment monitoring. The benefits of the same is listed below:

- Evaluate and quantify bitterness scores of new chemical entities (NCE).
- Optimizes and increases the formulation development process.
- Within the formulation, it measures the efficiency of complexion/coating.
- Various combinations of sweeteners, enhancers, exhausters, aromas and masking agents can be tested in less time.
- Benchmark analysis: compares the palatability of new formulations with competitor‘s products.
- Serving a quality control function for flavored products and excipients.
- Developing suitable matching bitter placebo for double blind clinical testing.
- During the scale-up process from small production batches to full-scale manufacturing, it defines consistency of organoleptic quality.

B. APPLICATION OF I-TONGUE

1. Foodstuffs Industry
   - Food quality control during processing and storage (water, wine coffee, milk, juices…)
   - Optimization of bioreactors.
   - Control of ageing process of cheese, whiskey.
   - Automatic control of taste.
2. Medicine
   - Non-invasive diagnostics (patient‘s breath, analysis of urine, sweat, skin, odor).
   - Clinical monitoring in vivo.
   - Identification of unpleasant taste of pharmaceuticals.
3. Safety
   - Searching for chemical/biological weapon.
   - Searching for drugs, explosives.
   - Friend-or-foe identification.
   - Environmental pollution monitoring
   - Monitoring of agricultural and industrial pollution of air and water.
   - Identification of toxic substances.
   - Leak detection.
4. Chemical Industry
   - Products purity.
   - In the future – detection of functional groups, chiral distinction.
5. Quality control of air in buildings, closed accommodation (i.e. space station, control of ventilation systems).
6. Legal protection of inventions – digital ―fingerprints‖ of taste and odors.

II. STATEMENT OF PROBLEM

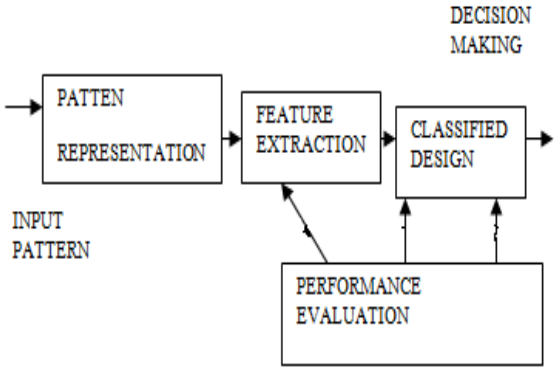


Fig. 1 Pattern Recognition Paradigm

K-means (MacQueen, 1967) is one of the simplest unsupervised learning algorithms that solve the well known clustering problem. The procedure follows a simple and easy way to classify a given data set through a certain number of clusters (assume k clusters) fixed a priori. The main idea is to define k centroids, one for each cluster. These centroids shoud be placed in a cunning way because of different location causes different result. So, the better choice is to place them as much as possible far away from each other. The next step is to take each point belonging to a given data set and associate it to the nearest centroid. When no point is pending, the first step is completed and an early groupage is done. At this point we need to re-calculate k new centroids as barycenters of the clusters resulting from the previous step. After we have these k new centroids, a new binding has to be done between the same data set points and the nearest new centroid. A loop has been generated. As a result of this loop we may notice that the k centroids change their location step by step until no more changes are done. In other words centroids do not move any more. Finally, this algorithm aims at minimizing an objective function, in this case a squared error function.

III. SOLUTION OF PROBLEM

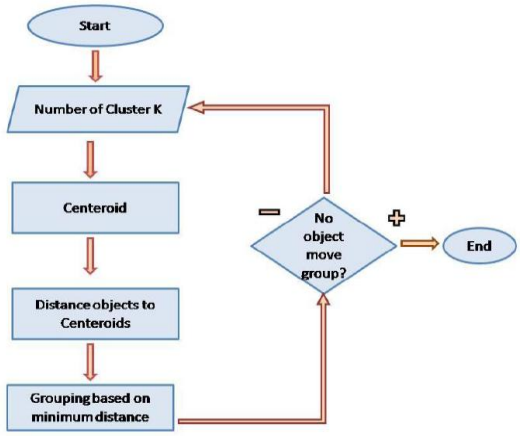


Fig.2. Algorithm of K-Means Clustering

IV. RESULTS AND DISCUSSION

So, the clustering algorithm explained above along with K-Means was implemented using the tea dataset obtained from iTongue experiment. As we have already mentioned that our aim is classify samples of tea classes along with the condition to maintain the maximum interclass distance, minimum intra class distance.

