ATLANTIS PRESS

Advances in Social Science, Education and Humanities Research, volume 207

3rd International Conference on Creative Media, Design and Technology (REKA 2018)

Internet of Speech: A Conceptual Model

Sasa Arsovski
Research Team
Imagineering Institute
79200 Nusajaya, Johor
City, University of London
London EC1V 0HB, UK

Hasmik Osipyan
Research Team
Imagineering Institute
79200 Nusajaya, Johor
City, University of London
London EC1V 0HB, UK

Adrian David Cheok
Research Team
Imagineering Institute
79200 Nusajaya, Johor
City, University of London
London EC1V 0HB, UK

Idris Oladele Muniru
Research Team
Imagineering Institute
79200 Nusajaya, Johor
City, University of London
London EC1V 0HB, UK

Abstract The interest to the concept of the Internet of Things is growing rapidly in the scenario of wireless telecommunications. The main idea behind this concept is the usage of the Internet as a common platform for connection of billions of Things around us. The main missing attribute in the Internet of Things concept is still the lack of interaction techniques with or between these Things. Hence, the Internet of Conversational Things makes this concept to happen and brings to the idea of direct interaction with all these devices. In this paper, we will describe and analyse the techniques of Internet of Conversational Things concept by examining the state-of-the-art approaches. At the end, we will come up with the new concept of Internet of Speech that will provide more connectivity between Things.
Keywords : Chatbots,Internet of Conversational Things, Internet of Speech

INTRODUCTION

The Internet of Things (*IoT*) is the general concept that allows Things (devices) to be controlled or interconnected remotely through the existing networking systems [1], [2]. In *IoT*, "Things" term can include a wide variety of devices that are using built-in sensors, monitoring devices, conversational agents, etc. Diversity of devices is conditioned by various domains that *IoT* is covering, such as healthcare [3], transportation [4], automation [5], etc. Whatever the domain, it is obvious that *IoT* use cases could broaden to nearly every aspect of our lives. In that scenario, as the *IoT* continues to develop more and more, the potential is estimated by the combination of existing interactive techniques, such as conversational agents. This idea brings to the general concept of Internet of Conversational Things (*IoCT*) where all components of the general system can interact with each other through the Internet in a real time manner.

In the world of IoCT, several research works focus on the direct interaction with Things (Humans-to-Things), receiving the information from the physical world (Things-to-Humans) and automation of tasks without human interaction (Thingsto-Things). Based on these concepts, interactive Things can create a valuable environment for interaction with users and define the communicative interface with the entire world.

- *Humans-to-Things or Things-to-Humans:*

The interfaces and techniques of interaction between Humans and Things are very crucial for *IoCT* development. For gesture-based interaction, recent research focuses on using the well-known wireless network signals, such as WiFi, Bluetooth, etc. The other direction for direct interaction with *Things* is Body Area Network (*BAN*) usage, which can extend the operation of the infrastructure network near to the user body
[6].

- *Things-to-Things:*

This type of interaction refers to the exchange of information and action between Things without Humans assistance. One of the famous applications for such interaction is remote monitoring [7]. This system includes the information obtained through the sensors, Radio Frequency Identification (RFID), communication channels and software for decision-making. Right now, the Things-to-Things interaction is very task specific, and the future of IoCT is going in the direction of making standardized platform for all types of interconnections.

There are still many open issues in this research direction that limit the full exploitation of *IoCT*. Some of these challenges are very important for further development of*IoCT*. Though, before going forward to solve the confidentiality issues (e.g., security, privacy) or Big Data issues (e.g., storage management), there is still a real issue of creating a fully learning environments with the possibility of real time *ThingsHumans-Things* interactions. In this context, the learning environment should enable Things to understand the meaning and context of the messages coming from the *Humans*.

In this paper, we will discuss the techniques and future directions of *IoCT* and will come up with the new concept of Internet of Speech (*IoS*). Although, for having a full interactive system speech and gestures are very important, in this research paper, we will only focus on the speech. Hence, the key novelty proposed in this paper is the new concept of *IoS*.

The rest of the paper is organized as follows. Sections 2 reviews the main concepts and technologies of *IoCT*. Section 3 reviews state-of-the-art *IoCT* approaches and applications in different domains along with the domains that are likely to drive the *IoCT* in the near future. Finally, we do the discussion and come up with a new concept of *IoS* in Section
4. Conclusion is made in Section 5.

