Unsupervised Activity Discovery and Characterization for Sensor-Rich Environments

A Thesis
Presented to
The Academic Faculty

by

Raffay Hamid

In Partial Fulfillment
of the Requirements for the Degree
Master of Science

College of Computing
Georgia Institute of Technology
December 2005

Unsupervised Activity Discovery and Characterization for Sensor-Rich Environments

Approved by:

Dr. Aaron Bobick
College of Computing
Georgia Institute of Technology, Advisor

Dr. Irfan Essa
College of Computing
Georgia Institute of Technology

Dr. Charles Isbell
College of Computing
Georgia Institute of Technology

Date Approved

To Ammi, Abbu, and Ayesha.

ACKNOWLEDGEMENTS

Many thanks indeed to my advisor, Aaron Bobick for all his insight and inspiration over the course of this thesis. I cannot thank Amos Johnson enough, who helped me immensely at each step of this work. Thanks to Samir Batta, Siddhartha Maddi and Graham Coleman for working so hard towards making this project a success. Many thanks to Irfan Essa, Charles Isbell and Jim Rehg for all their time and helpful suggestions. I want to thank Vivek Kwatra, Gabrial Brostow and Drew Steedly, whose advise always put me in good stead. Thanks to all my colleagues at Computational Perception Lab and Mobile Robot Lab for inspiring me to think about my research problem from different perspectives.

Without the love and support of my family, I never could have made it through this process.

Thank you all so very much.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

This thesis presents an unsupervised method for discovering and analyzing the different kinds of activities in an active environment. Drawing from natural language processing, a novel representation of activities as bags of event n-grams is introduced, where the global structural information of activities using their local event statistics is analyzed. It is demonstrated how maximal cliques in an undirected edge-weighted graph of activities, can be used in an unsupervised manner, to discover the different activity-classes. Taking on some work done in computer networks and bio-informatics, it is shown how to characterize these discovered activity-classes from a wholestic as well as a by-parts view-point. A definition of anomalous activities is formulated along with a way to detect them based on the difference of an activity instance from each of the discovered activity-classes. Finally, an information theoretic method to explain the detected anomalies in a human-interpretable form is presented. Results over extensive data-sets, collected from multiple active environments are presented, to show the competence and generalizability of the proposed framework.

CHAPTER I

INTRODUCTION

··· then purged with euphrasy and rue the visual nerve, for he had much to see ···
- John Milton (Paradise Lost: Book XI)

1.1 Motivation

As I look around myself, I see a group of fellow graduate students having a team meeting around a desk near mine. Some amongst them passionately move their hands as they emphasize a point, while others write equations on a white-board - the rest nod their heads in a "yay" or "nay" manner, either agreeing or disagreeing with the point being discussed.

Interestingly, this simple observation raises a fundamental question, *i.e.*, how do we actually understand these everyday activities in such efficient and effortless manner? This highly elliptical question has numerous facets along which it needs to be examined - including everything from the biological aspects of low-level image formation on retina, to the cognitive aspects of understanding streams of visual information. In this treatise, I would try to elucidate some of the computational A.I. aspects of this problem of everyday activity analysis.

The understanding of how we make sense of everyday activities has many ramifications, including one that involves coming up with systems that can be used for automatic surveillance or supporting users in ubiquitous environments.

1.2 Problem Statement

While the general theme of this treatise is activity analysis in sensor-rich environments, more specifically this work focuses on the following problems:

- How can we *represent* activities transpiring in a certain environment without having substantive prior knowledge about the activity-structure in that domain?

- How can we *discover* the different kinds of these activities in an unsupervised manner.
- How can we *characterize* these discovered activity-classes so that we can analyze them at different levels of details.
- What defines a *usual* or an *unusual* member of an activity-class, and finally,
- How can we *explain* the unusual activity-instances in a human interpretable form?

1.3 Solution Approach & Contributions

The approach to solving the key-questions enumerated above and the main contributions of this dissertation are:

- **Novel Activity Representation:** A novel representation of activities is presented as bags of discrete event n-grams - a perspective different from the previously used grammar driven approaches. This treatment of activities, motivated by some recent developments in natural language processing, allows one to analyze the global structural information of activities by simply considering their local event statistics.
- **Activity Similarity:** Based on this activity representation, the notion of similarity between two activities is formalized, taking into account their core structural and event-frequency based differences.
- **Unsupervised Activity-Class Discovery:** The problem of unsupervised activity-class discovery is posed as a graph theoretic problem, swhoing how finding maximal cliques in edge-weighted activity-graphs can be used to this end.
- **Activity-Class Characterization:** Taking on some of the previous work in the fields of computer networks, the problem of activity-class characterization at a wholestic scale is formalized as finding the *typical* content and structure of an activity-class.
- **Unsupervised Event-Motif Discovery:** Inspired by some of the recent work in the field of bio-informatics, predictably recurrent event subsequences (Event-Motifs) are extracted, using variable-memory Markov chains.

- **Anomalous activity detection:** An incremental method is proposed for classifying a new activity-instance and detecting whether it is a regular or an anomalous member of its membership class.

- **Anomaly Explanation:** An information-theoretic method is introduced that explains how an anomalous activity is different from regular activities in a human-interpretable form. Such explanations can be useful for large scale vision based surveillance systems.

1.4 Related Work

The problem of activity analysis has been taken on in numerous fields, including everything from human-computer interaction to computational perception, from ubiquitous computing to machine learning. Interestingly, some researchers have also tried to use these analyses for generation of new motion models in turn used for rendering different movements and behaviors on the console. One of the pioneer pieces of work in computational context analysis was [7] which analyzed the different activities of children in a sensor-rich environment. Researchers in the field of ubiquitous computing have done interesting work in creating context-aware applications including everything from mobile tour guides [1] to working with more recent paradigms of programming-by-demonstration [2]. In computer vision, most of the initial interest in activity understanding was focused on model based activity recognition [9] [35] where a set of target activities is modeled using some representation, followed by the learning of the model parameters given some training data. Different types of Dynamic Bayesian Networks have been extensively used for modeling various activities [39] [32]. Along these lines, some researchers have also made use of the context information to improve upon the recognition performance [23]. At the same time, these models of human behavior and motion have been used to synthesize new types of motions in the field of computer animation [8].

In the past, various approaches for activity representation have been fundamentally grammar-driven (see e.g. [16], [22]). In this work I propose to treat activities as bags of event n-grams, which allows the extraction of the global structural information of an activity, by simply considering its event statistics at a local scale. This treatment of activities, motivated by some recent developments in natural language processing [30], lets one to get away from actually scripting every single way in which an activity can be performed, and can be used for learning the different kinds of activity structures in an unsupervised manner.

Although the idea of discovering activity-classes has been previously explored in such fields as network intrusion detection [20], it has only recently been applied to everyday activities. My approach towards this problem is novel in a few key aspects. Unlike [15] which require a priori expert knowledge to model the activity-classes in an environment, I propose to discover this information in an unsupervised fashion. Since event-monograms, as used in [40] and [36], do not capture the temporal information of an activity, I use higher order event n-grams to capture this information more efficiently.

Numerous solutions to the problem of discovering important recurrent motifs in sequential data have been proposed (see e.g. [25] and [11] and the references therein). Work done in [38] and [28] present techniques for learning variable-memory Markov chains from training data in an unsupervised manner. The variable-memory elements in these Markov chains can be thought of as motifs that have good predictive power of the future events. However they presume the availability of pre-classified data. Moreover, their approach does not filter out the motifs that are common in multiple classes. Here, I modify the work done in [38] to handle data from multiple classes, finding motifs that are maximally mutually exclusive amongst activity-classes. This forms a nice continuum between the activity-class discovery, and characterization. Moreover, instead of sequentially finding individual motifs and masking them out from the sequences as proposed in [5], my scheme simultaneously finds all the motifs in the data in one pass. This allows one to find partially overlapping motifs.

Most of the previous attempts to tackle the problem of anomaly detection have focused on model-based anomaly recognition. These methods pre-define a particular type of activity as being anomalous, model it in some way, and then detect whether a new activity-instance is anomalous [15] [19]. While such an approach could prove to be useful for cases where the variance between different anomalous instances is not significantly large [21], for any reasonably unconstrained situation, anomalies are hard to completely define a priori. Since this is particularly true for everyday activities, I define an anomaly as “something different from regular”, with the hope of being able to model something regular more efficiently. Interestingly enough, there are studies done in Cognitive Science which show evidence that humans also learn about anomalies by considering the ”distance” of a new piece of information from the mental model of the class which they believe the new information belongs to [10] [12].

I formalize the problem of discovering activity classes as searching for edge-weighted maximal cliques in the graph of K activity-instances. Indeed, in the past, some authors have argued that maximal clique is the strictest definition of a cluster [4]. Finding maximal cliques in an edge-weighted undirected graph is a classic graph theoretic problem. Because combinatorially searching for maximal cliques is computationally hard, numerous approximations to the solution of this problem have been proposed (see [27] and the references within). For my purposes, I adopt the recently proposed approximate approach of iteratively finding dominant sets of maximally similar nodes in a graph (equivalent to finding maximal cliques) [26]. Besides providing an efficient approximation to finding maximal cliques, the framework of dominant sets naturally provides a principled measure of the cohesiveness of a class as well as a measure of node participation in its membership class. This measure of class participation may be used for an instance based representation of a clique [18].

It is argued that while facing a new piece of information, humans first classify it into an existing class [31] [29], and then compare it to the previous class members to understand how it varies in relation to the general characteristics of the membership class. Using this hypothesis as my motivation, I represent an activity class by a set of mutually disjunctive sub-classes, and then detect a new activity as a regular or an anomalous member of its membership sub-class. Some of the previous work done in bio-informatics on finding motifs, presumes the availability of pre-classified data [6]. Moreover, these approaches do not filter out the motifs that are common in multiple classes. My proposed scheme discovers activity classes in an unsupervised manner, and finds patterns that are maximally mutually exclusive amongst activity-classes. Besides activity-class characterization, these motifs can be used for finding interesting sub-sequences which need to be further analyzed more closely.

1.5 Thesis outline

I start in Chapter 2 by explaining a novel activity representation which does not require prior knowledge of the activity-structure. Using this novel activity representation, I explain a method of unsupervised activity-class discovery in Chapter 3, followed by a discussion about the different ways to characterize such discovered activity-classes in Chapter 4. In Chapter 5, I explain how these characterizations can be used for activity classification, unusual activity detection and explanation. The empirical analysis of my proposal for everyday activity analysis is presented in Chapter 6, followed by Chapter 7, which explains the shortcomings and conclusions of this treatise.

CHAPTER II

PREVIOUS WORK

The problem of activity analysis has been taken on in numerous fields, including everything from human-computer interaction to computational perception, from ubiquitous computing to machine learning. Interestingly, some researchers have also tried to use these analyses for generation of new motion models in turn used for rendering different movements and behaviors on the console. One of the pioneer pieces of work in computational context analysis was [7] which analyzed the different activities of children in a sensor-rich environment. Researchers in the field of ubiquitous computing have done interesting work in creating context-aware applications including everything from mobile tour guides [1] to working with more recent paradigms of programming-by-demonstration [2]. In computer vision, most of the initial interest in activity understanding was focused on model based activity recognition [9] [35] where a set of target activities is modeled using some representation, followed by the learning of the model parameters given some training data. Different types of Dynamic Bayesian Networks have been extensively used for modeling various activities [39] [32]. Along these lines, some researchers have also made use of the context information to improve upon the recognition performance [23]. At the same time, these models of human behavior and motion have been used to synthesize new types of motions in the field of computer animation [8].

In the past, various approaches for activity representation have been fundamentally grammar-driven (see e.g. [16], [22]). In this work I propose to treat activities as bags of event n-grams, which allows the extraction of the global structural information of an activity, by simply considering its event statistics at a local scale. This treatment of activities, motivated by some recent developments in natural language processing [30], lets one to get away from actually scripting every single way in which an activity can be performed, and can be used for learning the different kinds of activity structures in an unsupervised manner.

