International Journal of Engineering Science Invention (IJESI)
ISSN (Online): 2319-6734, ISSN (Print): 2319-6726
www.ijesi.org ||Volume 9 Issue 1 Series. II || Jan.2020 || PP 57-62

Machine Learning: Application To Pervasive Computing Systems

[1] Dr. Nlerum, Promise A. [2] Dr. Edward E. Ogheneovo

[1]Department of Computer Science and Informatics, Federal University Otuoke, Bayelsa State, Nigeria.
[2]Department of Computer Science, University of Port Harcourt, Port Harcourt, Rivers State, Nigeria.
Corresponding Author: Dr. Nlerum, Promise A

Abstract
This paper takes a critical look at basic machine learning and pattern recognition methods and also gives the underlying concepts of machine learning techniques that can be applied to problems in the domain of pervasive computing environments. The scope of this study covers the two primary types of machine learning; supervised and unsupervised learning methods. In the process of providing the fundamental knowledge of machine learning, we present some conceptual terms of machine learning and the steps required in developing machine learning systems with a great impact on context-aware pervasive computing environments.

Keywords: *Context-aware, Pattern recognition, Supervised/Unsupervised learning., Pervasive Computing*

Date of Submission: 20-01-2020 Date of Acceptance: 05-02-2020

I. Introduction

Pervasive computing environment is a physical space replete with mobile handheld, personal computers and all forms of devices that are free to move to and from this space. Pervasive computing environments are highly proactive in nature and observe environmental context at any given time. Weiser, M. (1999). At some point, a pattern exits, which could be tracked, studied and replicated when necessary. Machine learning principles can be used to study these patterns and use the results to predict future resource availability and/or usage in a pervasive computing environment.

The importance of pattern recognition in this regard cannot be overemphasized. Figure 1 shows a typical pattern recognition procedure. In all pattern recognition tasks there are a number of measurements made of an event or object. These raw materials are transformed in some way into a set of features, and the features are used by a decision procedure to assign the event to one category or another, Alpaydin, E. (2004). Thus the pattern recognition system is a classifier. This helps in context management and prediction of pervasive elements.

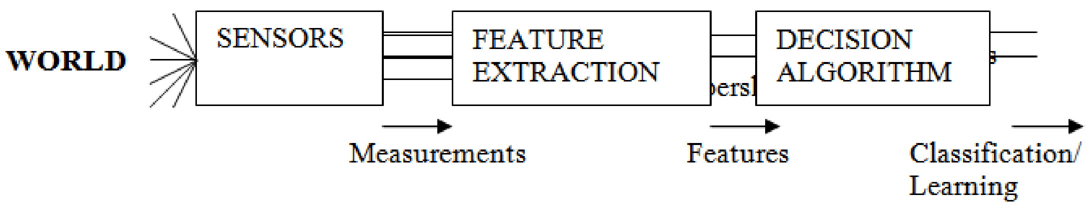

Figure 1: Pattern Recognizer

In general terms, machine learning can be clearly defined as set of methods that can automatically detect patterns (general regularities) in empirical data, such as sensor data or databases and then use the discovered patterns to predict future data, or execute other types of decision making under uncertainty, Bishop, C. M. (2006). Various disciplines utilize machine learning techniques, including the obvious disciplines like computer science and statistics, as well as many other fields from politics to geo-sciences Mitchell, T. M. (2006). Focus in machine learning research today majors on automatically learning patterns in large data set of complex data types.

So far, most of the machine learning approaches are merely reactive, adopting decisions based only on an earlier context. Research in anticipatory and proactive systems, including prediction of future situations is still in its beginnings, Karbassi, A. and M. Barth (2003). In these systems, a user is linked to a profile with a set of defining characteristics. Leveraging that profile allows us to improve the systems' efficiency by anticipating the user's needs and adapting the services accordingly. Several approaches have been proposed to achieve this anticipatory aspect. These approaches can be classified under two categories: passive or active:

Classification is the process of assigning to a particular input, the name of the class to which it belongs. *Learning is the process by which one entity acquires knowledge; usually that knowledge is already possessed by some number of other entities who may serve as teachers to the learner.* Alpaydin, E. (2004).