IOCT TECHNIQUES

Technologies that provide connectivity of *IoT* can be grouped into three categories:

- Technologies that allow "Things" to acquire background information,
- Technologies that enable the Things to process information about the context,

- Technologies that enable improvement of security and privacy.

The combination of *IoT* technologies, such as real-time communication and localization with the Conversational Agent (*CA*) technologies allows us to transform everyday "Things" into "Smart Objects" that can understand and react to their environment. This transformation brings to the concept of *IoCT*. The next step of defining the architectural principles of smart objects is the use of cloud speech recognition and speech synthesis technologies that allow increasing the interactivity of *IoCT*, and raise the level of interaction between people and smart objects.

CAs also can be grouped into several categories based on their design and application areas. Based on design, *CAs* are categorized into two groups:

- *Retrieval models* use predefined templates to select the appropriate response based on the context of input. These systems do not create any new text answers. System just chooses the response based on the set of rules and on the basis of recognizing lexical form of the input text that was constructed from a fixed predefined set of rules [8].
- *Generative models* do not rely on predefined answers. These models create new responses from scratch. They are typically based on the techniques of Machine Translation. However, instead of translating from one language to another, they translate the input text into an output text (response). These neural conversational models are built with Deep Machine Learning technologies such as Seq2Seq LSTM Neural Network [9]. Seq2seq learning uses recurrent neural networks to map variable-length input sequences to variable-length output sequences.

Based on the application area *CAs* are grouped into the following groups:

- *General CAs* are designed to work in the open domain. Social networks, such as Twitter, Facebook, etc. are examples of open domains. A large number of topics along with the need of *CAs* to have a certain amount of knowledge of the world to create reasonable answers make *General CA* construction more difficult.
- Task-oriented *CAs* are designed for a closed domain. Technical support for customers or shopping assistants are examples of the closed domain problem. *CAs* have the task to accomplish a specific goal, and they need to fulfil their specific task as efficiently as possible. Of course, users can direct the conversation to another topic but, unfortunately, the *task-oriented CA* cannot expect to generate answers that are outside the predefined domain. In order to generate reasonable response, *CA* should analyse the linguistic and physical context of question. In the long dialogues, it is necessary to store the previous dialogues and follow the information that is exchanged. *Task-oriented CA* usage introduces different types of commercial applications that are enhanced with speech processing.

There are several technologies available for creating a general *CAs*. According to our research, AIML 2.0 [10] is the one of the basic techniques used in the common *Task-oriented CA*. AIML 2.0 specification meets the needs and technical requirements of today's pervasive computing, but keeps AIML as simple as possible, especially for non-developers. In this paper, based on the described specification we propose a threelayer architecture model that is described in section 4.

RELATED WORKS

There is a lot of research in the area of *IoCT* in different applications from various domains starting from entertainment until healthcare. In this section, we will review the state-ofthe-art approaches and innovations in this developing area.

Recently, *CAs* become very popular in different social networking systems. One of the interesting direction in the networks is recommendation systems, where virtual agents are able to give advice depending on the context. For example, in [11], authors proposed new socially embedded search engine, SearchBuddies, that automatically answers the posted question statuses of the Facebook social network. This engine may be able to answer the questions in the way that would be more relevant than those provided through typical web searches. The authors showed that these CAs are useful approach for social and algorithmic search in social networks. There are also many research in the direction of creating real intelligent agent in different social networking systems (e.g., Facebook, Twitter) that will be able to behave as a human and get as much followers as possible [12], [13].

Computationally intensive healthcare interventions have been identified as one of the most powerful techniques to reduce the consequences of the main global health challenges. In this context, more effective interaction techniques are required for humans to interact with the deliverables in a very natural way. Hence, CAs are also used as a key element for effective communication in healthcare, especially for support. One of the nice examples of such virtual assistant is "SimSensei kiosk" [14] developed for healthcare decision support. It is able to engage users in 15-25 minutes interaction to identify potential distress indicators. The results showed that through face-to-face interaction, the user feels more comfortable talking and sharing the needed information with the virtual agent vs to real humans. For first aid diseases treatment the *DocBot* is implemented that is shown to be more interactive and user-friendly [15]. One of the recent technologies is the voice-enabled elder care robot, IBM Multi-Purpose Elder Care Robot Assistant (*IBM MERA*) [16], that was developed to provide assistance to the elderly and their caregivers. This system leverages on the speech recognition, natural language processing, and vital signs monitoring and gesture recognition capabilities to facilitate more natural Humans-to-Things interactions. Next step towards the development of this specific domain should be the socially assistive *CAs* that will improve natural and inspiring interaction.

