Information Abstraction from IoT Streaming Greenhouse Data

V R Srividhya
Research Scholar
Dept. of CSE, BMSCE
Bengaluru, India
srividhya.sridhar@gmail.com

Niketh S A
Post Graduate Student
Dept. of CSE, BMSCE
Bengaluru, India
nikethsa8@gmail.com

Dr. Kayarvizhy N
Associate Professor
Dept. of CSE, BMSCE)
Bengaluru, India
kayarvizhy@gmail.com

***Abstract*—Internet of Things (IoT) is a platform which gives the billions of computing devices, sensory devices that constantly produce and exchange the huge amounts of data over the network. Several interesting applications, like Greenhouse, involving extracting the higher level information from the raw data and representing it in human readable format, exists. An effective mechanism is required to process streaming data and inferring it, to get insight about the data and get some actionable information from processing and measurements. The aim is to represent the data from device point of view to user-centric point of view using Data analytics and Machine learning mechanism. Raw sensor data is created, based on the sensor threshold values. Parameters considered are temperature, humidity, co2 concentration, luminosity, radiation, soil moisture and soil temperature. Numerical values are converted to the strings to create higher level abstraction by applying set of rules. Abstracted data is cleaned by some cleaning processes. Then Latent Dirichlet Allocation (LDA),a topic extraction model, is applied to extract the hidden correlation among the data**

Keywords— Internet of Things, Greenhouse, Latent Dirichlet Allocation

I. Introduction

A. Greenhouse

Different crops have different seasons in which they grow and Different regions in the earth has extreme climatic conditions or different environmental status which makes it uncomfortable to grow the crops with specific requirement. To protect the plants from the several environmental conditions like ultraviolet radiation, wind, hailstorm, insect and pest attacks, a method called Greenhouse were developed as solutions to overcome the problem, making it possible to grow all vegetables and fruits in all the seasons throughout the year.

Greenhouse are structures made of transparent materials where the user use to grow the plants with the required climatic conditions. Few advantages are that yield can be 10-12 times more, off-season production, protected from insects and disease from them, less requirement of water[1]. The greenhouse can maintain the favourable environmental conditions/parameters like temperature, humidity, solar radiation etc., for the respected crops. The parameters inside the greenhouse sensed and measured using different sensors. In-house environmental status can be monitored from remote areas and in case of Automated Greenhouse, it can take care of itself by monitoring the in-house environmental status according the specified conditions.

B. Internet of Things

Internet of Things is a technology which gives platform to inter-related computing devices, sensory devices, digital machines, mechanical devices etc., to produce and exchange the tremendous volume of data via the Internet. The technology has increased the opportunities in innovation

The applications of IoT is not limited to a particular stream and are distributed in nature. Many of the IoT applications requires real-time-processing as they produce real time IoT data. These data are only limited to a specific domain or unused later for further purpose[2].

There are many streams of IoT where data is being generated daily like: Traffic Information [3], Parking spaces[4]. Smart Homes [5]. Infrastructure Application [6],Environmental Monitoring [7]. Metropolitan Scale Deployment [8].

C. Data abstraction & Semantic Representation

Data abstraction is the process of reducing multiple data to simplified version without altering the meaning. Whereas in Semantic representation, the data is represented in a meaningful form which can be understood by human.

Data abstraction methods can be obtained by Pre-processing Techniques like Signal Preprocessing [9],Piecewise Aggregation Approximation (PAA) [10], Symbolic Aggregate Approximation (SAX) that is applied to the PAA output[11],etc

Semantic Reasoning methods like Reasoning in which data is passed through the rule engine can be used. Based on the set of rule, operation on the data is carried out[12].Clustering process to group the same type of objects[13]. Frequent Item set Mining which involves the method to find the values that co-occur frequently[14] can also be used.

The data produced by these IoT devices remain unused for long and no information is gained from these data. In-order to get the information and use it for further future processing, abstraction is used. Depending on the frequency, the data is abstracted to string form from the numerical form.

