A

Dissertation On (Major Project-II)

"Friend Recommendation System by using Machine Learning"

Submitted in Partial Fulfillment of the Requirement
For the Award of Degree of

Master of Technology

In

Software Technology

By

Anuj Kumar
University Roll No. 2K14/SWT/506

Under the Esteemed Guidance of

Dr. Ruchika Malhotra
Associate Head & Associate Professor, Discipline of Software Engineering

2014-2017

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING
DELHI TECHNOLOGICAL UNIVERSITY
DELHI – 110042, INDIA

STUDENT UNDERTAKING

Delhi Technological University
(Government of Delhi NCR)
Bawana Road, Delhi- 110042

This is to certify that the thesis entitled **"Friend Recommendation System by using Machine Learning"** done by us for the Major project-II for the achievement of **Master of Technology** Degree in **Software Technology** in the **Department of Computer Science & Engineering**, Delhi Technological University, Delhi is an authentic work carried out by me under the guidance of Dr. Ruchika Malhotra.

Signature:
Student Name
Anuj Kumar
2K14/SWT/506

Above Statement given by Student is Correct.

Project Guide:
Dr. Ruchika Malhotra
Associate Head & Associate Professor,
Discipline of Software Engineering,
Department of Computer Science & Engineering, DTU

ACKNOWLEDGEMENT

I would like to express sincere thanks and respect towards my guide **Dr. Ruchika Malhotra, Associate head & Associate Professor, Discipline of Software Engineering, Department of Computer Science & Engineering, Delhi Technological University Delhi.**

I consider myself very fortunate to get the opportunity for work with her and for the guidance I have received from her, while working on this project. Without her support and timely guidance, the completion of the project would have seemed a far. Special thanks for not only providing me necessary project information but also teaching the proper style and techniques of documentation and presentation.

ANUJ KUMAR
M.Tech (Software Technology)
2K14/SWT/506

ABSTRACT

In this thesis, we have presented a friend recommendation model that is based on user lifestyle, which is generated using daily activities performed by the user.

Existing technology for friend recommendation is based on social networking graph, in which common friend of users decide the friend recommendation score. We believe the recommendation by social graphs is not the befitting real life friend selection for user. Thus, we have presented an approach of for friend recommendation that is based on matching life style and not on social graph.

To implement this semantic-based friend recommendation system, we have taken advantage of sensor-rich smartphones to recognize human activities. After this we have used these human activities to create user life document, which represents user's life style. Likeness between lifestyles of users is used to measure the similarity between users for recommending friends. First, we collected sensor data of user using mobile application and then we performed activity recognition. After finding activities of user we have created a life document of user, from which the algorithm Latent Dirichlet Allocation (LDA) [1] is used to select the life style. We further propose a similar metric algorithm in order to compute the likeness of life styles among users, and measure the impact the life style of users by a friend-matching graph.

When user requests for friends, then our algorithm will return people's list sorted by recommendation score, from which user can choose friends to send request. We have implemented this system for Android based smartphone and its performance has been evaluated on with large-scale experimental data.

Finally, the results manifest that the recommendations properly reflect the preferences of users in choosing friends. This approach exploits gradient boosting algorithm, Auto-regression Model, Signal Magnitude Area (SMA), tilt angle, standard deviation mean & median.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1: Introduction

1.1 PROBLEM STATEMENT

Presently, social networking recommendation of friend is dependent on social graph. The recommendation of friend by using the social graph is not the appropriate recommendation, because as per the research conducted good friendship have similar lifestyle. Hence, by the proposed friend recommendation model, recommendation of friend is dependent on life styles instead of the social graph model. The methodology behind is the implementation for discovery of life styles of users after tracking daily activity of user by using their smartphone wearable's sensor data, and recommendation of friend, having larger likelihood measurement of life between them.

In our daily lives, hundreds of activities form a significant series that define a user's life. In our approach, we used the word 'activity' to point to the actions taken in the time frame of seconds, like "walking", "sitting" and, "typing", meanwhile using word 'daily life style' for actions such as "shopping" or "office work". For example, "daily exercise" life style commonly contains of the "walking", but might also consists the "sitting" or "standing". For model, daily lives design a similarity between daily lives and life documents (Figure 1). Probabilistic topic models treats a document as combination of topics and collection of words for a topic [1]. In the same way, our daily lives documents as a combination of life styles topics and every life style as a collection of activities words.

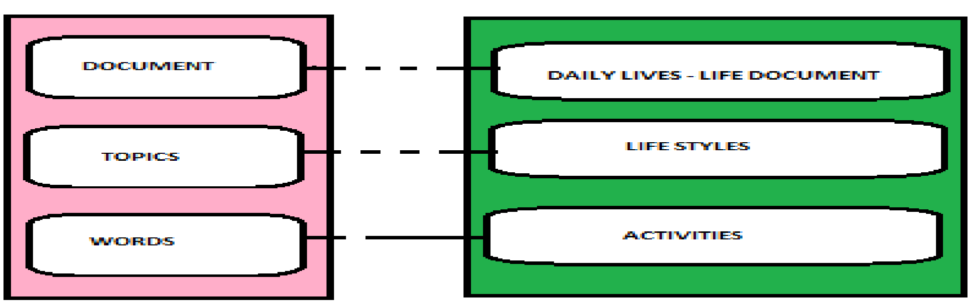

Fig 1: Relation between documents and Person's daily lives

1.2 WHAT ARE THE RECOMMENDATION SYSTEMS

Recommendation system is a type of filtering system, in which we filter the information and give preference information to the user. Nowadays recommendation system is very popular and being utilized in different areas like ecommerce, research articles, social network etc.

Recommendation system produces the result in two different ways. In the first way the preference result is derived from user past information is called Collaboration filtering. Such recommendation way is used to predict the user interest. While in the second way, one the based on the different characteristics of item recommend the additional item with similar characteristic is called content-based filtering.

These two approached have own strength and weaknesses. Nowadays recommendation systems replace the searching algorithm. A hybrid approach, which is defined, combination of these two approaches.

In the first approach we gather large amount of information and analyze the activities and behavior of user and give the preference information to the user. The main advantages of this approaches is that many algorithms are available to analyze and predict the preference information. The popular example of the approach is amazon product recommendation system.

In the second approach which is basal on the properties of the items and person's preference profile. By this approach, we create weight vector of item properties map to user's profile information, where the weight of item properties denote the importance of the feature.

In recommendation system approach, we defined the following phases of its process:-

1. **Information collecting phase**: In this phase, gather the relevant information of users for creating a model.
2. **Analyze collecting information**: In the second phase, analyze/learn the gathered information to define the model for predict the preference information to user.
3. **Prediction:** In the third phase, predict the information which user prefer. The preference information is the filtering information from gathered data for user interest.

4. **Feedback:** In the fourth phase, needed feedback from user to rate for given preference information, On basis of given feedback, we relearn the model for more better use rating.

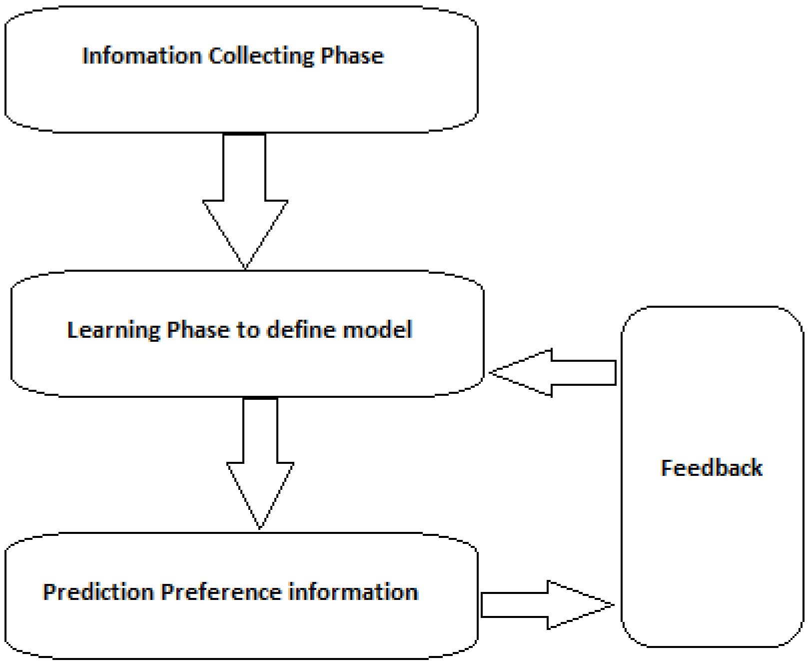

Fig 2: Recommendation System Model

1.3 THESIS MOTIVATION AND GOAL

Mostly Recommender system is designed to filter possible choices for users related to their interests from historical data. This system is widely used in recommendation in the field of news articles, products recommendation on e-commerce, friend recommendation in social network etc. Regarding friend recommendation many recommender system are based on user's friend graph, which is not appropriate in real life scenario.

