www.ijsrnsc.org

Available online at www.ijsrnsc.org

IJSRNSC

Volume-1, Issue-3, Aug- 2013
Research Paper

Int. J. Sci. Res. in Network Security and Communication
ISSN: 2321-3256

New Approach for Sampling Mobile Phone Accelerometer Sensor Data for Daily Mood Assessment

Dept. of Computer Application, IIPS, DAVV, Indore, India

Received: 16 Jun 2013 Revised: 28 Jun 2013 Accepted: 20 Jul 2013 Published: 30 Aug 2013

***Abstract*—**With the increasing stress and unhealthy in people's daily life, mental health problems are becoming a global concern. In particular, mood related mental health problems, such as mood disorders, depressions, and elation, are seriously impacting people's quality of life. However, due to the complexity and unstableness of personal mood, assessing and analyzing daily mood is both difficult and inconvenient, which is a major challenge in mental health care. In this paper, we propose a novel framework for assessing and analyzing daily mood of persons working in corporate organizations. It uses mobile phone data—particularly mobile phone accelerometer sensor data to extract human behavior pattern and assess daily mood. We also present a sampling approach for rapidly and efficiently computing the best sampling rate which minimizes the Sum of Square Error in order to handle the large data.

***Keywords*-** Mobile Phone Sensor, Mood Assessment, Sampling, Clustering, Daily Mood Assessment (DMA)

1. INTRODUCTION

As technologies and economy continue to develop dramatically, *well-being* goes well beyond keeping one from starving, cold, and physical diseases. People begin to pay more attention to other elements that affect people's feeling of *happiness*, of which mental health is an important aspect. Mood related mental problems, such as mood disorders, have become a global concern of human society.
The study of mood theory and the treatment of mood disorder has been an important topic of psychology. Different measures and treatments are applied according to different mood status and symptoms. However, the assessment of mood has long been a challenge in mood-related study. Current mood assessment is mainly based on traditional psychological measurements, like scales and psychological counseling. But these methods are faced mainly with two challenges in mood measuring.

One is the *subjectivity* of mood. Mood is with no doubt subjective, but for scientific analysis, objective mental states should be extracted from people's feelings and expressions. Self-reported data, suffering from its subjectivity, cannot serve as a reliable indicator of objective mood, without complex validation and process. The other is *inconstancy*. In comparison with emotion, mood is relatively long-time state, but it still changes in hours or days. Therefore, it is very difficult to use traditional methods to assess mood in such frequency. Psychological scales, due to their complexity, consume a large amount of time and energy of user, which is often annoying to do every day. Psychological counseling involves interaction between client and counselor (psychologist), and thus, it is more unrealistic to be taken frequently. Fortunately, the rapid increasing usage of mobile phones has opened a new possibility for daily mood assessment. Modern mobile phones are equipped with rich sensors, such as accelerometer, light sensor, sound sensor (microphone), location sensor (GPS), etc. Besides such sensory data, the mobile phone generates soft-sensor data—SMS and call information—that can also reflect people's daily behavior. What's more, the mobile phone is usually inseparable with its user, and then it can use its hard- and soft-sensors together to profile some significant behavior patterns of people.

In this paper, we propose a novel framework for daily mood assessment using mobile phone sensor data. First, we collect sensor data on mobile phone in a smart manner, which is both efficient and energy-friendly. Second, different types of sensor data and communication events are then combined together to model people's daily behavior pattern, including physical movements, minor motions etc..
Third, these features are used as input to model users' real-time mood.

The continuous arrival of data in multiple, rapid, time-varying and possibly unbounded streams appear to yield

Corresponding Author: *Rajesh Verma, IIPS, Indore*

fundamentally new research problems. Sampling techniques have been widely studied across several disciplines, but only a few of the techniques developed scale to support clustering of very large data.

We propose in this paper a new approach for sampling multiple source data streams which is applied to the Sum of Square Errors matrix. The elements of this matrix represent the error between the original and the interpolated measurement curve of a sensor when its data are sampled at a specific rate. In sensor networks, the rate at which data are collected at each node affects the communication resources and the computational load at the central server. The proposed approach seeks to attribute the best sampling rate for each sensor and to guarantee that the volume of transferred data fits a predefined bandwidth.

