Multimed Tools Appl
DOI 10.1007/s11042-013-1668-5

Smart CDSS: integration of Social Media and Interaction Engine (SMIE) in healthcare for chronic disease patients

Iram Fatima · Sajal Halder · Muhammad Aamir Saleem · Rabia Batool · Muhammad Fahim · Young-Koo Lee · Sungyoung Lee

Abstract Chronic disease may lead to life threatening health complications like heart disease, stroke, and diabetes that diminish the quality of life. CDSS (Clinical Decision Support System) helps physician in effective utilization of patient's clinical information at the time of diagnosis and medication. This paper points out the importance of social media and interaction integration in existing Smart CDSS for chronic diseases. The proposed system monitors health conditions, emotions and interests of patients from patients' tweets, trajectory and email analysis. We extract keywords, concepts and sentiments from patient's tweets data. Trajectory analysis identifies the focused activities after considering imperative location and semantic tags. Email analysis finds interesting patterns and communication trends from daily routine of patient. All these outputs are supplied to Smart CDSS into vMR (virtual Medical Record) format through social media adapter. This helps the health practitioners to understand the behavior and lifestyle of patients for better decision making about treatment. Consequently, patients can get continuous relevant recommendations from Smart CDSS based on their personalized profile. To verify and validate the working of proposed methodology, we have implemented a proof of concept prototype that reflects its complete working with potential outcomes.

Keywords Social media · CDSS · Personalization · Tweet · Trajectory · Email · Healthcare

I. Fatima · S. Halder · M. A. Saleem · R. Batool · M. Fahim · Y.-K. Lee (✉) · S. Lee
Department of Computer Engineering, Kyung Hee University, Yongin-si, Gyeonggi-do, Korea
e-mail: yklee@khu.ac.kr

I. Fatima
e-mail: iram.fatima@oslab.khu.ac.kr

S. Halder
e-mail: sajal@khu.ac.kr

M. A. Saleem
e-mail: aamir@oslab.khu.ac.kr

R. Batool
e-mail: rabia@oslab.khu.ac.kr

M. Fahim
e-mail: fahim@oslab.khu.ac.kr

S. Lee
e-mail: sylee@oslab.khu.ac.kr

Published online: 21 August 2013

1 Introduction

Chronic disease accounts for more than 75 % of healthcare expenditure and nearly an equivalent percentage of disease-related deaths [3]. Its disorders are generally characterized by long duration and slow progression. With enough care and supervision, health condition of these patients can be improved. CDSS (Clinical Decision Support System), based on health information systems, (e.g., EHRs, EMRs, PHRs, and CPOEs) assist physicians in the clinical processes of patient's care [34]. Our lab developed CDSS called Smart CDSS that caters the different aspects of patient's lifestyle in addition to clinical information [17]. Smart CDSS links health observations with contextual knowledge to influence the decisions of clinicians for improved healthcare. The most common applications of a Smart CDSS include alerts and reminders, diagnostic assistance, and prescription support [17, 34].

In the field of healthcare, recently researchers have realized the importance of social media as a potential domain for real time healthcare provisioning [18]. Social media and social interaction empowers users to know more about themselves including their health conditions. For example, research [10] investigated that four out of five users are using internet to find out personalized healthcare information related to the particular disease and its treatments. By knowing more about health, people will be more prepared to manage minefield of modern medical treatment. The challenge lies here is that how social media and interaction can be effectively utilized to manage health related issues.

In this paper, our aim is to improve the patients' health and lifestyle by utilizing his/her social interaction based on different social networks. For instance, after observing a patient's daily social media and interaction activities, our proposed SMIE (Social Media and Interaction Engine) finds some complications with his/her lifestyle like he/she usually sleeps late; does not exercise regularly; does not take medicine on time; eats too much. Obviously, these lifestyles are not good for chronic disease patients [9]. The proposed SMIE is integrated through social media adapter in the Smart CDSS. The social media adaptor is a bridge to connect the SMIE with Smart CDSS. The extracted knowledge form SMIE is converted into standard vMR (virtual Medical Record) [15] format to facilitate the Smart CDSS decision making and recommend changes in unhealthy lifestyles in better way.

In order to achieve above goals, our proposed SMIE includes several novel ideas and our contribution in this work is threefold. Firstly, tweet analysis extracts user interest, health conditions and sentiment from user's tweets. Twitter allows users to post a short text upto 140 characters into one tweet, so due to space limitation people use abbreviations, slangs and URL's. Our proposed approach process this information using natural language processing techniques with machine learning algorithms. As a result, entities and sentiments of user are returned for specific health condition to be used as knowledge for clinicians. Secondly, trajectory in terms of outdoor movement of the patient is tracked using GPS enabled location aware devices, such as smart phones. Usually a patient is prescribed to follow a particular schedule from practitioner based on ailment e.g., it may contain suggestions of daily exercise, avoidance of alcohol, and timely medication. Our proposed approach tracks all the movement related activities of a user and compare them with the prescribed schedule activities. Finally, email analysis investigates the patient's actions to identify significant behavior and communication trends in daily routines. It mines the frequent and periodic communication patterns that change over time to gain knowledge about their habits and preferences.

Therefore, learning about patients' lifestyles becomes an important step towards allowing Smart CDSS to provide personalized services more accurately and effectively. However, to the best of our knowledge, there is not any existing system available which utilizes these social networks with their potential role in decision making for healthcare. Our result shows a proof of concept that has been implemented to reflect the complete working flow of SMIE.

The rest of the paper is organized as follows: We briefly describe related works and their limitations in Section 2. Section 3 presents the relation between proposed SMIE and Smart CDSS. In Section 4, we describe the proposed SMIE based on patient's tweet, trajectory and email analysis. Section 5 illustrates implementation and experimental results. In Section 6 we discuss the significance of work with potential challenges and limitations. Finally the conclusion and future work are drawn in Section 7.