Nowadays, the friend recommendation in social network is not the appropriate friend recommendation for user, the reason behind is that each user has a different group of friends which vary according to user's interest and behavior. These characteristic is not captured by present way of recommendation.

To overcome this shortcoming, we have been motivated to give a proper solution to recommend appropriate friends on the basis of matching characteristic and behavior of users. Throughout this thesis, we have tried to find out user characteristic and behavior, which help the friend making and use this characteristic to drive an algorithm.

1.4 THESIS ORGANIZATION

This thesis is classified into six different chapters. Chapter 1 deals with the problem statement for the thesis, the problem statement is the recommendation of friend using social graph. Such recommendation is not the appropriate friend recommendation Technique. To overcome this shortcoming, this thesis describes a proposal for friend recommendation on basis of user matching life style.

Chapter 2 is describing the related work information.

Chapter 3 is containing the research details. In our research details, we explain the terms, which are being used in the thesis like "Topic model" ,"cosine similarity" ," SMA"," Tit Angle" etc. The topic model is also inferred to as probabilistic topic model. Here, we have used LDA algorithms as topic modeling algorithm in our approach. The "cosine similarity" is employed to compute the likelihood of two vectors (non-zero). In case, the vectors are the part of inner product space, it helps to computes the cosine of the angle between them. The cosine value for 0° is 1, and in another case it is smaller than 1. Thus it is not magnitude judgment, but just orientation. 'SMA' is total sum of the magnitude of the all three vectors of acceleration. 'Tilt Angle' is the Postural orientation infers to the comparative tilt of the body in space.

Chapter 4 is explaining the proposed approach for friend recommendation based on the matching of user lifestyle, with details architecture. It will also deal with the extraction of life style by topic model, where the Life style document consists of topics and topic consists of words. With Support development in the area of text mining, we have designed a model for user's daily lives as life document. Using the "Document", topics could be found by using this probabilistic topic model. Therefore, we followed this model to find out the probabilities of "life styles" from the "documents". Chapter 4 also deals with the activity recognition, in which Life styles is considered as a combination of motion activities that have distinct occurrence probability. Hence, we used two sensors, gyroscope and accelerometer, that are used to drive motion activities. We adopted one mainstream approach: Gradient Boosting Classifier for activities recognition. Chapter 4 also deals with Friend-matching graph to depicts the likelihood among user life style and how the impact other person.

Chapter 5 illustrates friend recommendation results. Using the representation of gray-scale image of 30 users, low recommendation score color is represented with green color followed with blue color and the white color, representing high recommendation score. The blocks that have green color which represents low recommendation score.

Chapter 6 is the conclusion of the thesis. It describes the benefits of that approach for recommending the friend on the basis of user's behavior and their life style instead of following the social graph. The recommendation of friend by using this approach is more appropriate. As, while surveying, we got potential feedback from user.

Chapter 2: Related Work

Amazon [2] makes suggestions to user on the basis of previously visited items and items that are being looked at. Similarly, movies suggested to user are based on former ratings and observation habits of the user.

A combined filtering friend suggestion design on the basis of personality matching was presented by Bian along with Holtzman [3]. Another friend suggestion method was suggested by Kwon and Kim [4] based on social and physical context. But authors failed to express the details about social and physical context and how the information is obtained.

Yu et al. [5] suggested friends on social network that are geographically linked through combination of social network structure and GPS [Global Position System].

Hsu et al. [6] analyzed link recommendation problem in weblog as well as similar social networks, and put forth a solution on the basis of collaborative recommendation that used a social networks link structure and recommendation based on content using common declared interest.

SFViz, a visual system was suggested by Gou et al [7]. It was for users support to elaborate and search friends under the intended of interest interactively. Gou et al reported a case study following the above system to find out the good opinion of friendship on the basal of people's music community behavior such as tagging.

Activity recognition work as the foundation for high level day to day routines extraction from sensor data that is widely researched through wearable sensors of various types. With the help of GPS data Zheng et al [8] tried to know users' transportation mode. Lester et al. applied to use data of wearable type of sensors to find out activities on the basis of the HMM (Hidden Markov Model) [9]. Li et al. found out dynamic transitions [10] as well as static postures using gyroscopes and accelerometers. Advancement of smartphones to the use of the abundant set of sensors on the smartphones enabled activity recognition. Reddy et al. [23] with the use of the accelerometer and built-in GPS on the smartphones helped to detect individual's transportation mode.

CenceMe [11] took into use many sensors, such as SoundSense, of the smartphone to capture user's state, activities, surroundings and habits. While EasyTracker [12] used GPS traces of smartphones available on transit vehicles in order to find routes served, infer schedules and locate stops, SoundSense uses the microphone in order to understand types of sound (e.g., voice, music) and uncover particular person's sound events.

Though a lot of steps have been taken to recognize activities by the use of smartphones, but there is very less work done on finding out day to day routines using these smartphones.

In this work, tried to use the probabilistic topic model to find living life styles through the smartphone and further use model to discover activities for best friends suggestion that support users to find good friends , who having similar matching of living-life styles.

Chapter 3: Research Background

3.1 Topic Modeling

Insight natural language processing and machine learning, define a topic model is a statistical design type for detecting the detailed topics, which comes during a group of documents. This model is mostly used as a text-mining tool for searching the hidden semantic structures in a text doc. In a given specific topic doc., one could find a specific word set in the doc irregularly. Hence, Topic models help to find an efficient way to analyze chunks of untitled text.

A "Topic" comprises of a group of words that comes repeatedly. By using contextual hints, topic models links words with likewise meanings and segregates words with more than one meaning.

This model is also inferred to as probabilistic topic model, which refers to statistical algorithms for searching the latent semantic structures of a large text body. Topic models can help to structure and offer details for us to understand extensive collections of text bodies which is unstructured. Initially originated as a text-mining tool, these models have been used to find out instructive design in data like genetic information, network, and images.

3.2 Latent Dirichlet Allocation

Latent Dirichlet Allocation (LDA) is a topic modeling algorithm. It is a generative statistical model that permits observation's sets to be presented by unobserved groups, presenting why few parts of the data are analogous. For instance, if a document contains the words gathered by observations, it define that every doc. is a combination of topics with small numbers and that every word's development is credited to one of the document topics. LDA [1] is an example of topic model and was first conferred as a graphical model for topic detection by David Blei, Andrew Ng, and Michael I. Jordan in 2003.

In LDA [1], each document may be seen as a collection of multiple topics, where each document is considered to have assigned topic sets. This is similar to probabilistic latent semantic analysis, with an exception that in LDA, the distribution of topic is presumed to have a sparse Dirichlet prior. The doc which covers only small topic's set, the sparse Dirichlet priors encode such documents.

3.3 Cosine Similarity

It is used to compute the likelihood of two vectors (non-zero). These vector are part of inner product space and it computes the cosine of the angle between them. The cosine value for 0° is 1 or is smaller than 1 in other cases. Thus it is not magnitude judgment, it is just orientation. If two vectors same orientation then cosine will be 1, it vectors are at 90° then value will be 0, and if two vectors opposed to each other, then value will be -1, irrespective of magnitude. The name orients from the word "direction cosine".

For instance, in information gathering and text mining, each word is notionally aligned a different dimension and a doc is defined by a vector where the value of every dimension reflects to the number of times that word appears in the document. Cosine similarity measure of, how likely two documents are to be in terms of their subject information. The technique is also used to compute the cohesion within clusters in the field of data mining.

3.4 Auto Regression Model (AR model)

In this approach, the AR model is used to model the time series signal of different activities [18], [19], [20], [22]. "The AR-model of a random process *y(t)* in discrete time *t* is defined by "as given in equation (1) [22]" below

$$y(t) = \sum_{i=1}^{p} a(i)\, y(t-i) + \varepsilon(t) \quad \text{---------------} \quad (1)$$

Where: $a_1, a_2, \ldots, a_p$ are the co-efficient of the model,

Where p the order of the model and, ε(t) the output uncorrelated error.

