A two-layer framework for activity recognition with multi-factor activity pheromone matrix

Liwen Xu*, Guoli Wang, and Xuemei Guo
Sun Yat-sen University, China

Abstract. With rapid population aging and increasingly indoor sensing technologies, mining effective information in sensor data is in need that we can analyse individual behaviour semantics, or even predict intentions. The model for indoor activity recognition (AR) is usually based on statistic while sensor data can impliedly reflect abundant information in order. Behaviour will trigger environment perception sensors. Inspired by information transmission in nature, persistent action keeps activity pheromone accumulating and inactive action keeps it volatilizing along with time shift. Different from statistic model, our framework proposes a method to construct multi factor features named activity pheromone matrix (APM). It has a double-layer model for recognizing daily activities include the high-overlapping. The experimental results show that our method can effectively promote the accuracy of activities recognition compared with the existing statistical models, even the high-overlapping activities in small areas.

1 Introduction

Ambient Assisted Living system is designed to help person get life supports and medical care by living in smart home. The activity analysis of daily living (ADLs) is an important part of the systems and divided into Activity Recognition (AR) and behavioural patterns discovery. Data is collected to model the behaviour patterns of target users, or guess the user’s behaviour intention in order to facilitate the system to adjust the response in advance. Sensor data are continuous time series. Commonly, artificial features can be divided into two main categories: statistical features and frequency-domain features. Statistical features are calculated by mean, variance, and covariance. Frequency-domain features are calculated by Fourier transform or wavelet transform. Traditional AR model features are often based on statistic features. However, a lot of hidden activity information is ignored and activity has been distorted. The accuracy of AR declines to a certain extent. How to fully excavate the hidden information of data ontology and construct effective features are difficulties of human indoor activity recognition. Generally, activity is contextually perceived and consists of sub-activities. Researchers regard sub-activity as atomic-level action [1,2]. The diversity of track makes it accidentally to generate new cluster during the training process and the ambiguity makes similarity between two activity clusters higher. Smaller inter-
cluster distance increases the probability of behaviour confusion. How to solve activity ambiguity is another difficulty of AR.

The wireless trigger-sensor network system is the experimental data collection environment of the target user daily activities in this paper. One record represents environmental change caused by a sub-activity. A complete activity has an active fragment, which is regarded as activity ontology and retains all original information. Researchers need to dig hidden data structures and relativity to construct explicit structures that can be used to maximize expression of activity. Hidden information contains behaviour-semantic related information, triggering order and temporal sequence. Most activity ambiguity refers to activity confusion in similar region. We define that high-overlapping activities have fragments with the same triggered sensors in the same region. For example, both washing and cooking will touch off the sensors in water channel or cupboard. Majority papers in ADLs based on sensor data show that traditional processing methods usually calculate the statistic distribution value. Feature vectors are constructed by some statistical numeric values, such as sensor-triggered count by space or time zone [3,4]. This model could have an effectively recognition, but simple construction of feature vector makes some identifiable information lost from the activity ontology.

In this paper, we propose a new framework based on the event trigger sensors data for AR. A novel method of feature construction is proposed. Multi-channel set different parameters for different activities, which makes the framework have a superiority to deal with high overlapping activity. The contribution of this paper about human activity recognition can be concluded as follows:

- **Activity Pheromone Matrix.** This paper proposes method of feature extraction from sensor data, to increase the inter-cluster distance for improving accuracy.
- **Time attenuation factor and time enhancement factor.** The method makes use of sensors spatial position, trigger frequency and trigger duration to construct model factors. Factors facilitate the distinction between high-overlapping activities.
- **Two-layer model with multi-channel.** According to the activity division, model sets personalized features to improve whole AR accuracy. It has reference significance.

