

Smart Home Automation in Air Conditioning using Data Mining and Image Processing

Ishaan Arora, Vanmathi C

School of Information Technology and Engineering , VIT University,
Vellore, India

Abstract

Any machine as we know is something that decreases human effort and increases comfort. Technology has been embedding into our daily usage appliances for a decent time now. We are surrounded by a number of intelligent systems who have been in a continuous run to increase our comfort using the best technologies available.

Smart Home automation in one of the major milestones now in the field of bringing the technology to our homes . Home automation is building computerization for a home, called a smart house. It includes the control and automation of lighting, warming, (for example, smart indoor regulators), ventilation, cooling ,and security, and additionally home apparatuses, for example, washer/dryers, broilers or iceboxes/freezers. Wi-Fi is regularly utilized for remote observing and control. Home gadgets, when remotely checked and controlled by means of the Internet, are a vital constituent of the Internet of Things. Current frameworks by and large comprise of switches and sensors associated with a central hub from which the overall functioning is controlled with a UI that is cooperated either with a cell phone software, tablet PC , a web interface, regularly however not generally by means of cloud computing.

In this paper we discuss about making a daily use appliance intelligent so it takes its decision and acts accordingly. The goal of the paper is to make the air conditioner set the appropriate temperature depending the need of user or users that are currently utilizing the service . The paper relies on data to take its necessary decision depending on 8 attributes like room temperature , weather , time of day and various such attributes and depending on various inputs given by the user in the past which are stored as datasets and then used for decision making in the future . While there are factors or attributes that can be easily observed recorded there are attributes which might need an embedding of a separate technology to bring out desiring results and make the technology or proposed concept even better . For the same we wish to add object recognition technology into it to take its decision more effectively and bring out better results.

Survey was carried out to collect large amount of data for analysis. On analyzing the given data 8 factors were found to be contributing towards the

result. Data was further analyzed through various algorithms to check the result through various aspects. Among the various algorithms, the most accurate algorithm was selected. Selected Algorithm is further coded in python.

Keywords: Machine Learning, RapidMiner, Classification, Decision Tree , Naive Bayes, Image Processing

I. INTRODUCTION

We are advancing every day with the technology in every field and aim of which is to ease our access to available facilities and hence home automation is one of them . Lights , Fans everything these days is available that can be operated using a Remote controlled device be it a mobile app or some Tablet. Sensing the need and urgency for such a process a technical solution is called for which will not only help the users but will also give other minds a new filed to dig deep into it and bring out new aspects.

The proposed system is expected to predict the desirable temperature class for the user . The predicted temperature is divided into 5 classes A,B,C,D,E to make the prediction and classification easier. This prediction is generated by a prediction algorithm that acts best among all algorithms for the desired dataset. The parameters, such as accuracy, F-measure, ensure an unbiased and a clear judgment over the classifiers. The data was created using surveys of students and expert research. The data presented demonstrates various factors that decide what temperature a user finds desirable when using an Air Conditioner. One important thing that is introduced is the involvement of Camera to calculate one of the factors. Another main point of focus is the selection of best classifier using the RapidMiner tool and other analysis. The developed application can be embedded into a system or into a mobile application and will play a vital role in effective temperature class prediction.

II. LITERATURE SURVEY

In [1], the design and development of an air conditioner on a card based PIC16F877A microcontroller is discussed and implemented to get

the desirable thermal feeling. Different sorts of aeration and cooling system control cards are accessible in the business sectors these days yet this control card in this framework is diverse relying upon microcontroller technology by changing control plan and parameters. It is more efficient to use for the buyer. The aerating and cooling framework is used to control the temperature for the esteems of cooling and heating procedure.. As indicated by the control plan and equipment frameworks, programming usage of the control framework are realized.

In [2], In this work, an Intelligent Air Conditioner plot is proposed and its adjustment to a Smart Home framework is exhibited. Right off the bat, working standards of customary Air Conditioner and the proposed smart Air Conditioner are compared and analysed.

Moreover, when it comes to classification and prediction algorithms, the testing must go on with other competing classifiers such as Naive Bayes and Neural Networks. Therefore, with the study of different algorithms and their respective specialties were studied and three among them were shortlisted apparently as the best algorithms for our datasets. [7]

In [3], an overview of object recognition has been demonstrated. Python and OpenCV have been utilized to make one of a kind Cat finder and recognizer. Additionally, cases of human face discovery and acknowledgment are covered . A methodology for training a detector for any class of Object and for recognizer have been developed .