The order of an AR model infers to the number of former values of *y(t)* used to calculate the current value of *y(t)*.

The order of AR model can be determined by executing a details analysis of the extent to which current value of signal *y(t)* is dependent on its former values. In this model fix order is 3.

3.5 Signal Magnitude Area

The SMA(Signal Magnitude Area) [13] , [14], [15], [16], [22] is defined as the total of the magnitudes of the acceleration vectors in all three direction. The unit of this measurement is g. where g is the acceleration by gravity. It is directly proportional to the utilized metabolic energy.

"The classification of the engaged activity is defined by setting of threshold value of SMA. "As given in equation (2) [22]" below:-

$$\mathrm{SMA} = \sum_{i=1}^{N} (|x(i)|) + (|y(i)|) + (|z(i)|) \quad \text{----------} \quad (2)$$

3.6 Tilt Angle

Postural orientation infers to the absolute tilt of the body in space. In this approach, the aim is to provide differentiate between the bodily properties of sitting and standing, as well as the various sub-postures associated with lying. Tilt angle (Φ) [13], [14], [17], [16], [22] is said to be an angle between the positive z-axis and g. "As given in equation (3) [22]" below: -

$$\Phi = \arccos(z) \quad \text{----------------} \quad (3)$$

If tilt angle is 0 to 60◦, it is categorized as upright.

3.7 Sensors

A sensor is an electronic component, whose purpose is to detect events or changes in its environment and send the information to other electronic component. Sensors have innumerable applications in wide variety of objects, from a touch sensitive buttons (tactile sensor) to lamps which dark or lighten by touching the basal. The use of Sensors area has been enlarging from fields like pressure, temperature or flow measurement because of enriching the micromachinery. For instance, In MARG (Magnetic, Angular Rate, and Gravity) sensors. Applications of uses sensor in our day-to-day life like cars, smartphone and medicine.

Sensors that are used in this project are ACCELEROMETER and GYROSCOPE. These sensors are already present in android devices. ACCELEROMETER sensor is used to measures the acceleration (m/s$^{2)}$ force all directions including the gravity. GYROSCOPE sensor is used to measures rate of rotation (rad/s) of all directions.

3.8 Activity Recognition

The goal and action of agent are captured through a series of observation by a process of Activity recognition. Given elderly assistance scenario need to be considered for understanding activity recognition better. An elderly man living alone in his house wakes up at dawn. He ignites the stove to prepare tea, turn on the oven, and picks bread and jam from the cupboard. When he is done with morning medication, an automated voice reminds him to switch off the oven. Later in the day, his daughter surfs a website where she scans a list, which was prepared by her father's house sensors network. She gets that her father is eating properly, eating medicine on time, and managing his life on his own. Because of its multiple feature nature, different areas may point to activity recognition, these activity recognition treated as recognition of plan, recognition of goal, recognition on intent, recognition of behavior and location-based services.

There are many kinds of activity recognition:

a) - Single-user, Sensor-based,

b) - Levels of sensor-based

c) - Multi-user ,Sensor-based

d) - Group activity Sensor-based.

Chapter 4: Proposed Approach

4.1 SYSTEM ARCHITECTURE OF FRIEND RECOMMENDATION

The Friend Recommendation system architecture follow client-server model. Where a phone used by a user is a client, while data centers are servers, as shown in the proposed system (Figure2).

On client side, every user's data (Accelerometer and Gyroscope sensor data) is recorded by his smartphone.

In this approach, activity recognition was done by using Gradient Boosting Classifier. To develop a reasonable activity classifier, Data and training phase is required. Three months were spent to build a large training data set by collecting raw data of 30 volunteer.

Collecting the weekly sensor data of 30 people for predefined activities for Gradient Boosting training, we applied it on rest of data for activity recognition. We are able to observe detailed human life-styles by probabilistic-topic model [1], because user continuously accumulates even more activities in life documents.

On the server side, the task of friend recommendation is fulfilled by the design of the Modules. Collection of data form smartphones & creation of life documents of user is done by data collection module. By the help of probabilistic topic model, user's life styles are defined with the help of life style analysis module. After that the indexing module keeps the user's life style in the database with format "life-style, respective user" rather than "user, life style". The friendship-matching graphical-record construction module can accordingly construct friendship matching visual-records to represent the likeness relationship among different user life styles. Then, the user impact ranking module depends on the friend-matching graph to calculate impacts of users. The user query module is used to intake user's request and present a ranked list of best friends as response.

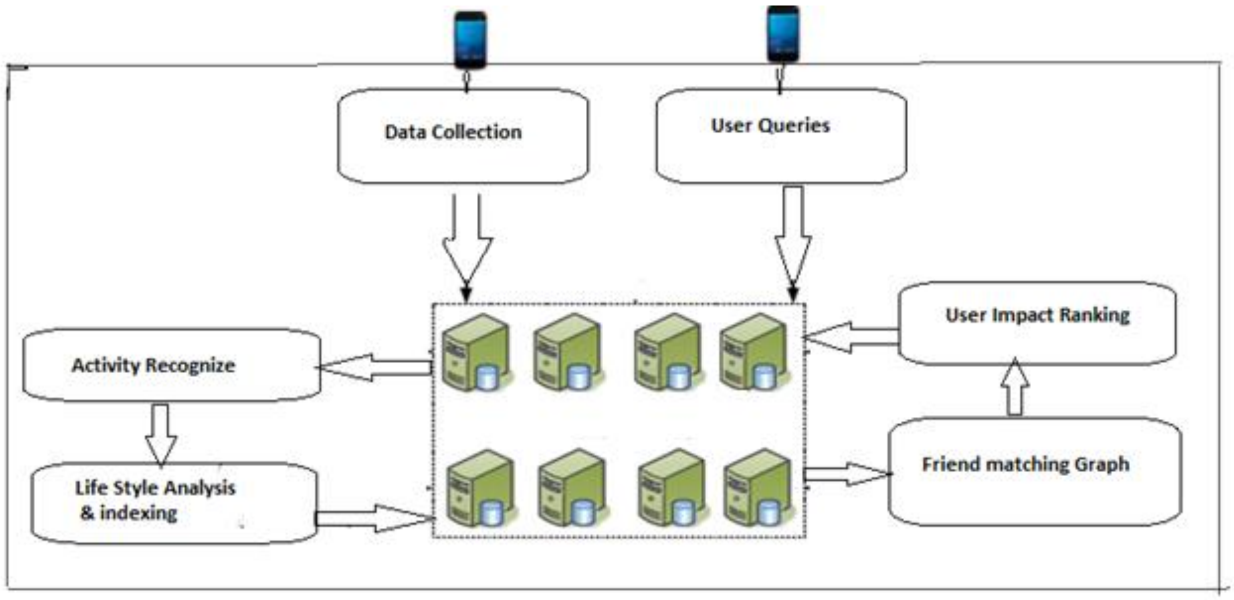

Fig 3: System Architecture of Friend Recommendation

4.2 TOPIC MODEL FOR LIFE STYLE EXTRACTION

4.2.1 LIFE STYLE MODEL

Everyday lives at two distinct levels are bundles of life styles and activities, where life style is a collection of activities. In case of considering everyday lives of person to living-life documents, where the living-life styles to topics and the activities to words is modeled. With the development support in the area of text mining, probabilities of underlying "topics" can be discovered from the "documents" by the probabilistic topic model. Thus, the use of probabilistic topic model to find out the probabilities of "life styles" from the "life documents". As distinct occurrence of words represent their information entropy variances, the repetition of vocab is especially important in probabilistic topic models.

