JOURNAL OF PURE AND APPLIED MICROBIOLOGY, Dec. 2015. Vol. **9**(4), p. 3203-3206

Anomaly Detection System in Determination of Heavy Metal Pollution using K-medoid Algorithm

J.Jabez[1] and B.Muthukumar[2]

[1]Computer Science Engineering, Sathyabama University, Chennai, Tamil Nadu, India.
[2]Faculty of Computing, Sathyabama University , Chennai, Tamil Nadu, India.

(Received: 11 June 2015; accepted: 15 August 2015)

Anomaly detection becomes a predominant tool in Medical Stream, Fraud detection, Environmental related issues , Network Intrusion and other rare events that have great significance but are hard to find. In the recent years various algorithms were proposed to detect the growth of anomaly in different field of toxicology. The knowledge discovery of anomaly to find out the heavy metal toxicity in plants that has ability to accumulate toxic heavy metals from the contaminated soil site has made a new revolution. Thus this paper proposes a new method known as K-Medoids-HNNN using the technique HHNN and K-Medoids clustering. This system achieves the higher detection rate in finding out the pollutants from the environment. At first the proposed system implements the K-Medoids clustering technique on the various training subsets. Afterwards a mono HHNN model is trained using the different training subsets to detect the Anomaly. The experimental results shows the K-Medoids-HHNN approach achieve better results rather than other framework in estimating the amout of heavy metal pollution by employing phytoaccumulating plant that is more effective in cleaning up heavy metals like chromium, copper and zinc from the polluted industrial site.

Key words: Anomaly Detection, K-Medoids Clustering approach, Atomic spectrophotometer.

Soil is an important resource which is composed of both inorganic materials and living organisms. It provides the basis for life, giving nutrients to plants, which allow animal life to exist. Throughout human history, soil has been repeatedly polluted, ruined and destroyed by the developmental activities of the human beings. Dumping of industrial wastes is one of the prime factors contributing towards soil pollution. Industrial wastes contain large amounts of various chemicals which get accumulated on the top layer of the soil also causes imbalance in soil fauna and flora. Such loss ultimately results in changes in the ecological balance of the environment.

Heavy-metal pollution of soils and waters caused by the mining and burning of fossil fuels is a major environmental problem and exposure to these metals can be toxic to living cells (Qian *et. al.*, 1999). The increasing use of wide variety of heavy metals in industries and agriculture has caused a serious concern of environmental pollution (Sinha *et. al.*, 2010).

The removal of pollutants has also been of great concern. In general, heavy metal toxicity can cause chronic degenerative diseases like mental disorders, pain in muscle and joints, gastro intestinal disorders, vision problems, chronic fatigue ,Genotoxicity and cancers can also occur.

The environment has been found to absorb pollutants or clean up itself by natural biological/biochemical activities hence the increasing use of plants, plankton and other biota to remediate the environment was studied by Agunbiade and Fawale, 2009; Hammaini *et. al.*, 2006.

Certain plants have evolved the capacity to take up and accumulate selected metals in their shoots and roots at levels that are toxic to ordinary

* To whom all correspondence should be addressed.
E-mail: jabez.it@sathyabamauniversity.ac.in

plants; (Ebbs and Kochiain , 1997). Plants have evolved interactions and association with organisms that can cause breakdown or transformation of certain pollutants in the plant root zones to products that no longer pose environmental hazards (Becker., 2000).

Phytoremediation is one method that has provided a lot of interest among the researchers in recent years .Phytoremediation is developing technology that can potentially address the problems of contaminated soil affected by urban industrial activities. Phytoremediation is a green technology for the sustainable remediation of surface soil contaminated with heavy metals. It is an innovative, novel & potentially inexpensive technology using metal polluted soils, sludge's & sediments (Salt *et.al.*,1998). This process is cost effective without creating disturbance to the landscape (Itanna and Coulman , 2003). In general, favorable plant properties for phytoremediation are to be fast growing, have high biomass, and are tolerant to pollution. High levels of plant uptake, translocation, and accumulation in harvestable tissues of the plant are important properties for the phytoextraction of inorganics (Pilon Smits, 2006).These advances are promising for improving the effective use of phytoremediation technology for cleaning up the soil in contaminated sites.

Over the last decade, different approaches have been proposed and developed to detect the intrusion analysis in data mining concept [Kumar S. August, 2005][Chen, *et al*., 2003,]. In early day's Rule based expert systems and statistical methods were two methods to detect the intrusion analysis.

