Glassbot: Personalized Wearable Agents Learning from Everyday Human Behaviors

Sang-Woo Lee*, Chung-Yeon Lee*, Dong-Hyun Kwak† and Byoung-Tak Zhang*†
*School of Computer Science and Engineering
†Interdisciplinary Program in Neuroscience, Seoul National University
Seoul National University
Seoul, 08826, South Korea
E-mail: {slee, cylee, dhkwak, btzhang}@bi.snu.ac.kr

***Abstract*—This paper introduces Glassbot, the agent on glass-type wearable devices with camera and audio sensors. We want to train Glassbot continuously in a wearable device by rapidly adapting deep neural networks from sensor data streams of user behaviors. In this paper, we describe our early works on dataset and online learning algorithms for Glassbot. We also discuss the future work of predictive learning and scheduling on Glassbot.**

I. Introduction

It is essential for building human-aware intelligent agent that can learn from human behaviors in the real world. Glass-type wearable devices, including Google Glass and Narrative Clip, have merits on this problem. It is noticeable that these devices can see and hear what the device user sees and hears; this property differentiates them from the classical agents in personal computers or the smartphones.

We introduce the concept of Glassbot, the agent on Glass-type wearable devices. We believe Glassbot is promising for the next-generation digital assistant, because video stream collected through wearable devices have a wealth of information, and recent deep learning technique can handle this to make the agent smarter. To implement Glassbot, we are paying attention to three following technique; 1) Acquisition of knowledge from real stream data, 2) continuous learning of knowledge, 3) predictive learning over daily life. Potential applications of Glassbot is as follows; 1) Context recognition, 2) Scheduling, 3) question answering on the context, 4) automatic diary writing.

We are currently doing one early work of Glassbot. In this paper, we describe our early works on dataset and online learning algorithms for Glassbot. We also discuss on the future work of predictive learning and scheduling on Glassbot.

II. Lifelog Dataset

The environment Glassbot encounters has two properties. First, there are hidden high-level contexts in the raw sensory stream, for example, an first-person view video stream recorded during a dating includes various types of high-level contexts, although the data is only a stream of pixels and audio signals. Second, the data streams are often non-stationary, for example, the life patterns of the weekday and weekend are different.

Fig. 1. Lifelog dataset collected through Google Glass

TABLE I
Statistics of the lifelog dataset of each participant

	Instances (sec/day)		Number of class		
	Training	Test	Location	Sub-location	Activity
A	105201 (13)	17055 (5)	18	31	39
B	242845 (10)	91316 (4)	18	28	30
C	144162 (10)	61029 (4)	10	24	65

TABLE II
Top-5 classes in each label of the lifelog dataset

Location	Sub-location	Activity
office (196839)	office-room (182884)	working (204131)
university (147045)	classroom (101844)	commuting (102034)
outside (130754)	home-room (86588)	studying (90330)
home (97180)	subway (35204)	eating (60725)
restaurant (22190)	bus (34120)	watching (35387)

We collected Google Glass lifelog dataset recorded over 46 days from three participants [1] (Figure 1). The 660,000 seconds of the first-person view video stream data reflects the behaviors of participant including the indoor activities, such as ‘studying in the library’ or ‘watching TV in the house’, and the outdoor activities, such as ‘walking on the road’ or ‘waiting for the arrival of the bus’. The participants were asked to notate what they were doing and where they were in real-time by using a life-logging application installed on their mobile phones. In this study, location, sub-location, and daily activity