Classification Rates

The tea dataset arranged in 120×401 matrix contains 401 impedance values corresponding to 401 frequency points for 120 samples of 8 different classes of tea. This dataset obtained at 3 different voltages is fed to our K-Means Clustering algorithm. This algorithm was implemented in MATLAB 7.6.We set the K-Means system parameters to the following value

1. **Distance: cityblock**
   Emptyaction: singleton
   Onlinephase: off
   Start: Matrix
   In this Matrix is the Median of the each cluster.
   No of Features=320

TABLE 1

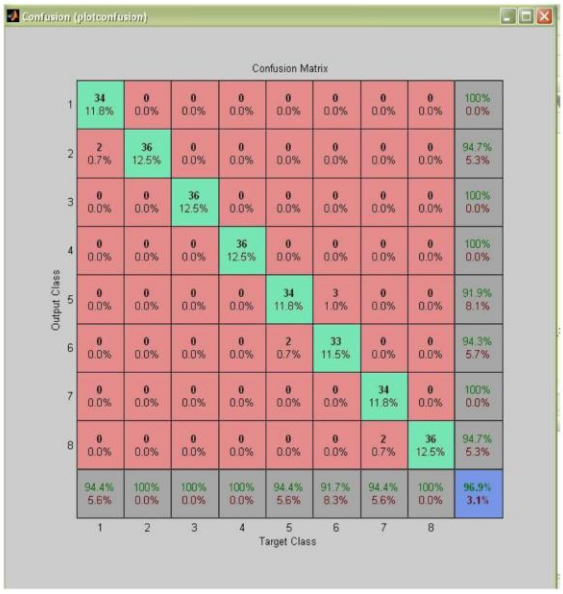
IMPLEMENTATION COST COM

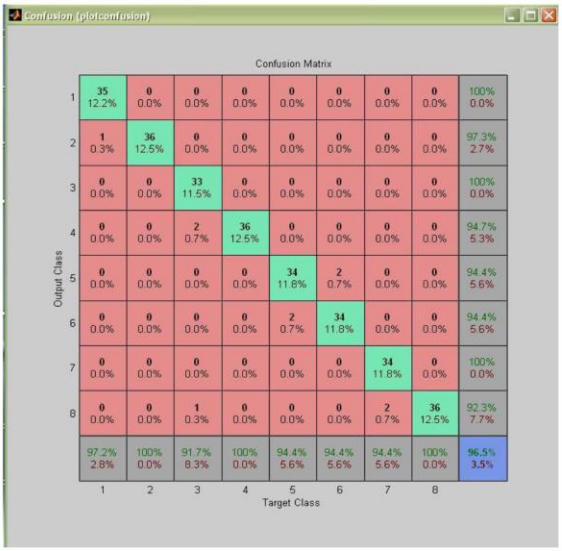
DATA	Classification (%) Rate		
Trail No.	1V (%)	10mv (%)	100 mV (%)
1.	93.8	94.8	96.2
2.	90.1	95.1	96.2
3.	90.3	95.5	96.9
4.	91.6	95.1	96.4
5.	93.3	96.5	95.8

TABLE 2

IMPLEMENTATION COST COM

DATA	Classification (%) Rate		
Trail No.	1V (%)	10 mV (%)	100 mV (%)
1.	87.0	91.0	92.0
2.	87.4	91.7	94.4
3.	88.0	94.4	94.1
4.	87.5	91.7	94.4
5.	88.0	93.4	92.0





2. **Distance: sqEuclidean**
   Emptyaction: singleton
   Onlinephase: off
   Start: Matrix
   In this Matrix is the Mean of the each cluster. No of Features=320

V. CONCLUSIONS

Clustering is an important task for machine learning which gives best discriminability among different subsets of features. It is usually a Classification problem with unsupervised learning paradigm. Recently unsupervised learning paradigms have gained tremendous attention, especially in the field of electrochemistry, bioinformatics. A novel impedance Tongue (iTongue) employing non specific multi-

electrode electrochemical impedance spectroscopy is used for classification of Indian black tea. Impedance response at logarithmic frequency interval (features) ranging from 15 MHz (high frequency range) to 20 Hz (low frequency range) of three different type of electrodes were measured by using standard electrochemical workstation, which is used as our features dataset. Further the dimensions of these feature dataset containing impedances at particular frequency intervals are reduced by using Principal Component Analysis (PCA). Our proposed algorithm uses features similarity to distinguish between different tea samples by using a K-Means Clustering as a classifier to find the optimal data locations to have the best discriminability with minimum intra-cluster distance and maximum inter-cluster distance among different tea classes.

REFERENCES

[1]. A. Riul, H.C. de Sousa et al., "Wine classification by taste sensors made from ultra-thin films and using neural networks", *Sensors and Actuators B: Chemical* vol. 98 pp. 77–82, 2004.

[2]. Anil Kumar Bag et al., "Rough Set Based Classification on Electronic Nose Data for Black Tea Application", *Proc. 2nd Int. Conf. on Advances in Comput. and Inform. Technology (ACITY)*, vol. 3, pp. 23-31, July 13-15, 2012.