The technologies in Home System Network (*HSN*) are growing rapidly by bringing the idea of next generation smart houses into the reality. The general model of this concept is

based on making interconnections between all house components via different appliances or sensors. There is already a trend to make smart houses based on the voice controlling model where you can speak with different objects in your house and get some feedback [17], [18].The *Amazon Echo* and *Google Home*, both though not fully conversational, have also shown how the smart home experience can become more natural and interactive through speech. *GENIO*, a more believable *CA* for home automation, was described in the work [19]. Here, the user can interact with their home appliances and request services and functionalities via much natural dialog.

IoCT is not only entered the domains such as smart environment, healthcare or entertainment; it is also covering domains such as transportation [20], [21], education [22], industry [23], [24], etc. However, we are not going to review all the domains separately in the scope of this paper. Based on our research, we have seen that one thing which is yet missing in all domains is the common prototype of connectivity between *Things*.

CONCEPTUAL MODEL AND DISCUSSION

IoT is a concept that describes the wireless and wired *Things* that are connected to the Internet for exchanging information. The ability of communication and cooperation of these *Things* for creating new services brings to the concept of *IoCT*.

IoCT leverages the huge connectivity provided by the ubiquitous devices to meet the demand of social, technological, political, and economic requirements. The *IoCT* ecosystem today is no longer people talking to people, nor is it about people accessing information; rather it is about using machines to talk to other machines on behalf of humans. This new concept currently uses a bewildering array of incompatible communication protocols that pose many interaction problems with information exchange, *Things-Humans-Things* communication, and cooperative event processing among different connected entities. Our state-of-the-art research shows the diversity of applications of *IoCT* that covers different domains of our daily life. Compared to *IoT*, that seems covered approximately all domains, *IoCT* is yet growing and trying to bring the IoT into the new level of connectivity. There are still several domains, such as software engineering, infrastructure management, manufacturing, etc. that luck *Things-HumansThings* communication. Even the applications based on the *CA* technology yet need the improvement in the communication methods. For example in "Education" domain where IoCT is already studied for several years, there is no yet the *CA* that has the capabilities of virtual teacher including full interactiveness, emotions along with gestures, etc. Therefore, there is an urgent need for a new approach of interaction that will cover all domains and will aim to unify several standards at the interface layer of the *IoCT* architecture. Speech is inherently qualified as the universal interface for such interactivity in the *IoCT* ecosystem.

Speech is universal and ubiquitous; it is more natural, faster and does not limit the situation of use. These characteristics coupled with the advances in Voice Computing (*VC*) that include Natural Language Processing (*NLP*), speech recognition, as well as growth in artificial intelligence, is a major indication that the future interface of interaction in the *IoCT* ecosystem is speech. Embedding *CAs* into different *Things* around us makes the interaction more natural and effortless. The Chatbot revolution and the gradual preference of conversational search over textual search will further drive the deployment of a real *CAs* in different domains. Integration of above-mentioned characteristics can also facilitate dynamic knowledge interaction. Current efforts in creating a more efficient and effective network of connected *Things* should be geared towards developing refined and capable systems that are able to interact with people, not just by reacting to schematic request but through dialog or conversation mainly based on the *IoS* concept.

In order to bring the concept of IoS to reality and convert everyday objects to more conversational Things or "Smart Objects" we propose three-layer architectural model (Figure. 1.). In the first layer, everyday object will be extended with wireless Internet connection and with HTML 5 supported browser. This extension can provide communication functionality and connect "Smart Objects" with other levels of the proposed architecture. In the second layer of the *IoCT* architecture, we propose implementation of the web based Task-oriented *CA* built with AIML 2.0 technology. This *CA* will be connected with large amounts of organized knowledge that is available on the Internet. This layer represents a "brain" of "Smart Objects".

The third layer introduces Cloud Speech recognition and speech synthesis functionalities to *CA*, hence, bringing the technical part of general *IoS* concept to reality. Cloud Speech API-s enable easy integration of speech recognition technologies into *CA*.

