Thesis for the Degree of Doctor of Philosophy

Intelligent Human Behavior Analysis based on Group and Individual Activities

Iram Fatima

Dept. of Computer Engineering
Graduate School
Kyung Hee University
South, Korea

Fall 2013

Abstract

Behavior refers to the action or reaction of any material under given circumstances and environment. Human behavior has been increasingly highlighted for social activities in many areas such as social computing, intrusion detection, fraud detection, event analysis, and group decision making. The activities performed by individuals can be characterized by duration, frequency, sequential order, temporal order, and other factors such as location, cultural background, age, and gender. With the boom of information technology we experience today comes an explosion in the amount and richness of data recorded related to human activities in both group and individual aspects.

In group activities diffusion process has attracted much attention and a lot of research efforts have been made in this field from all areas of academic interest, such as physics, mathematics and computer science. The collective diffusion process for these various information types seems chaotic. Each piece of information has separate spread process according to its type, associated constraints, and importance. For example, some information is independent of any competition (e.g., TV news) while some ideas, opinions, and products compete, with all other content (e.g., product adoption, political elections) for the scarce attention of the users. Previously, researchers simulated these types of information independently in separate diffusion models as a single objective problem to analyze a social network.

Moreover, usually inhabitants perform individual activities in a sequential manner characterized by preceding and following activities to identify their influence on each other. For example, taking medicine is very likely followed by eating, and brushing teeth is usually preceded the face washing activity. Thus, learning of user behavior by means of a sequence of actions is highly desirable and is not yet available.

In this thesis, first the diffusion is formulated as a multi-objective optimization problem to model the information spread closer to a real social network. A multi-objective diffusion model (MODM) is proposed that assigns a value of importance to each individual according to his information manipulation and propagation ability. The goal of MODM is to selfishly maximize the amount of information possessed by each individual during communication. The key difference from earlier studies is that multiple-objectives are achieved in terms of diverse information spread and calculation method to measure the propagation capabilities of individuals.

For behavioral analysis based on individual activities, the behavioral patterns are extracted from the day to day performed activities in a sequential manner with the help of data mining techniques. Sequential pattern mining algorithm is applied of behavior modeling from the activity log. In our proposed framework, each sequence is a set of activities performed in a temporal order of three days for consistent sequence prediction. Finally, the sequential activity trace is utilized for behavior learning to predict the future actions. A Conditional Random Fields (CRF) algorithm is designed for ongoing activities as labeled sequences and future actions as observations. Therefore, the analysis of the history information transmitted by users' activities helps in discovering the routine behavior patterns and future actions of inhabitants in a home environment. For empirical evaluation, the experiments are performed on two real datasets from the CASAS smart home. The identification of significant behavioral sequential patterns and precise action prediction enables the observation of the inherent structure present in users' daily activity for analyzing routine behavior and its deviations.

Contents

List of Figures

List of Tables

List of Acronyms

In alphabetical order:

ANMO- Average Normalized MultiObjective

ANMOS-Average Normalized MultiObjective Score

AL – Activity log

CRF- Conditional Random Fields

CS- Closure Count

HDF- Holland's hyperplane defined function

GA- Genetic Algorithm

GADM- Genetic Algorithm based Diffusion Model

HMM- Hidden Markov Model

ML- Machine learning

MODM- MultiObjective Diffusion Model

Chapter 1

Introduction

1.1 Overview

"The future influences the present just as much as the past."

— Friedrich Nietzsche (Philosopher, 1844–1900)

Humans are able to predict the future actions and identify the consequences of our actions and potential relationships to our goals and objectives. Even though uncertainty factor is present at different levels, these actions and activities refine human high-level behavior. Most of the time, our actions are purposeful to the certain goal and sensitive to the revelation of new information in the future. Humans are able to anticipate the possible outcomes of our performed actions and may possible to select intelligently appropriate actions that lead to desirable goals. The study of intelligence leads the human behavior as goal-directed adaptive actions [1]. This reasoning is needed not only as a basis for intelligently defining our own actions and activities, but also for being able to infer the intentions of others, their probable reactions to our behaviors, and rational possibilities for group behavior [2].

In general, behavior can be define as an action or reaction of any material under the given circumstances and environment. Similarly, it is true for the humans under the certain situations. Human behavior has been increasingly highlighted for social activities in many research areas such as social computing [3], intrusion detection, fraud detection [4], event analysis [5], and group decision making [6]. The set of activities as a human behavior is shown in Figure 1.1. In both natural and social sciences and its applications, multiple human behaviors from either one or multiple

subjects often interact with one another. Such behavior interactions may form interior driving forces that impact underlying social activities or situations, and may even cause challenging problems [7]. For example, during the online shopping process, the customer and the merchant communicate with each other to guarantee the success of an online transaction through the inspection of a trusted third party. Similarly the communication behaviors are widespread in many applications including interactions in social communities and multi-agent systems.

On the other hand, to the best of our knowledge, along with qualitative research in behavior sciences [8], human behavior representation has been a typical topic in the artificial intelligence (AI) community. The major efforts have been made for behavior coordination, action reasoning and composition [9]. For instance, the research [10] used the application-independent software connector to specify multi-agent societies rather than agent behaviors. In [11] authors discussed reasoning about action based on a modified situation calculus. Moreover, behavior modeling actually refers to behavior recognition and recreation [12] instead of representation and checking, which is different from focus of the thesis. Limited work can be identified on representation and inspection [13] complex behavior structures and communications. Following paragraphs provide more insight knowledge of group and individual activities along the challenges.

Group Activities: The growth of information technology has been shown best during the current decade. We also experience and our interaction most of the time generate the huge amount of data through many possible devices and connections.

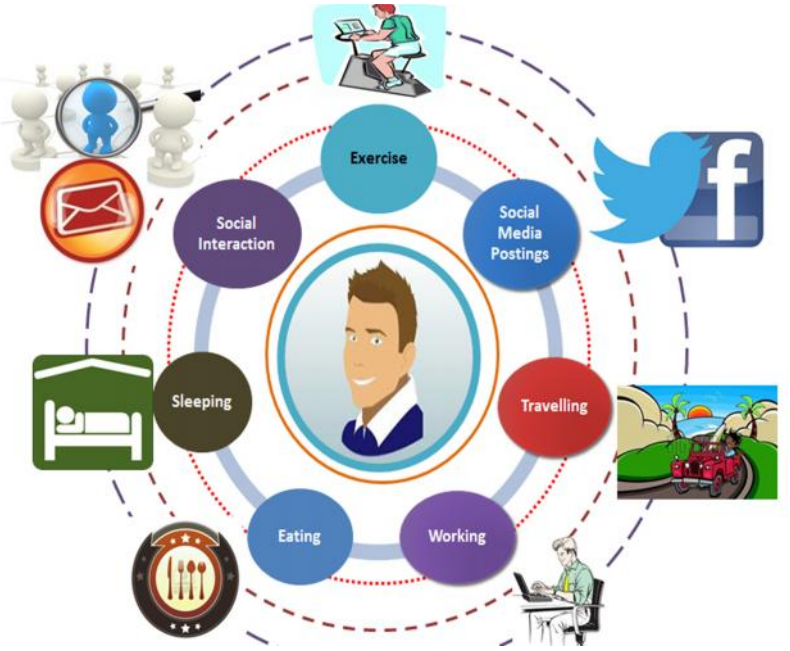

Figure 1.1: Set of Human Activities

The necessary ingredients for this phenomenon have been fully realized recently: ubiquitous Internet connectivity, virtually limitless data storage, and powerful mobile devices [14]. At the same time, all the ingredients are inexpensive, widely available, and the technology behind them reached a level of maturity where the general population uses the devices in everyday life. As a result, interactions through social network sources are used by nearly everyone, even in less developed parts of the world. The analysis of social network reveals location, demographic and other physical and social factors of people around them. Due to such data availability, social aspects of computing are gaining prominence. According to a recent Nielsen study [15], the average Facebook user spent nearly 7 hours and 45 minutes on the site per month more than on any other single site on the Internet. To put these statistics in context, the average amount of time a person spends online per month is 30 hours. All these group activities based on

human interactions with other human beings can be a great source to analyze the behavior of single individual within group or behavior and adaption of societies according the information in their surroundings.

Individual Activities: Epidemiological, clinical, and laboratory research have provided convincing evidence that physical activity based behavior analysis has numerous beneficial effects on physical health, psychological well-being, and overall quality of life. The activities performed by individuals can be characterized by duration, frequency, sequential order, temporal order, and other factors such as location, cultural background, age, and gender [16]. To recognize these activities, smart home technology is playing its role to identify the performed human individual activities. Important attributes that must be considered while analyzing the activities are their nature and temporal differences among different activities. For example, in Figure 1.2(a), eating requires more time than taking medication, which may only take a couple of minutes. Periodic variations may occur in daily, weekly, monthly, annual, and even seasonal activities [17]. For example, cleaning is more likely to occur on weekends than on weekdays. Parallel activities may occur, in which one user can perform more than one activity in a single time unit [18], as shown in Figure 1.2(b), where eating and taking medicine are performed in parallel with watching TV. Sequential activities are characterized by preceding and following activities to identify the influences of different activities on each other. For example, taking medicine is very likely followed by eating. Physical location also affects activity flow. For instance, kitchen cleaning involves a different sequence of steps than washroom cleaning. Cultural habits may influence activity patterns. For example, in some cultures, people prefer to take a nap after lunch, while others prefer to have a cup of coffee or tea [17].

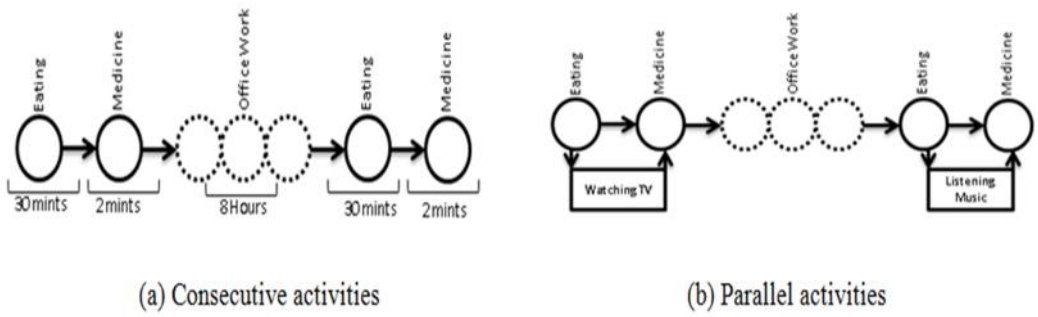

Figure 1.2: Sequences of Consecutive and Parallel Activities

A combination of evolutionary, data mining and machine learning is now poised to effectively answer questions related to human behaviors in the context of group and individual activities. This thesis focuses on methods that overcome the challenges of analysing the human lifelog to capture important phenomena with a significantly increased level of detail and predictive power. The underlying theme of our work is the unification of diverse data with the help of evolutionary algorithm over large numbers of individuals for group activities. Moreover, the set of individual activities are analysed with the help of data mining based reasoning for significant pattern discovery and action prediction. The proposed technique shows that the mined patterns can be leveraged in predictive models of human behaviour at diverse levels. Furthermore, the activity data is now available to machines in massive volumes either in form of group or individual format.

1.2 Motivation

The basic assumption of this work is that outcome behaviors are the result of interactions between individuals and physical environmental factors within a specific behavior-setting. However, the degree to which a person's decision to be physically active depends on physical surroundings and do not depend on personal influences in group activities for exchanging information.

Diffusion process for information exchange in social networks as group activities has attracted much attention and a lot of research efforts have been made in this field from all areas of academic interest, such as physics, mathematics and computer science [19-21]. It is widely believed that user-to-user exchanges, also known as “word-of-mouth”, can spread contents, ideas, or information widely and quickly throughout the network [22][23]. The fundamental purpose of each diffusion process is to differentiate a set the individuals on the basis of their social ability for information manipulation and propagation [24]. In most of the diffusion models the notion of ‘information’ is restricted to a single unit during the diffusion process to evaluate the importance of individuals in a social network. Conversely, in real social networks, individuals communicate their ideas and feelings in a spur of a moment with various people like family, friends, relatives, neighbors, and colleagues in homes, offices, universities, shopping malls, and hospitals. So in a true social network, a variety of information like news, rumors, gossips, stories, and announcements is manipulated and spread at the same time. The flow of behavior analysis based on group and individual activities is shown in Figure 1.3.

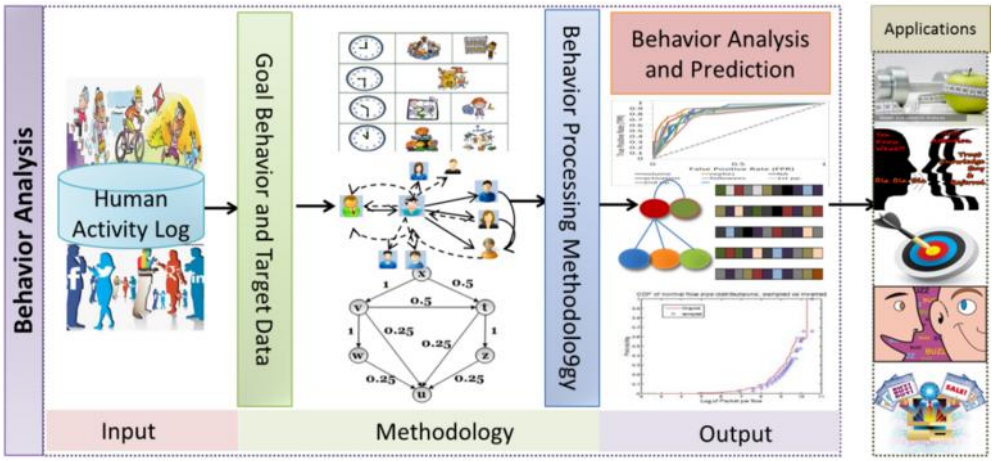

Figure 1.3: Flow of Behavior Analysis

On the other hand, individual activities are performed by inhabitants in a sequential manner characterized by preceding and following activities to identify their influence on each other [25]. For example, taking medicine is very likely followed by eating, and brushing teeth is usually preceded the face washing activity. Therefore, the activity log in terms of performed activities can be effectively analyzed to discover the sequential behavior patterns. The identified patterns provide the significant list of action that mostly occurs together in daily routine to support the health maintenance and functional capability of individuals. In a particular scenario, the daily routine of inhabitant Mr. Ben, if the significant sequential behavioral pattern is: (wakeup, exercise, bathing, breakfast, medication), reflects that Mr. Ben's activities includes daily exercise before breakfast and he is on constant medication. In this case, the care givers can easily identify the missing exercise and medication routines after analyzing his lifestyle based on frequently performed activities. Furthermore, assuming that human beings perform behaviors based on habits, it could be inferred that patterns describing past and present behaviors will define the future actions as well.

1.3 Problem Statement

The collective diffusion process in group activities for multiple types of information with in single diffusion model seems chaotic. Each piece of information[1] has separate spread process according to its type, associated constraints, and importance. For example, some information is independent of any competition (e.g., TV news) while some ideas, opinions, and products compete, with all other content (e.g., product adoption, political elections) for the scarce attention of the users. Previously, researchers simulated these types of information independently in separate diffusion models as a single objective problem to analyze a social network [22][23].

[1] Information and piece of information are interchangeable used throughout the thesis

Most of the approaches spread ‘information’ as a single unit with ‘active’ or ‘inactive’ status in order to group the individual in to two categories as shown in Figure 1.4 with the help of white and green colors. The white color demonstrate the inactivation and green reflects the activation if individuals within group activities. Therefore, at the end of the diffusion process the individuals with the ‘active’ status achieve the single objective of the diffusion whereas, ‘inactive’ individuals has no effect of the diffusion process [22][23]. However, it is intractable to distinguish among ‘active’ individuals in order to find the differences between them according to their network property. Furthermore, in a real social network people don’t lie between two status of either ‘active’ or ‘inactive’ to show their significance in the network. Instead, more granularity of individual importance is required to find the differences between them that can reflect their information propagation capability in the network. The situation becomes more complex for single objective diffusion models when individuals propagate multiple types of information.

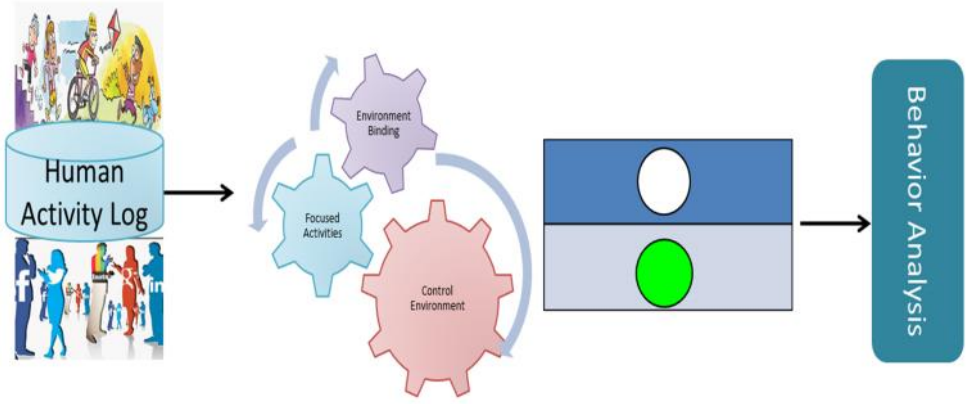

Figure 1.4: Conventional System of Behavior Analysis

At present, none of the existing diffusion models are able to comprehensively handle the aforementioned problems. Therefore, in this study, first the diffusion process is formulated as a multi-objective optimization problem to model the information spread closer to a real social network. A multi-objective diffusion

model (MODM) is proposed that assigns a value of importance to each individual according to his information manipulation and propagation ability. The goal of MODM is to selfishly maximize the amount of information possessed by each individual during communication. The key difference from earlier studies is that multiple-objectives are achieve in terms of diverse information spread and calculation method to measure the propagation capabilities of individuals in group activities.

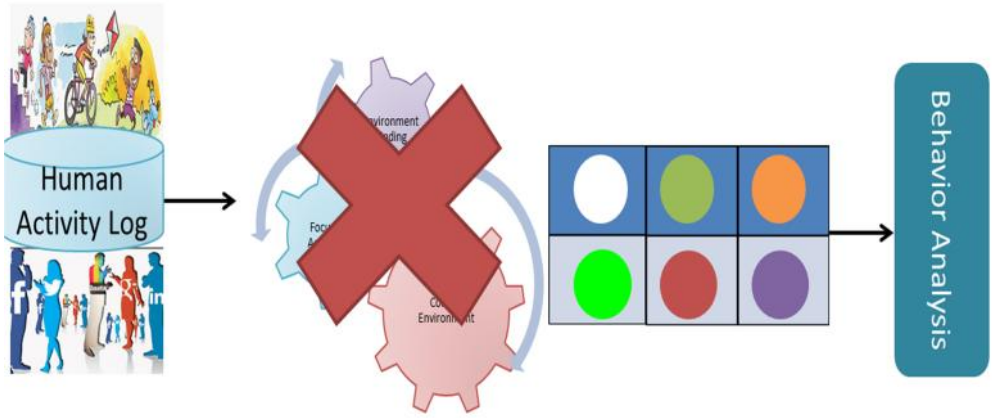

Figure 1.5: Proposed Approach for Behavior Analysis

In individual activities, learning of user behavior by means of a sequence of actions is highly desirable and is not yet available. The prediction about future actions allows caregivers to take proactive actions for the wellbeing of inhabitants after analyzing their healthy or unhealthy routines. Thus, according to the routine of Mr. Ben (i.e., Section 1.2), after his exercise activity the most likely activity is of having a breakfast and the framework can remind him to measure his blood pressure and heart rate just before breakfast, if required. However, there is a shortage of formal, systematic and unified behavior modeling and analysis methodologies based on daily life activities. So far, most of the existing applications [22-25] relate an action to the set of sensor values instead of relating the actions among themselves. In

conclusion dependability of physical environment can be avoaided while analyzing the human behavior along with the propagation of multiple types of information instead of single information as shown in Figure 1.5.

1.4 Contributions

Our contributions in this thesis dissertation for group activities are in four folds as described follows.

- First, a novel multi-objective diffusion model is proposed based on Genetic Algorithm (GA) where the GA's objective function is designed to handle more than one objective at a time in group activities.
- Second, a method to propagate the multiple types of information (i.e., independent, mutually exclusive, and competing) is introduced in binary schema formats. The HDF (Hyper Defined Function) is used to model the type of information in binary chromosome. This gives the flexibility in terms of modeling the set of information with variations in bits of string according to their types and constraints.
- Third, evolution fitness criterion of proposed GA is designed in order to calculate the multi-objective score (value of importance) of each individual. In real group activities, different aspects of information transmission are important to determine which individuals have high worth in multiple types of information spread. Therefore, the proposed evolution fitness criteria of GA are based on information score, influence, and diversity.
- Fourth, an information history log is maintained during the crossover operation of proposed GA for each individual to keep track of information flow during the whole diffusion process. This facilitates the evolution fitness criteria to calculate the true information value of each individual during their interactions from binary chromosome. Our proposed MODM, to achieve multi-objectives by propagating a set of information with various

evaluation criteria to the best of our knowledge has never been applied before. To validate the advantages of MODM, the extensive experiments are performed on a real world dataset. The solution given by proposed model, more realistically reveals the modeling of complex and nonlinear phenomena of information exchange to affect the total information worth of each individual. The experimental results show superiority of MODM over single information propagation and single evaluation criteria as discussed in Chapter 5.

In order to predict the future actions by using daily life activities of indiviuals, a novel and unified framework is proposed to analyze user behaviors.

- First, the behavioral patterns are extracted from the day to day performed activities in a sequential manner with the help of data mining techniques. The sequential pattern mining algorithm is applied by modifying it according to the requirements of behavior modeling from the activity log. In our proposed framework, each sequence is a set of activities performed in a temporal order of three days for consistent sequence prediction.
- Finally, the sequential activity trace is utilized for behavior learning to predict the future actions. A Conditional Random Fields (CRF) algorithm is designed for ongoing activities as labeled sequences and future actions as observations. Therefore, the analysis of the history information transmitted by users' activities helps in discovering the routine behavior patterns and future actions of inhabitants in a home environment.