D. Motivation

The scope to the IoT is growing day by day because of its feature Machine to Machine Communication. Billions of devices are connected to the internet and the numbers are increasing day by day. According to Gartner's Forecast IoT devices produce 2.5 quintillion bytes of data. The data production will increase as there will be 25 billion devices

978-1-7281-0630-4/18/$31.00 ©2019 IEEE

connected to the internet by 2020[15]. The connected devices produce a huge amount of data. The applications of these are distributed in nature and there is necessity of real-time processing of data. It requires novel methods capable of interpreting patterns and make accurate decisions to the current situations with minimum latency. Real time data are needed to be processed with a novel approach and represented in the form of high level knowledge. The term Real time data analysis notifies that the data has to be analyzed before it is stored [16].

Because of the large number of specific technologies, the valuable data or the information from the sensors remains unused or limited to specific application domains [2]

By the techniques like Data Mining and Knowledge Discovery in Database, the generated data can be converted to knowledge. The techniques provide solutions for finding the information hidden in the IoT data, that can be used to improve the performance of the system and quality of service. The results show that data analysis algorithms can be used to make IoT more intelligent, thus providing smarter services.

Efficient stream reasoning mechanisms are required to interpret the meaning of events in a context-aware fashion and share such meaning across applications. A sophisticated mechanism is required to make these data available to the end user in useful way.

E. Overview of technologies and methods used

Internet of Things technology, produce and exchange the data via the Internet. The technology has increased the opportunities in innovation and there are huge devices connected in internet resulting in producing tremendous volume of data.

Greenhouses are structures which are made of transparent materials where the user use to grow the plants with the required climatic conditions. The greenhouse can maintain the favourable environmental conditions/parameters like temperature, humidity, solar radiation etc., for the respected crops. The parameters inside the greenhouse sensed and measured using different sensors. Topic modelling is used to discover the topics and hidden semantic structure in a collection of documents.

II. RELATED WORK

There are various works done in the platform of Automation of Greenhouse, Data Abstraction Methods. D D Nangare et al.[17] constructed bamboo framed structure with green shade nets with different heights and different shades to determine the effect on the crops.

Several constraints were found while growing the plants in the open field like pest attack, that decreases the yield. And also due to the climatic conditions like extreme low/high temperature in the semi-arid region there was inferior quality of yield. Hence, there was a need of growing crops in the covered structure. There was significant increase in yield and quality of the fruits, due to conducive growth conditions in a protected environment, when they were grown in Greenhouse.

P. S. Asolkar et. Al.[18] monitored the in-house parameters Soil moisture, temperature, Light intensity, Humidity, CO2 Gas of the greenhouse with the help of At-mega32 microcontroller as central processing unit and GSM Communication with the sensors LM35, SY-HS220, Two copper probes, LDR and MQ5 gas sensor respectively. By implementing the monitoring system in Greenhouses, they can provide automatic controlling techniques which are effective for improvement in crop production compared to old growing methods resulting in reduction of human efforts required for growing crops in open field. They have presented Global system for mobile communication (GSM) system for controlling and monitoring of greenhouses. The method was found to be very effective as it allows parameters of greenhouse like moisture, humidity, temperature etc., to be controlled from a remote area to the desired location. They found that productivity of the crops was much better when the environmental conditions were controlled.

Payam Barnaghi et. al.[19] in their work presented the challenges in finding and extract actionable knowledge from raw data of Web of Things. The data are collected from the sources like Smartphones, Sensor Devices, Social medias, wearable devices, etc. which are represented in numerical form or the symbolic descriptions of occurrences in the physical world. The quality, quantity, validity of the data and trust of the data are the challenges in the big data especially in the situation where data is made available to multiple users as there will data related to environment, events and people and hence privacy and security are key concerns. When multiple parties access and process the data, dealing with these issues are difficult. The large number of data coming from the sensory devices requires more bandwidth and reliable Quality of Service solutions. Hence, methods like preprocessing which can be aggregation, summarization, abstraction can help to reduce the size of the data at source level. Knowledge discovery requires set of mechanisms. As different users provide variety of data, there can be inconsistency in the data like errors in reading the data. The data are dynamic in nature and is difficult to discover the knowledge from it.

