Classifying Activity Patterns using Activity Graphs from Sensors

Kisung Park, Yongkoo Han, and Young-Koo Lee

Department of Computer Engineering, Kyung Hee University,
Yongin-si, Gyeonggi-do, Korea
{kspark, ykhan, yklee}@khu.ac.kr

Abstract. Advancements in sensor technologies give us the opportunity to recognize activities of daily living. Activity pattern classification is a technique that predicts class labels of people based on activity patterns. Because people have both intrinsic and common activity patterns, mining discriminative features reflecting the intrinsic patterns is important for this classification problem. In this paper, we propose an effective method for classifying activity patterns. In order to mine discriminative features, we represent activities as a graph model and mine activity patterns in various periods. Experiments show the proposed method achieves high classification accuracy compared with existing graph classification techniques.

Keywords: Sensor, Activity, Classification.

1 Introduction

The advances of sensor technology make activity recognition possible. Activity recognition is a technique that automatically recognizes human activities by analyzing sensor data. Recently several studies [1-2] have been performed to analyze activity patterns from an activity database gathered by activity recognition.

Activity pattern classification is a technique that predicts class labels of people based on activity patterns. The class labels can be not only individual identification but also meaningful groups such as nationalities, genders, jobs, and hobbies. Therefore, activity pattern classification has many applications varying the class labels accordingly. For example, recommender systems can recommend similar items to users with the same hobby.

Activity patterns are styles in which people perform their activities and they reflect people’s life styles. People have both intrinsic and common activity patterns. The intrinsic patterns can be distinguished depending on activities in different frequencies [3-4], orders, and periods [5]. Therefore, in order to mine discriminative features, we need to explore the activity patterns by adjusting the frequencies in various periods.

Graph based modeling techniques [6] for activity data have been proposed. These modeling techniques generate activity graphs in various periods such as daily, weekly, monthly, and yearly. Nodes and edges of the activity graphs represent activities and the occurrence order between two activities, respectively. The nodes and edges have activity related multi-labels such as frequencies, time, durations, and locations of
activities. Therefore, graph based modeling is suitable to mine discriminative features since the modeling technique can reflect frequency, order, and period of activity patterns.

Then we can use graph classification techniques [7-8] to classify the activity graphs. The techniques represent graphs as frequent subgraph features and classify the graphs by using a machine learning classifier. However, they cannot achieve the high classification performance for activity graphs since they consider only frequency of the features.

Activity graphs modeled in different periods have different frequent subgraphs. Though some frequent subgraphs can be overlapped in activity graphs with different periods, some frequent subgraphs only occur in a specific period. Therefore, a novel classification is required to classify activity graphs considering the human activity patterns.

In this paper, we propose an effective method for classifying activity patterns using activity graphs. The proposed method represents activity data as a graph model with two kinds of periods, modeling and feature period. We generate activity graphs with feature periods having the range one to the modeling period so that we can mine discriminative features in all combinations of activity patterns in the modeling period. These discriminative features can be used as informative features for activity pattern classification since these patterns reflect individuals’ intrinsic lifestyles. Through experiments, we show that the proposed method can classify activity data better than existing graph classification methods.

2 Related Work

Various studies have been proposed to represent activity data collected from sensors such as statistics, sequence, and graph based modeling. Statistics based activity models [1] calculate average frequency and duration of each activity. Sequence based activity models [2] represent activities as daily sequences based on the occurrence time. Graph based activity models [6] generate activity graphs by combining the daily sequences in every monitoring period. The main advantage of this graph based activity model is that we can analyze activity pattern in various periods.

Graph classification studies [7-8] have been proposed. The existing techniques mostly adopt frequent subgraphs as features for classification. Many efficient frequent subgraph mining algorithms have been proposed such as FSG [3] and gSpan [4]. Though existing studies successfully classify graphs by using frequent subgraphs, we cannot directly apply the existing technique for activity graphs since activity graphs have a temporal characteristic.

3 Activity Graphs

In this section, we introduce how to generate activity graphs based on the graph model [6]. People perform similar activity patterns every day. We can regard the daily activity pattern as a sequence of activities such as \( S = a_1^{m_1} \rightarrow a_2^{m_2} \rightarrow \cdots \rightarrow a_n^{m_n} \),
where \( a_i^t \) represents an activity \( a_i \) occurred at time \( t \) and \( t_i < t_{i+1} \). Therefore, activity sequences are generated in each day. In order to represent activities of more than one day concisely, we combine corresponding activity sequences and generate activity graphs. The formal definition of an activity graph is shown in Definition 1.