2. WHY MOBILE PHONE SENSORS?

Today's Smartphone not only serves as the key computing and communication mobile device of choice, but it also comes with a rich set of embedded sensors, such as an accelerometer, digital compass, gyroscope, GPS, microphone, and camera. Collectively, these sensors are enabling new applications across a wide variety of domains, such as healthcare, social networks, safety, environmental monitoring, and transportation, and give rise to a new area of research called mobile phone sensing. First the availability of cheap embedded sensors initially included in phones to drive the user experience (e.g., the accelerometer used to change the display orientation) is changing the landscape of possible applications. Now phones can be programmed to support new disruptive sensing applications such as sharing the user's real-time activity with friends on social networks such as Facebook, keeping track of a person's carbon Footprint or monitoring a user's well being. Second, Smartphone's are open and programmable. In addition to sensing, phones come with computing and communication resources that offer a low barrier of entry for third-party programmers (e.g., undergraduates with little phone programming experience are developing and shipping applications). Third, importantly, each phone vendor now offers an *app store* allowing developers to deliver new applications to large populations of users across the globe, which is transforming the deployment of new applications, and allowing the collection and analysis of data far beyond the scale of what was previously possible.

2.1 Accelerometer

Accelerometer data is calculated by measuring forces applied to the sensor itself, including the force of gravity. There are three readings corresponding to three axes of coordinate system: the x-axis horizontal and points to the right, the y-axis vertical and points up, and the z- axis points towards the outside of the front face of the screen. Thus accelerometer reading at time t can be denoted as ($x^{(t)}$, $y^{(t)}$, $z^{(t)}$). Accelerometer may be the most common sensor available on smart phones, and it has relatively low energy cost. Accelerometer data can be used to detect user's physical movement, which is very important in behavior modeling. For these reasons, accelerometer plays as a central role in our system.

3. DATA AND MODEL DESCRIPTON

3.1 Feature Definition

This approach deals with the problem of daily mood assessment using mobile phone sensor and communication data.

Definition 1: Mood. Mood is defined as daily emotional status of a person. There are a number of structure theories about mood, most of which decompose mood into three dimensions. Thayer further states that the three dimensions are not equal - Two of them are basic dimensions and the other is a mix of two basic dimensions. We adopt Thayer's theory and choose three dimensions to represent an individual's daily mood, including *displeasure, tiredness and tensity.* Degree of *displeasure* of user i in day t is donated as $d_i^{(t)}$, tiredness as $ti_i^{(t)}$ and tensity as $te_i^{(t)}$. The three values $d_i^{(t)}$,$ti_i^{(t)}$ and $te_i^{(t)}$ are integer values ranging from 1 to 5 with 1 for least severe and 5 for most severe. For example, $d_i^{(t)}$ can be any value of "very pleasant", "pleasant", "medium", "unpleasant" and "very unpleasant", with "very pleasant" being 1 and "very unpleasant" being 5. Overall mood of user i in day t is $m_i^{(t)}$, where $m_i^{(t)}=(d_i^{(t)},ti_i^{(t)},te_i^{(t)})$. As explained above, the three dimension of mood is not totally independent. In fact, *displeasure* is an overall evaluation of mood states, whose value is affected by *tiredness* and *tensity*. Generally, people who feel less tired and more relax may feel more pleasant.

Modeling mood as three dimensions gives us flexibility to find relationships between daily behavior and a particular mood dimension, which is often more remarkable. For example, micro motion is more related to degree of tensity than degree of tiredness. The relationship between different dimensions can help validate user-report data, and thus increases the reliability of the data.

3.1.1Daily behavior features. We have defined a set of daily behavior features extracted from mobile sensor data and communication data. All features of user i in day t is

denoted as $X_i^{(t)}$ specifically, $x_{ij}^{(t)}$ is the value of feature j of user *i* at day *t*.

- Micro motion. Micro motion is defined as the following user behaviour—a user picks up the phone and does nothing useful for no longer than a few seconds. This feature can be extracted using accelerometer raw data.