II. Learning System

When a computer system improves its performance at a given task over time, without re-programming, it can be said to have learned something. One can suffice it to say that 'automatic performance improvement with experience is a rough-and-ready definition of learning'. Jaime G. Carbonell, R. S. (1983). Tom Mitchell, in 1997, loosely opined that, a computer program is said to learn from experience E with respect to some task T and some performance measure P, if its performance on T, as measured by P, improves with experience E

The term machine in this paper means almost the same as computers. In practice, Machine-learning algorithms attempt to achieve one or more of the following goals:

- provide more accurate solutions
- cover a wider range of problems
- obtain answers more economically
- simplify codified knowledge

The last goal presumes that simplification of stored knowledge is valuable for its own sake. For example, a system might rearrange its knowledge base so that it was more intelligible to human readers. Even if its performance at the task was no better, this could well be useful. It is the first two criteria (accuracy of solutions and range of applicability), which usually have the highest priority, however.

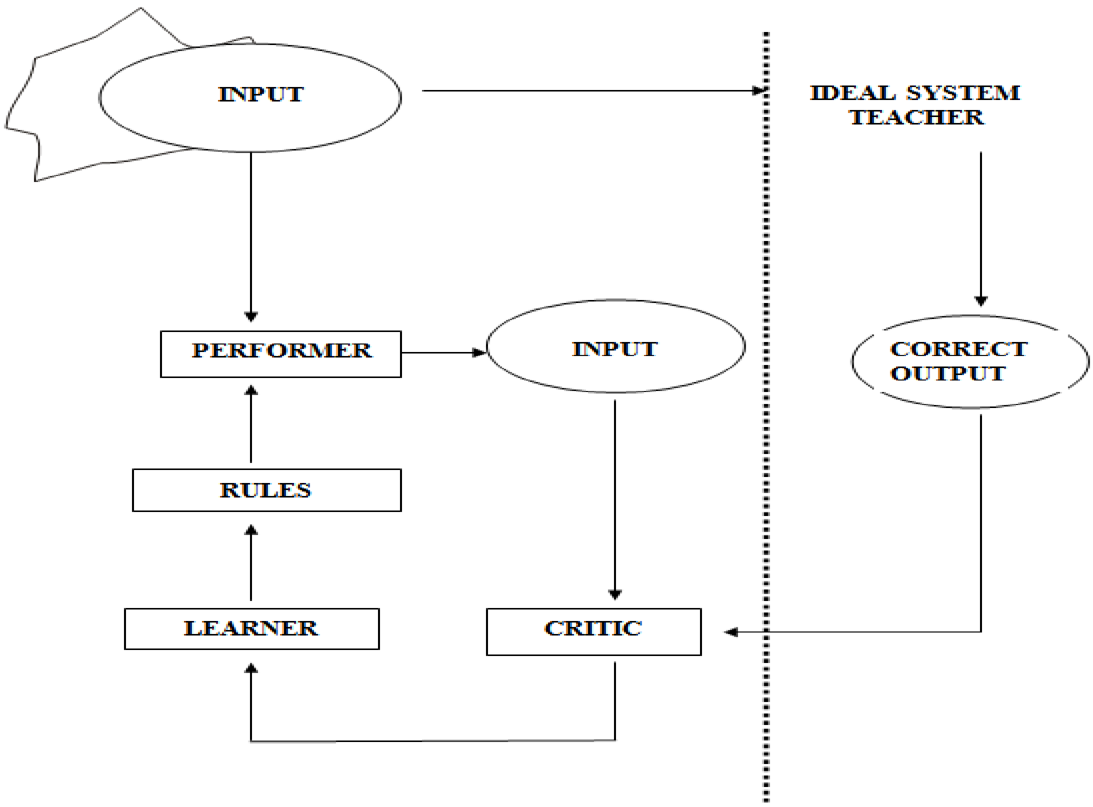

Fig.2: A framework for learning

All systems designed to modify and improve their performance share important common features. Fig. 2 is a diagram of the four major components of a typical learning system. Incidentally, this sketches a pattern recognizer, which learns to associate input descriptions with output categories. The system sketched has a feedback loop. The major components will be discussed by going round this feedback loop. To the left of the dotted line is the learning system and to the right is the ideal system teacher, whose behaviour it is trying to match.