The proposed architectural principles of "Smart Objects" "Things" as a part of *IoCT* will raise the level of not only *Humans-to-Things* but also Things-to-Things interaction. This new type of *Things-Humans-Things* interaction characterizes dialogue between the participants (*Humans* and "Smart Objects"*Things*), connected to the Internet, using a *CA* technology. We strongly believe that proposed three layers will be enough to realize the suggested concept of *IoS*.

However, the creation of really interactive system where all Things will be interconnected by some type of the speech has yet a lot of open questions. Do Things need to use the same natural language as *Humans*? Maybe it will be more productive and faster to implement *Things-to-Things* interaction in a different manner than *Things-to-Humans* and use the wrappers to convert these interaction methods on the first layer. Is it possible that one type of *Things* learn from the other type of *Things* in interactive mode? On the other hand, is there a way to extract whole knowledge of *CAs*? The possibility of extraction whole knowledge will ultimately save the time, cost and provide an efficient way to multiply

knowledge over the Internet. It can provide the model that based on the machine learning techniques will be trained with previous knowledge. Hence, the *CAs* will be able to learn from each other just by posing the different type of questions on the Internet or to *Humans*. The realization of the *IoS* concept by answering the mentioned questions will certainly affect the way we live and work. By applying the proposed concept, Things that converted to the "Smart Objects" will obtain a distinctive personality, intelligence, and decision-making ability. The fact, that *Things* can communicate with each other and with the *Humans* in some general way enables unsupervised learning and knowledge multiplying different opportunities. In our future work, we are going to focus on the proposed threelayer architecture by answering the posed questions.

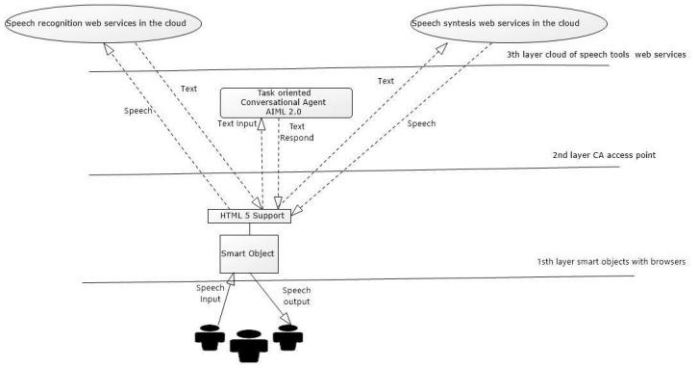

Figure. 1. Recommended architectural model

CONCLUSION

The use of *Things-Humans-Things* interaction is growing exponentially making the *IoCT* more popular. These interactions are organized through different *CAs/Robots* and aimed to provide different functionalities in different fields. In that context, *IoCT* is becoming the system of virtually endless opportunities and connections.

In this paper, we described the main concept and techniques of *IoCT* along with the state-of-the-art approaches and applications. Although it is quite new concept, it already covers most of the important domains of our life and continues progressing. We also discussed their advantages/disadvantages along with their future directions that can support other researchers in their next studies within this multidisciplinary area. We came up with the new concept of *IoS* where the general platform of interconnection will be developed to provide more connectivity between "Smart Things" through some type of general speech/protocol. We strongly believe that this concept can bring to the new machine learning technique that will be above pattern matching solution. In this new platform, *Things* should be able to learn not only from *Humans* but also from each other using the new digital intelligence that could be even out of the *Humans* understanding.

ACKNOWLEDGMENTS

This research was funded by the Khazanah Nasional Berhad Malaysia. We are grateful for this support.

REFERENCES

[1] J. Gubbi, R. Buyya, S. Marusic, and M. Palaniswami, "Internet of things (iot): A vision, architectural elements, and future directions," *Future generation computer systems*, vol. 29, no. 7, pp. 1645–1660, 2013.

[2] A. Whitmore, A. Agarwal, and L. Da Xu, "The internet of thingsa survey of topics and trends," *Information Systems Frontiers*, vol. 17, no. 2, pp. 261–274, 2015.

[3] S. R. Islam, D. Kwak, M. H. Kabir, M. Hossain, and K.-S. Kwak, "The internet of things for health care: a comprehensive survey," *IEEE Access*, vol. 3, pp. 678–708, 2015.

[4] L. Da Xu, W. He, and S. Li, "Internet of things in industries: A survey," *IEEE Transactions on industrial informatics*, vol. 10, no. 4, pp. 2233–2243, 2014.