[3]. B. Tudu, A. Jana et al., "Electronic nose for black tea quality evaluation by an incremental RBF network", *Sensors and Actuators B: Chemical* vol. 138, pp 90–95. 2009.

[4]. Bhattacharyya N., R. Bhuyan M. et al., "Electronic nose for black tea Classification and correlation of measurements with "Tea Taster" *IEEE Trans. Inst. Measurement*, vol. 57, pp. 1313-1321, Jul. 2008.

[5]. Bhondekar, A.P. and Dhiman et al., "A Novel iTongue For Indian Black Tea Discrimination*" Sensors and Actuators B: Chemical*, 148 (2). pp. 601-609. 2010.

[6]. G. Subramanya Nayak, Puttamadappa C, et al., "Classification of Bio Optical signals using K- Means Clustering for Detection of Skin Pathology", *Int. J. Comp. Application (IJCA)*, 1(2), pp.92-96, Feb., 2010.

[7]. K. Brudzewski, S. Osowski, et al., "Classification of milk by means of an electronic nose and SVM neural network", *Sensors and Actuators B Chemical*, vol 98,Issue 2-3, pp. 291–298, Mar. 2004.

[8]. Kiyoshi Toko, "Electronic sensing of tastes", *Electro analysis*, vol. 10, Issue10, pp. 657–669, Aug. 1998.

[9]. Legin Andrey, Rudnitskaya Alisa et al., "Electronic tongue for pharmaceutical analytics: quantification of tastes and masking effects" *J. Bio analytical Chemistry*, vol. 380, no.1, pp. 36-45, 29 Jul., 2004.

[10]. M. Pramod Kumar, Prof K V Krishna Kishore, "Simultaneous Pattern and Data Clustering Using Modified K-Means Algorithm" *Int. J. On Comp. Sci. and Eng.(IJCSE)* Vol. 02, No. 06,PP 2003-2008, 2010.

[11]. Mahanta P. K., "Biochemical analysis as a measure of dynamic equilibrium in genomic setup during processing of tea", *J. Bio-Sci*. vol.13, no 3 pp. 343–350, Sep.1988.

[12]. Maria Jamal, M R Khan et al. "Electronic Tongue and Their Analytical Application Using Artificial Neural Network Approach: A Review", *Masaum J. of Reviews and Surveys*, vol 1, Issue 1, Sep. 2009.

[13]. N. Bhattacharyya, R. Bandyopadhyay, et al., " Correlation of multi-sensor array data with Tasters panel evaluation for objective assessment of black tea flavor",*Int. Proc. ISOEN-2005, Barcelona, Spain*, pp. 13-15, Apr. 2005.

[14]. P. Devijver and J. Kittler, *Pattern Recognition: A Statistical Approach*, Prentice Hall, 1982.

[15]. P. Bhondekar, Mopsy Dhiman, et al., " A novel iTongue for Indian black tea discrimination", , *Sensors and Actuators B: Chemical , CSIO Chandigarh*, 2010.

[16]. R. Gutierrez-Osuna, H.T. Nagle et al., " Transient response analysis of an electronic nose using multi-exponential models", *Sensors and Actuators B: Chemical*, vol.61, pp. 170-182. 1999.

[17]. Seber, G. A. F., *Multivariate Observations*, Wiley, New York, 1984.

[18]. Tapas Kanungo et al., "An Efficient K-Means Clustering Algorithm: Analysis and implementation", *IEEE Trans. on pattern anal. and mach. Intell.*, vol 24, no.7, pp. 881-892, Jul. 2002.

[19]. Yu. Vlasov, A. Legin, A. et al., "Nonspecific sensor arrays "electronic tongue" for chemical analysis of liquids (IUPAC Technical Report)", *Pure and Applied Chemistry*, vol. 77, no 11, pp.1965-1983. 2005.

[20]. www.mathwork.com

AUTHOR PROFILE

MANDEEP SINGH SAINI Mr. Saini received his B.TECH From Nagpur University during 2001. They obtain M.Tech from GNDEC, Ludhiana (Punjab Technical University) during Apr. 2012. His field of interest is Digital Signal Processing and VLSI Architecture Design. He has published 22 Papers in International journal and 01 paper in Int. conference. They also obtain 10 years Technical Teaching experience in diff. Eng. Colleges.

.

MUKESH KUMAR CHOUDHARY:

Mr. Mukesh his B.TECH from M.S.E.C Bangalore Visvesvaraya Technological University-Belgaum during 2008.
Now he is a Research M.Tech. Student in IITT Eng. College, Pojewal under Punjab Tech. University.Jalandhar His field of interest is Digital Signal Processing. They also obtain 02 year Technical Teaching Experience form different Eng. Colleges.