ACTIVITY RECOGNITION FOR AMBIENT SENSING DATA AND RULE BASED ANOMALY DETECTION

E. Seyedkazemi Ardebili [1], S. Eken[2],*K. Küçük[1]

[1] Koceli University, Dept. of Computer Engineering, 41001 İzmit Koceli, Turkey - seyed.esk@gmail.com, kkucuk@kocaeli.edu.tr
[2] Koceli University, Dept. of Information Systems Engineering, 41001 İzmit Koceli, Turkey - suleyman.eken@kocaeli.edu.tr

Commission VI, WG VI/4

KEY WORDS: Smart Home, Anomaly Detection, Machine Learning, Human Activity, Internet of Things

ABSTRACT:

After a brief look at the smart home, we conclude that to have a smart home, and it is necessary to have an intelligent management center. In this article, We have tried to make it possible for the smart home management center to be able to detect the presence of an abnormal state in the behavior of someone who lives in the house. In the proposed method, the daily algorithm examines the rate of changes of a person and provides a number which is henceforth called NNC (Number of normal changes) based on the person's behavioral changes. We achieve the NNC number using a machine learning algorithm and performing a series of several simple statistical and mathematical calculations. NNC is a number that shows abnormal changes in residents' behaviors in a smart home, i.e., this number is a small number for a regular person with constant planning and for a person who may not have any fixed principles and regular in personal life is a big number.To increase our accuracy in calculating NNC, we review all common machine learning algorithms and after tests we choose the decision tree because of its higher accuracy and speed and finally, NNC number is obtained by combining the Decision Tree algorithm with statistical and mathematical methods. In this method, we present a set of states and information obtained from the sensors along with the activities performed by the occupant of the house over a period of several days to the proposed algorithm. and the method ahead generates the main NNC number for those days for anyone living in a smart home. To generate this main NNC, we calculate each person's daily NNC. That means we have daily NNCs for each person (based on his/her behaviors on that day) and the main NNC is the average of these daily NNC. We chose ARAS dataset (Human Activity Datasets in Multiple Homes with Multiple Residents) to implement our method and after tests and replications on the ARAS dataset, and to find anomalies in each person's behavior in a day, we compare the main(average) NNC with that person's daily NNC on that day , Finally, we can say, if the main NNC changes more than 30%, there is a possibility of an abnormality. and if the NNC changes more than 60% percent, we can say that an abnormal state or an uncommon event happened that day, and a declaration of an abnormal state will be issued to the resident of the house.

1. INTRODUCTION

Perhaps in the 1980s and 1995s, when the first experimental intelligent systems were in the making, in many human societies, talking about nowadays' smart homes may have meant talking about a dream or it was a fictional story or, at best, talking about a luxury item. But after only a few years, all kinds of smart homes are seen. Although the beginning of smart homes began even in the 1970s, but at a glance, the start of smart homes in the modern sense can be considered as a time from the beginning of intelligent objects to the human living environment. Ever since ordinary mobile phones moved to smartphones, it was at this time that we have witnessed a great leap in artificial intelligence and communication; moreover, the Internet of Things (IoT) has taken on more real meaning, and smart homes have emerged one after another.

In the early days of smart homes, building a primitive smart home that provided the simplest of services sometimes cost millions of dollars, which was not possible for most homes built due to the high cost of the system. However, the passage of time and the increasing attention of industries to the discussion of smart homes led to increased efforts to reduce costs and improve the quality and variety of parts production in the hardware sector, and this increase in production has gradually led to these smart home systems with facilities Much more advanced than before can be purchased for less than $500.

After solving the problems of producing all kinds of hardware and sensors related to home smartening, attention to the software dimension became more intense, and the reason for this can be considered the need for a smarter brain than before to manage and connect these components. In other words, the newly produced hardware needed software that could make the most of them. This paper tries to provide a method to help make the core of building management smarter by analyzing the information collected from a smart home.

Now that the most important part of a smart home is the home management center, to have a smarter management center, this center must have an effective and smarter relationship with home equipment and home residents, and for this communication, it must have a better understanding of home conditions and recognizing abnormalities in residents' behaviors is perhaps one of the most important parts. An attempt has been made to find an algorithm to understand the abnormal conditions in the house or more precisely in the behavior of the resident of the house (Pan et al., 2016). According to the available methods to find the abnormal state and even allows the study of the behavior of residents with surveillance camera images. In this method, in addition to trying to increase the accuracy of finding the abnormal state for people who usually do not have a fixed behavior pattern in life. The first is the simplicity and possibility

---

* Corresponding author

of implementation in the simplest smart computers at the smart homes, and the second is to increase the privacy of residents (Tapia et al., 2004).