Although the idea of discovering activity-classes has been previously explored in such fields as network intrusion detection [20], it has only recently been applied to everyday activities. My approach towards this problem is novel in a few key aspects. Unlike [15] which require a priori expert knowledge to model the activity-classes in an environment, I propose to discover this information in an unsupervised fashion. Since event-monograms, as used in [40] and [36], do not capture the temporal information of an activity, I use higher order event n-grams to capture this information more efficiently.

Numerous solutions to the problem of discovering important recurrent motifs in sequential data have been proposed (see e.g. [25] and [11] and the references therein). Work done in [38] and [28] present techniques for learning variable-memory Markov chains from training data in an unsupervised manner. The variable-memory elements in these Markov chains can be thought of as motifs that have good predictive power of the future events. However they presume the availability of pre-classified data. Moreover, their approach does not filter out the motifs that are common in multiple classes. Here, I modify the work done in [38] to handle data from multiple classes, finding motifs that are maximally mutually exclusive amongst activity-classes. This forms a nice continuum between the activity-class discovery, and characterization. Moreover, instead of sequentially finding individual motifs and masking them out from the sequences as proposed in [5], my scheme simultaneously finds all the motifs in the data in one pass. This allows one to find partially overlapping motifs.

Most of the previous attempts to tackle the problem of anomaly detection have focused on model-based anomaly recognition. These methods pre-define a particular type of activity as being anomalous, model it in some way, and then detect whether a new activity-instance is anomalous [15] [19]. While such an approach could prove to be useful for cases where the variance between different anomalous instances is not significantly large [21], for any reasonably unconstrained situation, anomalies are hard to completely define a priori. Since this is particularly true for everyday activities, I define an anomaly as "something different from regular", with the hope of being able to model something regular more efficiently. Interestingly enough, there are studies done in Cognitive Science which show evidence that humans also learn about anomalies by considering the "distance" of a new piece of information from the mental model of the class which they believe the new information belongs to [10] [12].

I formalize the problem of discovering activity classes as searching for edge-weighted maximal cliques in the graph of K activity-instances. Indeed, in the past, some authors have argued that maximal clique is the strictest definition of a cluster [4]. Finding maximal cliques in an edge-weighted undirected graph is a classic graph theoretic problem. Because combinatorially searching for maximal cliques is computationally hard, numerous approximations to the solution of this problem have been proposed (see [27] and the references within). For my purposes, I adopt the recently proposed approximate approach of iteratively finding dominant sets of maximally similar nodes in a graph (equivalent to finding maximal cliques) [26]. Besides providing an efficient approximation to finding maximal cliques, the framework of dominant sets naturally provides a principled measure of the cohesiveness of a class as well as a measure of node participation in its membership class. This measure of class participation may be used for an instance based representation of a clique [18].

It is argued that while facing a new piece of information, humans first classify it into an existing class [31] [29], and then compare it to the previous class members to understand how it varies in relation to the general characteristics of the membership class. Using this hypothesis as my motivation, I represent an activity class by a set of mutually disjunctive sub-classes, and then detect a new activity as a regular or an anomalous member of its membership sub-class. Some of the previous work done in bio-informatics on finding motifs, presumes the availability of pre-classified data [6]. Moreover, these approaches do not filter out the motifs that are common in multiple classes. My proposed scheme discovers activity classes in an unsupervised manner, and finds patterns that are maximally mutually exclusive amongst activity-classes. Besides activity-class characterization, these motifs can be used for finding interesting sub-sequences which need to be further analyzed more closely.

CHAPTER III

REPRESENTING ACTIVITIES AS BAGS OF EVENT *N*-GRAMS

Consider an active setting such as a loading dock with delivery vehicles, people, packages etc. Such active environments consist of animate and inanimate objects interacting with each other. The interaction of these objects in a particular manner constitutes an event, while a sequence of such events constitutes an activity. In the past, various approaches for activity representation have been fundamentally grammar-driven [16]. These representations have proven to be useful in cases where one has *a priori* information about the structure of activity and can actually construct a model whose parameters can be later learned given some data. While such a presumption can be safely made in constrained situations, it is far from being true in large scale uncontrolled settings. Observing this constrain, researchers have extensively used ergodic hidden markov models for activity representation [22]. However, most of the work in that direction has been done in a supervised learning framework where the availability of classes along with labeled training data is presumed.

3.1 Vector Space Model - VSM

Looking at an activity as a sequence of discrete events, two important quantities emerge:

- *Content* - events that span the activity, and
- *Order* - the arrangement of the set of events.

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) [30], in which a document is represented as a vector of its word-counts, in the space of possible words.

To use such a scheme, we must define a set of possible events (*event vocabulary*) that could take place in the situation under consideration. As this representation is designed

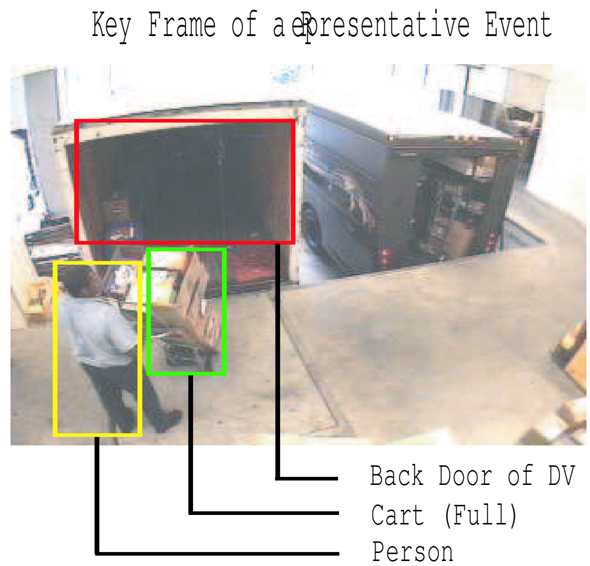

Figure 1. A Person pushes a Cart carrying Packages into the Back Door of a Delivery Vehicle.

to be manipulated by a perceptual system, the events must be chosen such that they are detectable from low-level perceptual primitives. A particular interaction of these perceptual primitives constitute an event. A key-frame of a representative event from one of the active environments that is explored in this work (Loading Dock) is shown in Figure 1.

While VSM captures the content of a sequence in an efficient way, it ignores its order. To stress upon this point, let us assume we are given a set of 10 events:

$$E = \{1, 2, ..., 10\} \tag{1}$$

Consider two sequences of exactly the same content, but different event order given as:

$$S1 = \{1, 2, 3, 8, 9, 10\} \tag{2}$$

and

$$S2 = \{8, 9, 10, 1, 2, 3\} \tag{3}$$

The VSM representations for both $S1$ and $S2$ are given in figure 2. As can be seen, while VSM captures the content of the sequences competently, it loses the order information of the sequences. Because the word content in documents often implies causal structure, this is usually not a significant problem. Generally activities are not fully defined by their event-content alone; rather, they have preferred or typical event-orderings. Therefore a model for capturing the order of events is needed.

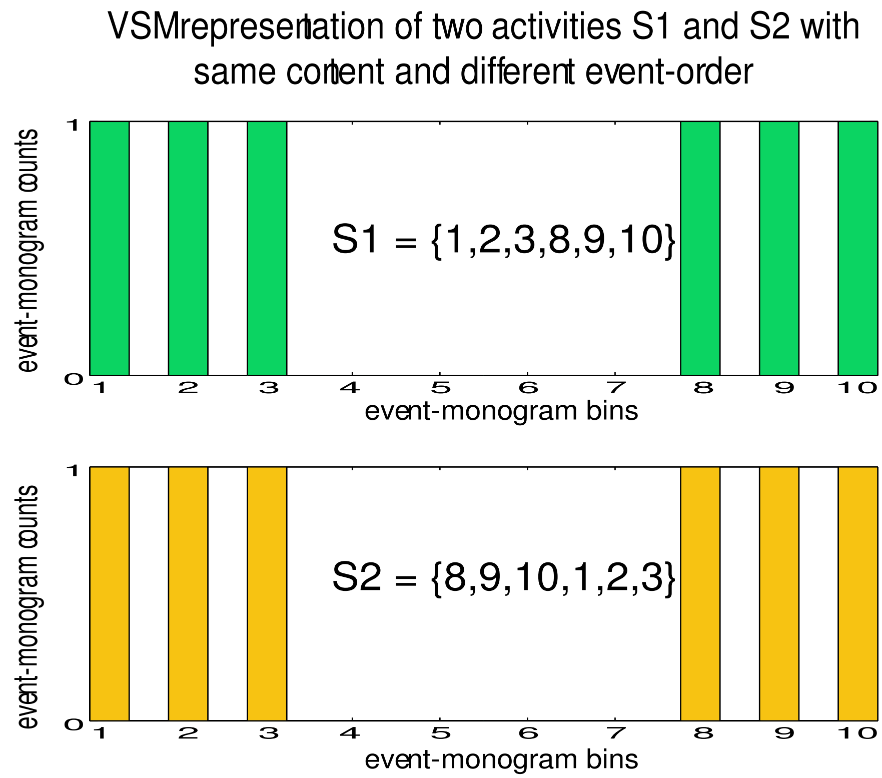

Figure 2. VSM representation of two sequences S1 and S2 which have the same event-content, while different event order. As it can be seen, the VSM representation does not capture such order differences.

3.2 Activities as Histograms of Event n-Grams

To this end, we consider histograms of higher order event n-grams (see figure 3), where we represent an activity by a sparse vector of counts of overlapping event n-grams in a high dimensional space of possible event n-grams. Our proposed scheme would capture the activity-structure for domains with substantive structural coherence. It is evident that higher values of n would capture the temporal order information of events more rigidly, and would entail a more discriminative representation.

3.3 Discussion

While the proposed representation of treating activities as bags of event n-grams captures both content and order of events, it does pose a problem of dealing with sparse very high dimensional data. For instance, if we defined K events that could take place in a situation, and we considered n-grams with n = 3, our activity would be living in a space of order

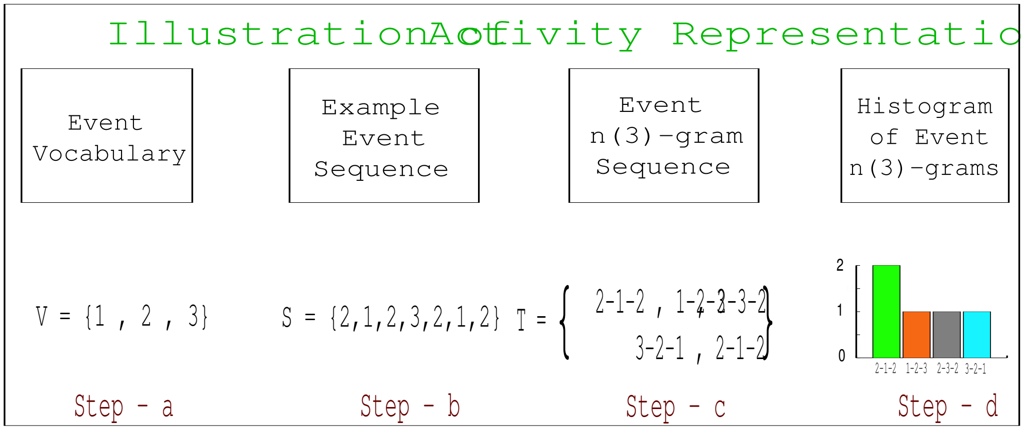

Figure 3. Transformation of an example activity from sequence of discrete events to histogram of event n-grams. Here the value of n is shown to be equal to 3. V is event vocabulary, S is event sequence, and T is sequence of overlapping n-grams. Step-d shows the non-zero n-gram counts of V.

$O(K^3)$. For even moderate values of K, learning and estimation in such a space can be infeasible.

One could partially solve the curse of dimensionality of the space using any of the plethora of dimensionality reduction techniques available. Such an approach is very similar to the Latent Semantic Analysis [37] of documents in the Information Retrieval community, where documents are mapped to a lower dimensional subspace (using Principal Component Analysis [34]), and document similarity is computed based on the inner product between vectors in the reduced dimensional subspace. However, PCA would lose some information (least significant in terms of L_2 norm) in the process of reducing dimensionality. Again, since the word content in documents often implies causal structure, this loss of information is usually not a significant problem for documents. On the other hand, activities are generally characterized by specific event orderings, therefore the loss of information in case of activities can prove to be more serious.