2 Related work

There are many existing work in literature focusing on development of CDSS, starting in 1960 with stand-alone environments. It is evident that CDSS is beneficial to assistant clinicians in diagnosing and therapeutic decisions of patients [34]. The Smart CDSS framework works on top of an SC^3 [16] environment working as a cloud service. The services of the Smart CDSS consist of healthcare recommendations to the user based on his/her activities and interactions. Recently, researches have realized the importance of social media and interaction in the domain of healthcare due to open availability of useful information [18]. Existing CDSS have potential usages in integrated environments but have failed to incorporate the knowledge extracted from social networks. Lacking of social aspects in healthcare decision creates thirst for behavioral knowledge regarding patient's lifestyle. Much research work has been done to analyze the tweets, trajectory, email and other social media resources for different application domains [1, 8, 22, 26–28, 30, 32, 39]. Chen et al. [7] analyzed URL recommendations on Twitter using data stream technique. System working is based on content sources, topic interest models and social voting to design URL recommender. They analyzed user modeling on Twitter for personalized news recommendation and enrich news with tweets to improve the semantic of Twitter activities. Celik et al. [6] studied semantic relationship between entities in Twitter to provide a medium where users can easily access relevant content for what they are interested in. It shows that Twitter is a suitable source as it allows for discovering trending topics with higher accuracy and with lower delay in time than traditional news media. Hassan et al. [12] used a novel approach of adding semantics in Twitter sentiment classification and explored three different approaches for incorporating them into the analysis; with replacement, augmentation, and interpolation. For each extracted entity from tweets, they added its semantic concept that represented more consistent correlation with positive or negative sentiment.

For trajectory analysis, mostly work is done for finding effective and efficient path tracking based on movement patterns. Yang et al. [37] used GPS for finding people preferences regarding attractive areas and movement patterns, which can lead to instructive insight to transport management, urban planning, and Location-Based Services (LBS). Zhu et al. [40] proposed Automatic Identification System (AIS) that uses trajectory mining techniques for finding the ship movement paths. Its purpose is self-navigation and collision avoidance. Braga et al. [5] designed a trajectory based tracking system named 'Captain'. This system is designed for tracking of short, yacht trajectories. The focus of this system is to record the movement path of the person by using the parameters of the pictures, temperature, and coordinates of the locations. In email analysis, the focus of existing work is on investigation of email network to identify importance of individuals on the basis of their communication patterns in network. The communication analysis is also used to analyze the huge amount of data such as e-mail habits [25], mobile phone usage patterns [4], and dominance behavior [13]. Christopher et al. [8]

analyzed the email contents to discover experts on particular topic. They proposed two approaches (a) content based approach consider emails text and (b) graph based approach that consider both text and communication network. They find answers to questions by following people with specific knowledge, skill, or experience. Yingjie et al. [38] examined email data by applying value patterns to cluster a social network. They applied statistical analyses, including hierarchical clustering, overlapping clustering, and correspondence analysis, to identify the value profiles of the employees. Paweł et al. [31] studied the email network to discover the importance of individuals according to their communication capacity. They scrutinized the delays in answering emails. They find implicit ranking about the importance of users and by measuring the procrastination in answering of messages.

The purpose of all this research work is to identifying routine practices and problems within the domain of interest to analyze the interaction network for current understanding and future predictions [20, 21, 30]. The main focus of tweet analysis in existing work is to monitor the publics' feelings towards their brand, and business. In trajectory analysis the emphasis of research is to suggest alternative path based on recommender systems for transport management, urban planning, and location-based Services. Email data is mostly utilized to identify communities or people with similar characteristics within a network. However, to the best of our knowledge none of the existing system deals with the processing of social media contents to extract concepts related to healthcare. In this paper we analyze tweets, trajectory and email data with intensions to extract information that can facilitate the Smart CDSS. We analyze the sentiments, track movement activities of users and identify recurring patterns correspond to recurrent behaviors from daily lifestyle. Furthermore we identify the typical actions and specific interaction information that may affect the patients' health and can be used as knowledge for clinicians.

3 SMIE and Smart CDSS

Smart CDSS is HL7 [14] standard base service that provides guidelines and reminders to chronic disease patients. Over the years, our research group consistently works on Smart CDSS where one of the important inputs mentioned for decision making is social media and social interaction. By monitoring patient's social activities, interests and emotions can be extracted that help the clinicians to provide better guidance to patients. The Fig. 1 depicts the overall architecture of Smart CDSS service with all intended modules.

The detail of Smart CDSS components are not explained here as they are elaborated in [17]. Our SMIE is served as a plug-in module and social media adapter plays a vital role in integrating the output of SMIE with Smart CDSS. Knowledge extracted from user's different social networks becomes input for social media adaptor. It is a cloud based service works as a connector, performs automatic mapping of knowledge extracted from social media into vMR [15]. A vMR for Smart CDSS is a standard data model for representing individual medical records and clinical information inputs and outputs. We used SNOMED [33] medical classification hierarchy to convert personal interests and health conditions into semantic codes that are recognized by Smart CDSS. Social media adaptor performs intelligent mapping of patient's personal data to the coded standard of vMR format as there are many codes of SNOMED exist for one disease under different conditions. So a close attention is paid to identify the tags of SNOMED from results of SMIE before its transformation into vMR. Figure 2 shows example of vMR input for diabetes and depression in dotted boxes. It shows the codes of SNOMED in solid boxes with their respective labels. vMR is further processed by inference engine of Smart CDSS to provide better recommendations to a patient.

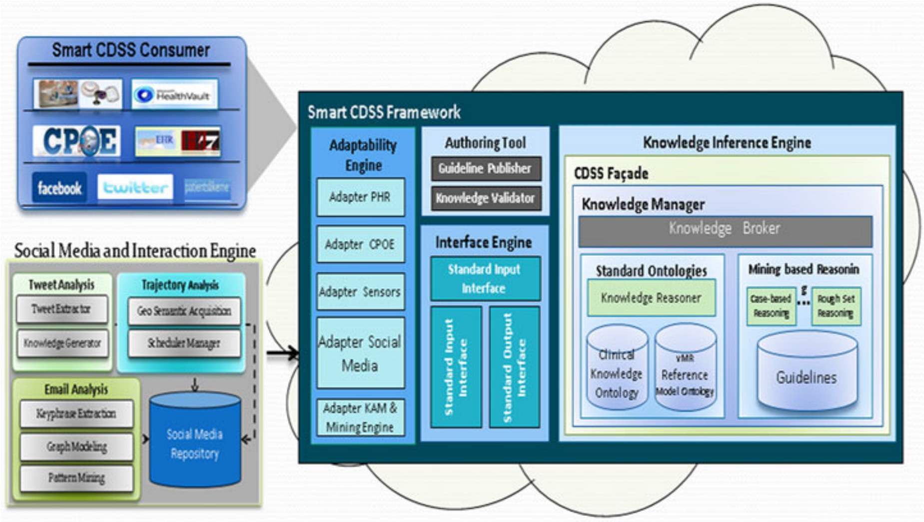

Fig. 1 Smart CDSS architecture

Physicians verify the recommendations provided by Smart CDSS before alerting the patient about newly suggested changes in his lifestyle.

