International Journal of Pure and Applied Mathematics
Volume 118 No. 18 2018, 3151-3157
ISSN: 1311-8080 (printed version); ISSN: 1314-3395 (on-line version)
url: http://www.ijpam.eu
Special Issue

ijpam.eu

The Analysis of Fruit Ripening Level from Diffusion of Gases by using Data Mining Techniques

[1]Udomsak Paeanpairoj, [2]Wutthipong Chinnasri

[1]College of Information and Communication Technology, Rangsit University.
udomsakpaeanpairoj@gmail.com
[2]College of Information and Communication Technology, Rangsit University.
wutthipong.c@rsu.ac.th

Abstract

The purpose of this research aims to analyze fruit ripening level from diffusion of gas by data mining techniques. Due to the fact that Ethylene is associated with fruit ripening level. A tool, Arduino board connected with MQ3, MQ6 and MQ8 detection sensors, has been developed for the assessment. These sensors are substituted for Ethylene one which is quite expensive. After collecting 160 data by this tool, data mining procedures are mainly used for data clustering and classification. The experimental result demonstrates that K-Means clustering algorithm classify durian ripening level into 4 types; 1) unripe, 2) pre-ripe, 3) ripe, 4) overripe. When it comes to classification, various types of algorithm, for instance, Decision Tree, K Nearest Neighbors (KNN), Neural Network, Naive Bayes and Support Vector Machine (SVM), have been compared by accuracy rate of their performance through cross validation procedure. It has been found that Support Vector Machine and Neural Network algorithm are similarly the most accurate. However, the accuracy of all algorithm in this experimental closely to 100%, these algorithms can be further developed to analyze fruit ripening level from diffusion of gas analysis.

Keywords: Ethylene; Ripening; Gas Sensor; Data Mining; Clustering; Classification

1. Introduction

Thailand locates in tropical climates area; therefore, there are various different kinds of fruit, which can be categorized in many ways according to chosen classification factors. This research emphasizes on the analysis of fruit ripening level, which it is divided into 2 types by fruit respiration rate [1]. Fruit which respiration rates change after full grown harvest, so called climacteric fruit, such as banana, mango, papaya and durian. On the other way, one which respiration rates is unchanged or has minimal change after full grown harvest, non-climacteric fruit, such as orange, pineapple, lemon and rambutan.

Problems related to fruit collection, transportation and trade, always come along with climacteric fruit. Whenever farmer, entrepreneur or distributor is fail to manage in harvesting and transporting fruit so it is not to meet consumers' requirement. It also wastes cost and time to store additional goods. Several researches study in ripening level of fruit control during harvesting and transporting procedure [2], [3], [4]. However, this research indicates that climacteric fruit and non-scientific identificial color change of fruit are able to be classified by its ripening level. Farmers and distributors is still able to specify how much of fruit ripening level though their experiences and its color observation. Even though, it is difficult to observe non-climacteric fruit, such as durian and jackfruit, which its color is unchanged or has minimal change. Farmers and distributors must have much more experiences on observation. This study conceives to analyze

fruit ripening level from diffusion of gases by using data mining techniques. Durian is used for the experiment as non-climacteric fruit samples which cannot be apparently notice its ripening level. As its fame known as "King of fruit", durian bring Thailand reputation to around the world [5]. Each consumer always picks different ripening level of durian. As per further study on other research, ripening level of durian can be classified into 3 levels; unripe, ripe and overripe [6]. In fact, portion of Thai people consume different level of ripening which is pre-ripe durian. Refer to the research which study on papaya's physical change, it is indicated that unripe level appeared to be apparently changed on its skin color to be yellow around 10 - 20% [7]. As well as other kinds of fruit, their skins turn from green to partial yellow. Due to these facts, this research chose to classify ripening level of durian into 4 levels; unripe, pre-ripe, ripe and overripe. Unripe one can be fried as crispy durian chips. Overripe one would go through a process to be preserved durians. On the other way, most of pre-ripe and ripe ones are distributed to consumers without any processing. Vendors mostly recognize durians ripening level by using wooden stick to tap it or observing its thorn color. Nonetheless, this method cannot be accurate so that they have to chop apart of it and let consumer press their finger on its pulp to re-examine whether its ripening level meet their requirements. Hence, these procedures are still inaccurate and unreliable for consumers. Furthermore, durian need to be partly chopped so that they can examine its ripening level. After studying climacteric fruit ripening process, it can be implied that continuously ripening of the fruit or leaving it in properly condition will let it have numerous change on flavor, color and savor. If fruit ripening process is considered through scientific change, it involves with ethylene. Ethylene appears to be low diffusion in Unripe durians, but high in ripe ones. Assuming that it is significant factor of fruit ripening process [8], [9]. In case that ethylene can be measured so fruit ripening level can be done in the same way. The research purposes to analyze fruit ripening level from diffusion of gases by using data mining techniques to synthesize and classify data. This study would help farmers, entrepreneurs and distributors to devise plan to spend less time for delivery and cost less money on stocking, including to efficiently select durian and measure its ripening level to sell consumers.