4. PROBLEM DEFINITION

4.1 Learning problem

Given a feature set $X_i^{(t)}$, and the previous mood state $m_i(t-1)$ the goal is to establish an assessment function *f that* outputs mood $mi^{(t)}$. More specifically, di $^{(t)}$, ti_i $^{(t)}$, *tei* $^{(t)}$.Formally, *f* is defined as:

$$F\ (X_i\ ^{(t)},\ \text{mi}\ ^{(t-1)})\ \text{->m}_i\ ^{(t)}$$

These large data can be effectively stored on a disk with limited storage space by assuming that sensor measurements are generated regularly at the same rate for every sensor. We cut out all streams into sampling periods (*T)*—also called temporal windows— of equal length. A typical sampling period corresponds to a length of one day. A set of measurements issued by a sensor during a sampling period is also referred as a *curve* from now on. A temporal window is composed of *p* measurements for each sensor (the number of measurements per sensor in the window is the same for all sensors). For each sensor, the number of measurements that have to be stored within a temporal window can vary from *m* (fixed parameter) to *p*. Let us define *s* as the maximum number of measurements which can be stored on the disk from the *N* sensors during a temporal window ($s < N * p$).

The problem consists of finding the best policy to summarize sensor curves within a time window which respects the constraints of the maximum number of data stored on disk (parameter *s*) and the minimum number of data taken from a sensor (parameter *m*). The goal is to answer to aggregate queries using the summary. Each curve *r* takes the value *Cr (t)* at time *t*. In this article, we concentrate on SUM aggregate queries:

$$C(t) = \sum_{r \in p(N)} C_r\ (t)$$

Where *P (N)* is a subset from {1, 2 . . . *r*. . . *N*}.

The data collection model should preserve detailed information as much as possible by reducing summarizing errors. Several criteria can be used to measure errors. In this article, we use the Sum of Square Errors (SSE), which is also the square of the *L2* distance between the original curve $C = \{c1, c2, \ldots, cp\}$ and $\hat{C} = \{\hat{c}_1, \hat{c}_2, \hat{c}_3 \ldots \hat{c}_p\}$ an estimation from the summarized curve . SSE is computed as follows:

$$SSE(C, \hat{C}) = \sum_{i=1}^{p} (c_i - \hat{c}_i)^2$$

5. TEMPORAL SAMPLING

Temporal sampling is the process of keeping fewer measurements from a curve, preserving the underlying information as much as possible. Note that if no temporal sampling is applied, then aggregating the whole population curves at each timestamp *t* consists of summing all values *Cr (t)* measured on each curve *r*.

From the observation of what is happening during sampling period $T - 1$, we determine a data collection model to apply to the *N* sensors on the next sampling period *T*. During a sampling period *T* , sensors send data to the DSMS. The engine computes on the fly SSE values for different assumptions of sampling rates between *m* and *p* values. It also summarizes data streams at the scheduled sampling rate and stores summarized data on a database. Then, the application solves an optimization problem to find the best sampling rate for each sensor during sampling period $T + 1$. Several sampling schemes can be considered to summarize data streams during a sampling period . Almost all methods of curve compression can be used depending on the handled data. For the needs of our application, we concentrate on three methods: regular sampling, curve segmentation and wavelet compression.

5.1 Regular sampling

Curves are sampled using a step *j* chosen between 1 and maxStep = _ *pm* _ (_*x*_ is the floor function). When $j = 1$, all measurements are selected, when $j = 2$ every two measurement is selected and so on. This technique is described as *regular* because selected points are temporally equidistant, and a jump *j* is made between two sampled measurements. For the estimation of the original curve from the sampled measurements, linear interpolation is used.

At the end of each sampling period *T* , different values of SSE are computed which correspond to different levels of summarization, indexed by *j* varying from 1 to maxStep. These SSE values are stored in a matrix *WN*∗maxStep of *N* rows and maxStep columns (*N* is the number of sensors connected to the server). Element *wi j* of the matrix corresponds to the SSE obtained when collecting data from sensor *i* with a summarizing level *j* . Instead of equally distributing the summarizing levels between all sensors for sampling period *T* +1, an optimization is performed to

assign different sampling rates to sensors from the SSE values computed during sampling period T . This problem can be formalized as an optimization problem minimizing the sum of SSEs on sensors:

6. SPATIAL SAMPLING

Spatial sampling is the process of selecting an appropriate set of sensors so that the sample allows estimating unknown quantities (queries) of the population (the N sensors). This is similar to a standard sampling survey problem.

