www.ijecs.in
International Journal Of Engineering And Computer Science ISSN:2319-7242
Volume - 3 Issue -9 September, 2014 Page No. 8276-8280

Smart Home Implementation Using Data Mining

Gayatri D. Kulkarni[1], Priyanka V. Gode[2],Jadi Pratapreddy[3],Madhura H. Deshmukh[4],Nitin R. Talhar[5]

[1]Student(UG),Department of Computer Engineering
A.I.S.S.M.S College of Engineering,Pune,Maharashtra,India
gayatri6882@yahoo.com

[2]Student(UG),Department of Computer Engineering
A.I.S.S.M.S College of Engineering,Pune,Maharashtra,India
gode.priyanka@yahoo.in

[3]Student(UG),Department of Computer Engineering
A.I.S.S.M.S College of Engineering,Pune,Maharashtra,India
jpreddy019@gmail.com

[4]Student(UG),Department of Computer Engineering
A.I.S.S.M.S College of Engineering,Pune,Maharashtra,India
madhura.22MD@gmal.com

[5]Assistant Professor,Department of Computer Engineering
A.I.S.S.M.S College of Engineering,Pune,Maharashtra,India
nrtalhar@gmal.com

Abstract: **In** ***this project we are attempting to design a home that acts as an intelligent agent. This paper mainly focuses on developing a system which will automate all domestic devices. Datasets are built for analysis purpose. Voice is recorded and then converted in textual form. Then patterns of text are matched with the data sets provided. Data sets consisting of different conditions will be the initial input to the system. According to the input provided at given instant, corresponding action will be performed by the system.***

Keywords: HMM (Hidden Markov Model) , Sphinx, Datasets, Naïve Bayes

1. Introduction

This project aims at creating an environment (automated home) that acts as an intelligent agent [1].The main purpose is to maximize the comfort of the inhabitants. People with physical disabilities will be benefitted using this product. From a particular location in the area covered by the system, they can give the voice commands very easily. These voice commands and the system time and date will be used to perform the basic domestic automation. According to the input given and the datasets trained the output action will be performed and then it will be depicted in the interface. The domestic automation will be easily performed by a person who is physically disabled. Product can be also be used by a normal person who is temporarily disabled or it can be used by old aged people who find it difficult to manage the basic home automation [8].

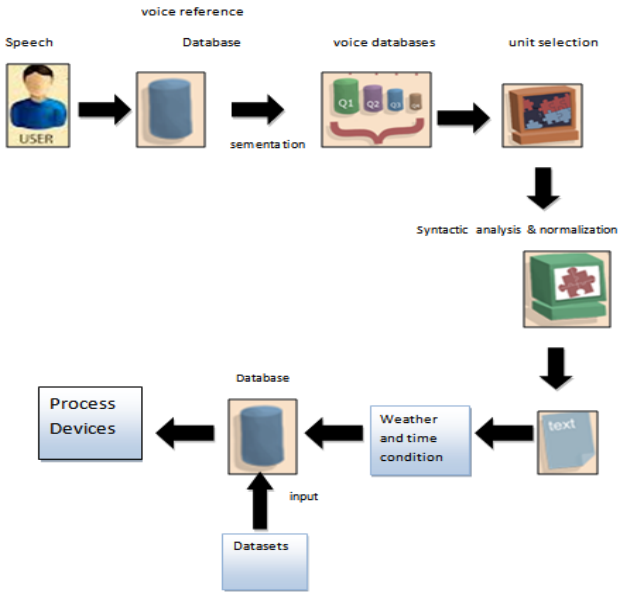

Figure 1: Architecture of the system

2. METHODOLOGY:

The methodology of this project is as follows:

1) The application which is designed accepts speech as an input, which the user will give as a voice input to the system.
2) First speech will be recognized, then processed and then converted to text. This voice input will be converted in textual form.
3) During the process of giving the voice commands to the system, care must be taken that there is minimum background sound or a proper working microphone must be connected to the system, so that the unnecessary sounds of the background will be automatically cancelled.
4) From the text input, conditions and keywords will be extracted. A condition and the keyword will be considered as one of the attributes of the dataset.
5) Other attributes considered in the datasets will be the system time and date. The system time and date must be properly adjusted so as to get the desired output.
6) The system considering the given input and the previous datasets used for its training will predict a proper operation (*action*).Finally the predicted operation (*action*) will be performed by the system.
7) Desired proper output i.e. user will get a proper operation performed (*action*). Improper date and time will result in an improper action performance.
8) This dataset will be given as an input to the system in the form of a database.

3. IMPLEMENTATION:

Naive Bayes is a probabilistic classifier based on the Bayes theorem, considering Naive (Strong) independence assumption. The assumption in Naive Bayes is that the effect of a variable value on a given class is independent of the values of other variables. This assumption is known as **Class Conditional Independence**. It is particularly suited when the dimensionality of the inputs is high. When more competent output is needed, as compared to other methods we can use Naïve Bayes implementation. Naïve Bayesian is used to create models with predictive capabilities [5].

3.1 Bayes Theorem

Probability (B given A)=Probability(A and B)/Probability(A).
For B given A=Number of cases where A and B work together divided by cases in which A works alone. Let X be a data tuple. In, Bayesian terms, X is considered "**evidence**". Let H be some hypothesis, such that the data tuple X belongs class C. P (H|X) is the posterior probability of H conditioned on X. In contrast, P (H) is the prior probability of H. Bayes' theorem states that:

P(H|X)=P(X|H)P(H)/P(X)
Posterior=Likelihood*Prior/Evidence

3.2. Bayesian Classifiers:

Approach:

Compute the posterior probability:

1. P (H|X1,X2,….,Xn) for all values of H using Bayes theorem.
2. To estimate P(X1,X2,…Xn |H) select all the tuples of class H in the training set.
3. Count all possible combinations of X1, X2 , …, Xn. However all combinations may not be present [5].

3.3. Why Naïve Bayes?

It relies on counting techniques, to count the probabilities. There are two types: numerical and categorical**.** Numerical Data can be classified in range of values like low, medium and high. Categorical data can be classified as meta-classes.

For the numerical values , the mean and deviation values are calculated .Thus Naïve bayes classifier handles the class priors. It also normalizes the observations and is based on assumptions. It handles the missing data and also has good computational complexity . There are two types:

1)Continuous variables
2) Discrete variables

4. DATABASE PREPARATION:

4.1. Speech to Text:

a) Use of HMM (Hidden Markov model):

System which is modeled with hidden(unobserved) states.

State is not directly visible but output dependent on time is visible .The reason behind the popularity of the HMM is it's training can be done automatically. It is feasible to use. Words are represented in the complete format. The parameters of this model are state transition probabilities , means that characterize state output distributions.HMM for the sequence of words is made by combining the individual trained HMM for single word [6].

Sphinx is a continuous- speech, speaker- independent recognition system which makes use of Hidden Markov acoustic models [7]. Sphinx consists of 3 types: Sphinx 2,Sphinx 3,Sphinx 4 and also Pocket sphinx. In the project we are using Sphinx-4 version. It is a complete rewrite of Sphinx and a pure Java speech recognition library .There are various high-level recognition interfaces in sphinx 4:

1) LiveSpeechRecognizer
2) StreamSpeechrecognizer
3) SpeechAligner

For this we have to set 4 attributes:

1)Acoustic model
2)Dictionary
3)Grammar/Language model
4)Source of speech

Our proposed system uses the microphone for the voice input, hence we can deduce that we are using LiveSpeech Recognizer as an interface of sphinx-4 [9].

4.2. System Time and Date:

System time and date can be found out by using Java code. Specific date packages need to be installed for finding the time and date at given instant. Date can be displayed in any of the specified formats. After receiving the date, month will be extracted which will be used as input specification to the datasets training.

