Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 109C (2017) 1028–1034

Procedia Computer Science

www.elsevier.com/locate/procedia

The 7th International Symposium on Frontiers in Ambient and Mobile Systems (FAMS 2017)

Toward a proposed framework for mood recognition using LSTM Recurrent Neuron Network

Huu Son, Hoang[a]

[a] University of Hagen, Hagen, Germany

Abstract

Affective analysis plays an important role in understanding human characteristics, predicting human behavior and diagnosing mental health problems. Although a large number of affective recognition researches have been published, predicting mood - the long lasting affect from real life context is still a challenge because of the complexity of correlations between mood and daily factors. We therefore aim at developing a framework for predicting mood considering several aspects on human and environment adaption. The framework consists of two components: a data acquisition tool designed for wearable devices and a mood model based on Long Short-Term Memory Recurrent Neuron Network (LSTM-RNN). In this paper, we are focused on presenting the later component to apply LSTM RNN for mood prediction in daily life.

Peer-review under responsibility of the Conference Program Chairs.

Keywords: Mood recognition;Thayer model; LSTM-RNN; Mobile computing;

1. Motivation

Affective states: emotions (in short term) or moods (longer lasting) provide valuable information about personal traits, sociability and well-being. They reflect environmental adapting abilities and warm risks of mental health problems. Moreover, affect can lead us to extend or intellectual facilities of consciousness, perception and reasoning [1].

Mood recognition in real life is still a big challenge. Different from emotions, natural moods are impacted by many factors either internal or external, even sometime these influences are not strongly clear. Due to the subjectivity and the inconstancy of personal mood, mood assessment based on traditional methods such as surveys or psychological counsellings are improper [3].

In last two decades, mood recognition based on daily context through mobile environment has been an interesting topic. For instances, MoodMiner [3] assesses mood according to Thayer model with three dimensions (displeasure, tiredness and tensity) based on location micro-motion, communication frequency and activities using factor graph and

* Corresponding author. Hoang Huu Son. Tel.: +4915773136916
E-mail address: hoanghuuson@gmail.com

1877-0509 © 2017 The Authors. Published by Elsevier B.V.
Peer-review under responsibility of the Conference Program Chairs.
10.1016/j.procs.2017.05.378

classification approach. Sato A. developed SNATSHOT study[4] that focused on correlations of stress, mental health and academic performance with ambulatory data that were gathered throw mobile and wearable combine to questionnaire methods. MoodScope[5] used mobile phones to predict two axes of daily emotion(activeness and pressure) using multi linear regression model with 93% accuracy rate. Although aforementioned studies have contributed to describe relationships of mood and various human factors as well as indicated mood assessment models, several limitations have existed. Firstly, traditional data mining techniques only highlight correlations between mood and human factors based on dataset and ignore understand psychological mechanisms of natural mood. Secondly, mood variation is a gradual and continuous process and should be dependent upon not only current factors but also historic mood states, so into mood recognition model, mood should be considered as a sequences instead of discrete states (as emotion).

Hence, to overcome these limitations, our research is conducted based on the assumption: mood is mentioned as a time series of continuous values. Utilizing the ambient information that can be recorded through wearable devices and the available previous states of mood, the change of mood can be assessed.

Hidden Markov Model, one of the most popular approaches used to sequence processing, observes sequence as probabilistically dependent upon a sequence of unobserved states. However Markov assumption is unreasonable for modeling mood mechanism, because mood is continuous and depends on more than one previous states. In our approach, RNN-LSTM is applied to assess mood states. RNN can generate mood states from influences of activities, environmental adaption as well as the correlations of mood to physiological signals and behavior. LSTM blocks allow the network flexibly capture historical states in both short term and long term to calculate current mood state.

Following the above idea, the paper is presented into 4 sections. Beside motivation and conclusion, section 2 presents fundamentals of mood model according to Thayer model, In section 3 describes about out proposal framework that consists a collection data tool and mood model based on LSTM-RNN.

2. Fundamentals of Thayer Model

There are several models to present mood that are used two dimensions (Positive Affect and Negatives Affect scale, Arousal and Valence) or three dimensions (Pleasure, Arousal and Dominance (PDA) or Valence, Arousal and Calmness)[6]. In this study, Thayer model has been selected because the model covers two main aspects of mood: psychology and biology[8].