2. The Proposed Method for Analysis of Fruit Ripening Level and Implementation

Arduino, microcontroller board, can be used as diffusion of gases detection for benefit on analysis. It can be connected with numerous kinds of sensor, including gas sensor. Providing that ethylene gas sensor connect with Arduino board, then it would detect ethylene gas to collect data for further analysis. Due to the fact that this sensor type is hard to find in general store in Thailand and price is higher than other ones as hydrogen gas sensor or alcohol gas sensor, so that this research detect gas by using MQ3, MQ6 and MQ8 gas sensors instead. As chemical formula analysis of ethylene, C_2H_4 [10] refer to 2 carbon atoms and 4 hydrogen atoms. MQ3 gas sensor is used for detection of alcohol and ethanal (Ethanol; C_2H_5OH) which its elements are ethylene (C_2H_4) and water (H_2O) [11]. Then, MQ6 gas sensor is used for detecting hydrogen gas. Nevertheless, it is no sensor tool to detect carbon gas (C) so this experiment provides MQ8 gas detection sensor for LPG, consisting of propane and butane, which carbon is main element of it [12]. During Arduino board is processing, MQ Gas Sensors will transfer Norazlan amount of gas volume data to collect and display the result before it is evaluated by using data mining techniques. On the experiment, plastic box has developed for fruit ripening level assessment which finish installing Arduino board and relevant gas sensors, along with ventilation fan to pull out durian's diffusion gases to be processed in figure 1.

Furnished plastic box in figure 1 is used for measuring alcohol, hydrogen and carbon gases level. When durian is put into this tool, amount of diffusion gases data will be displayed in diagram and figure. The researchers start collecting data when the number in diagram go to the highest and saturation point and then gather amount of durian's gas data.

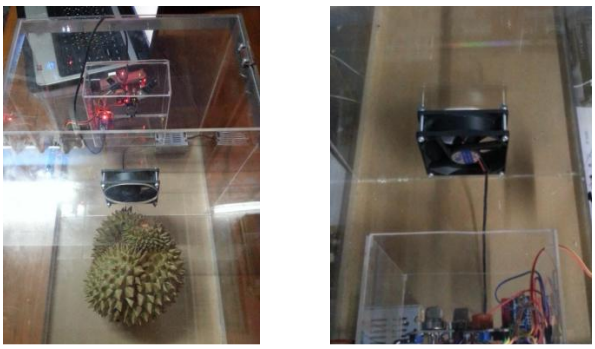

Figure 1: Display the plastic box for fruit ripening level measurement.

After gathering required data, it was analyzed by using clustering data mining technique to preliminary examine how many groups it can be classified. Classification technique was used for creating model to inspect fruit ripening level. To initiate, 160 attributes of data are collected and processed by cross validation technique [13]. In this experiment, various algorithms; Decision Tree, K Nearest Neighbors (KNN), Neural Network, Naïve Bayes and Support vector machine (SVM), were used to create models. After that, comparison has been made to find which one has the most accuracy rate. This research has been designed experiment into 3 steps; 1) create tools for gathering various gases amount data, 2) evaluate fruit ripening level by people and collect data from each type of sensors, 3) bring all gathered data for data mining.

