Behavioral Anomaly

MOTIVATION AND INTERESTS

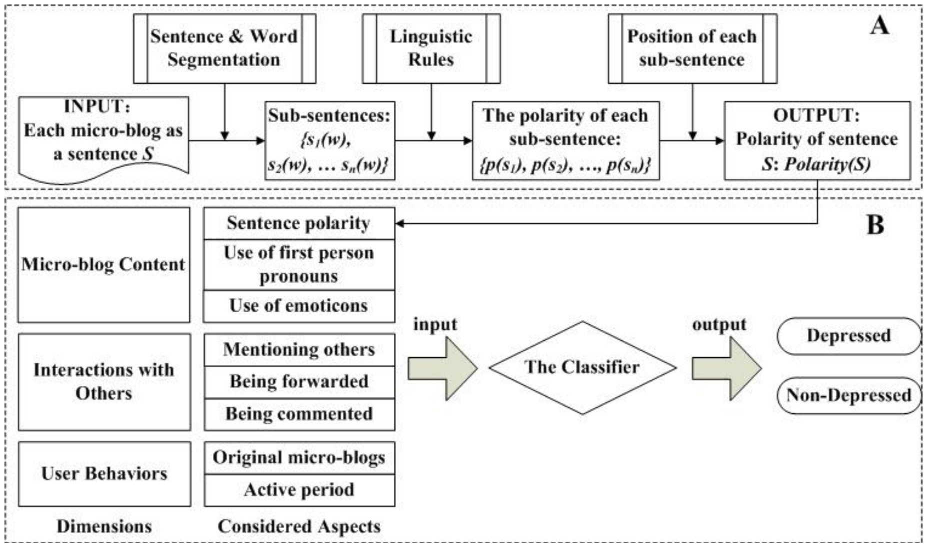
Human mind is one of the most complex systems ever created; add to it, it's not deterministic as well. So even after so much advancement in science and technology, we have not been able to state the condition of a human mind accurately. The branch of pathology has made great advancements and today we can predict and detect diseases with a fair deal of accuracy. But the same cannot be said about the diseases that afflict the brain. Depression is one such affliction, which affects many but most don't know it until its too late. Till now physicians rely greatly on the near and dear ones of the persons, to create a picture of the state of mind of the person concerned. But this method is prone to errors as the statements given by other persons are liable to errors and also there is the problem of insufficiency of data. Hence, we would take the approach of creating objective-self to help determine if the person concerned is moving into to a state of depression.

NOVELTY

Over the past years there have been many research conducted which have attempted to establish a relation between financial factors and depression symptoms of a person [5][6]. Recently with the rise of social media researchers have started looking into usage of applications like Facebook and Twitter to determine if they can be used as a tool to predict depression. There have been research in establishing correlation between Twitter usage patterns and user depression [7]. Although the analysis of these social media apps do uncover an interesting correlation between user behavior on the application and their mental state; these insights are not sufficiently reliable and accurate. There is significant scope to use variety of data collected by the ubiquitous smartphones and attempt to generate insights from it. Coupling the data collected from smartphones with the ones received from the social media we have attempted to make the process of depression detection more accurate.

CURRENT STATE OF RESEARCH

The current state of depression detection system consists of many monolithic approaches.We start our discussion with the oldest tool used by the practitioner; the questionnaire. The community of physician have long pondered over the idea of developing a good pathological test for detecting depression. In the earlier times there was a lack of electrochemical tests for any disease what so ever. So the earliest known physicians used to detect the disease by its symptoms. They correlated the symptoms with their earlier known experience and derived at the result. With the advancement of science, especially the field of pathology, a lot many novel tests were derived which could give conclusive results with low false positives and low false negatives.



But the above method had their own limitations. The above tests had their origin in chemistry and thus depended on agents. The problems of the nervous system are not so easily diagnosed by bio-chemical testing. Hence there was not a significant break through. The major change came in this field with the invention of electroencephalogram (EEG). This analyzed the electrical signals in the brain and gave good data for the physician to diagnose. But still the primary method of diagnosis remained the questionnaire. The physicians over time and using accumulated experience have developed a encyclopedia of questionnaire which can give them insight into the current state of affairs going in the brain. Based on the feedback, the physicians create a mental model of the map.Then they corroborate the facts with the EEG data and come to any conclusion.

