Mining Human Behavior for Health Promotion

Oresti Banos[1], Jaehun Bang[1], Taeho Hur[1], Muhammad Hameed Siddiqi[1],
Huynh-The Thien[1], Le-Ba Vui[1], Wajahat Ali Khan[1], Taqdir Ali[1], Claudia Villalonga[2], and Sungyoung Lee[1]

***Abstract*— The monitoring of human lifestyles has gained much attention in the recent years. This work presents a novel approach to combine multiple context-awareness technologies for the automatic analysis of people's conduct in a comprehensive and holistic manner. Activity recognition, emotion recognition, location detection, and social analysis techniques are integrated with ontological mechanisms as part of a framework to identify human behavior. Key architectural components, methods and evidences are described in this paper to illustrate the interest of the proposed approach.**

I. Introduction

Diseases linked to lifestyle choices are currently the biggest cause of death worldwide. In fact, diseases of enormous prevalence such as cardiovascular conditions, cancers, chronic respiratory disorders, obesity and diabetes, represent more than 60% of global deceases, half of which are of premature nature. Most of these diseases have been observed to be fairly associated to common risk factors, namely, tobacco and alcohol use, unwholesome diet and physical inactivity. These worrisome facts have been recently reported by the World Health Organization [1], which urges global action against what has been defined as "slow-moving public health disaster." The key elements of this action plan consist of lessening risk factors through enhancing prevention and control mechanisms, and raising public awareness and personalizing healthcare.

The assessment of personal health habits and risk factors is of key importance to understand the particular demands and necessities of each individual, as well as to specialize the services devised to potentially alter one's risk of disease or premature death [2]. People self-reporting mechanisms have been traditionally used to record lifestyle information shared with doctors during occasional checkups and hospitalizations. Similarly, questionnaires have been considered as a tool for inspecting on people's daily conducts. Nevertheless, these techniques are proved to be certainly inaccurate. Leading factors of imprecision relate to the honesty of the participants, their introspective ability, their level of understanding or the wording of questions [3]. Moreover, an increasing lack of interest in the report task is normally experimented due to the required continuous user involvement.

Mechanisms to autonomously and automatically record people conducts are of utter importance to help identify potential unhealthy behaviors. Digital technologies appear in this context as a perfect means to support human self-quantification [4]. Particularly boomed by the growth of wearable and mobile technology, the industry has recently shown a huge interest in the development of applications and systems for tracking health and wellness personal states [5]. These solutions are mostly oriented to the fitness domain, normally supporting the monitoring of primitive user physical routines and the delivery of simple guidelines. Examples of existing commercial solutions are *Fitbit Surge*, *Jawbone Up* or *Garmin Vivofit*, which are composed of instrumented gadgets and companion apps, which provide information based on the measured taken steps or hours of sleep. More prominent health and wellness systems have been shown at the research level, for example, to alert on physical conditions [6] or detect chronic illnesses [7], yet most of them are prototypes or work-in-progress.

The identification of human behavior is a very complex matter that demands the analysis of multiple factors. Likewise, it requires to approach the person observation from various perspectives, including physical, mental and social aspects. Accordingly, current domain-specific solutions are seen to be certainly insufficient to deal with the magnitude of this problem. Instead, more complete platforms combining diverse technologies to infer people lifestyle and provide more personalized services are required. In this direction is devised Mining Minds [8], [9], a novel digital health and wellness framework designed to seamlessly investigate on people's lifestyles, by processing human's daily living data generated through heterogeneous resources. This paper is particularly focused on the description of Information Curation Layer, the core part of the Mining Minds platform which is in charge of intelligently mining people behavior by analyzing person-centered multimodal sensor data.