We have limited our research to three algorithms. There are some factors that might not contribute significantly to the research but might be helpful for future research

III. METHOD DATASET AND VARIABLES

Dataset consists of 8 feature attributes and one target attribute. The weightage of each feature attribute differs. The dataset consists of around seven hundred instances. Abbreviations of attributes are explained in the following table.

TABLE I. DATASET RELATED VARIABLES

Time	Time of the day(Morning, Afternoon, Evening , Night)
Season	Season of the Year(Spring, Summer, Autumn)
People	Number of people in room(Numeric : 1 to 5)
Humidity	Current Humidity outside(Heavy, Mild, Low)
Lights	The working status of Lights in room (Binary : On or Off)

Fan Speed	Speed at which fan is working (Binary : 1 to 5)
Doors	The status of doors and windows of Room(Binary : Open or Close)
Outer Temperature	The weather outside in open air at that point of time in Degree Celsius (Numeric : 25 to 45)

All the variables above mentioned are used for the prediction. Although considered ,factors like Lights might not contribute effectively since radiation of lights won't cause any significant change in temperature . For one of the factors like People we have used Image processing (Human Detection) for counting the number of people in the room assuming the camera is embedded within the Air Conditioner

IV. HUMAN DETECTION

OpenCV is implemented with a pre-trained HOG(Histogram of Oriented Graphs) and Linear SVM(Support Vector Machines) model that can be used to perform Human detection in both images and video streams. A non maxima suppression algorithm is used to take numerous, overlapping area bounding boxes and reduce them to only a single bounding box. detectMultiScale is an important factor of this detection. It is used to detect objects of different sizes in the image. detectMultiscale constructs and image pyramid of scale = 1.05 . A larger scale makes the algorithm run faster. However if too large, accuracy will decrease.

The entire code is coded in Python. For now images present in a computer directory are used .Although HOG is better than it HAAR counter-part. It still requires parameters like detectMultiScale to be set properly. A sample result is shown below.

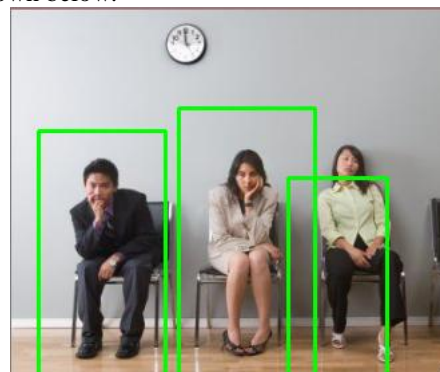


Figure 1. Sample Result of the Detection Code

The number of people in the room is the number of rectangles in the final output. The implementation is capable of providing a satisfying accuracy and

hence provides us the application/program the number of people in the room.

V. CLASSIFIERS

A. Naive Bayes

Naive Bayes is a probabilistic characterization which display the probability of class participations given the information and make a probabilistic model of information of each class. Here all factors of each class are autonomous of each other. Probability of each class is evaluated and the class with most highest probability is chosen. It is best implemented in separating spam messages.[4]

B. Decision Tree

Decision tree is the principle innovation utilized for classification and prediction. Decision tree learning is based in view of occurrence, which center around classification rules showing as decision trees deduced from a gathering of disorder and irregular instance. It works in a top-down recursive fashion and draws a conclusion from the leaf nodes by comparing internal nodes and by judging the downward branches according to different attributes of node.[5]

C. Random Forest

A Random Forest is a group of unpruned decision trees. It works by creating many decision trees and for classification and prediction. Random Forests are frequently utilized when we have huge Training datasets and an extensive number of information factors .[6]

Analysis of 3 different algorithms of classification known as 'Classifiers' is carried out and is judged on the basis of 4 parameters Accuracy, Precision, Recall , F-Measure . The four parameters used for judgment are as given below:

- Accuracy = $((\text{True Positive} + \text{True Negative}) / (\text{P} + \text{N})) * 100$
- Precision = $(\text{True Positive} / (\text{True Positive} + \text{False Positive})) * 100$
- Recall = $(\text{True Positive} / (\text{True Positive} + \text{False Negative})) * 100$

$$\text{F-Measure} = (2 * \text{Precision} * \text{Recall} / (\text{Precision} + \text{Recall})) * 100$$

VI. EXPERIMENT AND DISCUSSION

All the All the three classifiers were implemented on Rapidminer tool to find the best among them. The results of respective classifiers are given below.