For empirical evaluation, experiments are performed on real datasets. In order to show the significance of CRF the experiments are performed with other classification algorithm such as Hidden Markov Model (HMM), Neural Networks

(NN), and Support Vector Machine (SVM). The results show that our proposed framework identifies the significant behavioral sequential patterns and precise action prediction. This enables the observation of the inherent structure present in users' daily activity for analyzing routine behavior and its deviations.

1.5 Thesis Organizations

This dissertation is organized in to following chapters.

Chapter 1 - Introduction: In this chapter a brief introduction of human group and individual activities and their nomenclature is illustrated to analyze the behavior and future action prediction. Furthermore, the limitation of existing methods and current challenges are discussed in the behavior analysis domain. After that an overview of contributions is given for this dissertation.

Chapter 2 - Related Work: Chapter 2 gives an overview of related work in the area of behavior analysis based on group and individual activities. The available technologies and learning models for behavior analysis and prediction are briefly discussed. Furthermore, different application domains are reviewed to show the applicability of human behavior analysis.

Chapter 3 – Group Activity based Behavior Analysis and Prediction: Chapter 3 explains the proposed evolutionary algorithm for information diffusion in group activities. The problem is formulated into mathematical form and then explained the different information types, different evaluation criteria based on information categories.

Chapter 4- Individual activity based behavior analysis and prediction: Chapter 4 presents the data mining and machine learning methods to identify the significant behaviors from user log data and then how to use these behaviors to predict the future actions.

Chapter 5 - Implementation and Results: Chapter 5 provides the details about implementation and results of behavior analysis based on group and individual activities. This helps in identifying the behavioral importance of individuals in their group interactions. Similarly results based on behavior analysis of individual activities provide details about their significant sequential behavior patterns and future actions.

Chapter 6- Conclusion and future directions: Chapter 6 concludes the dissertation along with the future directions.

Chapter 2

Related Work

Humans are able to express, perceive, process, and memorize a rich set of behavioral cues that enable natural and multimodal communication and social information exchange via group and individual activities. In this chapter, first we briefly discussed about the existing methodologies behavior analysis based on group and individual activities. Finally, we explain a list of applications based on human behavior analysis. However, stronger emphasize is given to the methodologies and techniques related to group and individual activities based behavior analysis because they are the main focuses of this dissertation.

2.1 Information Cascade Based Behavior Analysis

Initially human beings are aware of cascading process at an anecdotal level [26]. The systematic study of cascading process starts in the middle of the 20th century as the diffusion of innovations. Hence information cascade is an important process of social sciences by which new idea and behavior spread within group of people. The initial research in this area was empirical but in the start of 1970s economists and mathematical sociologists such as Thomas Schelling and Mark began to formulate mathematical models for such mechanisms by which information spread within a group of people [26]. It could be new religion beliefs, adoption of new technology, spread of new medical innovations, and sudden success of new products in market, rise of particular candidate in media or in politics. All these phenomena share some hidden properties about the interaction of individuals within group or societies [27]. The actual process starts with the adoption of new phenomena with some early adopters and then more and more people begin to adopt the new concept as they observe their friends, neighbors, or colleagues doing so. As a result the new phenomena spread through the

population contagiously from person to person and turn into behavior with the dynamics of an epidemic. The example of the cascade of interaction in education system is shown in Figure 2.1. The flow of information starts from the regional specialist and goes down to students in order to improve the education system of schools.

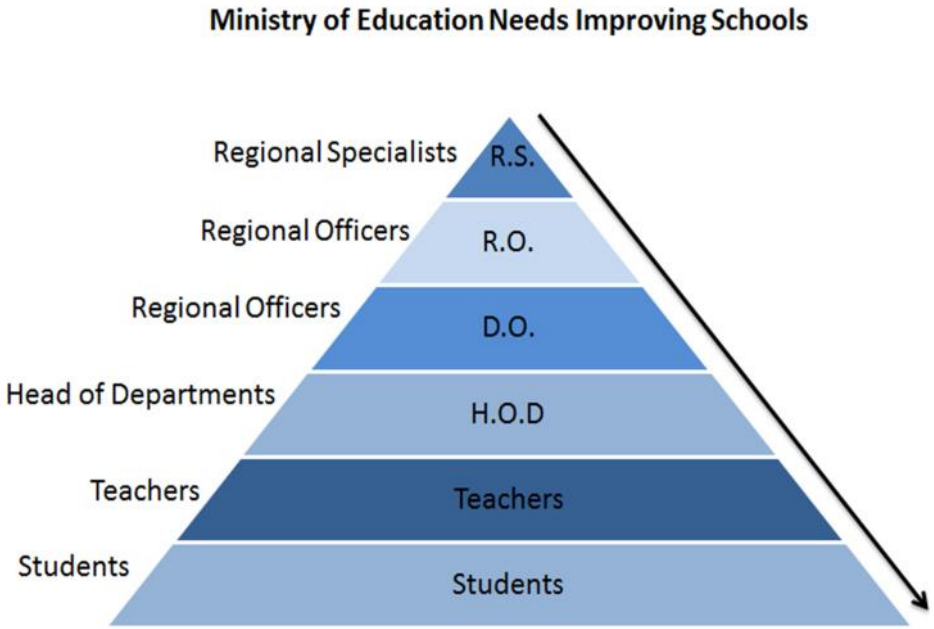

Figure 2.1: The Cascade of Interaction in Education System

So researchers start to put their efforts in designing and seeking the models that capture observed data in diffusion. This results in helping the domain experts, at a fundamental level, how the cascading process works for the spread of new ideas. Such models help the researchers to predict the success or failure of new innovations in early stages [28]. As a result they can work on the cause of underlying process in case of failure of spread if new ideas to increase the chances of success. Smart Models of information diffusion and influence maximization have been studied in many different forms, for example, the transmission of political opinions and news in political science [29], the diffusion of innovations in management science [30], the value of novel information in organizational behavior [31], and the propagation of obesity and smoking

behaviors in public healthcare [32]. In past, several models have been introduced to simulate information diffusion through a network [33-36]. The linear threshold model [33] and independent cascade model [34][35] are the most widely used diffusion methods. With the passage of time many extensions of the basic models have been proposed by a set of researchers. For the sake of basic understanding of our proposed diffusion model in this thesis, we discuss the basic working of these two models in particular.

2.1.1 Independent cascade Model (ICM)

Independent Cascade Model (ICM) is a stochastic information diffusion model where the information flows over the network through cascade. A lot of research [37][38] has been based on ICM model where the nodes have two states either "active" or "inactive". Initially whole group of people supposed to be inactive at the start of cascading process. The activation means that node is influenced during the diffusion process and inactive means either the node is unaware of the information or not influenced. The chances for the activation of the node depends of the influence probability other nodes have on that particular node. Let's consider the simple example of ICM, in which the chances that an individual node is activated by a newly neighbor does not dependent on the set of its all neighbors. In other words each node has a full chance to make its neighbor active without dependency of any other node. So if we have initial active set A_1 then the cascading process start in a series of time steps. At each time *t*, any activated node *v* can attempt to activate each inactive node *u* for which $v \in N(u)$. So the activation of *u* depends on the probability $p(u,v)$, if the probability threshold is low than the coming influence value then *u* becomes active at the next time step. Either *u* becomes active or not, nodes *v* and *u* have no contact with each other during the remaining diffusion process.

At the beginning of ICM process, information is provided to few nodes by assuming them the most important or highly influential nodes within a group. These nodes are

called seed nodes. It is assumed that after receiving the information these nodes become active and they propagate the information to their neighbors. In each discrete step, an active node tries to influence one of its inactive neighbors. Irrespective of activation or inactivation, the same node doesn't get a chance again to activate the same neighbor again till the end of diffusion process. The success depends on the propagation probability of their tie. Propagation Probability is a measure by which one can influence the other. Each edge has different value for propagation probability; however for experimental purpose sometimes the propagation probability consider to be same for each node.

2.1.2 Linear Threshold Model

Linear Threshold Model (LTM) is another well-known stochastic information diffusion model where the information flows over group of people through cascading process [39][40]. Similar to ICM, the nodes in LTM have two states either "active" or "inactive". Initially whole group of people supposed to be inactive at the start of cascading process. The activation means that node is influenced during the diffusion process and inactive means either the node is unaware of the information or not influenced. The chances for the activation of the node depend of the influence probability of all of its neighbors. . Let's consider the simple example of LTM, in which the chances that an individual node is activated depends influence of all its neighbors. In the linear threshold model [36] a node is influenced by each neighbor according to a given weight. Each node chooses a threshold value uniformly at random from interval [0, 1]; this represents the weighted fraction of node neighbors that must become active in order for a node to become active. Given a random choice of threshold and an initial set of active nodes, the diffusion process unfolds deterministically in discrete steps. In step '*t*', all nodes that are active in step *'t-1*' remain active and activation of any node depends on the total weight of its active neighbors that must be above threshold value. If we consider the generalization of the LTM then setting of influences weighs on neighbors can be set differently. Then the decision about the fraction of neighbors required to make a node

active can be chosen uniformly at random. This means that weighted fraction of v's neighbors that must adopt the behavior before *v* depends of uniform criteria that may vary from time to time.

In addition to ICM and LTM, recently researchers have introduced the GA based diffusion model that is independent of the propagation probability and weighted parameters.

2.1.3 GA based Diffusion Models

Genetic Algorithms mimic the same strategy to solve a problem as nature uses [41]. Every living organism built up from the tiny building blocks of life. These blocks are made of genes. Each gene represents characteristics of organism and decides about its skin color, eye color, height, body structure and face features. These genera are connected together into chromosomes. When two organisms mate they share their genes to offspring. The offspring may end up having half the genes from one parent and half from the other as a result of recombination. Mutation of a gene is very rare; normally it does not affect the development of offspring but very occasionally it express in the organism as complete new trait. So inspired by nature, GA use the same combination of selection, recombination and mutation to evolve a solution to a problem.

The flow of GA begins with the population initialization to encode the chromosomes according to problem domain [41]. These chromosomes are then evaluated to represent a solution to the target problem. Reproductive opportunities are applied to chromosomes which can result in better population for final solution than those which leads to poor outcomes. The flow of standard GA is shown in Figure 2.2 and details about the operators are given in subsequent subsections.

GA based diffusion model can be used for framework for a large, rich class of diffusion models with binary chromosome and one point crossover operation. These models can deal with both static and dynamic diffusion interaction among groups. Consider that dynamic group of people consists of a set of individuals $V = \{v_1, ..., v_n\}$ interacting with

each other at different time stamps. Each individual in the group of people is mapped with the help of binary chromosome as shown in Figure 2.3.

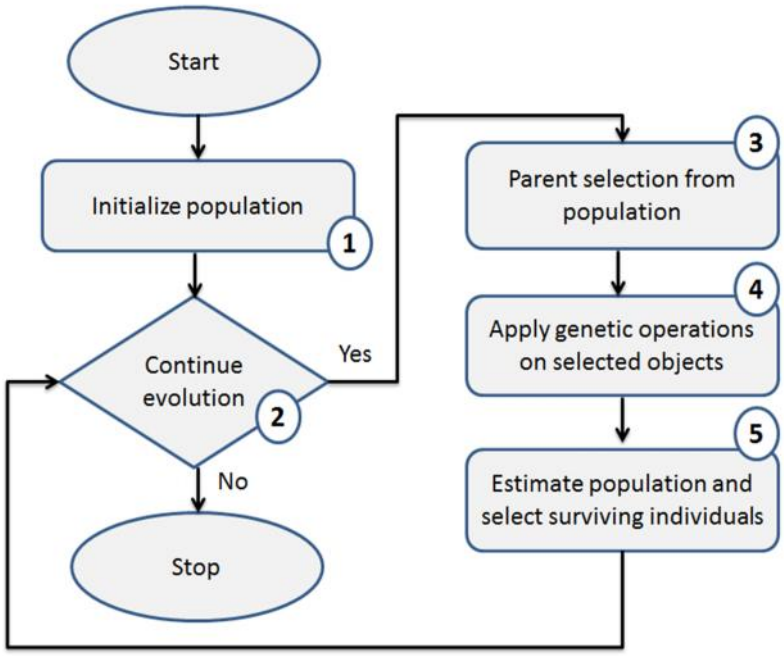

Figure 2.2: Flow of Genetic Algorithm

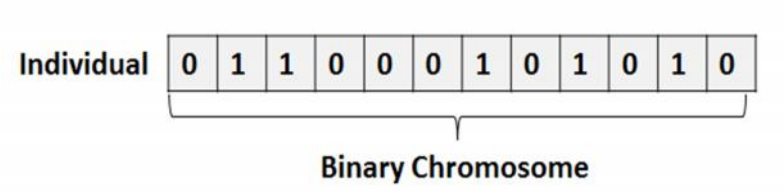

Figure 2.3: Individual Representation as Binary Chromosome in GA

Let that state of individual is a binary chromosome of some particular length in GA population. The initial values of the chromosome can be set to all zeros and GA population can be initialized with some random distribution. To evaluate the importance of each individual during the process of information diffusion, the objective

function is defined that assigns a score to any state of individual during all stamps of its interaction.

A genetic algorithm based diffusion model [42] was introduced to combine the advantage of a genetic algorithm paired with a form of Holland synthetic hyper plane-defined objective functions for a parameter free diffusion process. Chromosomes in the genetic algorithm represented individuals in a network, and the crossover operator modeled the interactions between them. Information diffusion proceeded in the crossover operation in all time stamps. During a crossover process, the tail of a chromosome containing some information is replaced when the sum of information is lower than that of the new tail. This causes the original information on the chromosome to be totally lost, although it may not conflicted with the new incoming information. The information value of each individual is calculated by adding the score of each gene within a chromosome.

In order to create realistic models for diffusion process, it is important to train with a true picture of the social interactions between individuals and the parameters that affect the propagation process. In the cases of independent cascade and linear threshold models, two kinds of data, a social network and probabilities to the edges capturing the degree of influence among individuals, is required. For example, if edge (*v, u*) has 0.45 probability that user *'v'* influences *'u'* and thus *v*'s action will propagate to *'u'* with a fix probability. In real life, edge probabilities are not available for a social network, so previous work either makes assumptions about these probabilities or uses other heuristics to calculate them. This poses a big problem of estimating this probability from the real data. GADM[42] proposed a parameter free diffusion process with the help of genetic algorithm and Holland hyper plane defined function. However, the above described methods use the diffusion process to achieve single objective according to their domain of interest. That's why, these methods are not appropriate to find the optimized diffusion solution for more than one objective under single diffusion process.

In addition to group activities, individual activities are one of the important aspects in user behavior analysis. Data mining and machine learning technologies can make use of information on the individual activity for example, activities of residents of a nursing home or elderly patients living alone at home. The analysis of these activities can play an important role in providing better healthcare services according to current significant action. So in subsequent sections we discuss the data mining and machine learning techniques for analyzing the individual activities.

2.2 Pattern Mining Based Behavior Analysis

Each human being has his own personality based on his social, cultural, religious and educational background. Therefore, everyone differs in terms of psychology, behavior, physiology, personality, strengths, interests and values [43][44]. This explains the reason why people will have their own distinct requirements. Generally speaking, it is hard to conclude single and common rule of behavior superficially within short term. But some patterns for certain kind of persons can be obtained after a long term observation. Traditionally, approaches of capturing user behavior including survey, interview, group discussion, experiment, observation, record and so on are adopted commonly. However, nowadays data mining has been introduced and converged with them.

Pattern mining is one of the most important topics in data mining which is used to find existing patterns in data. The main theme is to extract the chunks of knowledge scattered and hidden within different dimensions of data. Given a language of patterns, the traditional pattern mining task is to identify all patterns in this language that fulfill a given constraint on a database. The important question is: what are the important features that a reliable for finding the patterns from dataset and what are the needs of those patterns for different application domains. In order to resolve these goals many techniques have been developed so in this section we discuss the most well-known method for pattern mining with their basic understanding according to the scope of this thesis research.

2.2.1 Behavior Frequent Pattern Mining

Frequent pattern mining is one of the most common concepts in pattern mining branch of data mining. A lot of other mining tasks and theories evolve from this concept. In frequent pattern mining an itemset or a group of elements that represents together a single entity [43]. A frequent itemset is an itemset that occurs frequently according to the user defined threshold. Two major properties that help to find the frequent itemsets are (a) every subset of a frequent itemset is also frequent. This is made possible because of the anti-monotone property of support measure – the support for an itemset never exceeds the support for its subsets. (b) If we divide the entire database in several partitions, then an itemset can be frequent only if it is frequent in at least one partition. An itemset is maximal frequent if none of its immediate supersets is frequent. Furthermore, an itemset is closed if none of its immediate supersets has the same support as the itemset. The example of frequent itemset mining is shown in Figure 2.4.

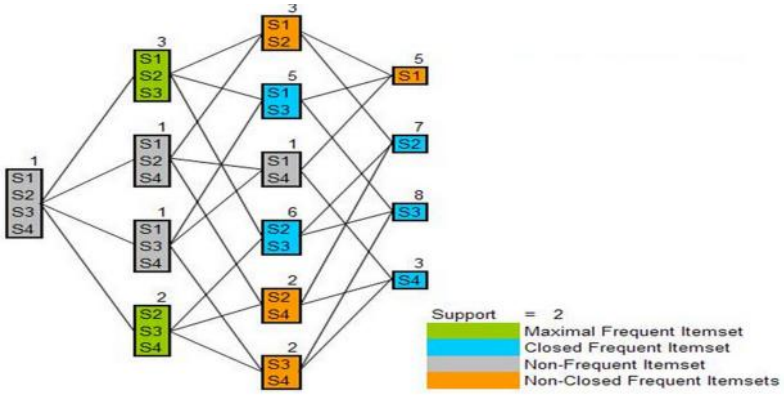

Figure 2.4: Example of Frequent Itemset Mining[1]

Many researchers used the frequent patterns to identify the behaviors from a list of application such as Huiyi *et al.*[45] analyzed the behavior of investors in stock market.

[1] http://www.dataminingarticles.com/data-mining-introduction/frequent-pattern-mining/

Real securities clearing dataset is used to mined frequent behaviors. The identified behavior results indicated that investors do not diversify their capital and investors avert risks. Renáta *et al.*[46] discovered hidden information from Web log data to obtain information about the navigational behavior of the users. The identified patterns are according to page sets, page sequences and page graphs. Behrooz *et al.*[47] discovered hidden correlations and identified usage patterns of users moving from demographics attributes in space and time. Their proposed approach is based on abstraction and refinement that use hand-crafted taxonomies on time. Vignesh *et al.*[48] worked on mining and prediction of mobile user movements and purchase transactions in the context of mobile commerce. The proposed framework consisted of three major components such as similarity inference model, personal mobile commerce pattern mine and mobile commerce behavior predictor. Li-Fu *et al*.[49] combined the request traversal paths and frequent purchase items to identify the frequent purchase behavior pattern.

2.2.2 Behavior Sequence Pattern Mining

Sequential pattern mining is a process to identify statistically relevant patterns between data examples where the values are delivered in a sequence [50]. A database of sequences stores a number of records with some ordered events, with or without concrete notions of time. A typical example of sequence database is collection of store items a customer purchased every week for one month. Thus, records in a sequence database can be of different lengths, and each event in a sequence can have one or more items in its set. A sequential pattern-mining algorithm mines the database for repeating patterns which can be later used by domain experts to find association between different instances of data or events data for purposes such as marketing campaigns, business reorganization, prediction and planning.

Valerie *et al*.[51] proposed sequential patterns learning to model the human behavior in order to enhance the independent lifestyle assistant. The work mainly focused on

enhancements of pattern learner. The identified behavior patterns depicted the life of elderly and helped the independent lifestyle assistant to configure automatically. Grazia *et al*.[52] observed that data coming from Multi-Agent System is inherently sequential where gents must be able to quickly adapt to the environment and infer knowledge from other agents' deportment. In order to identify the meaningful features as frequent patterns among the relational sequences, they devised a relational learning algorithm. They applied the similarity measure for relational sequences to adversary classification. Srikantaiah *et al*.[53] proposed a bidirectional growth based directed acyclic graph to understanding users' behavior in order to improve the quality of service offered by the World Wide Web. They used the periodicity and threshold of cyclic behavior that can be utilized to effectively prefetch Web pages. Date and time constrains are specified so their proposed approach can return more desirable, useful patterns. Andreas *et al*.[54] proposed that in multi agent environment If intentions of other agents or events about future can be recognized then agents can adapt the behavior to the situation for better performance. Their proposed approach is based on approach unsupervised symbolic learning in order to create frequent patterns in dynamic scenes.

2.2.3 Behavior Periodic Pattern Mining

In real life patterns are usually and it is essential to understand the complicated occurrence of periodic patterns that may have partial time span, multiple interleaving periods and spatiotemporal noises. The periodicity is usually defined as repeated occurrences of a list of items in a certain order with some frequency. Periodic patterns commonly appear in all kinds of time-series databases. For instance, trajectories of objects, weather, tides, stock market prices, and DNA sequences [55]. Due to the natural changes of in usage data of different application domains, some pattern may only be notable within a portion of the entire data sequence. The different periodic patterns may present at different places and for different durations. Zhenhui *et al*.[56]

addressed the problem of mining periodic behaviors for moving objects. They worked on two problems “how to detect the periods in complex movement” and “how to mine periodic movement behaviors”. They we proposed a two stage algorithm with a notion of reference spot and probabilistic model. Their findings identified periodic behaviors are statistically generalized from partial movement sequences through hierarchical clustering. Mayank *et al*.[57] studied social interactions that occur regularly typically correspond to significant. They proposed measure for ranking mined periodic subgraph close to being perfectly periodic, rather than by the number of its occurrences.

2.3 Behavior Prediction

Machine learning is the scientific discipline concerned with the development and design of algorithms that identify relations between data. Human behavior is recorded as a series of historical data with attributes. Then derive general principle of action which is called as process. However, human behavior is so complicated that either process or algorithm fails to capture its model alone. Thus, improved algorithm based on process is necessary for special situations. After a long period of accumulation, human beings could produce a series of sequential activities for certain tasks when they were studying, working and living, which is called Process. It turns out that human beings make decisions in certain activities getting accustomed to certain conditions or circumstances. Therefore, defining human being’s behavior as processes will contribute to user behavior analysis and prediction.