The primary take away from the above system is the model creation. We will use this key idea but our approach will be different. The questionnaire paradigm is good but there is a need of a system which can detect the onset of depression early. Social media gives us a good hope.

Social Media as a Measurement tool

Social media has taken the world by storm. The level of penetration of social media is amazing. Even if we look beyond the first wold countries, most of the people have access to mobile phones and thus access to social media. The two most important player in this area are Twitter and Facebook. Both allow us to gauge the state of mind of the person behind the words. The posts in Facebook and the tweets in Twitter are a good indicator as to what is the mental state of the person. Existing system do a good job in sentiment analysis of the text. The results are pretty good at least when English text is concerned. We will discuss the rational behind the sentiment analysis next.

Many studies have been done and some of the common tell-tale sign of depression have been noted. A very common signal is the higher propensity to use first-person singular pronoun rather than first-person plural pronoun.[10] The accepted psychology behind it is that depressed persons are generally lonely and so cannot identify themselves with any group. A statistics can be taken over this and we can get a probability count can be evaluated. It is observed that depressed person also suffer from insomnia. This can be taken into account by noting the time of the posts. If the person is all of a sudden posting a lot of tweets during the time from say mid-night to morning, then we can speculate that the person is having some sleep-disorder and there is a high likelihood that the person might be suffering form depression as well.

The words in english can be broadly divided into three category, for the problem in our domain takin into account the emotion associated with the word. We would like to note the words which have positive, negative, neutral emotions. Taking the emotions of the words into account, we get an overall polarity of the text,(post or a tweet).

The engine to derive the polarity is as follows:

The main challenge behind this engine is how to classify the word into positive, negative, neutral. Fortunately we have Wordnet and Sentiwordnet which provide us the words and their corresponding emotion.

The above system was initially designed to use normal text. But social media has it own lingo. For example the ‘smilence’ is not an actual English word. But in the cyberworld it means smile silently. Then we have to take into account degree modifiers. They should not be removed as stop words. We can associate a value to signify the strength of modifiers. We can come up with a table like one shown below:

<table>
<tr><td>Item</td><td>Example</td></tr>
<tr><td rowspan="2">Emotion Words:
Positive
Negative</td><td>pretty, love, happy</td></tr>
<tr><td>ugly, hate, sad</td></tr>
<tr><td>Degree Modifiers</td><td>most(2), over(1.75),-ish(0.75),insucient(0.5)</td></tr>
</table>

The above model has been implemented by many systems to detect sentiment. But it has some short-comings which we discuss below.

The system does not take into account the cyber family. We can represent each individual node as a node and friends as edges. This will give us a social graph. The count of outgoing edges can be give us a good measure the social status of the person concerned. Hence our system takes this aspect of social network into account.

Till now the existing model, act independently. For example, the tweet count may have reduced. But this can be due to the person being involved in some more urgent work. So we need to get a more holistic approach to identify the mental health. Hence our system takes care of a lot of factors in addition to sentiment analysis.

Mobiles can capture a lot of data like the location of person, the meeting scheduled, the amount of activity like walking, running done by the person. This data is used in creating a life log of the person . We will use this model,life-log, and map this with the sentiment determined by the social media system and give a more well-rounded result.

A recent research [7] conducted at Microsoft explores ways to use the social media postings to gain insight into one's mental state. The research was performed on a large corpus of tweets by people who were diagnosed with clinical depression. The probabilistic model trained on this data could potentially complement the traditional survey techniques to accurately predict depression.

The underlying idea behind such an approach is exploiting the current trend in human behavior where people share their emotions in social media postings. The postings collected from clinically depressed people were used to define features like emotion, time, linguistic style, user engagement (a measure that signifies how engaged the user is with the social platform) etc. Once the features were identified, the researchers then created a classification framework in order to decide if a given post is depression-indicative or not. Based on the results generated from this model, they come up with a social media depression index (SMDI) which could be used to measure the depression based on their tweets.