2.1.**The *Critic***compares the actual with the desired output. In order to do so, there must be an 'ideal system' against which the system's behaviour is measured. The job of the Critic is known as Credit assignment or alternatively 'blame assignment'. It must access deviations from correct performance.

2.2 The learneris the heart of the system. This is the portion that has responsibility for amending the knowledge base to correct erroneous performance.

2.3 The 'Rules'are the data structures that encode the system's current level of expertise. They guide the activity of the performance module.

2.4 The Performeris the part of the system that carries out the task. This uses the rules in some way to guide its activity. Thus, when the rules are updated, the behaviour of the system as a whole change for the better, if all goes according to plan.

Two other terms need to be defined before we begin our discussion of practical learning methods - *description language* and *training set.*

The description language is the notation or formalism in which the knowledge of the system is expressed. There are two kinds of description language, which are important:

The first is the notation used to represent the input examples. One of the simplest of input format is the ***feature vector:***Each aspect of the input example is measured numerically and the vector of measurements defines the input situation.

The second kind of description language is that chosen to represent the rules themselves. Here representation schemes capable of expressing structuraldescriptions where the relations between parts of an object are important as well as the elementary attributes (or features) of the object.

It is essential therefore to ensure that the description is capable of expressing the kinds of distinction that will be needed.

The notion of a training set is important in understanding how a machine learning system is tested. Basically, there is a database of examples for which the solutions are known. The system works through these instances and derives a rule or set of rules for associating input descriptions with output decisions. In Ian H. Witten, E. F. (2005), according to ***Bacon,*** rules must be tested on cases other than those from which they were observed. Therefore, there should be another database, the 'test set' of the same kind but containing unseen data. If the rules also apply successfully to these fresh cases, our confidence in them is increased.

Methods of selection and or placement of hyper-surfaces that supplement the training set technique can then be called a *training method* and a classifier whose hyper-surfaces are adjustable can be called a *trainable classifying system,* which comes under the popular heading of learning machines.

Learning is thus a task of constructing the regions or templates in the N-dimensional space in which labeled samples of the classes are contained. Our study will be based on this loose definition.

III. Context Awareness In Machine Learning

According to Schilit and Theimer. (1994), context means all information coming from the environment that is used by the application to adapt its behavior. Context comprises resources and services and defines their role in the application. Context awareness is one of the fundamental principles underpinning pervasive computing. Context prediction, a new trend in pervasive computing, is an open-ended research topic with a lot of challenges and opportunities for innovation. Perttunen, M. (1999).

The context acquisition step concerns itself with gathering contextual information from physical and logical sensors. The captured data is then transformed into more meaningful features during the extraction step for better interpretation. The classification step then, recognizes recurring patterns, called clusters in the feature space, built in the previous step. Classification uses a feature vector which can be assigned to multiple clusters with different degrees of belonging. This degree represents the probability that a feature vector belongs to a cluster. Labeling allocates the descriptive names for individual clusters or combinations of several clusters. Based on the observed history, the future context cluster can be predicted. The prediction step uses the vector of clusters generated by the classification step. The final goal is to generate a cluster of vectors for the future, which matches the current cluster provided by the classification step. This classification is used for supplying the cluster vectors predicted in the labeling step to provide context label, intended for use in dynamic applications, Nicole, A. (2013).