2. RELATED WORKS

Internet of Things (IoT) and smart homes are two present tendencies in the world. Due to these technologies moreover, the existent types of equipment in a smart home allow researchers to collect data from a person's daily activities and analyze these data and use them in the field of anomaly detection. In recent years, due to the growth of the IoT and on the other hand, a significant increase in the number of smart homes and as a result, the increase in the information collected from this home has led to various articles and research in this field. By reviewing the articles, it can be seen two general approaches in making homes as smart as possible. In some cases, trying to provide smarter methods for the building management system, this central system is able to do many things that the residents of the house from habit and involuntary daily doing is done by intelligent systems. Studies have shown that in this type of paper, intelligent management, in most cases, has saved time and money, and in addition, has led to the right consumption. But another approach that has been done with the aim of improving the welfare and comfort and increasing the security of home residents and turning the home environment into a pleasant environment for the home resident. The purpose of presenting the method of this article is in line with the second approach and to try to have a method that can detect any anomalies and adopt a newer process for further investigations.

To better understand the needs and situations, we examine the possibility of an abnormal state event in previous papers and use the insight presented in these papers for abnormal states in the IoT and smart homes. Contrary to some papers, it is possible to pay major attention to the possibility of disturbing the information received from the sensor, in this paper, the possibility of hacking is ignored, and it is considered that the information received from the sensors is correct (Ali et al., 2017, Ramapatruni et al., 2019).

In some papers, a home network is simulated to implement methods for better learning, and this information is used to try to find a priority sequence to the extent that some activities are omitted for better learning and to discuss the accuracy and effectiveness of the proposed methods, a comparison is made with the methods of learning user behavior by Hidden Markov Models (Yamauchi et al., 2020). In some papers, it has an approach similar to the proposed approach of this article, and it has been similar in terms of statistical studies, and this combination of statistical techniques with machine learning techniques is evident.. Those templates are being calculated by combining machine learning techniques and statistical techniques according to their network behaviors in a smart house, so in these method, statistical measurements sometimes are used after processing to build clusters (Fahad, Rajarajan, 2015, Spanos et al., 2019).

In research like this, several paths and even methods based on image processing have been presented; for example, it can be said that a approach that uses recorded images has the ability to detect abnormal behaviors, but the important point is that in some cases, it will not work for people who have a lot of unconventional behaviors, or if it works with a fundamental change, the designed machine may make a serious error. It is important to note that this method does not require a strong computer, which does not require heavy processing in this method.

3. ACTIVITY RECOGNITION WITH AMBIENT SENSING DATA

3.1 Data Description

The importance of selecting real human activities led to the selection of a real datasets from a set of human activities in a smart home, so The ARAS (Activity Recognition with Ambient Sensing) datasets has been selected for experiments and initial implementation of algorithms (Alemdar et al., 2013). The ARAS datasets identify human activities collected over two months from two real homes with multiple residents. ARAS datasets contain the ground truth labels for 27 various activities. Each house was equipped with 20 binary sensors of various kinds that communicate wirelessly using the ZigBee protocol (Al Baalbaki et al., 2015). A full month of data that contains the activity labels and the sensor data for both residents was gathered from each home, resulting in a total of two months of data.

3.2 Data Preprocessing

Due to the standardization of the information in the ARAS database and compliance with the needs of the algorithm, no significant changes were made. Only one column was added to specify the time of each activity to indicate the seconds the activity occurred.

3.3 Used Machine Learning Algorithms

Another part that is to be used in the formation of NNC number is the use of a machine learning method. In this section, there are two approaches. The first approach is based on each activity and is such that it divides a day into 24 hours and is checked for each hour using the decision tree algorithm and checking the status of the sensors in at that time, was it possible for the intended activity to occur at this time or not? If the activity is performed by the resident of the house, the number one is considered and otherwise, the number zero is considered for that hour, and this process is checked for all activities. At the end of the work, the information received from the activities performed by the resident of the house is checked by presenting 24 members. Any difference between the two increases the size of the NNC.

The second approach is done regardless of the activity type, i.e. seconds are selected completely randomly, and according to the status of the sensors and previous information, the activity of the resident is predicted and stored. Finally, these predictions are compared with the actual activities performed, and the degree of correspondence between the two causes the NNC number to be large or small. One way to increase the accuracy and segregation of the dependence of the occupants' activities on the occupants' occupants (in the multiple residents) is that the above process is done twice, once considering the activity of other residents at that time and once without considering the activity information of other residents (Mehr et al., 2016).

One of the most important parts of this method is selecting an algorithm from the machine learning algorithm that has the best performance in the fastest time. Choosing the best algorithm may not only increase performance accuracy but also provide scalability for the implemented method for any house with any number of inhabitants. First, after reviewing the available algorithms in machine learning, we selected a number of common algorithms and tested in practice on the ARAS datasets, and the results are presented in Table 1.

