



A Survey Paper on Artificial Intelligence and Machine Learning

¹Shilpa Veerabhadrapa, ²Puttaraj K, ³Puttaraj B, ⁴Manjunath Gulguli
¹Lecturer, ²BCA II Year Student, ³BCA II Year Student, ⁴BCA II Year Student
¹BCA Department,
¹KLE's JT BCA College Gadag, Gadag, Karnataka. INDIA, Pin-582101

Abstract: Artificial intelligence is defined as developing computer programs to solve complex problems by applications of processes that are analogous to human reasoning processes. It is that branch of computer science that studies and develops intelligent machines and software. This paper is a survey which contains emerging of the technology, AI Design Models and also k-means algorithm.

Index Terms - Artificial algorithm , AI Design Models, k-means algorithm and clustering algorithm.

I. INTRODUCTION

Artificial intelligence is defined as developing computer programs to solve complex problems by applications of processes that are analogous to human reasoning processes. It is that branch of computer science that studies and develops intelligent machines and software. Artificial intelligence was first proposed by John McCarthy in 1956 in his first academic conference on the subject. The idea of machine operating like human beings began to be the center of scientist's mind and whether if it is possible to make machines have the same ability to think and learn by itself was introduced by mathematician Alan Turing.

Alan Turing was able to put his hypotheses and questions into actions by testing whether "*machines can think*"? After series of testing (later was called as Turing Test) it turns out that it is possible to enable machines to think and learn just like humans. Turing Test uses the pragmatic approach to be able to identify if machines can respond as humans.

Artificial Intelligence is the field to study that describe the capability of machine learning just like humans and the ability to respond to certain behaviours also known as (A.I). The need of Artificial Intelligence is increasing every day. since AI was first introduced to the market, it has been the reason of the quick change in the technology and business fields. Computer scientist are predicting that by 2020 "85% of customer interactions will be managed without a human". This means that humans simple request will depend on computers and artificial intelligence just like when we use Siri or Galaxy to ask about the weather temperature. It is very important to be prepared for AI revelation just like UAE have by installing a state minister for AI in Dubai.

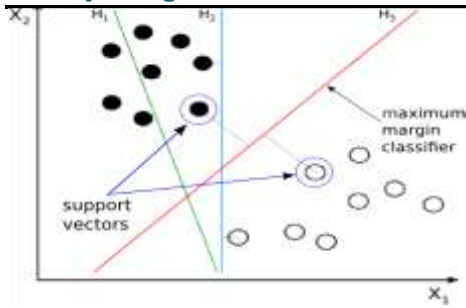
AI offers reliability, cost-effectiveness, solve complicated problems, and make decisions; in addition, AI restrict data from getting lost. AI is applied nowadays in most fields whether business or engineering. One of the great tools in AI is called "reinforcement learning" which is based on testing success and failure in real life to increase the reliability of applications. Unfortunately, AI is limited with its capability and functionality.

Although Artificial Intelligence made our lives much easier and saved us more time than ever, scientists are predicting that by the huge dependency on AI humanity could extinct. Scientists argue that by having a AI machines, people will be jobless and that will conclude in losing the sense of living. Since machines are learning and doing things more efficiently and affectively in a timely manner, this could be the reason of our extinction.

II. AI ALGORITHMS AND MODELS

AI is mainly based on algorithm and models as a technique which is designed based on scientific findings such math, statistics, and biology. AI works based on several models such as: Ant colony Algorithm, Particle Swarm Algorithm, Neural Network, Deep Learning and in this report, some of the most known models which are: Support Vector Machine, and the Artificial Neural Network.

Support Vector Machine (SVM) where it is used to build a classification model by finding an optimal hyperplane based on a set of training examples as shown in figure. It is also have been used for pattern classification and trend prediction lots of applications for instance: power transformer fault diagnosis, disease diagnosis and treatment optimization.