II. Mining Minds in a Nutshell

Mining Minds is a novel digital framework for personalized health and wellness support, comprising innovative services, tools, and techniques, working collaboratively to investigate on human's daily life data generated from heterogeneous resources. Mining Minds philosophy revolves around the concepts of data, information, knowledge and service curation, which refer to the discovery, processing, adaptation and evolution of both contents and mechanisms for the provision of high quality services. A multilayer architecture is particularly devised for Mining Minds (Fig. 1) motivated by the curation concepts. In a nutshell, the Data Curation Layer is responsible for processing and persisting the data obtained from multimodal sources of user's health and well-

[1]Authors with the Department of Computer Engineering, Kyung Hee University, 446-701 Yongin, Korea {oresti,...,sylee}@oslab.khu.ac.kr
[2]Claudia Villalonga is with the CITIC-UGR, 18015 Granada, Spain cvillalonga@correo.ugr.es

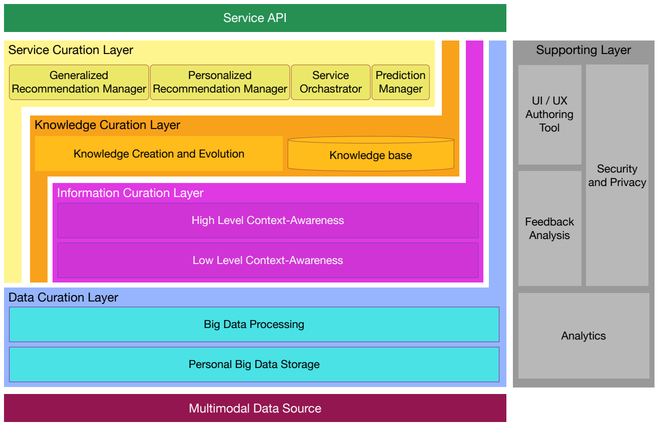

Fig. 1. Mining Minds Platform Architecture.

ness data. The Multimodal Data Source abstractly defines the possible sources, including, but not limited to, data from social networks, questionnaires, wearable biomedical devices or ambient intelligence systems. The Information Curation Layer uses the data processed by the Data Curation Layer to infer low-level and high-level person-centric information, mainly the user context and behavior, and, to some extent, their physical, mental and social state. The Knowledge Curation Layer leverages the information extracted by the Information Curation Layer to nurture and evolve the health and wellness knowledge created primarily by human experts. The Service Curation Layer uses the data, information and knowledge to create advanced health and wellness support services, mostly in the form of smart recommendations. The Supporting Layer enriches all contents and processes in terms of security and privacy and provides analysis of user experience, feedback and trends to guarantee the highest personalization.

III. Information Curation for Behavior Mining

One of the main challenges of Mining Minds refers to the transformation of multimodal data into actual interpretable information. This is performed through the so-called Information Curation Layer (see Figure 2), which encompasses the core technologies devised for the inference and modeling of the user's context. This layer is composed by two main modules, namely, Low Level Context Awareness (LLCA) and High Level Context Awareness (HLCA). LLCA is in charge of converting the wide-spectrum of data obtained from the user interaction with the real and cyber-world, into abstract concepts or categories, such as physical activities, emotional states, locations and social patterns. These categories are intelligently combined and processed at HLCA in order to determine and track context.

LLCA comprises four main components, respectively, Activity Recognizer, Emotion Recognizer, Location Detector and SNS Analyzer. Each LLCA component includes various models operating on different sensing modalities. The decisions delivered by each model are fused for each category, i.e., activities, emotions and locations. The diverse nature of the social data makes the SNS Analyzer a polyvalent component, which outputs can serve for identifying the previous categories apart from supporting the detection of other personal information, like user interests and preferences. The LLCA components require multimodal sensory data in order to operate. Therefore, the Input Adapters subscribe to the Data Curation Layer which dispatches the data after curation, i.e., including relevant metadata like modality, source, user or timestamp. Each Input Adapter determines the particular model to be used depending on the nature of the data and routes the data for its processing. Once the LLCA main components have identified the different categories, the Output Adapters add to them the relevant metadata like the user or the timestamp. The LLCA outputs are given to HLCA for a more comprehensive analysis of the user's context. The Mappers interpret the low-level information and transform it into the corresponding ontological concepts. The Context Recognizer, which is the main component of HLCA, is in charge of building and classifying the high-level context. LLCA and HLCA information is served to Knowledge and Service Curation Layer for the creation of personalized health and wellness recommendations.