Algorithm	K-Nearest Neighbors		Support Vector Machine		Decision Tree		Logistic Regression		Random Forest	
Resident	*R1*	*R2*	*R1*	*R2*	*R1*	*R2*	*R1*	*R2*	*R1*	*R2*
Accuracy	0.68	0.78	0.73	0.78	**0.86**	**0.95**	0.33	0.52	0.87	0.95
Precision	0.61	0.67	0.65	0.67	**0.79**	**0.93**	0.01	0.02	0.76	0.92
Recall	0.47	0.50	0.48	0.50	**0.75**	**0.90**	0.04	0.04	0.75	0.90
F1-score	0.51	0.56	0.53	0.56	**0.76**	**0.91**	0.02	0.03	0.75	0.91
	LightGBM		Gradient Boosting		CatBoost		XGBoost		Gaussian NB	
Resident	*R1*	*R2*	*R1*	*R2*	*R1*	*R2*	*R1*	*R2*	*R1*	*R2*
Accuracy	0.60	0.32	0.86	0.91	0.61	0.77	0.83	0.89	0.51	0.64
Precision	0.41	0.10	0.85	0.88	0.30	0.44	0.78	0.95	0.21	0.11
Recall	0.37	0.09	0.74	0.73	0.28	0.39	0.67	0.72	0.25	0.17
F1-score	0.37	0.09	0.77	0.77	0.27	0.40	0.71	0.79	0.20	0.12

Table 1. Comparison of machine learning algorithms

4. RULE BASED ANOMALY DETECTION

To make a final decision about the anomalies, we introduced the daily NNC and main NNC. Daily NNC is composed of the five sub-NNCs that each of these sub-NNCs increase the daily NNC by one unit when they detect any abnormal changes. In the following, we define the sub-NNCs. We compute the daily NNC with two machine learning methods as well as the three statistical methods (see Table 2), and we get the main NNC from the average of that person's daily NNC.

In parts based on the decision tree algorithm and in homes with in multiple residents, we have two approaches to creating a model for the machine learning part. Once the information of the activities of another resident is part of the input features of the decision tree algorithm, once is not. For a better understanding, we can say this. We have two residents in the home, and there are 20 sensors at home. In the first experiment, we use 20 sensors as features of a decision tree, and then we decided to check that what will happen if we add the activities of the second resident as the extra feature of the model, so in the second model, we have 21 inputs. Adding the activities of the second resident as input, improved accuracy on average around 10%, but we find that if the second resident has anomalies in her/his activities, it degraded the final results. So we decided to use both of the models to compute the daily NNC.

The method suggested in this paper starts to calculate the daily NNCs, and Based on the 15 days, creates a tool to compute the main NNC (from day 16). our method computes the average NNC for the last 15 days, and in the day that we want to check the anomaly, it computes the daily NNC, and if the variation of daily NNC that day from the main NNC (average NNC in the last 15 days) is higher than 30% We can say, there may be abnormalities in a person's behaviors.

5. DISCUSSION AND CONCLUSION

In this paper, we present a concept in the form of a number to measure the rate of change in the behaviors of the resident of the smart home. We call this number NNC and it will be a small number and even tends to zero in regular residents with a routine schedule. Using NNC, in addition to being able to detect the presence of abnormal conditions in the short term, in the long term, the behavior of residents can be detected, and If the NNC does not change constantly and remains constant around a number, it can be concluded that normal behaviors may appear from the resident. In addition, every resident of the house can have a more regular life by trying to make NNC smaller. In this paper, we have a daily NNC for each day and the main NNC is calculated from the average of the daily NNC.

In short, to find anomalies in a day, we compare the daily NNC with the main NNC.

In more specialized reviews in the proposed method, it is possible to diagnose diseases such as depression. Moreover, in the complementary stages, we can assign tasks to this system based on the protocols that the residents have planned. In the end, although these efforts to make the home smarter with the aim of increasing the security and well-being of human beings, we should not neglect this point, as this smartening is valuable to the extent that it does not invade human privacy and cause annoyance (Tang et al., 2019). Perhaps the importance of providing an NNC number for smart homes where single-person families or the elderly live alone is much more obvious, and with each passing day, more and more people feel the need for a system that can better understand smart home conditions and residents' behaviors (Galvão et al., 2017). There are a lot of smart devices in every person's life today, and the most important limitation in implementing this algorithm today is the lack of access to information from outside sensors. In the near future, perhaps on the one hand, by increasing the number of sensors in smart homes and on the other hand by connecting all available smart equipment and the possibility of using smart device information outside the home such as a smart watch or smart car information. Also, it is planned to increase the accuracy values of used algorithms. All codes can be accessed from: https://github.com/seyedkazemi/NNC.

REFERENCES

Al Baalbaki, B., Pacheco, J., Tunc, C., Hariri, S., Al-Nashif, Y., 2015. Anomaly behavior analysis system for zigbee in smart buildings. *2015 IEEE/ACS 12th International Conference of Computer Systems and Applications (AICCSA)*, IEEE, 1–4.