Machine Learning (ML)[8] is the scientific discipline that studies how computer simulates or implements human’s behavior, in order to acquire new knowledge or skills and improve its performance by reorganizing the existing knowledge structure (model).

Now we state other techniques which has used the above described methods and tools along with some extensions and variations for analyzing the human behavior. Nugent *et al.* [58] analyzed the user’s interaction with technology and environment in order to

provide useful information relating to lifestyle trends and how the environment can be adapted to improve the user's experience. They proposed homeML, an XML based cross-system standard, to support information exchange between intra- and inter-institutional levels. Their proposed XML-based schema improved the accessibility and analysis of the collected data for meaningful analysis of person's life within smart home environments. Rashidi *et al.* [59] applied data mining techniques to solve the problem of sensor selection for activity recognition along with classifier selection in smart homes. They examined the issue of selecting and placing sensors effectively in order to maximize activity recognition accuracy. Chikhaoui *et al.* [60] applied sequential pattern mining for person identification in a multiuser environment. Their proposed approach is utilized for audiovisual and image files collected from heterogeneous sensors in smart homes. Fusion techniques play an important role to achieve high accuracy as compared to single classifiers and successfully produced more accurate results in different application domains such as image processing [61], and gene functional classification [62]. In the context of activity recognition, Xin *et al.* [63] addressed the fusion process of contextual information derived from the sensor data. They analyzed the Dempster-Shafer theory and merged with a weighted sum to recognize the activities of daily living. Rongwu *et al.* [64] proposed classifier fusion as a learning paradigm where many classifiers are jointly used to solve the prediction problem. They used seven wearable sensors including five accelerometers and two hydrophones. Their used classifiers are Linear Discriminant Classifier (LDC), Quadratic Discriminant Classifier, k-Nearest Neighbor (k-NN) and Classification and Regression Trees (CART).

So far, most of the applications where a learning process is involved have treated it as an action to map the overall situation instead of relating the actions among themselves. They process independent pieces of information instead of complete and comprehensive representation of user behavior. However, some of the research groups started to create methods to relate user actions. Fernández *et al.* [65] applied the workflow mining technique to infer human behaviors. Their approach involved an expert user who

can identify the changes in behavior of dementia patients. They validated their approach on synthetic data to identify the deviation from normal behavior. Aztiria *et al.* [66] focused on automatic discovery of user behavior as a sequence of actions. Their developed approach is based on discovery of frequent sets, identification of topology and temporal relations of performed activities with other constraints. Doctor *et al.* [67] focused on developing an application based on set of fuzzy rules to represent the users" patterns. They recorded changes caused by users in the smart environment and generated the membership functions that mapped the data into fuzzy rules. A survey of all these works can be found in [68,69]. The focus of all the above mentioned research is to discover the behavior patterns; however a step towards predicting the future actions from a set of performed activities is still need to be explored for better analysis of human lifestyle and intended services.

Our objective is to overcome the limitation of existing methods by introducing a unified framework for behavior analysis of inhabitants that ranges from activity log to action prediction in order to support the smart home inhabitants in performing their daily tasks and providing personalized services adapted to their needs.

2.4 Analysis of Existing Methods

In the existing work, the information propagation is based on single type of information. Most of the approaches spread 'information' as a single unit with 'active' or 'inactive' status in order to group the individual in to two categories. However, in a real social network people don't lie between two status of either 'active' or 'inactive' to show their significance in the network. Each piece of information has separate spread process according to its type, associated constraints, and importance. So, more granularity of individual importance is required to find the differences between them that can reflect their information propagation capability in the network. The situation becomes more complex for single objective diffusion models when individuals propagate multiple types of information. Therefore, in this study, first the diffusion process is formulated as a

multi-objective optimization problem to model the information spread closer to a real social network. The key difference from earlier studies is that multiple-objectives are achieve in terms of diverse information spread and calculation method to measure the propagation capabilities of individuals in group activities.

On the other hand, identification of significant user behavior by means of a pattern mining has been done with different application domains such as multi-gent system, web usage and market analysis. However identification of user sequential behavior patterns from daily life activities is highly desirable and is not yet available. The prediction about future actions allows caregivers to take proactive actions for the wellbeing of inhabitants after analyzing their healthy or unhealthy routines. However, there is a shortage of formal, systematic and unified behavior modeling and analysis methodologies based on daily life activities. So far, most of the existing applications relate an action to the set of sensor values instead of relating the actions among themselves.

2.5 Applications of Behavior Analysis

In this section, we discuss the application point of view to highlight some important application domains that rely on behavior analysis. We discuss lifestyle analysis, fraud detection, social computing, intrusion detection, group decision making, and event analysis areas. In these applications, we presented that the analysis of behavior and prediction about future actions can play a significance role to provide useful services.

2.5.1 Lifestyle Analysis

Lifestyle identifies the inner behavior of human being by which they make sense of their hobbies, opinions, likes and dislikes. The collection of these habits in a lifestyle defines the influencing factors and inherited traits of different individuals. Most of the aspects of lifestyles cannot be captured with measuring scale. However these factors have a

direct impact in the decision-making of day to day life [70-73]. Let's consider the example of changing consumer behavior with their changing lifestyles. The factors that are involved in a lifestyle of consumer affect the adaption of a new product, loyalty towards brand and the motivation behind the purchase. For launching a new product during a business campaign lifestyle data takes the market analysis a step further. This data recognizes the following factors

- The way people live
- What is their geography
- What influence other friends have on them
- What is the income and occupation of average individual

Analysis of lifestyle enables researchers to see insides people's interests, opinions, and activities. All these factors effect on buying behavior of individuals. Lifestyle affects individuals from a broad perspective that may vary from adoption or usage of a new product or the attachment for the incoming event like new year arrival, valentine's day, or mother's day.

2.5.2 Fraud Detection

Fraud means illegal use of credit cards, tax returns, cell phones, electronic passwords that is a significant problem for business and governments [73-75]. The detection of fraud, security and safety measures to prevent from fraud is not simple task. It grows like an adaptive crime so intelligent means and methods are required to detect and prevent any unwanted activity. These methods exist in the areas of Knowledge Discovery in Databases (KDD), data mining, machine learning and statistics. All these methods have been successfully used in different areas of fraud crimes.

With the increase of white collar crimes, fraud is correlated to social, ethical and legal norms. There is a widespread belief that richer get richer; so many people come to believe that to pass up any opportunity either legal or illegal is like to miss the boat. The

pressure of getting better and better from social perspective compel people to avail the unlawful methods if they are available and in the access with ease.

Even the big companies do not grow with the exponential curve in every quarter. To compensate the pressure and to avoid the negative corporate value, downsizing of employees honest executives turn to the dark side. So in order to avoid such situation companies build an analytics team to predict the pre-emptive strike without disrupting the customers' loyalty and organizations operations. The following steps can play a very effective role to deal with fraud within company environment.

- **Proactive testing:** Regular testing of transactions and daily critical dealings
- **Proper authorization processes:** Checking of rights of employees according to their role in organization
- **Segregation of duties:** Set the duties of employees and give rights according to those duties instead of giving rights to do multiple tasks beyond the scope of job description.
- **Adequate record keeping:** Maintenance of detail document for every sensitive decision and action
- **Physical controls over assets and records:** Maintaining a register or physical evidence of decisions increases the risk to prevent the fraud instead if all records keep electronically.

2.5.3 Social Computing

The online interaction of individuals generates a massive amount of data that can be analyzed to infer the personalities of individuals who produced them. It is very easy and attractive for people to interact with other individuals in virtual world to exchange ideas and feelings. In result the data generated during these interactions can be used to study the lifestyle and behavior of individuals. In parallel to understand the people from their physical space and belongings, we are now able to know users by studying their online

activities and behavior form their electronic interactions [76-78]. This method of lifestyle analysis can replace the traditional method of data collection. This is a big hassle to collect data intentionally and to avoid the biasness of human awareness during collection duration. On the other hand the data collected from online sources seems to be gold mine for scientist due to its diversity and easy access. This data can be processed easily by applying different intelligent and machine learning algorithms by keeping the semantics of the data. Furthermore, the major sources to analyze the social life of individuals are stated as follows

- Social Networks like Facebook, twitter
- Semantic Web
- Mobile Social
- Social Media Sources

2.5.4 Intrusion Detection

With the increase of computer use in everyday life, the reliable need of computer security is growing with each passing day. Researchers are working on most reliable and user friendly methods for access the computer system and network. The most widely used methods are fingerprint and iris scans. Human computer interaction identifies that how individuals interact with computational devices [79-81]. The interaction pattern of each individual is unique that depicts the behavior of individual and his intentions towards the usage of machine. So the study of these diverse patterns of interactions helps the researchers to devise very strong on-intrusive authentication mechanism. Mainly following two paradigms dominate the research.

- **Network based systems:** Designed for network traffic to determine anomalies within acceptable boundaries.
- **Host based systems**: Designed for individual systems being monitored to determine whether the activity on the system is acceptable.

2.5.5 Group Decision Making

The integration of computer in communication from last two decades has revolutionized the method of communication and its effects. So face-to-face communication methods are no longer the only method used by small groups within organization to discuss the problems. These days, the method of electronic messaging and computer conferencing are more acceptable than using the physical methods of sending or receiving postal services. The convergence of the world into global village encourage the use the virtual methods for group decision by considering the diverse aspects of different people from all around the world [82-84]. This results in increase the efficiency and affectability of decision and reducing the cost of arranging group meetings along with necessary arrangements. It is very easy to exchange documents, databases, and messages very easily. So it's very easy among people to do collaboration and to start the joint work for mutual interests which not only beneficial for the companies, individuals but also for nations. Understanding human biases, however, can help improve the quality of decisions.

2.5.6 Event Analysis

The range of applications for behavioral event data is increasing with the increase the linkage between different activities of individuals from their social interaction to mobile usage. So a list of fields get benefits from the results of event analysis which reveals the underlying reasons of particular events [85][86]. These fields include anthropology, psychology, sociology, and political science. For example, US Defense department applied event analysis to identify the epidemics in health and to reveal the reason of terrorist activities. Social event behavior naturally has two interpretations:

- The interpretation related to time and place of event that are used to record for event occurrence to analyze the behavior behind that particular event.

- A graphical representation of each individual as a node either in combination of location or event-location with the behavior being encoded at the node as labels.

2.6 Summary

This chapter explained the related the work in order to state the difference between exiting work in comparison to the proposed methodology. For group activities, exiting work is based on the single information propagation whereas in the proposed method we spread multiple types of information in single diffusion process. To achieve this goal we proposed the MultiObjective Diffusion Model (MODM) that processed the multiple types of information based on multi-evaluation criteria. Furthermore, the existing work based on individual activities treated the daily life activities within smart environment. Most of the work has been done to analyze the human behavior in context to his interaction with physical objects of smart environments on daily basis. However, the proposed method overcomes the dependence of smart environment while analyzing the daily life activities in addition to predict the future action from past significant behaviors. First we apply the sequential pattern mining algorithm to identify the significant behaviors and then these behaviors served as features for Conditional Random Fields (CRF) to predict the future activities.

Chapter 3

Group Activity based Behavior Analysis and Prediction

This chapter presents group activity-based behavior analysis to propagate multiple types of information according to multi-evaluation criteria. Multiple information propagation in one diffusion process is a challenge due to the handling of diverse information based on their natures and preferences. Our proposed MultiObjective Diffusion Model (MODM) is able to propagate multi-information according to their respective criteria.

3.1 Multi-Objective Diffusion for Social Network

A social network is a group of people and can be illustrated as a graphical representation of interactions between a set of vertices. Some famous social networks include online social networks, where vertices are user accounts and edges represent friendships among accounts. Similarly, in communications networks, vertices represent e-mail addresses or telephone numbers, and edges represent e-mails sent or telephone calls with the time of interaction. A typical social network tends to expand over time, with newly added nodes and edges being incorporated into the existing graph with time intervals.

Definition 1: A dynamic social network $G = (g_1, g_2, ..., g_T)$, is a directed multi-graph, where $g_i = (V_i, E_i)$ represent the bag of vertices V_i and edges E_i, at a particular time interval $t_i \in T$. A node $v \in V_i$ shows an individual and an edge $(u, v) \in E_i$ represents an interaction between two individuals during their communication.

The information propagation in any social network depends on the type of the diffusion model. A diffusion model accepts as input a graph structure and state of every individual

at a time '*t*'. It returns a new state of the individual on time '*t+1'* according to its interaction with other individuals. The process continues until all the interactions between individuals are exhausted. The conventional diffusion models can be roughly divided into two categories (1) an independent cascade model and (2) a linear threshold model. In both models, the diffusion process can be regarded as a single-objective optimization problem (*Ω*, *D*) as described below.

Definition 2: Single-objective diffusion model determines a set of individuals *I** for which

$$\boldsymbol{D(I^*) = max_{I \in \Omega} D(I)} \qquad (\mathbf{3.1})$$

where *Ω* is the a unit of information propagated among all individuals *I* in the network *G*, and *D* is assumed to be the objective function for optimization.

The single objective diffusion models have been widely applied as most conventional diffusion processes are based on this single-objective optimization problem. However, they have some disadvantages such as, (1) the single objective diffusion models attempt to solve the problem of diffusion in unitary format to fulfill a single criterion and thus optimize a network on one direction. (2) The diffusion process based on a single objective may fail to maintain the monotonicity property of information during individuals' interaction. (3) Many single-objective algorithms require some priori information about the influence of vertices in the form of edge weights in the network; this influential information is mostly missing in real world networks. (4) A single-objective optimization cannot optimize the multiple types of information on one evaluation criteria (5) A single diffusion model returned by single-objective algorithms may not be suitable for networks with multiple potential diffusion measures. The difficulty in selecting an appropriate criterion in single-objective diffusion model can be handled using a more natural approach that considers the diffusion process as a multi-objective optimization problem which can be defined as follows.

Definition 3: Multi-objective diffusion model determines a set of individuals *I** for which

$$D(I^{*}) = max_{I \in \Omega} (D_1(I), D_2(I), \dots, D_n(I)) \quad \textbf{(3.2)}$$

where Ω is a set of multiple types of information(e.g. news, gossips, rumors, and reports) and '*m'* is the number of objective functions for evolution fitness criteria. In the above equation D_i represents the i^{th} objective function of multi-information. Compared to the single-objective diffusion process, the multi-objective diffusion process has the following advantages.

- The optimal solution obtained by the multi-objective diffusion process defined by (Ω, D_1, ..., D_m) always contains the optimal solutions of the single-objective diffusion process defined by (Ω, D_1), ..., (Ω, D_m).
- The multi-objective diffusion process can always find a set of individuals, that are as good as or better than those of the single-objective diffusion process. Most specifically, in some situations where the best solution corresponds to a tradeoff between different objectives, only the multi-objective diffusion process is able to find it.
- The multiple objectives can measure characteristics of a diffusion process from various perspectives, and thus avoid the risk that a single-objective may only be suitable to a certain kind of information spread. Moreover, the multi-objective optimization process achieves a balance of the multiple objectives and can effectively avoid being trapped in a single dimension of the diffusion process.
- The number of individuals with high information value balanced by multi-objectives is able to keep the diffusion process dynamic, avoiding giving importance to a single information value of a single evaluation criterion.
- The multi-objective diffusion model reveals network analysis from different angles, which help to discover complex and comprehensive information diffusion trends in social networks.

3.2 The Proposed Multi-Objective Diffusion Model

For the multi-objective diffusion problem, both mathematical programming and heuristic approaches can be applied to solve it. The designed Genetic Algorithm (GA) is a type of heuristic approach to solve the problem of multi-objective diffusion. Compared to mathematical programming techniques GA has many advantages [87], such as simultaneously generating a set of candidate solutions and easily dealing with a discontinuous and concave solution [87]. Conventional evolutionary multi-objective algorithm is designed for numerical optimization problems. When it is applied to the multi-objective diffusion, many components of GA need to be redesigned. This is not a trivial task, because the design of these components directly determines the desired output in terms of algorithm performance.

Concretely, the multi-objective diffusion process with GA faces following challenges: (1) Modeling of multiple types of information, it is a set of information with different adoption and diffusion criteria, and (2) selection of optimization evaluation criteria. The fitness functions should reflect the semantic characteristics of individuals from different aspects, such as score, influence and diversity. Ideal fitness functions have better contain intrinsic conflicts, such that the optimal set of individuals can be obtained through the trade-off of multiple evaluation criteria. (3) Effective genetic representation, the genetic representation should be delicately designed according to the characteristics of the diffusion process, since it determines the algorithm performance and scalability to a large extent. Now it is shown that how the multi-objective GA with binary string chromosomes and one-point crossover can be used as the framework for a multi-objective diffusion model. For this purpose, our proposed architecture is represented in Figure 3.1. It has six major modules (a) multiple types of information (b) schema generation, (c) population initialization, (d) information history log, (e) evolution fitness criteria and (f) GA stochastic operations. The detail of each

component is described in subsequent sections, and the pseudo code of MODM is shown in Algorithm 1.

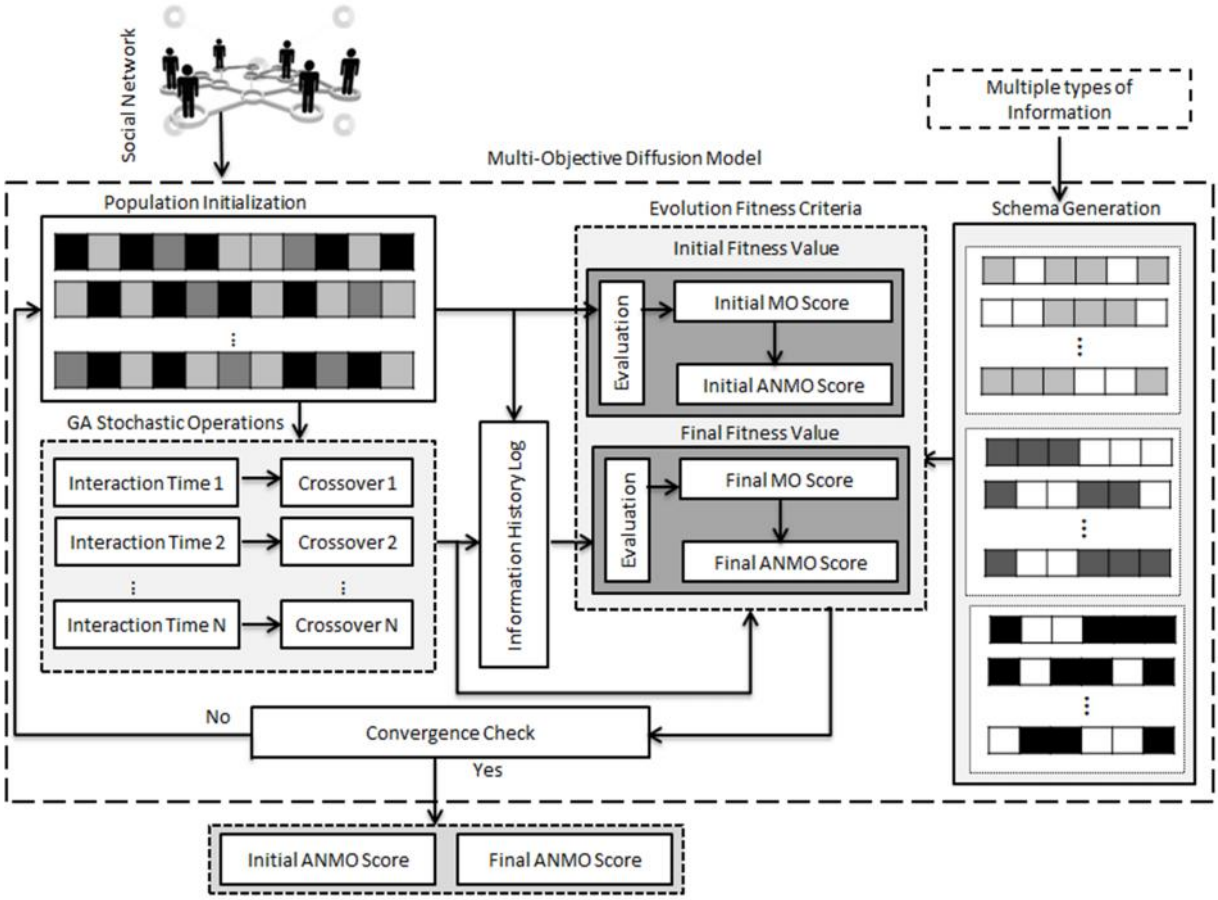

Figure 3.1 The Proposed Architecture of MODM

3.2.1 Multiple types of information

Social networks empower and inspire multiple types of information with separate diffusion processes based on their characteristics. In MODM three kinds of information is propagated: (a) independent information, (b) mutually exclusive information, and (c) competing information. The difference between each information type is as follows.

3.2.1.1 Independent information

Independent information spread autonomously without any constraint, and an individual can hold many independent pieces of information. For example, information about different news is independent and can diffuse independently in the network without any spreading constraints. In the health care domain, information about diseases like diabetes, heart attack, stroke, and blood pressure is independent of each other. In the cellular market, information about cell phones such as Android phones, Apple phones, and Windows smart phones is also independent.

3.2.1.2 Mutually exclusive information

Unlike independent information an individual can hold only one piece of information from a set of mutually exclusive information. Upon selecting a piece of information from mutually exclusive set, he automatically denies other pieces of information from the rest of the set. For example, two music concerts are going to be held at the same time (i.e., mutually exclusive), and an individual can choose only one to attend. Upon selecting one concert, he denies the other. In the healthcare domain a patient can choose between surgery or laser therapy (mutually exclusive) to cure his illness. In the cellular market a user can choose between postpaid or prepaid connection with his contact number.

3.2.1.3 Competing information

Similar to mutually exclusive information, an individual can hold only one piece of information from a list of competing information. However, competing information can be updated with certain constraints. For example, two music concerts are going to be held at the same time, and an individual can attend only one. One concert is popular among people, so when an individual knows both pieces of information he would update his choice to the most popular one and inform others about it. In the healthcare domain, usage of antibiotic is updated according to their ranking in different seasons,

side effects and environments. In the cellular market a user can update his data plan according to recent cost effective and usability offers.

