International Journal of Information, Control and Computer Sciences
ISSN: 2517-9942
Vol:10, No:2, 2016

Performance Comparison of Situation-Aware Models for Activating Robot Vacuum Cleaner in a Smart Home

Seongcheol Kwon, Jeongmin Kim, Kwang Ryel Ryu

***Abstract*—**We assume an IoT-based smart-home environment where the on-off status of each of the electrical appliances including the room lights can be recognized in a real time by monitoring and analyzing the smart meter data. At any moment in such an environment, we can recognize what the household or the user is doing by referring to the status data of the appliances. In this paper, we focus on a smart-home service that is to activate a robot vacuum cleaner at right time by recognizing the user situation, which requires a situation-aware model that can distinguish the situations that allow vacuum cleaning (Yes) from those that do not (No). We learn as our candidate models a few classifiers such as naïve Bayes, decision tree, and logistic regression that can map the appliance-status data into Yes and No situations. Our training and test data are obtained from simulations of user behaviors, in which a sequence of user situations such as cooking, eating, dish washing, and so on is generated with the status of the relevant appliances changed in accordance with the situation changes. During the simulation, both the situation transition and the resulting appliance status are determined stochastically. To compare the performances of the aforementioned classifiers we obtain their learning curves for different types of users through simulations. The result of our empirical study reveals that naïve Bayes achieves a slightly better classification accuracy than the other compared classifiers.

***Keywords*—**Situation-awareness, Smart home, IoT, Machine learning, Classifier.

I. Introduction

RECENTLY, interest in the Internet of Things (IoT) is increasing rapidly. The IoT is a technology that connects various devices that have communication function and sensors through the Internet. Connected devices exchange data over the Internet and provide services to people. The smart home, as a part of the IoT, is an automated house with installed sensors that provides intelligent services to users in the home.

In order to provide appropriate services for the smart home environment, it is important to infer the user situation. There were some researches about the smart home inferring the situation of the user [1]-[3] with various sensors, such as microphones, cameras, and infrared sensors, etc. For our research, we used a smart meter, which is an electronic power instrument, because the sensors used in previous works are hard to attach in the common home. The smart meter measures all of the electric power used, detecting which appliances are used in the home by the gap of measured power consumption [4]. We collected the status of home appliance data from the smart meter and inferred the situation of the user. In our case study, to activate the robot vacuum cleaner automatically, we need to develop a situation-aware model that would determine whether the vacuum cleaner can be activated or not according to user's situations. The situation-aware model should deactivate the vacuum cleaner if the user is in an inconvenient situation for cleaning, such as dining, cooking, taking a rest, or folding laundry.

Previous works [1]-[3] developed situation-aware models by predefined rules, but these rule-based models need to store too many rules if the smart home environment is complex, and it is hard to change the rules autonomously. Reference [5] used Bayesian network as a classifier, which models causal relationships between the home appliances and user situations. They developed a situation-aware model with some user feedbacks by adjusting conditional probability table (CPT) entries. However, some particularly initialized models could not be adapted well. To solve this problem, model should be initialized with reasonable data. Since classification performance varies according to which classifier is used, it is important to select a suitable classifier for given data. This paper aims to compare the performances of some classifiers: The Bayesian network, Naïve Bayes, decision tree, and logistic regression. To obtain the training and test data, we simulated a few users whose life patterns and probabilities for using home appliances are different. We cannot learn the whole model for each user, so the model determined by the selected classifier should be well-suited to users having various life patterns. We compared the learning curves of the aforementioned classifiers and the classification performance results when applying the learned models to users having different life patterns.

The rest of this paper is organized as follows. Section II reviews the previous works mentioned in detail. Section III describes the architecture of the simulator we used to obtain data and the classifiers we compared. In Section IV, we compared the learning curves and performances of the classifiers through simulation-based experiments. Finally, Section V provides a conclusion and recommendation for future work.

II. Related Works

Researches about inferring the user situation in the smart

Seongcheol Kwon and Jeongmin Kim are with the Department of Electrical and Computer Engineering, Pusan National University (phone: +82-51-510-3531; e-mail: ksc4323@naver.com, jeongminkim.islab@gmail.ac.kr).

