

R&D Theme

Data-driven modeling

Progress until FY2022

1. Outline of the project

We began to work on constructing a "high-low mix database" to construct a data-driven model for personality subtyping. Specifically, we selected the items of the questionnaire and of measuring daily behavior using smartphones that constitute the "low" portion and began preliminary research. The contents of the brain image database, which forms the "high" part, were formulated.

2. Outcome so far

Theme 1: (1) Based on successfully identifying subtypes of healthy subjects using various psychiatric symptoms in collaboration with Theme 3, nine clinical evaluation scales were selected for items in a large-scale Internet/smartphone survey. Questionnaires on comfort and vitality were developed in collaboration with Theme 2 and the social implementation team to start data collection in the next fiscal year. We confirmed that the data structure was suitable for building the subtyping model. In collaboration with Theme 2, we conducted a preliminary survey after receiving approval from an ethics committee.

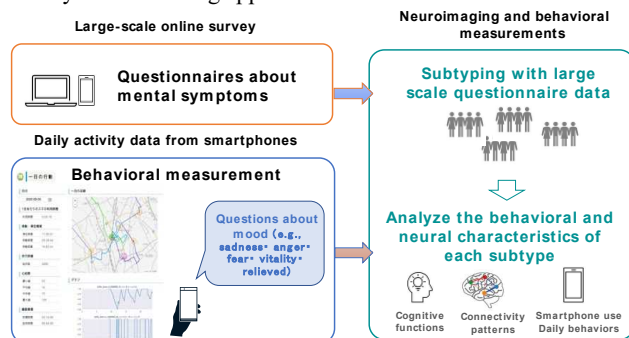


Figure 1: High-low mix database

(2) In collaboration with Theme 3, we adopted the protocol for multicenter disease research of the Human Connectome Project, an international standard for developing brain data protocols. We have completed the construction of an environment that enables us to conduct the preprocessing pipeline for neuroimaging data.

Category	Name	Summary
Clinical rating scales	AQ (Autism-Spectrum Quotient-Japanese version)	Autism Spectrum
	ASRS (Adult ADHD Self-report Scale)	Attention-Deficit/Hyperactivity Disorder
	BIS11 (Barratt Impulsiveness Scale11)	Impulsivity (attention, motor impulsiveness, planning)
	CE5-D (Center for Epidemiologic Studies Depression)	Depression (general population)
	LSAS (Liebowitz Social Anxiety Scale)	Social Anxiety
	OCI (Obsessive-Compulsive Inventory)	Scale to assess the severity of obsessive-compulsive symptoms and symptom dimensions
	SPQ (Schizotypal Personality Questionnaire)	Schizotypal Personality
	STAI-Y (State-Trait Anxiety Inventory-Form JYZ)	State anxiety and trait anxiety
	IGDS (Internet Gaming Disorder Scale-Short Form)	Internet Gaming Disorder
Awareness	FFMQ (Five Facet Mindfulness Questionnaire)	Mindfulness
	MALIA (Multidimensional Assessment of Interoceptive Awareness)	Awareness of interoceptive sense
Compassion	SCS (Short Form of the Japanese version of the Self-Compassion Scale)	Compassion for self
	SAPAS (Japanese Version of the Standardised Assessment of Personality - Abbreviated)	Personality disorder tendencies (Low tendency to compassion)
Positive	DPEs (Japanese version of Dispositional Positive Emotion Scale)	Emotion (compassion)
	IHS (Interdependent Happiness Scale)	Interdependent happiness
Loneliness	SWLS (Satisfaction With Life Scale)	Satisfaction with life
	UCLA Loneliness Scale	Loneliness among the elderly
	SES (Socio Economics Status)	Socio Economics Status
Others	ADEXI (Adult Executive Functioning Inventory)	Adult Executive Function (Working memory and inhibitory function)
	AALS (Anger Arousal and Lengthiness Scale)	Arousal and length of anger
	NAS (Nonattachment Scale)	Nonattachment

Table 1: Questionnaire measurements

There 2: A preliminary survey was conducted in preparation for a large-scale survey to construct high-low mix data in the next fiscal year and beyond. We confirmed that the created database was suitable for developing the subtyping model. In consultation with other research projects, we designed the experimental design, including the implementation period, subject segments, questionnaire items, and types of daily behavioral data (experience sampling and behavioral data) to be obtained. A data collection application was created, and a preliminary survey was conducted after receiving approval from an ethics committee. In the first stage, a screening questionnaire was administered to 50,000 participants, and 2,000 participants were selected based on their responses. In addition, approximately 100 respondents were selected among those who responded to the questionnaire, and a behavior measurement survey was conducted.

Event	Event
Battery level	Sensor: Heart rate
Battery charge	Sensor: Illuminance
Network status (Mobile/ Wi-Fi)	Sensor: Ambient pressure
Wi-Fi AP	Sensor: Proximity sensor
Amount of data (Wi-Fi)	Sensor: Relative ambient humidity
Amount of data (Mobile)	Sensor: Significant motion
OS version	Sensor: Fixed detector sensors
Device name	Sensor: Step counter
Carrier name	Sensor: Step detector sensors
Time	Sensor: Acceleration
Screen status (On/Off)	Sensor: Uncalibrated acceleration
Lock status (On/Off)	Sensor: Acceleration without gravity
Package name	Sensor: Rotation vector
Application Launch/ Quit (Ui-Logger)	Sensor: Rotation vector (Gyroscope)
Touch home button	Sensor: Rotation vector (Geomagnetic)
Power (On/Off)	Sensor: Gravity
Daily usage time (total time of [screen on- screen off])	Sensor: Geomagnetic field sensor
Application anomaly detection	Sensor: Uncalibrated geomagnetic field sensor
Application list	Sensor: Gyroscope
Application Launch/ Quit (Other Application)	Sensor: Uncalibrated gyroscope
Application update	Sensor: Proximity detector sensors
Screen brightness	Sensor: GPS
Volume	Screen touch event
Event that turn to airplane mode	Sleep
Event that press camera button	Usage time details (Array of [screen on- screen off])
Event that change device configuration (orientation, locale)	Walking details (Array of [walking start- walking stop])
Event that connect/disconnect wired headsets	Notification setting status
Event that shutdown device	Sensor list
Sensor: Ambient temperature	Height
Sensor: Heart rate detection	Weight

Table 2: Behavioral measurements

There 3: The contents of the "high" neuroimaging database were examined. In neuroimaging data collection, the type of imaging data and scanning parameters are crucial. We examined these points regarding the quality, scanning time, and other factors of open data. In addition, various preprocessing methods have been proposed day by day. We collected information on these methods and selected appropriate ones. We also determined the database structure in terms of data storage method, preprocessing method, operation method, and model construction. The type of neuroimaging data and the scanning parameters were determined to be compatible with the plan for the subtyping model constructed in Theme 1. We confirm that the dataset will contain sufficient data to build the subtyping model using neuroimaging data by the method for stratifying diseases (Hierarchical supervised/unsupervised learning).

3. Future plans

Begin collecting data for the "high" and "low" categories. The "low" data, questionnaires about personal characteristics, will be conducted online. Collect smartphone logs, daily behavior data, and mood sampling through smartphones and wearable devices over several months. Neuroimaging data, the "high" data, will also be collected to build a high-low mix database, and these data will also be used for subtyping. (TANAKA Saori:ATR, NAKAMURA Hajime:KDDI Research, SAKAI Yuki:XNef))