4 Social Media and Interaction Engine (SMIE)

The proposed Social Media and Interaction Engine (SMIE) architecture is composed of three components as shown in Fig. 3. (a) Tweet analysis: to analyze the user tweets in order to generate knowledge about user interest and health conditions. (b) Trajectory analysis: to track the trajectory of daily routine activities for user monitoring. (c) Email analysis: to gain knowledge about preferences, needs and habits of the user from frequent and periodic communication patterns. The details of each component are described in the following subsections.

```
<vmrInput>
    <templateId root="2.16.840.1.113883.3.795.11.1.1"/>
    <patient>
        <clinicalStatements><!--Health problem>
            <problems>
            <problem><!--Depression -->
                <id root="107c2dc0-67a5-11db-bd13-0800200c9a66"/>
                <problemCode System="2.16.840.1.113883.6.96"
                  codeSystemName="SNOMED CT" code ="291746001">
            </problem>
            <problem><!-- Diabetes-->
                <id root="10dca459-3249-29db-bd31-0800200c9a66"/>
                <problemCode System="2.16.840.1.113883.6.96"
                  codeSystemName="SNOMED CT" code =" 73211009"/>
            </problem>
          </problems>
        </clinicalStatement>
    </patient>
</vmrInput>
```

Fig. 2 vMR snippet for diabetes and depression

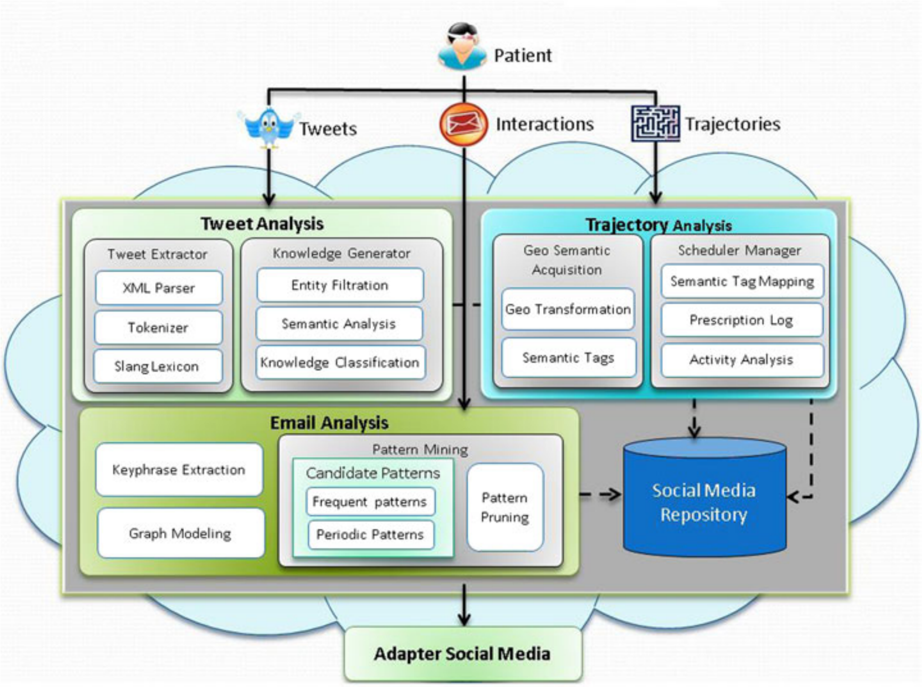

Fig. 3 The proposed architecture of SMIE

4.1 Tweet analysis

Tweet analysis monitors user's social activities, interests and emotional status. It is composed of tweet extractor and knowledge generator modules. The details of each module are as follows:

4.1.1 Tweet extractor

Tweet extractor fetches the tweets from the social network and apply preprocessing step to make it understandable for knowledge generator module. The extracted streams of tweets data are in XML format so we parse it through Document Object Model (DOM). In the body of tweets, user's posts are free text that may have abbreviated words and slangs. This makes the data hard to manipulate and sometime causes noise. We assume all tweets may have this issue that is resolved by passing the data to slang lexicon. Slang lexicon contains text explanation of mostly used abbreviations (e.g., plz, btw, brb, lol) over social media posts and chats. After this step, we process the plain text containing replacement of slangs and abbreviations for generating the knowledge by using natural language processing techniques.

4.1.2 Knowledge generator

Knowledge generator process the plain text through Alchemy API [2] and stores the extracted keywords, and participating entities into social media repository. Alchemy API is a cloud-based text mining platform rich in extracting entities and keyword from plain texts up to 28 major types (e.g., health condition, person, drug). On the basis of entities and keywords, sentimental analysis and tagging is performed to get the contextual information and emotional state from the tweets.

After getting all the required information and analysis of tweet data into the form of concept tags, we classify the tweets into health related and non-health related tweets. Finally, this information is stored into social media repository as an output of tweet analysis and provided to Smart CDSS in the format of vMR whenever needed. The pseudocode for tweet extractor and knowledge generator is depicted in Algorithm 1.

```
Algorithm 1: Tweet Analysis
Input:   UPT – User Posted Tweets
Output:  TE – Tweet Entities
         KW –Keywords of Tweets
         ST – Sentiments of Post
Tweet Extractor
  parseTweet = alchemyAPI.ParseXML(UPT)
  for i=1: length(parseTweet) do
    plainText = slangLexicon(parseTweet(i))
  end
end
Knowledge Generator
  for j=1: length(plainText) do
    TE = entitiesExtract(parseTweet(i))
    KW = keywordExtract(parseTweet(i))
    ST = SentimentsExtract(parseTweet(i))
  end
  for k=1: length(ST) do
    if (concept(ST) = "health")
      store(TE,KW,ST)
    end
  end
end
```

4.2 Trajectory analysis

Trajectory analysis monitors and tracks the user's performed activities through smartphone embedded GPS sensor. It is composed of geo semantics acquisition and scheduler manager modules. The details of each module are as follows.

4.2.1 Geo semantics acquisition

A smartphone-based application is developed to collect the geo points of visiting locations and semantic tagging. We process the static location coordinates (i.e., longitude, latitude) through Google API [11] and acquire the geographic tags. Geographic tagging information is not sufficient to get the true semantics of the visitors to a particular location. For instance, the restaurant may be a workplace for a chef or meal place for customer at the same time. To resolve this issue semantic tags are introduced and users provide the context about the certain location. These geo points and semantic tags are stored in the repository reside over a cloud server thorough a web service method. In routine life the importance of a specific task is measured by the amount of time spend to perform it. We define two parameters, *ATT* (Activity Time Threshold) for lasting duration of performed activity and *DT* (Distance Threshold) as imperative location for providing marginal adjustment of coordinates at specific location. The duration of an activity is set to 30 minutes. According to Ulf et al. [36] the minimum time for performing some significant activity is approximately 30 minutes. For imperative location, *DT* is set to accommodate the relative margin of a particular location according to user lifestyle. In this way, we are able to make it more robust by giving very close coordinate points over a location.