Kwang Ryel Ryu is with the Department of Electrical and Computer Engineering, Pusan National University (phone: +82-51-510-2453; e-mail: krryu@pusan.ac.kr).

home have been made in various forms. Reference [1] used cameras and microphones as sensors to infer user situation. Cameras used for a 3D video tracking system detected where the user is, how fast the user moves and the user position (sitting, standing, lying down). Microphones were used for noise detection to infer whether the user was speaking or not. Data were encoded into the observation code and then they classified the user situation: individual work, introduction, aperitif, siesta, presentation, and game. References [2] and [3] used many sensors, such as infrared sensors and ultrasonic sensors, to detect the user location, posture, and time. They assumed some sequences of the user situation (scenario) and predefined rules about using appliances when a particular situation happens. References [1]-[3] used sensors that are hard to attach in the home. Furthermore, without continuous observation, it is hard to collect user situation data, which is a target value for classifying. To solve these problems, [5] used only a smart meter to get the status of home appliances and collected user feedbacks that contain the user situation at that time. The collected feedbacks were used for learning the Bayesian network model, which infers the user situation from the status of home appliances data and determines whether an automated robot vacuum cleaner activates or not. The learning process, e.g., if the user gives a feedback to clean the deactivated vacuum cleaner, [5] used an algorithm for adjusting CPT entries of Bayesian network to activate the robot vacuum cleaner at that time. However, a problem prevented learning situation-aware models from integrating with some particularly initialized models.

III. Simulator and Classifiers

A. Simulator

We assumed a network considering the causal relationship between user situations and home appliances, as shown in Fig. 1.

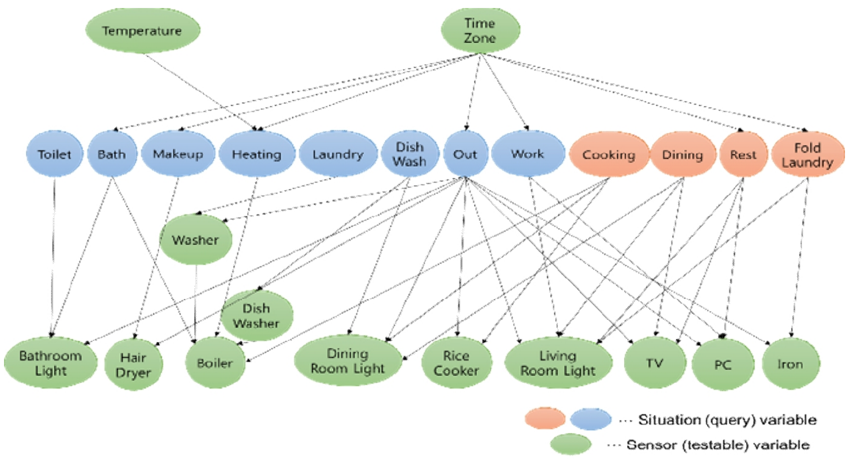


Fig. 1 The assumed network topology

The green nodes mean sensor variables that can be observed through the smart meter, while other nodes mean the user situation variables. Among the user situation variables, the red-colored nodes mean the situation in which the user feels it is inconvenient to vacuum. Each edge means that there is a relation between two nodes. We used Fig. 1 network topology to generate reasonable data through simulation.

The architecture of the simulator we used for obtaining data is divided into four parts, as shown in Fig. 2: File reader, scheduler, device controller, and data generator. The input file contains some information about the life pattern of the user, probabilities that the situation will occur, and operation probabilities of home appliances when the situation occurs. The simulator reads an input file and sets the daily schedule of the user stochastically. After scheduling, with the probabilities read from input data, the device controller determines the on-off status of home appliances whenever a situation transition occurs. Finally, the data generator combines the user situation data with home appliance status data for learning models.