Frieder Ganz, Daniel Puschmann, et al.[20] proposed techniques for Information Processing and information Abstraction in IoT. They have presented a number of techniques for processing the data and to transform it to higher level abstraction from raw unstructured-data. As the sensor devices generate numerous amount of data, a particular technique cannot be applied for the dynamic data to extract the information. They have provided a survey of the requirements and solutions to extract meaningful and higher level information from raw sensor data. Several methods have been presented from semantic web, machine-learning, pattern recognition and data mining to abstract the information. Pre-processing methods like signal pre-processing which cuts the certain parts of the signal before and after certain frequency value. Mathematical/Statistical Pre-processing methods like min-max value and mean-median values, the values are filtered out. Dimensionality Reduction methods like Discrete Fourier Transformation, Piecewise Aggregation Approximation and Symbolic Aggregate Approximation are used to reduce the length and size. Feature Extraction methods such as Clustering, Semantic Reasoning & Representation: The metadata, data and its related context information are represented in a linked graph model. The technical solutions they provided were to use the tools Rapid Miner, WEKA, SAMOA and Orange.

Altti Ilari Maarala et al.[12] presented method for Semantic Data provisioning and Reasoning for the Internet of Things, the goal was to represent the data produced by the IoT nodes to the machine understandable manner, as it facilitates interoperability with various systems and applications and to present the data such that its meaning can be explained and shared efficiently. Real time requirements of IoT data heterogeneity and dynamic nature are challenges for applying the technologies like reasoning and interoperability. They have specified the approaches which are able to deliver semantic data from IoT nodes to distributed reasoning engines. Reasoning over such data are performed. They collected Global Positioning System (GPS) data from taxi drivers. They detected several events from the observations like turns, traffic jams, stopping for a long time, speeding, sudden acceleration and strong acceleration, deceleration etc. The experiments were carried out with both distributed reasoning nodes and single reasoning node. The operations were carried out by collecting a set of messages from IoT nodes. Reasoning is performed over the data by reasoning engine. RDF database stores the reasoned knowledge.

Daniel Puschmann et al.[21] introduced a novel approach for recognizing the relation between the heterogeneous type of data. The aim was to take out the information from real-world data and to find the correlation between them. To reduce the dimensionality of streaming data, they have used Piecewise aggregate approximation (PAA). The data sets used were of traffic and Weather data. The traffic data streams and weather data streams were collected from the city of Aarhus. PAA takes equal sized frames as parameter and calculates the averages of the frames in the original data. They are represented by their mean value. Symbolic Aggregate Approximation determines the equivalent symbol for the coefficient. It is applied to the PAA output. The obtained information from the pattern along with mathematical values are translated to higher level abstraction. Virtual documents are created to group the higher level abstractions from different sources within a certain time frame. Then to find the correlation between the latent Dirichlet allocation is used.

DM Blei et. al. [22] described generative probabilistic model which is called latent Dirichlet allocation(LDA) for discrete data. The goal of the work was to find the short representations of the discrete data which gives well processing and essential statistical relationship that helps for the tasks like summarization, classification, similarity novelty detection, etc. They stated that LDA can be viewed as dimensionality reduction technique. The proposed method reduces all documents which are in paragraphs to vector of real numbers. The real numbers represents the counts in ratio.

III. EXISTING SYSTEM

Methods like Adaptive Clustering for IoT data streams[13] clusters the data from different sources and use probability distribution techniques to plot the graph. A new category is found by the turning points of the graphs. Reasoning nodes[12] perform reasoning over the data and stores the data according to the rule sets provided. To find the abstraction of the raw data, methods like Piecewise aggregate approximation are used to split the input data into window segments[21] and then Symbolic aggregate approximation to determine the equivalent symbol for the coefficient. There are methods that extract the abstraction from the existing data which are stored in the database, like Cloud based centralized architecture[23] which stores the raw data from the sensors. The data will be huge and these architectures will cause delay for providing the required service and wastes many resources.