**Definition 1. Activity graph** An activity graph \( G = (V, E, L, l) \) consists of a set \( V \) of activity nodes and a set \( E \) of edges, where an edge \( e \in E \) represents the order between two activity nodes in \( V \). \( L \) is a set of node and edge labels and \( l \) is a function assigning labels to nodes and edges.

We adopt multiple sequence alignment (MSA) [9] to combine the activity sequences among various graph modeling approaches. The number of combining sequences are determined depend on the modeling period \( p \) that we want to observe. For example, we generate weekly activity graphs by combining activity sequences every seven days when \( p = 7 \).

### 4 Classifying Activity Graphs based on Activity Patterns

A set of graph \( G_p = \{g_1, g_2, ..., g_n\} \) is the activity graphs with modeling period \( p \), where \( g_{i+1} \) has \( p \) days later activity patterns from \( g_i \). We define the activity graph classification problem in Definition 2.

**Definition 2. Activity graph classification** Given the set of activity graphs \( G_p = \{g_1, g_2, ..., g_n\} \) for people, and a set of class labels \( \{y_1, y_2, ..., y_m\} \), an activity graph classification is the problem that predicts the class label \( y_i \) for the activity graph \( g_i \).

**Definition 3. Feature period** Given the modeling period \( p \) and the activity sequences in \( p \), \( S = \{s_1, s_2, ..., s_p\} \), feature period \( p_f \) is the period in the range of 1 to \( p \).

In order to mine discriminative features of activity patterns for this classification problem, we find frequent activity patterns from activity graphs generated in various feature periods. A frequent activity pattern \( f \) is a subgraph that appears no less than the minimum support threshold in the activity graph set. It has been shown in [7-8] that frequent subgraphs can have high discriminative powers for classification.

Given the set of activity sequences \( S = \{s_1, s_2, ..., s_p\} \) and the modeling period \( p \), we generate activity graphs \( G_1, G_2, ..., G_p \) with the feature period, from one to \( p \). Then, we mine frequent activity patterns \( F_1, F_2, ..., F_p \) by performing frequent subgraph mining to each graph set generated in \( p_f \). Figure 1 shows the processing steps of frequent activity patterns.

The activity graphs \( G_1, G_2, ..., G_p \) have many common subgraphs since the graphs are generated from the same sequence set. Therefore, a number of duplicate frequent patterns are extracted. For example, the number of graphs in \( G_i \) is \( mC_i \) since we generate one graph by combining \( i \) activity sequences in Figure 1. Many of the graphs are similar to each other. These duplicate subgraphs cause to increase the running time and degrade the performance of classification. We select top-\( k \) patterns having high discriminative power among the extracted frequent activity patterns.

discriminative power can be estimated by feature evaluation functions such as information gain.

Fig.1. Processing steps of mining frequent activity pattern.

We represent graphs as feature vectors with the top-k frequent activity patterns. We make a classifier model by training machine learning algorithm, such as the support vector machine and the decision tree, with the graphs assigned the class labels.

5 Experiments

We use a real activity dataset for experiments. The dataset is gathered by 8 Kyung Hee university students for a year. The dataset for each student contains more than 2,000 activities. We generate activity graphs, where the modeling period p is set to 7. We performed 2-folds cross validation using an SVM classifier. We compared the classification performance between top-k and the proposed method.

Figure 2 shows the results of classifying activities with different number of selected features. In the all settings, the proposed method outperforms the existing top-K. Especially, the proposed method achieves high performances when the number of features is set to around 80.

Fig.2. Processing steps of mining frequent activity pattern.
5 Conclusion

Classification is an important technique for analyzing activity data. An effective method for classifying activity patterns has been proposed. We represent activity data as a graph model to mine discriminative features of activity patterns in various periods. These discriminative features can be used as informative features for activity pattern classification. Experimental results have shown the proposed method can classify activity data better than existing graph classification methods.

Acknowledgements. This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MEST) (No.2012R1A2A2A01047478).

References

5. Han, J., Dong, G. and Yin, Y.: Efficient Mining of Partial Periodic Patterns in Time Series Database. In: Proceeding of the 15th International Conference on Data Engineering, Sydney, Australia (1999)