Algorithm 1: Multi-objective Diffusion Model

Input: *sng* – social network graph
mt – maximum trials
ct – convergence threshold

Output: I_{ANMOS} – Initial average normalized multi-objective score
F_{ANMOS} – Final average normalized multi-objective score

```
Begin
g = load(sng)
nv = getUniqNodes(g)
edg= getUniqEdges (g)
ts =  getUniqTStamps(g)
itr=1
While(ct||isequal(mt,itr))
    schNum = rand( )
    sch = SchemaGen(schNum)
    for j = 1: size(nv)
        ini_node =  rand( )
        I_pop= concatenate ( ini_node ,multiObjscore(ini_node, sch ))
    end
    I_ANMOS  = fanmsCal( I_pop, I_ANMOS, itr)
    for 1: size(ts)
        F_pop = fGAOperation( I_pop , sch, edg)
    end
    F_ANMOS =  fanmsCal ( F_Pop , F_ANMOS, itr)
    itr= itr+1
end
End
```

3.2.2 Schema Generation

Schema is generated using well known Holland's hyperplane defined functions (HDFs) [88]. The characteristics of HDF are shown below.

- Generated from elementary building blocks,

- Nonlinear, no separable, and nonsymmetrical (and, so, resistant to hill climbing), Scalable in difficulty, and in a canonical form.
- Difficult to reverse engineer
- Include all finite functions in the limit of encoding information

Elementary Schema

S1	#	#	#	#	#	#	0	0	1	#	#	#	#	#	#	#	#	#	#
S2	#	#	#	#	#	#	#	#	#	0	0	1	#	#	#	#	#	#	#
S3	#	#	#	#	#	1	1	0	#	#	#	#	#	#	#	#	#	#	#
S4	#	#	#	#	#	#	#	#	#	#	1	0	1	#	#	#	#	#	#

Forward Refinement

Fs1	#	#	#	#	#	#	0	0	1	1	#	#	#	#	#	#	#	#	#
Fs2	#	#	#	#	#	#	#	#	#	0	0	1	#	#	1	#	#	#	#
Fs3	#	#	#	#	#	1	1	0	#	#	#	0	0	#	#	#	#	#	#
Fs4	#	#	#	#	#	#	#	#	#	#	1	0	1	#	#	0	0	#	#

Backward Refinement

Bs1	#	#	#	#	#	0	0	0	1	#	#	#	#	#	#	#	#	#	#
Bs2	#	#	#	#	#	#	#	1	#	0	0	1	#	#	#	#	#	#	#
Bs3	#	#	#	0	#	1	1	0	#	#	#	#	#	#	#	#	#	#	#
Bs4	#	#	#	#	#	#	#	1	#	#	1	0	1	#	#	#	#	#	#

Figure **3.2** Elementary Schemas with Forward and Backward Refinement

Representative elementary schemas with their forward and backward refinements are shown in Figure 3.2 where '#' represents don't care bits. The reasons to set the above requirement for HDF during schema generation are as follows:

- Long schema without the combination of small schema is difficult to find. During the corresponding operation if any of the bits is missed then that schema is considered to be lost during the information exchange step.
- Schema with gradual refinements makes the information growth more close to real world. One small schema can grow with the combinations of several other schemas in order to extend the information.

The worth of schema increased as its occurrences include in other positive scored schema.

Each schema is a set of binary values that gives flexibility in terms of modeling multiple types of information with variations in bits of the schema, where '*' represents don't care terms that are either zero or one. Schema generation begins from simple binary strings and becomes more complex to second and third levels by combing previous levels. Each schema string has a start position, length, encoding and score which one schema unique from others. The sample schema of each information type is shown in Table 3.1.

Table 3.1 HDF Based Schema Generation

Information type	Start position	Length	Encoding	Score
Independent	4	12	**11*0001*11	0.23
	25	9	11**001*0	0.15
Mutually exclusive	17	7	1*00*01	0.14
	17	7	1*10*01	0.14
Competing	15	10	0111**1100	0.26
	15	10	0110**0100	0.36

The encoding of independent information has no constraints, while mutually exclusive information is a set of identically scored information with the same start and length positions. However, the encoding for each piece of mutually exclusive information is different, so an individual can hold only one piece of information from whole set. Once an individual receives any mutually exclusive information, he cannot update it. In the case of competing information start and end position are same with different objective score and encoding. From a set of competing information an individual can choose only one piece of information at a time that can be replaced according to its score in later time stamps.

3.2.3 Population Initialization

The selection of the surviving individuals depends on the objective function for a target problem. After each evolution step the fitness of chromosomes is checked by using the objective function to deicide about its worth in the next evolution. Objective function helps to summarize a single merit to measure that how close a given solution is to achieving a desired aim in particular problem domain. There are two main classes for fitness function (a) fix fitness function: fitness function does not change for optimizing a particular solution or testing of a solution with a fixed set of test cases and (b)mutable fitness function: it is used for targeting the niche differentiation or gradually evolving the set of test cases.

For analyzing the individual activities the methods used in this thesis are based on pattern mining and machine learning. Pattern mining helped to identify the significant behavior where machine learning helped to predict the future actions based on the identified patterns learning. In the subsequent section these two approaches are discussed within the scope of the thesis

In order to initialize the population of GA to start the evolution process, encoding of chromosomes is one of the major problems. Encoding depends on the targeted problem so in this section some encodings schemes are introduced, which have been already used with some success. In the proposed GA design, the binary encoding scheme is used.

Binary encoding is the most widely used technique to model a problem because first works about GA used this type of encoding. In this encoding scheme each gene is represented with 0 or 1 to show the associated characteristics of target problem. For example if the problem is knapsack problem then the problem statement is "*there are things with given value and size. The knapsack has given capacity. Select things to maximize the value of things in knapsack, but do not extend knapsack capacity*". So binary encoding of each bit says, if the corresponding thing is in knapsack.

In the proposed GA population, each individual is represented with a binary chromosome of length β to characterize its state during the diffusion process. This can initially choose according to some random distribution. Each chromosome is a set of n pieces of information, where each piece of information is represented by a short binary string that is used to indicate its spread in a network. Depending on the initial state string, each individual in the network knows certain types of information. A vector (x, y, l) is used to describe a piece of information, where *'x'* is its start-point on a chromosome *'y'* is its score in the form of a real value between [0, 1], and *'l'* is the length of the information. If the length of the chromosomes is β, then $x + l \leq \beta$, $x \in [1, \beta]$. The sample chromosome of length β=20 is shown in Figure 3.3 and contains two pieces of information I_1(2,0.25,7), and I_2(12,0.45,8). If a chromosome contains an encoding of information, then the corresponding individual 'carries' the corresponding information. An objective value of a chromosome is the sum of all the scores of information it contains. All the chromosomes have the same length in one diffusion process. The process of initialization is shown in Algorithm 1 from lines 9 to 12.

Locus	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Value	0	0	1	1	1	1	1	1	0	0	0	1	0	0	1	0	1	1	0	1

Figure 3.3 A Representative Chromosome of Length β=20

3.2.4 Information History Log

In addition to the score assigned by a schema to each information type, other important aspects in the calculation of an individual's information worth are the diversity and influence of information. Information diversity is measured by the types of information retained by an individual at the end of the diffusion process whereas; the influence is determined by the frequency of the information generated for each individual during the diffusion process. For example, if someone receives the same information multiple times, the chances of adaptation for that information are high compared to information received fewer times from the same neighbors. Thus, gaining knowledge from all three

resources, the set of highly ranked individuals in a network can be extremely useful from the perspectives of viral marketing, propagating one's point of view, as well as setting which topics dominate the public agenda. To keep track of the history of the information flow from time t_1 to t_n, information history log is maintained to keep record of information identity, type, score, and influence as shown in Table 3.2.

Table 3.2 Representative History Log

Information	Information Type	Score	Influence
**11*0001*11	Independent	0.23	5
11**001*0		0.15	1
1*00*01	Mutually Exclusive	0.14	2
1*10*01		0.14	6
0111**1100	Competing	0.26	1
0110**0100		0.36	4

Algorithm 2: Information History Log

Input: *os* – offspring
sch – schema
Output: *ihl* – information history log
Begin

```
for i=1: size(sch)
    if (os ⊆ sch[i])
        if (os ⊆ ihl )
            ihl.inf= ihl.inf + sch[i].inf
        else
            ihl.type= sch[i]
            ihl.score = sch[i].score
        end
    end
end
```

End

After each information exchange among individuals, the information history log is updated according to the new incoming information and influence of the existing ones

as shown in Algorithm 2. This facilitates the evolution fitness criteria for calculation of an individual information worth based on the spread of numerous pieces of information.

3.2.5 Evolution Fitness Criteria

The fitness criteria guide the search process to quantify the optimality of the diffusion process. Keeping the maximum amount of information as a foundational quality by assigning a relative importance to individual criteria of a fitness function is defined as follows.

$$F(x) = \sum_{i=1}^{n} w_i \, f_i(x) \tag{3.3}$$

where '*x*' an individual, *F(x)* is a combined fitness function, $f_i(x)$ is the i^{th} evaluation criteria, w_i is a constant weight for $f_i(x)$, and '*n*' is the total number of evaluation criteria. In order to combine multiple evaluation criteria into a scalar fitness function a weighted sum approach is defined. Our objective is to maximize all the individual evaluation functions. In the proposed model, the evaluation criteria are score, influence and diversity of information. Score is the HDF generated value for each piece of information. Influence is maintained in the information history log by keeping a record of the number of times a piece of information is received by a particular individual. Diversity measures the total types of information retained by an individual. More specifically, the evolution fitness criterion is defined as:

$$F(x) = argmax\,[w_{scr} fScore(x) + w_{inf} fInfluence(x) + w_{dve} fDiversity(x)] \tag{3.4}$$

where,

$$fScore(x) = \sum_{i=1}^{n} (score.info_i)$$

$$fInfluence(x) = \sum_{i=1}^{n} (influence.info_i)$$

$$fDiversity(x) = Count(info.)$$

$$w_{scr} + w_{inf} + w_{dve} = 1$$

If constant weights in (4) are used, the search direction in genetic algorithms becomes fixed. Therefore, a selection procedure with random weights is proposed to search for optimal solutions by utilizing various search directions for each evaluation criteria. In equation 4, $F(x)$ is the weighted sum of all evaluation criteria that assigned multi-objective score to one individual at time '*t'*.

3.2.6 GA Stochastic Operations

Crossover and mutation are two basic operators of GA. The selection and implementation of these operations depends on the encoding scheme and target problem. These operations play an important role for the performance of GA.

At each time '*t*' that a pair of individuals interacts, they exchange information according to type and category, which is modeled by a crossover operation. A uniform crossover point '*c*' is selected at random from the integer range [1, *θ*].

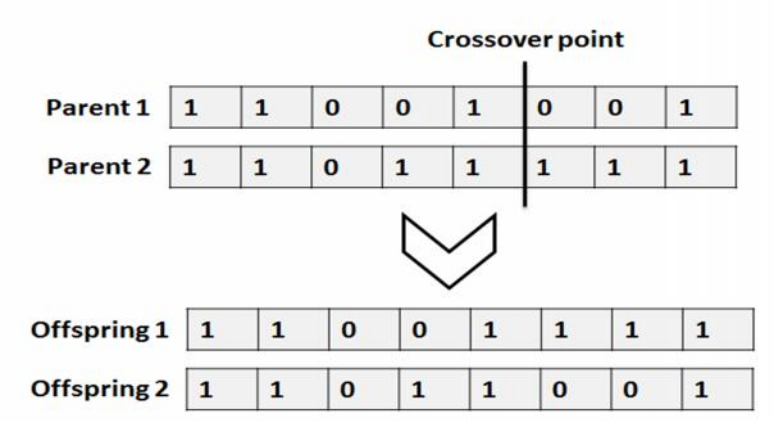

Figure 3.4 Single Point Crossover for Binary Encoding

Two new state strings $Ofsp_1$ and $Ofsp_2$ are created by swapping the tails of interacting individuals, where the tail is defined as all positions including and after index *c*. For example, if the value of *c*=6 is selected randomly as crossover point, the bits around that point (11001|001 and 11011|111 produce 11001111 and 11011001) are exchanged as

Shown in Figure 3.4. The idea here is to swap some alleles and the resultant bit-strings are the offspring to represent the information exchange as shown in Algorithm 3. If the edges in the social network are bidirectional, exchange of individual interaction roles is performed, and the crossover repeats. If there are multiple interactions at the same time for a single individual, one of the middle chromosome generated in each interaction is randomly chosen as a parent for crossover operation. This parent will adopt all other information existing on the other offsprings and bring benefit to it. The multi-objective score of each new offspring is evaluated according to equation 4.

Algorithm 3: *fGAOperation* - GA Stochastic Operations

Input: I_{pop} -population
s*ch* **-** schema
edg - interacting edges
Output: F_{pop} **-** final populations
Begin

```
crosPnt= rand( )
for i=1: size(edg)
    ofsp1 = concatenate (Ipop.edg[i] (crosPnt: end), Ipop.edg[i+1](1: crosPnt))
    ofsp2 = concatenate (Ipop.edg[i+1] (crosPnt: end), Ipop.edg[i] (1: crosPnt))
    if (ofsp1.multiObj(sch)> ofsp2.multiObj(sch))
        newOfsp= ofsp1
    else
        newOfsp= ofsp2
    end
    if (Ipop.(edg[i+1]< newOfsp)
        ofsp = newOfsp
    else
        ofsp = Ipop.edg[i+1]
    end
     Fpop=ofsp
end
```

End

If any of offspring has a higher multi-objective score than their parents, the corresponding parent's state string is replaced in the next iteration. In the case of ties in

the multi-objective scores of the original and an offspring, the original state string is retained as shown in lines 10 to 14 of Algorithm 3.

After the crossover operation, the multi-objective score of each individual is calculated for a particular generation. The high score of an individual could be the result of its network characteristic or based on its randomly assigned initial values. To avoid the latter bias GA is repeated multiple times and the Average Normalized Multi-Objective Score (*ANMO*) is calculated for each individual to show his information worth irrespective to the start of diffusion process. The value of *ANMO* is iteratively calculated till the end of diffusion process. In each interaction the value of *ANMO* for each individual is updated according to his previous *ANMO* and recent multi-objective score in current population. Previous *ANMO* value is normalized with numbers of GA iterations and new objective score is normalized with the maximum score of the network in current population. The complete workflow to calculate the *ANMO* score is shown in Algorithm 4.

Algorithm 4: *fanmsCal*- Average Normalized Multi-objective Score Calculation

Input: *pop* – population
P_{ANMOS} – Previous average normalized multi-objective score
itr – iteration
Output: U_{ANMOS} – Updated average normalized multi-objective score
Begin
1 ***for*** *j = 1: size(pop)*
2 | Max_{score} *= max(pop)*
3 | $Prev_{val}$*=* P_{ANMOS}*[j]/ itr*
4 | cur_{val}*=* P_{ANMOS}*[j]/* Max_{score}
5 | U_{ANMOS}*[j]=* $Prev_{val}$ *+* cur_{val}
6 **end**
End

The stopping criterion for GA is either a fixed number of generations or convergence to a predetermined threshold value. After performing all interactions among all individuals the convergence test guides the MODM to continue or to stop.

3.3 Summary

In summary, a multi-objective diffusion model is proposed that propagates multiple pieces of information with evolution fitness criteria by designing an evolutionary algorithm. In order to propagate multiple types of information in one diffusion process, the set of information is modeled into a binary schema where each schema represents one type of information with its associated score. Furthermore, information history log is maintained for each individual to keep track of all incoming and outgoing information in all time stamps. This helps to predict a more accurate class of information diffusion by holding the monotonicity property about information. The information value of each individual is calculated based on evolution fitness criteria for each information type. Evolution fitness criteria utilize the benefits of score generated by the schema and information history maintained in the information history log. Our experimental results on a real world dataset show that our model is able to simulate the rich class of diffusion model and predict the information flow in the multi-objective environment. Finally, the results show that a few individuals in the network always obtain a high information rank irrespective of the start of the diffusion process.

Chapter 4

Individual Activity based Behavior Analysis and Prediction

This chapter presents the proposed framework for individual activity based behavior analysis and prediction. Activity prediction becomes a challenge due to unavailability of the labeled data. Unlike the conventional methods that are unable to handle complex situations of activity log with high class-accuracy, this model is able to predict the future actions based on past learning.

4.1 The Proposed Framework

In the proposed approach, an activity is defined as set of active sensors at a particular time that perform a certain task in a smart home environment. The proposed framework consists of two major modules, as shown in Figure 4.1: (1) Behavior pattern discovery: to identify the sequential behavior patterns from the activity log. (2)Behavior Prediction: to predict the future actions by utilizing the significant behavior of inhabitants' daily life. The details of each module are described in the following sections.

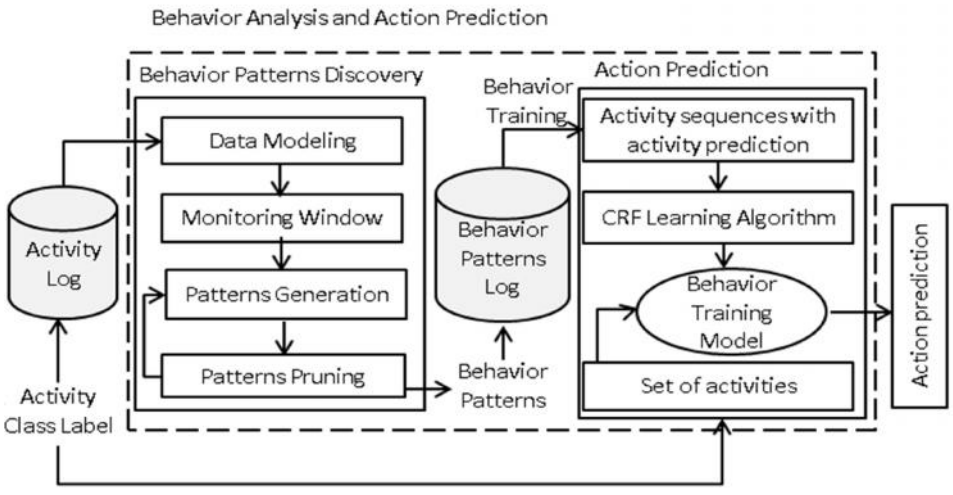

Figure 4.1. The Architecture of the Proposed Framework.

4.1.1 Behavior Patterns Discovery

Representing the inhabitants' actions by means of ordered sequence of activities facilitates our understanding of the significant behavior patterns in daily lifestyles. Sequential pattern mining is a process to identify statistically relevant patterns between data examples where the values are delivered in a sequence. A activity log of sequences stores a number of records with some ordered events, with or without concrete notions of time. A typical example of sequence activity log is collection of store items a person performed every week for one month. Thus, records in a sequence activity log can be of different lengths, and each event in a sequence can have one or more activities in its set. A sequential pattern-mining algorithm mines the activity log for repeating patterns which can be later used by domain experts to find association between different instances of activities for purposes such as healthcare, recommender systems and prediction and planning.

Therefore, the objective of this module is to identify the set of actions that frequently occur together. One intuitive way for behavior pattern generation is to apply a sequential pattern mining technique. For this purpose, a repository of activity log "*Al*" is given where activities are stored in sequential order with respect to activity time. Let $D = \{a_1, a_2, \ldots, a_m\}$ is a set of m activities performed in a particular day in a temporal manner T. Let each sequence in the "*Al*" be $S = \{D_1, D_2, \ldots, D_n\}$, where D_i is a set of performed sequences of activities on different days. For instance a set of sequential activities is defined as an individual who comes to the bedroom to sleep is likely to read or watch TV before the sleep activity. The sample activity log is shown in Table 4.1. In our proposed data modeling, the monitoring window is a list of activities performed in three days ordered by activity time.

Table 4.1. Representative Repository of an Activity Log.

Sequence ID	Days	Activities
	1	Read, Sleep
S1	2	Kitchen, Master Bedroom , Read
	3	Kitchen, Master Bedroom , Watch TV

	4	Read, Sleep, Chores
S2	5	Master Bedroom , Read, Sleep
	6	Kitchen, Master Bedroom, Watch TV, Master Bathroom
	7	Master Bedroom , Read, Watch TV, Kitchen
S3	8	Read, Sleep, Master Bathroom, Sleep
	9	Watch TV, Master Bathroom, Sleep

Here, the problem is to discover all sequential patterns with a specified minimum support, where the support of a pattern is the number of data-sequences that contain the pattern as shown in Equation (8). Therefore, a sequence pattern is a non-empty set of "*Al*" and a day D_i is said to contain pattern *P* if $P \subseteq D_i$:

$$Supp(P) = \frac{the\ number\ of\ instances\ containing\ P\ in}{the\ number\ of\ instances\ in\ Al} \quad (4.1) \qquad (4.1)$$

The pseudocode for the frequent sequential behavior patterns is shown in Algorithm 1 for activity log *Al* and support threshold ∝. Here, S_k is the candidate set for level *k,* genrated by *fGenCanSet*(*Al*) method and *fGenActivitySequence* (S_k) method is assumed to generate the candidate sets *CS* from the large activities of the preceding level. The downward closure *count* (CS) accesses a field of the data structure that represents candidate set *CS*, which is initially assumed to be zero. Therefore, all the activities in an element of a sequential pattern necessarily present in a single day activities for the data-sequence to support the pattern. A pattern is regarded as persistent if it has the highest support.

This demonstrates the most significant behavior of inhabitant due to its high continued or repeated ratio as compared to other identified patterns under same support threshold. The analysis of frequent user behaviors *Bp* reveals the significant habits of inhabitants from their daily routines and provides the basis for behavior learning to predict their future actions.

