Activity Recognition: An Evolutionary Ensembles Approach

Muhammad Fahim, Iram Fatima, Sungyoung Lee and Young-Koo Lee
Ubiquitous Computing Laboratory
Department of Computer Engineering, Kyung Hee University, Korea
{fahim, iram.fatima, sylee}@oslab.khu.ac.kr, yklee@khu.ac.kr

ABSTRACT
Activity recognition is an emerging field that demands active research in ubiquitous computing for analyzing complex scenarios such as concurrent situation assessment and domination of major over the minor activities. In this paper, an evolutionary ensembles approach using Genetic Algorithm (GA) as a homogeneous learner has been proposed. This approach values both minor and major activities by processing each of them independently. It consists of two phases. The first phase is preprocessing of sensory data and extraction of feature vectors. Evolutionary ensembles are designed in second phase to learn different daily life activities. Finally, multiple ensembles output is pooled on central node as a complete rule profile for all performed activities. The proposed approach was evaluated on six different types of activities from Intelligent System Laboratory (ISL) dataset. It shows a higher accuracy as compared to single learner GA.

Author Keywords
Activity Recognition, Genetic Algorithm, Ensemble learning.

ACM Classification Keywords
I.5 Pattern Recognition, I.1.2 Algorithms

General Terms
Algorithms, Design, Performance, Reliability

INTRODUCTION

Activity Recognition is a highly active research area due to a large number of potential applications such as healthcare, virtual reality, security & surveillance, human computer interface, and motion analysis. It can be divided into two categories, visual and ubiquitous sensing technologies [1, 2, 3]. In first category, camera based techniques are used to monitor the behavior of inhabitants. These are not practical due to privacy reason, day/night vision problem and jumble environment. In second category, ubiquitous sensors are deployed in smart home environment to recognize the daily life activities. They are cost-effective, easy to install and less intrusive to the privacy of inhabitants. Activity recognition is a big challenge by using ubiquitous sensors due to complex and highly diverse life styles.

In recent years, evolutionary methods [4, 5, 6] proved more robust and flexible for real world applications. In these methods, GA is one of the most successful technique, to obtain the high accuracy in reasonable amount of time. Furthermore, it solves the learning challenges posed by training phases due to imbalance datasets [5, 6]. This paper proposed an evolutionary ensembles approach integrated with GA to utilize its easy learning and robustness in the field of activity recognition.

During the training phase of activity recognition, minor activities are usually neglected by major activities in case of single learner approach. This problem arises due to few occurrences of minor activities in the datasets. For example instances of drink and breakfast activities are few as compared to washroom activity [7]. To give equal importance to each minor and major activity, we use the concept of ensemble learning. It oftenly achieves the higher accuracy as compare to single learner [8]. In the proposed approach, ensemble learning explores the search space in stochastic manner to extract the rule set. It designed individual learner for major and minor activity. For empirical evaluation, proposed approach has been tested on a real world dataset.

The remaining paper is structured as follows. In the next section we discussed the existing approaches for activity recognition. After that, we present our evolutionary ensembles approach. We explain the training and testing phase of proposed approach and empirical evaluation followed by discussion. Finally, the conclusion and future work are described.

RELATED WORK

These days, activity recognition in home environment is available at commercial level such as: Quiet Care Systems [9] and e-Neighbor [10]. Many researchers have proposed different machine learning techniques to recognize various kinds of activities.

SAGAware'11, September 18, 2011, Beijing, China.

Kasteren et al. [7] proposed a Bayesian probabilistic model and analyzed the various parameters to determine the effect on the performance. During experiments they observed that increasing the number of sensors have no high impact on accuracy.

Cook et al. [11] designed a model, which is based on Markov process using finite set of states. Its previous states depended process and transitions based on the probability values. They evaluated their approach in a smart apartment test bed.

Chen et al. [12] proposed a hybrid model of activity and preference model. They analyzed the relationship between these two models. The activity model maximized the expectation values for context selection and preference model developed the Dynamic Bayesian Network.

A. O. Puig et al. [4] investigated the capabilities of evolutionary systems. They proposed rule-based system for imbalance datasets and successfully solved the challenges posed by learning. These datasets have the tendency to cause small disjuncts and can mix few instances of one class with instances of another class. They showed that evolutionary systems are more robust and flexible as compared to instance based learners, decision trees, and support vector machines.

Some of the existing works [7, 12] do not consider the learning challenges posed by minor activities during training phase. The purpose of this paper is to develop more accurate and robust approach for recognizing the daily life activities as compared to single learner method. It gives equal importance to each activity and train as an individual learner. It has been evaluated on six different types of activities for both single learner and evolutionary ensemble approach.