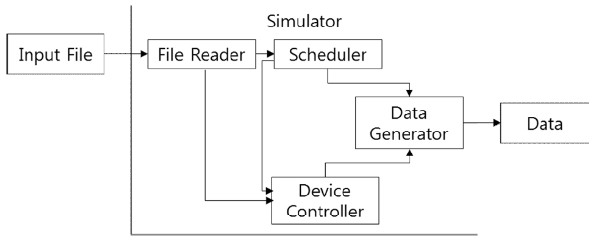
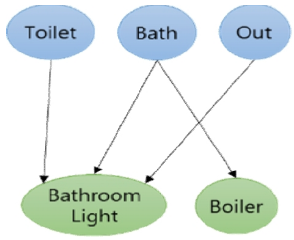


Fig. 2 Architecture of the simulator for generating data

We simulated three users: a housewife, a student, and another student who feels inconvenient when the robot vacuum cleaner activated while he works at home. The housewife and the student have different life patterns, so their daily schedules are dissimilar. Because of this dissimilarity, their patterns of using appliances are also different. For example, the housewife folds laundry in the morning, so she usually does not use the living room light when folding laundry; however, the student folds laundry at night, so he usually turns on the living room light. We generated 10-day simulated data and simplified the data by compressing continuous data that have same sensor value and class value to one data. It means that we collected data when the situation transition happened because the continuous data have some dangerous to overfit.

B. Classifiers

As mentioned in Section I, we selected four candidate classifiers to compare: The Bayesian network [6], Naïve Bayes [7], decision tree [8], and logistic regression [9]. The Bayesian network, also called a causal relationship graph, is a probabilistic model based on Bayes' theorem. It is a directed acyclic graph that represents the relation between two nodes as a connected edge and CPT. Fig. 3 shows a simple example of a Bayesian network. In Fig. 3, each CPT entry means the operation probability of the bathroom light according to the value changes of the connected parent node (Toilet, Bath, Out). The whole Bayesian network topology is exactly same as in Fig. 1, and each node has its own CPT. If the on-off statuses of sensor variables are given, we can calculate the probability of the user situation by CPT. The situation with the highest probability means the classification result of the current user situation.



T	*B*	*O*	*bl*	$\neg bl$
T	T	T	0.5	0.5
T	T	F	0.73	0.27
		⋮		
F	F	F	0.02	0.98

Fig. 3 A simple example of Bayesian network

Naive Bayes is a probabilistic model that simplified the Bayesian network with a strong independence assumption between sensor variables. Fig. 4 shows the Naïve Bayes model we used, and the situation simplified whether the vacuum cleaner could activate or not. The process of inferring is similar to the Bayesian network, but the computational cost is much lower because of the complexity.

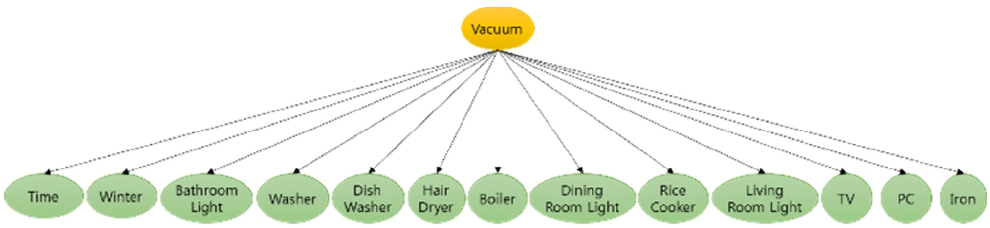


Fig. 4 The Naïve Bayes model we used

A decision tree is a tree-shaped directed acyclic graph. After learning the decision tree model, the component of every leaf node was a class label we wanted to classify. The root node is a variable with the highest information gain and the other nodes are sorted by the entropy in decreasing order, from root node to leaf node. The decision tree helps in making a complex decision by making several simple decisions. Our decision tree has root and internal nodes that represent the on-off status of home appliances, and leaf nodes that represent whether the vacuum cleaner can activate or not.

The last classifier we want to compare in our paper is logistic regression. Logistic regression has a linear decision boundary; in our case, the least probability that a vacuum cleaner can activate, which was made by selecting the probability where classes are maximally overlapping. Generally, decision boundary of logistic regression is hard to achieve by calculating, so use of some search algorithms is needed.

IV. Experiments

To compare classifiers that we aforementioned, two experiments were conducted. The first experiment compared four types of measures described in Table I, with learning curves of each classifier when learned models with user data were self-applied. The measures shown in Table I were calculated by four types of classification results: true positive, false negative, true negative, and false positive. The aim of first experiment was to measure how fast the model converged with less data and how accurate the model was. The second experiment compared measures when the learned model with one user then applied to another user whose life pattern is different. The goal of the second experiment was to find which classifier operates well for users having various life patterns. We obtained averages of all measures after 10 experiments. Training data used in each of the 10 experiments and test data were generated through simulation.