3. Experimental Result

After developing tools to collect gases amount from durian, the study found that unripe one which cut from its tree would have high respiration rate (ethylene) that is similar to ripe durian at peak rate. It can be demonstrated from gases amount on durian. Data is collected and processed in data mining progress after specialists have been invited to evaluate fruit ripening level. The result of this experiment is as following:

K-means clustering has been used for classification group of fruit ripening level. As per relevant studies, it is found that K-Means is efficient for accurate group number classification [14], [15] by assigning mean or centroid of each cluster, finding distance from data to each mean and then grouping data with nearest neighbor mean of each cluster. After calculating distance, mean would be changed and classification group goes on until data group in each cluster is stable. There is no initializing label in this experiment so as to using K-Mean clustering technique to classify 3 gas data group. Primarily, the theory of this research divide fruit ripening level into 4 types as Thai people commonly get used to. Nonetheless, number of clustering group, K value, is assigned to be 2 to 6 for the purpose of what the result of K-means clustering shall be. The result displays as following table 1.

As per table 1, k=2 can be classified as 1) unripe or overripe, 2) pre-ripe or ripe. However, the result is not efficient due to their extremely different on how it has been chosen in the market. If k=3, it will be separated to 3 cluster group; 1) unripe, 2) pre-ripe, 3) ripe or overripe that are full of errors, for examples; unripe (2) in cluster_0, overripe (3) in cluster_1 and ripe (6) in cluster_2. When it comes to k=4, it can be classified into 4 groups; 1) unripe, 2) pre-ripe, 3) ripe, 4) overripe which small error include unripe (2) in cluster_0 and overripe (2) in cluster_3. However, numerous data in column k=5 and k=6, which are much different from each other such as unripe or overripe, appear in wrong clustering group. In conclusion, the most perfect k variable, or numbers of clusters, for durian ripening level classification by 3 type of gases detecting, is k= 4.

Table 1: Display k variable in K-means clustering algorithm.

k / cluster	k=2	k=3	k=4	k=5	k=6
cluster_0	unripe(40), ripe(1) overripe (39)	unripe(2), ripe(34) overripe(37)	unripe(2), ripe(38)	unripe(25), overripe(2)	unripe(13), overripe(8)
cluster_1	unripe(40), ripe(39) overripe(1)	unripe(38), overripe(3)	ripe(40)	ripe(40)	unripe(23), overripe(1)
cluster_2		pre-ripe(40), ripe(6)	pre-ripe(40)	overripe(26)	ripe(38)
cluster_3			unripe(38), overripe(2)	unripe(15), overripe(12)	unripe(4), ripe(1), overripe(19)
cluster_4				pre-ripe(40)	pre-ripe(40)
cluster_5					ripe(1), overripe (12)

Durian ripening level classification needs significant constituent element, gas amount detected by sensors. Due to this, properly algorithm has been chosen for the assessment. Decision Tree, K-Nearest Neighbor (KNN), Neural Network, Naive Bayes and Support vector machine (SVM) algorithm has been used for this research and compare which one is the best.

Firstly, default parameter set had been assigned in every algorithm to create model, and to compare each other for finding the most efficient and accurate one. Experiments with optimize parameter set was later performed and reiterated. The result is shown in Table 2.

Table 2: Display efficiency of algorithm in comparison.

Algorithm		Unripe accuracy	Pre-ripe accuracy	Ripe accuracy	Overripe accuracy	Average accuracy
Decision Tree	Default Parameter Set	100%	97.50%	97.50%	97.50%	98.12%
	Optimize Parameter Set	100%	100%	100%	97.50%	99.38%
K Nearest Neighbor s (KNN)	Default Parameter Set	100%	100%	100%	97.50%	99.37%
	Optimize Parameter Set	100%	100%	100%	97.50%	99.38%
Neural Network	Default Parameter Set	100%	100%	99.99%	99.99%	99.998%
	Optimize Parameter Set	100%	100%	100%	99.99%	99.999%
Naive Bayes	Default Parameter Set	100%	97.50%	97.50%	100%	98.75%
	Optimize Parameter Set	100%	97.50%	100%	100%	99.38%
Support vector machine (SVM)	Default Parameter Set	100%	100%	97.50%	100%	99.38 %
	Optimize Parameter Set	100%	100%	100%	100%	100%