According to Thayer theory, moods are defined as insight feelings that last for a long period of time (an hour or a day) and that often have not particular causes[7]. Thayer proposed two dimensions model for mood assessment: Energetic Arousal (EA) and Tense Arousal (TA)[7]. EA is a dimension characterized by range of feeling from tired or sleepiness at the low end to alert and awake at the high end. The high level of EA associates positive affect tone, having motivation and ready taking action. The low level of EA indicates that to have reduced resources and need to rest and recuperate, or seek alternative energy resources[8]. On the other hand, TA is a dimension characterized by range of feeling from calmness or stillness at the low end to tension and anxiety at the high end. High levels of TA are associated with negative affective tone, and low levels of TA correlates to good mood with calm or peaceful states. Both TA and EA are measured following to self-assessment method based on a questionnaire named The Activation Deactivation Adjective Check List (AD-ACL) and mapped into continuous values from 0 to 10[7].

Mood mechanism is complicated and a combination of psychological aspects and biological aspects[7]. In cognitive process, nervous system takes ambient information through scenes, processes and triggers reactions including perception, behaviors, thought, feeling, emotion and mood. Hence mood is influenced by external events that appraisal to person: social interaction, cognition, memory, and stress events in daily life. In addition, mood relates to biological processes insight body in general and neuron processes in particular. Therefore, factors impact to health also have correlations with mood such as health status, nutrition, sleep quality, physical activity, time of day and stress[8]

3. A proposed Framework for Mood Modeling

In this study we have proposed a framework for mood modeling consisting of two main components: A data collection tool and a mood detection model based on LSTM Recurrent Neurons Network.

Table 1. List of features collected through smart watches and mobile phones on an hourly basis

Factors	No. of features	Features
Mood states	2	TA, EA
Activities	4	Type, duration, eating, drinking
Location	3	Type, duration,number of surrounding people
Weather	3	Weather type, temperature, sensitivity
Heart rate	8	HR_avg, HRV_arg, HR_min, HR_max, SDHR, SDHRV, hr_diff_avg, hr_diff_var
Pedometer	6	%waking, %runing, %NOT_Moving, number of footstep, moving speed, Carlo consuming
Sleep	1	Sleep efficiency score

3.1. Data Acquisition Tool

In this proposed framework, to obtain proper a dataset, an integrated application named CollectionInformation was developed and available on Google Appstore [1] for Android mobile and Samsung smart-watch. A Tizen component on smart-watch records heart rate and pedometer signals and samples in every minutes and mornitors sleep quality throws Sheath application. An Android component on mobile phone records and samples unobtrusive data in each minute (GPS signals, user-mobile interaction, Internet connection protocol, Bluethoth connection) and obtrusive data in each hour (mood annotation, activities drinking kind, eating). User's characteristics (profile, personality, general routine) are gathered through survey. The data are locally stored in mobile phone for each user and globally stored at server for all users. A web form is designed to manage annotated information as well as display user information [2].

A group of Vietnamese students studying in Germany are participants. Initially, each user created his/her account and completed a survey to gather profile information, medical indexes, personality and daily routines. Samsung wearable devices (mobile and smart watch) are equipped for participants. Information of user is recorded on working day (from Monday to Friday) in a month.

after preprocessing and extracting features from raw signals and obtrusive data, all daily data collected from mobile and smart watch is split into an hour window size that includes list of features (table 1).In total, 27 features (in 7 factors) are considered for the analysis.

3.2. Features Normalization

All features are normalized into range of [-1,1]. A feature with a normalized value of -1 means that it causes completely negative impact on mood (bad mood). On contrary, if the normalized value equals to 1, it causes completely positive impact on mood (good mood).

Based on assumption that informal (e.g going to party), physical, and nutrition activities can improve mood while formal activities (studying or working) in a long period probably increase stress, tension causing bad mood[12]. With location type, natural and informal social places can probability toward good mood, while official places or loneliness can raise bad mood.