We consider a finite population (the N sensors). Each sensor can be identified by a label. Let {1, 2, . . . , r, . . . , N } be the set of these labels. The aim of survey sampling is to study a variable of interest which is the curve in our context. The curve takes the value Cr (t) for a sensor rat instant t. The purpose is more specifically to estimate a function of interest of the Cr's for all or a subset of sensors. Note that we concentrate on SUM aggregate queries in this article. There are many strategies of sampling in literature: simple random sampling, stratified sampling, sampling with unequal probabilities, balanced sampling, etc. We use the simple random sampling in our framework. Let S be a sample containing n sensors, πr be the inclusion probability, i.e., the probability that sensor r is in the sample S, πrq be the second-order inclusion probability, i.e., the probability that both sensors r and q belong to the sample S. We use the simple random sampling in our framework where the same inclusion probability πr is assigned to each sensor. An unbiased estimator often used is the Horvitz-Thompson estimator. We can calculate the SUM estimate of curves as follows:

Variance of this sum can be estimated as,

$$\hat{V}_{ar}(\ \hat{C}_{HT}(t)) = \sum_{r\in S}\sum_{q\in S}\left(\frac{Cr(t)}{\Pi r}\frac{Cq(t)}{\Pi q}\frac{\Pi rq-\Pi r\Pi q}{\Pi rq}\right)$$

For simple random sampling

$$\Pi_r = \Pi_q = \frac{n}{N} \text{ and } \Pi_{rq} = \frac{n(n-1)}{N(N-1)}$$

7. MASTER ALGORITHM

In both sampling techniques above (temporal sampling and spatial sampling), summarized estimates are used to answer the SUM aggregate queries. Interpolation in the first case estimates affected by sampling error in the second case. We propose a method that combines these two approaches in order to reduce summarizing errors as much as possible. We call this "Method for Adaptive Spatial Temporal summaries" or MASTER".

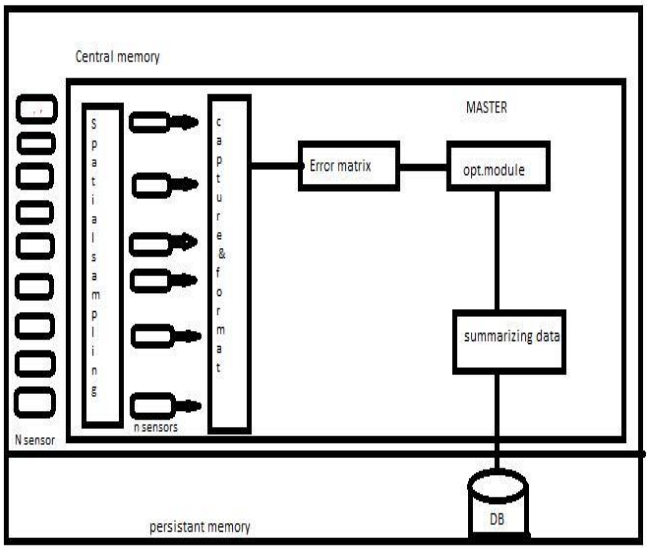

Fig: Global architecture for MASTER

The principle of the spatio-temporal approach is represented in Fig. 5 and is as follows: A survey technique is integrated into the temporal sampling approach to make up a spatio-temporal summary. As with traditional approaches, a survey plan is set up to select a sample with a size of n sensors from the whole population, i.e., the N sensors. Then, we apply a sampling method to collect measurements at variable granularities in time for each sensor in the spatial sample and store them in a database. For the needs of our application, we concentrate on one method: regular sampling.For the three sampling methods presented in Sect. 4, the worst case complexity of computing the errors for different summarizing levels is O(mp) (curve segmentation).