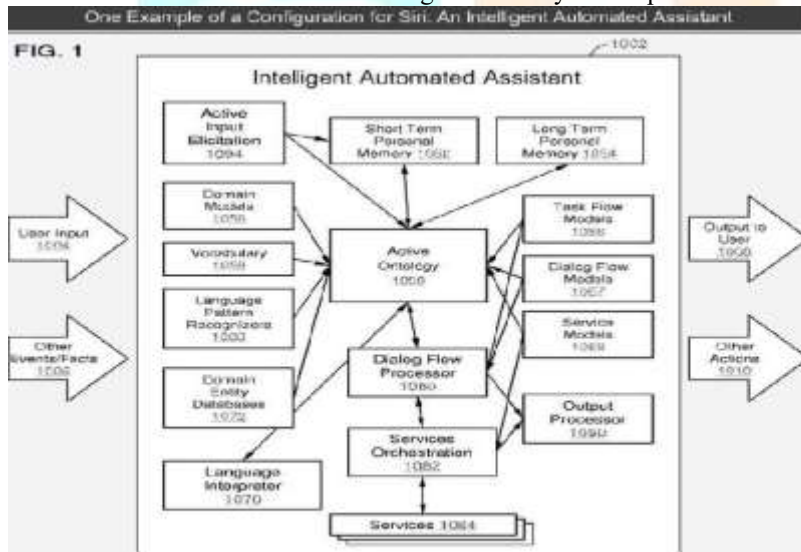


Describes how SVM algorithm being represented in AI. Artificial Neural Network (ANN) is a representative model of understanding thoughts and behaviours in terms of physical connection between neurons. ANN has been used to solve variety of problems through enabling the machine to build mathematical models to be able to imitate natural activities from brains perspective as shown in figure below .By using this algorithm, the machine will be able to identify the solution of any problem just like human’s brain.

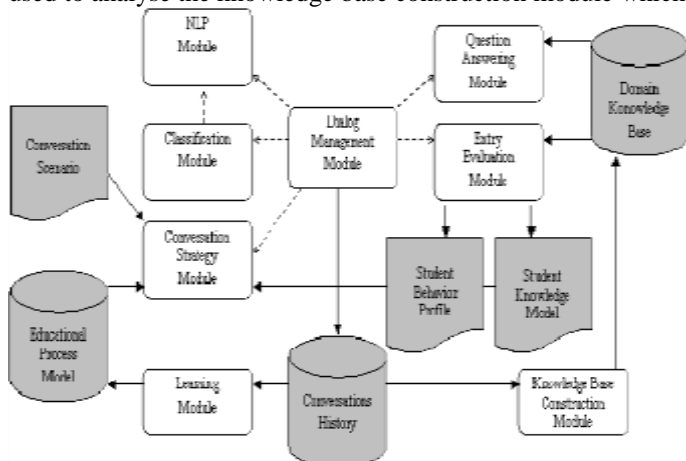


Describes Phases of Siri Virtual Assistant:

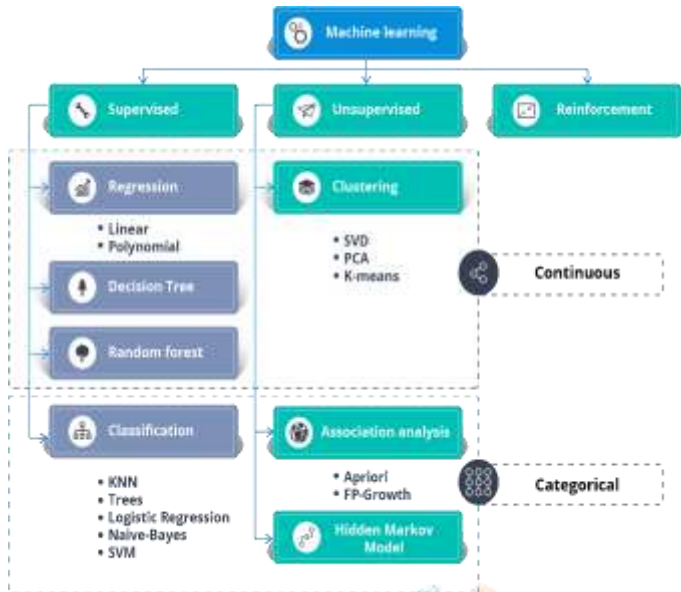
Siri is the well-known virtual assistant which uses voice recognitions and typed command in order to perform a certain task within a device. Siri is considered one of AI most used applications. The application simply takes the input from the user such as (e.g. Call dad) and try to find the most related keywords used in this command. Siri tries to eliminate inconsistent result through using the language pattern recognizer and from there to active ontology by searching through the contacts, then it tries to relate the contact named “Dad” and perform the task which is in this case is “Calling” and finally the output of this action will be “calling dad” and to consider all the possible situations.



In another scenario the architecture of the virtual assistant is shown in (figureA-5) as we can see the flow of the system starts by taking the input from the user, after that the system decide the conversation strategy module to be used which is a respond from the dialog management module, meanwhile a classification module response to an NLP module. Finally, using the conversation history database is used to analyse the knowledge base construction module which will response back to the domain knowledge based as explained in detail.



III. MACHINE LEARNING



IV. PREPARE YOUR PAPER BEFORE STYLING

Before you begin to format your paper, first write and save the content as a separate text file. Keep your text and graphic files separate until after the text has been formatted and styled. Do not use hard tabs, and limit use of hard return to only one return at the end of a paragraph. Do not add any kind of pagination anywhere in the paper. Do not number text heads—the template will do that for you.

Finally, complete content and organizational editing before formatting. Please take note of the following items when proofreading spelling and grammar. The K-Means clustering algorithm is interesting and different from other MapReduce algorithms. It is an iterative algorithm (that is, it requires multiple MapReduce phases) that you execute many times with different centroids until it converges (meaning that K optimal clusters are found after many iterations of the same MapReduce job).