IV. PROPOSED SYSTEM

The proposed system takes the raw data from the sensor in real time. Based on the set of rules, the data is divided accordingly and given abstraction. The abstracted form of data are then cleansed by applying some of the cleansing methods. Then the data is passed to LDA algorithm, a topic modelling method is applied to automatically discover the topics in the data and to know the correlation among the parameters of the Greenhouse.

V. METHODOLOGY

This section demonstrates the various techniques applied for proposed method.

A. *Overview of the Workflow*

Data Creation: In order to process the live streaming data, Greenhouse has to set up where multiple parameters are sensed and given output. Here, instead of setting a greenhouse,

Virtual sensor programs are created which acts like greenhouse sensors.

Data Creation: In order to process the live streaming data, Greenhouse has to set up where multiple parameters are sensed and given output. Here, instead of setting a greenhouse, Virtual sensor programs are created which acts like greenhouse sensors.

1. Data Abstraction: The raw data obtained from the sensors are translated to higher level abstraction based on set of rules.

2. Data Cleansing: The abstracted data has to be cleaned for further processing.

3. Latent Dirichlet Allocation (LDA): We train and incrementally update an LDA model based on the abstracted documents and virtual documents that are created manually to identify the correlation with the parameters..

Fig.1 shows the workflow of the model. Virtual sensor programs are created which acts like the sensors of the greenhouse producing real time data. While generating the data, randomness is introduced so that there will be some outlier. The generated data are stored in separate files for different sensors. With a specific time interval, the data is read from the file and is converted to the respective sensor parameters unit. There are 2 types of sensors namely Analog sensor and Digital sensors. A 10 bit analog sensor gives data in the range from 0 to 1023 which is then converted by the Analog to Digital converter to normal form of the data as per the parameter. Whereas the digital sensor gives the data in normal form. The value is converted to string format with the help of threshold range of the respective parameters. The semantic or abstract form of the value is stored in a separate file. The data has to be cleaned for further processing. There are 3 techniques applied in preprocessing which are Tokenizing, Removing stop words and stemming the words. In order to obtain the correlation among the data, LDA is applied.

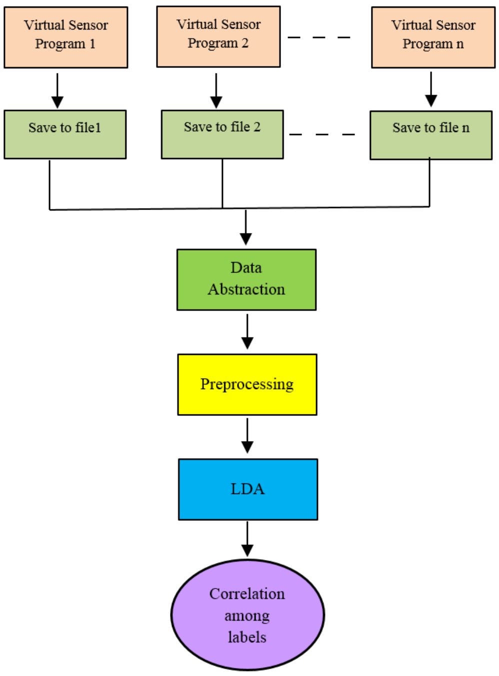

Fig 1: Workflow of the Model

B. *Data Creation*

The virtual sensors are based on the various IoT sensors used in greenhouse. P. S. Asolkar used LM35 temperature sensor and MQ5 gas sensor for greenhouse monitoring [18]. K Rangan developed an embedded system approach to monitor greenhouse [24]. Lijun Liu in their greenhouse used BH1750 sensor for measuring luminance[25]. Teemu Ahonen in their greenhouse monitoring system used SHT75 sensor for measuring temperature and humidity, TAOS TSL2561 luminosity sensor which converts light intensity to voltage[26].

