Achieving Personal Agents and Pursuing Collaboration

General Manager of
Service Innovation Department
Tomoyoshi Ohno

In 2013, NTT DOCOMO announced its goal of becoming a Smart Life Partner to better serve its customers. Since then, we have undertaken the development of "personal agents" that seamlessly fit into people's lives as a key objective. A personal agent has a close relationship with the user providing 24/7 support to enrich that person's life in all sorts of scenarios. Rather than just an agent acting on behalf of a person, a personal agent helps the user in deciding what action to take in a particular situation.

There are three main elements to achieving personal agents: technology, big data, and engineers. In addition to establishing the requisite technology, it is also important that big data related to services be collected in some way and that engineers use that technology and data to develop actual personal agents.

As described in the talk "Personal Agents and Robot Initiatives" given at the 2015 DOCOMO R&D Open House, we have been developing three core technologies for achieving personal agents. These are natural language processing technology for achieving "ear and mouth" (aural and vocal) functions, image recognition technology for achieving an "eye" (visual) function, and user understanding/behavior-anticipation technology for achieving a "heart" (thoughtful) function. A "core technology" does not mean only in-house developed technology—it can also mean outside technologies that we can incorporate as needed and accumulated know-how as well. For this reason, we are proactively using open source products.

The basis of these core technologies is machine learning targeting big data, and in particular, deep learning*1, which is fast becoming the dominant form of machine learning. Deep learning is also included in these core technologies. By the way, Google's computer Go program AlphaGo™*2 incorporates deep learning, and it is famous for beating a professional Go player for the first time in 2015. In the case of games, playing rules and what constitutes winning or losing are clearly defined, so it is possible to adopt techniques for automatically learning winning methods by having two computers compete against each other. Given abundant computer power, it is relatively easy to uncover winning strategies. However, in the case of personal agents that we are developing, replies to user utterances proceed against the background of overall context, the personal relationship between the two speakers, etc., and there may not necessarily be only one correct reply. As a result, the application of deep learning is currently at the research stage and it is apparent that the development of practical personal agents is difficult. On the other hand, in the fields of speech recognition and image recognition in which labeled data are easy to obtain, the application of deep learning has entered the practical stage and has reached the point at which computers are coming to exceed the recognition abilities of human beings.

Since the appearance of deep learning, it has become possible to obtain superior results in an efficient manner by applying machine learning to large volumes of data compared with the conventional problem-solving approach using human knowledge and experiences. The issue here is how to go about acquiring such large volumes of data for learning purposes. We can look to natural language processing technology, the core technology of NTT DOCOMO's Shabette-Concier voice agent service launched in March 2012 (more than 36 million packages installed and more than 1.4 billion accesses as of the end of April 2016), for an answer. In the course of providing this service, this technology iteratively performs data collection, data analysis, and reflection of analysis results, thereby acquiring huge volumes of data with which machine learning can be performed in parallel. In this way, speech recognition performance and Q&A performance are being dramatically improved compared with conventional methods. In other words, to improve service performance and quality using deep learning, it is important to provide the service quickly and repeat the above cycle continuously to acquire a massive amount of data.

The widespread use of devices that can support this collection of big data is also important for achieving personal agents. Connecting the many things that surround us to the cloud will enable personal agents to learn about the wide variety of situations that people come to be in. We can envision the ultimate personal agent that can provide support for even the future behavior and health of the user.

From here on, we plan to actively pursue +d*3 (collaboration) that makes use of the core technologies needed for achieving personal agents. We have already partnered with TOMY in the joint development of OHaNAS®*4, an interactive conversational toy released in October 2015, using a natural-language dialog platform provided by NTT DOCOMO. The +d initiative can be thought of as the meeting of minds between two professional enterprises. OHaNAS is a typical example of this approach. Putting each other's strengths to work creates new value!

At present, we are promoting +d with taxi companies too. Our goal here is to predict the demand for taxis 30 minutes into the future by combining NTT DOCOMO demographic statistics with a taxi company's operational data plus other external data such as area characteristics and weather data and performing deep learning. I believe that predicting the future in this way is precisely an application of artificial intelligence using our core technologies that we have been refining continuously to achieve personal agents.

Going forward, we seek to create new value by leveraging the core technologies for achieving personal agents and collaborating with other companies on the mutual use of big data. In this way, we hope that the world I have described above will one day become commonplace. I look forward to the challenges that this endeavor brings.

*1 Deep learning: Machine learning using a neural network with a many-layer structure.
*2 AlphaGo™: A trademark or registered trademark of Google, Inc.
*3 +d: Name of NTT DOCOMO initiative for creating new value together with partner companies.
*4 OHaNAS®: A registered trademark of TOMY Company, Ltd.