Another potential solution to the problem of dimensional complexity is to consider only those dimensions which are of significant importance, where the dimension importance can be estimated by using some inductive learning techniques based on some training data as proposed in some of the work in Network Intrusion Detection community [20]. Unfortunately, unlike the problem of network intrusion detection, where large sets of network activity data are usually available, we do not have such large data sets, and therefore using inductive learning techniques for our problem would not provide a good estimation of dimension importance.

In this work, a solution to this problem is proposed by constructing a similarity metric that encompasses all the non-zero components of the activity vector, hence giving equal importance to all the dimensions with values greater than zero, and no importance to dimensions with values equal to zero. Since the activity vector is highly sparse, this allows us to reduce the dimensionality dramatically without loosing any information, allowing us to be able to use higher values of n-grams for our analysis (see Chapter 4 for more details).

Another interesting question regarding the proposed representation is what is the optimal value of n to be used for activities in a certain domain. For instance, in environments with high quantum of structure, the activities would follow certain order more strictly, and one could use smaller values of n to capture this order information. On the other hand, higher order of n would be needed to capture the structural information for domains where activities have more stochastic element to them. We propose a potential solution to this problem of finding the optimal value of n, by optimizing over the lengths of predictably recurrent event subsequences (Event motifs) using variable-memory Markov chains (see Chapter 5 for more details).

CHAPTER IV

UNSUPERVISED ACTIVITY-CLASS DISCOVERY

This chapter begins by presenting a novel sequence comparison metric. The proposed view of the similarity between a pair of sequences consists of two factors, the core structural differences and differences based on the frequency of occurrence of event n-grams. Having established a notion of similarity between a pair of activities, the problem of activity-class discovery is posed as a graph theoretic problem of finding maximal cliques in edge-weighted activity graphs where the weights between activity-nodes are proportional to the similarity between the corresponding activities sequences.

4.1 Activity Similarity Metric

Sequence comparison is a well-studied problem and has numerous applications in such fields as text retrieval, bio-informatics etc. [13]. Our view of the similarity between a pair of sequences consists of two factors, the *core structural differences* and differences based on the *frequency of occurrence* of event n-grams.

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.

Let A and B denote two sequences of events, and let their corresponding histogram of n-grams be denoted by H_A and H_B. Let Y and Z be the sets of indices of n-grams with counts greater than zero in H_A and H_B respectively. Let α_i denote different n-grams. $f(\alpha_i|H_A)$ and $f(\alpha_i|H_B)$ denote the counts of α_i in sequences A and B respectively. We define the similarity between two event sequences as:

$$sim(A,B) = 1 - \kappa \sum_{i \in Y,Z} \frac{|f(\alpha_i|H_A) \; - \; f(\alpha_i|H_B)|}{f(\alpha_i|H_A) \; + \; f(\alpha_i|H_B)} \tag{4}$$

where $\kappa = 1/(||Y|| + ||Z||)$ is the normalizing factor, and $||\cdot||$ computes the cardinality of a set. While our proposed similarity metric conforms to: (1) the property of *Identity of indiscernibles*, (2) is *commutative*, and (3) is *positive semi-definite*, it does not however follow *Cauchy-Schwartz inequality*, making it a divergence rather than a true distance metric.

4.2 Activity-Class Discovery

It is argued that while facing a new piece of information, humans first classify it into an existing class [29] [31], and then compare it to the previous class members to understand how it varies in relation to the general characteristics of the membership class. Using this hypothesis as our motivation, we represent an activity space by a set of mutually disjunctive classes, and then detect a new activity as a regular or an anomalous member of its membership class.

4.2.1 Activity-Class as Maximal Clique

This work asserts that the activity-instances occurring in an environment do not span the activity-space uniformly. Rather, there exist disjunctive activity-sets with high internal similarity while low similarity across the sets. This assertion is backed by the assumption that the detected events, constituting activities in an environment, encode the underlying structure of activities [29].

Starting off with a set of K activity-instances, let us consider this activity-set as an undirected edge-weighted graph with K nodes, each node representing a histogram of n-grams of one of the K activity-instances. The weight of an *edge* is the similarity between a pair of nodes as defined in § 4.1. We can now formalize the problem of discovering activity-classes of as searching for edge-weighted maximal cliques[1] in the graph of K activity-instances [4]. Along these lines, a maximal clique in the graph is found, proceed by removing that set

[1]Recall that a subset of nodes of a graph is a *clique* if all its nodes are mutually adjacent; a *maximal* clique is is not contained in any larger clique, whereas a *maximum* clique has largest cardinality.

of nodes from the graph, and repeating this process iteratively with the remaining set of nodes, until there remain no non-trivial maximal cliques in the graph. The leftover nodes after the removal of maximal cliques are dissimilar from most of the regular nodes, and are hence anomalous (for details on anomaly detection, please see § 6.1).

4.2.2 Maximal Cliques using Dominant Sets

Finding maximal cliques in an edge-weighted undirected graph is a classic graph theoretic problem. Because combinatorially searching for maximal cliques is computationally hard, numerous approximations to the solution of this problem have been proposed [27]. For our purposes, we adopt the approximate approach of iteratively finding *dominant sets* of maximally similar nodes in a graph (equivalent to finding maximal cliques) as proposed in [26]. Besides providing an efficient approximation to finding maximal cliques, the framework of dominant sets naturally provides a principled measure of the cohesiveness of a class as well as a measure of node participation in its membership class. We now give an overview of dominant sets showing how they can be used for our problem.

Let the data to be clustered be represented by an undirected edge-weighted graph with no self-loops $G = (V, E, \vartheta)$ where V is the vertex set $V = \{1, 2, ...K\}$, $E \subseteq V \times V$ is the edge set, and $\vartheta : E \rightarrow \mathbb{R}^+$ is the positive weight function. The weight on the edges of the graph are represented by a corresponding $K \times K$ symmetric similarity matrix $A = (a_{ij})$ defined as:

$$a_{ij} = \begin{cases} sim(i,j) & if\ (i,j) \in E \\ 0 & otherwise \end{cases} \tag{5}$$

sim is computed using our proposed notion of similarity as described in §4.1. To quantize the cohesiveness of a node in a cluster, let us define its "average weighted degree". Let $S \subseteq V$ be a non-empty subset of vertices and $i \in S$, such that,

$$awdeg_S(i) = \frac{1}{||S||} \sum_{j \in S} a_{ij} \tag{6}$$

Moreover, for $j \notin S$, we define Φ_S as:

$$\Phi_S(i,j) = a_{ij} - awdeg_S(i) \tag{7}$$

Intuitively, $\Phi_S(i,j)$ measures the similarity between nodes j and i, with respect to the average similarity between node i and its neighbors in S. Note that $\Phi_S(i,j)$ can either be positive or negative.

Now let us consider how weights are assigned to individual *nodes*[2]. Let $S \subseteq V$ be a non-empty subset of vertices and $i \in S$. The weight of i w.r.t. S is given as:

$$w_S(i) = \begin{cases} 1 & if\ ||S|| \ = \ 1 \\ \sum_{j \in S \setminus \{i\}} \Phi_{S \setminus \{i\}}(j,i) w_{S \setminus \{i\}}(j) & otherwise \end{cases} \tag{8}$$

Moreover, the total weight of S is defined to be:

$$W(S) = \sum_{i \in S} w_S(i) \tag{9}$$

Intuitively, $w_S(i)$ gives a measure of the overall similarity between vertex i and the vertices of $S \setminus \{i\}$ with respect to the overall similarity among the vertices in $S \setminus \{i\}$. We are now in a position to define *dominant sets*. A non-empty sub-set of vertices $S \subseteq V$ such that $W(T)$ > 0 for any non-empty $T \subseteq S$, is said to be *dominant* if:

1. $w_S(i) > 0$, $\forall\ i \in S$, i.e. internal homogeneity
2. $w_{S \bigcup \{i\}}(i) < 0\ \forall\ i \notin S$, i.e. external inhomogeneity.

Effectively, we can state that the dominant set in a edge-weighted graph is equivalent to a cluster of vertices in that graph.

4.2.3 Dominant Sets Using Replicator Dynamics

We now turn our attention to finding a dominant set in an edge-weighted graph with adjacency matrix A. For this purpose, consider the following quadratic program which is a generalization of Motzkin-Straus program [24]:

---

[2]Note that here the term weight is being used to describe both the edge-weights and the node-weights. However, these two are different quantities.

$$maximize\ f(x) = \frac{1}{2}x^T Ax \tag{10}$$

subject to $x \varepsilon \Delta$. where

$$\Delta = x\ \in \mathbb{R}^n : \sum_{i=1}^{n}\ x_i = 1\ and\ x_i \geq\ 0, \forall i \tag{11}$$

is the standard simplex in $\mathbb{R}^n$. If S is a dominant sub-set of vertices, then its weighted characteristics vector x^S, defined as:

$$w_i(S) = \begin{cases} \frac{w(i,j)}{W(S)} & if\ |S|\ \in\ S \\ 0 & otherwise \end{cases} \tag{12}$$

is a strict local maximizer of f in Δ. Conversely, if x^* is a strict local maximizer of f in Δ then its support $\sigma = \sigma(x^*) = \{i\ \in\ V\ :\ x_i^*\ \neq 0\ \}$is a dominant set. By the virtue of the above result, we can find a dominant set by first localizing a solution of Equation 10 with an appropriate continuous optimization technique, and then picking up the support set of the solution found. The clustering algorithm we use basically consists of iteratively finding a dominant set in that graph by solving Equation 10 and finding its support, then removing the support from the graph, until all the vertices have been clustered.

Because solving Equation 8 combinatorially is infeasible, we use a continuous optimization technique proposed in [26] which applying replicator dynamics. Let $W = (w_{ij})$ be a non-negative real-valued $n\ x\ n$ matrix. The discrete time version of the replicator equation can be given as [24]:

$$x_i(t+1) = x_i(t)\frac{(Wx(t))_i}{x(t)^T Wx(t)} \tag{13}$$

According to the fundamental theorem of natural selection [14], if $W = W^T$, then the function $F(x) = x^T Wx$ is strictly increasing along any non-constant trajectory of the replicator dynamics of equation 13. In other words, $\forall\, t\ >\ 0$, $F(x(t+1)) > F(x(t))$. Finally, let $W = A$, the adjacency matrix, then the replicator system, starting from any arbitrary initial state will eventually converge to a maximizer of function given in Equation 10. This will correspond to a dominant set in the graph and hence to a cluster of nodes.

CHAPTER V

ACTIVITY CLASS CHARACTERIZATION

Having clustered a given set of activities, we are now looking for a way to represent their characteristics in terms of some general, tractable and concise form. Such a class model is required to:

- **Classification** of a new activity instance, and for the
- **Comparison** of the new activity-class member to the general characteristics of the membership class to analyze its normality.

There are many ways to approach the problem of creating a representative model of a class. One of these is the generative approach which presumes a stochastic process that creates class instances, where the objective is to learn the particular distribution which dictates this underlying process. However for our problem, since the parametric form of the underlying distribution is unknown, this direction cannot be adopted. Even if we approximate the actual distribution through some known parametric form, the large dimensionality of the activity space and the availability of small activity samples, makes learning such a distribution without over-fitting infeasible.

Let us therefore resort to the idea of instance based approach for activity class characterization. In this regard, two related approaches are investigated here:

- **Typical Class Member:** We formulate the problem as that of finding the node which is the "best representative" of the rest of the cluster nodes - essentially converting the problem of learning, into one of search. From now on we will call the best representative member of a class as the "Typical Member".
- **Event Motifs:** We formalize the problem as finding predictably recurrent activity subsequences using variable-memory Markov chains. These subsequences are generally called Event Motifs and are maximally mutually exclusive amongst activity-classes.