Maro Tamaki used SEN 13322 sensor to measure the solar radiation inside the greenhouse [27].Based on these sensors, the virtual sensor program creates the data within a specific range. Digital sensors directly senses the environment and gives the output. Whereas analog sensors, based on the bit resolution, range of sensing, etc., they sense the environment and give the output. 10 bit sensor creates values within the range 0 to 1023. According to the type of sensor, analog or digital, different programs are written which generates the data with an instance of 100 milliseconds. The data are stored in separate files in the database.

Table 1: List of Sensors taken to consideration

Name	Operating Range	Interface	Accuracy
LM35	-55°C to +155°C	Analog	0.5° C
SHT75	0 - 100% RH	Digital	±1.8% RH
MQ5	200-10000 ppm	Analog	-
BH1750	1 – 65535 lx	Digital	-
TGS4161	350 to 10000 ppm	Analog	±20% ppm
SEN 13322	0% - 100%	Digital	±0.5%

C. *Data Abstraction*

The data is read from the every file in 3 minutes interval each. To get the data in the respective unit form, there has to be equivalent converter to convert the analog signal to respective parameter unit value. Different sensors have got different analog to respective unit converter methods. According to the parameter and the unit conversion, the data is converted to meaningful unit form. Based on the rules provided in table 2, the statistical data is converted to the equivalent string format.

Table 2: Rule set for Data Abstraction

Parameters	**Low**	**Normal**	**High**
CO_2	<700 ppm	700 – 1000 ppm	>2000ppm m
Inside Humidity	<40% RH	40 – 45 % RH	>45% RH
Outside Humidity	<50 % RH	50 – 55 % RH	>55 % RH
Luminosity	<1000 lux	1000 - 2000 lux	>2000 lux
Radiation	<977 W/m^2	977 – 1000 W/m^2	<1000 W/m^2
Soil Moisture	<20% RH	20 – 30 % RH	>30 % RH
Soil Temperature	<20 °C	20 – 25 °C	>25 °C
Temperature Inside	<20 °C	20 – 25 °C	>25 °C
Temperature Outside	<25 °C	25 – 30 °C	>30 °C

D. Pre processing

Data cleansing method contains three steps. Tokenization, Stop words and stemming.

Tokenizer: Tokenization is the act of breaking down the set of strings into pieces. It may contain words, symbols, phrases, The abstracted data is given as the input to this model. It breaks the strings to form words.

Stop Words: Here, Stop words remove the same words which appear twice, as there is no common words of natural language in the doc.

Stemming: Stemming process removes the derivationally related form and inflectional form of a common rooted word.

E. Latent Dirichlet Allocation

Latent Dirichlet Allocation is general probabilistic model which is used for extraction of topics in a text document. LDA is trained over the set of documents which are manually created. Hidden relationship and correlation between the parameters can be obtained from LDA

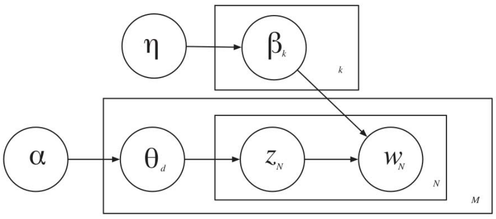

Fig 2: Plate model Representation of LDA

M is the total document set, N is the collection of the words represented by vector WN, w is the particular word in the document, z_{ij} is the topic that is most likely to have generated w_{ij}, k represents the number of topics which is fixed at the initial, θ is topic distribution, α is per document topic distribution, β is the multinomial distribution of words which represents the topics, j is the word count and i is document count.

VI. RESULTS

The parameters which are monitored inside the greenhouse are considered namely temperature, humidity, co2 concentration, luminosity, radiation, soil moisture and soil temperature. In order to get the real-time data, virtual sensor programs are written which acts as sensors in the greenhouse and creates the data according to the conditions specified in the table 2 . There can be situation where greenhouse sensors gives some error values such as out of range, inconsistent value etc. The analog sensors gives the output in the range of 0-1024. In some cases, it may go beyond this range or it might give unrealistic values. For example: The normal room temperature exists within 20° C to 35° C. If the output of the sensor gives the temperature is beyond this value, then we can assume that there is some error with the sensor values. So, the virtual sensor programs are introduced with an outlier concept where 80% of the values are within the normal range and 20% of the values are out of this range. The data creation is continuous and can be only stopped by breaking it or with any external command. The raw data created by different virtual sensor programs are stored in different files each.