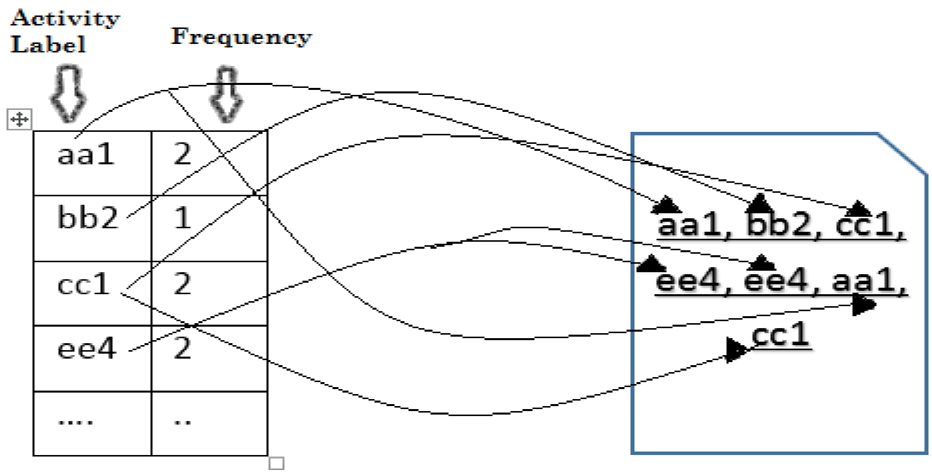

Fig 4: Pouch-of Activity Modeling for Life Document

Following this observation, replacing the initial sequences of activities by raw data along with probability distributions, we present the "bucket-of-activity" model by the figure 3 which describe the bag of activities module for user Life Style document. . Thereafter, comprising a mixture of activity words is user's life document representation.

Consider $w = [w_1; w_2; ::: ; w_W]$ represent a set of activities, in which w_i represent i^{th} activity and W represent the total activities.

Consider $z = [z_1; z_2; ::: ; z_Z]$ represent a life styles set, in which z_i represent i^{th} life style and Z represent total life styles.

Consider $d = [d_1; d_2; ::: ; d_n]$ represent a life documents set, in which d_i represent the i^{th} life document and total users is n.

Consider, $p(w_i|d_k)$ represent the probability measurement of the activity w_i in a particular living-life document d_k, $p(w_i|z_j)$ represent the probability of the degree the activity w_i adds to the life style z_j, and $p(z_j|d_k)$ represent the probability measurement of the living-life style z_j enrolled in the living-life document d_k.

“Based on the probabilistic topic model defined “as given in equation (4) [21]” below

$$p(w_i|d_k) = \sum_{j=1}^{Z} p(w_i|z_j)p(z_j|d_k) \qquad \text{-----------------------} \qquad (4)$$

Observe that “bucket-of-activity” presentation for the life document d_k, “$p(w_i|d_k)$ can be easily calculated “as given in equation (5) [21]” below : -

$$p(w_i|d_k) = \frac{f_k(w_i)}{\sum_{i=1}^{W} f_k(w_i)} \qquad \text{-----------------------} \qquad (5)$$

Where $f_k(w_i)$ denotes the frequency of w_i in d_k.

“The user’s life styles in place the life style vector, represented “as given in equation (6) [21]” below:-

$$\boldsymbol{L_k = [p(z_1|d_k);\ p(z_2|d_k);\ :::;\ p(z_Z|d_k)].} \qquad \text{-----------------------} \qquad (6)$$

To find out the life style vector of every user by using the user’s life documents is the objective. $p(w_i|d_k)$ can be calculated easily, but $p(w_i|z_j)$ and $p(z_j\ |d_k)$ are hard to compute due to the hidden characteristic of life styles.

$p(w_i|d_k)$, calculated by using Activity recognition, and but $p(w_i|z_j)$ and $p(z_j\ |d_k)$ are computed by algorithm LDA [1] decomposition algorithm.

4.2.2 ACTIVITY RECOGNITION

Derivation of $p(w_i|d_k)$ is facilitated by classification or recognition of the user activities. A combination of motion activities, which have distinct occurrences probability that usually reflects Life styles. Therefore, users’ motion activities are inferred using two motion sensors, accelerometer and gyroscope. In this model we have adopted one mainstream approach “Gradient Boosting Classifier” for activities recognition.

The activity recognition flowchart is shown in Figure 4 which describes the steps of recognizing activities of human. The raw data has been collected on the smartphones, and then features are gathered to categories the pre-processed data, thus, further improving recognition accuracy. After testing multiple features like mean, standard deviation, median, Auto regression coefficient, and the combination of them on the data, we found that the standard deviation, Auto regression coefficient [18], [19], [20], [22] , SMA [13] , [14], [15], [16], [22] tilt angle [13], [14], [17], [16], [22] are the most representative feature for characterizing motion activities.

This approach using 39 features including (mean_accX, mean_accY, mean_accZ, mean_gyroX, mean_gyroY, mean_gyroZ, median_accX, median_accY, median_accZ, median_gyroX, median_gyroY, median_gyroZ, std_accX, std_accY, std_accZ, std_gyroX, std_gyroY, std_gyroZ, model_AR_acc_coefficeint (9), model_AR_gyro_coefficeint (9), SMA_acc, SMA_gyro, and tilt angle) for Gradient Boosting Classifier Activity Recognition System. ‘mean_acc_#’: Mean value in all three directions of accelerometer sensor. ‘median_acc_#’: median value in all three directions of accelerometer sensor, ‘sd_acc_#’: standard deviation value in all three directions of accelerometer sensor, similarly take same value for gyroscope sensor data. ‘AR_acc_coefficeint’(9) : Auto regression co-efficient [34] of all three direction with order 3 , we mean AR_acc_XC1, AR_acc_XC2,and AR_acc_XC3 (three coefficient for X direction of sensor data) , similar taken for Y and Z direction .

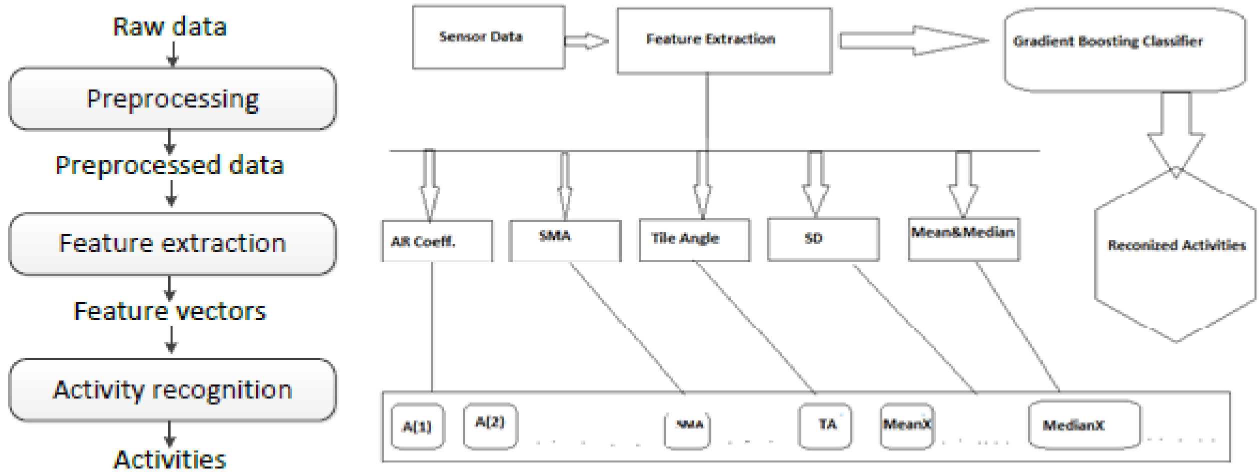

Fig 5: The Flowchart of Activity Recognition

In our experiment including with feature mean, median, Standard deviation. AR Coefficient, SMA [13], [14], [15], [16],[22] , tilt angle [13], [14], [17], [16],]22] we got accuracy of 87% in result of activities recognition. We have used the same Gradient Boosting Classifier for activities recognition and calculated the accuracy by using the library “sklearn.ensemble” and “sklearn.metrics” respectively. Source code as follow: -

```
from sklearn.ensemble import GradientBoostingClassifier
from sklearn.metrics import accuracy_score

/* Gradient Boosting Classifier for activities recognition */
clf = GradientBoostingClassifier(n_estimators=500, learning_rate=0.3, max_depth=1,
random_state=0).fit(X_train, y_train)
    predictions =  clf.predict(X_test)

/* measure Accuracy  score */
 Score  = accuracy_score(y_test, predictions)
```

Fig 6: Source Snippet for activities recognition and compute accuracy

The significance of this accuracy of the activity recognition will impact more to define the analogous of user life style consequence. We can recommend more appropriate friend which will be based on user life style and behavior.