TABLE I
Four Types of Measures

Measure	Description
Accuracy	The proportion of correctly classified examples
F-measure	The harmonic average of precision and recall
False Positive Rate	The proportion of examples labeled False among examples classified as True
False Negative Rate	The proportion of examples labeled True among examples classified as False

In this section, we will show the only meaningful results among whole experiments. The experiment results of learned model with housewife data and self-applied are shown in Fig. 5. The learning curves show that the Bayesian network classified more accurate to negative situation, means cannot activate vacuum cleaner, than positive. The Bayesian network usually misclassified the rest situation to outside situation in the

International Journal of Information, Control and Computer Sciences
ISSN: 2517-9942
Vol:10, No:2, 2016

morning when all devices were turned off. That affected the false positive rate of the Bayesian network to be higher than other classifiers.

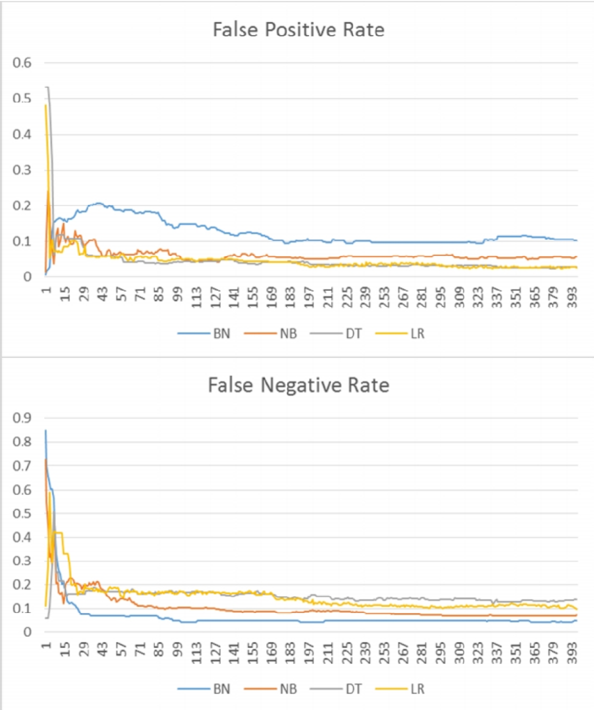


Fig. 5 Learning curves of false positive rate and false negative rate for self-applied housewife model

Self-applying the learned student model experiment results are shown in Fig. 6. Student data is difficult to learn because the probabilities of operating appliances have a lot of uncertainty. As well as the PC and the living room light sensors were commonly used together in rest situation and work situation, so overall classifiers did not perform well. The student uses PC not only the rest situation but also the work situation. The Bayesian network could classify the rest situation in afternoon more accurate than other classifiers because of its high dependency to timezone node. The timezone node is a parent node of the rest node so it affects the probability of the rest situation happening. The result of high F-mesure in Bayesian network because it modeled the causal relationship between timezone and rest well.

The self-applying experiment results of learned model with student who does not want to activate the vacuum cleaner in the work situation are shown in Fig. 7. By changing the work situation to do not activate the vacuum cleaner, the data used for experiment were much simpler than original student data. However, Fig. 7 shows that Bayesian network has much less accuracy than the other models because it could not classify the dining situation.

The second experiment was applying the learned model to the other users and measuring the classification performance. Table II shows the results of the experiments when the model learned housewife data applied to the other users. Naïve Bayes performs better than other the other models because it classifies the rest situation well.