Algorithm 1: Frequent Sequential Behavior Patterns

Input: *Al*: Activity log
∝: Support threshold
Output: *Bp*: Behavior patterns
Begin
1 S_1 = *fGenCanSet*(*Al*)
2 *k*=2
3 While (S_{k-1} != *Null*)
4 *CS* = *fGenActivitySequence*(S_k)
5 for *j*= 1:*length* (*CS*)
6 if (*Supp*. (*CS* (*j*)) > ∝**)**
7 *Count* (*CS* (*j*)) = *Count* (*CS* (*j*)) +1
8 S_k=*CS*(*j*)
9 *k = k+1*
10 end
11 end
12 *Bp* = *Union*(S_k)
13end
End

4.1.2 Action Prediction

The objective of this module is to predict the next action from the set of activities that occur together. For the learning process of action prediction, required data is extracted from the behavior pattern log. Let"s consider activities shown in Table 4.2 are occurring together in different sets of actions and the same set of activities with their relationships among them is presented in Figure 4.2. It is obvious that the occurrence of each activity depends on the set of previous actions. For example, "Meditate" comes after "Sleep" or "Chores", whereas, "Sleep" comes after "Watch TV" or "Master Bathroom" and there could be the repetitive actions in the same sequence. Therefore, a decision about the next activity depends on the transition of previous actions. For instance, "Kitchen" activity follows by "Meditate" or "Sleep" represents breakfast while "Kitchen" activity after "Enter Home" represents dinner. So it is clear that a set of previous actions provide remarkable evidence to identify the meaningful behavior in terms of forthcoming action. In our proposed approach, sequences of 8 to 10 activities are considered to predict the next action.

Table 4.2. Representative Sequences from Behavioral Patterns.

Sequence	Prev. Activity 3	Prev. Activity 2	Prev. Activity 1	Next Activity
1	Kitchen	MasterBedroom	WatchTV	Sleep
2	—	WatchTV	MasterBathroom	Sleep
3	WatchTV	MasterBathroom	Sleep	Chores
4	—	Sleep	Chores	Meditate
5	MasterBathroom	Sleep	Meditate	Kitchen

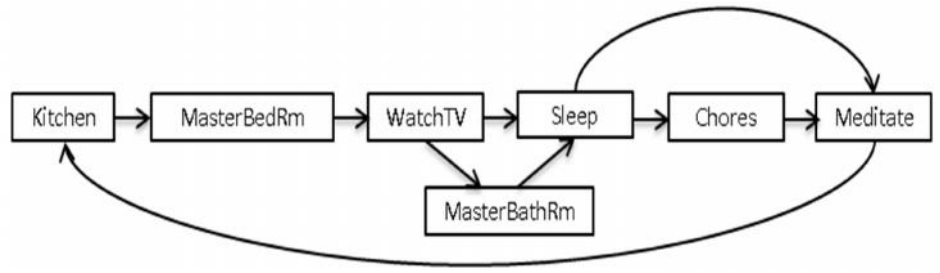

Figure 4.2. Set of Sequences with Activity Relationships.

Once the sequences of activities are selected, then these activities can be used for the learning process of action prediction. In the proposed framework, CRF is designed as a learning classifier for predicting the future actions. It is a discriminative and generative probabilistic model for labeling the sequences under the conditional probability $p(y|x)$. It is modeled as undirected acyclic graph that allows arbitrary, non-independent relationships among the observation sequences. A CRF flexibly capture the relation between a pair of observations and label sequences that do not explicitly model the marginal probability of observations. It uses a potential function instead of a joint probability. Suppose there are finite label sequences $Y = (y_1,y_2,\ldots,y_{T-1},y_T)$ and observations $X = (x_1,x_2,\ldots,x_T)$. In Figure 4.3 a design of CRF is shown for the activity sequences presented in Table 4.2.

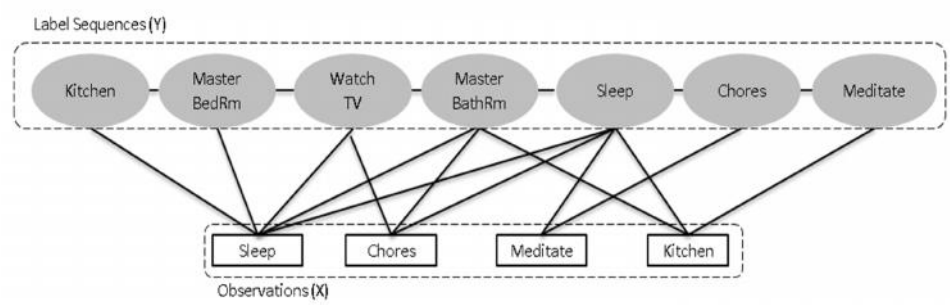

Figure 4.3. The Design of CRF for Activity Sequences.

In addition to CRF, other classification algorithm such as HMM, NN and SVM are used to justify the better performance of CRF as compare to other existing algorithms. . Brief description of each classifier with preferred settings for our experiments is given as the following.

4.1.3 Action Classification Methods

Action classification methods concern the construction and study of systems that can learn from activity log. In this section, the brief introduce the most relevant supervised machine learning classifier (i.e., ANN, HMM, CRF, and SVM) is given according to the scope activity prediction.

Artificial Neural Networks (ANNs): An artificial neural network (ANN) is inspired by the structure and functional aspects of biological neural networks [89]. The group of artificial interconnected neurons uses to compute the complex relationships between input and outputs.

Modern ANNs are non-linear statistical data modeling tools to find the statistical structure in an unknown joint probability distribution between labeled variables. If the example of multilayer neural network with back propagation learning algorithm[8], is taken then the structure of the network, number of hidden layers, and number of neuron in each layer effects the learning of different class labels. The activation of the neurons in the network depends on the activation function [8]. Training of multi-layer neural network through back propagation learning method and weights are updated by the following equation:

$$\Delta w_{ki} = -c\left[-2\sum_{j}\{(y_{j(desired)} - y_{j(actual)})f'(act_j)w_{ij}\}\, f'(act)_i x_k\right] \quad (4.2)$$

Where Δw is the weights adjustment of the network links and c is the learning rate of neural network such that (0 < c <=1). In order to learn class labels more rapidly, the recommended value of c = 0.1. The number of hidden layers and number of neurons in each hidden layer along with activation function affect the performance of the ANNs. The example of tangent sigmod function as an activation function as given below:

$$\varphi(x) = \tanh\left(\frac{x}{2}\right) = \frac{1 - \exp(-x)}{1 + \exp(-x)} \quad (4.3)$$

The total number of epochs for learning of the network depends on the target problem. The multi-layer neural network can be seen as an intuitive representation of a multi class problem. The number of correctly classified class labels depends on the number of training instances during the learning phase.

Hidden Markov Models (HMMs): It is a generative probabilistic graph model that is based on the Markov chains process [90]. HMM gets its name from two defining properties. First, it assume that he observation at a particular time is generated by some process whose state is hidden from the observer. Second, it assumes that the state if this hidden process satisfies the Markov property.

For class label problem, HMM is based on the number of states and their transition weight parameters. The learning of HMM find the best set of state transitions and output probabilities for a given set of sequences. The maximum likelihood estimation is used to learn the parameters. In most of the experimental setting, parameters are learned thorough observation and following parameters are required to train the model:

$$\lambda = \{A, B, \pi\} \quad (4.4)$$

Where λ is graphical model for class labels, A is a transition probability matrix, B represents the output symbol probability matrix, and π is the initial state probability [90]. HMM usually assume that the state transition matrices and output models are not depend on time, in other words the model is time invariant. There exist no tractable algorithms which can learn the maximum likelihood accurately for a given set of sequences. However, a local maximum likelihood can be derived efficiently using the Baum–Welch algorithm or the Baldi–Chauvin algorithm. If Baum-Welch algorithm is considered to determine the states and transition probabilities during training of HMM, then the i^{th} classification of a class label is given as:

$$\lambda_i = \{A_i, B_i, \pi_i\}, \qquad i = 1, \dots, N \tag{4.5}$$

HMM can model complex Markov processes where states can emit the observations according to some probability distribution (e.g., Gaussian). HMM can also be generalized for continuous state spaces and Markov process are treated as linear dynamical system with linear relationship with relevant variables and observed variables follow the Gaussian distribution. In such models of HMM the exact inference is tractable however, in general, exact inference in HMMs with continuous latent variables is infeasible and required an approximate method to infer the outcome variables.

Conditional Random Fields (CRFs): It is a discriminative probabilistic graph model for labeling and segmenting structured data, such as sequences, trees and lattices [91].

In the CRF model, the conditional probabilities of next action with respect to previous activity observations are calculated as follows:

$$p(y_{1:T}|x_{1:T}) = \frac{1}{Z(x_{1:T}, w)} exp\left\{\sum_{j=1} w_j F_j(x_{1:T}, y_{1:T})\right\} \tag{4.6}$$

In Equation (9), Z denotes normalized factor and $F_j(x_{1:T}, Y_{1:T})$ is a feature function that is computed as:

$$F_j(x_{1:T}|y_{1:T}) = \sum_{t=1}^{T} f_j(y_{t-1}, y_t, x_{1:T}, t) \quad (4.7)$$

In Equation (5.3), the feature function depends on known observations $x_{1:T}$ and is determined by any combination of input values instead of considering all arguments. To make the inference in the model, the most likely activity sequence is computed as follows:

$$y^*_{1:T} = \text{argmax}_{y'_{1:T}} p(y'_{1:T}|x_{1:T}, w) \quad (4.8)$$

Hence, the learning capability of CRF in terms of sequences of actions is able to capture long-range transition among activities collected from behavior patterns log for future action prediction.

Support Vector Machines (SVMs): SVM is statistical learning method to classify the data through determination of a set of support vectors and minimization of the average error [92]. It can provide a good generalization performance due to rich theoretical bases and transferring the problem to a high dimensional feature space. The basic SVM process the data by considering it two class problem that makes it a non-probabilistic binary linear classifier. Given a set of training examples, the SVM train it on the basis of two classes and builds a model that assigns a new example into one category or the other. SVM represents the data samples as points in space so that the examples of the different categories are identified by clear gap. New examples are then mapped into that space based on the gap they fall on nearby category.

For a given training set of sequence value and class labels, the binary linear classification problem require the following maximum optimization model using the Lagrangrian multiplier techniques and Kernel functions as:

$$\text{Maximize (w. r. t } \alpha) \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=0}^{n} \sum_{j=1}^{n} \alpha_i y_i \alpha_j y_j K(x_i, x_j) \quad (4.9)$$

$$\text{Subject to: } \sum_{i=1}^{n} \alpha_i y_i = 0, \quad 0 \leq \alpha_i \leq C \quad (4.10)$$

Where K is the kernel function that satisfies $K(x_i, x_j) = \Phi^T(x_i)\Phi(x_j)$. The example of radial basis function (RBF) for recognizing labeled sequences as.

$$K(x_i, x_j) = \exp\left(\frac{-\|x_i - x_j\|^2}{(2\sigma^2)}\right) \quad (4.11)$$

In SVM for multi-class problem, “one-versus-one” method is adopted to classify the different class labels. Classification of the final class label is based on the voting mechanism and maximum vote of a class determined the final class label.

4.2 Summary

A unified framework for activity recognition based behavior analysis and action prediction is proposed. This facilitates the service provider about inhabitants’ significant behavior in order to perform meaningful interventions. In the proposed framework, first, the kernel fusion method is introduced to overcome the learning effects of different kernel function for recognition of individual activities. Furthermore, the recognized activity log is utilized for behavioral pattern discovery with the help of frequent sequential mining technique on a set of activities that are performed in temporal sequence of three days. Finally, CRF is investigated for the actions that occur together in order to predict the next activity from current situation. Our study found that identification of behaviors patterns and prediction of forthcoming action with high precision signifies the possibility of helping people by analyzing the long-term data of one’s behavior to fulfill his needs in the current circumstances and in future.

Chapter 5

Implementation and Results

This chapter provides the details about implementation and results of behavior analysis based on group and individual activities. This helps in identifying the flow of information among different individuals in their group interactions. Similarly behavior analysis of individual activities provides details about their significant sequential behavior patterns and future actions.

5.1 Results and Evaluation for Group Activities

This section will validate the effectiveness of MODM through experiments on a real social network. The goal of the experiments is to estimate the information value of each individual over multiple random state initializations using HDF schema and information history log. This identifies that whether all individuals receive same *ANMO* score as a result of their interaction or their score varies according to their relative position in the network and information processing capabilities.

5.1.1 Dataset Description

MODM is applied on publically available Enron email dataset [93]. The snippet of dataset is shown in Figure 5.1.

```
852139323 laurinh@prodigy.net  m..love@enron.com
852139323 laurinh@prodigy.net  queue@ebay.com
883935960 educanto@msn.com  d..thomas@enron.com
884050320 jan_levine@travisintl.com  breco@ix.netcom.com
884050320 jan_levine@travisintl.com  abreco@ix.netcom.com
884070321 jan_levine@travisintl.com  abreco@ix.netcom.com
894050320 jan_levine@travisintl.com actbars@aol.com
884050320 jan_levine@travisintl.com actbars@aol.com
884050320 jan_levine@travisintl.com apetrello@nabors.com
```

Figure 5.1 Snippet of Enron Email Dataset

It is the large dynamic repository of e-mails of the former Enron Corporation where vertices represents email addresses and directed time stamped edges represent interaction between individuals. It has 84,716 e-mail addresses, 1,326,771 total timestamps, and 215,841 unique timestamps covering a period of approximately four years. Moreover, the out-degree and in-degree of the network are in a power-law distribution as shown in Figure 5.2.

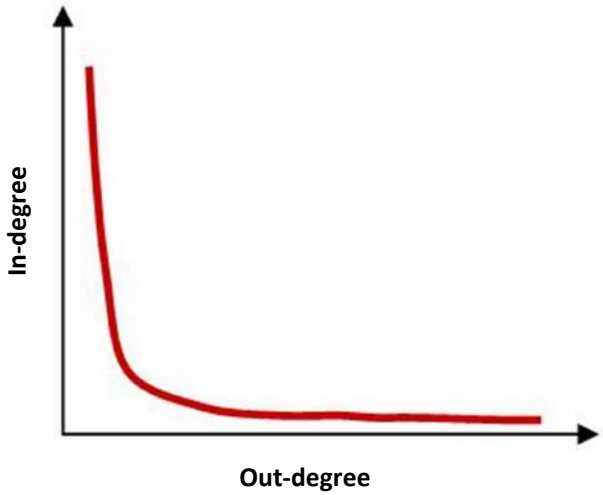

Figure 5.2 Power Law Distribution of In-degree and Out-degree

5.1.2 Experimental Setup

The proposed method has been implemented in MATLAB 7.6. The configuration of the computer is an Intel Pentium(R) Dual-Core 2.5 GHz with 3 GB of memory and Microsoft Window 7. The experiments start by generating the schema for each information type with their scores, and at the start of the diffusion process, each individual is initialized with random initial values in chromosome. The basic steps of implementation are as follows.

- **Step 1:** Start by generating the schema for each information type with their Score
- **Step 2:** Each individual is initialized with random binary values in chromosome

- **Step 3:** Calculate the relative importance of each individual
- **Step 4:** Model the interaction among individuals
- **Step 5:** Calculate the relative importance of each individual and check the convergence of the solution. If solution is not converged then go to step 1 and continue the process.

During the interaction of individuals at each time stamp initial multi-objective score is calculated. At the end of single iteration MODM processes all time stamped edges, the final multi-objective score of each individual is normalized relative to the maximum multi-objective score in the population. Multiple trials of GA are run in order to avoid any biasness caused by the random assignment of initial values to individuals. At the end of diffusion process, the Average Normalized Multi-Objective score (*ANMOS*) of each individual is calculated over many trials to show the relative information worth of whole population.

5.1.3 Results and Discussion

In this section, five experiments are performed to show the usefulness of MODM in comparison with: (a) single information propagation, (b) single evaluation criteria, (c) weighted and neutral weight factor of evaluation criteria, (d) conventional network measures, and (e) an existing approach GADM [42]. In all experiments x-axis represents the *ANMO* score and y-axis represents the *Fn*(*ANMOS*) that is defined as the proportion of individuals having *ANMO* score. The details of the experiments are as follows.

5.1.4 Comparison of MODM and single information propagation

To show the significance of MODM in comparison to single information diffusion process, each information type is separately propagated from a set of multiple types of information. The evaluation criterion during this experiment is based on equation 4, with equivalent weights. In Figure 5.3(a), the initial *ANMO* score of each information type and MODM multiple types of information is illustrated. The similar diffusion curve for initial *ANMO* in all information types show the similar start of

diffusion process in all cases and depend on the inherent position of the network. However, for the final *ANMO* score, the diffusion curves in Figure 5.3(b) show that MODM is better obtaining the maximum diffusion objective as compared to individual information types. The diffusion curve of mutually exclusive information is strongly clustered and shows no dispersion in the *ANMO* score maximization. The diffusion process for competing information finished quickly while independent information is better as compared to the other two information types. However, neither of them can reach to maximum *ANMO* score. The result shows that MODM can be applied to model the information exchange based on a single information type. However, the multi-objective design of the proposed model combined the benefits of multiple types of information propagation in single diffusion process that demonstrate the better information maximization during the diffusion process.

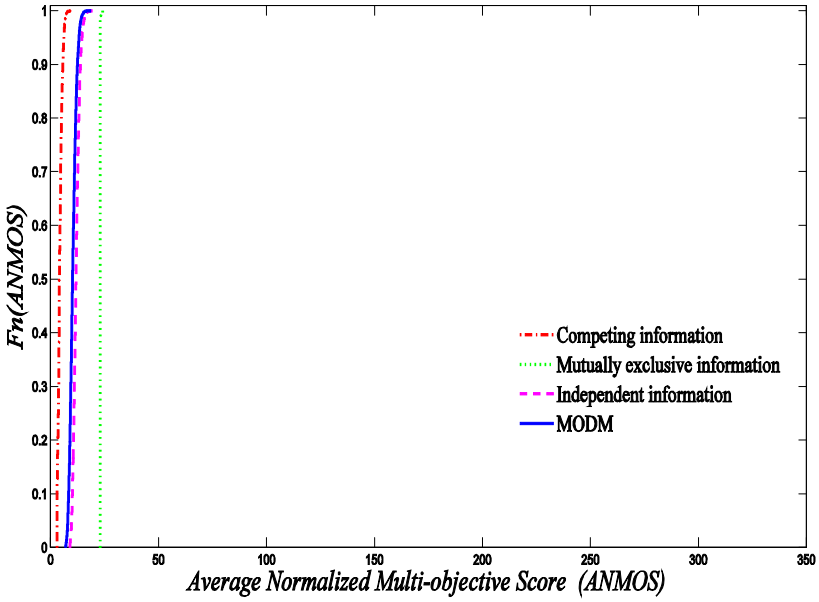

Figure 5.3(a) Initial *ANMO* Score for MODM and Single Information Types

5.1.5 Comparison of MODM and single evaluation criteria

In this experiment, the importance of MODM is estimated in comparison of single evaluation criteria. Multiple types of information are propagated in the network, but calculate their diffusion values based on single evaluation criteria. In Figure 5.4(a),

the initial *ANMO* scores are shown for all evaluation measures; again, the diffusion curves are very similar to each other. This shows that no matter whether the experiment is performed on information propagation or evaluation criteria the initial *ANMO* curves are highly clustered, and there is no information dispersion in the network. In Figure 5.4(b), the diffusion curves in terms of the final *ANMO* score are illustrated for all measures. This signifies the MODM capability to propagate the information on single evaluation criteria in addition to its original goal of achieving the multi-objectives during the diffusion process. The influence measure has very low dispersion; it means that the influence of the information is not changing rapidly during the interaction of individuals. Diversity measure is better than influence, as information history log helps in calculation of the diversity of each individual after each interaction. The diffusion curve for score measure gets clustered after a certain limit, however it shows better performance than the influence and diversity measure. The diffusion curve for MODM represents its similar start with other evaluation criteria; however it achieves highest *ANMO* score at the end of the diffusion process that represents its significance in getting the objective of information maximization with high dispersion.