The characterization of an activity-class using its Typical Member allows one to represent the general characteristics of the class at a wholestic scale, which in turn comes in handy for the overall explanation of an unusual class-member. The Event Motif characterization lets one take a more granular look at the structure of an activity-class, which in turn allows one to analyze a new class member at a more local scale. In the following, a detailed explanation of each one of these characterizations is provided.

5.1 Typical Class Member

The question of typicality is closely related to the idea of how similar a node is to the other members of the cluster. There are many ways in which this idea of similarity of a node with respect to other nodes could be exploited to find the typical node. The classic graph theoretic literature provides a potential answer to this problem in terms of finding the "centroid" of the cluster, *i.e.* finding the node which minimizes the maximum distance (inverse of similarity) between the rest of the nodes and itself (also known as the Min-Max algorithm). While this method is theoretically sound, it is prone to noisy clusters and would work well only in cases where the clusters are well-behaved.

Another method proposed in graph theory for such a problem relates to finding the maximum in-degree of every node of the cluster, labeling the node with maximum in-degree as the typical node. One could consider the number of nodes maximally close to a particular node as it in-degree, transforming our undirected graph into a directed one. Stated otherwise, the idea is to consider that node as the typical member of the cluster, to which most nodes are maximally similar. Indeed, the approach of labeling a node as typical or not, based on its in-degree usually works very well. It still however retains some major problems in terms being completely agnostic about the more global structure of the cluster. More specifically, due to the maximization operation which we have to do to transform of our undirected graph to a directed one, we are forced to look at a very local view of our landscape which of course could lead to problems.

The idea of fining the best representative member of a cluster has been studied in some other fields such as Computer Networks, where the problem is very similar to finding the web-page which best represents a collection of web-pages (see e.g. [18]). Along the lines of [18], we propose the idea of Typical nodes (mentioned as "Authoritative Sources" in [18]) and "Similar to Typical (STT)" nodes (mentioned as "hubs" in [18]). Typical and STT nodes exhibit a mutually reinforcing relationship - a good STT node is one which is closer to a Typical node, while a Typical node is one closer to more STT nodes. Like [18], we associate a non-negative Typicality weight x^p and a non-negative STT weight y^p to each node in the cluster where p denotes the index of nodes in a cluster. Naturally, if p is closer to many nodes with large x values, it should receive a large y value. On the other hand if p is closer to nodes with large y values, it should receive large x value. We define two coupled processes to update the weights x^p and y^p iteratively, i.e.

$$x^p \leftarrow \sum_{q:(q,p)\in E} y^p \tag{14}$$

and

$$y^p \leftarrow \sum_{q:(q,p)\in E} x^p \tag{15}$$

As we iterate the above two equations k times in the limit $k \leftarrow \infty$, x^p and y^p converge to x^* and y^*. The node which has the largest component in the converged vector x^* would correspond to the node which has the greatest Typical weight and hence is the best representative of the nodes of clusters. x^* can be computed from the Eigen Analysis of the matrix A^TA where A is the symmetric similarity matrix of all the nodes of the cluster. Essentially x^* is the principal eigenvector (the one with greatest corresponding Eigen value) of A^TA, the largest component of which corresponds to the Typical Node of the cluster (for the proof, please refer to [18]).

5.2 Discovering Event Motifs

Let us now turn our attention towards finding interesting recurrent event-motifs in these discovered classes. Some of the previous work done in bio-informatics on finding motifs, presumes the availability of pre-classified data [6]. Moreover, these approaches do not filter out the motifs that are common in multiple classes. The scheme proposed here discovers activity classes in an unsupervised manner, and finds patterns that are maximally mutually exclusive amongst activity-classes.

5.2.1 A Definition of Motif

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. Following [38], we assume that a class of activity-sequences can be modeled as a variable-memory Markov chain (*VMMC*). We define an event-motif for an activity-class as one of the variable-memory elements of its *VMMC*. We cast the problem of finding the optimal length of the memory element of a *VMMC* as a function optimization problem and propose our objective function in the following.

Let Y be the set of events, A be the set of activity-instances, and C be the set of discovered activity-classes. Let us define a function $\mathcal{U}(a)$ that maps an activity $a \in A$ to its membership class $c \in C$. Let us define the set of activities belonging to a particular class $c \in C$ as $A_c = \{a \in A : \mathcal{U}(a) = c\}$. For $a = (y_1, y_2, ..., y_n) \in A$ where $y_1, y_2, ...y_n \in Y$, let $p(c|a)$ denote the probability that activity a belongs to class c. Then,

$$p(c|a) = \frac{p(a|c)p(c)}{p(a)} \propto \prod_{i=1}^{n} p(y_i|y_{i-1}, y_{i-2}, ..., y_1, c) \tag{16}$$

where we have assumed that all activities and classes are equally likely. We approximate Eq 16 by a *VMMC*, M_c to get:

$$\prod_{i=1}^{n} p(y_i|y_{i-1}, y_{i-2}, ..., y_1, c) = \prod_{i=1}^{n} p(y_i|y_{i-1}, y_{i-2}, ..., y_{i-m_i}, c) \tag{17}$$

where $m_i \leq i-1 \; \forall \; i$. For any $1 \leq i \leq n$, the sequence $(y_{i-1}, y_{i-2}, ..., y_{i-m_i})$ is called the *context* of y_i in M_c([38]), denoted by $\mathcal{S}_{M_c}(y_i)$. We want to find the sub-sequences which can effeciently characterize a particular class, while having minimal representation in other classes. We therefore define our objective function as:

$$\mathcal{Q}(M_c|A_c) = \gamma - \lambda \tag{18}$$

where

$$\gamma \;=\; \prod_{a \in A_c} p(c|a) \tag{19}$$

and

$$\lambda = \sum_{c' \in C \setminus \{c\}} \prod_{a \in A_{c'}} p(c'|a) \tag{20}$$

Intuitively, γ represents how well a set of event-motifs can characterize a class in terms of correctly classifying the activities belonging to that class. On the other hand, λ denotes to what extent a set of motifs of a class represent activities belonging to other classes. It is clear that maximizing γ while minimizing λ would result in the optimization of $\mathcal{Q}(M_c|A_c)$. Note that our motif finding algorithm leverages the availability of the discovered activity-classes to find the maximally mutually exclusive motifs. This shows the usefulness of our activity discovery framework as a pre-step to the motif finding scheme.

5.2.2 Objective Function Optimization

We now explain how we optimize our proposed objective function. [38] describe a technique to compare different *VMMC* models that balances the predictive power of a model with its complexity. Let s be a context in M_c, where $s = y_{n-1}, y_{n-2}, ..., y_1$, and $y_{n-1}, y_{n-2}, ..., y_1 \in Y$. Let us define the suffix of s as $\textit{suffix}(s) = y_{n-1}, y_{n-1}, ...y_2$. For each $y \in Y$, let $N_{A'}(y, s)$ be the number of occurrences of event y in activity-sequences contained in $A' \subseteq A$ where s precedes y, and let $N_{A'}(s)$ be the number of occurances of s in activity-sequences in A'. We define the function $\Delta_{A'}(s)$ as

$$\Delta_{A'}(s) = \sum_{y \in Y} N(s, y) log \left(\frac{\hat{p}(y|s)}{\hat{p}(y|\textit{suffix}(s))} \right) \tag{21}$$

where $\hat{p}(y|s) = N_{A'}(s, y)/N_{A'}(s)$ is the maximum likelihood estimator of $p(y|s)$. Intuitively, $\Delta_{A'}(s)$ represents the number of bits that would be saved if the events following s in A', were encoded using s as a context, versus having $\textit{suffix}(s)$ as a context. In other words, it represents how much better the model could predict the events following s by including the last event in s as part of context of these events.

We now define the function $\Psi_c(s)$ (bit gain of s) as

$$\Psi_c(s) = \Delta_{A_c}(s) - \sum_{c' \in C \setminus \{c\}} \Delta_{A_{c'}}(s) \tag{22}$$

Note that higher values of $\Delta_{A_c}(s)$ imply greater probability that an activity in A_c is assigned to c, given that s is used as a motif. In particular, higher the value of $\Delta_{A_c}(s)$, higher will be the value of γ. Similarly, higher the value of $\sum_{c' \in C \setminus \{c\}} \Delta_{A_{c'}}(s)$, higher the value of λ. We include a sequence s as a context in the model M_c iff

$$\Psi_c(s) > K \times log(\ell) \tag{23}$$

where ℓ is the total length of all the activities in A, while K is a user defined parameter. The term $K \times log(\ell)$ represents added complexity of the model M_c, by using s as opposed to $suffix(s)$ as a context, which is shorter in length and occurs at least as often as s. The higher the value of K the more parsimonious the model will be.

Equation 23 selects sequences that both appear regularly and have good classification and predictive power - and hence can be thought of as event-motifs. Work done in [28] shows how the motifs in a *VMMC* can be compactly represented as a tree. Work done in [3] presents a linear time algorithm that constructs such a tree by first constructing a data structure called a *Suffix Tree* to represent all sub-sequences in the training data A, and then by pruning this tree to leave only the sequences representing motifs in the *VMMC* for some activity-class. We follow this general approach by using Eq 23 as our pruning criterion.

CHAPTER VI

ACTIVITY CLASSIFICATION, ANOMALY DETECTION AND EXPLANATION

As mentioned in § 4.2, it is argued that while facing a new piece of information, humans first classify it into an existing class [29] [31], and then compare it to the previous class members to understand how it varies in relation to the general characteristics of the membership class. Using this hypothesis as our motivation, we represent an activity space by a set of mutually disjunctive classes, and then detect a new activity as a regular or an anomalous member of its membership class. Unlike [40] we do not wish to re-analyze the entire data set for every new activity instance. Therefore, we present an incremental approach to classification and detection for a new activity instance.

6.1 Activity Classification and Anomaly Detection

Given $||C||$ discovered activity-classes, we are now interested in finding if a new activity instance is regular or anomalous. Each member j of an activity-class c has some weight $w_c(j)$, that indicates the participation of j in c. We compute the similarity between a new activity-instance τ and previous members of each sub-class by defining a function $A_c(\tau)$ as:

$$A_c(\tau) = \sum_j sim(\tau, j) w_c(j) \quad \forall j \in c \tag{24}$$

Here $w_c(j)$ is the same as defined in Equation 8. A_c represents the average weighted similarity between the new activity-instance τ and any one of the discovered sub-classes c. The selected membership sub-class c^* can be found as

$$c^* = \arg\max_{\forall c} A_c(\tau) \tag{25}$$

Once the membership decision of a new test activity has been made, we now focus our attention on deciding whether the new class member is regular or anomalous. Intuitively speaking, we want to decide the normality of a new instance based on its closeness to the previous members of its membership sub-class. This is done with respect to the average closeness between all the previous members of its membership sub-class. Let us define a function $\Gamma(\tau)$ as:

$$\Gamma(\tau) = \sum_{j \in c^*} \Phi_{c^*}(j, \tau) w_{c^*}(j) \tag{26}$$

where Φ in is defined by Equation 7. We define a new sub-class member τ as regular if $\Gamma(\tau)$ is greater than a particular threshold. The threshold on $\Gamma(\tau)$ is learned by mapping all the anomalous activity instances detected in the training activity-set to their closest sub-class (using Equation 24, 25), and computing the value of Γ for both regular and anomalous activity instances. We can now observe the variation in $false$ *acceptance rate* (FAR) and *true positives* (HITS) as a function of the threshold Γ. This gives a "Receiver Operating Curve" (ROC). The area under this curve is indicative of the confidence in our detection metric $\Gamma(\tau)$ [17]. Based on our tolerance for HITS and FAR we can now choose an appropriate threshold.

6.2 Anomaly Explanation

We now address the question of characterizing the anomalous members. We first review (as explained in Chapter 5) the characterization of a model for the regular members of a sub-class against which its anomalous members could be compared [31]. We then find the most informative features of our space in terms of discriminability between the regular and the anomalous sub-class members.