K-means clustering is the most commonly used unsupervised machine learning algorithm for partitioning a given data set into a set of k groups. Where k represents the number of groups pre-specified by the analyst. It classifies objects in multiple groups such that objects within the same cluster are as similar as possible whereas objects from different clusters are as dissimilar as possible.

K-means algorithm

K-means algorithm can be summarized as follow:

- Specify the numbers of clusters (K) to be cleared (by the analyst)
- Select randomly k objects from the data set as the initial cluster centers or means
- Assigns each observation to their closest centroid, based on the Euclidean distance between the object and centroid
- For each of the k-clusters update the cluster centroid by calculating the new mean values of all the data points in the cluster. The centroid of a Kth cluster is a vector of length p containing the means of all variables for the observations in the Kth cluster; p is a number of variables.

Iteratively minimize the total within sum of square. that is, iterate steps 3 and 4 until the cluster assignments stop changing of the maximum number of iterations is reached. By default, the R software uses 10 as the default value for the maximum number of iterations.

Computing K-means clustering in R Data

We'll use the demo data sets "USArrests". The data must contains only continuous variables, as the K-means algorithm uses variable means. As we don't want the K-means algorithm to depend to an arbitrary variable unit, we start by scaling the data using the R function scale() as follows:

K-means clustering the advantages and disadvantages

K-means clustering is very simple and fast algorithm. it can efficiently deal with very large data sets.

k-means clustering algorithm

$$J(V) = \sum_{i=1}^c \sum_{j=1}^{c_i} (\|x_i - v_j\|)^2$$

Where,

- $\|x_i - v_j\|$ is the Euclidean distance between x_i and v_j
- c_i is the number of data points in i^{th} cluster
- c is the number of cluster centers.

Algorithmic steps for k-means clustering

Let $X = \{x_1, x_2, x_3, \dots, x_n\}$ be the set of data points and $V = \{v_1, v_2, \dots, v_c\}$ be the set centers.

- 1) Randomly select 'c' cluster centers.
- 2) Calculate the distance between each data points and cluster centers.
- 3) Assign the data point to the cluster center whose distance from the cluster centers.
- 4) Recalculate the new cluster center using:

$$v_i = (1/c_i) \sum_{j=1}^{c_i} x_j$$

Where, 'c_i' represents the number of data points in ith cluster

- 5) Recalculate the distance between each data point and new obtained cluster center
- 6) If no data point was reassigned then stop, otherwise repeat from step 3).

Advantages

- Fast, robust and easier to understand.
- Relatively efficient: $O(tknd)$, where n is # objects, k is # clusters, d is # dimensions of each object, and t is # iterations. Normally, $k, t, d \ll n$.
- Gives best result when data set are distinct or well separated from each other.

Disadvantages

- The learning algorithm requires a priori specification of the number of the cluster centres.
- The use of Exclusive Assignment- If there are two highly overlapping data then k-means will not be able to resolve that there are two clusters.
- The learning algorithm is not invariant to non-linear transformations i.e. with different representation of data we get different results (data represented in form of cartesian co-ordinates and polar co-ordinates will give different results).
- Euclidean distance measures can unequally weight underlying factors.
- The learning algorithm provides the local optima of the squared error function.
- Randomly choosing of the cluster center cannot lead us to the fruitful result. P1.refer fig
- Applicable only when mean is defined i.e. fails for categorical data.
- Unable to handle noisy data and outliers.
- Algorithm fails for non-linear data set.

IV. CONCLUSION

It is difficult to make a machine like humans which can show emotions or think like humans in different circumstances. AI nowadays is being implemented in almost every field of study through several models such as SVM and ANN. We should be able to proceed with knowing and understanding the consequences of every technological trend. We are in the AI revelation era and therefore, We should adopt into this change and welcome it too by embracing AI and moving, The learning algorithm requires a priori specification of the number of the cluster centres.

REFERENCES

- [1] Xiangjie Kong, Kai Ma, Shen Hou, Di Shang, Feng Xia, "Human Interactive Behavior: A Bibliographic Review", *Access IEEE*, vol. 7, pp. 4611-4628, 2019.
- [2] Xiaomei Bai, Mengyang Wang, Ivan Lee, Zhuo Yang, Xiangjie Kong, Feng Xia, "Scientific Paper Recommendation: A Survey", *Access IEEE*, vol. 7, pp. 9324-9339, 2019
- [3] Wei Wang, Jiaying Liu, Zhuo Yang, Xiangjie Kong, Feng Xia, "Sustainable Collaborator Recommendation Based on Conference Closure", *Computational Social Systems IEEE Transactions on*, vol. 6, no. 2, pp. 311-322, 2019.