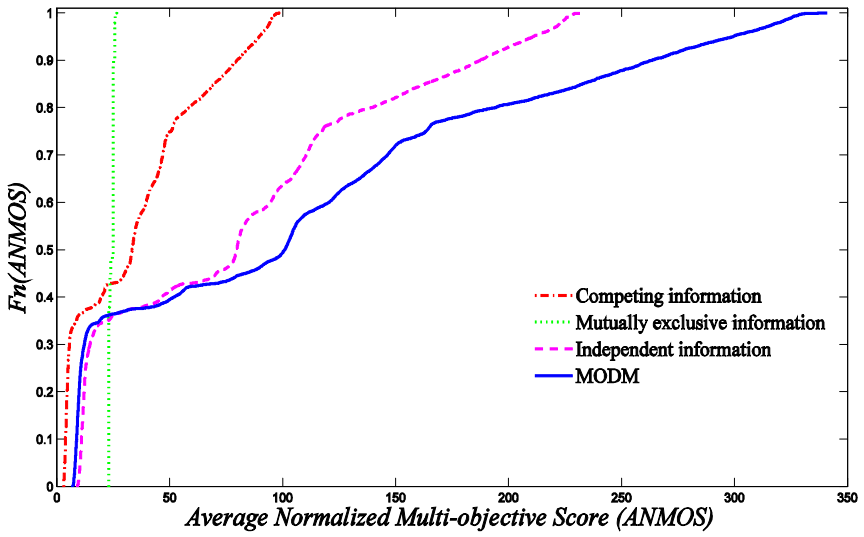

Figure 5.3(b) Final *ANMO* Score for MODM and Single Information Types

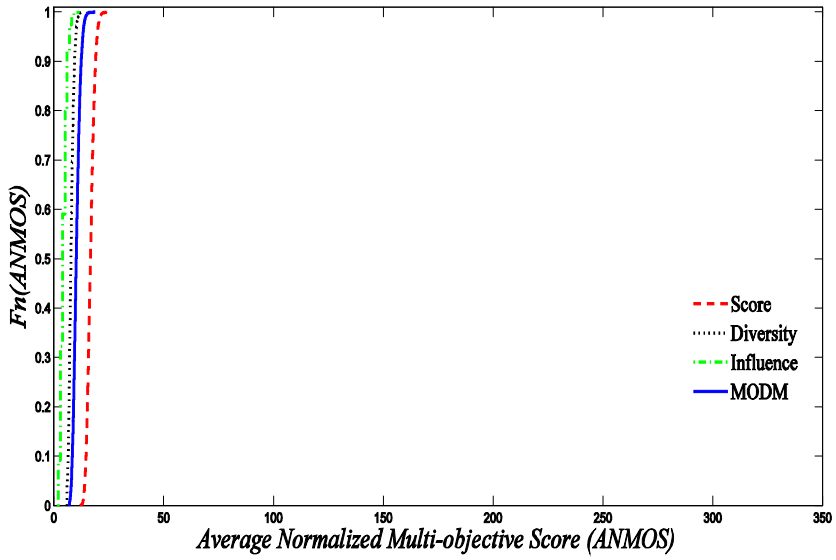

Figure 5.4(a) Initial *ANMO* Score for MODM and Single Evaluation Criteria

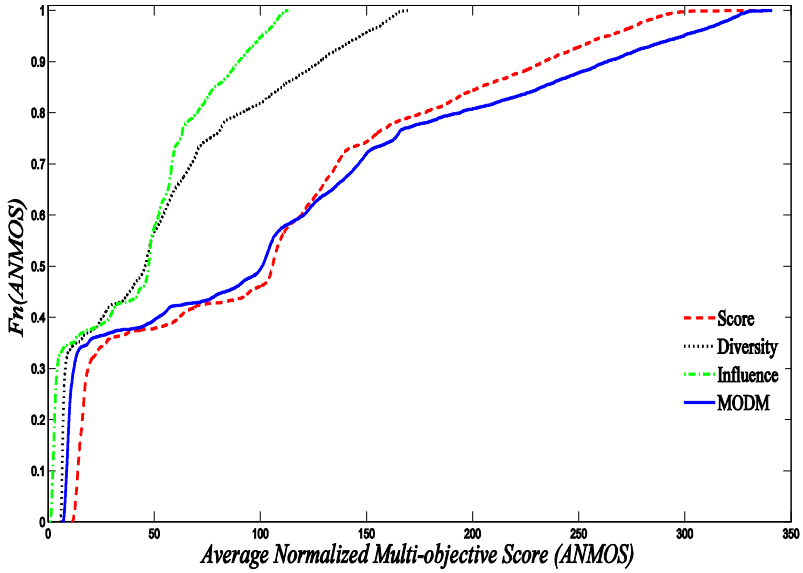

Figure 5.4(b) Final *ANMO* Score for MODM and Single Evaluation Criteria

5.1.6 Comparison of evaluation criteria with weighted and neutral weight factors

This experiment shows the effects of weight factor attached to each evaluation criteria in equation 4 of MODM. In order to give high importance to particular evaluation criteria, its weight factor can be increased at the start of diffusion process. In this experiment first, multiple types of information is propagated by assigning the equivalent weights (a neutral weight factor) to each evaluation criteria and then higher weight of 0.6 is assigned to each evaluation criteria while the remaining weight of 0.4 is equally distributed in others. Diffusion curves for weighted and neutral weight factor are shown in Figure 5.5 (a), (b) and (c) for score, influence and diversity measures respectively. The high diffusion curves for each evaluation criteria illustrate that giving a high weight to a particular evaluation criterion thereby created a high importance in multi-objective optimization. The results of this experiment show that the proposed MODM gives flexibility to users to assign a high importance of any evaluation criteria in order to mold the diffusion process in the intended dimension for the analysis of a underlying social network.

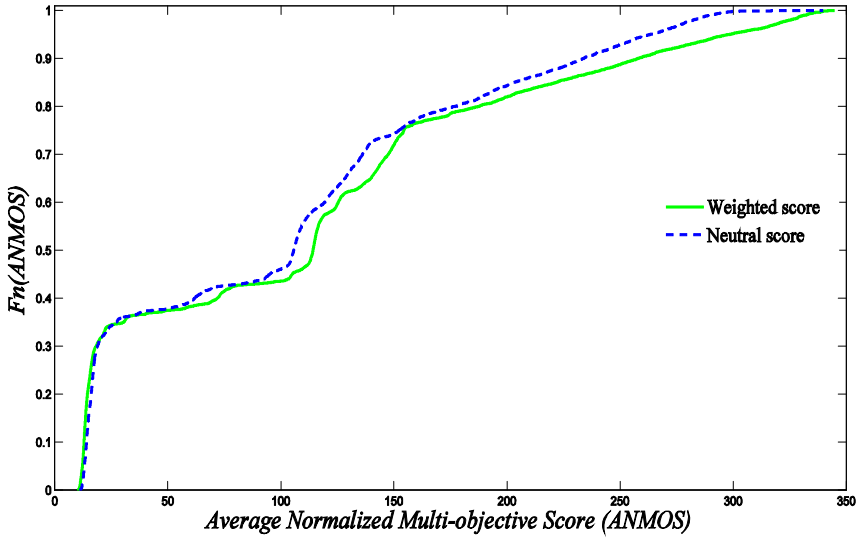

Figure 5.5(a) Score Comparison for Weighted and Neutral Weight Factor

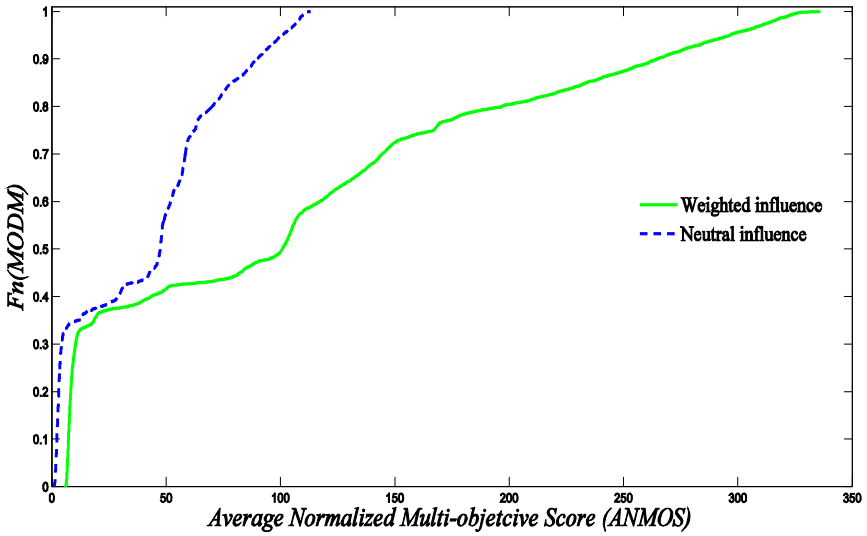

Figure 5.5(b) Influence Comparison for Weighted and Neutral Weight Factor

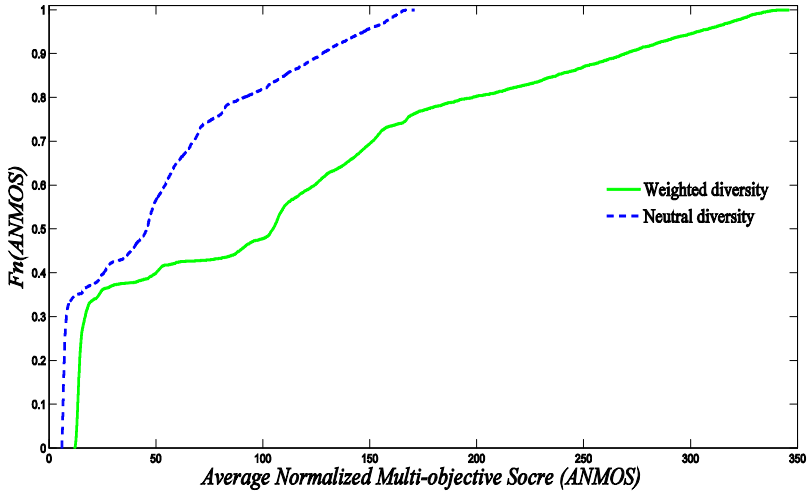

Figure 5.5(c) Diversity Comparison for Weighted and Neutral Weight Factor

5.1.7 Comparison of MODM and conventional network measures

In this experiment, the *ANMO* score of MODM is compared with five conventional network measures such as, (a) betweenness centrality, it is the centrality measure for each vertex of network that quantify the control of an individual on the communication with others in a social network [94]. Individuals, who have a high

probability to occur on a randomly chosen shortest path between two randomly chosen individuals for interaction, have a high betweenness. (b) PageRank, it is link analysis algorithm that assigns a rank to each individual within the social network based on its importance in communication [95]. If an individual has a lot of links with other high degree nodes then he is assigned high ranking. (c) Clustering coefficient, it is a measure of the degree to which vertices in a social network tend to cluster together [96]. It quantifies position of an individual that how close his neighbors are to form a complete community. (d) In-degree shows the number of incoming edges, and (e) out-degree represents the number of outing edges for communication. It is shown in Table 5.1 that high *ANMO* score cannot be explained by conventional social network measures. The correlations between *ANMO* score and other network measures are very poor as shown in Figure 5.6. Intuition might suggest that an individual who receives email from many people (an individual with high in-degree) would be an accumulator of information with corresponding high *ANMO* score, but the correlation between *ANMO* and in-degree is 0.41 which shows a week correlation. Some of the measures show high correlation with each other, for example the correlation between in-degree and out-degree is 0.6401. The highest correlation of 0.8206 exist between PageRank and in-degree, however correlation between out-degree and PageRank is comparatively low. This shows that *ANMO* score and most of the conventional network measures are weekly correlated with each other so the high *ANMO* score in diffusion process is not dependent on the conventional measures of network.

Table 5.1 Correlation Comparison with Network Measures

Network measures	*ANMO*	Out-degree	In-degree	Clustering coef.	PageRank
Betweenness centrality	-0.006	0.0073	-0.0143	-0.0255	-0.0051
PageRank	0.188	0.4242	0.8206	-0.0588	-
Clustering coefficient	0.146	-0.1136	-0.0387	-	-
In-degree	0.418	0.6401	-	-	-

Out-degree	0.294	-	-	-	-

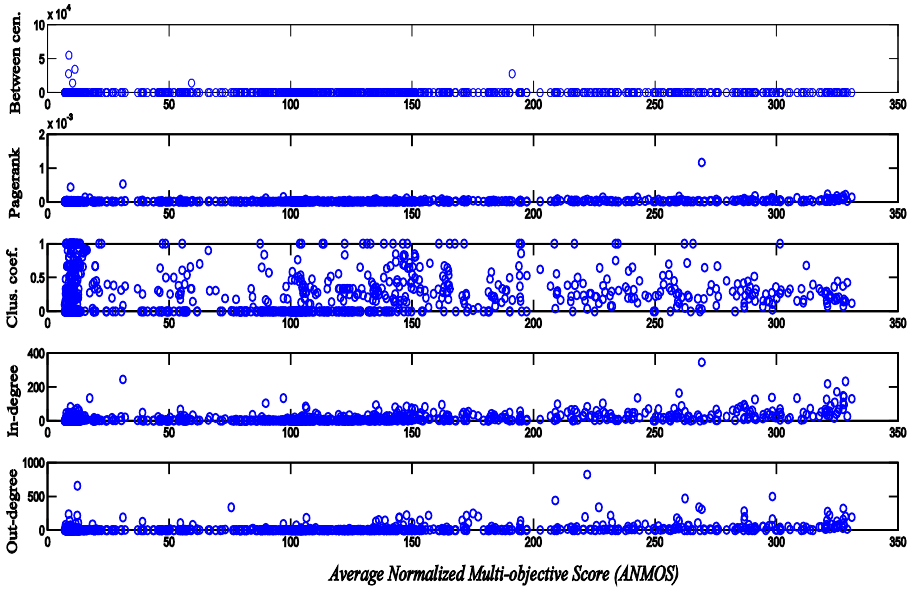

Figure 5.6 *ANMO* Correlation with Conventional Network Measures

5.1.8 Comparison of MODM and GADM

This experiment shows the effectiveness of MODM in comparison of GADM, which used evolutionary algorithm to diffuse the information in the network as a single objective optimization problem. The settings kept unchanged during the experiments. In Figure 5.7(a), the diffusion curves for the initial score are presented, the internal processing of both the algorithms are different. However upon beginning the diffusion process both algorithms show a similar state, initial scores are highly clustered and there is no dispersion. Figure 5.7(b) shows the final score curves, the diffusion curve of MODM started at the similar position of GADM however, MODM finished with high information maximization in more dispersion as compare to GADM diffusion curve. It shows that MODM outperforms the GADM in achieving a high diffusion rate in terms of an information maximization objective.

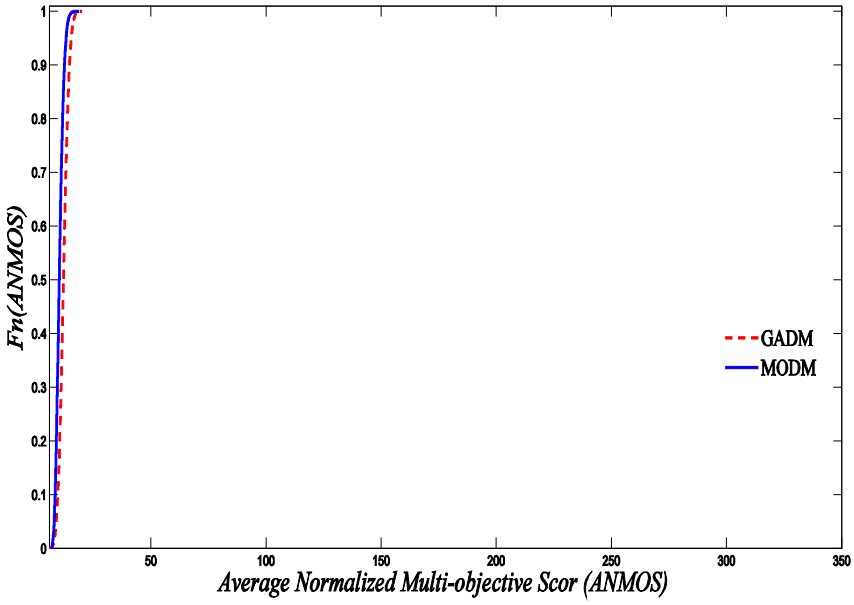

Figure 5.7 (a) Initial Score of MODM and GADM

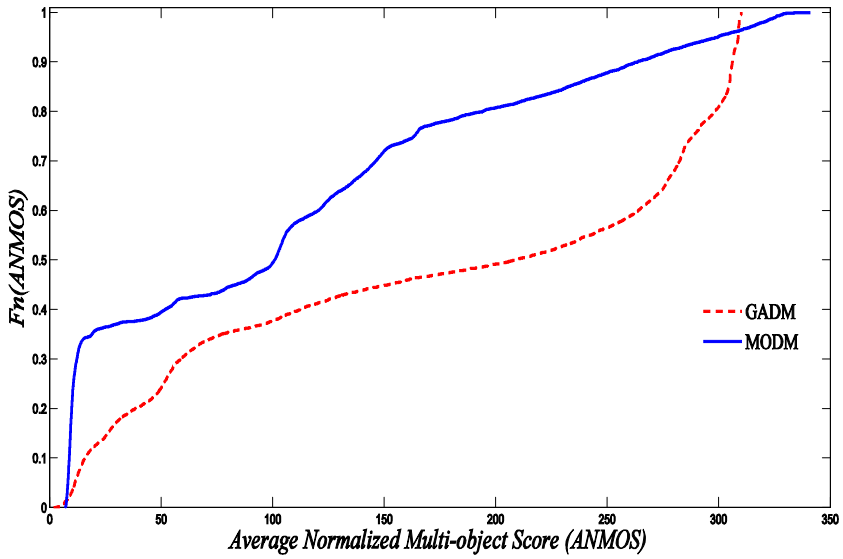

Figure 5.7 (b) Final Score of MODM and GADM

5.2 Evaluation and Results for Individual Activities

In this section, the results are presented to evaluate and validate our proposed framework to measure usefulness of behavior patterns and reliability of action prediction.

5.2.1 Datasets Description

The experiments are performed on the *Milan2009* and *Aruba* datasets collected in the CASAS smart home, a project of Washington State University, with full-time residents [97]. In the case of *Milan2009*, 31 motion sensors, one door sensor, and two temperature sensors were deployed at various locations and 15 activities were performed for 62 days. For *Aruba*, 31 motion sensors, three door sensors, five temperature sensors, and three light sensors were deployed and 11 activities were performed for 220 days. The details description of the datasets and annotation method can be found in [97].

Table 5.2 Characteristics of the Annotated Activities of CASAS Smart Home Datasets.

Milan2009							
Activities	Num.	Time	Sensor	Activities	Num.	Time	Sensor
Idle	-	911.233	5760	Evening Medicines	19	10.56	250
Bedto Toilet	89	379.37	1255	Guest Bathroom	330	952.31	10601
Sleeping	96	37,217.9	22172	Kitchen Activity	554	7,526.81	128942
Leave Home	214	4,229.47	4946	Master Bathroom	306	1,946.33	15071
Watch TV	114	5,919.72	23688	Master Bedroom	117	2,168.97	27337
Chores	23	684.82	7587	Meditate	17	109.94	1315
Desk Activity	54	743.74	7628	Morning Medicines	41	45.97	1023
Dining Rm Act	22	330.37	4295	Read	314	10,942.75	50281
Aruba							
Idle	-	59,495.15	903669	Enter Home	431	48.84457	2041

Meal Preparation	1,606	12,588.53	299300	Housekeeping	33	670.6926	11010
Bedto Toilet	157	428.833	1483	Leave Home	431	45.75227	1954
Relax	2,919	97,813.58	387,851	Respirate	6	51.38585	571
Sleeping	401	139,659.9	63,792	Wash Dishes	65	465.5383	10682
Eating	257	2,610.955	19,568	Work	171	2,920.759	17637

In Table 5.2, the characteristics of the *Milan2009*, and *Aruba* dataset are shown. The "Num." column shows activities count, "Time" column shows the time in seconds and "Sensor" column shows generated sensor events. It is very obvious of from that table that the number of annotated activities varies significantly. Some activities in both the datasets are annotated with high frequency like "Kitchen Activity" from Milan2009 and "Relax" from Aruba dataset. On the other hand some activities annotation frequency is very low like "Evening Medicines" from Milan2009 and "Respirate" from Aruba have very few occurrences as compare to other activities. Hence, these characteristics of activity annotation can affect the results in case of significant behavior mining using sequential pattern mining algorithm. The chances to discover the activities with very low annotation frequency in significant behavior are very low as these are not significant in the dataset.

5.2.2 Performance Measures

In order to evaluate our proposed framework, four standard metrics of precision, recall, F-measure and accuracy are used as performance measures. They are calculated using the values of the confusion matrix [98] and computed as:

matrix [98]

$$\text{Precision} = \frac{1}{Q}\sum_{i=1}^{Q}\frac{T_{P_i}}{N_{l_i}} \tag{5.1}$$

$$\text{Recall} = \frac{1}{Q}\sum_{i=1}^{Q}\frac{T_{P_i}}{N_{G_i}} \tag{5.2}$$

$$\text{F} - \text{Measure} = \frac{2}{\cdot\, precision \cdot recall} \tag{5.3}$$

$$Accuracy = \frac{\sum_{i=1}^{Q} TP_i}{Total} \quad (5.4)$$

where Q is the number of performed activities, TP is the number of true positives, NI is the total number of inferred labels and NG is the total number of ground truth labels.

5.2.3 Experiments and Discussion

This section presents that activity log can provide the basis to analyze the behavior of individuals from their daily routines. The proposed method has been implemented in MATLAB 7.6. The configuration of the computer is an Intel Pentium(R) Dual-Core 2.5 GHz with 3 GB of memory and Microsoft Window 7. For clear understanding of user behavior, 8 to 10 of the most significant sequential behaviors for *Milan2009* and *Aruba* are represented in Figures 5.8 and 5.9, respectively. Here, red and gray bars show sequential patterns in monitoring windows of three days with user specified support. In both the datasets the annotation frequency of activities varies significantly as shown in "Num." column of Table 5.3. To discover the sequential patterns we the support is set to 0.8% in order the get the most significant behaviors of daily life after patterns pruning. As a result the activities that are annotated with low frequency cannot be identified in the sequential behavior patterns. For instance, "Chores", "Dining Rm Activity", "Eve Meds" and "Meditate" have very few occurrences in the *Milan2009* data. Similarly, the occurrences of "House Keeping", "Respirate" and "Wash Dishes" in *Aruba* are much fewer compared to the other annotated activities.

It is obvious from Figure 5.8 that in sequential behavior the existence of "Kitchen Activity", "Master Bathroom" and "Read" are more noticeable due their high annotation ratio in the dataset. The performed experiments on the *Milan2009* activity log represent that the most prominent user behavior such as "Read" activity symbolizes that the user habit of reading before sleep is significant as compared to "Watching TV". Similarly, "Kitchen Activity" shows its sequence prior to "Desk Activity" and "Leave Home" that represents the users" eating behavior.

Furthermore, “Bed to Toilet” and “Master Bathroom” activities show bathroom usage habits before and during “Sleep”.

In Figure 5.9, a list of ten most significant sequential behaviors is shown for *Aruba*. The most obvious activities are “Enter Home”, “Meal Preparation”, “Relax” and “Leave Home”. The behavior analysis of the *Aruba* activity log shows the most substantial activities in the user routine such as “Meal preparation” illustrate the user’s habit of cooking after coming home. The “Relax” activity represents his behavior of relaxing before sleep and “Bed to Toilet” characterizes user’s habit to go to the toilet during sleep. Although, the “Work” activity is not noticeable in most of the activity patterns but “Enter Home” activity after “Work” signifies that user go out of home for “Work” activity most of the times.
The above frequent sequential patterns can be effectively utilized to analyze the lifestyle of inhabitants in terms of significant routine discovery. Furthermore, these routine behaviors facilitate the personalized service providers (*i.e*., caregivers) to estimate the forthcoming action of inhabitants in order to take proactive actions for their better lifestyle.