With the research subject student, activities such as physical activities, relaxing, shopping, informal communication with friend (or family) help toward good mood, while formal activities like working, study (at home and university), formal meeting, or conference worsen mood. Similar to activities, we categorize location types as 1.) Formal places: University, office, 2.) Informal places; natural area, shopping center, coffee shop or restaurant (except working activities) and 3.) At home. Based on category of places and duration, number of surrounding people, some rules are defined to mapping features of location to -1 (toward bad mood), 0 (not relevant), or +1 (toward good mood). For

[1] https://play.google.com/store/apps/details?id=com.sonhh.dat.collectinformation
[2] http://aofire.xyz/sonhh/MoodDatatabase/index.php

example a rule: "if (Location="university and duration $\gtrsim$ 4) then x =-1" means that staying at university more than 4 hours probability toward a bad mood.

Weather features also are mapped to mood appraisal based on assumption: "Although weather affect little to mood[12], there are several special weather conditions that considerably impact to mood". Some special weather conditions are defined based on temperature and weather type: 1) Good weather (mood appraisal = 1): comfortable temperature (18C-25C) and sunny; clear sky after rainy. 2) Bad weather (mood appraisal =-1): low temperature ($< 4^oC$) ; high temperature ($> 30^oC$); heavy cloudy; heavy rainy.

Moreover, weather significantly impacts to person's mood if person is outside[9] so an sensitivity index are defined as a percentage of outside time in an hour. Hence, appraisal of weather to mood can be calculated as equation:

$$\begin{aligned} x &= sensitivity.appraisal(weather\_condition) \\ sensitivity &= \frac{\#outdoor}{\#outdoor + \#indoor} \end{aligned} \tag{1}$$

Sleep efficiency score (SEC) is defined as the ratio of total minutes asleep to total minutes in bed[11]. SEC is used to measure quality of sleep, sleep is considered as good if $ESC \geq 85\%$ $(x = +1)$, neutral if $50\% \geq ESC < 85\%$ $(x = 0)$ and bad if $ESC < 50\%$ $(x = -1)$.

Features of heart rate, pedometer, mobile interaction are continuous values and have various ranges. Hence, all features are mapped as into [-1,1] following to equation 3:

$$F(j)_{norm} = \frac{F(j) - Z(j)_{mean}}{Z(j)_{max} - Z(j)_{min}} \tag{2}$$

where by: $F(j)_{norm}$ are normalized values, $Z(j)_{mean}$,$Z(j)_{max}$ and $Z(j)_{min}$ are mean, maximum, and minimum values of feature j, respectively.

Finally all feature at the time t are combined as a temporal normalization vector $X^t = (x_1^t, x_2^t, ..., x_n^t)$ where $x_i^t \in [-1, 1]$

3.3. Using LSTM-RNN for mood prediction

An LSTM-RNN is an artificial neural network that is extension of Recurrent Neural network (RNN)[13]. RNN allows to generate a continuous sequence by establish loop connection, so information of current state is used in next state. A LSTM block is multi layers network and applied to replace memory states in RNN that excels at remembering values for either long or short durations of time.

An architecture of LSTM block consist an internal cell (C) and three activate gates to control information flows in the block (fig 1). An internal cell contains hidden states and has connections with gates to updated state by the time. An input gate (i) controls the extent to which a new value flows into the memory. A forget gate (f) controls the extent to which a value remains in memory. An output gate (o) controls the extent to which the value in memory is used to compute the output value.

In this research, mood recognition model is designed as a multiple neuron network with two layers: Recurrent layer and output layer.

Recurrent layer consists m LSTM blocks to store internal information in brain effecting to mood. At the time t, each block is activated by a temporal input vector X^t that are normalized in above subsection. There are 2 connections through the time: 1) A beep hole connection is established to allow internal memory states to control gates[14], 2) Output connections represent the correlation between LSTM blocks in that the operation of each block is influenced by the result of others ($h^t = (h_1^t, h_2^t, ..., h_m^t)$). an Operation of each blocks is defined as a series of below equals (Eqs. 3)

$$\begin{aligned} f^t &= \sigma(X^t U_f + h^{t-1} V_f + C^{t-1} W_f + b_f) \\ i^t &= \sigma(X^t U_i + h^{t-1} V_i + C^{t-1} W_i + b_i) \\ z^t &= g(X^t U_z + Y^{t-1} V_z + b_z) \\ C^t &= f^t \odot C^t + i^t \odot z^t \\ o^t &= \sigma(X^t U_i + h^{t-1} V_i + C^t W_i + b_o) \\ h^t &= o^t \odot h(C^t) \end{aligned} \tag{3}$$