4.2.2 Scheduler manager

After geo semantics acquisition of activities into trajectory repository, the semantic tag mapper checks the tagging of locations after a predefined time interval. We define scheduler time interval span over a one complete day (i.e., 24 h). If mapper found any of semantic tag is missing, the user is asked to provide the semantics of missing tag location. Scheduler manager also contains the complete user profile and able to analyze the activities of the user. It provides recommendation in real time according to the patient's prescription. This prescription is composed of taking medication on right time, doing exercise on right time and recommendation of food over the current location according to his/her dietary plan. The complete pseudo code of data acquisition and scheduler manager is explained in Algorithm 2.

Algorithm 2: Trajectory Analysis

```
Input:   ATT – Activity Time Threshold
         DT – Distance Threshold
Output: [LAT LOG] – Geo locations
         TS – Time Stamp
         ST – Semantic Tag
Geo Semantics Acquisition
  while(true) do
    delay(ATT)
    [tempLat tempLog] = fetch(GPS.location)
    if !(GPS.Find) then
      [LAT LOG] = lastPosition(GPS.location)
    else
      [LAT LOG] = newPosition(GPS.location)
    end
    if isequal([tempLat tempLog], [LAT LOG])
      spentTime += timeInterval
    elseif (spentTime > ATT) && dist([tempLat
      tempLog], [LAT LOG])< DT then
      imperLoc = convertToGeoTags([LAT LOG])
      TS = (time.start, time.end);
      ST = user.input(location.Context);
    end
  end
end
Scheduler Manager
  isMissSemanticTag = alert.user(location.context)
  userProfile = smRepository(ST, Activity)
  recommendation = profile.Build(userProfile)
  alerts.user(recommendation)
end
```

4.3 Email analysis

This module mines users' frequent and periodic communication patterns that change over time. We propose a two-phase strategy to identify the hidden structures shared across different dimensions in dynamic network of emails, such as type of communication, time of communication, and communication intervals. The structural features are extracted from each dimension of the email network and integrate them to find out robust patterns about user's behavior. Email analysis is composed of keyphrase extraction, graph modeling and pattern mining modules. The details of each module are as follows:

4.3.1 Keyphrase extraction

The relevant set of keyphrases is extracted from the subjects of the email by applying KEA++ (Keyphrase Extraction Algorithm) [23]. We set the vocabulary to the SNOMED classification in the SKOS (Simple Knowledge Organization System) format. The remaining parameter settings of KEA++ that affects the results of the algorithm are: *Max. Length of Phrases*: After analyzing the SNOMED classification, we set the value of this parameter to five words. *Min. Length of Phrase*: The minimum phrase length is one word in SNOMED classification (i.e., Diabetes), which is the top level. We set the value of this parameter to two words because setting the value to one word provides many irrelevant keyphrases. *Min. Occurrence*: KEA++ recommends two words for this parameter in short contents.

4.3.2 Graph modeling

This module helps in data modeling and parameter settings before applying the mining technique. It extracts a population of interest from large email communication data by removing noise. The extracted information is modeled in graphs based on user defined communication intervals and extracted keyphrases. In each graph, nodes are the individuals with keyphrases as node label and directed edge represents the communication between them. We keep a set of keyphrases as node label to incorporate all the contents of a particular communication. Graph is modeled as dynamic graph where the time of communication is presented as edge labels. Parameters set the thresholds of frequency and periodicity to identify the patterns of interest. For that, it is necessary to define a demanded minimum level (minimum confidence), so that all those sets of actions that have higher confidence level than the minimum confidence are considered as basic frequent periodic patterns.

4.3.3 Pattern mining

This module identifies the set of actions that frequently and periodically occur together. Frequent patterns are mined by using the FP tree based approach [19] while periodic patterns are mined using PSEMiner [24] with integration optimization. First we identify the frequent patterns and then these patterns are checked for periodicity. After discovering basic frequent patterns, an aspect to be considered is that a pattern cannot be regarded as candidate pattern if it lies in all periods but does not meet frequency criteria. We will discuss the values of these criteria in implementation and result section.

Pattern pruning reflects the common characteristics with some unusual association among candidate patterns. For that, the starting point is to transform the candidate patterns into integrated frequent periodic patterns set to make them useful comprehensively. Briefly explained, the process infers meaningful actions from the email data and then splits the string of actions into periodic sequences based on some frequency support to avoid any redundant information. The complete pseudo code of email analysis is described in Algorithm 3. The patterns of interest after pattern pruning are converted into vMR format that is passed to the Smart CDSS through social media adapter.

```
Algorithm 3: Email Analysis
Input:   ES – Email Subject
         SNOMED – Medical Classification Hierarchy
         EIN – Email Interaction Network
         PT – Parameter Threshold
Output: PI – Patterns of Interest
Email Analysis
  keyphr = KEA++(ES, SNOMED)
  sender = getSender(EIN)
  reciver = getReciver(EIN)
  interactTime = getInteractTime(EIN)
  graph = modelGraph(keyphr, sender, reciver, interactTime)
  while(PSEMiner(!PT)) do
    CandFreqPatt = getFrequentPatterns(graph,PT)
    CandPerodPatt = getPeriodicPatterns(graph,PT)
    PI = patterPruning(CandFreqPatt, CandPerodPatt,PT)
  end
end
```

5 Implementation and results

For prototype implementation of our system, we investigated 6,000 tweets and after filtering 4,000 health related tweets, we processed them by our tweet analysis module. A step wise proposed mechanism is applied on the free text to convert it into the useful information. We implemented tweet analysis in Java SE 7 platform by using Alchemy API. Initially, Tweets are extracted and abbreviated words are removed by passing it through slang lexicon. After it keyword, concepts, entities and tweets classification are done in our proposed knowledge generator module as discussed in Section 4.1.2. The extracted result for a user's scenario is presented in Table 1.