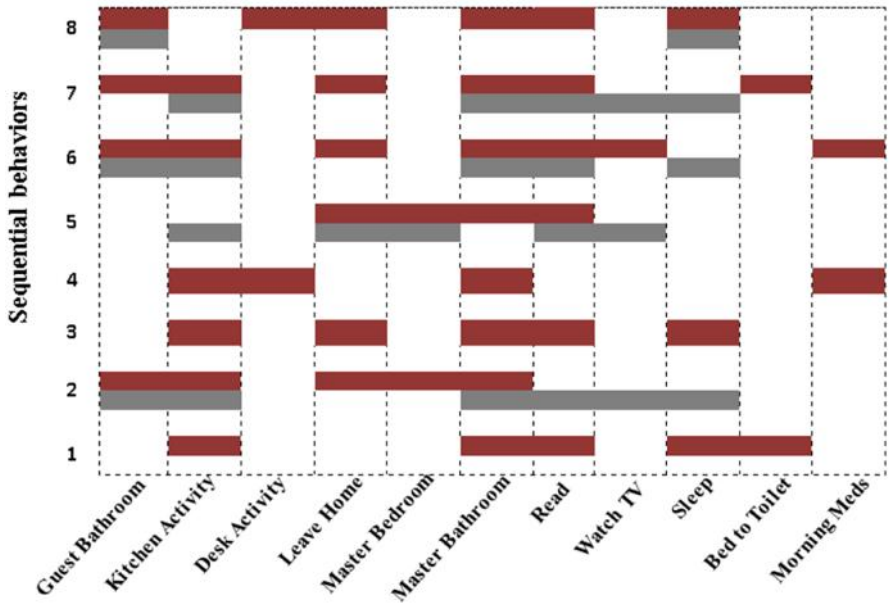

Figure 5.8 Sequential Behavioral Patterns for *Milan2009.*

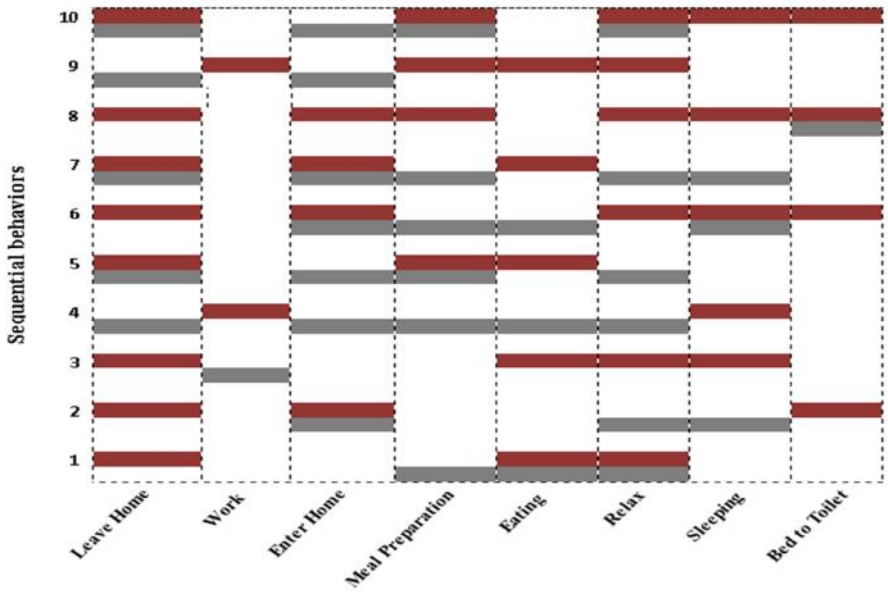

Figure 5.9 Sequential Behavioral Patterns for *Aruba.*

To evaluate the activity prediction method, the classification of future actions from the past sequential behaviors is used. The goal of the experiments is to determine how well our method performs in predicting the future actions of inhabitants. In our proposed approach, the sequences of 8 to 10 activities are consider to predict the next action. The representative activity sequences with predicted activity for *Milan2009* and *Aruba* are shown in Figures 5.10 and 5.11, respectively.

Here, each activity is symbolized by a specific color and "set of activities" represents the routine behaviors in different sequences and "predicted action" represents the outcome for the particular behavior. For example, in the daily routine of the inhabitant in *Milan2009*, activities prior to "Sleep" are "Master Bedroom", "Bed to Toilet", "Watch TV", and "Read", and subsequent activities to "Sleep" are "Kitchen", "Morning Meds" and "Desk Activity". This represents that prediction of "Sleep" as a forthcoming action depends on the order of priory performed activities as shown in Figure 5.10. Similarly for *Aruba*, "Sleep" activity is in-between the

"Eating", "Relax", "Bed to Toilet", and "Meal Preparation", so the prediction for "Sleep" as a future action depends particularly on the order in which these activities are performed in the daily routine of the inhabitant, as shown in Figure 5.11.

The proposed method is compared with the results of Hidden Markov Model (HMM)[90], Neural Network (NN)[89] and Support Vector Machine (SVM)[92] which are well known models for labeling sequences. The quality of a predicted activity is determined based on how closely the predicted activity resembles inhabitant's real future action. The precision, recall, and F-measure is computed, as shown in Figure 5.12, 5.13 and in Table 5.3. For both the datasets, the performance of NN is very low. The performance of HMM is better than SVM for Milan2009 however SVM outperformed HMM for Aruba. In conclusion, CRF performed better in comparison to HMM, NN and SVM for all datasets, the increase of 6.61% and 6.76% in F-measure is achieved as compare to HMM for *Milan2009* and *Aruba* respectively.

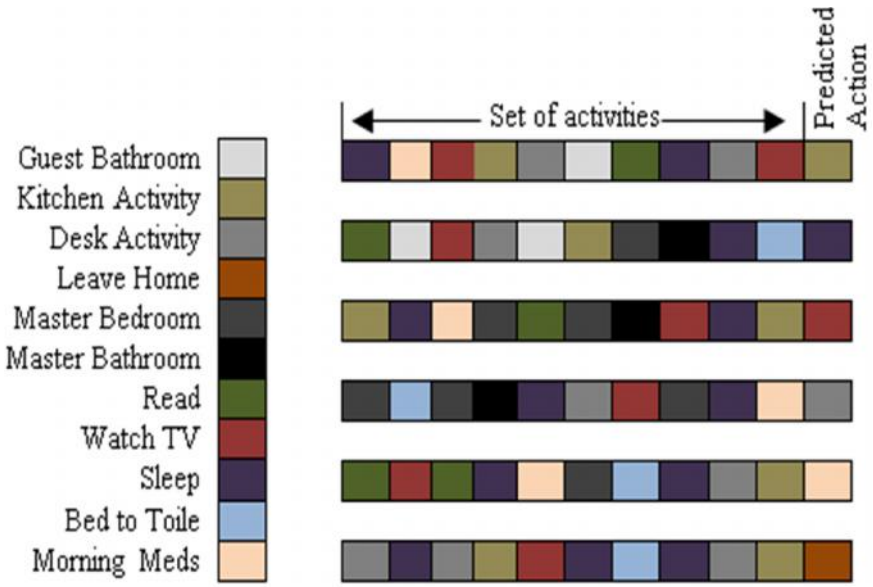

Figure 5.10 Behavioral Predictions for *Milan2009.*

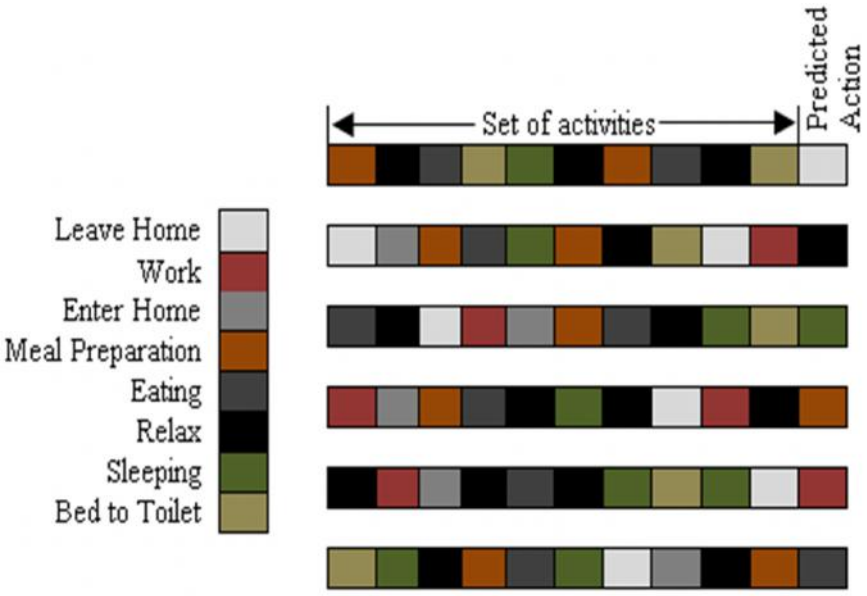

Figure 5.11 Behavioral Predictions for *Aruba.*

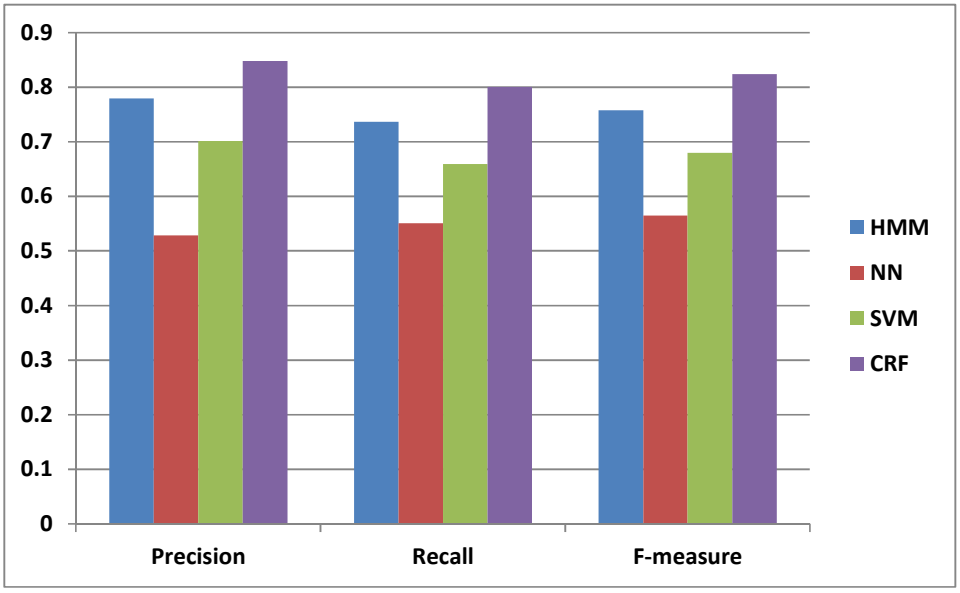

Figure 5.12 Action Prediction Comparison of CRF with HMM, NN, and SVM for Milan 2009

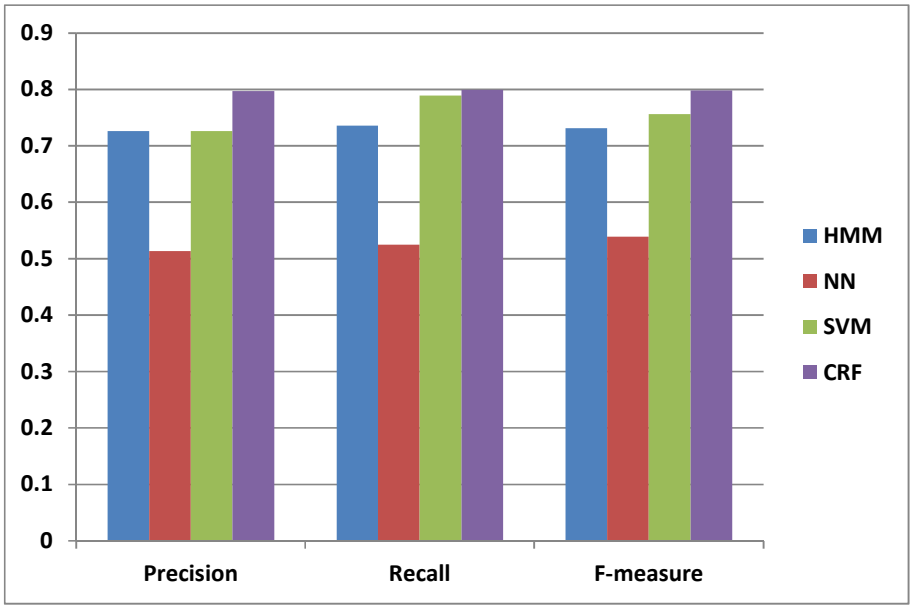

Figure 5.13 Action Prediction Comparison of CRF with HMM, NN, and SVM for Aruba

Table 5.3 Accuracy Performance for Action Prediction.

Dataset	Model	Precision	Recall	F-measure
Milan2009	HMM	0.7796	0.7363	0.7574
	NN	0.5284	0.5511	0.5645
	SVM	0.7018	0.6589	0.6769
	CRF	**0.8478**	**0.8006**	**0.8235**
Aruba	HMM	0.7261	0.7356	0.7308
	NN	0.5138	0.5249	0.5393
	SVM	0.7263	0.7891	0.7563
	CRF	**0.7971**	**0.7996**	**0.7984**

5.3 Summary

This chapter gives the details about implementation strategy that what were the system setting, operating systems and programming languages used to implement the proposed system for group and individual activities. The datasets used to

validate the proposed systems are selected from openly available datasets such as Enron email dataset for group activities and for individual activities datasets are selected from CASAS smart home project. The CASAS smart home is a well-known smart home project to collect the daily life activities. All the openly available CASAS datasets are analyzed and have selected the largest data Milan2009 and Aruba. The results of group activities show that there exist a significant the correlation between the importance of each individual and his information processing capability. This indicates that some individuals in the network are naturally and significantly better connected in terms of receiving information irrespective of the starting position of the diffusion process. For individual activities the main objective is to process the lifelog of individual in order to analyze his lifestyle. The extraction of significant behavioral patterns and precise activity predictions with 6.76% increase in F-measure. All this collectively help in understanding the users actions to gain knowledge about their habits and preferences.

Chapter 6

Conclusion and Future Work

This chapter concludes the research carried out in this dissertation. The subsequent sections summarize the contributions made in this dissertation to the area of group and individual activities based human behavior and analysis. In the end, conclusion about this study is given with potential future directions that can be explored to extend the research carried in this dissertation.

6.1 Conclusion

Behavior analysis and prediction is an emerging field of research that enables a large number of human centric applications. Therefore many researchers have been spending their time and efforts in proposing practical solutions for behavior analysis. So far two of the most important challenges in the area of behavior analysis are how to analyze significant behavior from the daily life activities performed within a group or individually and how to construct learning models which are able to resolve the uncertainties between the human behaviors.

In this thesis, multi-objective diffusion model is proposed that propagates multiple pieces of information with evolution fitness criteria by designing an evolutionary algorithm. In order to propagate multiple types of information in one diffusion process, the set of information is modeled into a binary schema where each schema represents one type of information with its associated score. Furthermore, information history log is maintained for each individual to keep track of all incoming and outgoing information in all time stamps. This helps to predict a more accurate class of information diffusion by holding the monotonicity property about information. The information value of each individual is calculated based on evolution fitness criteria for each information type. Evolution fitness criteria utilize the benefits of score generated by the schema and

information history maintained in the information history log. Our experimental results on a real world dataset show that our model is able to simulate the rich class of diffusion model and predict the information flow in the multi-objective environment. Finally, the results show that a few individuals in the network always obtain a high information rank irrespective of the start of the diffusion process.

For analysis of individual activities personalized service providers need to know the common behaviors and preferences of the inhabitants in leveraging the use of technology for different application domains. In this thesis, a unified framework for behavior analysis and action prediction is proposed. This informs the service provider about inhabitants' significant behavior in order to perform meaningful interventions. The proposed model is based on data mining based reasoning. First the data mining techniques is applied to identify the significant behavior and then these behaviors are used for the learning of machine learning module to predict the future actions. In the proposed framework, the recognized activity log is utilized for behavioral pattern discovery with the help of frequent sequential mining technique on a set of activities that are performed in a temporal sequence of three days. Finally, CRF is investigated for the actions that occur together in order to predict the next activity from a current situation. Our study found that identification of behavior patterns and prediction of forthcoming action with high precision signifies the possibility of helping people by analyzing the long-term data of one's behavior to fulfill his needs in the current circumstances and in future.

6.2 Future Work

In the future, I shall enhance the MODM with a more realistic class of diffusion model to better understand the dynamics of diffusion process based on the underlying network. I shall investigate the possible use of genetic programming to learn about a diffusion model that matches an observed spread.

I intend to combine the benefits of group and individual activities in order to unified understanding of human behavior. I shall analyze the human activity log first to process

the individual activities and identify the communities based on behavior segregation and then analyze the trend of group activities within identified behavioral communities. Additionally, the action prediction will help us to identify the future behavior of communities which could be use intelligently to recommend the intended group activities.

Bibliography

1. C.L. Baker, N.D. Goodman, and J.B. Tenenbaum. Theory-based social goal inference. In Proceedings of the thirtieth annual conference of the cognitive science society, pages 1447-1452, 2008.

2. D. A. Baldwin and J. A. Baird. Discerning intentions in dynamic human action. Trends in Cognitive Sciences, 5(4) pages 171- 178, 2001.

3. F. Wang, K. M. Carley, D. Zeng, and W. Mao. Social computing: From social informatics to social intelligence. IEEE Intelligent Systems, 22(2), pages 79–83, 2007.

4. T. Fawcett and F. Provost. Adaptive Fraud Detection. Data Mining and Knowledge Discovery, vol(1), pages 291-316, 1997

5. J. Frooman. Socially irresponsible and illegal behavior and shareholder wealth: A meta-analysis of event studies. Business & Society, 36, pages 221-249, 1997

6. S. Kiesler, and L. Sproull. Group decision making and communication technology. Organizational Behavior and Human Decision Processes, 52(1), pages 96–123, 1992

7. T. M. Jones. Ethical decision making by individuals in organizations: An issue contingent model. Academy of Management Review. 16, pages 366-395, 1991

8. R. Bell, Y. Koren, and C. Volinsky. Modeling relationships at multiple scales to improve accuracy of large recommender systems. In KDD, New York, NY, USA, pages 95-104, 2007.

9. S. Asur and B.A. Huberman. Predicting the future with social media. In Web Intelligence and Intelligent Agent Technology (WI-IAT), 2010 IEEE/WIC/ACM International Conference on, volume 1, pages 492-499, 2010.

10. J.M. Serrano, and S. Saugar. An architectural perspective on multiagent societies. In: Proceedings of AOSE-10 at AAMAS-10, pages 85–90, 2010

11. Y. Gu, and M. Soutchanski. Decidable reasoning in a modified situation calculus. In: Proceedings of IJCAI-07, pp. 1891–1897 (2007)

12. K. Subramanian. Task space behavior learning for humanoid robots using Gaussian mixture models. In: Proceedings of AAAI-10, pages 1961, 2010

13. B. Christel and P.K. Joost. Principles of Model Checking. MIT Press, Cambridge 2008

14. L. Backstrom and J. Leskovec. Supervised random walks: predicting and recommending links in social networks. In WSDM 2011, pages 635-644, 2011

15. http://www.nielsen.com/us/en/newswire/2010/facebook-users-average-7-hrs-a-month-in-january-as-digital-universe-expands.html

16. F. Iram, F. Muhammad, L. Young-Koo and L. Sungyoung. Analysis and effects of smart home dataset characteristics for daily life activity recognition, The Journal of Supercomputing, pages 760–780, 2013

17. R. Parisa, J.C. Diane, B.H. Lawrence, S.Maureen. Discovering activities to recognize track in a smart environment. IEEE Trans Knowl Data Eng, 23(4), pages 527–539, 2011

18. D.J. Cook, A. Crandall, G. Singla and B. Thomas. Detection of social interaction in smart spaces. Cybern Syst 2(41), pages 90–104, 2010

19. D. Agrawal, C. Budak, and A. Abbadi. Information Diffusion in Social Networks: Observing and Influencing Societal Interests. In proceedings of 37th International Conference on Very Large Data Bases, pages 1512-1513, 2011

20. S. Fox. The social life of health information. Technical report, Pew Internet & American Life Project, 2011

21. E. Gilbert, and K. Karahalios K. Predicting tie strength with social media. In Proceedings of the 27th International Conference on Human Factors in Computing Systems. pages 211-220, 2009

22. M. Rodriguez, J. Leskovec, and A. Krause. Inferring networks of diffusion and influence. In Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining, pages 1019-1028, 2010

23. S. Aral, and D. Walker. Creating social contagion through viral product design: A randomized trial of peer influence in networks. Management Science, 57(9) pages 1623-1639, 2011

24. E. Bakshy, I. Rosenn, C. Marlow, and L. Adamic. The role of social networks in information diffusion. In proceeding of 21st ACM conference on the World Wide Web, pages 519-528, 2012

25. P. Rashidi and D.J. Cook. Mining and Monitoring Patterns of Daily Routines for Assisted Living in Real World Settings. In Proceedings of the 1st ACM International Health Informatics Symposium, pages. 336–345, 2010

26. J. Kleinberg. Cascading behavior in networks: algorithmic and economic issues. Algorithmic Game Theory, 2007

27. N. Ramasuri, and N. Yadati. A Shapley Value-Based Approach to Discover Influential Nodes in Social Networks. IEEE T. Automation Science and Engineering 8(1), pages 130-147, 2011

28. D. Hayes. Cascade Training and Teachers' Professional Development, ELT Journal, 54(2), pages 135-45, 2000

29. K. Saito, M. Kimura, K. Ohara, and H. Motoda. Learning continuous-time information diffusion model for social behavioral data analysis. In: Proceedings of the advances in machine learning, pages 322–337, 2009

30. E.M. Rogers. Diffusion of innovations. Free Press, New York, 1995

31. S. Aral, E. Brynjolfssen, and M.V. Alstyne. Productivity effects of information diffusion in networks. MIT Center for Digital Business, paper 234, 2007

32. N. Christakis, and J. Fowler. The spread of obesity in a large social network over 32 years. N Engl J Med 357, pages 370–379, 2007

33. D. Kempe, J. Kleinberg, and E. Tardos. Maximizing the spread of influence in a social network. In: Proceedings of the 9th international conference on knowledge discovery and data mining, pages 137–146, 2003

34. D Kempe, J. Kleinberg, and É. Tardos. Influential nodes in a diffusion model for social networks. In: Proceedings of the 32nd international conference on automata, languages and programming, pages 1127– 1138, 2005

35. K. Saito, M. Kimura, K. Ohara, and H. Motoda. Learning continuous-time information diffusion model for social behavioral data analysis. In: Proceedings of the advances in machine learning, pp 322–337, 2009

36. K. Saito, M. Kimura, K. Ohara, and H. Motoda. Learning continuous-time information diffusion model for social behavioral data analysis. In: Proceedings of the advances in machine learning, pages 322–337, 2009

37. S. Kazumi, K. Masahiro, O. Kouzou, M. Hiroshi. Learning Continuous-Time Information Diffusion Model for Social Behavioral Data Analysis , pages 322-337, 2009

38. M. Richardson, and P. Domingos. Mining knowledge-sharing sites for viral marketing. In Proceedings of KDD 2002, pages 61–70, 2002

39. K. Masahiro, S. Kazumi, and M. Hiroshi. Solving the Contamination Minimization Problem on Networks for the Linear Threshold Model. PRICAI. pages 977-984, 2008

40. H. Xinran, S. Guojie, Wei Chen, and J. Qingye: Influence Blocking Maximization in Social Networks under the Competitive Linear Threshold Model Technical Report. CoRR abs/1110.4723, 2011

41. J. Horn, N. Nafpliotis and D. E. Goldberg. A Niched Pareto Genetic Algorithm for Multiobjective Optimization. In Proceedings of the First IEEE Conference on

Evolutionary Computation, IEEE World Congress on Computational Intelligence, pages 82-87, 1994

42. M. Lahiri, and M. Cebrian. The genetic algorithm as a general diffusion model for social networks. In: Proceedings of the 24th AAAI conference on artificial intelligence, pages 494–499, 2010

43. H. Carl Mooney and J.F. Roddick. Sequential pattern mining -- approaches and algorithms. ACM Comput. Surv. 45, 2, Article 19 (March 2013), 39 pages. 2013

44. S. Laxman and P. S. Sastry, “A survey of temporal data mining,” Acad. Proc. Eng. Sci., Indian Acad. Sci., vol. 31, no. 2, pp. 173–198, 2006.