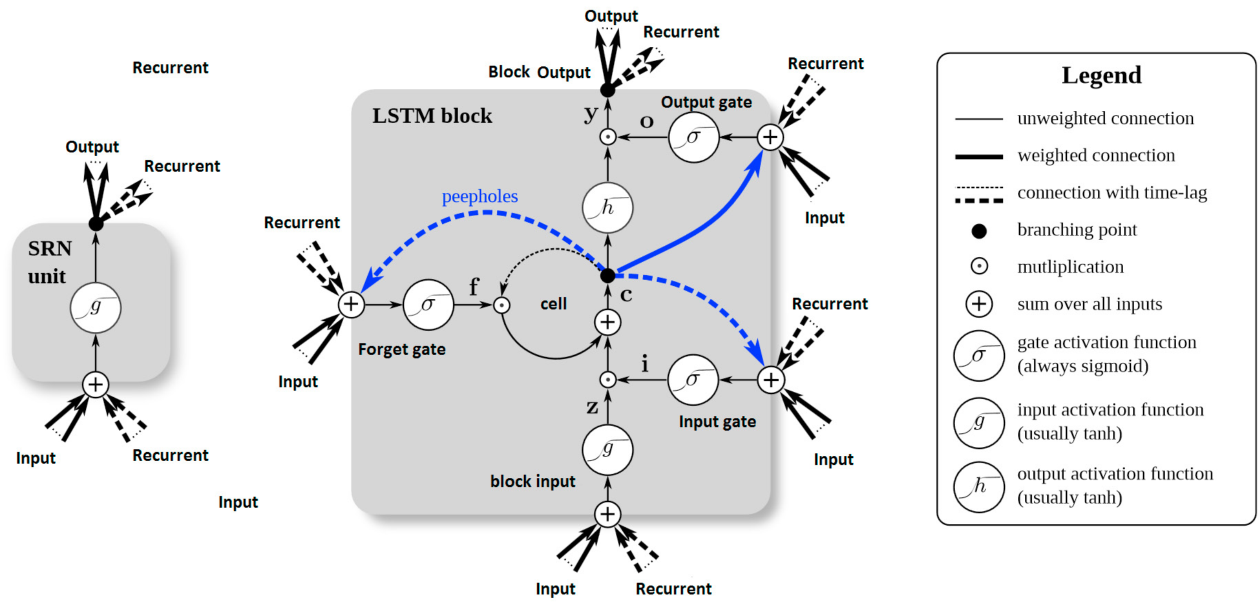

Fig. 1. A LSTM block with the hidden layers of a recurrent neural network using beep holed connection (Source: [15])