The important and vital sign of user social data is to find his feelings, emotional status and views about particular issues over a certain time period. Collectively it is known as sentiments analysis. For sentiment analysis we normalize user sentiment per day. We took all sentiments of a whole day and took average of them to get overall impression of a particular day. We assign +1 to positive sentiment, -1 to negative and 0 to neutral sentiment. Equation 1 shows normalization of sentiments for 1 day.

$$S_d = \sum_{t=1}^{n} \left(\frac{S_t}{n} \right)$$

Where, S_d is user sentiment for a day, n is total number of tweets and S_t is the number of sentiments. Figure 4 shows the results of user overall sentiment for 1 week. It is obvious

Table 1 Knowledge extracted from tweets

No.	Tweet	Keywords	Concept	Entity	
				Text	Sentiment
1	I feel my high blood pressure is at an unsafe level every time I'm at work. ☹ It's seriously going to give me a depression one of these days	High blood pressure	Hypertension, Orthostatic, hypotension, Blood pressure	Depression	Negative
2	I am Diabetic. Here's how it works. My insulin pump and continuous glucose meter (CGM). Plzzz help me	Glucose meter, Insulin pump, Diabetic	Insulin, Diabetes mellitus, Glucose, Hypoglycemia, Diabetes, Glucose meter	Diabetes	Negative
3	Wide awake, I've got a headache & work in the morning. .	Wide awake	Hypertension Psychology	Headaches	Negative
4	I am healthy and feeling good after having high blood pressure now	Blood pressure	Hypertension, Orthostatic, Hypotension, Blood pressure	High blood pressure	Positive
5	I thought I was in dream and in reality I was in a coma.	Dream	Mind Psychology	Coma	Negative

from Fig. 4 that the first 3 days of the user's overall sentiments are positive; however it is neutral on forth day and negative for 5th and 6th day. On 7th day it is again positive.

Similarly, for trajectory analysis we developed android based smartphone application and deployed over Samsung Nexus S running Android 4.04 to get the trajectory information. The smartphone is GPS receiver enabled and programmable through Android SDK. The application has different interfaces according to the users, recommendation and alerts. User's interfaces provide a way to get the semantic tagging of the certain locations. On the basis of this tagging Smart CDSS respond respective recommendations. Furthermore, practitioners

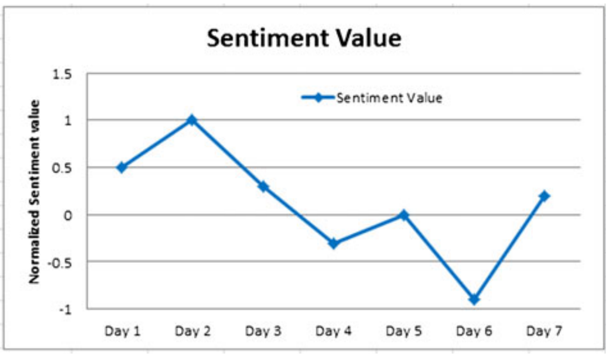

Fig. 4 User sentiments analysis

can visualize the trajectory tracks and recommend some exercise or diet information through the help of Smart CDSS. Figure 5 shows the interfaces of trajectory tracking application.

A user scenario for the results of trajectory analysis is given in Table 2 to get the trajectory information and semantic tagging. A user visit different places and put semantic tags of these locations. On the basis of spent time, we categorized the focused activities and ignore some activities which durations are less than 30 minutes. The details about the duration of time are provided in Section 4.2. In Table 2, geo locations, spent time in minutes, imperative locations, semantic tagging and focused activities are shown. On the basis of

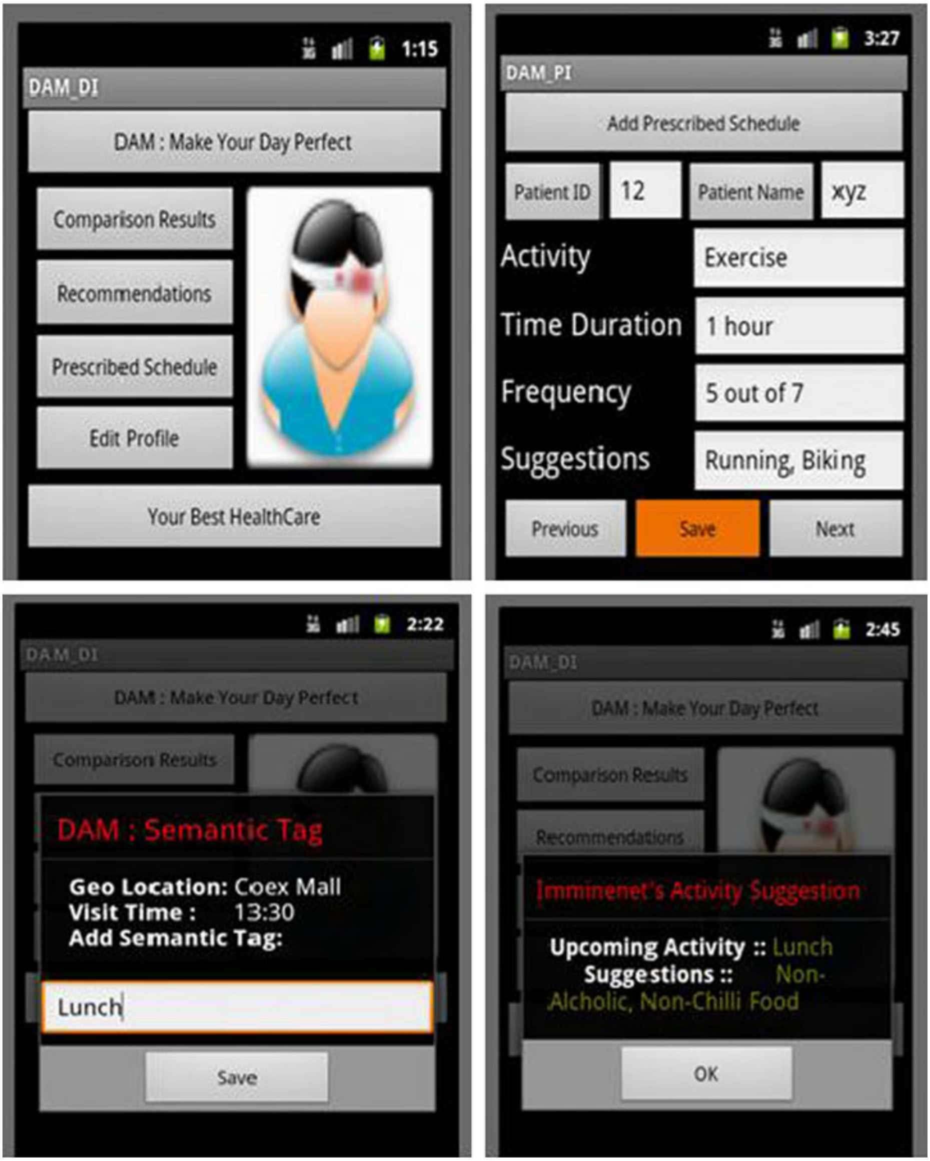

Fig. 5 Trajectory analysis application

Table 2 Spent time, imperative locations, semantic tags and focused activities

No.	GPS coordinates	Spent time	Imperative locations	Semantic tags	Focused activities
1.	37.245084, 127.078889	35	GS-25	Dpt:Store Shopping	–
2.	37.231183, 127.084773	250	Samsung Head Office	Office Work	Office
3.	37.274813, 126.971653	49	Pizza Hut	Lunch	Lunch
4.	37.778501, 127.741562	28	–	–	–
5.	37.245806, 127.072489	400	I-Park Apartments	Home	Home
6.	37.249683, 127.079026	36	GS-25	Shopping	–
7.	37.249674, 127.080722	55	Sports Complex	Exercise	Exercise

imperative locations semantic tags provide the context of user intentions. For instance, imperative location is 'Pizza Hut' that may be a working place or a restaurant for lunch/dinner. Semantic tag differentiates it that is consequently helpful in recommendation. Either this kind of food is suitable for user or what is the frequency of eating pizza as meal in his daily routine. All this information is stored into social media repository.