6.2.1 Activity-Class Modeling

As mentioned in § 5.1, because of the huge dimensionality of our feature space and the availability of meager (and sparse) training data, we resort to the idea of activity-class representation using class prototype(s) (*the exampler view* [33]) to model the regular members of an activity-class. We formulate this problem as finding the member that is the "most representative" of the rest of the sub-class members. Fining the best representative member of a cluster in terms of its similarity to other cluster members has been studied in other fields. For instance [18] finds the *most authoritative nodes* in a cluster by iteratively assigning *authority weights* to each node member. An advantage of using the dominant sets framework for discovering constituent sub-class structure of an activity class is that it naturally provides a principled measure of a node's representativeness of its membership activity-class, defined by $w_S(i)$ in Equation 8. We propose using the member node of a activity-class with maximum weight $w_S(i)$ as the representative model of the sub-class. This most representative node is used to explain the anomalous members of the activity-class.

6.2.2 Explanatory Features

We now focus on the problem of finding the features that can be used to explain an anomalous activity in a maximally- informative manner. We are interested in features of a sub-class with minimum entropy and substantive frequency of occurrence. The entropy of a tri-gram indicates the variation in its observed frequency, which indicates the confidence in the prediction of its frequency. The frequency of occurrence of a tri-gram suggests its participation in a sub-class. We want to analyze the extraneous and the pertinent features in an activity that made it anomalous with respect to the most explanatory features of the regular members of the membership activity-class. We now construct our approach mathematically (a figurative illustration is given in Figure 4).

Let α_i denote a particular tri-gram i for an activity, and c denote any of of the $||C||$ discovered sub-classes. If R denotes the *Typical Member* member of c as described in §5.1, and τ denotes a new anomalous sub-class member, then we can define the difference between their counts for α_i as:

$$D(\alpha_i) = f_R(\alpha_i) - f_\tau(\alpha_i) \tag{27}$$

where $f(\alpha_i)$ denotes the count of a tri-gram α_i. Let us define the distribution of the probability of occurrence of α_i in c as:

$$P_c(\alpha_i) = \frac{\sum\limits_{k \in c} f_k(\alpha_i)}{\sum\limits_{i=1}^{M} \sum\limits_{k \in c} f_k(\alpha_i)} \tag{28}$$

where M represents all the non-zero tri-grams in all the members of sub-class c. Let us define multiset χ_c^i as:

$$\chi_c^i = \{f_k(\alpha_i)|k \in c\} \tag{29}$$

We can now define probability $Q(x)$ of occurrence of a particular member $x \in \chi_c^i$ for α_i in c as:

$$Q(x) = \psi \sum_{j \in c} \begin{cases} 1 & if\ f(\alpha_i) = \ x \\ 0 & otherwise \end{cases} \tag{30}$$

where ψ is the normalization factor. Let us define Shannon's Entropy of a tri-gram i for a sub-class c by $H_c(\alpha_i)$ as:

$$H_c(\alpha_i) = \sum_{x \in \chi_c^i} Q_c(x) ln(Q_c(x)) \tag{31}$$

We can now define the notion of *predictability*, $PRD_c(\alpha_i)$, of the values of tri-gram α_i of cluster c as:

$$PRD_c(\alpha_i) = 1 - \frac{H_c(\alpha_i)}{\sum\limits_{i=1}^{M} H_c(\alpha_i)} \tag{32}$$

It is evident from this definition, that α_i with high entropy $H_c(\alpha_i)$ would have high variability, and therefore would have low predictability.

We define the explainability of a tri-gram $\alpha_i \in c$ that was frequently and consistently *present* in the regular sub-cluster as:

$$\xi_c^P(\alpha_i) = PRD_c(\alpha_i) P_c(\alpha_i) \tag{33}$$

Intuitively, ξ_c^P indicates how much an α_i is instrumental in representing a sub-class c.

Similarly, we can define the explainability of $\alpha_i \in c$ in terms of how consistently was it *absent* in representing c.

$$\xi_c^A(\alpha_i) = PRD_c(\alpha_i)(P_c^{max}(\alpha_i) - P_c(\alpha_i)) \tag{34}$$

where $P_c^{max}(\alpha_i)$ is the maximum probability of occurrence of any α_i in c.

The first term in both Equation 33 and 34 indicates how consistent α_i is in its frequency over the different members of a cluster. The second term in Equation 33 and 34 dictates how representative and non-representative α_i is for c respectively.

Given an anomalous member of a sub-class, we can now find the features that were frequently and consistently *present* in the regular members of the sub-class, but were deficient

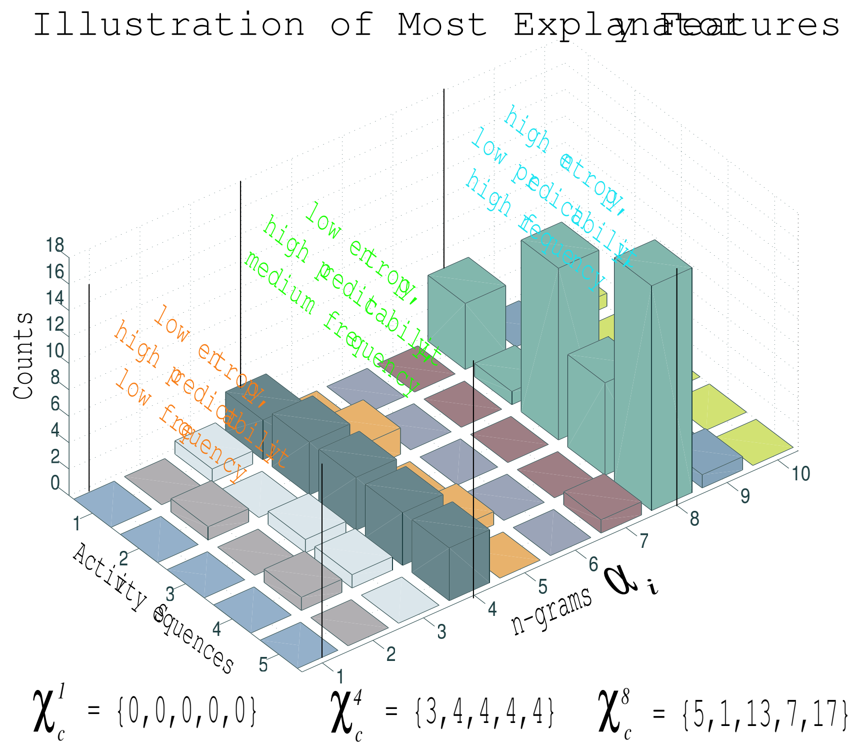

Figure 4. Five simulated activity sequences are shown to illustrate the different concepts introduced in § 6.2.2. α_1 has low value of P_c, its entropy H_c is low and therefore its predictability is high. α_4 has medium P_c, its entropy H_c is also low and its predictability is high. Finally α_8 has high P_c, but its entropy H_c is high which makes its predictability low. α_1 could be useful in explaining the extraneous features in an anomalous activity, while α_4 could be useful in explaining the features that were deficient in an anomaly.

in the anomaly τ. To this end, we define the function $DEFICIENT(\tau)$ as:

$$DEFICIENT(\tau) = \arg\max_{\alpha_i}[\xi_c^P(\alpha_i)D_c(\alpha_i)] \tag{35}$$

Similarly, we can find the most explanatory features that were consistently absent in the regular members of the membership sub-class but were *extraneous* in the anomaly. We define the function $EXTRANEOUS(\tau)$ as:

$$EXTRANEOUS(\tau) = \arg\min_{\alpha_i}[\xi_c^A(\alpha_i)D_c(\alpha_i)] \tag{36}$$

We can now explain anomalies based on these features that were

- *Deficient* from an anomaly but were frequently and consistently *Present* in the regular members
- *Extraneous* in the anomaly but were consistently *Absent* from the regular members of the activity-class.

CHAPTER VII

EXPERIMENTS AND RESULTS

To test the competence of the proposed framework, experiments on extensive data-sets collected from two active environments were performed. For both experimental setups, the value of n for the n-grams was chosen to be equal to 3 (tri-grams). This is done with the understanding that it encodes the past, present and future information of an event (essentially following second order Markov assumption).

7.1 Loading Dock Scenario

We collected video data at the Loading Dock area of a retail bookstore. To visually span the area of activities in the loading dock, we installed two cameras with partially overlapping fields of view. A schematic diagram with sample views from the two cameras is shown in Figure 5. Daily activities from 9a.m. to 5p.m., 5 days a week, for over one month were recorded. Based on our observations of the activities taking place in that environment, an event vocabulary of 61 events was constructed. Every activity has a known starting event, i.e. *Delivery Vehicle Enters the Loading Dock* and a known ending event, i.e. *Delivery Vehicle Leaves the Loading Dock*. We used 150 of the collected instances of activities, that were manually annotated using our defined event-vocabulary of 61 events. The interaction of some low-level perceptually distinguishable primitives constitute each of these 61 events. For the Loading Dock environment, we used 10 such primitives: *Person, Cart, Delivery Vehicle(D.V.), Left Door of D.V., Right Door of D.V., Back Door of D.V., Package, Doorbell, Front Door of Building, Side Door of Building.*

7.2 House Scenario

To test our proposed algorithms on the activities in a house environment, we deployed 16 strain gages at different locations in a house, each with a unique identification code.

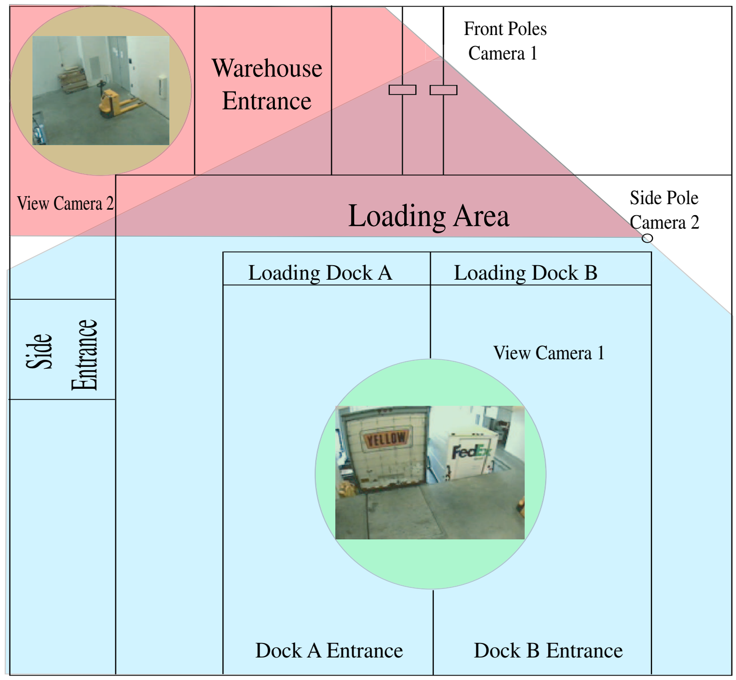

Figure 5. A schematic diagram of the camera setup at the loading dock area with overlapping fields of view (FOV). The FOV of camera 1 is shown in blue while that of camera 2 is in red. The overlapping area of the dock is shown in purple.

These transducers registered the time when the resident of the house walked over them. The data was collected daily for almost 5 months (151 days - each day is considered as an individual activity). Whenever the person passed near a transducer at a particular location, it was considered as the occurrence of a unique event. Thus our event vocabulary in this environment consists of 16 events. Figure 6 shows a schematic top-view of this environment.