4.3. Detecting conditions of the particular day:

The month which is found from the code , is provided as one of the attributes of the dataset. Conditions like humidity and also the temperature can be found by using specific java packages.

5. TRAINING DATASETS:

Attributes of the datasets:

a) Date and time table
b) Weather conditions table
c) Room table
d) Action to be performed table

All the attributes of the datasets are then matched for further procedure. Date and time will be the output obtained from the java code. Weather conditions are the output of the Naïve Bayes, where season is detected [3].

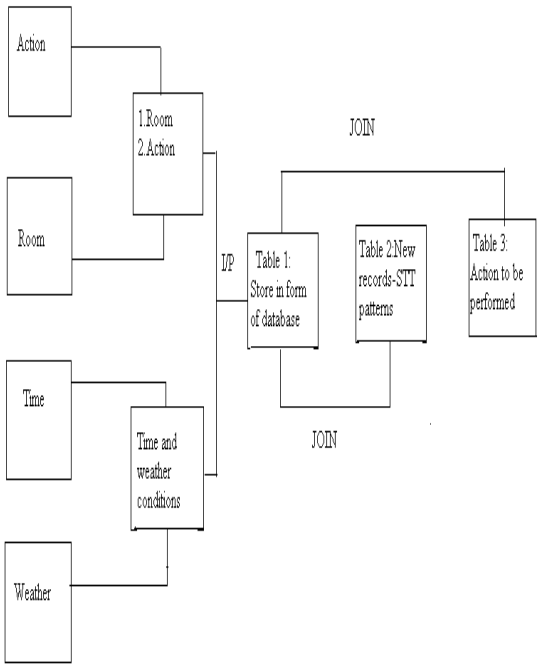

Figure 2: Tables of database

5.1. TABLES USED:

By using java code system time and date can be predicted. From the information provided table is prepared. Date can be provided in dd-mm-yyyy format and time in hh:mm:ss. A 24 hour format is used. On basis of that , it will decide whether it is morning or evening time. If evening time light will be switched on. Thus system time and date is predicted.

Table 1: Time and date

Date	Time
29-11-2006	12:30:47

5.2. NAÏVE BAYES:

Using code , temperature, humidity at given instant can be predicted. Once the temperature is predicted, the range of the temperature will be given and then words will be assigned to them ,on the basis of which attributes will be assigned to the

values which will be used as tables for the prediction on the basis of the Naïve Bayes algorithm. Various attributes like windy weather, temperature and humidity are given.

5.2.1. Attributes:

Table 2: Naïve Bayes attributes

Temperature	Humidity	Windy	Class (Fan)
Hot	High	False	Y
Hot	High	True	N
Hot	High	False	Y
Mild	High	False	N
Cool	Normal	False	Y
Cool	Normal	True	N
Mild	High	False	Y
Cool	Normal	False	N
Mild	Normal	False	Y
Mild	Normal	True	Y
Mild	High	True	N
Hot	Normal	False	N

From these attribute values are provided, the probability for Yes and No is to be calculated [2].

Y=0.5, N=0.5

Probability of each attribute:

5.2.1.1. Temperature:

Table 3: Temperature Conditions

Temperature	Y	N
Hot	3/8	1/4
Cool	2/8	1/4
Mild	3/8	2/4

5.2.1.2. Humidity:

Table 4: Humidity Conditions

Humidity	Y	N
High	5/8	2/4
Normal	3/8	2/4

5.2.1.3. Windy:

Table 5: Windy Weather

Windy	Y	N
True	2/8	3/4
False	6/8	1/4

Consider the dataset as <cool, normal, true>

1) Calculate the probability for 'Y':

P (cool |Y) =1 + (0.5*2)/8=0.25
P (normal |Y) = 3 + (0.5*2)/8=0.5
P (true |Y) =1 + (0.5*2)/8=0.25