4.2.3 LDA FOR LIFE STYLE EXTRACTION

"Life style documents of users, the Equation-1, can be represented in the form of problem matrix decomposition represented "as given in equation (7) [21]" below: -

$$p(w|d) = p(w|z)\, p(z|d) \quad \text{------------------------} \quad (7)$$

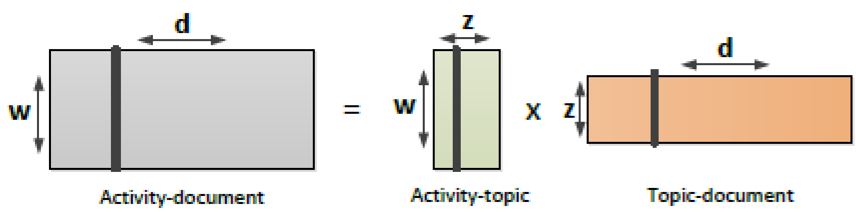

Fig 7: Matrix Decomposition for Life Styles Analysis

Here $p(w|d) = [p(w|d_1);\ p(w|d_2);\ :::;\ p(w|d_n)]$ gives the activity-document matrix (Figure 5), which having the probability measurement of every activity over every living-life document, whereas $p(w|d_k) = [p(w_1|d_k);\ p(w_2|d_k);\ :::;\ p(w_W|d_k)]^T$ gives the kth position of column in the matric of human's action document presenting the probabilities measurement of activities across the life document d_k of user k.

This is an actual LDA [1] model, as represented in above matrix decomposition problem. The decomposition of matrix process can be easily completed more effectively by incremental iteration.

4.3 FRIEND-MATCHING GRAPH AND USER IMPACT

To denote the likeness between user's life styles and how they affect other people in the graph is the Friend-matching graph. The likeness of life styles among two users is denoted using the link weight. A user's affinity to other user can be obtained on the basis of the friend-matching graph. Using this, we can find out how probable it is for this user to be chosen as friend of any other user.

4.3.1 SIMILARITY OF METRIC

A similarity of metric is used to compute likeness among two life style vectors.

Let L_i = $[p(z_1|d_i); p(z_2|d_i);\ :::;\ p(z_Z|d_i)]$ and $L_j = [p(z_1|d_j); p(z_2|d_j); :::; p(z_Z|d_j)]$ represent the life style vectors of user we and user j.

Similarity is impacted by not only user's whole life style vectors, but also the most imperative life styles (value within vector). If most of their life styles are totally distinct, two users do not share much likeness. Hence, the likeness of living-life actions among person i and perosn j, represented by Sim(i; j), is defined as following [21]:

$Sim(i; j) = Sim_c(i;j) * Sim_d(i;j)$

Where $Sim_c(i;j)$ is computing the likeness of the combined life style vectors of users, $Sim_d(i;j)$ is used to emphasize the likeness of users on their dominant life styles. Follow the cosine similarity metric for $Sim_c(i;j)$, which is [21],

$Sim_c(i; j) = \cos(L_i;L_j)$

To calculate $Smin_d(i; j)$, initially define the set of user's possessive life styles.

We define possessive life styles as a set where the probability distribution must be greater than or equal to an already defined value [lambda]

These conditions assure that more probable living life-styles to be included in the set of living-life styles are with larger probabilities. We sorted the life style vector L_i in the decreasing order in respect to the probabilities of life styles to find D_i.

$L_i = [p(z_{i1} \mid d_i); p(z_{i2} \mid d_i);\ \_\ \_\ \_\ ; p(z_{iZ} \mid d_i)]$

“The similarity metric $Sim_d(i; j)$ for computing the likeness of the dominant living life-style sets among two users is denoted “as given in equation (8) [21]” below

$$Sim_d(i,j) = \frac{2|D_i \cap D_j|}{|D_i| + |D_j|} \quad \text{-------------------} \quad (8)$$

The range of $Sim_d(i; j)$ is [0; 1].

As an instance, for showing the calculation of two users’ living-life style likeness, let us presume that there are two person in the system, having the living -life style vectors

L_1 = [0:3; 0:1; 0:2; 0:3; 0:1] and L_2 = [0:2; 0:1; 0:4; 0; 0:3], respectively.

The count of living-life style topics is considered as 5. Initially calculate $Sim_c(1; 2)$ = $\cos(L_1;L_2)$ = 0:6708. Given lambda = 0:8, the possessive living-life style sets of two users are calculated, D_1 = {z1; z_4; z_3} and D_2 = { z_3; z_5; z_1 }. Hence, the possessive living-life style likeness calculated as $Sim_d(1; 2)$ = (2*2) /(3 + 3) = 0:67. At last, the likeness of user 1 and 2 is

Sim (1; 2) = $Sim_c(1; 2)$ * $Sim_d(1; 2)$ = 0:45.

4.3.2 FRIEND-MATCHING GRAPH CONSTRUCTION

Def: Friend-matching visual representation is a undirected weighted graph G = (V;L;W), in which V = {v_1; v_2;… ; v_n} is the users set and n is the count of perosn, L = { l(i; j) } is the links set among users, whereas W : L --> R is edges weights set. There exist an link l(i; j) connecting user i and user j only if the likeness Sim(i; j) >= Sim_{thr}, here Sim_{thr} is the already defined likeness limit.

The likeness, that is, w(i; j) = Sim(i;j), gives the weight of the edge.

Table 1: Friend-Matching Table

	V1	V2	V3	V4	V5	V6	V7
V1		0.87	0.99	0.43	0.67	0.43	0.87
V2	0.87		0.53	0.99	0.67	0.53	
V3	0.99	0.53		0.43			
V4	0.43	0.99	0.43		0.76		
V5	0.67	0.67		0.76		0.99	
V6	0.43	0.53			0.99		0.67
V7	0.87					0.67	

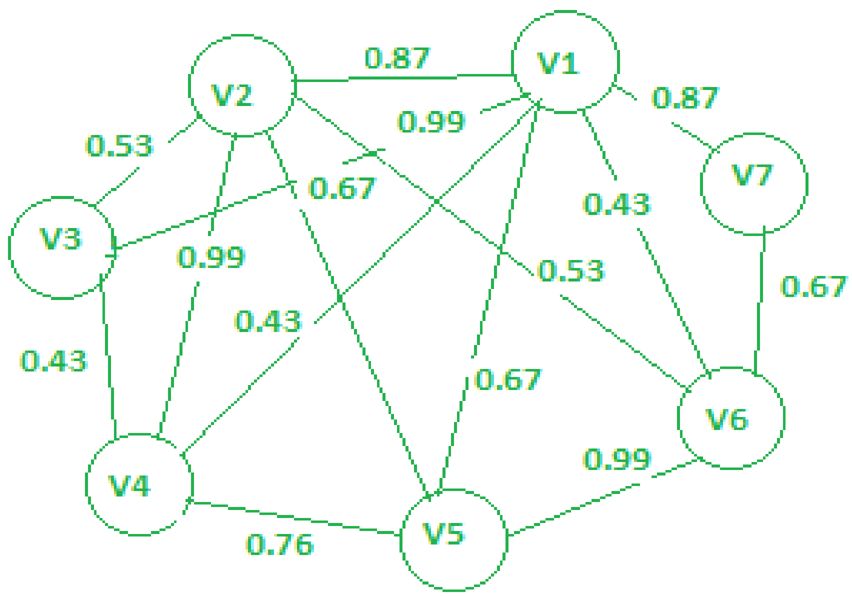

Fig 8: A Representation of Friend-Matching Graph

Above figure represent a relationship of friends with link on basis of their living-life styles of 7 persons. A connecting edge between two users shows the relationship value of living life styles (e.g., e(1; 7) by the edge weight 0.87).

4.3.3 USER IMPACT RANKING

A user's ability to establish friendships in the network is his impact ranking. Once the ranking of a user is obtained, this module gives guidelines to ones who receive the recommendation list on how to select friends. The ranking itself should be independent from the user query. In other meaning, the rank is based only on the graph structure of the friend-matching graph that contains two aspects:

1. How the edges are connected?
2. How much weight is on every edge?

Additionally, as the recommended friends are the ones who not only share sufficient similarity with the query user, but are also popular ones, through whom the query user can increase their own impact rankings.

4.4 QUERY AND FRIEND RECOMMENDATION

The server will do infusion the human living life style vector on receiving a human's request and on its basis will suggest likeness friends for better friendship to the user. The response results are filtered with best matching living life style.

Our recommendation mechanism is to recommend likely friends to a user query. Server measures the suggestion valuation for all the user and descending sorts with respect to their suggestion scores. The higher scores values person will be returned to the user query.

