Research Statement

Iram Fatima
+8210-8371-9540
iram.fa@gmail.com

Research Interest: The study of human behavior lies under a wide domain of research fields like psychology, sociology, anthropology, and neuroscience, each use different methodologies, scope, and evaluation criteria to understand aspects of human behavior. Computer science offers a complementary perspective on the study of human behavior by focusing physical and cyber activities of humans. My research analyses the human behavior and predict the future actions by using machine learning, data mining and evolutionary algorithms. The abstract view of my research is as follows:

1. To understand the cyber activities, the fundamental purpose of information diffusion within a social network is to differentiate a set of the individuals on the basis of their behavior for information manipulation and propagation. My proposed Multi-Objective Diffusion Model (MODM) allows the modeling of complex and nonlinear phenomena of multiple types of information exchange, and calculates the information worth of each individual from different aspects of information spread such as score, influence and diversity. I designed evolutionary algorithm to achieve the multi-objectives in single diffusion process.
2. Discovery of users' common behaviors and prediction of future actions from past physical activities become an important step towards allowing an environment to provide personalized service. I developed a unified framework for activity recognition-based behavior analysis and action prediction. For this purpose, first kernel fusion method is proposed for accurate activity recognition and then identified the significant sequential behaviors of inhabitants from recognized activities of their daily routines. Moreover, behaviors patterns are further utilized to predict the future actions from past activities.

To pursue this research, considerable challenges have been resolved by using Genetic Algorithm (GA), Support Vector Machine (SVM), Conditional Random Fields (CRF), and sequential pattern mining. Its main theme is utilizing the human activity log in various formats, for creating state of the art algorithms that is independent of environmental factors and domain experts' interventions. As a result, human behavior can be studied from many perspectives and at many scales including his whole lifestyle. Besides the methods and techniques above, I am intimately familiar with all major methods in machine learning and graph mining, including graphical models, graph pattern mining, and social network analysis.

MultiObjective Diffusion Model

The proposed multi-objective diffusion process with Genetic Algorithm (GA) faces the 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 behavior of individuals from different aspects, such as score, influence and diversity. Ideal fitness functions have to better contain intrinsic conflicts, such that the optimal set of individuals could be obtained through the trade-off of multiple evaluation criteria. (3) For 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.

To resolve above issues (1) the information is modeled by dividing into three categories (a) independent information (b) mutually exclusive information (c) competing information. Independent information spreads autonomously, without any constraint, and an individual can hold many independent pieces of information. Unlike independent information, an individual can hold only one piece of information from a set of mutually exclusive information. Similarly 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. (2) 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 in total, whereas the influence is determined by the frequency of the information generated for each individual during the diffusion process. (3) GA with binary string chromosomes to represent the individual's within the network and dynamic one-point crossover has been used to model the interactions between them

Physical Activity based Behavior Analysis and Prediction

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. It can be recognized over the collected sensory data and annotated either at the micro level (i.e., book reading) or macro level (i.e., leisure) from the daily life of inhabitants. The proposed framework consists of two major modules, (1) SVM based kernel fusion for activity recognition: to recognize the daily life activities using decision fusion of four individual SVM kernel functions, where each kernel is designed to learn the performed activities in parallel. (2) Behavioral analysis and action prediction: to identify the sequential behavior patterns and then predict the future actions by utilizing the significant behavior of inhabitants' daily life.

SVM based Kernel Fusion for Activity Recognition: Activity recognition is a multi-class problem so I adopted a one-versus-one approach to recognize the performed activities. A single kernel function is inadequate to perform well for all annotated activities so in order to achieve better accuracy in the classification. I trained multiple kernels (linear, polynomial, Gaussian (RBF) and multi-layer perceptron) and fuse the individual results at decision level with the help of max rule.

Behavior Analysis and Action Prediction: Representing the inhabitants' actions by means of ordered sequence of activities facilitates our understanding of the significant behavior patterns in daily lifestyles. In order to identify the set of actions that frequently occurs together a sequential pattern mining technique is applied. I utilized the repository of activity log. I discovered all sequential patterns with a specified minimum support, where the support of a pattern is the number of data-sequences that contain the pattern. The analysis of frequent user behaviors reveals the significant habits of inhabitants from their daily routines and provides the basis for behavior learning to predict their future actions. For the learning process of action prediction, set of previous actions provide remarkable evidence to identify the meaningful behavior in terms of forthcoming action. I investigated CRF as a learning classifier for predicting the future actions. It is modeled as undirected acyclic graph that allows arbitrary, non-independent relationships among the observation sequences. 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.

Future Research

In future, I am intended to work on challenges that lie ahead to realize more realistic applications of machine learning, data mining and evolutionary techniques towards human based computing. My main emphasis will be on following topics.

1. **Classifier Ensemble Optimization**: It is not possible that one classifier always performs better than all the other classifiers for every possible situation. This observation has motivated towards combining multiple classifiers to take advantage of their complementary performance for high accuracy. Therefore, I am working on a method for activity recognition by optimizing the output of multiple classifiers with Genetic Algorithm (GA). Our proposed method combines the measurement level output of different classifiers for each activity class to make up the ensemble in order to get better accuracy in activity recognition.
2. **Social network analysis based on multi-product marketing**: Human behavior analysis for marketing the multiple products under a network is the area being ignored by research community. It is not always realistic to do market analysis for the selection of efficient seed set based on the propagation of single products. I am envisioned to analyze the different social networks by introducing the marketing of multiple products in order to analyze the individual behave on these products under their preferences and intensions to adopt those products.
3. **Linkage of human behavior within physical and cyber activities:** A life of human being is surrounded by a set of actions including his physical and cyber activities. I am interested to study the collective behavior of human beings. For this purpose, the main idea is to identify the communities within the network of humans based on their physical activity behaviors and then use the information propagation methodology to introduce the single information within all communities and then identify the interesting behavior based on variations of information adaption according to communities' segregation and people preferences.
4. **Semantic interpretation of physical activities based on set of individuals**: Activities are performed by set of individuals in collaborative, parallel, interleave, and sequential manner. The correlation between these activities in important in order to understand the true lifestyle of individuals and their behavior. The current machine learning algorithms process the activities of single individuals and analysis of activities based on a network of individuals is still an open research area. So I intend to design a semantic model of human lifestyle that can accommodate the interaction of single individuals with others in their dealing life dealings.