Refer to Table 2, default parameter set has been assigned to each type of algorithm. It is found that Neural Network, Support vector machine and K Nearest Neighbors (KNN) algorithm are subsequently accurate than other ones. During the default parameter set assignment, Naive Bayes and Decision Tree are less accurate. However, optimized parameter has been adjusted and accuracy rate is raise to more than 99%.

4. Conclusion

Gas detection sensors, MQ3, MQ6 and MQ8, are developed as tools for measuring durian's gas amount to collect data from 160 durian samples. According to the data, it has been analyzed by clustering data mining process for group data classification. K-means algorithm used for categorizing group in this experiment. It reveals that fruit ripening level from diffusion of gases can be separated into 4 group of data; unripe, pre-ripe, ripe and overripe, which the result is similar to studies method and match with the types of durian level for consumer purchase.

Furthermore, the reason why the chosen models inaccurately categorize group, is durian's unstable respiration rate which raises when durian is unripe, and decrease when it is ripe. Later, it comes up to peak rate once it is overripe. Due to the fact that the respiration rate and diffusion gases at different ripening level are almost the same, the algorithms conduct and display inaccurate results.

Classification technique has been used on durian ripening level from diffusion of gases analysis. Initially, model was created to measure ripening level. This study applies all of 5 algorithms; Decision Tree, K Nearest Neighbors (KNN), Neural Network, Naïve Bayes and Support vector machine (SVM). As a result of rapid miner default parameter, Neural Network, following with Support vector machine and K Nearest Neighbors, are the most accuracy algorithm. On the other hand, Naive Bayes and Decision Tree are lower at accuracy rate. Parameter is later modified in Rapid miner. This experiment refers that the most accurate classification algorithm model is Support Vector Machines (SVM) consisted with relevant studies, a Comparative Efficiency of Neural Network Classification for the Diagnosis of Heatstroke [16], Developing and Effective Automatic Thai Document Categorization [17] and Thai Social Media Alert System for Business by Support Vector Machine [18]. Hence, SVM is the most appropriate and accurate method to assess fruit ripening level. In future research, tools will be developed for durian ripening level measurement by Arduino board. Algorithm would be considered other facts further than accuracy rate, for instance, difficulty of using models for developing and data processing speed.

Reference

[1] T. Kaushlendra, P. Saurabh, M. Murtaza and K. Tanushri, 2015. Fruit ripening of climacteric and non climacteric fruit. *Journal of Environmental and Applied Bioresearch*, 4(1), pp. 27-34.

[2] R.B.H. Wills, D.R. Harris, L.J. Spohr and J.B. Golding, 2014. Reduction of energy usage during storage and transport of bananas by management of exogenous ethylene levels. *Postharvest Biology and Technology*, 89, pp. 7-10.

[3] Michael M Blanke, 2014. Reducing ethylene levels along the food supply chain: a key to reducing food waste. *Journal of the science of food and agriculture*, 94(12), pp. 2357–2361.

[4] D. Noiwan, W. Kumpoun, K. Whangcha, K. Saengnil and J. Uthaibutra, 2016. Effect of Coating Materials on Storage Life of Mango cv. Mahajanaka. *Agricultural Science Journal*, 5(37), pp. 132-135.

[5] S. Subhadrabandhu and S. Ketsa, Durian: King of Tropical Fruit, CABI Publishing, 2001.

[6] P. Arancibia-Avila, F. Toledo, Y.-S. Park, S.-T. Jung, S.-G. Kang, B.-G. Heo, S.-H. Lee, M. Sajewicz, T. Kowalska and S. Gorinstein, 2018. Antioxidant properties of durian fruit as influenced by ripening. *LWT - Food Science and Technology*, 41(10), pp. 2118-2125.