THE PROPOSED APPROACH

Sensors are deployed in the home environment to make it smart for recognizing the daily life activities. The proposed approach comprises of two stages, namely, pre-processing and feature extraction (extraction of temporal feature vectors and uniformity of the sensory data), and evolutionary ensembles (learning and extraction of rule sets) as shown in Figure 1.

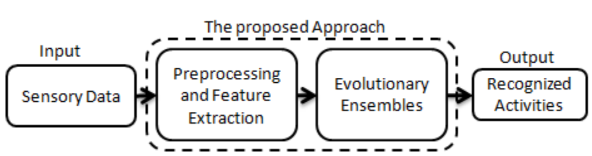

Figure 1. The main components of proposed approach

PREPOROCESSING AND FEATURE EXTRACTION

Different sensors in smart home environment sensed the data and stored into log files along with time stamp over the server. In order to apply evolutionary ensembles approach; unnecessary information is removed in preprocessing step like removing the multiple header lines from sensory and annotation data log files. In our model, feature vector is a synchronization of sensory data logs with annotation data logs and extraction of active sensors value with time stamp. The pseudocode for extracting the feature vector is described in Figure 2. This step labeled the sensory data for training of evolutionary ensembles.

```
Input: Rad: Read Annotation Data, Rsd: Read Sensory Data
Output: Active sensors feature vector
Feature Vector (Rsd , Rad)
begin
  for (each instance in Rad )
    adTimeStamp ← "Rad";     // Activity time stamp
    for (each instance in Rsd )
      sdTimeStamp ← "Rsd"; //Sensor data time stamp
      if (adTimeStamp ≥ sdTimeStamp)
        f vector_activeSensor[index,1] = SensorValue
        fvector_activeSensor[index, 2] = Time
      end if
    end for
  end for
end
```

Figure 2. Pseudocode for feature extraction

EVOLUTIONARY ENSEMBLES

The proposed approach is based on central and ensemble nodes. It learns the concept at each ensemble and combines on central node as shown in Figure 3.

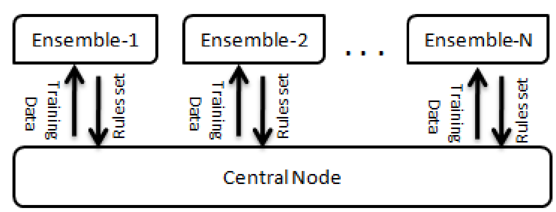

Figure 3. Evolutionary ensembles with central node

It consists of central node and n-evolutionary ensembles. The number of ensembles is dependent on the number of activities performed in the smart home environment. Detail of the components is given in the following sections.

CENTRAL NODE

It is responsible to provide the training data for ensembles and collects output in the form of rule set. After collecting the outputs of ensembles, some conflicting rules may exist by residing in more than one class. It is removed from the complete rule profile on the basis of majority class.

ENSEMBLES

Ensembles are used to develop an independent learning model for each activity, which treats the major and minor activities independently. The proposed ensembles are based on GA. The following sections discussed the design of ensembles.

Chromosome encoding: Chromosome is encoded by using Michigan [4] approach because it provides faster convergence rate for rule extraction. Chromosome is a combination of encoded sensor values followed by an activity class. The structure of the chromosome is as follows:

0	1	0	0	1	1	...	4

$$\underbrace{0\ 1\ 0\ 0\ 1\ 1\ \ldots}_{\text{Sensor values}}\ \underbrace{4}_{\text{Activity class}}$$

The size of chromosome is fixed and depends on the number of deployed sensors. Each sensor is modeled on the chromosome index. e.g., value 0 at index 1 represents the refrigerator sensor off state.

Stochastic operators

Initially random population is generated and following five steps are performed to obtain the next generation.

(1) Fitness function. In given population, it measures the quality of rules. We evaluated the fitness of each individual by following formula:

$$\text{F} = \sum_{i}^{n}\sum_{j}^{m}[reward\ (Candidate_i\ |example_j) - payoff\ (Candidate_i\ |example_j)] \quad (1)$$

where,

$$reward = \begin{cases} 1 & if\ candidate \equiv example\ \cap classLabel \equiv Correct, \\ 0 & otherewise. \end{cases}$$

$$payoff = \begin{cases} 1 & if\ candidate \equiv example\ \cap classLabel \equiv Incorrect, \\ 0 & otherewise. \end{cases}$$

The fitness function *"F"* evaluates the individual candidate rule. It is comprised on reward and payoff mechanism [13] to check the quality.