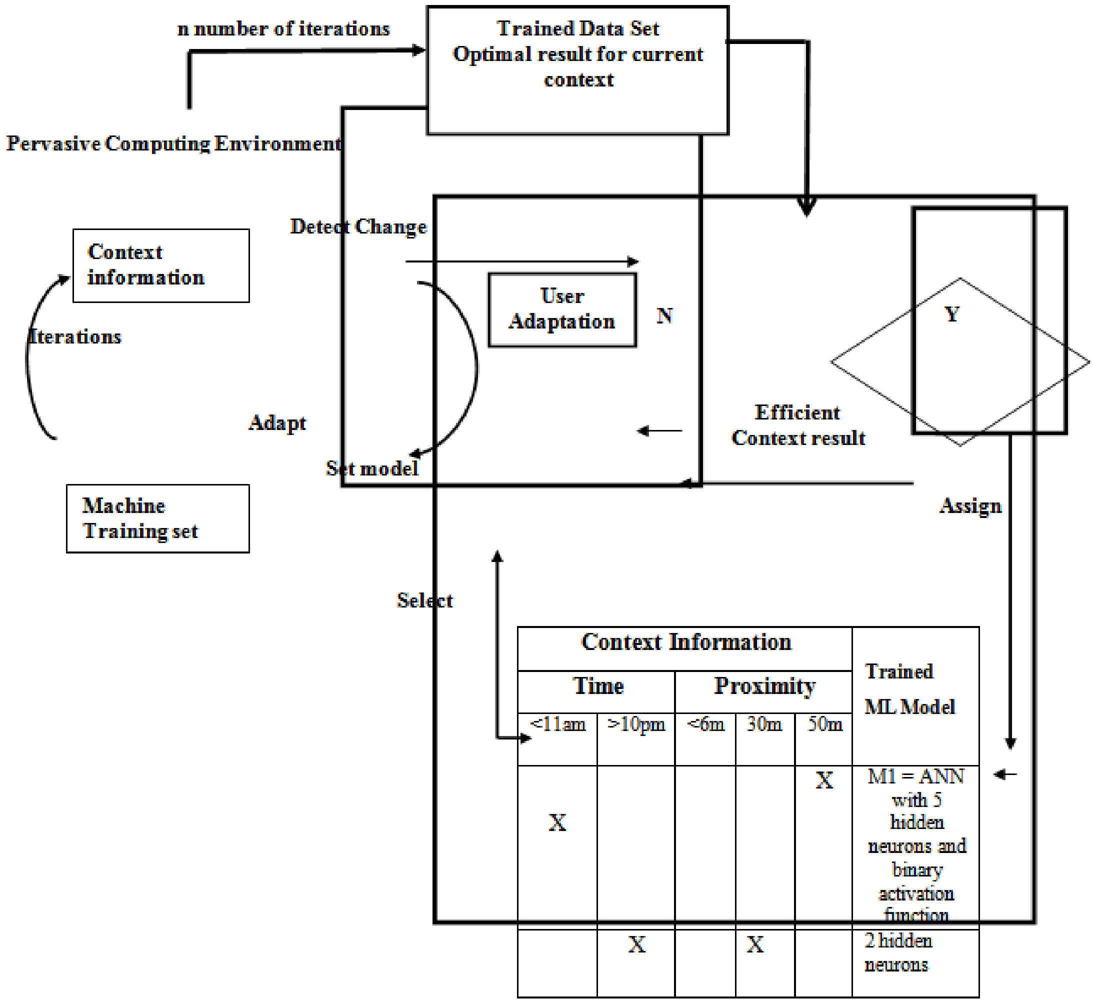

Figure 3: Schematic illustration of a pervasive architecture to store and retrieve machine learning models based on context.

To make predictions, we implemented a neural network with five hidden neurons that were trained by considering the full range of all of these variables. For example, the model was generic enough to predict effect of distance on resource availability, independent of the time and of the temperature and humidity of the environment. However, we observed that the mean absolute error of the predictions provided by the generic model was higher at distances above 50m.

IV. Parametric/Non Parametric Approach To Machine Learning

According to Friedberg, R. M. (1958), in the statistical classification approach, if the unknown is the parameter values of a known distribution function $P(X / C_k)$, then the parametric technique can be applied.

Suppose that $\varphi_1, \varphi_2, \ldots, \varphi_n$ are the m training subsets of patterns corresponding to the m classes, then with the knowledge that $P(X / C_j)$ are normal, the estimates of the parameters X_j(or μ_j) and S_j(or Σ_j) are defined by the following sample statistics:

$$\langle X\rangle_j = E(X) = 1/m_j \Sigma_{X \in \varphi_j} X \quad (1)$$

$$\langle S\rangle_j = 1/m_j \Sigma_{X \in \varphi_j} (X - \langle X\rangle_j(X - \langle X\rangle_j)` \quad (2)$$

Here E(X) denotes expected value X, M_j is the number of patterns in the training subset $M \in \varphi_j$, and $\langle X\rangle_j$ and $\langle S\rangle_j$ denote the sample mean (or, centre of gravity) and sample covariance matrix respectively of the jth class. If on the other hand, both function and parameter values are unknown, the nonparametric techniques in general should be used.