### 2 Related works

With the trend of intelligent computing and ubiquitous computing, the paper number keeps increasing for AR based on location data or sensor-triggered data stream [5]. Tapia [6] proposed that Naive Bayes model could be used to classify and identify activities in tagged sensor data. Z Zhao [7] proposed an activity recognition self-adaption model for creating a personalized model for target users AR. K Rieping [3] drew on the experience of the text-language classification algorithm. Latent Dirichlet Allocation is a document-topic generation model consisting of words, themes and documents in three-layer structure. One record is regarded as one word in the special documents, and the activity sequence fragment could be seen as the topic of the documents. He segments the dataset into shape-let by a certain time window, construct time series pattern. In paper [8], researchers use sequential patterns to discovery the latent structure also through the very model. L Liu [9] change the ubiquitous sensors data into time series shape-let. W Jiahui [10] adopts the same dataset as we used in this paper. His major research is focused on implicit structure and association analysis between data through Hierarchical Dirichlet Process and Adaboost-HMM (Hidden Markov model) algorithm. He [11] also adopts the statistic vector to process the binary data. Like Johnson N [12] promoting in 1995, models calculate feature vector statistically from sensor data sequences in most paper, which. Our framework innovatively uses activity pheromone matrix as feature, which is described in section III.
High-overlapping activities share major triggered sensors and frequency so that the sequence form is similar. Motion overlapping causes motion self-occlusion [13], which is a hard question due to trajectory randomness. After comparing motion history image with directional motion history image, Mar Ahad [14] adopts seven higher order Hu moments to calculate feature vector in one action. V Nen [15] presented on-line version of the Minimum-distance classifier to achieve higher classification accuracy than static off-line one. Zhu C and Sheng W [16] process dataset ingeniously in multilevel and multi-dimension. One event sensor frequency differs from another, which was used to distinguish activities in same region.

3 FSDAR for sensor data to activities recognition

3.2 F1SDAR resume

High-overlapping activities data could be excavated to distinguish by distinctive frequency and duration time. This paper proposes a novel multi-granularity and multi-dimensional framework. As shown in figure 1, dataset is sliced into independent activity fragments called ontology to uniquely represent activity. In FSDAR, it prefers to combine location information and spatio-temporal factors to construct activity pheromone matrix as model features, rather than the feature vectors based on statistical rules in traditional methods. It also emphasizes context semantic relevance on sub-activities, which mapped to the data structure is between adjacent data. There is a two-layer model for coarse-grained and fine-grained classification. The coarse-grained classification model is used to identify whether the activity belongs to the high-overlapping. High-overlapping activity uses the feature which is the activity pheromone matrix with enhancement and attenuation time factors (APMTA) with signal TAF in figure 1. And activity pheromone matrix (APM) features if not. After changing the fragments into feature matrixes, those will be input fine-grained classification model, which is used for classifying activity category.

Fig.1. Framework for sensor-data to activities recognition, FSDAR.

3.2 Classification for high-overlapping activities

Coarse-grained classification model identifies whether the activities are highly overlapping. We add temporal constraint into an activity pheromone matrix of high overlapping activity to improve activity specificity. Without losing generality, sensors record is in formula 1.

\[ s_t = [d,t,t,v,(x,y)] \]
where \( d \) represents date and \( t_i \) represents time. \( v \) is binary value and \((x,y)\) represents the context-awareness application location. Environment-embedded Sensor is fixed position and trigger, which means an active behaviour in the vicinity. Activity ontology slice constructed represent as:

\[
A = \frac{1}{2} \sum_{t_i \in T} [1 + (-1)^{v_i}] s_i, \ i = 1, 2, \ldots, n
\]  

(2)

\( T \) is trigger duration time. Activities in one category are composed of specified sensors. First extract fragments from activity ontology. In text analysis, model extract keywords and feature vector though TF-IDF algorithm, Vector Space Model (VSM) respectively. The basic idea is to simplify documents as vector, value being weight of keywords. Similarity calculation between two documents uses cosine similarity formula in formula 3. Activity ambiguities are discriminant through activity similarity.

\[
\text{CosSim}(act_1, act_2) = \frac{\langle act_1 \times act_2 \rangle}{\|act_1\| \times \|act_2\|} = \frac{\sum_i (act_1 \times act_2)}{\left(\sum_i act_1^2 + \sum_i act_2^2\right)^{1/2}}
\]  

(3)