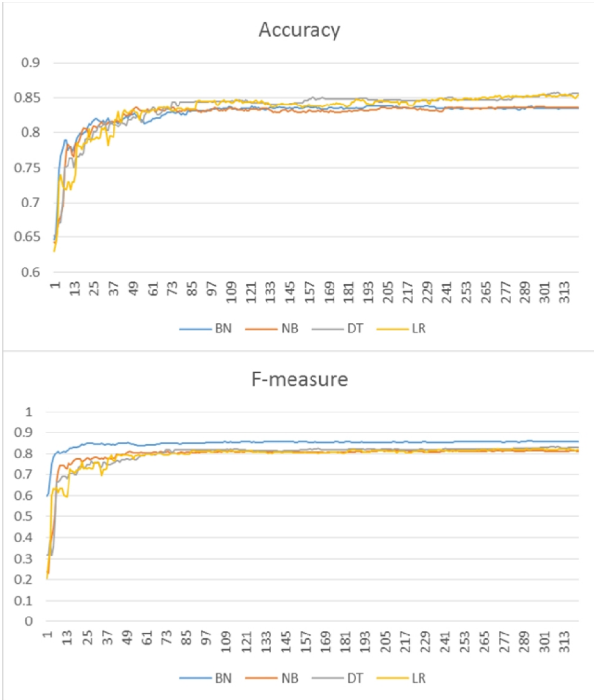


Fig. 6 Learning curves of accuracy and F-measure for self-applied student model

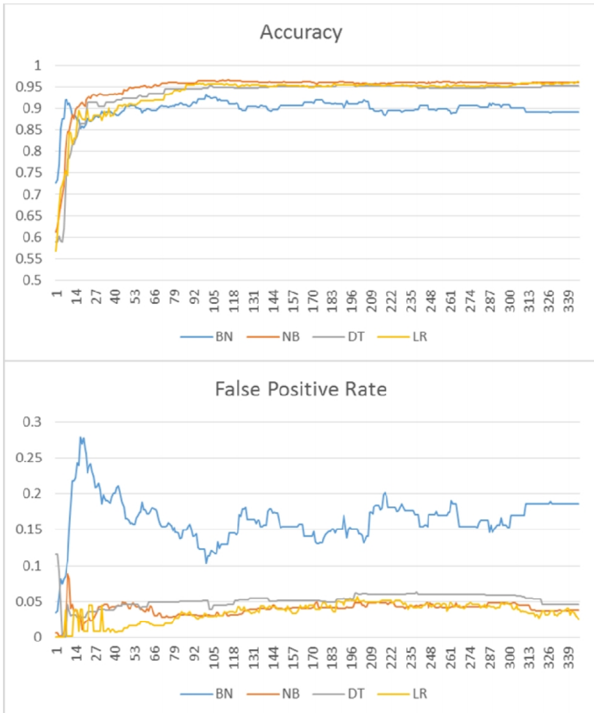


Fig. 7 Learning curves of accuracy and false positive rate for self-applied student who feels inconvenient when the robot vacuum cleaner activated while he works

TABLE II
AVERAGE PERFORMANCE WHEN APPLYING THE LEARNED HOUSEWIFE MODEL TO THE OTHER USERS

Applied user	classifier	Accuracy	False positive rate	False negative rate	F-measure
Student	BN	0.793	0.262	0.171	0.828
	NB	**0.849**	0.173	**0.122**	**0.832**
	LR	0.795	0.159	0.274	0.737
	DT	0.786	**0.091**	0.403	0.688
Student (work)	BN	0.867	0.260	**0.005**	0.882
	NB	**0.905**	0.014	0.177	**0.896**
	LR	0.857	**0.006**	0.290	0.835
	DT	0.789	0.008	0.413	0.736

*BN: Bayesian network, NB: Naïve Bayes LR: Logistic regression, DT: Decision tree

TABLE III
AVERAGE PERFORMANCE WHEN APPLYING THE LEARNED STUDENT MODEL TO THE OTHER USERS

Applied user	classifier	Accuracy	False positive rate	False negative rate	F-measure
Housewife	BN	0.812	0.211	0.154	0.781
	NB	**0.899**	**0.027**	0.174	0.891
	LR	0.857	0.141	0.144	0.878
	DT	0.875	0.158	**0.103**	**0.896**
Student (work)	BN	0.901	0.170	**0.029**	0.907
	NB	**0.930**	0.008	0.132	**0.925**
	LR	0.906	**0.007**	0.180	0.898
	DT	0.899	0.027	0.174	0.891

*BN: Bayesian network, NB: Naïve Bayes LR: Logistic regression, DT: Decision tree