(2) Selection. It is an important operator for the selection of chromosome to generate the offspring in next generation. The ranked based selection [13] is applied for the selection of parent. In our case, one parent is selected from the rank based selection and the second is chosen randomly from the entire population for the better exploration of search space.

(3) Crossover and Mutation. The crossover operator exchanges the useful information in the new generation. A uniform single point crossover [13] has been applied as a reproduction operator. It keeps the same semantic of the chromosome after crossover operation. Mutation is applied to prevent premature convergence of the population [14]. In proposed approach, it inaugurates the diversity in rule set to increase the fitness of individual. The mutation operator is applied on randomly selected chromosome with locus after the crossover operator. It randomly chooses the value from 0 and 1.

(4) Elitism. Elitism preserves the good candidate rules of each population. The proposed approach maintains the elitism and shifts top ranked rules to the next generation and is discussed in empirical evaluation and discussion section.

(5) Stopping criteria. The stopping criterions for ensembles are; all training instances passed correctly, or for fixed number of generations is shown in last column of Table 1 and complete pseudocode for ensembles is depicted in Figure 4.

Input: **Set of training instances, Activities label, Stochastic operators**
Output: **Compact rule set**

```
begin
  nEnsembles ←count (activities label)        //generate ensembles
  foreach (nEnsembles)
    population ←  new random Population
    while (generation! = count || flagConverge)
      //Evaluating fitness function
      [flagConverge, pFitness] = fRankFitness(training
                                  data, activityLabel);
      if (flagConvergence) then
        return population;
      else
        while (length (new Population) = length (population)
          //Crossover operator
          parentOne = fRankedPupulation (pFitness)
          parentTwo = randomSelection (population)
          offSpring = fOnepointCrossover(parentOne, parentTwo);
          new population(index) = offSpring;
        end while
        //Mutation operator
        index = randomly select (chromosome);
        locusIndex = randomly Select (chromosome(index));
        gene = chromosome(index, locusIndex);
        mutatedChromsome = flip(gene);
        newPopulation (index)= replaced(mutatedChromosome)
        count = count +1;
      end if
    end while
  end foreach
end
```

Figure 4. Pseudocode for ensembles

TRAINING PHASE

The objective of training phase is to learn the daily living activities by extracting the compact rule set from sensory data as discussed in ensembles section. The training algorithm counts the number of performed activities and generates the equal number of ensembles. In each ensemble, stochastic operators are applied iteratively until stopping criteria matched. Finally, ensembles outputs are gathered at the central node.

TESTING PHASE

After training the ensembles, rule set has been maintained on the central node. It reflects most relevant rules over the training data for activity recognition. To test the trained model, input is the active sensors value at any instance of time. At central node, rule is activated on the basis of inputs to recognize the activity.

EMPERICAL EVALUATION AND DISCUSSION

To evaluate the performance of evolutionary ensembles approach, we use ISL dataset [7] that was recorded in intelligent systems laboratory and is publically available with its annotation. They collected data from fourteen binary sensors, which are attached to the doors; cupboards, a refrigerator, and a toilet flush. A volunteer performed common house hold activities during the 28 days period. These activities include leave house, prepare breakfast, dinner, and beverage, shower, and use washroom. These selected activities are based on the Katz ADL (Activities of Daily Life) index, to assess the physical capabilities of elderly people [15]. The 70% of the dataset is used for training while remaining 30% is used for testing and validation purpose. The parameters analyzed during the experimental phase of ensembles are depicted in Table 1. The last row of the table shows the optimal parameters of fitness function according to equation 1.

Parameters / Exp.	Size of Chromosome	Crossover Rate	Mutation Rate	Elitism Rate	No of Generations
1.	14	0.5	0.03	3	300
2.	14	0.4	0.02	2	250
3.	14	0.2	0.01	2	200
4.	14	0.2	0.005	2	150