For email analysis, we consider the email network of user's working environment. In our experiments we divide the communication graph on monthly basis and set the support threshold to 0.8 % in order the get the most significant behaviors of daily life after patterns pruning. The results in terms of extracted patterns are shown in Figs. 6 and 7. The solid edges represent sent email, dotted edges represent email response and edge labels shows the time of communication. In Fig. 6 the user's position in extracted pattern represents him the bridge of communication. He receives the carbon copy of each email communicated with a specific time over repeated days. Although he is not directly involved in the task being discussed in this communication but he is responsible for that in addition to his own work. Consequently, this pattern shows user's busy lifestyle in his working environment which may lead to skipping his meal and exercise routines as identified by the trajectory analysis module.

The behavior pattern shown in Fig. 7 identifies strong ties among user and his coworkers. Consider different time stamps of the day t_1 to t_4 which lies within a particular hour with frequency of daily occurrence for last 10 days. It reflects extensive and unusual communication which may lead to unhealthy lifestyle and results in muscle fatigue and stress.

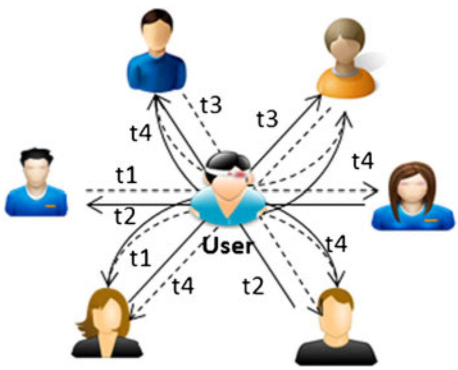

Fig. 6 Communication pattern as bridge in emails

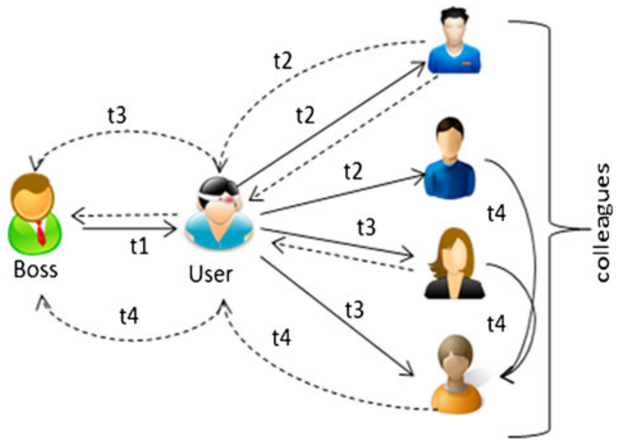

Fig. 7 Extensive communication pattern

One very obvious behavior in response of email has been observed based on weekdays and weekends. It is obvious from the results shown in Fig. 8 that average response time on weekdays is very quick as compared to weekends. Average response time in weekdays lies between 1 and 2 h while in weekends it reaches up to 4 h. Similarly the ratio of email received in weekdays is high as compare to weekends as shown in Fig. 9. In weekdays communication ratio lies among 15 to 30 emails while it drops down to 8 on weekends. Email response and communication rate show less working stress on weekends so Smart CDSS system can recommend the healthy activities (e.g., hiking, walking) for weekend after looking into user's lifestyle, disease and personal preferences.

All extracted knowledge from tweet, trajectory and email analysis modules is stored in social media repository and get aligned with Smart CDSS through social media adaptor in vMR format to make it understandable for clinicians for effective decision making.

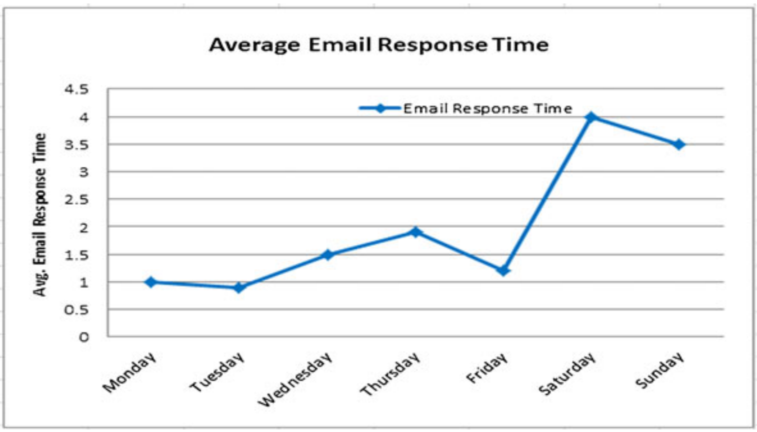

Fig. 8 User email response pattern

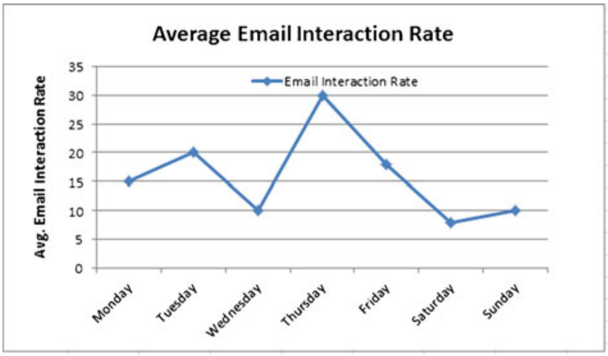

Fig. 9 User email communication rate

6 Discussion on significance, challenges and limitation of the work

With the growing use of social media and social interaction, healthcare is one of the most potential domains to get benefited into a list of applications like medication, clinical observation, behavioral profiling, lifestyle analysis, patient education and personalized care. However, fear of open and unauthenticated content of social media and interaction is an open debate that mostly discouraged its use for healthcare. The major concerns are privacy, security, accuracy, and trustworthiness of information disseminated through different medium of social media. However, besides all these reservations, the importance of information shared through social sources can never be ignored for its reusability in healthcare. The concern of privacy and security arises when integrating the personal information with other system of social sources. The issues related to security and privacy have important consequences for users and can only be resolved by careful content sharing as shared information can be utilized by others to make decisions about users. Research [29] shows the correlation between perceived security and privacy with development of new contents and relations within social network. In recent report [35], Twitter admitted about scanning and importing user's phone contacts into database to identify the behavior of users. Most users are unaware by the fact that Twitter uses this method for new users to search their friends. In real social networks, many people besides friends and acquaintances are interested in the information people post on different health related issues. Therefore, annonymization technique plays a significant role to resolve the issue, patient demographic information and social identities should be annonymized to ensure his security and privacy concerns.

