Unsupervised Activity Discovery and Characterization From Event-Streams

Raffay Hamid, Siddhartha Maddi, Amos Johnson, Aaron Bobick, Irfan Essa, Charles Isbell
College of Computing / GVU Center, Georgia Institute of Technology, Atlanta, GA 30332-0280 USA

Topic: estimation, prediction, and sequence modeling. **Oral/Poster**

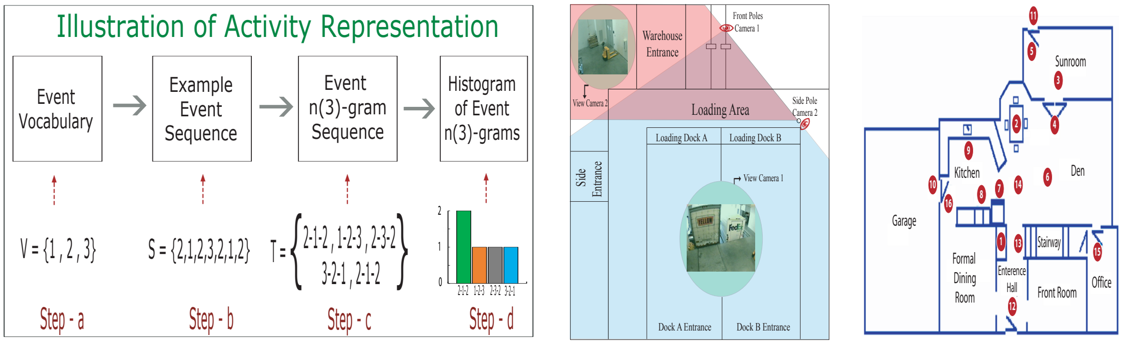
Introduction: Recognizing what is happening in an environment has many potential applications, ranging from automatic surveillance systems to supporting users in ubiquitous environments. A key step to this end is to discover the kinds of similar activities that frequently occur in a particular domain. Equally important is the question of finding efficient characterizations for these different kinds of activities. We are interested in the study of *activity class discovery and characterization*, in the context of analyzing everyday activities. We present a novel representation of activities as bags of discrete n-grams, . We then demonstrate how disjunctive activity groups can be discovered in an unsupervised manner. Finally, we lay out a framework for unsupervised discovery of predictably recurrent event motifs for activity class characterization.

Representation: We introduce a novel representation of activities as bags of discrete event n-grams of events - a perspective different from the previously used grammar driven approaches. This treatment of an activity is similar to the representation of a document as a set of words - also known as the Vector Space Model (VSM) [3], in which a document is represented as a vector of its word-counts, in the space of possible words. This representation allows us to analyze the global structural information of the activities by simply considering its local event statistics. To use such a scheme, we must define a set of possible events ($event\ vocabulary$) that could take place in the situation under consideration. To capture the temporal ordering of events in activities in a better way, we consider histograms of higher order event n-grams, where we represent an activity by a (sparse) vector of counts of overlapping event n-grams in a (very) high dimensional space of possible event n-grams (see figure 1 (A).)

Activity Similarity: Based on this activity representation, we formalize the notion of similarity between two activities, taking into account their core structural and event-frequency based differences. The core structural differences relate to the distinct n-grams that occurred in either one of the sequences in a sequence-pair, but not in both. We believe that for such differences, the number of these mutually exclusive n-grams is of fundamental interest. On the other hand, if a particular n-gram is inclusive in both the sequences, the only discrimination that can be drawn between the sequence pair is purely based on the frequency of the occurrence of that n-gram.

Activity Discovery: We start off by assuming that we are initially given a set of K activities. We consider this activity set as undirected edge-weighted graph, with K nodes, each representing a histogram of n-grams of one of the K activities. The weights of the graph nodes represent the similarity between pairs of nodes as defined earlier. We formalize the problem of discovering sub-classes of activities as searching for maximal cliques in the graph of K activities [2]. We proceed by finding the maximal clique in the graph, removing that set of nodes from the graph, and repeating this process iteratively with the remaining set of nodes, until we are left with the few nodes that did not belong to any of the discovered sub-class.

Event Motif Discovery Our proposed scheme discovers activity-classes in an unsupervised manner, and finds patterns that are maximally mutually exclusive amongst activity-classes. From the perspective of activity discovery and recognition, we are interested in frequently occurring event-sequences that are useful in predicting future events, and can therefore be used for activity class characterization. We assume that a class of activity-sequences can be modeled as a variable-memory Markov chain (*VMMC*) [1]. We define an event-motif for an activity-class as one of the variable-memory elements of its *VMMC*. We want to find the sub-sequences which can efficiently characterize a particular class, while having minimal representation in other classes. We find all such sub-sequences in linear using the *Prediction Suffix Tree* data structure.



(A) **(B)** **(C)**

Figure 1: (A) Transformation of an example activity from sequence of discrete events to histogram of event n-grams. Here the value of n is shown to be equal to 3. V is event vocabulary, S is event sequence, and T is sequence of overlapping n-grams. Step-d shows the non-zero n-gram counts of V.(B) A schematic diagram of the camera setup at the loading dock area with overlapping fields of view (FOV). (C) A schematic diagram of the strain-gage setup in the house scenario. The red dots represents the positions of the strain gages.

Experiments and Results: To test our proposed algorithms, we performed experiments on data-sets collected from two active environments, i.e. the Loading Dock area of a retail bookstore, and a House environment. A schematic diagram representing both these environments is shown in Figure 1 and Figure 1. For both of our experimental setups, we chose the value of n for the n-grams to be equal to 3 (*tri-grams*). Of the 150 training activities, we found 7 classes (maximal cliques), with 106 activities as part of any one of the discovered class, while 44 activities being different enough to be not included into any non-trivial maximal clique. Of the 151 activities captured over a little more than 5 months, we found 5 activity-classes (maximal cliques), with 131 activities as members of any one of the discovered class, and 20 activities being dissimilar enough not to be a part of any non-trivial maximal clique.

The discovered activity-classes both for the Loading Dock and the House data-sets, are semantically coherent and divide their respective activity space quite efficiently. The fundamental differences between various classes are dictated by the fact whether the activities were of delivery or pick-up, how many people were involved in the activity, how many packages were moved, and what type of delivery vehicle was used. For the house environment, these differences consist of how long does a person stay in the house, and what time of the year it is.

The discovered motifs of membership classes efficiently characterize these classes. The discovered motifs for activity-classes where package *delivery* occurred, have events that are particularly related to the process of delivering packages. A similar characterization of the *pick-up* activities is done by the representative discovered event-motifs. The motifs for the House environment capture the position where the person spends most of her time and the order in which she visits the different places in the house.

Conclusions: Our discovered activity-classes are semantically meaningful, which implies that our defined similarity metric is an efficient way to compute the correspondence between event-sequences and that the framework of dominant sets is useful in transforming the domain knowledge embedded in the system in terms of our defined vocabulary in a meaningful manner. We demonstrated how variable-memory Markov chains can be used to extract event-motifs that can compactly characterize activity-classes.

References

[1] Weinberger et al. A universal finite memory source. In *IEEE Trans. Inform. Theory*, 1995.

[2] M. Pavan and M. Pelillo. A new graph-theoretic approach to clustering and segmentation. In *CVPR*, 2003.

[3] G. Salton. *The SMART Retrieval System - Experiment in Automatic Document Processing*. Prentice-Hall, Englewood Cliffs, New Jersey, 1971.