The data are stored in the files with the numerical range of 0-1023 if the sensor is analog or else the respective unit sensor output value. The data is read in the interval of 3 minutes. In the case of analog sensor, the raw sensor values are converted to respective parameter unit according to the sensor conversion formula and checked with the respective conditions provided. Then the value is assigned with the equivalent text phrases and saved in different files. For example: if the analog value for the parameter "inside temperature" is 163, the equivalent temperature value the greenhouse which is 30° C is given converted to text phrase "inside_temperature_is_normal".

Then the text file is read by the LDA module which calculates the probability of occurrences of every phrases. The correlation among the parameters of the greenhouse is found by it. LDA reads 10 text files at a time in which each individual text file contains 10 samples of abstracted data. The LDA algorithm was tuned to produce 1 topics on each run with various data sets.

Results obtained from LDA with changes in the parameter.

```
[(0, '0.111*"co2_is_normal"
+0.111*"humidity_inside_is_normal"
+0.111*"humidity_outside_is_normal"
+0.111*"luminosity_is_normal"
+0.111*"radiation_is_normal"
+0.111*"soil_moisture_is_normal"
+0.111*"soil_temperature_is_normal"
+0.111*"temperature_inside_is_normal"
+0.111*"temperature_outside_is_normal"')]
```

Fig 3: LDA results when all the parameters are kept normal

10 files containing 10 samples of abstracted text phrases each are passed to LDA. At first all the parameters are kept normal and checked with the results. Fig. 3 shows the result in which probability of every parameter were kept to normal. In this test, the probability of every parameter was obtained same

```
[(0, '0.111*"humidity_inside_is_normal"
+0.111*"humidity_outside_is_normal"
+0.111*"soil_moisture_is_normal"
+0.111*"soil_temperature_is_normal"
+0.067*"co2_is_high"
+0.067*"luminosity_is_high"
+0.067*"radiation_is_high"
+0.067*"temperature_inside_is_high"
+0.067*"temperature_outside_is_high"
+0.045*"co2_is_normal"
+0.045*"luminosity_is_normal"
+0.045*"radiation_is_normal"
+0.045*"temperature_inside_is_normal"
+0.045*"temperature_outside_is_normal"')]
```

Fig 4: LDA results with variable parameters

In second test, the parameters like inside temperature, outside temperature, luminosity, radiation and co2 of some files are changed from normal to high. The result obtained are shown in Fig. 4. It can be noted that inside temperature is high when the outside temperature is high which has got the same probability and are related. Radiation is high and luminosity is high has got the same probability and are related to each other and co2 are correlated to each other. In the third test, inside humidity and outside humidity are changed from normal to low in some of the files and obtained same probability which is shown in Fig. 5. In the fourth test, inside temperature, outside temperature, co2, luminosity and radiation are changed from normal to high. Inside humidity and outside humidity are changed from normal to low and noted that the parameters like inside temperature, co2, radiation and luminosity changes when the outside temperature is changed. Inside humidity is affected by the outside humidity. The result obtained is shown in Fig. 6.

```
[(0, '0.111*"co2_is_normal"
+0.111*"luminosity_is_normal"
+0.111*"radiation_is_normal"
+0.111*"soil_moisture_is_normal"
+0.111*"soil_temperature_is_normal"
+0.111*"temperature_inside_is_normal"
+0.111*"temperature_outside_is_normal"
+0.078*"humidity_inside_is_low"
+0.078*"humidity_outside_is_low"
+0.034*"humidity_inside_is_normal"
+0.034*"humidity_outside_is_normal"')]
```

Fig 5: LDA results with variable parameters

It can be noted from the above results that if the parameters are related to each other, the probability associated with the parameters changes and if they are highly related, then probability are very much closer.

The following correlations were found out.

1) Outside temperature is high is highly correlated with inside temperature being high. This is also related to both inside and outside humidity as high and other factors like radiation and CO2 both being high.