Learning process can be termed supervised or non-supervised depending on whether the correct classifications of the patterns observed are known or not.

4.1 BAYESSIAN SUPERVISED LEARNING METHOD

Supervised learning is fairly common in classification problems because the goal is often to get the computer to learn a classification system that we have created. It often leaves the probability for inputs undefined. This model is not needed as long as the inputs are available, but if some of the input values are missing, it is not possible to infer anything about the outputs.

Let $(x_1, x_2, \ldots, x_n)$ be a sequence of independent identically distributed feature vector observed from the same pattern class C_j then according to Bayes' theorem, the function $P_o(x)$ changes to the a posteriori density function:

$$P(\varphi | X_1) = \frac{P(X_1 | \varphi)\, P_o(\varphi)}{P(X_1)} \qquad (3)$$

After X_1 and X_2 are observed, the a posteriori density function of φ is:

$$P(\varphi | X_1, X_2) = \frac{P(X_2 | X_1, \varphi)\, P_o(\varphi | X_1)}{P(X_2 | X_1)} \qquad (4)$$

After the nth observation is observed:

$$P(\varphi | X_1, \ldots X_n) = \frac{P(X_n | X_1, \ldots, X_{n-1}, \varphi)\, P(\varphi | X_1, \ldots, X_{n-1}))\, P_o(\varphi)\,.}{P(X_n | X_1, \ldots, X_{n-1})} \qquad (5)$$

The required probability density function with the knowledge of

$P(\varphi | X_1, \ldots, X_n)$ is:

$$P(X_{n+1} | X_1, \ldots, X_n.C_j) = \int P(X_{n+1} | X_1, \ldots, X_n, C_j, \varphi)\, P(\varphi | X_1, X_n, C_j)\, d\varphi \qquad (6)$$

$n = 1, 2, \ldots$

where $P(X_{n+1} | X_1 \ldots X_n.C_j.\ \varphi)$ is known.

The a posteriori density function on the average becomes more concentrated and the estimate converges to the true value of the parameter, so long as the true value is not excluded by the a priori density function of the parameter φ.

4.2 NON-SUPERVISED LEARNING

In non-supervised learning, the learning observations (as their correct classifications are not known) are considered as coming from the mixture distribution with the probability distribution of each class as component distributions. The problem of learning is then reduced to a process of successive estimation of some unknown parameters in either a mixture distribution of all possible pattern classes or of a known decision boundary, Ghahramani, Z. (2008).

The mixture distribution is defined as:

$$P(X) = \sum_i P(X | Z_j^n)\, P(Z_i^n),\ i = 1, 2, \ldots .w \qquad (7)$$

Where $P(X | Z_i^n)$ denotes the ith – partition conditional distribution, $P(Z_i^n)$the mixing parameter for the ith – partition Z_i and $W(m^{,}$ m = number of class distribution) is the number of ways $Z_1^n, Z_2^n, \ldots Z_w^n$) in which the set of training observations

$X_1, X_2, \ldots X_n$ can be partitioned constituting the overall mixture distribution.

If $P(X | \varphi, P)$ represents the parameter-conditional mixture distribution where:

$\varphi = \varphi_1, \varphi_2, \ldots \varphi_w$ and

$$P = \{P(Z_1^n), P(Z_2^n), \ \ldots P(Z_w^n) \quad \} \qquad (8)$$

are the two sets of parameter and $P(X | \varphi_1 Z_1^n)$ the ith parameter-conditional distribution, then in terms of the set of parameters the above equation becomes

$$P(X | \varphi, P) = \sum_i P(X | \varphi_I, Z_i^n)\, P(Z_i^n),\ i = 1, 2, \ldots .\ w \qquad (9)$$

The problem of non-supervised learning is thus reduced to that of fin ding a unique solution for φ and P, given $P(X | \varphi, P)$.

If we assume that there are two pattern classifiers $C_1$and C_0having the respective known form of the probability density functions $P(X | C_1)$ and $P(X | C_2)$, and the parameter φ is obtained by the equation:

$$P(\varphi | X_1, \ldots X_n) = \sum_i P(\varphi | X_1, \ldots\ X_n, Z_i^n)\, P(Z_i^n \ | X_1, \ldots, X_n) \qquad (10)$$

$$I = 1, 2, \ldots w,\ w = 2^n$$

The problem is therefore reduced to that of supervised learning for each of the 2^n partitions.