3.3 APM and APMTA construction

Animals will secrete an external hormone in order to leave some information to other individuals of the same species called pheromones. Traditional data process usually constructs statistic features via regional information and time shapelet, missing abundant information. Sequential fragments contain triggering duration time, rough position and sensor ID. Information has a specific order. The sequence transformation matrix of data structures and data characteristics can be viewed as information left over from the activities called active pheromone matrix. Data structure is not sensitive to the isolated point. Similarity is a common barrier for AR. Both sequence and continuous sensor-triggered duration increase the self-characteristics between various categories. It effectively improves the recognition accuracy for high-overlapping activities. The value “1” returned indicates that the sensor triggering state starts, while “0” indicates that the sensor trigger state is over. We consume that \( i \)-type sensor was triggered in the \( k1 \) record and ended in \( k2 \) record, \( s_{k1} = [d_{k1}, t_{k1}, k1, 0, (x_{k1}, y_{k1})] \), \( s_{k2} = [d_{k2}, t_{k2}, k2, 0, (x_{k2}, y_{k2})] \). From \( v_{k1} = 0 \) to \( v_{k2} = 0 \), the difference value between time-stamp represents a complete sensor-triggered duration. For high-overlapping activities in small region, they have similar sensor set and frequency while sequence and duration obvious difference.

Extracting event data from the raw sequence ontology, we add the temporal coefficient to preserve triggering pheromone. Figure 2 shows the time factors how to play in pheromone matrix generation. For the \( t \) moment, \( W_{t-1} \) represents historical pheromone and \( W_t \) represents synthesized pheromone. \( C_t \) represent newly generated pheromone from the real-time container that temporarily save the triggering sensor type. When remains in triggering state, sensors in container will continuously accumulate pheromone with the time coefficient combined factor \( \sigma \). As triggering state have been over, position will be removed from such a container and the corresponding pheromone begin to attenuate with the time coefficient combined factor \( \rho \). Both of the two temporal factors enhance the impact of main triggered sensor. This structure looks like a simplified version of Long Short-Term Memory (LSTM) with a forget gate to reduce historical memory. The specific formula is as follows:

\[
W_t = k(1 + \sigma)^{\Delta T} \times C_t + (1 - \rho)^{\Delta T} \times W_{t-1}
\]

(4)

\[
\Delta T = T_t - T_{t-1}
\]

(5)
The APM with time factors increases distance between samples. Experimental results of
validation effectiveness are shown in next section. We assume that $W_t$ represents the
pheromone before time $t$ and the $C_t$ represents the immediate pheromone matrix at time $t$.
At the next moment, APM is shown below:

$$W_{t+1} = W_t + k \times C_t$$  \hspace{1cm} (6)

And the APM with time factors is shown below:

$$\bar{W}_{t+1} = (1 - \rho)^{\Delta(T_{t+1} - T_t)}\bar{W}_t + k \times (1 + \sigma)^{\Delta(T_{t+1} - T_t)}C_t$$  \hspace{1cm} (7)

After n moments, $W_t$ has changed into APM and APM-TA respectively shown below
after iteration:

$$W_{t+n} = W_t + k \times \sum_{i=t}^{t+n-1} C_i$$  \hspace{1cm} (8)

$$\bar{W}_{t+n} = (1 - \rho)^{\Delta(T_{t+n} - T_t)}\bar{W}_t + k \times \sum_{i=t}^{t+n-1} (1 + \sigma)^{\Delta(T_{t+1} - T_t)}C_i (1 - \rho)^{\Delta(T_{t+n} - T_i)}$$  \hspace{1cm} (9)

4 Experiment

4.1 Dataset

Some well-known datasets for AR are open on the Internet, such as CASAS [17,18],
KASTEREN-dataset [19], WSUCASAS [20] and so on. We model on Aruba dataset. All
data pre-process and model training is based on Python 2.7 language with the Core(TM)
i5–6400T processor. Three sensor categories are in the dataset with the respectively
symbols M, D, T. Records are sensor data triggered by an individual’s behaviour for 220
days. The numerical difference of samples and records from different activity categories is
quite large. The active category named Resperate only has 6 samples while Relax has 2919
samples in total. On the other hand, the type of triggered sensors is also quite different.
Data imbalance is an important reason to affect accuracy of AR.