2) Humidity inside being low is related to humidity outside being low, normal luminosity and normal soil moisture and CO2 being normal.

3) Also, temperature is normal, co2 and humidity normal is related to all other attributes being normal.

```
[(0, '0.110*"soil_moisture_is_normal"
+0.110*"soil_temperature_is_normal"
+0.078*"co2_is_high"
+0.078*"humidity_inside_is_low"
+0.078*"humidity_outside_is_low"
+0.078*"luminosity_is_high"
+0.078*"radiation_is_high"
+0.078*"temperature_inside_is_high"
+0.078*"temperature_outside_is_high"
+0.034*"co2_is_normal"
+0.034*"humidity_inside_is_normal"
+0.034*"humidity_outside_is_normal"
+0.034*"luminosity_is_normal"
+0.034*"radiation_is_normal"
+0.034*"temperature_inside_is_normal"
+0.034*"temperature_outside_is_normal"')]
```

Fig 6: LDA results with variable parameters

VII. CONCLUSION

There are numerous number of devices connected to the internet which are regularly produce and exchanging huge amount of data and either limited to specific domain or unused later. There is an effective methodology required to process the real time data and extract the information in it.

We have implemented a novel approach to extract the Information hidden in the raw IoT data from Greenhouse and to find the correlation among the data. The process involves collecting the raw data from the sensory devices in a particular frequency and representing the equivalent text form for respective sensors data. The semantic form of the data is further preprocessed with the techniques tokenization, removal of stop words and stemming. Further the abstracted data is given to LDA, a topic modelling method to find the correlation among different parameters of the Greenhouse. Hence, LDA can be used to find the correlation among the greenhouse parameters

ACKNOWLEDGEMENT

We are grateful to BMS College of Engineering for having provided us with the facilities needed for the successful completion of this Survey paper. The work reported in this paper is supported by the college through the Technical Education Quality Improvement Programme [TEQIP-III] of the MHRD, Government of India.

REFERENCES

[1] S. Pandey and A. Pandey, “GREENHOUSE TECHNOLOGY,” *Int. J. Res. - GRANTHAALAYAH*, vol. 3, pp. 1–3, 2015.

[2] R. Tönjes *et al.*, “Real time iot stream processing and large-scale data analytics for smart city applications,” in *poster session, European Conference on Networks and Communications*, 2014.

[3] M. Lakshminarasimhan, “Advanced Traffic Management System Using Internet of Things.”

[4] A. Khanna and R. Anand, “IoT based smart parking system,” in *Internet of Things and Applications (IOTA), International Conference on*, 2016, pp. 266–270.

[5] M. Al-Kuwari, A. Ramadan, Y. Ismael, L. Al-Sughair, A. Gastli, and M. Benammar, “Smart-home automation using IoT-based sensing and monitoring platform,” in *Compatibility, Power Electronics and Power Engineering (CPE-POWERENG), 2018 IEEE 12th International Conference on*, 2018, pp. 1–6.

[6] J. Gubbi, R. Buyya, S. Marusic, and M. Palaniswami, “Internet of Things (IoT): A vision, architectural elements, and future directions,” *Future Gener. Comput. Syst.*, vol. 29, no. 7, pp. 1645–1660, 2013.

[7] S. Li, H. Wang, T. Xu, and G. Zhou, “Application study on internet of things in environment protection field,” in *Informatics in Control, Automation and Robotics*, Springer, 2011, pp. 99–106.

[8] A. Zanella, N. Bui, A. Castellani, L. Vangelista, and M. Zorzi, “Internet of things for smart cities,” *IEEE Internet Things J.*, vol. 1, no. 1, pp. 22–32, 2014.

[9] A. Arora *et al.*, “Exscal: Elements of an extreme scale wireless sensor network,” in *Embedded and Real-Time Computing Systems and Applications, 2005. Proceedings. 11th IEEE International Conference on*, 2005, pp. 102–108.

[10] E. J. Keogh and M. J. Pazzani, “Scaling up dynamic time warping for datamining applications,” in *Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining*, 2000, pp. 285–289.