Accuracy and validation of information is another concern for such systems. Social media and interaction sources are sometimes criticized for not following the standards of other types of research. The major concerns are identification of relevant contents about a particular disease and elimination of irrelevant information. Similarly, the knowledge extracted from social media contents are not reliable enough to take any serious conclusions therefore clinicians' validation of information is necessary before reaching any decision. Another big challenge for contents of social media and interaction is interoperability; Smart CDSS accepts data in vMR format while data fetched from social sources are in unstructured format. Standard medical classification (e.g., SNOMED) is required to resolve the issue. The tags of standard medical classification are mapped on the output of SMIE before transformation into vMR. In this paper, we deal with the trustworthiness by keeping the results of SMIE as unpublished knowledge in the social media repository until it is approved by the clinicians before sending any recommendation to patient.

7 Conclusion and future work

Building a healthcare system with effective utilization of social media and interaction sources where the information is coming from diverse modalities is a collaborative and challenging issue. In this paper, we demonstrated SMIE to extract user interests, health conditions, emotions and lifestyle from different social networks. Our proposed SMIE analyzed user's tweet, trajectory and email analysis to identify user's habit and preferences for the domain of healthcare. The output of SMIE is provided to smart CDSS by converting it into vMR format with the help of SNOMED codes. Integration of SMIE with smart CDSS envisioned facilitating the practitioners for effective and accurate decision making. Hence, the impact of the proposed system is to overcome the barrier of adopting the benefits of patient's social media and interaction knowledge for health related issues. In order to show the significance of the proposed SMIE, we demonstrated its complete workflow with potential outcome of each module. This reflects the importance of each social network and its usefulness in decision making process.

In future, we intend to identify the integration of more diverse social network in SMIE and how to align its output for wellness of ordinary people. Furthermore, we will enhance the validity of the proposed system by comparing its usability with existing work in the domain of healthcare.

Acknowledgment This research was supported by the MSIP (Ministry of Science, ICT&Future Planning), Korea, under the ITRC(Information Technology Research Center) support program supervised by the NIPA(National IT Industry Promotion Agency) (NIPA-2013-(H0301-13-2001))

References

1. Abel F, Gao Q, Houben G, Houben K (2011) Analyzing temporal dynamics in twitter profiles for personalized recommendations in the social web. In Proceedings of the ACM WebSci′11, 3rd International Conference on Web Science, Koblenz, Germany, pp. 1–8
2. Alchemy API [Online]: www.alchemyapi.com, Dec. 2012
3. Alwan A (2011) Global Status report on noncommunicable diseases. WHO report
4. Amit AN, Siva G, Gautam D, Dipanjan C, Koustuv D, Sougata M, Anupam J (2006) On the structural properties of massive telecom call graphs: findings and implications. In Proceedings of the 15th ACM international conference on Information and knowledge management, USA, pp. 435–444
5. Braga RB, Martin H (2011) CAPTAIN: A context-aware system based on personal TrAckINg. In Proceedings of the 17th International Conference on Distributed Multimedia Systems, Florence, Italy, pp. 130–133
6. Celik I, Abel F, Houben G (2011) Learning semantic relationships between entities in twitter. In Proceedings of the Web Engineering, pp. 167–181
7. Chen J, Nairn R, Nelson L, Bernstein M, Chi E (2010) Short and tweet: experiments on recommending content from information streams. In Proceedings of the 28th international conference on Human factors in computing systems. ACM, pp. 1185–1194
8. Christopher SC, Maglio PP, Cozzi A, Dom B (2003) Expertise identification using email communications. In Proceedings of the 12th international conference on Information and knowledge management (CIKM ′03). ACM, pp.528–531
9. Fatima I, Fahim M, Lee YK, Lee S (2010) Socially interactive cloud based CDSS for u-life care. In Proceedings of the 5th International Conference on Ubiquitous Information Management and Communication, Korea
10. Fox S. Health topics [Online]: http://pewinternet.org/Reports/2011/HealthTopics.aspx, Dec. 2012
11. Google maps API [Online]: https://developers.google.com/maps/, Dec 2012