Li [2004] introduced a method using the genetic algorithm to find out the anomalies in the network intrusion [Patcha, and Park, 2007] . The method also has incorporated with the categorical and quantitative features of network in order to take the classification rules. On the other hand with the addition of quantitative features, the detection rate is increased in IDS but there are no results in the experimental setup. In [14] Goyal and Kumar explain Genetic Algorithm support to categorize all kind of smurf attack with the training dataset. The experimental detection rate is 100% and false positive rate is very low at 0.2% [Chetan Kumar and Anup Goyal, 2008,].

The research work [Depren *et al*., 2005] uses the genetic algorithm to identify the anomalous using the information theory [Chetan Kumar and Anup Goyal, 2008,][Manikopoulos and Papavassiliou, 2002,].

Proposed Work

In this work, our main aim is to model and develop more flexible, to identify small changes in patterns, and more accurate method using K-Medoid algorithm to evaluate the heavy metal toxicity in Mentha spicata L. using the soil physical parameters as model clusters.

The physical parameters on the abiotic factors of polluted area which are taken into consideration like PH, porosity, soil texture , soil temperature etc., When all these raw data are substituted in the algorithm. When the parameter exists the permissible limits then K-means algorithm act as a tool in finding out the heavy metal pollution level. The data set is trained by using set of physical parameters which are found commonly in all types of soil. For example

Data	Normal	Abnormal
PH	5-7	less than 5
Temp	25-32 C	> 35

We utilize the algorithm to discover the highly polluted site from the normal site. The important parameters used to evaluate the anomaly characterization are soil PH, soil temperature mineral contents and heavy metal contents.K-mediod is tested for different clusters of polluted site.Once the Polluted site is specific and accurate, the experimental plant is introduced to remove the toxic metals from the soil.

Toxic heavy metals like chromium, copper and zinc can be removed by phytoaccumulation method.ie) plant absorbs water, mineral nutrients along with heavy metals. Thus heavy metal gets accumulated in the biomass. The plant biomass is dried, powdered and made into ash .The percentage of heavy metal present is found using the instrument known as atomic absorption spectrometer .

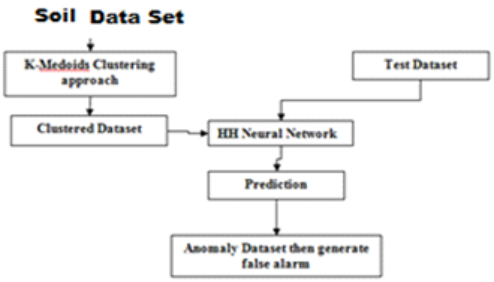
The k-means algorithm is more effective to outliers, because if an object exists with a large value, the distribution of data may be destroyed [Chetan Kumar and Anup Goyal, 2008].

The procedure of k-Medoids is defined below

Algorithm

1. Input dataset
2. Enter number of cluster k
3. Choose k objects randomly

4. Assign every object to nearest mediod
5. Find the sum of distance between cluster mediod and object.
6. Exchange the mediod object with non-mediod object
7. Again calculate the location of k-mediods
8. Perform k-Medoids for 4-7 times repeatedly until the medoid get constant.



K-Medoids algorithm

This algorithm is one of the most efficient clustering algorithm from other clustering approaches like k-means. A medoid is one of the data point and act as the other entire data points. In this situation the k-medoids are strong to outliers as well as noise because the partitioning method is done as per the concept of minimizing the total of dissimilarities between every object in cluster. Here the heavy metal polluted area is consider as the cluster and mint plant (Mentha, Spiketa, L.) is used as phyto accumulator in absorbing the heavy metal from the polluted soil. The instrument used is automic spectro photo meter for the analysis of heavy metals . The K-mediod algorithm is used to find out mathematically the percentage of pollution present in the study area via., plant. The K-mediod algorithm is used to find out the heavy metal percentage that have accumulated in the experimental plant.

Accumulation of heavy metals

After sorting the polluted area, The common Mint plant (Mentha Spicata L.) were collected from the study area was used for K-mediod algorithm to find out the amount of toxic heavy metal accumulation. On the other hand, treated mint plants were removed and the dried samples were powdered. Later, the ash was digested and the amount of metal accumulation was analyzed using Atomic Absorption Spectrophotometer. Using K-Mediod algorithm the amount of heavy metal accumulated in the polluted site plant can be determined . The algorithm in our present study showed about 99% accuracy.