Table 1. Parameters setting of ensembles

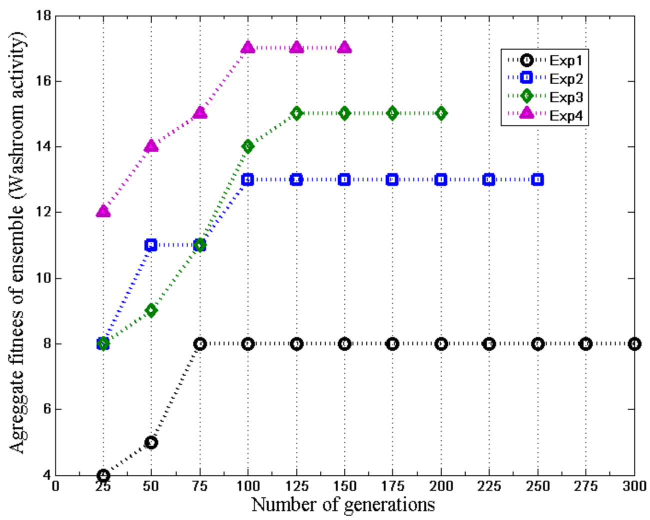

Figure 5. Activity ensemble convergence

Figure 5, shows the analysis of different parameters setting with the fitness function for single activity (i.e. Washroom). It is obvious that parameters of experiment 4 are stable and aggregate fitness is high as compared to the others. These optimal parameters are used in each ensemble and also for single GA. Before presenting our evolutionary ensembles approach, let us show single GA results, while treated as a single learner for all the activities. Table 2, shows the accuracy of single GA for each activity over 28 day's data.

Activities	Experiment results (%age)		
	1	2	3
Leave House	92.00	90.00	93.00
Shower	42.17	40.95	40.95
Sleeping	68.34	69.64	77.85
Breakfast	50.64	41.08	46.67
Dinner	43.92	44.8	48.04
Drink	49.98	47.98	46.9
Wash Room	70.65	66.91	66.91

Table 2. Individual accuracy of single GA

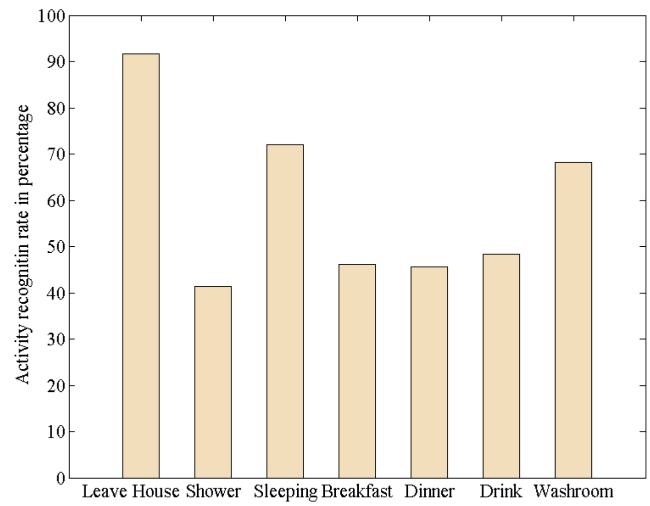

Figure 6. Average accuracy of single GA

Figure 6, shows the average accuracy of single GA experiments. In our approach, seven ensembles processed the data independently for daily living activities. Table 3, shows the accuracy of each activity determined by applying our evolutionary ensemble approach over 28 day's data.

Activity	Experiment Results (%age)		
	1	2	3
Leave House	100.00	100.00	100.00
Shower	48.65	45.00	53.21
Sleeping	99.00	98.00	99.00
Breakfast	61.98	62.50	62.48
Dinner	55.32	57.00	53.54
Drink	49.63	48.25	48.62
Wash Room	84.65	84.14	84.65

Table 3. Individual accuracy of evolutionary ensembles

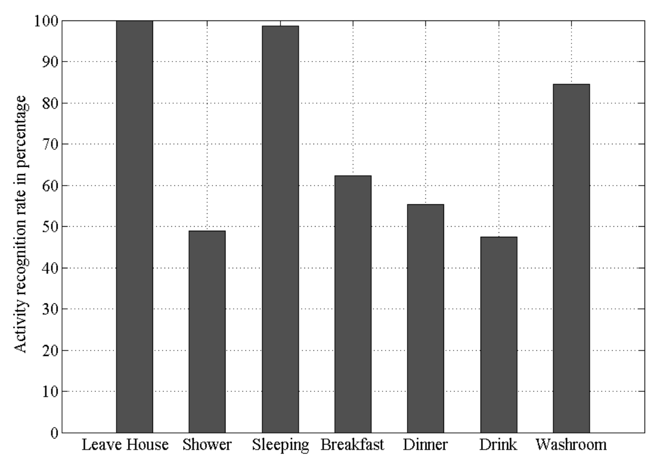

Figure 7. Average accuracy of evolutionary ensembles

Figure 7, shows the average accuracy of experiments by processing the information independently through evolutionary ensembles approach.