However, we can reduce computational overhead for some summarizing methods. In the case of regular sampling (complexity O(1)), errors can be computed incrementally. Indeed, to update the matrix of errors in the case of regular sampling, each sensor should send m SSEs (we recall that maxStep is the lower limit of the sampling step). However, it is not necessary to send all maxStep SSEs. Indeed, for a curve sampled at a step j, the sensor sends maxStep− M j SSEs, M j is the number of multiples of j that are lower than maxStep. The DSMS collect sampled data and can therefore compute the missing SSES to update the matrix. For example, if a curve is sampled at steps from j = 2 to maxStep = 20, the sensor sends at the end of the sampling period maxStep − M2 = 20 − 10 = 10 SSEs as M2 = card({2, 4, 6, . . . , 20}) = 10. In our experiments, for a threshold s = 14,400 and maxStep = 20, each sensor sends an average of 9 SSEs instead of 20. Even if the segmentation is the most efficient summarizing method for

power consumption curves, we apply regular sampling in MASTER because of its low complexity.
The sensors that are not part of the spatial sample are not observed. The curves in the temporal sample are reconstructed by linear interpolation. The remaining curves are estimated with the average curve of the sample.

Algorithm 6.1 describes MASTER steps.

Algorithm 7.1 MASTER

Input: T, max Step, s

Output: Sample

Step1: Apply the spatial sampling to the whole population (the N sensors) and select n sensors

Step2: Define sensor curves by reading all measurements sent by the n selected sensors during sampling period T

Step3: Compute SSE matrix (Wn*max Step) from curves of the n selected sensors

Step4: Compute the best sampling rates respecting threshold s for each sensor by applying the linear programming solver to the SSE matrix

Step5: Apply the temporal sampling to the n selected sensors on sampling period T + 1 with the sampling rates from period T

8. CONCLUSION

In this paper, we propose a novel approach to assessing Individual's daily mood using mobile phone. We first extract several features from mobile phone data, and then propose a method based on factor graph for assessing mood using these features. We have built a system for data collection and model implementation. We presented CLUSMASTER, a new approach for combining clustering and sampling techniques on data streams. The goal is to assign the best sampling rate to each individual sensor in a network in a rapid and efficient way by minimizing the sum of square errors (SSE). . Moreover, our approach overcomes the main limitations of the well-known data stream clustering methods. Compared to BIRCH and STREAM methods, CLUSMASTER proposes a dynamic optimization stage which continuously adapts with the content of the stream. The proposal of efficient and fast sampling techniques able to deal with this new environment of sensor networks will certainly yield future research challenges. The analysis of results obtained by our approach clearly testifies the power and benefits introduced by combining clustering and sampling techniques. Clustering strategies as the ones presented in this work are recommended for dealing with the problem of sampling and storing massive data streams generated by multiple sources.
Future work on CLUSMASTER may include the following: experiments on larger data sets generated by a larger number of sensors on a longer sampling period, extension of the approach to sensors producing multi- dimensional numerical data, investigation and test of other clustering techniques.

9. REFERENCES

[1] Yuanchao Ma, Bin Xu, Yin Bai, Guodong Sun(2012) Daily mood assessment based on mobile sensing,2012 ninth international conference on wearable and implantable body sensor networks.

[2] Nicholas D Lane, Emiliano Miluzzo, Hong Lu, Daniel Peebles (2010), A survey of mobile phone sensing. IEEE Communications Magazine 2010

[3] Aggarwal CC (2010) A segment-based framework for modeling and mining data streams. In: Knowledge and information systems, pp 1–29. Springer

[4] Aggarwal CC, Han J, Wang J, Yu PS (2003) A framework for clustering evolving data streams. In: Proceedings of the 29th international conference on very large data bases (VLDB'2003), pp 81–92

[5] Bash BA, Byers JW, Considine J (2004) approximately uniform random sampling in sensor networks. In: Proceeedings of the 1st international workshop on Data management for sensor networks (DMSN'04), pp 32–39 (Toronto)

[6] Csernel B, Clerot F, and Hebrail G (2006) Streamsamp: datastream clustering over tilted windows through sampling. In: ECML PKDD 2006 Workshop on knowledge discovery from data streams, Berlin.