45. T. Huiyi, H.J. Cai, and L. Yong. Frequent Patterns of Investment Behaviors in Shanghai Stock Market. n proceeding of International Conference on Computer Science and Software Engineering, Volume: 4, pages 325 – 328, 2008

46. R. Iv´ancsy and I. Vajk. Frequent pattern mining in web log data. Acta Polytechnica Hungarica, 3(1), 2006.

47. O.T. Behrooz, A.Y. Sihem, T. Alexandre, B. Aurélie, G. Éric, and R. Marie-Christine. Towards a Framework for Semantic Exploration of Frequent Patterns. Proceedings of the 3rd International Workshop on Information Management for Mobile Applications, pages 7-14, 2013

48. S. Vignesh, P. Robert, U. Vignesh, D. Bharathidasan, S. Rajasekaran. Mining Frequent Patterns and Prediction of User Behavior in Mobile Commerce. International Journal of Application or Innovation in Engineering & Management (IJAIEM) ,Vol. 2, Issue 3, pages,238-242 , 2013

49. H. Li-Fu, H. Chuin-Chieh, K. Yi-Chen, Mining Frequent Purchase Behavior Patterns for Commercial Websites, Computational Collective Intelligence. Semantic Web, Social Networks and Multiagent Systems, pp 732-742, 2009.

50. J. Ayres, J. Gehrke, T. Yiu, and J. Flannick. Sequential PAttern Mining Using Bitmaps. In Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Edmonton, Alberta, Canada, pages 429-435, 2002

51. V. Guralnik and K.Z. Haigh. Learning models of human behavior with sequential patterns. In Proceedings of the AAAI-02 workshop “Automation as Caregiver”. (2002) pages 24–30, 2002

52. G. Bombini, N.D. Mauro, S. Ferilli, and F. Esposito, Classifying Agent Behaviour through Relational Sequential Patterns. Agent and Multi-Agent Systems: Technologies and Applications Lecture Notes in Computer Science Volume 6070, pages 273-282, 2010

53. K.C. Srikantaiah, K.N. Krishna, K.R. Venugopal, and L.M. Patnaik. Bidirectional Growth based Mining and Cyclic Behaviour Analysis of Web Sequential Patterns, International Journal of Data Mining & Knowledge Management Process (IJDKP), 03(02), pages 49 - 68. 2013

54. A.D. Lattner, A. Miene, U. Visser, and O. Herzog, Sequential Pattern Mining for Situation and Behavior Prediction in Simulated Robotic Soccer, RoboCup 2005: Robot Soccer World Cup IX Lecture Notes in Computer Science Volume 4020, , pages 118-129, 2006

55. F. Rasheed. Efficient Periodic Pattern Mining in Time Series & Sequence Databases. Ph.D. Dissertation. University of Calgary, Calgary, Alta., Canada, Canada. AAINR75499. 2011

56. Z. Li, B. Ding, J. Han, R. Kays, and P. Nye. Mining periodic behaviors for moving objects. In *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining* (KDD '10). pages 1099-1108. 2010

57. M. Lahiri and T.Y. Berger-Wolf. Mining Periodic Behavior in Dynamic Social Networks. Proc. of the 8th IEEE International Conference on Data Mining (ICDM 2008), Pisa, Italy. Pages, 373 - 382 December 2008.

58. C.D Nugent, D.D.Finlay, R.J. Davies, H.Y.Wang, H. Zheng, J. Hallberg, K. Synnes, and M.D. Mulvenna. HomeML: An open standard for the exchange of data within smart environments. Lect. Notes Comput. , 4541, pages 121–129. 2007

59. P. Rashidi, Cook, D.J., and E.M. Schmitter. Discovering activities to recognize and track in a smart environment. IEEE Trans. Knowl. Data Eng, 23, pages 527–539, 2011

60. B. Chikhaoui, S. Wang, and H. Pigot. A New Algorithm Based on Sequential Pattern Mining for Person Identification. In Proceedings of the 4th International Workshop on Knowledge Discovery from Sensor Data, pages 19–28, 2010

61. Q. Wang, and Y. Shen. The Effects of Fusion Structures on Image Fusion Performances. In Proceedings of the 21st IEEE Instrumentation and Measurement Technology Conference, pages 468–471, 2004

62. B. Chen, F. Sun, and J. Hu. Local Linear Multi-SVM Method for Gene Function Classification. In Proceedings of the 2nd World Congress on Nature and Biologically Inspired Computing, pages 183–188, 2010

63. X. Hong, C. Nugent, M. Mulvenna, S. McClean, B. Scotney, and S. Devlin. Evidential fusion of sensor data for activity recognition in smart homes. Perv Pervasive Mob. Comput, 5, pages 236–252, 2009

64. R.W. Xu, and L. He, GACEM: Genetic algorithm based classifier ensemble in a multi-sensor system. Sensors, 8, pages 6203–6224, 2008

65. C. Fernández, J.P.Lázaro, and J.M. Benedí, Workflow mining application to ambient intelligence behavior modeling. Lect. Note. Comput. Sci. 5615, pages 160–167, 2009

66. A. Aztiria. Learning frequent behaviours of the users in intelligent environments. JAISE, 2, pages 435–436, 2010

67. F. Doctor, H. Hagras, and V. Allaghan. A fuzzy embedded agent-based approach for realizing ambient intelligence in intelligent inhabited environments. IEEE Trans. Syst. Man Cybern, 35, pages 55–65, 2005

68. A. Aztiria, A. Izaguirre, and C. Juan. Learning patterns in ambient intelligence environments: A survey. Arti. Intell. Rev. 34, pages 35–51, 2010

69. C. Liming, H. Jesse, D.N. Chris, J.C. Diane, and Z.W. Yu. Sensor-based activity recognition: A survey. IEEE Trans. Syst. Man Cybern. Part C, 99, pages 1–19, 2012

70. J. B. J. Bussmann, W. L. J. Martens, J. H. M. Tulen, F. C. Schasfoort, H. J. G. van den Berg-Emons, H. J. Stam, Measuring daily behavior using ambulatory accelerometry: The Activity Monitor, Behavior Research Methods, Instruments, & Computers August 2001, Volume 33, Issue 3, pp 349-356

71. M.G. Jennifer, and B.P. Arthur. Feeling good-doing good: A conceptual analysis of the mood at work-organizational spontaneity relationship, Psychological Bulletin, Vol 112(2), 1992

72. K.W. Axhausen, A.a Zimmermann, S. Schönfelder, G. Rindsfüser, and T. Haupt, Observing the rhythms of daily life: A six-week travel diary, Transportation, Volume 29, Issue 2, pages 95-124, 2002

73. T. Fawcett, and F. Provost, Adaptive Fraud Detection, Data Mining and Knowledge Discovery, Volume 1, Issue 3, pages 291-316, 1997

74. P.K. Chan, FL Miami, W. Fan, A.L. Prodromidis, and S.J. Stolfo. Distributed data mining in credit card fraud detection, Intelligent Systems and their Applications, IEEE (Volume:14 , Issue: 6),pages 67 – 74, 1999

75. K. Yufeng, L. Chang-Tien, S. Sirwongwattana, and H. Yo-Ping, Survey of fraud detection techniques. IEEE International Conference on Networking, Sensing and Control. Pages 749 – 754, 2004

76. W. Fei-Yue, K.M. Carley, Z. Daniel,and W. Mao, Social Computing: From Social Informatics to Social Intelligence, Intelligent Systems, IEEE (Volume:22 , Issue: 2), pages 79 – 83, 2007

77. Benjamin, and S. Lorna. Structural analysis of social behavior. ,Psychological Review, Vol 81(5), Sep, pages 392-425, 1974

78. M. Pantic, A. Pentland, A. Nijholt, and T. S. Huang, Human Computing and Machine Understanding of Human Behavior: A Survey, Artifical Intelligence for Human Computing Lecture Notes in Computer Science Volume 4451, 2007, pp 47-71

79. S.A. Hofmeyr, S. Forrest, and A. Somayaji, Intrusion detection using sequences of system calls, Journal of Computer Security, Volume 6, Number 3, pages 151-180, 1998

80. D.E. Denning, SRI International, An Intrusion-Detection Model, Software Engineering, IEEE Transactions on (Volume:SE-13 , Issue: 2), pages, 222 – 232, 1987

81. Y. Dit-Yan, and D. Yuxin. Host-based intrusion detection using dynamic and static behavioral models, Pattern Recognition, Volume 36, Issue 1, Pages 229-243, 2003

82. S. Alper, D.Tjosvold, and K.S. Law, Interdependence and Controversy in Group Decision Making: Antecedents to Effective Self-Managing Teams, Organizational Behavior and Human Decision Processes, Volume 74, Issue 1, Pages 33-52, 1998

83. B.B. Baltes, M.W. Dickson, M.P. Sherman, C.C. Bauer, J.S. LaGanke, Computer-Mediated Communication and Group Decision Making: A Meta-Analysis, Organizational Behavior and Human Decision Processes, Volume 87, Issue 1, Pages 156-179, 2002

84. K.L. Kraemer and J.L. King. Computer-based systems for cooperative work and group decision making. ACM Comput. Surv. 20, 2, pages 115-146., 1998

85. M.J. Stampfer, M.D. Frank, J. E. Manson, M.D. Eric, B. Rimm, Sc.D., and W. C. Willett, M.D, Primary Prevention of Coronary Heart Disease in Women through Diet and Lifestyle, N Engl J Med 343, pages 16-22, 2000

86. Weiner, and Bernard, A cognitive (attribution)-emotion-action model of motivated behavior: An analysis of judgments of help-giving., Journal of Personality and Social Psychology, Vol 39(2), 1980

87. D.E. Goldberg. Genetic algorithms in search, optimization and machine learning. Addison Wesley Longman, Reading, 1989

88. H. Holland. Building blocks, cohort genetic algorithms, and hyperplane-defined functions. Evol Comput 8(4), pages 373–391, 2000

89. S. Helal, E. Kim, and S. Hossain. “Scalable Approaches to Activity Recognition Research”. In proceedings of the workshop of How to do good activity recognition research? Experimental methodologies, evaluation metrics, and reproducibility issues, pages 450-453, 2010

90. T.L.M. Kasteren, G. Englebienne, and B.J.A. Krose “An Activity Monitoring System for Elderly Care using Generative and Discriminative Models”, Personal and Ubiquitous Computing, vol. 14, no. 6, pages 489-498. 2010

91. B.J.A. Krose, T.L.M. Kasteren, C.H.S. Gibson, and T. Dool, “CARE: Context Awareness in Residences for Elderly”. In the proceeding of 6th International

Conference of the International Society for Gerontechnology, pages 101-105, 2008

92. A. Fleury, M. Vacher, and N. Noury "SVM-Based Multimodal Classification of Activities of Daily Living in Health Smart Homes: Sensors, Algorithms, and First Experimental Results". IEEE Transactions on Information Technology in Biomedicine, vol. 14, no. 2, pages 274-83, 2010

93. http://www.cs.cmu.edu/~enron/ [Oct 25, 2013]

94. .A. Bader, S. Kintali, K. Madduri, and M. Mihail. Approximating betweenness centrality. In Proceedings of the 5th workshop on algorithms and models for the web-graph, pages 134–137, 2007

95. Y. Jing, and S. Baluja. PageRank for product image search. In: Proceedings of WWW, pages 307–316, 2008

96. Z. Liu, C. Wang C, Q. Zou, and H. Wang Clustering coefficient queries on massive dynamic social networks. In: Proceedings of the 11th international conference on web-age information management, pages 115–126, 2010

97. CASAS Smart Home Project. Available online: http://www.ailab.wsu.edu/casas/ (accessed on 16 February 2013).

98. T.L.M. Kasteren, H. Alemdar, and E. Cem. Effective Performance Metrics for Evaluating Activity Recognition Methods. In Proceedings of ARCS 2011 Workshop on Context-Systems Design, Evaluation and Optimization, Como, Italy, 22–23, 2011.

Appendix

List of Publications

Journal Publications

1. **Iram Fatima**, Muhammad Fahim, Young-Koo Lee and Sungyoung Lee, "**A Genetic Algorithm-based Classifier Ensemble Optimization for Activity Recognition in Smart Homes**", Transaction on Internet Information Systems (TIIS) (SCIE, IF: 0.35)

2. **Iram Fatima**, Sajal Halder, Muhammad Aamir Saleem, Rabia Batool, Muhammad Fahim, Young-Koo Lee and Sungyoung Lee, "**Smart CDSS: Integration of Social Media and Interaction Engine (SMIE) in Healthcare for Chronic Disease Patients**", The Journal of Multimedia Tools and Applications (SCIE, IF:1.014), 2013

3. **Iram Fatima**, Muhammad Fahim, Young-Koo Lee and Sungyoung Lee, "**Analysis and effects of smart home dataset characteristics for daily life activity recognition**", The Journal of Supercomputing (SCI, IF:0.971), 2013

4. **Iram Fatima**, Muhammad Fahim, Young-Koo Lee and Sungyoung Lee, "**MODM: Multi-Objective Diffusion Model for Dynamic Social Networks using Evolutionary Algorithm**", The Journal of Supercomputing (SCI, IF:0.971), 2013

5. **Iram Fatima**, Muhammad Fahim, Young-Koo Lee and Sungyoung Lee, "**A Unified Framework for Activity Recognition-Based Behavior Analysis and Action**

Prediction in Smart Homes", Sensors (SCIE, IF: 1.953) , EISSN: 1424-8220, Vol. 13, Issue 2, 2013

6. M Hussain, AM Khattak, WA Khan, **I Fatima**, MB Amin, Z Pervez, R Batool, MA Saleem, M Afzal, M Faheem, MH Saddiqi, SY Lee, K Latif "**Cloud-based Smart CDSS for chronic diseases**" The Journal of Health and Technology, 1-23, 2013

7. Muhammad Fahim, **Iram Fatima**, Sungyoung Lee and Young-Tack Park, "**EFM: Evolutionary Fuzzy Model for Dynamic Activities Recognition using a Smartphone Accelerometer**", International Journal of Applied Intelligence, ISSN:1573-7497, (SCI, IF: 1.853), 2013

8. Muhammad Fahim, **Iram Fatima**, Sungyoung Lee and Young-Koo Lee, "**EEM: Evolutionary Ensembles Model for Activity Recognition in Smart Homes**", International Journal of Applied Intelligence, (SCI, IF: 1.853), ISSN:1573-7497, June 14, 2012

Conference Publications

1. **Iram Fatima**, Muhammad Fahim, Rabia Batool, Muhammad Aamir Saleem, Young-Koo Lee, Sungyoung Lee, "**CDSS: Integration of Social Media Interaction Engine (SMIE) in Healthcare for Chronic Disease Patients**". International Conference on Advanced IT, engineering and Management (AIM 2013), Feb. 21-23, Seoul, Korea

2. **Iram Fatima**, Muhammad Fahim, Young-Koo Lee, Sungyoung Lee, "**Classifier Ensemble Optimization for Human Activity Recognition in Smart Homes**", The

7th International Conference on Ubiquitous Information Management and Communication (ICUIMC IMCOM 2013), Kota Kinabalus, Malaysia, Jan 17-19,

3. **Iram Fatima**, Muhammad Fahim, Young-Koo Lee, Sungyoung Lee, "**Effects of Smart Home Dataset Characteristics on Classifiers Performance for Human Activity Recognition**". In 4th International Conference of Computer Science and its Applications (CSA 2012) Jeju, Korea, Nov 22-25, pp 271-281, 2012

4. **Iram Fatima**, Asad Masood Khattak, Young-Koo Lee and Sungyoung Lee, "**Automatic Documents Annotation by Keyphrase Extraction in Digital Libraries using Taxonomy**", In FutureTech 2011 Conference, Crete, Greece, June 28-30, 2011

5. **Iram Fatima**, Muhammad Fahim, Young-Koo Lee, Sungyoung Lee, "**CDSS: Integration of Social Interaction and Smart Space for Chronic Disease Patients**". In proceedings of Korea Computer Congress (KCC'11), Kyungju, Korea.

6. **Iram Fatima**, Young-Koo Lee, Sungyoung.Lee, "**Domain Specific Annotation of Digital Documents through Keyphrase Extraction**" The 35th Conference of Korea Information Processing Society (KIPS), Jeju, Korea. 14 May, 2011.

7. **Iram Fatima**, Muhammad Fahim, Guan Donghai, Young-Koo Lee and Sungyoung Lee, "**Socially Interactive Cloud Based CDSS for u-life care**", 5th International Conference on Ubiquitous Information Management and Communication (ACM, ICUIMC'11), Seoul, Korea, February 2011.

8. Muhammad Fahim, Le Ba Vui, **Iram Fatima**, and Sungyoung Lee, Yongik Yoon, "**A Sleep Monitoring Application for u-lifecare using Accelerometer Sensor of Smartphone**". Published in 7th International Conference on Ubiquitous Computing and Ambient Intelligence, Carrillo - Guanacaste (Costa-Rica), Dec 2-3, 2013

9. Muhammad Aamir Saleem, **Iram Fatima**, Kifayat Ullah Khan, Young-Koo Lee and Sungyoung Lee, "**Trajectory Based Activity Monitoring and Healthcare Provisioning**", The Tenth IEEE International Conference on Pervasive Intelligence and Computing (PiCom2012), Changzhou, China, December 17-19, 2012

10. Muhammad Fahim, **Iram Fatima**, Sungyoung Lee, Young-Koo Lee, "**Activity Recognition Based on SVM Kernel Fusion in Smart Home**". In 4th International Conference of Computer Science and its Applications (CSA 2012) Jeju, Korea, Nov 22-25, pp 271-281, 2012

11. Muhammad Aamir Saleem, **Iram Fatima**, Kifayat Ullah Khan Young-Koo Lee and Sungyoung Lee, "**Personal Tracking using Static Trajectory Locations**", International Conference on Information Technology(YSEC-2012) Sunchon, South Korea April, 26-28, 2012

12. Muhammad Fahim, **Iram Fatima**, Sungyoung Lee and Young-Koo Lee, "**Activity Recognition: An Evolutionary Ensembles Approach**", International Joint

Conference on Pervasive and Ubiquitous Computing, Workshop SAGWare, Beijing, China, Sep, 2011

13. Muhammad Fahim, **Iram Fatima**, Sungyoung Lee and Young-Koo Lee, "**Daily Life Activity Tracking Application for Smart Homes using Android Smartphone**", The 14th International Conference on Advanced Communication Technology (ICACT'12),Phoenix Park,Pyeongchang,Korea, Feb 19-22, 2012

14. Muhammad Hameed Siddiqi, **Iram Fatima**, Young-Koo Lee, Sungyoung Lee, "**Comparison of Error and Enhancement: Effect of Image Interpolation**". In the proceedings of Korea Computer Congress (KCC'11), Kyungju, Korea

15. Muhammad Fahim, Muhammad Hameed Siddiqi, **Iram.Fatima**, Sungyoung.Lee and Young-Koo Lee, "**Daily Life Monitoring Application for Diabetic Patients Using Android Smartphone**", The 35th Conference of Korea Information Processing Society (KIPS), Jeju, Korea. 14 May, 2011.

16. Asad Masood Khattak, Khalid Latif, Zeeshan Pervez, **Iram Fatima**, Sungyoung Lee and Young-Koo Lee, "**Change Tracer: A Protégé Plug-in for Ontology Recovery and Visualization**", The 13th Asia-Pacific Web Conference (APWeb2011), (LNCS Conference), Beijing, China, April 18-20, 2011

17. Irshad Ahmad, Muhammad Hameed Siddiqi**, Iram Fatima**, Sungyoung Lee, Young-Koo Lee, "**Weed Classification Based on Haar Wavelet Transform via k-Nearest Neighbor (k-NN) for Real-Time Automatic Sprayer Control System**". Published in the proceedings of the 5th International Conference on Ubiquitous Information Management and Communication (ACM, ICUIMC'11), Seoul, Korea

18. Asad Masood Khattak, Zeeshan Pervez, **Iram Fatima**, Sungyoung Lee, Young-Koo Lee, “**Towards Efficient Analysis of Activities in Chronic Disease Patients**”, the 7th International Conference on Ubiquitous Healthcare, Jeju, Korea, October 2010.

Patent

이승룡, 이영구, ***Iram Fatima***, "**동적 소셜 네트워크를 위한 다중 목적 보급 모델장치**", 출원인: 경희대학교 산학협력단, 출원번호: 10-2012-0097343, 2012 년 9 월 3 일