4.2 High-overlapping recognition

We first extract the activity fragments from the active ontology. In text analysis, TF-IDF
algorithm extracts the keyword of the article, and the Vector Space Model(VSM) model
constructs the feature vector of keywords. the VSM model is used to extract uniquely
feature representation vector from the active fragment. The smaller the angle between the
vectors is, the more similar the active fragments are. Key words response sensor types,
values response sensor-triggered frequency. Activities similarity is shown below in figure 3.
As we can see from the similarity matrix, Enter_Home is highly similar to Leave_Home in
0.869, and Wash_Dishes is highly similar to Meal_Preparation in 0.785. From the
perspective of life experience, it is logical because activities are similar in its scope and significance.

|                | Enter_Home | Work | Eating | Housekeeping | Relax | Wash_Dishes | Sleeping | Leave_Home | Bed_to_Toilet | Meal_Preparation |
|----------------|------------|------|--------|--------------|-------|-------------|----------|------------|---------------|-----------------|
| Enter_Home     | 1.0000     | 0.0023 | 0.0000 | 0.0083       | 0.0000 | 0.0000      | 0.0000   | 0.008690   | 0.0000        | 0.0161          |
| Work           | 0.0023     | 1.0000 | 0.0011 | 0.0448       | 0.0220 | 0.0039      | 0.0015   | 0.0214     | 0.0133        | 0.0044          |
| Eating         | 0.0000     | 0.0011 | 1.0000 | 0.2751       | 0.094  | 0.0407      | 0.0000   | 0.0000     | 0.0000        | 0.1229          |
| Housekeeping   | 0.0083     | 0.0448 | 0.2751 | 1.0000       | 0.2258 | 0.02094     | 0.0827   | 0.0119     | 0.2337        | 0.1791          |
| Relax          | 0.0000     | 0.0220 | 0.0084 | 0.2258       | 1.0000 | 0.0034      | 0.0010   | 1.0000     | 0.0000        | 0.0013          |
| Wash_Dishes    | 0.0000     | 0.0039 | 0.0407 | 0.2094       | 0.0034 | 1.0000      | 0.0010   | 0.0000     | 0.0000        | 0.7846          |
| Sleeping       | 0.0000     | 0.0015 | 0.0000 | 0.0827       | 0.0000 | 0.0010      | 1.0000   | 0.0000     | 0.0000        | 0.0000          |
| Leave_Home     | 0.8690     | 0.2014 | 0.0000 | 0.0119       | 0.0000 | 0.0010      | 1.0000   | 0.0000     | 0.0000        | 0.0153          |
| Bed_to_Toilet  | 0.0000     | 0.0133 | 0.2337 | 0.0000       | 0.0000 | 0.0680      | 0.0000   | 1.0000     | 0.0000        | 1.0000          |
| Meal_Preparation| 0.0161    | 0.0044 | 0.1229 | 0.1791       | 0.7846 | 0.0010      | 0.0153   | 1.0000     | 0.0000        | 1.0000          |

Fig. 3. Similarity matrix of training samples. This is a symmetric matrix to describe the similarity.

4.3 Time factors

The purpose of using two-layer model is to classify the easy confusing activities on the coarse-grained model, and accordingly decide whether activity fragment is charged to be APM or APMTA as the fine-grained model feature. Compared to APM, APMTA is added enhancement and attenuation time factors. As shown in figure 4, Meal_Preparation and Wash_Dishes are used as confusing activity group in the experiment. Compare numerical values to see whether distance between two similar activities have increased through temporal factors. Active specificity increased so that active similarity decreased. Enhancement time factor \( \sigma \) and attenuated time factor \( \rho \) are constant of the exponential function with time as variable. Factor \( \sigma \) and \( \rho \) restraint each other. If \( \sigma \) is too large, the current pheromone effect is also large. It causes prophase distortion. If \( \rho \) is too large, pre-existing pheromone have been excessively volatilized, and that easily leads to post distortion. Whether it is prophase distortion or post distortion, features cannot accurately express activity.