Alemdar, H., Ertan, H., Incel, O. D., Ersoy, C., 2013. Aras human activity datasets in multiple homes with multiple residents. *2013 7th International Conference on Pervasive Computing Technologies for Healthcare and Workshops*, IEEE, 232–235.

Ali, W., Dustgeer, G., Awais, M., Shah, M. A., 2017. Iot based smart home: Security challenges, security requirements and solutions. *2017 23rd International Conference on Automation and Computing (ICAC)*, IEEE, 1–6.

Fahad, L. G., Rajarajan, M., 2015. Anomalies detection in smart-home activities. *2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA)*, IEEE, 419–422.

	Based on statistics		Based on machine learning
S1	For each activity, for each hour, we extract the minimum and maximum continuous spent time during the last 15 days. We compare with test days' results if for each hour for an activity she/he spent more than the maximum we increase the NNC by one unit and also if the spent time for activities is less than the minimum one we increase the NNC by one unit.	M1	We randomly selected 7200 seconds from the test day, and then we generate the prediction for activities for each 5% of error in prediction. We increase the NNC by one unit.
S2	For each activity for each hour of the day, we racked from 0 to 15 based on that activity's frequency in that specific hour in the last 15 days. Then we compare each activity in each hour of the test day with an above mentioned rank list if activity in a specific hour is done in test day which has a frequency less than three times the NNC increase 1 unit, and if an activity that has frequency more than 11 times in rank list which is not done in the test day the NNC increase by one unit.	M2	With the trained model, for each hour of test day, we predict the activities (If an activity lasts for more than a minute in an hour, we consider that activity to be done) and compare them with the actual label of activities in that hour for each activity prediction error we increase the NNC by one unit.
S3	We compute NNC utilizing the following three components: i- For each of the 27 activities, we calculate the sum of spend time, then we compared with the average of those activities in the last 15 days. If the variation of the sum of activities is more than 50% of the average one, then NNC increases by one unit. ii- For activities that done in a test day, if the frequency of activities is less than three times in the last 15 days, we increase the NNC by one unit meanwhile, for activities that did not do on the test day if the frequency is more prominent than 12 times in last 15 days we increase the NNC by one unit. iii- Considering the last 15 days, we compute the daily sum for each activity; then, we extract the maximum and minimum spent time for each activity in the last 15 days. If in the test day, he/she spent less than the minimum, the NNC increase by one unit, and if the spent time is higher than the maximum, the NNC increase by one unit.		

Table 2. Rule-based anomaly list

Galvão, Y. M., Albuquerque, V. A., Fernandes, B. J., Valença, M. J., 2017. Anomaly detection in smart houses: Monitoring elderly daily behavior for fall detecting. *2017 IEEE Latin American Conference on Computational Intelligence (LA-CCI)*, IEEE, 1–6.

Mehr, H. D., Polat, H., Cetin, A., 2016. Resident activity recognition in smart homes by using artificial neural networks. *2016 4th international istanbul smart grid congress and fair (ICSG)*, IEEE, 1–5.

Pan, Z., Pacheco, J., Hariri, S., 2016. Anomaly behavior analysis for building automation systems. *2016 IEEE/ACS 13th International Conference of Computer Systems and Applications (AICCSA)*, IEEE, 1–8.

Ramapatruni, S., Narayanan, S. N., Mittal, S., Joshi, A., Joshi, K., 2019. Anomaly detection models for smart home security. *2019 IEEE 5th Intl Conference on Big Data Security on Cloud (BigDataSecurity), IEEE Intl Conference on High Performance and Smart Computing,(HPSC) and IEEE Intl Conference on Intelligent Data and Security (IDS)*, IEEE, 19–24.

Spanos, G., Giannoutakis, K. M., Votis, K., Tzovaras, D., 2019. Combining statistical and machine learning techniques in iot anomaly detection for smart homes. *2019 IEEE 24th International Workshop on Computer Aided Modeling and Design of Communication Links and Networks (CAMAD)*, IEEE, 1–6.

Tang, S., Gu, Z., Yang, Q., Fu, S., 2019. Smart home iot anomaly detection based on ensemble model learning from heterogeneous data. *2019 IEEE International Conference on Big Data (Big Data)*, IEEE, 4185–4190.

Tapia, E. M., Intille, S. S., Larson, K., 2004. Activity recognition in the home using simple and ubiquitous sensors. *International conference on pervasive computing*, Springer, 158–175.

Yamauchi, M., Ohsita, Y., Murata, M., Ueda, K., Kato, Y., 2020. Anomaly Detection in Smart Home Operation From User Behaviors and Home Conditions. *IEEE Transactions on Consumer Electronics*, 66(2), 183–192.

Revised September 2020