Chapter 5: Results & Analysis

5.1 DATA COLLECTION AND PREPROCESSING DETAILS

We have collected user's real time sensor data for twelve different activities. For collection of these row data, made a android application which is deployed on user's smartphone. Our application records accelerometer and gyroscope sensor data in x, y and z direction mapping with respective activities.

These raw data collected from 30 users for 2 months. Using this raw data we have made a feature vector having 39 features including mean, median, standard deviation, AR coefficients, SMA and Tilt Angle. This feature vector defines user activity.

To define life style of user, we have prepared a user activity document with the help of feature vector. The twelve different activities that we have used in our recommendation system are as follows:

1. Activity of walking
2. Activity of walking downstairs
3. Activity of walking upstairs
4. Activity of sitting
5. Activity of laying
6. Activity of standing
7. Activity of sit to stand
8. Activity of stand to sit
9. Activity of stand to lie
10. Activity of sit to lie
11. Activity of lie to stand
12. Activity of lie to sit

Collected raw data table format for one activity are as follows:

User ID	ACC X	ACC Y	ACC Z
1	1.420	-0.340	-0.125
2	1.002	-0.204	-0.108
3	0.683	-0.061	-0.108
4	0.733	-0.083	-0.120
5	0.956	-0.263	-0.137
6	1.050	-0.402	-0.144
7	1.013	-0.415	-0.104
8	0.950	-0.393	-0.105
9	0.950	-0.359	-0.102
10	0.952	-0.315	-0.086
11	0.913	-0.213	-0.055
12	0.912	-0.125	-0.026
13	0.950	-0.111	-0.063
14	0.969	-0.130	-0.104
15	0.652	-0.075	0.222
16	0.652	-0.075	0.222
17	0.716	-0.055	0.201
18	0.809	-0.141	0.213
19	0.809	-0.141	0.213

Table -2: Accelerometer Sensor of WALKING activity

User ID	GYRO X	GYRO Y	GYRO Z
1	-0.275	1.642	-0.0821
2	-0.675	0.670	-0.083
3	-1.133	-0.391	0.118
4	-1.290	-0.763	0.105
5	-1.204	-0.759	0.034
6	-0.853	-0.632	-0.087
7	-0.566	-0.653	-0.118
8	-0.351	-0.733	-0.091
9	-0.175	-0.485	0.066
10	-0.127	-0.409	0.169
11	-0.166	-0.479	0.314
12	-0.300	-0.574	0.445
13	-0.497	-0.525	0.457
14	-0.724	-0.367	0.425
15	-0.790	-0.201	0.368
16	-0.743	-0.191	0.289
17	-0.700	-0.218	0.278
18	-0.674	-0.252	0.262
19	-0.693	-0.427	0.264

Table-3: Gyroscope Sensor of WALKING activity

As our experiment on given training and test data, first taken only the features Mean, Median and Standard Deviation with 18 features only. With this features we got accuracy of 64% in result of activities recognition. We used Gradient Boosting Classifier for activities recognition.

Below is the list of the feature names (18 features) for gyroscope and accelerometer sensor data in the direction of x y and z.

- **mean_accX:** Mean of X accelerometer sensor's direction data.
- **mean_accY**: Mean of Y accelerometer sensor's direction data.
- **mean_accZ:** Mean of Z accelerometer sensor's direction data.

- **mean_gyroX** : Mean of X gyroscope sensor's direction data.
- **mean_gyroY** : Mean of Y gyroscope sensor's direction data.
- **mean_gyroZ** : Mean of Z gyroscope sensor's direction data.

- **median_accX**: Median of X accelerometer sensor's direction data.
- **median _accY**: Median of Y accelerometer sensor's direction data.
- **median _accY**: Median of Z accelerometer sensor's direction data.

- **median _gyroX:** Median of X gyroscope sensor's direction data.
- **median _gyroY:** Median of Y gyroscope sensor's direction data.
- **mean _gyroZ:** Median of Z gyroscope sensor's direction data.

- **std_accX:** SD(Standard deviation) of X accelerometer sensor's direction data.
- **std_accY:** SD of Y accelerometer sensor's direction data.
- **std_accZ:** SD of Z accelerometer sensor's direction data.

- **std_accX:** SD of X accelerometer sensor's direction data.
- **std_accY:** SD of Y accelerometer sensor's direction data.
- **std_accZ:** SD of Z accelerometer sensor's direction data.

In another experiment, we added a feature AR Coefficient [18], [19], [20] of accelerometer and gyroscope sensor data in the direction of x y and z. With this features, we got accuracy of 83% in result of activities recognition. We used Gradient Boosting Classifier for activities recognition.

Below is the list of AR Coefficient feature for accelerometer and gyroscope sensor data in x y and z directions.

- **AR_AccX_C1:** first AR co-efficient C1 of X accelerometer sensor's direction data.
- **AR_AccX_C2:** first AR co-efficient C2 of X accelerometer sensor's direction data.
- **AR_AccX_C3:** first AR co-efficient C3 of X accelerometer sensor's direction data.

- **AR_AccY_C1:** first AR co-efficient C1 of Y accelerometer sensor's direction data.
- **AR_AccY_C2:** first AR co-efficient C2 of Y accelerometer sensor's direction data.
- **AR_AccY_C3:** first AR co-efficient C3 of Y accelerometer sensor's direction data.

- **AR_AccZ_C1:** first AR co-efficient C1 of Z accelerometer sensor's direction data.
- **AR_AccZ_C2:** first AR co-efficient C2 of Z accelerometer sensor's direction data.
- **AR_AccZ_C3:** first AR co-efficient C3 of Z accelerometer sensor's direction data.

- **AR_ GyroX _C1:** first AR co-efficient C1 of X gyroscope sensor's direction data.
- **AR_ GyroX _C2:** first AR co-efficient C2 of X gyroscope sensor's direction data.
- **AR_ GyroX _C3:** first AR co-efficient C3 of X gyroscope sensor's direction data.

- **AR_ GyroY _C1:** first AR co-efficient C1 of Y gyroscope sensor's direction data.
- **AR_ GyroY _C2:** first AR co-efficient C2 of Y gyroscope sensor's direction data.
- **AR_ GyroY _C3:** first AR co-efficient C3 of Y gyroscope sensor's direction data.
- **AR_ GyroZ _C1:** first AR co-efficient C1 of Z gyroscope sensor's direction data.
- **AR_ GyroZ _C2:** first AR co-efficient C2 of Z gyroscope sensor's direction data.
- **AR_ GyroZ _C3:** first AR co-efficient C3 of Z gyroscope sensor's direction data.

In our last experiment, we added two more features SMA and tile angle got 84% accuracy in result of activities recognition.

Feature vector table having format as follows:

X_mean_acc	Y_mean_acc	Z_mean_acc	X_std_acc	Y_std_acc	Z_std_acc	X_median_acc
1.021	-2.370	-7.280	1.980	1.551	1.421	9.822
9.810	-2.343	-1.648	2.533	1.740	1.362	9.442
1.014	-2.532	-3.826	2.771	2.002	1.584	9.863
1.026	-2.327	-2.713	2.583	2.005	1.742	9.592
9.875	-2.172	-3.622	2.304	1.803	1.401	9.671

Table 4: Feature vector table 1

Z_median_acc	X_mean_gyro	Y_mean_gyro	Z_mean_gyro	X_std_gyro	Y_std_gyro	Z_std_gyro	X_median_gyro
-1.101	-4.989	1.466	2.878	5.354	6.327	2.580	-5.320
-3.892	-8.251	-1.506	3.775	5.132	7.495	3.152	5.731
-4.518	-5.023	2.404	-1.272	5.704	9.011	3.436	2.753
-6.183	-9.143	8.293	-1.591	5.020	8.332	3.553	1.292
-4.586	-3.190	-1.991	-2.020	4.631	7.521	3.458	4.411

Table 5: Feature vector table 2

Z_median_gyro	X_ARCoef1_acc	X_ARCoef2_acc	X_ARCoef3_acc	Y_ARCoef1_acc	Y_ARCoef2_acc	Y_ARCoef3_acc	Z_ARCoef1_acc
3.631	3.511	1.351	-6.911	-6.567	1.511	-7.878	-8.090
5.025	2.453	1.286	-5.276	-8.155	1.295	-6.367	-3.290
3.514	2.765	1.343	-6.148	-8.741	1.397	-7.316	-7.381
3.515	2.232	1.268	-4.802	-7.304	1.338	-6.383	-5.264
4.280	2.429	1.374	-6.139	-5.080	1.311	-5.588	-5.234