[7] Krongyut, W, Srilaong, V, Uthairatanakij, A, Wongs-Aree, C, Esguerra, E. B. and Kanlayanarat, S, 2011. Physiological changes and cell wall degradation in papaya fruits cv. 'Kaek Dum' and 'Red Maradol' treated with 1- methylcyclopropene. *International Food Research Journal*, 4(18), pp. 1251-1259.

[8] M. Liu, . J. Pirrello, C. Chervin, J.-P. Roustan and M. Bouzayen, 2015. Ethylene Control of Fruit Ripening: Revisiting the Complex Network of Transcriptional Regulation. *Plant Physiology*. 169, pp. 2380-2390.

[9] Nath, P., Bouzayen, M., Mattoo, A. K. and Pech, J. C., 2014. Fruit Ripening: Physiology, Signalling and Genomics, United Kingdom: CPI Group (UK) Ltd.

[10] A. Eberhardt, K. Schmitt and J. Wöllenstein, 2017. Rotating interference filter spectrometer

for the detection of ethylene in the ripening process of climacteric fruit. in *19th International Conference on Solid-State Sensors, Actuators and Microsystems (TRANSDUCERS)*, Kaohsiung, Taiwan, 2017.

[11] D. O. Adams and S. F. Yang, 1979. Ethylene Biosynthesis: Identification of 1-aminocyclopropane-1-carboxylic Acid as an Intermediate in the Conversion of Methionine to Ethylene. *Proceedings of the National Academy of Sciences of the United States of America*, 76(1), pp. 170-174.

[12] Hashim, N., 2008. Study on Combustion Characteristics of Different Compositions of LPG," *UMP*, pp. 1 - 56.

[13] RapidMiner, 2017. Operator Reference Guide. [Online]. Available: https:// docs. rapidminer. com/ latest/ studio/ operators/ validation/ cross_validation. html. [Accessed 14 December 2017].

[14] Surawat,Sripaoraya and Saichon,Sinsomboonthong 2017. Efficiency Comparison of Data Mining Classification Methods for Chronic Kidney Disease: A Case Study of a Hospital in India. *Science and Technology Jouranl*. 25(5). pp. 839-853.

[15] Werayut, Pimpaporn and Phayung,Meesad. 2014. A Comparative Effciency of Clustering Using Dynamic Feature. *Information Technology Journal*, 10(2), pp. 43-51.

[16] Paranya Ammaruekarat and Phayung Meesad. 2010. A Comparative Efficiency of Neural Network Classification for the Diagnosis of Heatstork. in *The 6TH National Conference on Computing and Information Technology*, Bangkok Thailand.

[17] Nivet Chirawichitchai, Parinya Sanguansat and Phayung Meesad. 2011. Developing and Effective Automatic. *NIDA Development Journal*, 51(3) pp. 187 - 205.

[18] Supatta Viriyavisuthisakul1, Parinya Sanguansat, Pisit Charnkeitkong and and Choochart Haruechaiyasak. 2016. Thai social media alert system for business by Support vector machine. *Panyapiwat Journal*, 8(Special Issue), pp. 223 - 233.

Authors Biography

Udomsak Paeanpairoj[1] received the Computer Science B.S. (2015) from SuanSunandha Rajabhat University and Under Studying in Information Technology M.S. at College of Information and Communication Technology, Rangsit University, Thailand. He is Chief Petty Officer First Class of Royal Thai Navy. Currently, he is working for Office of The Navy Comptroller. He is interest in Technology and innovation. When he was studying in B.S., He research about Pali-Dictionary application. He won the “Thailand Start up award” competition of in June 2017.

Wutthipong Chinnasri[2] received the M.S. (2006) in Information Technology from King Mongkut's Institute of Technology Ladkrabang and Ph.D. (2013) in Computer Education from King Mongkut's University of Technology North Bangkok. Since 2007, he has been working as an assistant professor of Information Technology at College of Information and Communication Technology, Rangsit University, Thailand. Currently, his position is director of Master of Science program in Cybersecurity Management. His main research interests are focused on Artificial Intelligence, Computer Education, and Information Technology related field.