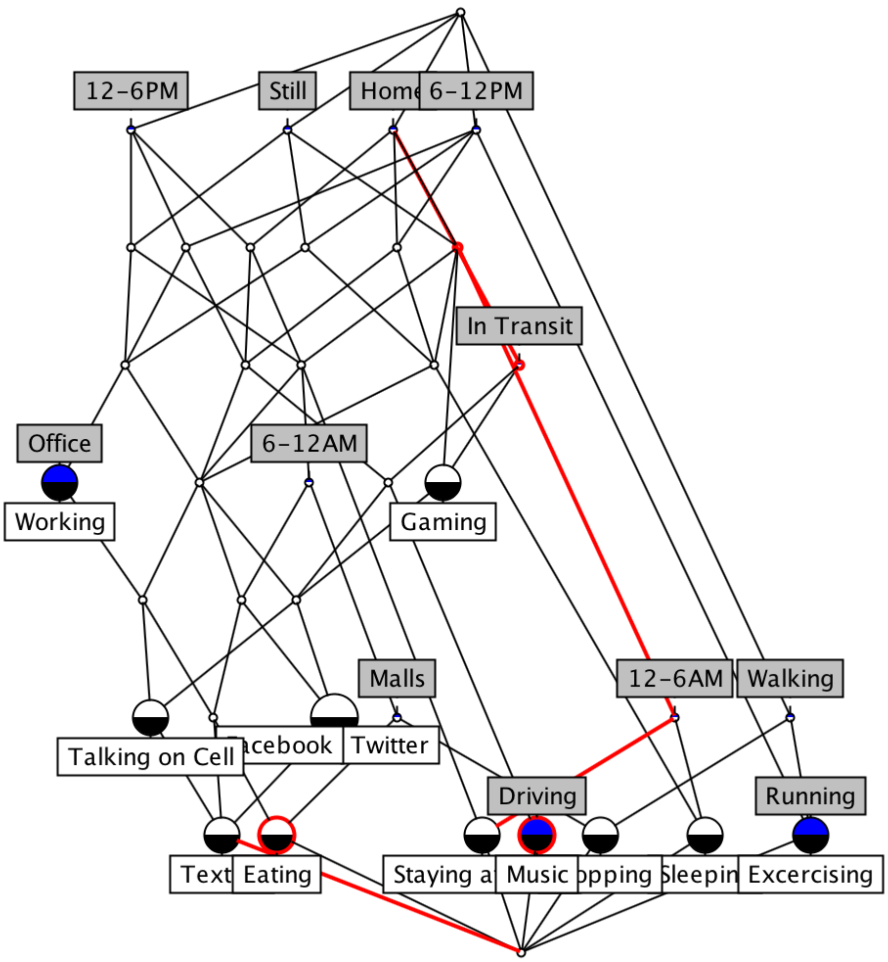
This research sheds light on the potential of social postings as means of getting insights about a person. By gathering a sufficiently large data set and defining appropriate features on the data, a model can be constructed to predict depression.

Formal Concept Analysis:

There is a limit to the amount and types of data that we can obtain pertaining to life logs of a person. In order to predict certain events that might occur we base our model on formal concept analysis.

As given by Belohlavek in [8], Formal Concept Analysis is the analysis of data which describes the relationship between particular set of objects and a particular set of attributes. FCA produces two types of outputs, a concept lattice and a collection of attribute implications. A concept lattice is a collection of formal concepts in the data that are hierarchically ordered by a subconcept-superconcept relation. An attribute-implication describes a particular dependency that is valid on the data.

We use the concept pair of {Life Events, Life Logs} as given in Oh[2] to construct a cross table of Life Events like Working, Exercising, Eating, Sleeping etc., and Life Logs that are made up of a stream of Life Logs containing Spatial data, Temporal data, Sensory data, Activity Level etc. These life log data sources describe the environment around the user and thereby helps us understand the context which is used to organize different Life Events.