[11] J. Lin, E. Keogh, S. Lonardi, and B. Chiu, “A symbolic representation of time series, with implications for streaming algorithms,” in *Proceedings of the 8th ACM SIGMOD workshop on Research issues in data mining and knowledge discovery*, 2003, pp. 2–11.

[12] A. I. Maarala, X. Su, and J. Riekki, “Semantic data provisioning and reasoning for the internet of things,” in *Internet of*

Things (IOT), 2014 International Conference on the, 2014, pp. 67–72.
[13] D. Puschmann, P. Barnaghi, and R. Tafazolli, "Adaptive clustering for dynamic IoT data streams," *IEEE Internet Things J.*, vol. 4, no. 1, pp. 64–74, 2017.
[14] Y. Qin, Q. Z. Sheng, N. J. Falkner, S. Dustdar, H. Wang, and A. V. Vasilakos, "When things matter: A survey on data-centric internet of things," *J. Netw. Comput. Appl.*, vol. 64, pp. 137–153, 2016.
[15] D. Evans, "The internet of things: How the next evolution of the internet is changing everything," *CISCO White Pap.*, vol. 1, no. 2011, pp. 1–11, 2011.
[16] A. Akbar, F. Carrez, K. Moessner, J. Sancho, and J. Rico, "Context-aware stream processing for distributed IoT applications," in *Internet of Things (WF-IoT), 2015 IEEE 2nd World Forum on*, 2015, pp. 663–668.
[17] D. Nangare, J. Singh, V. Meena, B. Bhushan, and P. Bhatnagar, "Effect of green shade nets on yield and quality of tomato (Lycopersicon esculentum Mill) in semi-arid region of Punjab," *Asian J. Adv. Basic Appl. Sci.*, vol. 1, no. 1, pp. 1–8, 2015.
[18] P. Asolkar and U. Bhadade, "Analyzing and predicting the green house parameters of crops," *Int. J. Comput. Appl.*, vol. 95, no. 15, 2014.
[19] P. Barnaghi, A. Sheth, and C. Henson, "From data to actionable knowledge: Big data challenges in the web of things [Guest Editors' Introduction]," *IEEE Intell. Syst.*, vol. 28, no. 6, pp. 6–11, 2013.
[20] F. Ganz, D. Puschmann, P. Barnaghi, and F. Carrez, "A practical evaluation of information processing and abstraction techniques for the internet of things," *IEEE Internet Things J.*, vol. 2, no. 4, pp. 340–354, 2015.
[21] D. Puschmann, P. Barnaghi, and R. Tafazolli, "Using LDA to Uncover the Underlying Structures and Relations in Smart City Data Streams," *IEEE Syst. J.*, vol. 12, no. 2, pp. 1755–1766, 2018.
[22] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation," *J. Mach. Learn. Res.*, vol. 3, no. Jan, pp. 993–1022, 2003.
[23] R. Lea and M. Blackstock, "City Hub: a cloud based IoT platform for Smart Cities," 2014.
[24] K. Rangan and T. Vigneswaran, "An embedded systems approach to monitor green house," in *Recent Advances in Space Technology Services and Climate Change (RSTSCC), 2010*, 2010, pp. 61–65.
[25] L. Liu and Y. Zhang, "Design of greenhouse environment monitoring system based on Wireless Sensor Network," in *Control, Automation and Robotics (ICCAR), 2017 3rd International Conference on*, 2017, pp. 463–466.
[26] T. Ahonen, R. Virrankoski, and M. Elmusrati, "Greenhouse monitoring with wireless sensor network," in *Mechtronic and Embedded Systems and Applications, 2008. MESA 2008. IEEE/ASME International Conference on*, 2008, pp. 403–408.
[27] M. Tamaki *et al.*, "Daily CO2 changes monitored by an environment sensor unit in a greenhouse with bitter gourd," in *Intelligent Informatics and Biomedical Sciences (ICIIBMS), 2015 International Conference on*, 2015, pp. 109–112.