7.3 Discovered Activity Classes

7.3.1 Loading Dock Scenario

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. The visual representation for the similarity matrices of the original 150 activities and the arranged activities in 7 clusters is shown in Figure 7. These discovered activity-classes were then provided to our motif finding framework which discovered multiple motifs of various lengths, ranked by their respective bit-gains (class-characterization ability). Analysis of the discovered classes reveals a strong

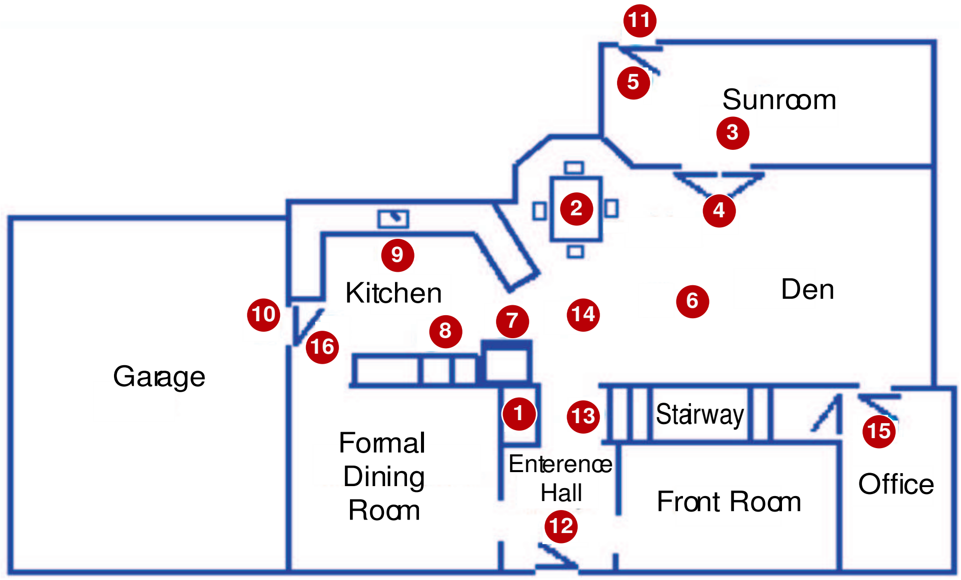

Figure 6. A schematic diagram of the strain-gage setup in the house scenario. The red dots represents the positions of the strain gages.

structural similarity amongst the class members, while the discovered motifs show ability to characterize the membership class efficiently. A brief description of the discovered activity-classes is given in following:

- **Sub-Class 1:** UPS® delivery-vehicles that picked up multiple packages using hand carts.
- **Sub-Class 2:** Pickup trucks (mostly Fed Ex®) and vans that dropped off a few packages without needing a hand cart.
- **Sub-Class 3:** Delivery trucks that dropped off multiple packages, using hand carts, that required multiple people.
- **Sub-Class 4:** A mixture of car, van, and truck delivery vehicles that dropped off one or two packages without needing a hand cart.
- **Sub-Class 5:** Delivery-vehicles that picked up and dropped-off multiple packages using a motorized hand cart and multiple people.
- **Sub-Class 6:** Van delivery-vehicles that dropped off one or two packages without needing a hand cart.

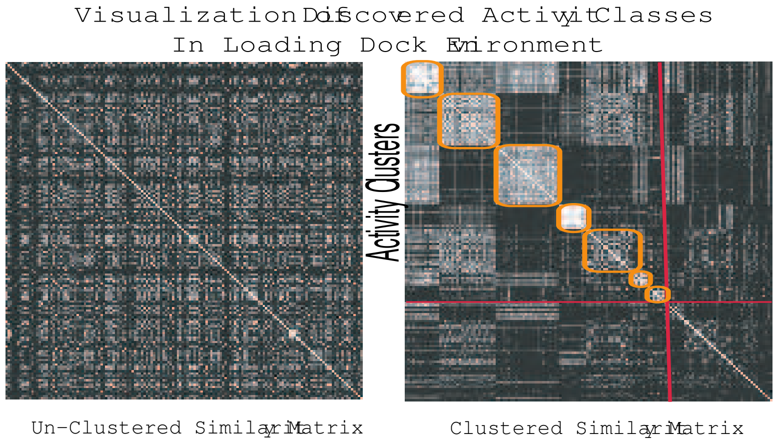

Figure 7. Each row represents the similarity of a particular activity with the entire activity training set. White implies identical similarity while black represents complete dissimilarity. The activities ordered after the red cross line in the clustered similarity matrix were dissimilar enough from all other activities as to not be included in any non-trivial maximal clique.

- **Sub-Class 7:** Delivery trucks that dropped off multiple packages using hand carts.

7.3.2 House Scenario

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 (see Figure 8). A brief description of the discovered activity-classes is given below:

- **Sub-Class 1:** Activities lasting for the entire length of days where the person's trajectory spans the entire house space. Most of the time was spent in the area around the Kitchen and the Dining Table.

- **Sub-Class 2:** The person moves from from kitchen to the stairway more often. Further more, as opposed to cluster 1, the person does not go from the Office to the Sum Room area.

- **Sub-Class 3:** The person spends more time in the areas of Den and the living-room. Moreover, he visits the Sun-room more often.

- **Sub-Class 4:** The person spends most of the day in the Kitchen and the Dining Room.The duration for which she stays in the house is small for this sub-class.

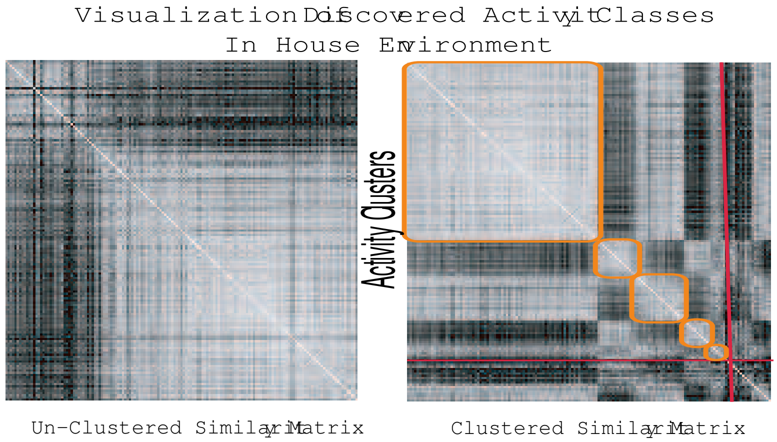

Figure 8. Visualization of similarity matrices before and after class discovery for the House Environment.

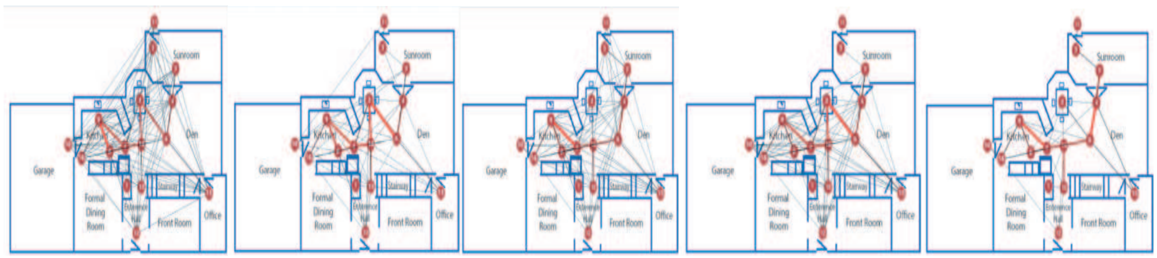

Figure 9. Visualization of the structural differences between the discovered activity-classes. Thick lines with brighter shades of red indicate higher frequency.

- **Sub-Class 5:** The person moves from Dining Room to the Sun Room more often. The duration for which she stays in the house is significantly smaller than any other sub-class.

To better illustrate the structural differences in the discovered activity-classes, a visualization of the normalized frequency-counts of the person's trajectory between different locations is shown in figure 9.

7.4 Discovered Event Motifs

7.4.1 Loading Dock Scenario

The highest big-gain event-motifs found for the 7 discovered activity-classes in the Loading Dock domain are given below:

- **Sub-Class 1:** Person places package into back door of delivery vehicle $\longrightarrow$ Person enters into side door of building g Person is empty handed $\longrightarrow$ Person exists from side door of building g Person is full handed $\longrightarrow$ Person places package into back door of delivery vehicle.

- **Sub-Class 2:** Cart is full $\longrightarrow$ Person opens front door of building $\longrightarrow$ Person pushes cart into front door of building $\longrightarrow$ Cart is full $\longrightarrow$ Person closes front door of building $\longrightarrow$ Person opens front door of building $\longrightarrow$ Person exists from front door of building $\longrightarrow$ Person is empty handed $\longrightarrow$ Person closes front door of building.

- **Sub-Class 3:** DV drives in forward into LDA $\longrightarrow$ Person opens left door of DV $\longrightarrow$ Person exists from left door of DV $\longrightarrow$ Person is empty handed $\longrightarrow$ Person closes the left door of delivery vehicle.

- **Sub-Class 4:** Person opens back door of DV $\longrightarrow$ Person removes package from back door of DV $\longrightarrow$ Person removes package from back door of DV $\longrightarrow$ Person removes package from back door of DV $\longrightarrow$ Person removes package from back door of DV $\longrightarrow$ Person removes package from back door of DV.

- **Sub-Class 5:** Person closes front door of building $\longrightarrow$ Person removes package from cart $\longrightarrow$ Person places package into back door of DV $\longrightarrow$ Person removes package from cart $\longrightarrow$ Person places package into back door of DV $\longrightarrow$ Person removes package from cart $\longrightarrow$ Person places package into back door of DV.

- **Sub-Class 6:** Person Removes Cart From Back Door of DV $\longrightarrow$ Person Removes Package From Back Door of DV $\longrightarrow$ Person Places Package Into Cart $\longrightarrow$ Person Places Package Into Cart $\longrightarrow$ Person Removes Package From Back Door of DV $\longrightarrow$ Person Places Package Into Cart $\longrightarrow$ Person Removes Package From Back Door of DV $\longrightarrow$ Person Places Package Into Cart.

- **Sub-Class 7:** Person closes back door of DV $\longrightarrow$ Person opens left door of DV $\longrightarrow$ Person enters into left door of DV $\longrightarrow$ Person is empty handed $\longrightarrow$ Person closes left door of DV.

7.4.2 House Scenario

The highest big-gain event-motifs found for the 5 discovered activity-classes in the House scenario are given below:

- **Sub-Class 1:** Alarm $\longrightarrow$ Kitchen entrance $\longrightarrow$ Fridge $\longrightarrow$ Sink $\longrightarrow$ Garage door (inside).
- **Sub-Class 2:** Stairway $\longrightarrow$ Fridge $\longrightarrow$ Sink $\longrightarrow$ Cupboard $\longrightarrow$ Sink.
- **Sub-Class 3:** Stairway $\longrightarrow$ Dining Table $\longrightarrow$ Den $\longrightarrow$ Living-room Door $\longrightarrow$ Sun-room $\longrightarrow$ Living-room door $\longrightarrow$ Den.
- **Sub-Class 4:** Den $\longrightarrow$ Living-room door $\longrightarrow$ Den $\longrightarrow$ Kitchen Entrance $\longrightarrow$ Stairway.
- **Sub-Class 5:** Fridge $\longrightarrow$ Dining Table $\longrightarrow$ Kitchen Entrance $\longrightarrow$ Fridge $\longrightarrow$ Sink

7.4.3 Subjective Assessment of Evaluation

The method defined above would, by construction, find activity-classes and the characterizing event-motifs. This begs the question as to how valid are the discovered activity-classes and the characterizing event-motifs. Our final goal is to design a system that would be able to discover and characterize human-interpretable activity-classes. Keeping this thought in mind, we performed a limited user test to subjectively assess the performance of our system involving 7 participant. For each participant, 2 of the 7 discovered activity classes were selected from the Loading Dock environment. Each participant was shown 6 example activities, 3 from each of the 2 selected activity-classes. The participants were then shown 6 motifs, 3 for each of the 2 classes, and were asked to associate each motif to the class that it best belonged to. Their answers agreed with our systems 83% of the time, *i.e.*, on average a participant agreed with our system on 5 out of 6 motifs. The probability of agreement on 5 out of 6 motifs by random guessing[1] is only 0.093.

[1] According to the binomial probability function the chance of randomly agreeing on 5 out of 6 motifs is $C_5^6(0.5)^1(0.5)^5$.

7.5 Discussion regarding Discovered Classes and Motifs

The discovered activity-classes both for the Loading Dock and the House data-sets, are subjectively semantically coherent and divide their respective activity space discriminatively. The fundamental differences between various classes in the Loading Dock environment are dictated by the fact whether the activities were of deliver 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.