[5] C. Gomez and J. Paradells, "Wireless home automation networks: A survey of architectures and technologies," *IEEE Communications Magazine*, vol. 48, no. 6, 2010.

[6] E. Jovanov and A. Milenkovic, "Body area networks for ubiquitous healthcare applications: opportunities and challenges," *Journal of medical systems*, vol. 35, no. 5, pp. 1245–1254, 2011.

[7] A. J. Jara, M. A. Zamora-Izquierdo, and A. F. Skarmeta, "Interconnection framework for mhealth and remote monitoring based on the internet of things," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 9, pp. 47–65, 2013.

[8] D. Britz, "Deep learning for chatbots, part 1–introduction," 2016.

[9] I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to sequence learning with neural networks," in *Advances in neural information processing systems*, 2014, pp. 3104–3112.

[10] R. Wallace, "Aiml 2.0 working draft," 2013.

[11] B. J. Hecht, J. Teevan, M. R. Morris, and D. J. Liebling, "Searchbuddies: Bringing search engines into the conversation." *ICWSM*, vol. 12, no. 138-145, p. 2, 2012.

[12] S. M. Rodrigo and J. G. F. Abraham, "Development and implementation of a chat bot in a social network," in *Information Technology: New Generations (ITNG), 2012 Ninth International Conference on*. IEEE, 2012, pp. 751–755.

[13] T. Hwang, I. Pearce, and M. Nanis, "Socialbots: voices from the fronts, interactions, v. 19 n. 2," *March+ April*, 2012.

[14] D. DeVault, R. Artstein, G. Benn, T. Dey, E. Fast, A. Gainer, K. Georgila, J. Gratch, A. Hartholt, M. Lhommet *et al.*, "Simsensei kiosk: A virtual human interviewer for healthcare decision support," in *Proceedings of the 2014 international conference on Autonomous agents and multiagent systems*. International Foundation for Autonomous Agents and Multiagent Systems, 2014, pp. 1061–1068.

[15] A. K. Tripathy, R. Carvalho, A. Puthenputhussery, N. Chhabhaiya, and B. Anthony, "Mediassistedgesimplifying diagnosis procedure & improving patient doctor connectivity," in *Technologies for Sustainable Development (ICTSD), 2015 International Conference on*. IEEE, 2015, pp. 1–6.

[16] (2016) Ibm develops a voice-enabled eldercare robot. [Online]. Available: http://www.speechtechmag.com/Articles/News/SpeechTechnology-News-Features/IBM-Develops-a-Voice-Enabled-EldercareRobot-115261.aspx

[17] J. Vanus, J. Koziorek, and R. Hercik, "The design of the voice communication in smart home care," in *Telecommunications and Signal Processing (TSP), 2013 36th International Conference on*. IEEE, 2013, pp. 561–564.

[18] M. Vacher, B. Lecouteux, and F. Portet, "Multichannel automatic recognition of voice command in a multi-room smart home: an experiment involving seniors and users with visual impairment," in *Fifteenth Annual Conference of the International Speech Communication Association*, 2014.

[19] P. Saini, B. De Ruyter, P. Markopoulos, and A. Van Breemen, "Benefits of social intelligence in home dialogue systems," in *IFIP*

Conference on Human-Computer Interaction. Springer, 2005, pp. 510–521.

[20] F. Weng, P. Angkititrakul, E. E. Shriberg, L. Heck, S. Peters, and J. H. Hansen, "Conversational in-vehicle dialog systems: The past, present, and future," *IEEE Signal Processing Magazine*, vol. 33, no. 6, pp. 49–60, 2016.

[21] M. T. Garip, M. E. Gursoy, P. Reiher, and M. Gerla, "Congestion attacks to autonomous cars using vehicular botnets," in *NDSS Workshop on Security of Emerging Networking Technologies (SENT), San Diego, CA*, 2015.

[22] A. Iftene and J. Vanderdonckt, "Moocbuddy: a chatbot for personalized learning with moocs," in *RoCHI–International Conference on HumanComputer Interaction*, 2016, p. 91.

[23] M. Lasek and S. Jessa, "Chatbots for customer service on hotelswebsites," *Information Systems in Management*, vol. 2, no. 2, pp. 146–158, 2013.

[24] R. Etemad-Sajadi, "The influence of a virtual agent on web-users' desire to visit the company: The case of restaurant's web site," *International Journal of Quality & Reliability Management*, vol. 31, no. 4, pp. 419– 434, 2014.