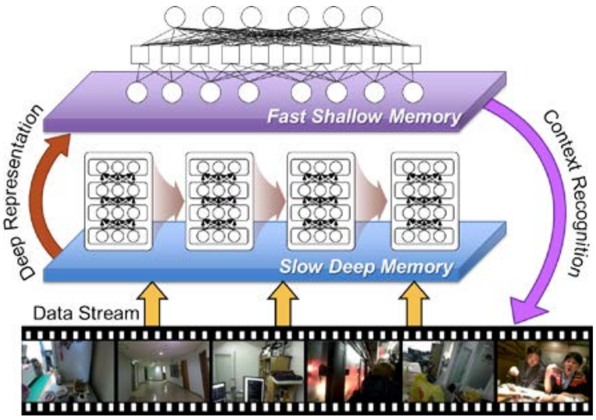

Fig. 2. Dual memory architecture (DMA) for lifelong learning of Glassbot agent

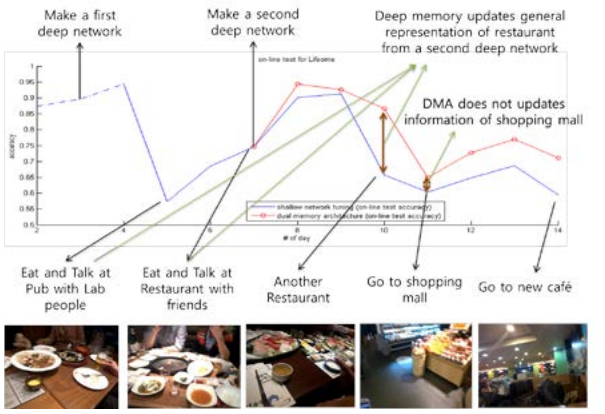

Fig. 3. The story of daily life in the lifelog dataset the effectiveness of DMA on this scenario.

are labeled. A frame image of each second is used as one instance. Table I summarizes the dataset statistics and Table II presents the distribution of the five major classes in each class type.

III. Lifelong Learning

We are interested in adapting the context-aware activity recognizer continually and rapidly from human behaviors gathered through wearable devices. To treat these properties, two algorithmic techniques are required. First, the deep learning method is necessary to handle raw-level data efficiently [2]. Second, an online learning algorithm is required to keep track of fast-changing life patterns of user behavior [3].

To address these isseus together, we utilized the concept of complementary learning systems (CLS) theory a framework that suggests a dual learning system structure in the brain [4]. According to the CLS theory, there are two critical areas in the brain that affect online learning: the neocortex and hippocampus, which complement each other's functionality.

Inspired by the CLS theory, we propose a dual memory architecture (DMA) (Figure 2) [1]. The DMA trains two memory structures: one is an ensemble of DNNs, and the other consists of a shallow network that uses hidden representations of the DNNs as input. These two memory structures are designed to use different strategies. The ensemble of DNNs learns new information to adapt its representation to new data, whereas the shallow network aims to manage non-stationary distribution and unseen classes more rapidly.

In our experiments, the proposed DMA outperformed other online learning methods on two datasets: the CIFAR-10 image-stream dataset and the lifelong dataset. Comparative learning methods often failed to keep the information of old data. Figure 3 illustrate the story of daily life in the lifelog dataset and the effectiveness of DMA on this scenario. As DMA can learn deep representation of new events, DMA outperforms comparative results and robustly learns from non-stationary environment.

IV. Predictive Learning and Scheduling

We want to learn the daily story of the user and predict future daily activity based on the knowledge of the daily story. The typical application of predictive learning is scheduling. By understanding the current context and extracting a series of daily schedule from lifelog dataset, digital assistant can recommend and predict user behaviors [5]. Predicted user behaviors can be used for question answering about user's schedule or multimodal context [6].

Currently, we collect more data for simulating the scheduling service of Glassbot. In addition to video data, the schedule of the user will be augmented. The continuity of the dataset will also be enhanced.

Acknowledgment

This work was supported by the Naver Corp. and partly by the Korea Government (IITP-R0126-16-1072-SW.StarLab, KEIT-10060086-HRI.MESSI, KEIT-10044009-RISF, BMRR-UD130070ID).

References

[1] S.-W. Lee, C.-Y. Lee, D. H. Kwak, J. Kim, J. Kim, and B.-T. Zhang, "Dual-memory deep learning architectures for lifelong learning of everyday human behaviors," in *Twenty-Fifth International Joint Conference on Artificial Intelligencee*, 2016, pp. 1669–1675.

[2] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems 25*, 2012, pp. 1097–1105.

[3] P. Ruvolo and E. Eaton, "Ella: An efficient lifelong learning algorithm," in *Proceedings of The 30th International Conference on Machine Learning*, 2013, pp. 507–515.

[4] J. L. McClelland, B. L. McNaughton, and R. C. O'Reilly, "Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory." *Psychological review*, vol. 102, no. 3, p. 419, 1995.

[5] E.-S. Kim, K.-W. On, B.-T. Zhang, and C. R. A. I. Center, "Deepschema: Automatic schema acquisition from wearable sensor data in restaurant situations," in *Twenty-Fifth International Joint Conference on Artificial Intelligence*, 2016, pp. 834–840.

[6] J.-H. Kim, S.-W. Lee, D.-H. Kwak, M.-O. Heo, J. Kim, J.-W. Ha, and B.-T. Zhang, "Multimodal residual learning for visual qa," *Advances in Neural Information Processing Systems 29*, 2016.