Table 6: Feature vector table 3

Z_ARCoef2_acc	Z_ARCoef3_acc	X_ARCoef1_gyro	X_ARCoef2_gyro	X_ARCoef3_gyro	Y_ARCoef1_gyro	Y_ARCoef2_gyro	Y_ARCoef3_gyro
1.241	-3.482	-8.166	1.639	-8.045	8.664	1.443	-6.169
1.196	-4.214	6.182	1.566	-7.582	-1.401	1.181	-4.456
1.222	-4.253	-1.936	1.483	-7.744	1.148	1.166	-5.194
1.260	-4.511	-6.079	1.521	-8.311	2.586	1.209	-5.272
1.031	-1.966	-9.951	1.500	-7.919	3.523	1.232	-4.784

Table 7: Feature vector table 4

Z_ARCoef1_gyro	Z_ARCoef2_gyro	Z_ARCoef3_gyro	sma_acc	sma_gyro	tilt_angle
7.554	1.442	-6.218	7.071	-4.551	1.641
9.291	1.158	-5.032	7.318	2.277	1.593

Table 8: Feature vector table 5

We have used 70% user data as a training dataset and 30% data used to verify model. Accuracy of final model is 84% on testing data.

DATA SUMMARY FOR OUR MODEL:

Total User	Total Activities	Total Features	Total Data Records For Model	Training Data Size (67%)	Testing Data Size (33%)
30	**12**	**39**	**20000**	**13400**	**6600**

Table 9: Data Summary for model

5.2 FRIEND RECOMMENDATION RESULTS

Figure 7 illustrates the image representation of the likeness matrix for 30 users. A block color represents recommendation score between two users. Low recommendation score color is represented with green color, which is followed with blue color and the white color, representing high recommendation score. The blocks in the diagonal is having green color, representing low recommendation score, which we have set manually because it is not a good design if a user is recommended to himself. In figure-9 , User 1 has higher matching living life style with user 10 and user 5 has higher matching living life style with user 4, which is represented with white color, and user 8 and user 4 have no matching in lifestyle with each other, which is represented by dark blue color [Figure 7 : Friend recommendation matching graph].

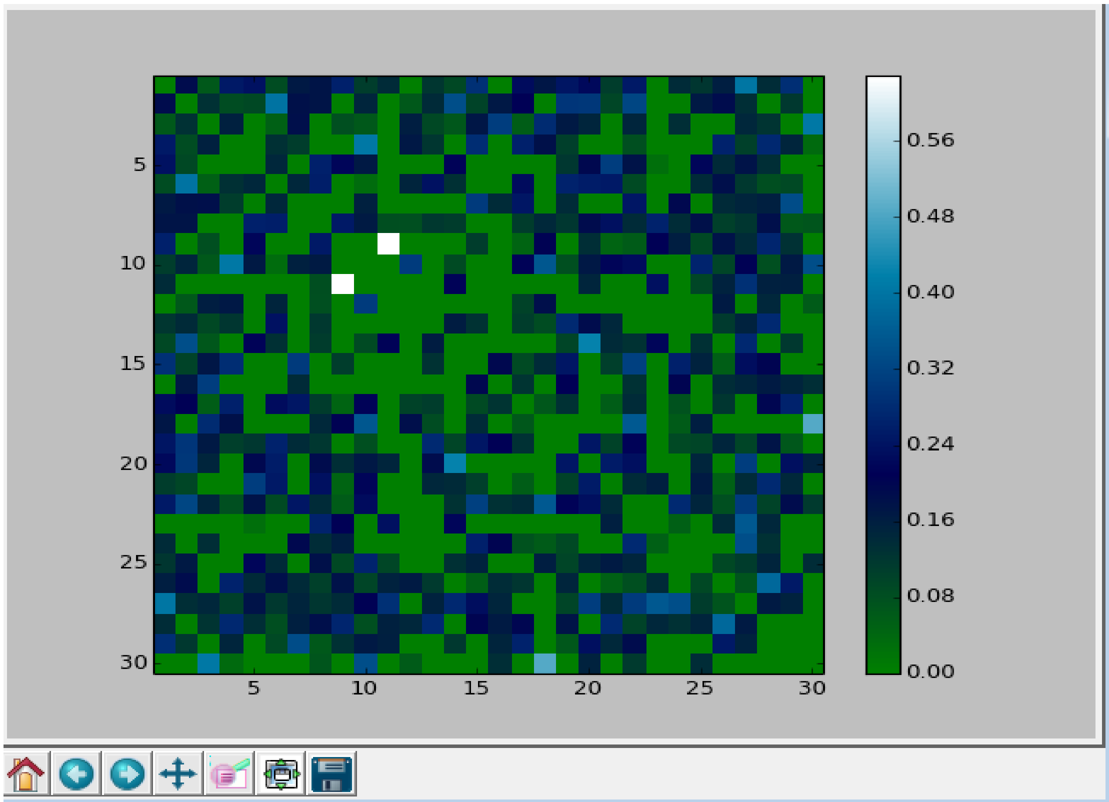

Fig 9: Friend recommendation results

5.3 IMPLICATION OF THESE RECOMMENDATION RESULTS

Currently, the friend recommendation mechanisms are based on the social graph. In a research, in lot of cases user accept the friend request and after some days they remove that person from their friend list. The most of reason collected is difference in their user life style.

In our approach, we have intended to focus on this point. Our approach extracted the life style of user and return the list of members on user query, these users have more similar life style. With our approach we could suggest potential or appropriate friends. Appropriate friend bring the boots and happiness in the life. Appropriate friend always give good advice that reduce the stress, our self-confidence is improved and an appropriate friend always encourages to change or use unhealthy lifestyle habits.

Chapter 6: Conclusion

6.1 CONCLUSIVE RESULTS

In our experiment, we collected 30 user's sensor data mapping with respectively 12 activities as Activity of walking ,Activity of walking downstairs, Activity of walking upstairs, Activity of sitting , Activity of laying , Activity of standing ,Activity of position of sit to position of stand ,Activity of position of stand to position of sit ,Activity of position of stand to position of lie ,Activity of position of sit to position of lie , Activity of position of lie to position of stand and Activity of lie to sit. While considering 39 features in our model we got 84% accuracy in activities recognition. Data Summary for our activities recognition model as following:

Total User	Total Activities	Total Features	Accuracy Measure for model of Activity recognition
30	12	39	84%

Table 10: Data summary for activities recognition for model

We defined the user lifestyle document on the basis of their activities by using LDA model and recommend friend to users which having most similar life style. Life style matching graph of 30 users with 30 users as following:

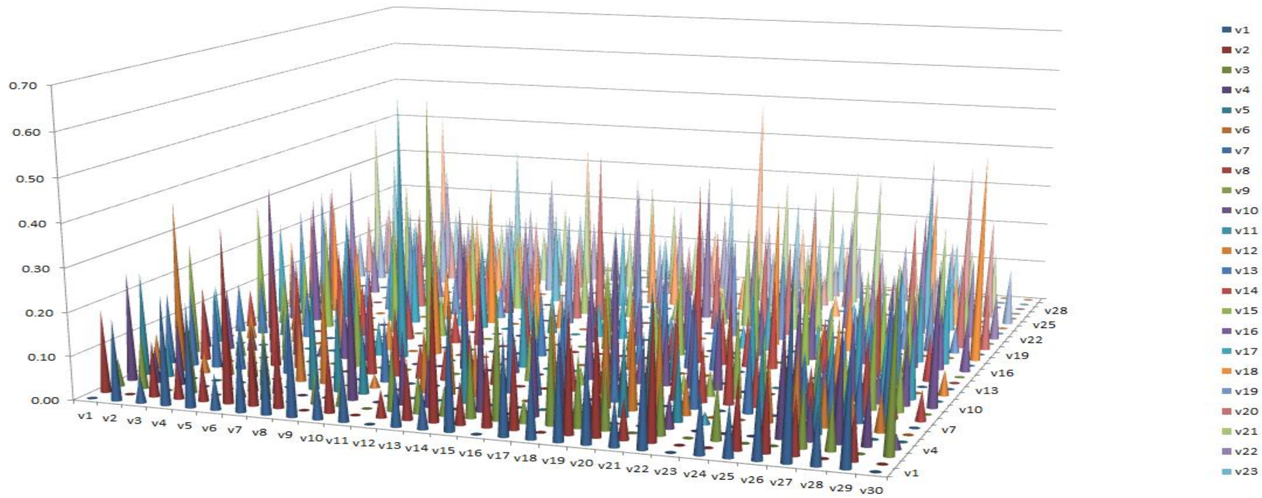

Fig 10: Life style matching of 30 users

According the feedback module we got average (1.6) rated for this model for friend recommendation as following: -