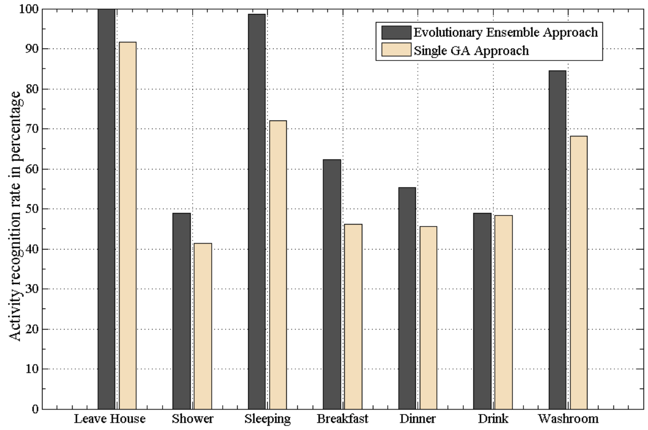

Figure 8. Comparisons of single GA and evolutionary ensembles

From Figure 8, it is obvious that evolutionary ensembles approach is more accurate. Our proposed approach processes the information independently and combined on central node for identification of daily life activities. Furthermore, it has a potential to extend for the other activities as new sensors are deployed in the environment.

CONCLUSION AND FUTURE WORK

In this paper, an evolutionary ensembles approach is proposed for recognizing the daily living activities. First we preprocessed the information and extract the features i.e. sensors values are based on activity time synchronization. Then the rules are extracted by evolutionary ensembles and further pooled at central node to predict the true class. We used the ISL dataset for testing, comprising of the six variant activities (leave house, take shower, sleeping, breakfast, dinner, drink and washroom). Our proposed approach is able to get the average accuracy 71.22% for recognizing the daily life activities.

Our future work will focus the accuracy of proposed approach and comparison with other activity recognition techniques for the development of stable framework. We also plan to test our approach on continuously sensing smart environments.

ACKNOWLEDGMENT

This research was supported by the MKE (The Ministry of Knowledge Economy), Korea under the ITRC (Information Technology Research Center) support program supervised by the NIPA (National Industry Promotion Agency) (NIPA-2010-(C1090-1021-0003)).

REFERENCES

1. Hoey, J., James J. Little Value-Directed Human Behavior Analysis from Video Using Partially Observable Markov Decision Processes. In *IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.29, no.7,* (2007) 1118-1132.
2. Turaga, P., Chellappa, R., Subrahmanian, V.S., Udrea, O. Machine Recognition of Human Activities: A Survey. In *IEEE Transaction on Circuits and Systems for Video Technology, Vol.18, no.11*, (2008) 1473-1488.
3. Logan B., Healey J., Philipose M., Tapia E. M., and Intille S. A Long-Term Evaluation of Sensing Modalities for Activity Recognition. In *UbiComp*, *LNCS 4717* (2007) 483-500.
4. Puig A. O., Mansilla E. B. Evolutionary rule-based systems for imbalanced data sets. In *Journal of Soft Computing - A Fusion of Foundations, Methodologies and Applications*. (2008) 213-225.
5. Maslov I. V., Gertner I. Multi-sensor fusion: an evolutionary algorithm approach. In *Journal of Information Fusion, vol. 7* (2006) 304-330.
6. Alberto F., Salvador G., Luengo J., Ester B., Herrer F. Genetics-Based Machine Learning for Rule Induction: State of the Art, Taxonomy, and Comparative Study. In *IEEE Transaction on Evolutionary Computation* (2009) 913 - 941.
7. Kasteren T., Noulas A., Englebienne G., Ben K. Accurate Activity Recognition in a Home Setting. In *UbiComp*, (2008) 21-24.
8. Diego, M. E., Miguel, A. R., and Antonio, P. An evolutionary ensemble-based method for rule extraction with distributed data. In *Lecture Notes in Computer Science* 5572, (2009) 638-645.
9. Quiet Care, http://www.security-tsi.com/ [July 10, 2011].
10. e-Neighbor, http://www.healthsense.com/ [July 10, 2011]
11. Rashidi, P., Cook, D. J., and Schmitter, E. M. Discovering Activities to Recognize and Track in a Smart Environment. In *IEEE Transaction on Knowledge and Data Engineering.* (2011) 527-539.
12. Chen Y., Lu C., Hsu K., Fu L., Yeh Y., and Kuo L. Preference Model Assisted Activity Recognition Learning in a Smart Home Environment. In *IEEE/RSJ International Conference on Intelligent Robots and Systems October*, (2009) 4657-4662.
13. Mitchell, T., Machine Learning, McGraw Hill. (1997)
14. Mendes, J. M., Resende, C. A random key based genetic algorithm for the resource constrained project scheduling problem. In *Journal of Elsevier Computers & Operations Research* (2009) 92-109.
15. Katz, S., Down, T.D., and Cash, H.R. Progress in the development of the index of adl. Gerontologist, (1970) 20-30.