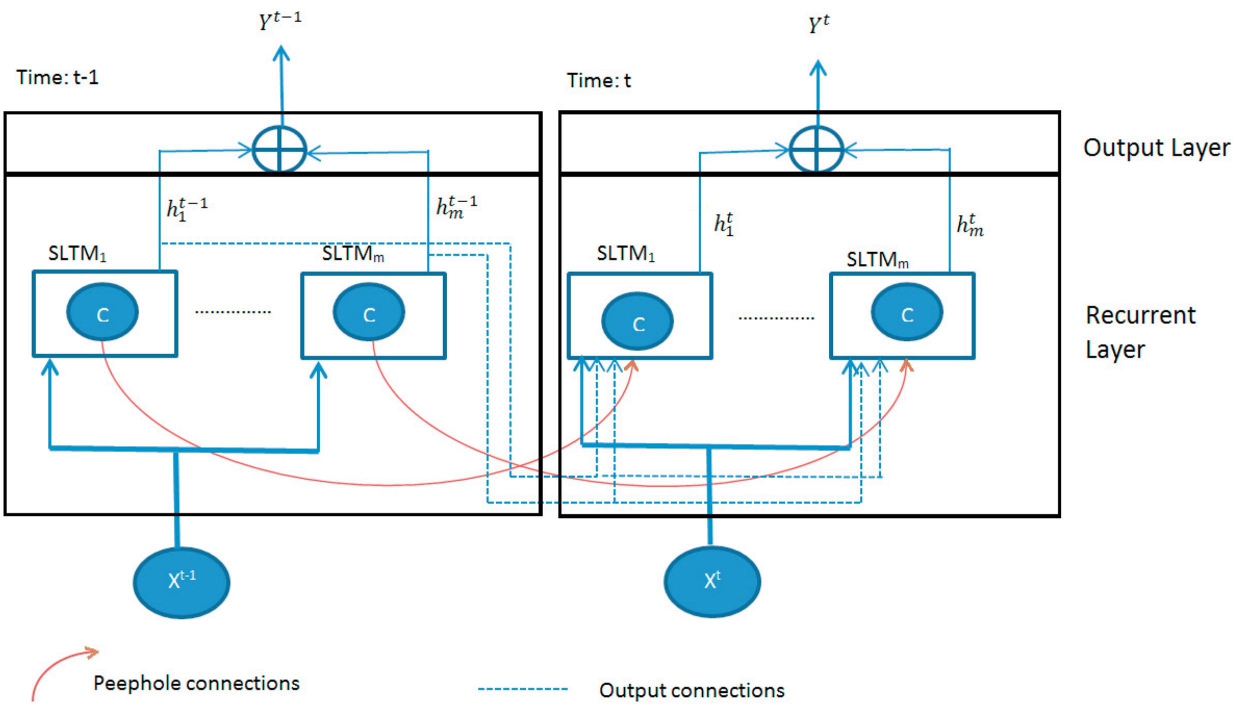

Fig. 2. A Neuron network including m memory blocks and their connections for mood prediction

The operation of the LSTM block at the time t can be explained as: In first step, value of 2 gates: forget gate (f^t) and input gate (i^t) are calculated from input vector (x^t), previous output $y^{(t-1)}$ and previous memory states (C^{t-1}). A sigmoid activation function $\sigma(x) = \frac{1}{1+e^{-x}}$ allows value of gates in range [0,1] corresponding the percentage of information pass through these gates. In second step, input block (z^t) is calculated from input vector and previous output. In third step, cell state (C^t) is updated from the previous state and the input block according to portions f^t and i^t. In forth step, the output gate value(o^t) is calculated from input vector, previous outputs and current cell state using sigmoid function to identify the percent of cell state through the gate. Finally, cell state is pushed through $h(C)$ function and multiply it by output gate values to get output values.

In these equations, U_*, V_*, W_* ($* = f, i, z, o$) is denoted weight parameter matrices using in the block. U are the matrices of weight from input vector to gates, V are the matrices of weight from previous output and W are matrices of peephole connection. b_* terms denote bias values. $g(.)$ and $h(.)$ functions are tangent function ($tanh(x) = \frac{\exp(x)-\exp(-x)}{\exp(x)+\exp(-x)}$) to identity input and output blocks into range of [-1,1].

In output layer, mood state is calculated as a mean of output of LSTMs in hidden layer. Simplify, the operation of node is defined as equations:

$$\begin{aligned} y^t &= f(H) = H \\ H &= \sum_{k=1}^{m} W_k h_k^t + b_o \end{aligned} \tag{4}$$

Where by:

- $y^t = (EA^t, TA^t)$ is output vector.
- h_k^t (k=1,2,..,m) are output vector of $LSTM_k$,
- W_k (k=1..m) are weights and b_o is bias.

3.4. Loss functions

The loss function is defined as a measure of how far away a particular solution is from an optimal solution to the problem to be solved. A RNN can be unfolded into sequence of feed forwarded neurons, so the loss function is a sum of single point loss:

$$Loss = \frac{1}{T}\sum_{t=1}^{T} l(y^t, \overline{y}^t) \tag{5}$$

Where T is the number of time steps in the sequence. T is assigned as a day because mood trajectories usually follow daily events [10]. y and $\overline{y}^t$ are the target mood dimension value and the predicted values receptively at step t, l is a function to compute the distance between the target and the prediction. In this research, two most common distances are computed as mean square error (mse) and mean absolute error (mae):

$$\begin{aligned} l_{mse}(y^t, \overline{y}^t) &= (y^t - \overline{y}^t)^2 = (EA^t - \overline{EA}^t)^2 + (TA^t - \overline{TA}^t)^2 \\ l_{mae}(y^t, \overline{y}^t) &= |y^t - \overline{y}^t| = |EA^t - \overline{EA}^t| + |TA^t - \overline{TA}^t| \end{aligned} \tag{6}$$

According to the gradient descent, each weight change in the network should be proportional to the negative gradient of the loss function with respect to the specific weight ($\Delta W = -\eta \frac{\partial Loss}{\partial W}$). Hence, the network can be trained based on Back propagation through time for LSTM network that was introduced by K. Greff [15].