	Home	Office	In Transit	Malls	Still	Walking	Running	Driving	12-6AM	6-12PM	12-6PM	6-12AM
Sleeping	X				X				X	X		
Staying at home	X				X				X		X	
Working		X			X					X	X	
Eating	X	X		X	X					X	X	X
Gaming	X		X		X							
Texting	X	X	X		X					X	X	X
Facebook	X		X		X					X	X	X
Twitter	X		X		X					X	X	X
Shopping	X			X	X	X					X	X
Excercising						X	X			X		
Music	X		X					X		X	X	
Talking on Cell	X	X	X		X					X	X	

We use these concepts to capture the what, where, why, how context of that person, using this we can recognize situations that the person is in and also any changes in his behavior can be found out easily. Binary relations between the life events and life logs must be found to use FCA to detect different life events. There is a need to discretize the life logs that have continuous values as the cross table requires binary values. By using the breakpoints mentioned in the looks table in Oh [2]. We can determine binary values for these breakpoints by using Piecewise Aggregate Approximation(PAA) [2].

The result of the cross table of objects and attributes is a concept lattice, which will be traversed at certain points in order to identify what is the behavior of the user. Oh in [2] shows that FCA is a vastly better in terms of accuracy than compared to other supervised learning algorithms when trying to determine the user's life event.

We use the Life Logging app created by Oh in [2] to detect life events of the user and to determine if there are any variations in the users behavior, any small variation could be included in a report to physicians to validate whether there are any red flags that indicate deterioration in the mental health of the user. In our project we propose on using data from different social media apps, media players present on the phone, calendar events etc., all in order to gain as much relevant data of the user's behavior.

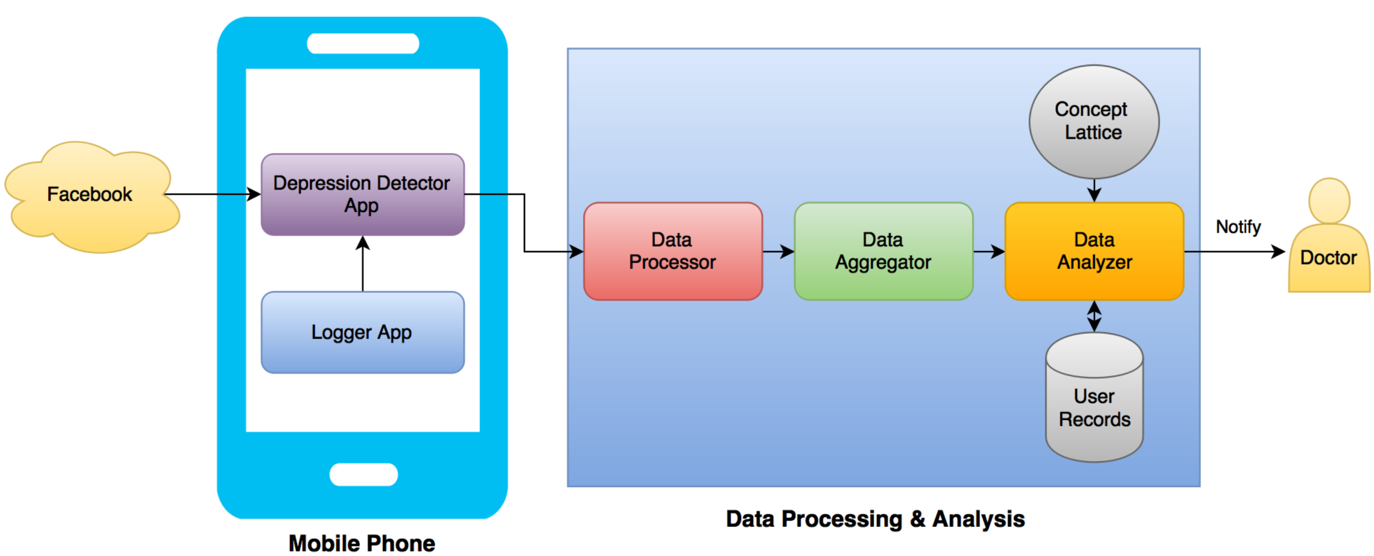
Technology

We used the following technologies in the development of this project:

- Colibri: Formal concept lattice java library used to generate concept lattice.
- json-simple: Library for parsing JSON content.