TABLE IV
AVERAGE PERFORMANCE WHEN APPLYING THE LEARNED STUDENT (WORK) MODEL TO THE OTHER USERS

Applied user	classifier	Accuracy	False positive rate	False negative rate	F-measure
Housewife	BN	0.812	0.208	0.157	0.781
	NB	**0.880**	**0.160**	**0.093**	**0.901**
	LR	0.848	0.167	0.141	0.872
	DT	0.866	0.163	0.115	0.888
Student	BN	0.837	**0.171**	0.158	**0.862**
	NB	**0.845**	0.230	**0.052**	0.839
	LR	0.834	0.238	0.057	0.818
	DT	0.833	0.234	0.064	0.816

*BN: Bayesian network, NB: Naïve Bayes LR: Logistic regression, DT: Decision tree

As shown in Table III, when we applied the learned model with the student data to housewife, the Bayesian network did not perform well because if misclassified the bath situation many times. The bath situation occurs in the morning in the student data while it occurs in the evening in the housewife data.

The learned model with the student data have little probability with bath in the evening, so the Bayesian network could not classify the bath situation well. Overall, the mostly misclassified situations were the rest situation and the work situation. They share the PC and the living room light sensor together, but the label that whether can activate vacuum cleaner or not is different. The Bayesian network misclassified these situations especially many times because of its high dependency between timezone node and the situations. Naïve Bayes classified the rest situation and the work situation relatively well, so it shows better performance than the other models in terms of accuracy.

V. CONCLUSION AND FUTURE WORKS

We did some simulation-based experiments to find out the classifier which performs well to our generated data for our case study. As a result of comparing the classifiers, Bayesian network, Naïve Bayes, decision tree, and logistic regression, with the users having different daily life pattern, Naïve Bayes was suitable to our data. It could be applied to other users with suitable level of performance and the model can be learned easily.

As a future work, we should learn a model that customized to a user with user feedback data based on well-initialized model.

ACKNOWLEDGMENT

This work was supported by Institute for Information & communications Technology Promotion(IITP) grant funded by the Korea government(MSIP) (No.10043907, Development of high performance IoT device and Open Platform with Intelligent Software)

REFERENCES

[1] O. Brdiczka, J. L. Crowley, and P. Reignier, “Learning situation models in a smart home,” *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, vol. 39, no. 1, pp. 56–63, 2009.

[2] K. Wongpatikaseree, M. Ikeda, M. Buranarach, T. Supnithi, A. O. Lim, and Y. Tan, “Activity recognition using context-aware infrastructure ontology in smart home domain,” in *Knowledge, Information and Creativity Support Systems (KICSS), 2012 Seventh International Conference on*. IEEE, 2012, pp. 50–57.

[3] P. Chahuara, F. Portet, and M. Vacher, “Context aware decision system in a smart home: knowledge representation and decision making using uncertain contextual information,” in *The 4th International Workshop on Acquisition, Representation and Reasoning with Contextualized Knowledge (ARCOE-12)*, 2012, pp. 52–64.

[4] J. Kim and H. Kim, “The prediction and classification of family type by using power consumption model over smart grid,” *Computer*, vol. 29, no. 1, pp. 170-175, 2014.

[5] E. T. Ha, S. Kim, J. Kim, and K. R. Ryu, “Learning a Bayesian Network for Situation-Aware Smart Home Service: A Case Study with a Robot Vacuum Cleaner,” *International Science Index Computer and Information Engineering*, vol. 2, no. 2, pp. 877-884, 2015

[6] N. Friedman, D. Geiger, and M. Goldszmidt, “Bayesian network classifiers,” *Machine learning*, vol. 29, no. 2-3, pp. 131–163, 1997.

[7] K. P. Murphy, “Naive bayes classifiers,” *University of British Columbia*, 2006.

[8] S. R. Safavian and D. Landgrebe, “A survey of decision tree classifier methodology,” *IEEE transactions on systems, man, and cybernetics*, vol. 21, no. 3, pp. 660–674, 1991.

[9] C.-Y. J. Peng, K. L. Lee, and G. M. Ingersoll, “An introduction to logistic regression analysis and reporting,” *The Journal of Educational Research*, vol. 96, no. 1, pp. 3–14, 2002.