Clustering and unsupervised Learning: Clustering analysis is the assignment of a set of observation into subsets (clusters) so that the observations within the same cluster are similar according to some pre-designated criterion or criteria, Timothy, Jason and Shepard, P. J. (1998).
It is a supervised learning that automatically forms clusters of similar things. It can be described as an automatic classification. The main difference is that, in classification, we know what we are looking for, but for clustering, we do not have such information. Since it also produces the same result as classification but without having predefined classes. Clustering is sometimes called unsupervised classification.

V. Conclusion

The above approach has explored the Bayesian learning techniques of supervised and non-supervised learning. Although many other Learning techniques exist; the Stochastic Approximation method, Parameter Adjustment, etc. for example. Bayes' rule is optimal in the sense that it minimizes the expected loss or the probability of misclassification. Apparently, measurements with higher complexity provide more information for the classification rule.

Machine learning is a broad area that is applied in various fields. Its application in broad areas such as data mining and Artificial Intelligence, and other various fields led to numerous existing techniques and algorithms. In general, the methods have aided lots of decision-making systems by learning from data of different types. The recent interest in the Internet of things and sensor networks, which is expected to grow rapidly to 50 billion devices and beyond, has sparked a great interest for the application of Machine Learning in pervasive environments. In view of this, this study is rudimentary introduction to common Machine Learning techniques that have been applied in the area ofcontext-aware pervasive computing.

References

[1]. Abbeel, P. and A.Y. Ng, "Apprenticeship learning via inverse reinforcement learning," in Proceedings of the Twenty-first International Conference on Machine Learning, Banff, Alberta, Canada, 2004, pp. 1–8.Google Scholar
[2]. Alpaydin, E. (2004). Introduction to Machine Learning. Massachusetts, USA: MIT Press.
[3]. Bishop, C. M. (1995). Neural Networks for Pattern Recognition. Oxford, England: Oxford, University Press.
[4]. Bishop, C. M. (2006). Pattern Recognition and Machine Learning (Information Science and Statistics). New York, New York: Springer Science and Business Media.
[5]. Friedberg, R. M. (1958). A learning machine: Part, 1. IBM Journal, pp. 2-13
[6]. Ghahramani, Z. (2008). Unsupervised learning algorithms are designed toextract structure, from data. 178, pp. 1-8. IOS Press.
[7]. Ian H. Witten, E. F. (2005). Data Mining Practical Machine Learning and Techniques (Second edition ed.). Morgan Kaufmann.
[8]. Jaime G. Carbonell, R. S. (1983). Machine Learning: A Historical and Methodological Analysis. Association for the Advancement of Artificial Intelligence, 4 (3), 1-10.
[9]. Karbassi, A. and M. Barth, (2003) "Vehicle route prediction and time of arrival estimation techniques for improved transportation system management," in Intelligent Vehicles Symposium, Columbus, Ohio, USA. pp. 511–516.Google Scholar
[10]. Mitchell, T. M. (2006). The Discipline of Machine Learning. Machine Learning Department technical report CMU-ML-06-108, Carnegie Mellon University.
[11]. Nicole, A. Context-aware web search using dynamically weighted information fusion. Concurr. Comput. Pract. Exp. 2013, 25, 672–685.
[12]. Perttunen, M., Riekki, J., Lassila, O. (2009). Context representation and reasoning in pervasive computing: A review. Int. J. Multimed. Ubiquitous Eng., 4, 1–28.
[13]. Schilit B. and Thelmer M. (1994) Disseminating Active Map information to Mobile Hosts. IEEE Network Vo8(5) 22-32.
[14]. Timothy Jason and Shepard, P. J. (1998). Decision Fusion Using a Multi-Linear Classifier. In Proceedings of the International Conference on Multisource -Multisensor Information Fusion.
[15]. Weiser, M. (1999). The computer for the 21st century. ACM Sigmobile Mob. Comput. Commun. Rev.1999, 3, 3–11.

Dr. Nlerum, Promise A"Machine Learning: Application To Pervasive Computing Systems " *International Journal of Engineering Science Invention (IJESI)*, Vol. 09(01), 2020, PP0 57-62