A. Activity Recognizer

The identification of the user physical actions is performed through the Activity Recognizer. This component builds on several sensing modalities as they happen to be available to the user, such as inertial sensors, video and audio. The output of this component corresponds to elementary activity categories such as "sitting" , "jogging" or "cycling". Activity recognition models consist of various steps combining signal processing and machine learning techniques to define a specific human activity recognizer, here capable of distinguishing among various commonplace activities. In the implemented version three types of activity recognition models are supported based on the used data source. The first model, [10], builds on body motion data, namely, acceleration, obtained from smartphones embedded sensors. The captured signals are partitioned into windows of three seconds, from which time and frequency features are extracted and further used as input to a gaussian mixture model for the classification process. The second model, [11], also operates on acceleration data, but rather collected through wearable sensors attachable to diverse body parts, including limbs and torso. The information is processed as in the previous case, but here including a sophisticated decision fusion mechanism to unify into a single recognized activity the categories yielded by several support vector classifiers. The last model, [12], processes video streams normally collected through CCTV cameras. Distance transform and principal component analysis are used together with support vector machines and gaussian kernels to recognize some postures and interactions. This model is particularly devised for determining actions that involve various individuals, such as "hugging".

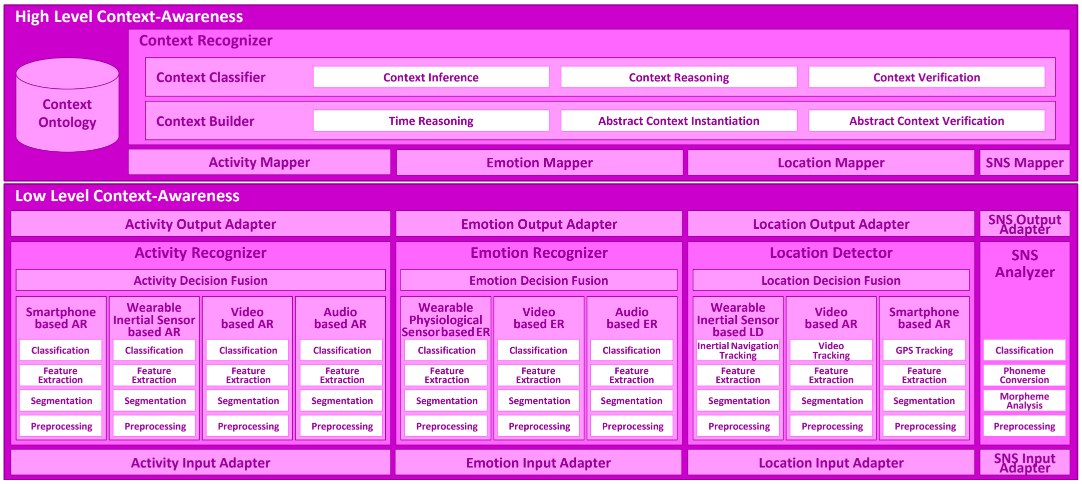

Fig. 2. Information Curation Layer Architecture.

B. Emotion Recognizer

The Emotion Recognizer is defined to infer user emotional states, such as "happiness" or "disgust", by using video and audio data as well as human physiological data. At the moment the system supports three kinds of emotion recognition models. One of the implemented models develops on the smartphone audio data processed during call conversations [13]. Non-speech reduction techniques are employed to differentiate among voice and silences. Mel frequency cepstral coefficients are extracted from sequential frames of 64ms size. Support vector machines and radial basis function kernels are used for the classification task. Anonymization techniques are further considered in order to preserve security and privacy principles. The second model, [14], also represents an audio-based approach but here rather devised for analysis of utterances captured through regular microphones. The model uses low-level descriptors and functionals frequently used in this domain, such as signal energy, pitch or statistical metrics, which are input to a support vector machine. The third model, [15], performs the emotion recognition based on the analysis of video frames normally captured through smartphone or web cameras. The implemented technique consists of an active contour level method for the face detection, a wavelet transform coupled with optical flow for the feature extraction, and a hidden markov model for the classification.

C. Location Detector

The user situation is determined by the Location Detector, which builds on the data collected through indoor and outdoor positioning sensors to specify the exact location or direction of the user. The current model just employs the smartphone GPS sensor to determine the person situation. User-centric points of interest are simply identified by using personalized maps. These maps, that build on top of Google Maps technology, are created by requesting users to provide the location of personal relevant places such as their "home" or "office".