CONCLUSION

The study has shown that though plants Mentha spicata L.and Chrysanthemum cinerariaefolium L. can effectively be employed in the phytoremediation of soil which is polluted by harmful, toxic heavy metals like Chromium and Copper. Properties such as high productivity, high absorption capacity and high metal removal potential establish plants Mentha spicata L. and Chrysanthemum cinerariaefolium L. immense potential for use in phytoremediation technologies. The resulting metal treated biomass can be safely smelted later.

A fundamental understanding of both uptake and translocation processes in normal plants and metal hyper accumulators, regulatory control of these activities, and the use of tissue-speciûc promoters offer great promise that the use of molecular biology tools can give Scientists the ability to develop effective and economic hytoremediation plants for soil metals.

We can conclude that plant menthe, spicta , L.experiment in heavy metal polluted has proven to be favored by K-mediod Algorithm.

REFERENCES

1. Agunbiade F.O and Fawale A.T. Use of Siam weed biomarker in assessing heavy metal contaminations in traffic and solid waste polluted areas. *Int. J. Environ. Sci. Tech.* 2009; **6**(2): 267-276.
2. Alves-Foss, J. C. and Taylor, " 'Low cost' network intrusion detection. In: Proceedings of the new security paradigms workshop" 2001.
3. Becker P.R. Concentration of chlorinated hydrocarbons and heavy metals in Alaska Arctic marine mammals. *Mar.Pollut.Bull*. 2000; **40**(10): 819–829.
4. Chen, Chang, Shyu, M. L., S. C., Sarinnapakorn, K., & L. 2003, "A novel anomaly detection scheme based on principal component classifier". In Proceedings of the IEEE Foundations and new directions of data mining workshop.
5. Chetan Kumar and Anup Goyal, "GA-NIDS: A Genetic Algorithm based Network Intrusion Detection System" 2008.

6. Depren, O., Topallar, M., Anarim, E., & Ciliz, M. K., "An intelligent intrusion detection system (IDS) for anomaly and misuse detection in computer networks. Expert Systems with Applications" 2005.
7. Ebbs S.D and Kochian L.V. Toxicity of Zn and Cu to Brassica species : Implacations for phytoremediation. *J. Environ. Qual*, 1997; **26**: 776-781.
8. Hammaini A., Gonzalez F., Ballester A, Bla zquez M.L, Munoz JA. Biosorption of heavy metals by activated sludge and their desorption characteristics. *J. Environ. Manag*. 2006; **84**(4), 419-426
9. Itanna F., Coulman B. 'Phyto-extraction of copper, iron, manganese, and Zinc from environmentally contaminated sites in Ethiopia with three grass species'. *Communication in soil science and plant analysis*, 2003; **34**(1,2): 111-124.
10. Kumar S. August 2005, "Classification and detection of computer Intrusions". PhD Thesis, Department of Computer Science, Purdue university,.
11. Li, W. 2004, "Using Genetic Algorithm for Network Intrusion Detection". "A Genetic Algorithm Approach to Network Intrusion Detection". SANS Institute.
12. Lu, W. and Traore, I. 2004 ,"Detecting New Forms of Network Intrusion Using Genetic Programming", Computational Intelligence.
13. Manikopoulos, C., & Papavassiliou, S. "Network intrusion and fault detection:A statistical anomaly approach". IEEE Communications Magazine 2002.
14. Patcha, A., & Park, J. M, "An overview of anomaly detection techniques 2007:
15. Pilon Smits Elizabeth A Hand John L Freeman. Environmental cleanup using plants:biotechnological advances and ecological considerations 203,*Front Ecol Environ*; 2006; **4**(4): 203–21 .
16. Qian J.H., Zayed A., Zhu Y-L., Yu M., and Terry N. Phytoaccumulation of trace elements by wetland plants: III. Uptake and accumulation of ten trace elements by twelve plant species. *Journal of Environmental Quality*, 1999; **28**: 1448-1455.
17. Salt D.E., Smith R.D and Raskin I. Phytoremediation. *Annual Review of Plant Physiology and Plant Molecular Biology.* 1998; **49**: 643-668.
18. Sinha-Khetriwal D., Kraeuchi P., Widmer R. Producer responsibility for e-waste management: Key issues for consideration.*The Swiss experience.J Environ Manage*. 2010; **90** (1):153–65