Figures 7 and 8 show that the activities performed in the Loading Dock environment are structurally more well defined than those performed in the House environment. This is because our vocabulary for the Loading Dock environment consists of semantically meaningful events, which can encode the underlying activity structure efficiently. For the House environment, the events are simply the locations where a person went, and are not particularly designed to encode the underlying structure of the activities.

The discovered motifs of membership classes efficiently characterize these classes. Note that the discovered motifs for activity-classes where package *delivery* occurred, have events like *Person Places Package In The Back Door Of Delivery Vehicle* and *Person Pushes Cart In The Front Door of Building→ Cart is Full.* On the other hand event-motifs for activity-classes where package *pick-up* occurred, have events such as *Person Removes Package From Back-Door Of Delivery Vehicle* and *Person Places Package Into Cart.* Similarly, 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.

7.6 Detected Anomalies

We performed experimental analysis on the activities from the Loading Dock scenario. As mentioned in § 7.3.1, 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. We know give a detailed explanation of how, using these initially detected anomalous activities, we can

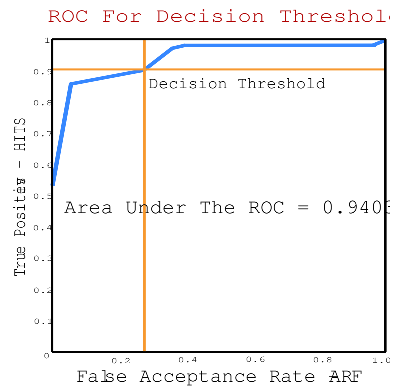

Figure 10. ROC obtained by varying Γ over a range of values.

learn a threshold for detecting new anomalous activity-class members, how valid are these detected anomalies from a human view-point, and finally, what explanations did we get for detected anomalous activities using based on the selected key-features of the activity-classes.

7.6.1 Learning Threshold for Anomalies using ROC

Using the 7 discovered activity-classes and the anomalous activities, the anomalous activities were first classified into one of the 7 activity-classes using Equation 24 and 25. Based on these activity-class labels, Γ as defined in Equation 26 was computed for all 150 activities. The *ROC* that was obtained is shown in Figure 10. The area under the obtained ROC was 0.94, which indicates a confidence of 94% in the proposed detection metric [17].

7.6.2 Analysis of Detected Anomalies

Analyzing the detected anomalous activities reveals the interesting fact that there are essentially two kinds of activities that are being detected, (1) ones that are truly *alarming*, where someone must be notified, and (2) those that are simply *unusual* delivery activities with respect to the other regular activities. Key-frames for three of the truly alarming anomalous activities are shown in Figure 11. Figure 11-a shows a truck driving out without closing it's back door. Not shown in the key-frame is the sequence of events where a loading-dock personnel runs after the delivery vehicle to tell the driver of his mistake. Figure 11-b shows a delivery activity where a relatively excessive number of people unload the delivery vehicle. Usually only one or two people unload a delivery vehicle, however as can be seen from Figure 11-b, in this case there were five people involved in the process of unloading. Finally, Figure 11-c shows a person cleaning the dock floor which is very unusual.

It is interesting to see that the proposed algorithm can detect the alarming activities. On the other hand detection of unusual activities means that the system has not seen enough instances of the activities to start considering that group as regular. Moreover, in an uncontrolled environment such as a loading dock, variance between activities is high. It is therefore plausible to believe that as the training data starts spanning the space of all regular activities, the detected number of unusual activities would reduce.

7.6.3 User Study For Detected Anomalies

To analyze how intuitive the detected anomalies are to humans, a user test involving 7 users was performed. First 8 regular activities for a subject were selected so she could understand the notion of a regular activity in the environment. 10 more activities were selected, 5 of which were labeled as regular by the system while the rest of the 5 were detected as anomalies. Each of the 7 users were shown these 10 activities and asked to label every one of them as a regular instance or an anomaly based on the regular activities previously shown. Each of the 10 activities were given labels based on what the majority agreed upon. 8 out of 10 activities labeled by the users, corresponded with the labels of the system. The probability of the system choosing the correct label 8 out of 10 times by chance is 4.4% [2]. This highlights the interesting fact that the anomalies detected by the proposed system fairly match the natural intuition of human observers.

7.6.4 Noise Sensitivity

The results presented thus far were generated using activities with hand-labelled events. However, using low-level vision sensors to detect these events will generate noise. This invites the question as to how well would the proposed system perform over noisy data. In the following, the noise analysis to check the stability and robustness of the proposed framework is presented; allowing one to make some predictions about its performance on

[2] Given that the probability of correctly choosing the true label by simply guessing is 0.5, the binomial probability states that the chance of an 8 out of 10 success is $C_8^{10}(0.5)^8(0.5)^2 \approx .0439$

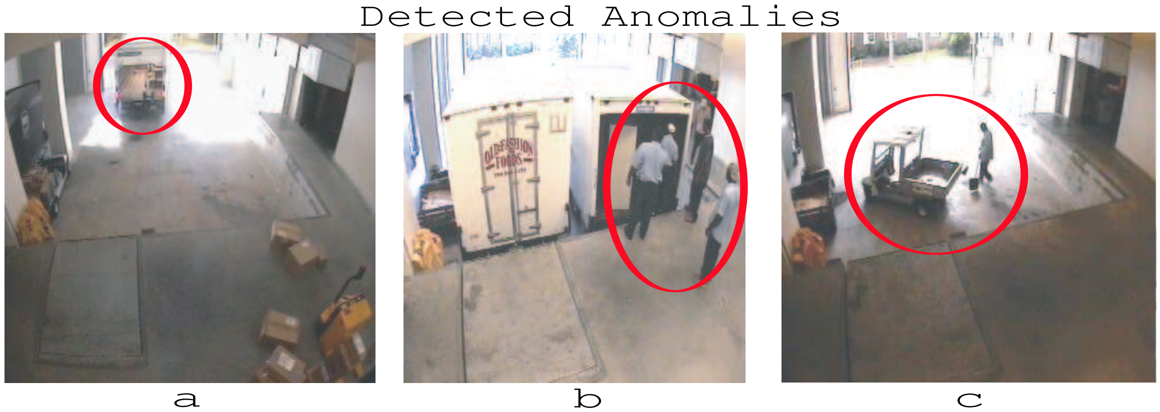

Figure 11. Anomalous Activities - (a) shows a delivery vehicle leaving the loading dock with its back door still open. (b) shows an unusual number of people unloading a delivery vehicle. (c) shows a person cleaning the loading dock floor.

data using low-level vision.

Given the discovered activity-classes and the learned detection threshold using the training set of 150 activity-instances, various types and amounts of noise to the 45 test sequences was added, and the following two tests were performed:

1. ***Regular Classification Rate:*** what percent of activities classified as regular members in the 45 ground truth test activities maintain their correct activity-class and regular-membership labels in the face of noise.
2. ***Anomaly Detection Rate:*** what percent of 45 ground truth test activities detected as anomalies still get detected as anomalies in the face of noise.

Different amounts of noise using four types of noise models, *Insertion Noise*, *Deletion Noise*, *Substitution Noise* and *Swap Noise* was synthetically generated. We generated one noisy event-symbol using a particular noise model, anywhere within a window of a time-period for each activity in the testing data set. For instance *Insertion Noise* of time period 10 would insert one event-symbol between any two consecutive event-symbols, every 10 symbols. The classification performance of the proposed system under such noise model is shown in Figure 12. The system performs robustly in the face of noise and degrades gracefully as the amount of noise increases. Likewise, the anomaly detection capability of our system in the face of synthetically generated noise is shown in Table **1**. The reason for such high detection

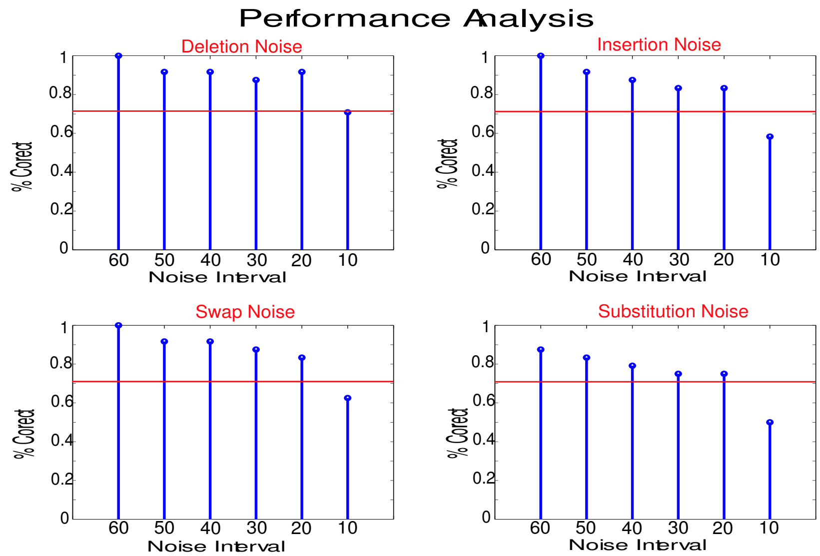

Figure 12. Performance Analysis - Each graph shows system-performance under synthetically generated noise using different generative noise models. The X-axis represents the noise interval where the amount of noise is inversely proportional to the noise interval. The Y-axis represents the percentage of regular test activities that remain regular members of the original sub-classes in the face of noise. The horizontal line in all these graphs shows the classification performance using automatically detected events as described in § 7.6.5.

Table 1. The average detection rate of the system in the face of noise.

Noise Model	%age Correct
Insertion Noise	100%
Deletion Noise	99%
Swap Noise	97%
Substitution Noise	100%

rate even with large amount of synthetic noise is that it is unlikely that an anomaly would transform into something regular when perturbed randomly.

7.6.5 Automatic Event Detection

To move one step closer towards using low-level vision, we wrote a feature-labelling software that a user uses only to label the various objects of interest in the scene such as the doors of the loading dock, the delivery vehicles and its doors, people, packages and carts. We assign each object a unique ID during labelling. The ID numbers and object locations are stored in an XML format on a per-frame basis. We also wrote event detectors that parsed the XML data files to compute the distances between these objects for the 45 test activities. Based on the relative locations and velocities of these objects, the detectors automatically decided when one of the 61 events took place.

The horizontal line in Figure 12 shows the *Regular Classification Rate* of our system over these automatically generated event sequences, *i.e.* 70.8%. The results for *Anomaly Detection Rate* for the automatically generated event sequences is 90.48%.

7.7 Anomalous Activity Explanation

Figure 13 shows the explanation generated by the system for the three anomalous activities (shown in Figure 11). The anomaly shown in Figure 11- (a) was classified to a activity-class where people frequently carry packages through the front door of the building. There was only one person in this anomaly who delivers the package through the side door. This is evident by looking at the extraneous features of the anomaly (Figure 13 - (b)) where the tri-gram *Person Full Handed* → *Person Exits from Side Door of Building* → *Person Empty Handed* captures this difference. The second tri-gram of Figure 13 - (b), (*Person Full Handed* → *Person Exits from Garage Back Door* → *Person Full Handed*) shows the fact that there was another person who went out of the garage to tell the driver of the delivery vehicle that his back door was still open.

