THIR2 at the NTCIR-13 Lifelog-2 Task: Bridging Technology and Psychology through the Lifelog — Personality, Mood and Sleep Quality

**Pouneh Soleimaninejadian[†], Yewen Wang[‡], Haoyue Tong[†], Zehui Feng[†], Min Zhang[*†], Yiqun Liu[†], Shaoping Ma[*†]**

†: Department of Computer Sci. & Tech., Tsinghua University   ‡: Department of Automation, Tsinghua University

z-m@tsinghua.edu.cn

❖ Introduction

➢ Traditionally, self-surveys and self-reported activities and behaviors were used to analyze a person's mind

➢ 4 psychological categories to understand and model the life-logger.

➢ Study of big five personality traits
  ➢ Openness to experience, Conscientiousness, Extraversion, Agreeableness, Neuroticism

➢ Music mood and style detection

➢ Sleep quality prediction

➢ Mood detection

Thayer's 2D mood model
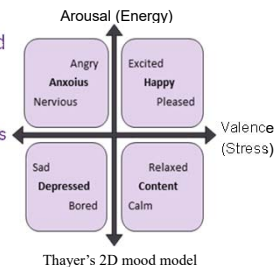
## ❖ Study of Five Personality Traits

40 participants' **3 days**' lifelog data and the NEO-FFI questionnaire.

➢ **Data:**
  ➢ Heart Rate and Mood Record each 3 hours between 8am to 23pm. Mood: Nervous, Angry, Excited, Pleased, Relaxed, Calm, Sad and Bored.
  ➢ Panoramic images of office and bedroom everyday.
  ➢ NEO-FFI test result is applied here for labeling.

Feature	Type	Description
Gender	Binary	1 for female, -1 for male
Moody index	Decimal [0,1]	Moodier the person is, higher value this feature has. Extraction: is the variance of ordinate of mood record
Optimistic index	Decimal [-1,1]	More optimistic the person is, higher value this feature has. Extraction: is the variance of abscissa of mood record
Heart rate Stability	Decimal [0-1]	More stable the person is, lower value this feature has. Extraction: is the variance of heart rate records
Room tidiness Index	Decimal [0-1]	Neater the room is, higher value this feature has. Extraction: Total confidentiality scores (room's tidiness tags)
Room decorative index	Decimal [0-1]	More decorative the room is, higher value this feature has. Extraction: Total confidentiality scores (room's decoration tags)

➢ **Model:** 5 Logistic Regression Models for 5 dimensions

➢ **Experiments:** training: 38 samples (20% cross validation) ; test: 2.
 ➢ Binarizing original NEO-FFI test score [0,5]: threshold 3.
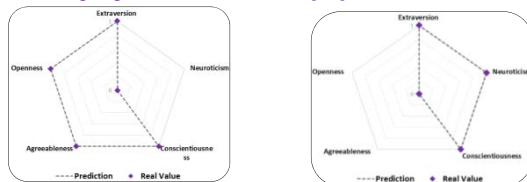
❖ Music Mood and Style Detection

Data Sources: NTCIR Lifelog dataset on **user 1**'s music listening history of 763 songs in **45 days** and retrieved data from online sources.

➢ **Data and Experiment Design:**
  ➢ Lifelog data: Activities, Biometrics, Time stamp
  ➢ **Mood labels** by Thayer 2D Model for each listened music:
    ➢ **Pleased and relaxed:** positive valence
    ➢ **Nervous and sad:** negative valence
    ➢ **Bored and calm:** negative arousal
    ➢ **Angry and excited:** positive arousal
  ➢ **Style labels**: Metal, Jazz, Soul, Pop, Easy Listening, Soundtrack, R&B, Country, New Age, Rock, International, Vocal Pop, Electronic, Folk

➢ **Results:** Best results (Accuracy 85% for Music mood and 80% for music style) achieved by AdaBoost.M1 + Decision Tree in both tasks. (80% as training set, 20% as test set)
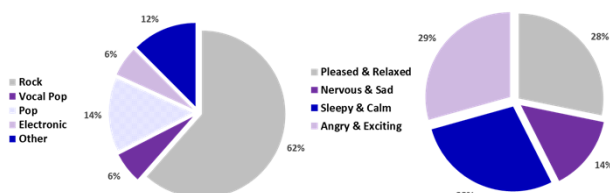
❖ Sleep Quality and Mood Prediction

Dataset: NTCIR dataset on **2 users** + extended dataset on **5** participants. Total of **473 days** of data on sleep quality prediction experiment and **256 days** of data on mood prediction.

➢ **Features:**

Feature	Exp.	Description
User	All	7 binary variables show if the instance is related to a specific user
Weekend	All	A binary variable which indicates if the record belong to a weekend
Home /Work/ Commuting	All	3 integer variables for minutes at Home/ Work/ Commuting
Total Calories	All	An integer variable for calorie consumed during 24 hours
Total Steps	All	An integer variable for steps taken during 24 hours
Average HR	All	An integer variable for average daily HR during 24 hours
Calories in Time	SQ	5 integer variables for calories burned during selected time windows
Steps in Time	SQ	5 integer variables for steps taken during selected time windows
HR in Time	SQ	5 integer variables for average heartrate during selected time windows
Wakeup Time	Mood	An integer between 1 to 24 indicating waking hour
Sleep Duration	Mood	A whole number for sleep Duration in minutes
Sleep Quality	Mood	A decimal showing sleep quality score
Average Arousal	Arousal	A decimal showing average user arousal reported up to last day
Last Day Arousal	Arousal	A binary variable for last day's user arousal record
Average Valence	Valence	A decimal showing average user valence reported up to last day
Last Day Valence	Valence	A binary variable for last day's user valence record

➢ **Models:** Linear Regression model for sleep quality prediction, and two Logistic Regressions for mood dimensions (Thayer's 2D mood model).

➢ **Experiments Design and Results:**
  ➢ For each of the 3 datasets (sleep quality, Mood-Valence, Mood-Arousal)
    ➢ Training: Test = 9: 1
  ➢ For sleep quality
    ➢ Labels: "Poor", "Border Line" and "Good".
    ➢ Test accuracy: 78%
  ➢ For mood prediction
    ➢ Test accuracy: Mood-Valence: 76%; Mood-Arousal: 73%

Visualization and Insight

➢ Provides visual insights on historical data, and gives insights on user's psychological life

❖ Conclusions

➢ We propose novel methods to psychologically understand the user and track user's mental health, Including:
  ➢ Make personality evaluations based on objective data which is time-saving and can obtain real-time evaluation.
  ➢ Determination of music mood and style based on biometrics of the audience.
  ➢ Prediction of the life logger's sleep quality and mood based on lifelog data such as physical activities and biometrics.

❖ Future Work

➢ Enlarge and Diversify the Sample set
➢ More features: Make use of culture differences, daily activities, hobbies, age and more environmental features
➢ Improve the models
➢ Intervention: Giving Suggestions to users for better sleep quality and mood during the day

❖ Acknowledgement

Email: z-m@tsinghua.edu.cn
Website: http://www.thuir.cn/group/~mzhang/