System Architecture

• **Data Processor** - Processes the data sent by the app and extracts the relevant information (attributes) from it.
• **Data Aggregator** - This component helps aggregate the 5-min interval data into larger chunks which can be easily analyzed by the Data Analyzer.
• **Data Analyzer** - This component takes in the attributes and feeds to the concept lattice to find the relevant events. The events determined for the user are stored in the user records database.



A sample of the JSON data snippet is given below:


```
[
  {
    "time_band": "2",
    "context": "time",
    "timeWindow": "20151121_107",
    "week": 1,
    "longTime": "1448124617"
  },
  {
    "context": "activity",
    "activityType": "{standing still=1}",
    "activityLevel": "0.0"
  },
  {
    "context": "location",
    "longitude": "-117.8305316",
     "venueType": "[[1013, 34], [1013, 34], [94, 1013,
34]]",
   "latitude": "33.6427187",
    "venueName": "[Palo Verde Housing, University of
California Irvine]"
  },
  {
    "context": "media",
    "playtime": "0"
  },
  {
    "context": "application",
    "app_count": 1,
    "app_name": "[off]",
    "app_duration": "[300]"
  },
  {
    "context": "soundSetting",
    "sound": "SILENT"
  },
  {
    "context": "photo",
    "photo_count": "0"
  },
  {
    "context": "phoneStatus",
    "phone_oncount": "0",
    "phone_offcount": "0"
  },
  {
    "context": "light",
    "lightValue": "9.005901639344263"
  },
  {
    "context": "step",
    "stepCount": "0"
  },
  {
    "context": "calendar",
    "location": "UNKNOWN",
    "event": "UNKNOWN"
  },
  {
    "context": "transition",
    "change": "false"
  },
  {
    "context": "call",
    "callEndTime": "[]",
    "callStartTime": "[]",
    "callCount": 0,
    "callContactExist": "[]",
    "callType": "[]"
  },
  {
    "context": "lifeevent",
    "event": "Meal"
  }
]
```

BACK END

The backend of the application receives the life-log of the user in a JSON format with the different attributes that are already created using the concept lattice. The data first is processed and outputs attributes. These attributes are analyzed using the analyzer provided in the Colibri library and persisted in a JSON file, if any continuous deviation from the norm is noticed, we send a notification through either email to his doctor. The email is sent using the JavaMail API and is configured with the email address of his doctor. The JSON based life-log is sent from the front-end and will consists of the entire days lid-logs that have been collected for the user. This JSON is parsed in the backend using a JSON parser (JSON.simple library) in the Data Processor which takes into account the various attributes related to a particular event that is logged and can determine which event in it would fit in in the concept lattice.

The Data Analyzer then reads this processed JSON file which is aggregated into an hourly log of events using the Data Aggregator. The analyzer determines the exact object or event that takes place throughout the day by running it against the concept lattice, this gives us a log of activities throughout the day for the user. The log of this current day is compared with a baseline set of events that is precomputed by averaging out events/activities that have occurred for a period of 15 days and have been approved by a doctor to have the least amount of noise. Any deviation from this baseline would trigger a logger which would log the deviations, if these deviations are frequently occurring then we raise a flag which would send email notifications to the doctor. The doctor can then determine whether this person is showing signs of depression or not.

SUMMARY

We have created a model that can be used in order to determine the different aspects of a person's mental health by processing and analyzing a person's life-log. The life-log that is created is part made up of the life-log from the app created by Oh in [2] and our own idea of using social media feeds to gain interesting insight into the mental state of that person. We use this combination of life-logs and formal concept analysis to identify deviations in the behavior of the person to be able to notify the proper experts in the field of mental health about certain symptoms of people who are depressed.