D. SNS Analyzer

The SNS Analyzer processes the information generated by the user during their interactions in regular social networks such as Facebook or Twitter, including posts, mentions, messages, traces and even global trends. From here, personal and general interests, conducts, situation and sentiments may be determined. Concretely, the actual version of this model [16] implements an instant messaging processor, which performs a morphemic and a phonemic analysis of the text. For the learning process, support vector machines are considered in combination with string kernels. Currently, the system operates on top of *KakaoTalk*, a multi-platform texting app widely used in Korea for communication and social networking.

E. Context Recognizer

High-level context is determined by the Context Recognizer, which represents, infers, reasons, categorizes and verifies the user context. The context representation is performed through ontologies, while its classification is done through ontological inference and reasoning. Whenever new information is detected by LLCA, a new instance is created by the Context Builder and categorized into one of the different defined high-level contexts by the Context Classifier. The main concept of the Context ontology is the `Context` class and its subclasses define the different high-level contexts, e.g., `Sleeping`, `HavingMeal` or `OfficeWork`. To model the different low-level information, the `Activity`, the `Location` and the `Emotion` classes have been described, as well as their subclasses, e.g., `LyingDown` or `Running`, `Home` or `Office`, and `Happiness` or `Neutral`. These primitive classes are related to the `Context` class via the object properties `hasActivity`, `hasLocation` and `hasEmotion`. Each `Context` subclass is defined through existential and universal axioms that define the necessary and sufficient conditions of the equivalent anonymous class. For example, the `Sleeping` class is defined as `Context ⊓ (∃ hasActivity`

LyingDown) ⊓ (∃ hasLocation Home) ⊓ (∀ hasActivity LyingDown) ⊓ (∀ hasLocation Home) ⊓ (∀ hasEmotion Neutral) and the descriptive logic for the OfficeWork class is Context ⊓ (∃ hasActivity Sitting) ⊓ (∃ hasLocation Office) ⊓ (∀ hasActivity Sitting) ⊓ (∀ hasLocation Office) ⊓ (∀ hasEmotion (Anger ⊔ Neutral ⊔ Happiness)).

IV. Models Performance

The implemented models are trained and evaluated in existing datasets for each particular domain as described in the aforementioned works and summarized in Table I. The accuracy results yielded by each of these models are shown in Figure 3. Concretely, the smartphone-based AR (S-AR) recognizes the categories Resting, Walking, Jogging, Riding the Bus, Taking the Subway with a 85.4% accuracy. More activities can be identified through the wearable-based AR (W-AR), which attains up to 99.5% accuracy when detecting Lying down, Standing, Sitting, Walking, Running, Climbing stairs, Cycling, Stretching, and Strength-training. The video-based AR (V-AR) yields a 92.4% for the recognition of Standing, Walking, Hugging, Shaking, Pointing, Pushing, Kicking, and Punching. Diverse kind of emotions can be detected by the implemented models. The smartphone-based ER (S-ER) supports the detection of Happiness, Anger, Sadness, Fear with a 89.1% accuracy. These emotions plus Disgust, Boredom and Neutral can be recognized by the audio-based ER (A-ER), which yields a 84.6% accuracy. A performance of 99.1% is attained by the video-based ER (V-ER), which recognizes the previous emotions and also Surprise. The implemented SNS analyzer, here particularly devised for emotion recognition purposes, detects Happiness, Anger, Sadness, and Neutral with a performance of 78.0%. The recognition of the high-level context for all potential combinations of the low-level categories has been satisfactorily tested in Protégé using the HermiT reasoner.

TABLE I

Characteristics of the datasets used for the evaluation.

Modality	Subjects	Ages	Data Type	Data Amount
S-AR	10	25-29	Acceleration + GPS	25.5 hours
W-AR	10	25-29	Acceleration + GPS	25.5 hours
V-AR	7	20-28	Video sequences (RGB)	0.38 hours
S-ER	8	20-30	Audio (WAV)	1.11 hours
A-ER	10	21-35	Audio (WAV)	1.19 hours
V-ER	100	18-30	Images (RGB)	50000 images
SNS	7	20-30	MMS	2000 messages

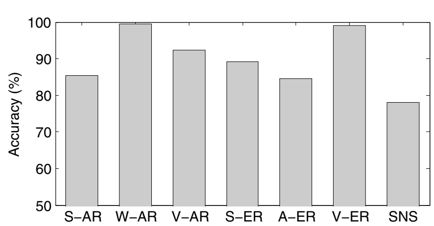

Fig. 3. LLCA models performance. S-AR=Smartphone-based AR; W-AR=Wearable-based AR; V-AR=Video-based AR; S-ER=Smartphone-based ER; A-ER=Audio-based ER; V-ER=Video-based ER; SNS=SNS Analyzer.