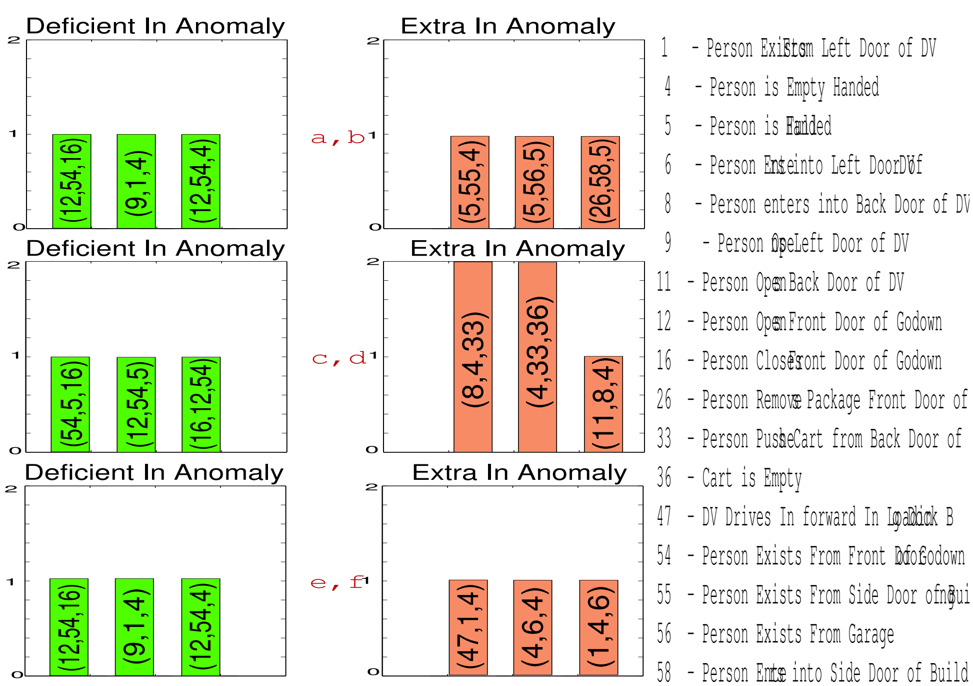

Figure 13. Anomaly Explanation - explanations generated by the system for the three anomalies in Figure 11.

The membership activity-class of anomaly in Figure 11- (b) has people frequently carrying packages through the *front* door of the building. In this anomaly, all of the workers go to the *side* door of the building. Moreover, majority of events in this anomaly were related to carts that is not one of the general characteristic of its membership activity-class. This is shown in Figure 13 - (d) by the tri-grams *Person Enters Back Door of DV* $\rightarrow$ *Person Empty Handed* $\rightarrow$ *Person Pushes Cart from Back Door of DV*, and *Person Empty Handed* $\rightarrow$ *Person Pushes Cart from Back Door of DV* $\rightarrow$ *Cart Empty*. Similarly Figure 13- (e) and Figure 13- (f) explain how anomaly in Figure 11- (c) was different from its membership activity-class.

CHAPTER VIII

CONCLUSIONS AND FUTURE WORK

This work examines the various aspects of everyday activity analysis, including activity representation, unsupervised activity discovery, activity-class characterization, anomalous activity detection and explanation and a maximally informative form. Results over extensive data-sets, collected from multiple active environments are presented, to show the competence and generalizability of the proposed framework.

The proposed representation of activities as bags of event n-grams is an attempt towards extracting *structure from statistics*, as opposed to having to script the different ways in which activities can take place. Such an approach can be helpful in large uncontrolled settings where explicitly encoding the activity structure is infeasible if not impossible. While the proposed representation of activities captures both content and order of events in activities, it does pose the problem of dealing with sparse very high dimensional data. It is evident that higher values of n would capture the temporal order information of events more rigidly, and would entail a more discriminative representation. This discriminative power would however come at the cost of an exponential growth in the dimensionality of the space. A through analysis of how to strike a balance between the discriminative power of the proposed representation and the dimensional growth that comes along with it still remains an open question and is left for future work. Moreover, the proposed representation does not cater for activities where sub-activities can happen parallel in time. Extending the bags of events representation to parallel activities is another problem which remains open. Besides, currently, there is no notion of the coherence of n-grams' events in terms of the temporal duration elapsed between them. This idea of "soft n-grams" where different n-grams would be weighted based on the time elapsed between the occurrence of events of that n-gram is also left for future work.

Making use of this representation, it is shown how different activity-classes can be discovered in an unsupervised manner, by exploiting the notion of maximal cliques in an edge-weighted graph. The empirical analysis of the discovered activity-classes in both the examined active-settings, shows that these activity-classes are meaningful to human observers. At the same time, it is evident that some of these discovered classes are more meaningful than the others. This begs the question as to how can one learn which characteristics of activities are important in terms of making a class more coherent or meaningful. Since the validity of these classes ultimately depends on how a human observer interprets it, there is a strong motivation to somehow include a human observer in the discovery of these classes, as opposed to going through it in a purely unsupervised manner. The extension of the current framework into an *Active Learning* paradigm where a user is brought into the loop is left for the future work.

The analysis of the detected anomalous activities clearly indicate that some of these activities are more anomalous than the others. Once again, being different than a majority of activities that are similar to each others does not necessarily make an activity *anomalous*. Moreover, in the current scheme, there is a strong assumption being made that the activity classes in the active setting are well-behaved and very coherent. More than often, this assumption does not necessarily hold true, which raises an important question that for such not-so-well-behaved environments, how does the proposed framework distinguish between a truly anomalous activity and a benign outlier? It is plausible to hypothesize that to answer this question, a human observer must be brought into the loop. This is again left as future work.

This work has explored the characterization of the discovered activity-classes, both from a holistic view-point and from a more granular perspective. While the event-motifs found for each of the activity-classes are maximally mutually exclusive amongst the different classes, such inter-class comparisons are not considered while finding the Typical class members of the discovered classes. Again, this would be justified only if the assumption that the activity-classes are well-behaved and are structurally cohesive. For domains where this assumption does not hold true, inter-class analysis should also be carried out for finding the Typical class members.

The idea of explaining why an anomalous activity is detected as such is interesting and can be helpful in large scale surveillance systems. While the selection of the important features in terms of which the detected anomalies are explained is a step in the right direction, a lot remains to be done in terms of learning the dependence of these features amongst one another and how that can effect the interpretability of such explanation. Moreover, in certain cases, the presence or absence of just one event in an activity, regardless of the content of the regular members of its membership class can go along a long way explaining why that activity was anomalous. Incorporating such factors in the current framework of anomalous activity explanation remains an open problem.

REFERENCES

[1] Abowd, G., "Cyberguide: A mobile context-aware tour guide," *ACM Wireless Networks*, vol. 3, no. 9, pp. 421–433, 1997.

[2] Anind K. Dey, Raffay Hamid, C. B. I. L. D. H., "a cappella: programming by demonstration of context-aware applications," in *SIGCHI*, pp. 33–40, 2004.

[3] Apostolico, A. and Bejerano, J., "Optimal amnesic probabilistic automata.," *J. of Comp. Biology*, vol. 7, pp. 381–393, 2000.

[4] Auguston, J. and Miker, J., "An analysis of some graph theoretical clustering techniques," *J. ACM*, vol. 17(4), pp. 571–588, 1970.

[5] Bailey, T. and Elkan, C., "Fitting a mixture model by expectation maximization to discover motifs in biopolymers," in *Proc Int. Conf. Intell. Syst. Mol. Biol., pp. 28-36*, 1994.

[6] Bejerano, G. and Yona, G., "Modeling protein families using probabilistic suffix trees," in *In the Proc. of RECOMB*, 1999.

[7] Bobick, A., Intille, S., Davis, J., Baird, F., Pinhanez, C., Campbell, L., Ivanov, Y., Schuette, A., and Wilson, A., "The kidsroom: A perceptually-based interactive and immersive story environment," in *Vismod*, 1996.

[8] Brand, M.E.; Hertzmann, A., "Style machines," in *SIGGRAPH*, pp. 183–192, 2000.

[9] Brand, M., Oliver, N., and Pentland, A., "Coupled hidden markov models for complex action recognition," in *CVPR*, 1997.

[10] Chi, M. T. H., "Conceptual change within and across ontological categories," *Cognitive models of Science: Minnesota Studies in Philosophy of Science*, pp. 129–186, 1992.

[11] Chiu, B., Keogh, E., and Lonardi, S., "Probabilistic discovery of time series motifs.," in *SIGKDD*, 2003.

[12] et. al., G., "Concept formation in dimensional spaces," *Lund University Cognitive Studies*, no. 26, 1995.

[13] Gusfield, D., *Algorithms on Strings, Trees, and Sequences: Computer Science and Computational Biology.* Cambridge University Press; 1st edition, 1997.

[14] Hafbauer, J. and Sigmund, K., *Evolutionary Games and Populations Dynamics.* Cambridge University Press, Cambridge, UK., 1998.

[15] Hongeng, S. and Nevatia, R., "Multi-agent event recognition," in *In Proc. of IEEE ICCV*, 2001.

[16] Ivanov, Y. and Bobick, A., "Recognition of visual activities and interactions by stochastic parsing," *PAMI Journal*, vol. 22, no. 8, pp. 852–872, 2000.

[17] Johnson, A. and Bobick, A., "Relationship between identification metrics: Expected confusion and area under a roc curve," in *In Proceedings of IEEE International Conference on Pattern Recognition*, 2002.

[18] Kleinberg, J., "Authoritative sources in a hyperlinked environment," *Journal of the ACM*, vol. 46, 1999.

[19] Koral Ilgun, R. A. Kemmerer, P. A. P., "State transition analysis: A rule-based intrusion detection approach," *IEEE Transaction on software engineering*, pp. 188–199, 1995.

[20] Lee, W. and Stolfo, S., "A framework for constructing features and models for intrusion detection systems," *ACM Transactions on Information and System Security*, vol. 3, November 2000.

[21] Mei Chen, Takeo Kanade, D. P. H. A. R., "Anomaly detection through registration," in *Pattern Recognition 32(1): 113-128*, 1999.

[22] Minnen, D., Essa, I., and Starner, T., "Expectation grammars: Leveraging high-level expectations for activity recognition," in *IEEE Conference on CVPR. Madison, WI.*, 2003.

[23] Moore, D., Essa, I., and Hayes, M., "Context management for human activity recognition," in *Proceedings of Audio and Vision-based Person Authentication*, 1999.

[24] Motzkin, T. and Straus, E. G., "Maxima for graphs and an new proof of a theorm of turan.," *Canad. J. Math.*, vol. 17.

[25] Oates, T., "Peruse: An unsupervised algorithm for finding recurring patterns in time series," in *IEEE ICDM, Japan.*, 2002.

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

[27] Raghavan, V. and Yu, C., "A comparison of the stability characteristics of some graph theoretic clustering methods.," *IEEE Trans. on PAMI*, vol. 3, pp. 393–402, 1981.

[28] Ron, D., Singer, Y., and Tishby, N., "The power of amnesia: learning probabilistic automata with variable memory length.," *Machine Learning*, vol. 25, pp. 117–149, 1996.

[29] Rosch, E., C.Mervis, Gray, W., Johnson, D., and Boyes-Braem, P., "Basic objects in natural categories," *Cognitive Psychology*, vol. 8, 1976.

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

[31] Schank, R., *Dynamic Memory: A Theory of Reminding and Learning in Computers and People*. Cambridge University Press, 1983.

[32] Shi, Y., Huang, Y., Minen, D., Bobick, A., and Essa, I., "Propogation networks for recognizing partially ordered sequential action," in *In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2004.

[33] Smith, E. E. and Medin, D. L., *Categories and Concepts, pp-147*. Harvard University Press, 1981.

[34] Smith, L., *A tutorial on Principal Components Analysis*. 2002.

[35] Starner, T., Weaver, J., and Pentland, A., "Real-time american sign language recognition using desk and wearable computer based video," *PAMI*, vol. 20, pp. 1371–1375, December 1998.

[36] Stauffer, C. and Grimson, W., "Learning patterns of activity using real-time tracking," *IEEE PAMI*, vol. 22, no. 8, pp. 747–757, 2000.

[37] T. Landauer, P. F. and Laham, D., "Introduction to latent semantic analysis.," *Discourse Processes*, vol. 25, pp. 259–284, 1998.

[38] Weinberger, M., Rissanen, J., and Feder, M., "A universal finite memory source," in *IEEE Trans. Inform. Theory, vol. IT-41, pp. 643–652, 48*, 1995.

[39] Yamato, J., Ohya, J., and Ishii, K., "Recognizing human action in time-sequential images using a hidden Markov model," in *CVPR1992*, pp. 379–385, 1994.

[40] Zhong, H., Shi, J., and Visontai, M., "Detecting unusual activity in video," in *In Proc. of IEEE CVPR*, 2004.