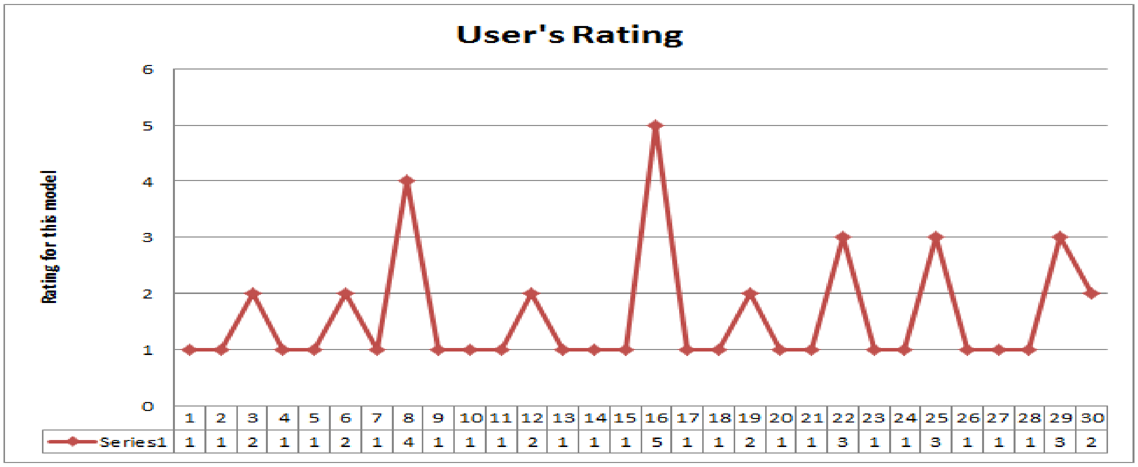

Fig 11: User rating graph recommendation results

6.2 CONCLUSIVE SUMMARY

The proposal that we have presented contain design and implementation of Friend Recommendation System. This is a friend recommendation system based in matching life-style which can be used to recommend friend in social networking applications. This approach is different from the friend recommendation mechanisms, which are currently using by different social networking applications which rely on social graphs in existing social networking technology. Friend Recommendation System extracts life styles from user-centric data collected from sensors on the smartphone and recommended potential friends to users if they share similar life styles.

Implemented Friend Recommendation System can run on the Android-based smartphones, and can learn things on runtime. We have measured its performance on real small series of data. The results show that the recommendations accurately reflect the preferences of users in choosing friends .The future task would be to determine our proposed system on bigger- series field experiments. We would modify the life infusion of user by LDA and apply multiplication of the matric vector method in ranking of user impact. We would be adding more feedback module to evaluate the impact of friendship relations. Also, we would add more devices for discovering more users' activities.

REFERENCES

[1] D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent Dirichlet Allocation. Journal of Machine Learning Research, 3:993-1022, 2003.

[2] Amazon. http:// www.amazon.com .

[3] L. Bian and H. Holtzman. Online friend recommendation through personality matching and collaborative filtering. Proc. of UBI-COMM,pages 230-235, 2011.

[4] J. Kwon and S. Kim. Friend recommendation method using physical and social context. International Journal of Computer Science and Network Security, 10(11):116-120, 2010.

[5] X. Yu, A. Pan, L.-A. Tang, Z. Li, and J. Han. Geo-friends recommendation in gps-based cyber-physical social network. Proc. Of ASONAM, pages 361-368, 2011.

[6] W. H. Hsu, A. King, M. Paradesi, T. Pydimarri, and T. Weninger. Collaborative and structural recommendation of friends using weblog-based social network analysis. Proc. of AAAI Spring symposium Series, 2006.

[7] L. Gou, F. You, J. Guo, L.Wu, and X. L. Zhang. Sfviz: Interestbased. friends exploration and recommendation in social networks. Proc. of VINCI, page 15, 2011.

[8] Y. Zheng, Y. Chen, Q. Li, X. Xie, and W.-Y. Ma. Understanding Transportation Modes Based on GPS Data for Web Applications. ACM Transactions on the Web (TWEB), 4(1):1-36, 2010.

[9] J. Lester, T. Choudhury, N. Kern, G. Borriello, and B. Hannaford. A Hybrid Discriminative/Generative Approach for Modeling Human Activities. Proc. of IJCAI, pages 766-772, 2005.

[10] Li, J. A. Stankovic, M. A. Hanson, A. T. Barth, J. Lach, and G. Zhou. Accurate, Fast Fall Detection Using Gyroscopes and Accelerometer-Derived Posture Information. Proc. of BSN, pages 138-143, 2009.

[11] E. Miluzzo, N. D. Lane, S. B. Eisenman, and A. T. Campbell. Cenceme-Injecting Sensing Presence into Social Networking Applications. Proc. of EuroSSC, pages 1-28, October 2007.

[12] J. Biagioni, T. Gerlich, T. Merrifield, and J. Eriksson. EasyTracker: Automatic Transit Tracking, Mapping, and Arrival Time Prediction Using Smartphones. Proc. of SenSys, pages 68-81, 2011.

[13] M. J. Mathie, A. C. F. Coster, N. H. Lovell, and B. G. Celler, "A pilot study of long term monitoring of human movements in the home using accelerometry," J. Telemed. Telecare, vol. 10, pp. 144–151, 2004.

[14] D. M. Karantonis, M. R. Narayanan, M. Mathie, N. H. Lovell, and B. G. Celler, "Implementation of a real-time human movement classified using a tri-axial accelerometer for ambulatory monitoring," IEEE Trans. Inf. Technol. Biomed., vol. 10, no. 1, pp. 156–67, Jan. 2006.

[15] M. Mathie, B. Celler, N. Lovell, and A. Coster, "Classification of basic daily movements using a triaxial accelerometer," Med. Biol. Eng. Comput., vol. 42, pp. 679–687, 2004.

[16] C.V. Bouten,K. T.Koekkoek, M.Verduin, R.Kodde, and J. D. Janssen, “A triaxial accelerometer and portable data processing unit for the assessment of daily physical activity,” IEEE Trans. Biomed. Eng., vol. 44, no. 3, pp. 136–147, Mar. 1997.
[17] P. H. Veltink, H. B. J. Bussmann, W. de Vries, W. L. J. Martens, and R. C. van Lummel, “Detection of static and dynamic activities using uniaxial accelerometers,” IEEE Trans. Rehabil. Eng., vol. 4, no. 4, pp. 375–385, Dec. 1996.
[18] L. Bao and S. S. Intille, “Activity recognition from user-annotated acceleration data,” in Pervasive Computing, Berlin/Heidelberg, Germany: Springer-Verlag, 2004, pp. 158–175.
[19] N. Ravi,N.Dandekar, P. Mysore, and M. L. Littman, “Activity recognition from accelerometer data,” in Proc. 20th Nat. Conf. Artif. Intell., 2005, pp. 1541–1546.
[20] U. Maurer, A. Smailagic, D. Siewiorek, and M. Deisher, “Activity recognition and monitoring using multiple sensors on different body positions,” and monitoring using multiple sensors on different body positions,” pp. 113–116.
[21] Zhibo Wang, Student Member, IEEE, Jilong Liao, Qing Cao, Member, IEEE, Hairong Qi,Senior Member, IEEE, and Zhi Wang, Member, IEEE, “Friendbook: A Semantic-based Friend Recommendation System for Social Networks” – 2013
[22] A. M. Khan,Y.K. Lee, andT.-S. Kim, “Accelerometer signal-based human activity recognition using augmented autoregressive model coefficients and artificial neural nets,” in Proc. 30th Annu. IEEE Int. Conf. Eng. Med. Biol. Soc., 2008, pp. 5172–5175.
[23] S. Reddy, M. Mun, J. Burke, D. Estrin, M. Hansen, and M. Srivastava. Using Mobile Phones to Determine Transportation Modes. ACM Transactions on Sensor Networks (TOSN), 6(2):13, 2010.