4. Conclusion

In summary, a framework based on LSTM -RNN for mood prediction in daily life has been proposed. Natural mood of human is considered as sequence of states according to time series. Mood variation can be predicted with daily ambient context of user such as activity, environmental adaption, physiological signals and behavior. The proposed model based on LSTM-RNN shows its suitability for modeling mood considering multiple aspects. In that, two dimensions of mood are modeled as LSTM networks with two layers: A recurrent states layer that consists m memory cells to compute hidden values and a output layer that computes result as a mean of hidden values. Comparing with other previous mood assessment studies, our proposal is more sensitive to the way that human brain do. Moreover, this proposal seems to be an empirical aspect of the Thayer model that indicates the mechanism o mood as an intersection of psychology and biology.

Currently, the collection tool was completed and data acquisition process has been started and in progress. We now start learning and testing the proposed mood model with the available collected dataset and hope to report the promising result soon.

References

1. Picard, R. W. (2003). Affective computing: challenges. International Journal of Human-Computer Studies, 59(1), 55-64.
2. Zeng, Z., Pantic, M., Roisman, G. I.,& Huang, T. S. (2009). A survey of affect recognition methods: Audio, visual, and spontaneous expressions. IEEE transactions on pattern analysis and machine intelligence, 31(1), 39-58.
3. Ma, Y., Xu, B., Bai, Y., Sun, G., & Zhu, R. (2012, May). Daily mood assessment based on mobile phone sensing. In 2012 ninth international conference on wearable and implantable body sensor networks (pp. 142-147). IEEE.
4. Sano, A. (2015). Measuring College Students Sleep, Stress, Mental Health and Wellbeing with Wearable Sensors and Mobile Phones (Doctoral dissertation, Massachusetts Institute of Technology).
5. LiKamWa, R. (2012). MoodScope: Building a mood sensor from smartphone usage patterns (Doctoral dissertation, Rice University).
6. Handayani, D., Yaacob, H., Rahman, A. W. A., Sediono, W., & Shah, A. (2014, November). Systematic review of computational modeling of mood and emotion. In Information and Communication Technology for The Muslim World (ICT4M), 2014 The 5th International Conference on (pp. 1-5). IEEE.
7. R. E. Thayer, The Biopsychological of Mood and Arousal, New York: OXFORD UNIVERSITY PRESS, 1989.
8. Glover, D. L. (2012). Mood & learning: The relationship between energetic and tense arousal states and perceived study quality (Doctoral dissertation, CALIFORNIA STATE UNIVERSITY, LONG BEACH).
9. Kts, L., Realo, A., & Allik, J. (2011). The influence of the weather on affective experience. Journal of individual differences
10. Johnson, E. I., Husky, M., Grondin, O., Mazure, C. M., Doron, J., & Swendsen, J. (2008). Mood trajectories following daily life events. Motivation and Emotion, 32(4), 251-259.
11. Reed DL, Sacco WP. Measuring sleep efficiency: what should the denominator be? J Clin Sleep Med 2016;12(2):263266
12. David Hume. Chapter 8 Emotions and moods, PART TWO The Individual, Objective learning, 258-297
13. Olah, C. (2015). Understanding lstm networks. GITHUB blog, posted on August, 27, 2015.
14. Gers, F. A., Schraudolph, N. N., & Schmidhuber, J. (2002). Learning precise timing with LSTM recurrent networks. Journal of machine learning research, 3(Aug), 115-143.
15. Greff, K., Srivastava, R. K., Koutnk, J., Steunebrink, B. R., & Schmidhuber, J. (2016). LSTM: A search space odyssey. IEEE transactions on neural networks and learning systems.