Fig. 4. High-overlapping active segments distance in Euclidean Measure and Cosine Similarity.

4.4 Experiment analysis

Our paper proposed a novel feature matrix APM compared to traditional feature vector, its effectiveness stemming from its more implicit active information to enhance the identification performance of activities. We first extract a multi-dimensional column vector as features. Five dimensions represent the total number of spatial triggering sensors in different regions of the active fragments. Another dimension is the number of sensor types triggered and the last dimension is the duration time.
\[ V_i = [n_{i1}, n_{i2}, n_{i3}, n_{i4}, n_{i5}, N_{i}, t_i]^T \]  

In the comparative experiment, all the active ontology fragment was transformed into APM feature matrices with different models. Activities completed in overlapped regions generally have an inferior performance in figure 5(a) while the activity recognition accuracy has been significantly improved intuitively in figure 5(b). However, the performance of activity Wash Dishes is still very low because the positive and negative samples are out of proportion. Its totality is 0.04% in the proportion of Meal Preparation. Selecting algorithm non-sensitive to the unbalanced dataset or oversampling minority activity is effective to reduce the impact of dataset unbalanced. Both Naive Bayes algorithm and Decision Tree algorithm will have better performance on minor-sample events than others.

![Fig. 5](image-url) Single-layer AR model with input feature: feature vector based on statistic and feature matrix based on activity pheromone. There are five common classifiers to identify: Support Vector Machine, Decision Tree, K-Nearest Neighbours, Logistic Regression, Naive Bayes.

![Fig. 6](image-url) The experimental results with different active pheromone matrices. The left one is the result of single layer model with APMTA, with factors \( \sigma = 0.1 \) and \( \rho = 0.15 \); the right one is the result of dual-layer model with APMTA, with factors \( \sigma = 0.1 \) and \( \rho = 0.15 \).

Table 1. Framework two-layer model recognition performance. This is a table that partial parameters combination and model accuracy were displayed.

| \( \rho \) | 0.000 | 0.050 | 0.100 | 0.150 | 0.000 | 0.000 | 0.050 | 0.100 | 0.150 | 0.100 |
| \( \sigma \) | 0.000 | 0.000 | 0.000 | 0.000 | 0.150 | 0.200 | 0.250 | 0.100 | 0.100 | 0.200 |
| Precision | 0.732 | 0.758 | 0.742 | 0.759 | 0.746 | 0.750 | 0.744 | 0.764 | 0.738 | 0.751 | 0.766 |
| Recall | 0.795 | 0.835 | 0.814 | 0.837 | 0.824 | 0.828 | 0.818 | 0.828 | 0.810 | 0.829 | 0.843 |
| F1 score | 0.720 | 0.775 | 0.752 | 0.777 | 0.765 | 0.769 | 0.758 | 0.779 | 0.749 | 0.770 | 0.784 |

Compared to the two-layer model, the single-layer model cannot adaptively identify high-overlap activities and adopt multi-channel processing. As shown in figure 6, if all the features are in the form of APMTA, the recognition accuracy of some activities with highly discrimination turns into lower. The reason is the time factors will lead to the former distortion, weaken the specificity of the isolated activity. Model is sensitive to attenuation factor and enhancement factor. As results shown in figure 6,
the accuracy of easy confusing activities do improve. When the time factor is added, the effect promotion is universal. For different parameter combination of attenuation factors and enhancement factor, the model effect is different. The framework model performance evaluation according to the parameter settings is shown in table 1. The APMTA is equal to APM as both $\rho$ and $\sigma$ are zero. Obviously, the experimental results of APMTA are superior to the APM on precision, recall, and F1 score.

5 Conclusion

In this paper, a novel framework FSDAR is proposed for activity recognition in WSN. It can apply to wearable devices or indoor location-based for the method of feature matrix construction being general.

Traditional statistic feature vector can be extract easily with abundant information losing. Pheromones matrix combines spatial and temporal information to distinguish activities, even the high-overlappings. Results show that feature matrix has a common improvising than feature vector. Temporal factors emphasize the tiny specificity so that it effectively increase the sample distance between similar activities. The framework chooses a two-layer model with multi-channel to avoid time factor weakening specificity of simple activity with few core sensors. Through datasets effective test, the experiments show that novel feature extraction improve accuracy of recognition. The appropriate time factor effectively increases the cluster distance. The experiment results emphasize the effectiveness of our new framework comparing with traditional AR model.