12. Hassan S, Yulan H, Harith A (2012) Semantic sentiment analysis of twitter. In Proceedings of the 11th international conference on The Semantic Web—Volume Part I (ISWC′12), Philippe, pp.508–524
13. Hayley H, Yan H, Friedland G, Gatica-Perez D (2011) Estimating dominance in multi-party meetings using speaker diarization. IEEE Trans Audio Speech Lang Process 19(4):847–860
14. HL7 organization home. http://www.hl7.org/, Dec. 2012
15. Huang C, Noirot LA, Heard KM, Reichley RM, Dunagan WC, Bailey TC (2006) Implementation of virtual medical record object model for a standards-based clinical decision support rule engine. In Proceedings of the AMIA Symposium
16. Hung LX, Lee SY, Truc PTH, Vinh LT, Khattak AM, Han MH (2010) Secured WSN-integrated cloud computing for u-life care. In Proceedings of the 7th Annual IEEE CCNC 2010 Conference,Las Vegas, USA, pp. 1–2
17. Hussain M, Ali W K, Afzal M, Lee S (2012) Smart CDSS for smart homes. In Proceedings of the10th International Conference On Smart homes and health Telematics(ICOST 2012). Artimino, Italy, pp. 266–269
18. H. R. Institute, Social media likes healthcare, from marketing to social business [Online]: http://www.pwc.com/us/en/health-industries/publications/health-care-social-media.jhtml, Dec. 2012
19. Jiawei H, Hong C, Dong X, Xifeng Y (2007) Frequent pattern mining: current status and future directions. J Data Min Knowl Discov 15(1):55–86
20. John O, Anis Y, Ole-Christoffer G (2012) An adaptive approach to learning the preferences of users in a social network using weak estimators. J Inf Process Syst 8(2):191–212
21. Joseph KYN (2012) Ubiquitous healthcare: healthcare systems and applications enabled by mobile and wireless technologies. J Converg 3(2):15–20
22. Juyoung K, Hwan-Seung Y (2010) Mining spatio-temporal patterns in trajectory data. J Inf Process Syst 6(4):521–536
23. KEA++ [Online] http://www.nzdl.org/Kea/, Dec 2012
24. Lahiri M, Berger-Wolf, TY (2008) Mining periodic behavior in dynamic social networks. In Proceedings of the 8th IEEE International Conference on Data Mining (ICDM 2008), Pisa, Italy, pp. 373–382
25. Lahiri M, Tanya YB (2010) Periodic subgraph mining in dynamic networks. J Knowl Inf Syst 24(3):467–497
26. Menaouer B, Baghdad A, Nada M (2012) Dynamic knowledge mapping guided by data mining: application on Healthcare. J Inf Process Syst 9(1):1–30
27. Naseem I, Mubarak M, Vangalur A (2013) Publishing and discovering context-dependent services. J Hum Centric Comput Inf Sci 3(1). doi:10.1186/2192-1962-3-1
28. Neil YY, Stephen YFK (2012) An intergrated approach for internet resources mining and searching. J Converg 3(2):37–44
29. Nour MA, Dhanapal DD, Lai FW (2011) Perceived security, privacy, and trust concerns within social networking sites: The role of Information sharing and relationships development in the Malaysian Higher Education Institutions' marketing. In Proceedings of the international conference on Control System, Computing and Engineering (ICCSCE), pp.426–431
30. Pankaj K (2012) Incorporating the social web environment in agile requirements engineering. J Converg 3(3):31–38
31. Pawel L, Mikolaj M (2012) Measuring the importance of users in a social network based on email communication patterns. In Proceedings of the IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining ASONAM, pp. 86–90
32. Roman YS, Qun J (2011) A human-centric integrated approach to web information search and sharing. J Hum Centric Comput Inf Sci 1(2). doi:10.1186/2192-1962-1-2
33. SNOMED [Online]: www.ihtsdo.org/snomed-ct, Dec. 2012
34. Soumeya A, Michel D, Claire R, Philippe B, Eric L (2012) A UMLS based knowledge acquisition tool for rule-based clinical decision support system development. J Am Med Inf Assoc 8(4):351–360
35. Twitter admits peeking at address books, announces privacy improvements http://www.foxnews.com/tech/2012/02/16/twitter-admits-peeking-at-address-books-announces-privacy-improvements/, Dec 2012
36. Ulf B, Schiele B et al. (2009) Daily routine recognition through activity spotting. In Proceedings of the 4th International Symposium on Location and Context Awareness LoCA, pp. 192–206
37. Yan Z, Qingquan L, Qingzhou M (2009) Mining time-dependent attractive areas and movement patterns from taxi trajectory data. In Proceedings of the 17th International Conference on Geoinformatics, pp.1–6
38. Yingjie Z, Kenneth RF, William AW (2010) Automatic text analysis of values in the enron email dataset: clustering a social network using the value patterns of actors. In Proceedings of the 2010 43rd Hawaii International Conference on System Sciences (HICSS ′10). IEEE Computer Society, Washington, DC, USA, pp. 1–10
39. Yishui Z, Qun J (2012) An adaptively emerging mechanism for context-aware service selections regulated by feedback distributions. J Hum Centric Comput Inf Sci 2(15). doi:10.1186/2192-1962-2-15
40. Zhu F (2011) Mining ship spatial trajectory patterns from AIS database for maritime surveillance. In Proceedings of the 2nd IEEE International Conference on Emergency Management and Management Sciences (ICEMMS), Beijing,China, pp. 772–775

Iram Fatima received her BS from the Department of Computer Science, Government College University (GC University) in 2004. She received his MS in Information Technology from the School of Electrical Engineering and Computer Science, National University of Sciences and Technology in 2009. Since March 2010, she has been working on his PhD in the Department of Computer Engineering at Kyung Hee University, Korea. Her research interests include healthcare, social networks, clinical decision support system and activity recognition.

Sajal Halder received his B.Sc in Computer Science & Engineering from Dhaka University, Bangladesh. Now he is a Masters student in Computer Engineering Department, Kyung Hee University, Korea. His research interest are graph mining, social network analysis, trajectory mining and recommendation system in real life applications.

Muhammad Aamir Saleem received his Bachelor's degree in Information Technology from National University of Sciences and Technology, Pakistan in August, 2011. Since September 2011, he has been perusing Master degree in Computer Engineering from Kyung Hee University, South Korea. His research interests are Data Mining, Trajectory Mining and eHealth.

Rabia Batool received her BS degree in Information Technology from National university of Sciences and Technology Pakistan in 2011. She is pursuing her MS in Computer Engineering from the Department of Biomedical Engineering at Kyung Hee University, Korea. Her research interests include natural language processing, tweet analysis, health informatics, and interoperability.

Muhammad Fahim received his B.S. with distinction from Gomal University, Pakistan in 2007. He got M.S. from National University of Computer and Emerging Sciences (NUCES), Pakistan in 2009. Since March 2010, he has been working on his PhD degree at the Department of Computer Engineering at Kyung Hee University, Korea. His research interests include activity recognition in smart homes and smartphone, digital signal processing, Machine Learning and artificial intelligence.

Young-Koo Lee received his B.S.,M.S., and Ph.D. in Computer Science from Korea Advanced Institute of Science and Technology (KAIST), Korea in 1988, 1994 and 2002, respectively. Since 2004, he has been an assistant professor at the Dept. of Computer Engineering, College of Electronics and Information, Kyung Hee University, Korea. From 2002 to 2004, he was a Post Doctoral Fellow Advanced Information Technology Research Center (AITrc), KAIST, Korea, and a Postdoctoral Research Associate at Dept. of Computer Science, University of Illinois at UrbanaChampaign, USA. His research interests are Ubiquitous Data Management, Data Mining, Activity Recognition, Bioinformatics, On-line Analytical Processing, Data Warehousing, Database Systems, Spatial Databases, and Access Methods.

Sungyoung Lee received his Ph.D. degrees in Computer Science from Illinois Institute of Technology (IIT), Chicago, Illinois, USA in 1991. He has been a professor in the Department of Computer Engineering, Kyung Hee University, Korea since 1993. Before joining Kyung Hee University, he was an assistant professor in the Department of Computer Science, Governors State University, Illinois, USA from 1992 to 1993. His current research focuses on Ubiquitous Computing, Cloud Computing, Intelligent Computing, and eHealth.