2) Calculate the probability for 'N':
P (cool |N) = 2 + (0.5*2)/8=0.37
P (normal |N) = 3 + (0.5*2)/8=0.5
P (true |N) = 3 + (0.5*2)/8=0.5

3) Calculation of final probability:

Probability for 'Y':0.5*0.25*0.5*0.25=0.015625
Probability for 'N':0.375*0.5*0.5*0.5=0.046875

0.046875>0.015625
P (true |N) = 3 + (0.5*2)/8 = 0.5
Hence answer is 'N' i.e. do not switch on the fan when temperature is high, humidity is normal and windy weather [2].

5.3. Merging Datasets:
The output of speech to text gives the room condition .From time entity, we get the condition for switching the light. And from the naïve bayes we have condition for the fan.Thus all these conditions will be merged in one table to perform the action as required [3].

Table 6: Final database

ROOM	TIME	ACTION	
LIVING ROOM	12:20:49	LIGHT (OFF)	FAN (ON)
BEDROOM	23:30:45	LIGHT (ON)	FAN (OFF)

Thus the above table consists of room, time and action based on the results of the above tables. Thus, according to the room, time and other conditions in the datasets, device will be turned on or off and the change will be depicted in the interface. The interface will consist of applications probably done by using java coding i.e. by using JSP and Servlets.

6. CONCLUSION AND FUTURE WORKS:
This paper presented an approach to create a smart home system which can be handled by inhabitants of the house especially the handicapped ones. The development of our project is based on data mining techniques .From a specific location in the area covered by the system, one can give the voice commands very easily and then the system performs actions.

The proposed system is speaker independent . A speaker dependent system can be devised in future. Also speaker input provided should be accessed in different languages as desired by the speaker [4].

[1] R.VINAY CHAND, M.VEDA CHARY**2**," Wireless Home Automation System with Acoustic Controlling, " International Journal of Engineering Trends and Technology, 9- Sep 2013
[2] Eric Meisner , " Naive Bayes Classifier example "
[3] Carlos Ordonez,Sofian Maabout,David Sergio Matus evich ,Wellington Cabrera, Extending ER models to capture database transformations to build data sets for data mining",Data and Knowledge Engineering of science direct,2014
[4] Jagmeet Kaur, Parminder Singh ,"Punjabi Speech Synthesis System for android mobile phones",Inter-national journal phones", International Journal of Engineering and Computer science , 9 September 2014
[5] Victor Lavrenko and Nigel Goddard," Naïve Bayes", School of Informatics
[6] http://en.wikipedia.org/wiki/Hidden_Markov model
[7] http://en.wikipedia.org/wiki/Sphinx
[8] S.P.Pande ,rof.Pravin Sen," Home Automation system for disabled people usingBCI",International Conference on advances in Engineering andTechnology-2014(ICA ET-2014)
[9] http://cmusphinx.sourceforge.net/wili/tutorialsphinx4

Author Profile

Gayatri D. Kulkarni pursuing Bachelor's degree from Savitribai Phule Pune University in A.I.S.S.M.S College of Engineering, Pune, Maharashtra ,India.

Priyanka V. Gode pursuing Bachelor's degree from Savitribai Phule Pune University in A.I.S.S.M.S College of Engineering, Pune, Maharashtra ,India.

References:

Jadi Pratapreddy pursuing Bachelor's degree from Savitribai Phule Pune University in A.I.S.S.M.S College of Engineering, Pune, Maharashtra, India

Madhura H. Deshmukh pursuing Bachelor's degree from Savitribai Phule University in A.I.S.S.M.S College of Engineering ,Pune ,Maharashtra, India.

Nitin R .Talhar, Assistant Professor, completed Master's from CSE (IT), SCOE,Pune.Attended various conferences at national and international level .Research in realtime and object-based Streaming Techniques with application to communication systems in 2009.