A. Naive Bayes

It is one of the best classifier and works well with small datasets using probability function and gives us the best accuracy of 80% . It is a quite fast algorithm.

B. Decision Tree

It relies on calculating the Gini Index and Gain Ratio. It works well with mediocre sized datasets by calculating the most favorable attribute on which the target attribute is dependent.

C. Random Forest

As Random Forest works best when we have humongous datasets and hence won't go well with our dataset of 700 instances still helps us provide an output accuracy of 69%

TABLE II. COMPARISON RESULTS OF DIFFERENT CLASSIFIERS

	Accuracy	Precision	Recall	F-Measure
Naive Bayes	80	83.3	71.126	76.73
Decision Tree	74	70.182	65.5	67.76
Random Forest	69	56.52	58.1	57.29

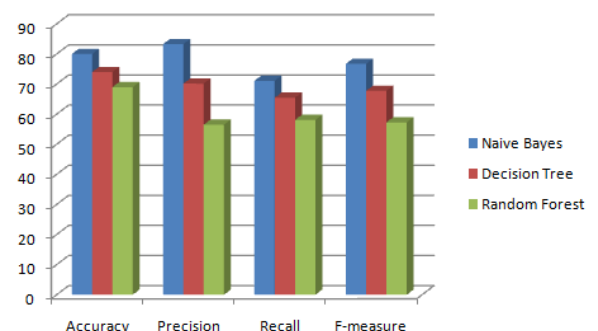


Figure 2. Bar graph representing Table II

VII. RESULTS

The experiments carried out in Rapidminer tool over the datasets collected have shown that Naive Bayes works best and provides us the highest accuracy although Decision Tree is equally competitive but fails to produce desired results. Random Forest is an effective algorithm but as our dataset isn't huge therefore this algorithm isn't appropriate.

VIII. CONCLUSION,LIMITATIONS AND FUTURE SCOPE

The system proposes a whole new concept involving different major technologies like image processing and data mining and many more. Involvement of sensors for the factors like doors ,

fan speed etc. can contribute highly towards even efficient working of the proposed system.

The proposed system has a wide future scope. The system once enabled and has been working for a good time will be synchronized with every individual personal panel with their data stored on cloud using a Mobile application. So when the user moves to a new place let it be a hotel or new living place the Air Conditioner can synchronize with the data of user stored on cloud and the Air Conditioner doesn't have to be trained for the efficient usage

There is a huge scope of improvement in the system. It does involve lot of complexities as involvement of multiple sensors and technologies makes it a little complex but altogether is a huge step towards innovation.

IX. ACKNOWLEDGMENT

I would like to thank our guide Prof. Vanmathi C., Assistant Professor (Selection Grade) who provided us insight and expertise that greatly assisted the research. I would also like to thank Dinakaran M. (Head, B.Tech (IT), SITE, VIT University) for all the help with our research that greatly improved manuscript. Further I would also like to show my gratitude to Prof. G Jayaraman (Dean, SITE, and VIT University) for sharing his pearls of wisdom with us during the course of this research . This research was supported/partially supported by VIT University, Vellore.

X. REFERENCES

- [1] Thein Thein Soe, Kyaw Soe Lwin, Zaw Min Naing, Aung Soe Khaing , "Design and Development of Microcontroller Based Air Conditioning System", IJSETR, May-2014
- [2] Hanife Apaydın Özkan; Aydın Aybar, "A smart AirConditioner in Smart Home" ,EEEIC, 2016
- [3] Joseph Howse, "Training detectors and recognizers in Python and OpenCV", IEEE ISMAR , 2014
- [4] Amit Gupte, Sourabh Joshi, Pratik Gadgul, Akshay Kadam," Comparative Study of Classification Algorithms used in Sentiment Analysis", International Journal of Computer Science and Information Technologies, Vol. 5 (5) , 2014
- [5] Shahrukh Teli, Prashasti Kanikar " A Survey on Decision Tree Based Approaches in Data Mining", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 5, Issue 4, 2015
- [6] Eesha Goel , Er. Abhilasha , " Random Forest: A Review", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 7, Issue 1, January 2017.
- [7] Ashmeet Singh and R Sathyaraj. "A Comparison between Classification Algorithms On Different Datasets Methodologies Using Rapidminer". International Journal of Advanced Research in Computer and Communication Engineering 5.5 (2016):