V. Conclusions

The analysis of people lifestyles is of much interest to prevent high prevalence diseases. A novel approach for identifying human behavior in a holistic manner has been presented in this paper. The proposed system combines low level context-awareness models devised to recognize activities, emotions and location primitives from multimodal person-centric data. This information is further considered to sophisticatedly infer high level context categories by using ontological reasoning techniques. The proposed system partakes of Mining Minds, an innovative digital framework for personalized healthcare and wellness support. Undergoing work includes the completion of the implementation of the proposed architecture as well as the evaluation of the whole platform on a large scale testbed.

Acknowledgment

Work funded by the Industrial Core Technology Development Program, Korean Ministry of Trade, Industry and Energy (MOTIE), under grant number #10049079.

References

[1] WHO, "Global status report on noncommunicable diseases 2014," World Health Organization, Tech. Rep., 2014.

[2] R. E. Soler et al., "A Systematic Review of Selected Interventions for Worksite Health Promotion: The Assessment of Health Risks with Feedback," *Am. J. Prev. Med.*, vol. 38, no. 2, pp. 237 – 262, 2010.

[3] R. J. Shephard, "Limits to the measurement of habitual physical activity by questionnaires," *Br. J. Sports Med.*, vol. 37, no. 3, pp. 197–206, 2003.

[4] M. Swan, "Health 2050: the realization of personalized medicine through crowdsourcing, the quantified self, and the participatory biocitizen," *J. Pers. Med.*, vol. 2, no. 3, pp. 93–118, 2012.

[5] A. C. Powell et al., "In search of a few good apps," *J. Am. Med. Assoc.*, vol. 311, no. 18, pp. 1851–1852, 2014.

[6] O. Banos et al., "Physiodroid: Combining wearable health sensors and mobile devices for a ubiquitous, continuous, and personal monitoring," *Sci. World J.*, vol. 2014, no. 490824, pp. 1–11, 2014.

[7] J. J. Oresko et al., "A smartphone-based platform for real-time cardiovascular disease detection via electrocardiogram processing," *IEEE Trans. Inf. Tech. Biomed.*, vol. 14, no. 3, pp. 734–740, 2010.

[8] O. Banos et al., "An Innovative Platform for Person-Centric Health and Wellness Support," in *Proc. Int. Conf. on Bioinf. & Biomed. Eng.*, 2015.

[9] O. Banos et al., "Mining minds: an innovative framework for personalized health and wellness support," in *Proc. Int. Conf. on Perv. Comp. Tech. for Healthcare*, 2015.

[10] M. Han et al., "Comprehensive context recognizer based on multimodal smartphone sensors," *Sensors*, vol. 12, no. 9, pp. 588–605, 2012.

[11] O. Banos et al., "Human activity recognition based on a sensor weighting hierarchical classifier," *Soft Computing*, vol. 17, pp. 333–343, 2013.

[12] T. Huynh-The et al., "A novel human behavior recognition framework based on a multi-level assessment model," in *Adv. Multim. Comp.*, 2015.

[13] J. Bang et al., "Call speech emotion recognition for emotion based services," *J. KIISE, Soft. and App.*, vol. 41, no. 3, pp. 208–213, 2014.

[14] B. V. Le et al., "Hierarchical emotion classification using genetic algorithms," in *Symp. Inf. and Com. Tech.*, 2013.

[15] M. H. Siddiqi et al., "Facial expression recognition using active contour-based face detection, facial movement-based feature extraction, and non-linear feature selection," *Multimedia Systems*, 2014.

[16] K. H. Han et al., "Short text emotion recognition based on complex keywords," in *Korean Inf. Proc. Soc.*, 2013.