References

1. Atallah L and Louis B, et al. Sensor Positioning for Activity Recognition Using Wearable Accelerometers IEEE Transactions on Biomedical Circuits & Systems 5.4(2011):320-329.
2. Jin S Y and H J Essential Body-Joint and Atomic Action Detection for Human Activity Recognition Using Longest Common Subsequence Algorithm. International Conference on Computer Vision Springer-Verlag 2012:148-159.
3. Rieping K, Englebienne G, and Kröse B. Behavior analysis of elderly using topic models Pervasive & Mobile Computing 15 C(2014):181-199.
4. Sadek S, Alhamadi A and Michaelis B, et al. A Fast Statistical Approach for Human Activity Recognition International Journal of Intelligence Science 2.1(2012):9-15.
5. Chen L, Hoey J, Nugent C D, et al. Sensor-Based Activity Recognition[J] IEEE Transactions on Systems Man & Cybernetics Part C 2012 42(6):790-808.
6. Tapia E M, Intille S S, Larson K Activity Recognition in the Home Using Simple and Ubiquitous Sensors Pervasive Computing Springer Berlin Heidelberg 2003:158-175.
7. Zhao Z, Chen Y and Liu J, et al. Cross-people mobile-phone based activity recognition International Joint Conference on Artificial Intelligence AAAI Press 2011:2545-2550.
8. Chikhaoui B, Wang S, Pigot H. ADR-SPLDA: Activity discovery and recognition by combining sequential patterns and latent Dirichlet allocation[J] Pervasive & Mobile Computing 2012 8(6):845-862.
9. Liu L, Peng Y and Liu M, et al. Sensor-based human activity recognition system with a multilayered model using time series shapelets[J] Knowledge-Based Systems 2015, 90(C):138-152.
10. Wen J, Indulska J, Wang Z. Discovering Latent Structures for Activity Recognition in Smart Environments[C] Ubiquitous Intelligence and Computing 2014 IEEE Intl Conf
on and IEEE, Intl Conf on and Autonomic and Trusted Computing and IEEE, Intl Conf on Scalable Computing and Communications and ITS Associated Workshops. IEEE 2014:140-147.

11. Wen J, Zhong M, Wang Z. Activity recognition with weighted frequent patterns mining in smart environments[J]. Expert Systems with Applications 2015 42(17–18):6423-6432.

12. Johnson N, Hogg D. Learning the distribution of object trajectories for event recognition[C] British Conference on Machine Vision BMVA Press 1995:583-592.

13. Ahad M A R, Tan J K, Kim H S, et al. Analysis of Motion Self-Occlusion Problem Due to Motion Overwriting for Human Activity Recognition[J] Journal of Multimedia 2010 5(1).

14. [14] Ahad A R, Ogata T, Tan J K, et al. Performance of Multi-directional MHI for Human Motion Recognition in the Presence of Outliers[J] Conference of the IEEE 2007:2366-2370.

15. Könönen V, Mäntyjärvi J, Similä H, et al. Automatic feature selection for context recognition in mobile devices[J] Pervasive & Mobile Computing 2010 6(2):181-197.

16. Zhu C, Sheng W. Motion- and location-based online human daily activity recognition[J] Pervasive & Mobile Computing 2011 7(2):256-269.

17. Cook D, Crandall A, Thomas B, et al. CASAS: A Smart Home in a Box[J] Computer 2013 46(7):62-69.

18. Cook D. [n. d.]. CASAS smart home project. http: www.tp-ontrol.hu/index.php/TP Toolbox ([n. d.]).

19. Kasteren T V, Noulas A, Englebienne G. Accurate activity recognition in a home setting[C] International Conference on Ubiquitous Computing ACM 2008:1-9.

20. Cook D. [n. d.]. WSU Datasets. http: ailab.wsu.edu/casas/datasets/index.html. ([n. d.]).