



City Research Online

City, University of London Institutional Repository

Citation: Arsovski, S., Wong, S. & Cheok, A. D. (2018). Open-domain neural conversational agents: The step towards artificial general intelligence. *International Journal of Advanced Computer Science and Applications*, 9(6), pp. 402-408. doi: 10.14569/IJACSA.2018.090654

This is the accepted version of the paper.

This version of the publication may differ from the final published version.

Permanent repository link: <http://openaccess.city.ac.uk/20185/>

Link to published version: 10.14569/IJACSA.2018.090654

Copyright and reuse: City Research Online aims to make research outputs of City, University of London available to a wider audience. Copyright and Moral Rights remain with the author(s) and/or copyright holders. URLs from City Research Online may be freely distributed and linked to.

City Research Online:

<http://openaccess.city.ac.uk/>

publications@city.ac.uk

Open-Domain Neural Conversational Agents: The Step Towards Artificial General Intelligence

Sasa Arsovski, Sze Hui Wong, Adrian David Cheok
Imagineering Institute
Iskandar Puteri, Malaysia

Abstract—Development of conversational agents started half century ago and since then it has transformed into a technology that is accessible in various aspects in everyday life. This paper presents a survey current state-of-the-art in the open domain neural conversational agent research and future research directions towards Artificial General Intelligence (AGI) creation. In order to create a conversational agent which is able to pass the Turing Test, numerous research efforts are focused on open-domain dialogue system. This paper will present latest research in domain of Neural Network reasoning and logical association, sentiment analysis and real-time learning approaches applied to open domain neural conversational agents. As an effort to provide future research directions, current cutting-edge approaches applied to open domain neural conversational agents, current cutting-edge approaches in rationale generation and the state-of-the-art research directions in alternative training methods will be discussed in this paper.

Keywords—Artificial intelligence; deep learning; neural networks; open domain chatbots; conversational agents

I. INTRODUCTION

Artificial General Intelligence (AGI) is achieved when machines are capable to conduct intellectual tasks [1]. Hence it is undeniable that being able to interact with machines in natural language is an important feature of Artificial General Intelligence. Human's ability to communicate and conduct cognitive behavior in natural language play an important role in completing most intellectual tasks. The oldest conversations agents created could be traced back to the success of ELIZA [2] which was first designed with the aim to be a virtual therapist. However due the lack of human knowledge and basic coding, it failed to reach farther than a short conversation. The next conversational agents that followed the creation of ELIZA have improved ELIZA model as a base. Conversational agent *Parry*, also known as "ELIZA with attitude" was designed to simulate a patient with schizophrenia. *Jabberwacky* [3], is the one of the earliest attempts to create a chatbot via human interaction. *Alicebot* [4], an improved version of ELIZA, took the attention of conversational agents research. *Alicebot* uses heuristical pattern matching rules to understand questions and make reply to the human's input. Winning the Loebner Prize, for *Alicebot* was a great success at its time of launch. However, due to its occasional glitch in exposing its mechanistic aspects in short conversations, *Alicebot* was unable to pass the Turing Test. The earliest precursor of the commercial conversational agent in the market would be *SmarterChild* [5], but it is more widely known as the precursor to the various well-known conversational agents in the industry such Apple's Siri or Samsung's S voice.

The first few years of the 21st century witnessed creations of conversational agents which do not only rely on natural language processing. For example, IBM's Watson integrated Machine Learning methodologies along with natural language processing to pull intelligent insights from data. Watson was the only conversational agent with implemented Artificial Intelligence (AI) technologies to gain wide commercial pick up. Apple's Siri incorporates speech recognition to create a better user experience via voice command. Siri's speech recognition uses deep neural network to rank the confidence level of voice inputs and makes a response accordingly. The success of Siri subsequently inspired the creation of other intelligent personal assistants like Microsoft's Cortana [6] and Amazon's Alexa [7]. Despite the success of these chatbots, they are basically goal-oriented dialogue system [8], which sole purpose is to help people solve day-to-day problems using natural language processing. In order to create a conversational agent which is able to pass the Turing Test and move AI research towards full AGI, more research effort towards open-domain dialogue system are presented in [9]. Open-domain conversational agent, Microsoft Tay bot is designed to imitate the way teenagers communicate in America. Tay was created to learn through interactions with human users on Twitter, but it was suspended shortly after subsequent controversy where the bot began to post inflammatory and offensive tweets through Twitter [10].

After almost 50 years since the introduction of chatbots and numerous surveys over the past several decades, chatbot technology is still in the wrap with much more room for improvement. The improvement of computer hardware as well as the use of smartphones, which increased the availability of data, have acted as the catalyst to recent advancement in AI research such as neural conversational agents. The introduction of Sequence to Sequence (Seq2Seq) Long Short Term Memory (LSTM) Neural Network framework [11] has greatly improved Natural Language Processing applications such as neural machine translation [12]. Researchers apply the same methodology to generate neural conversational agents [13].

Seq2Seq is a neural network model that uses one LSTM, a special kind of RNN to read the input sequence encoder, for every time step and the results is a vector. Then it uses another LSTM to extract the output sequence from that vector-decoder. The goal of *LSTM* is to estimate the conditional probability $p(y_1, \dots, y_{T'} | x_1, \dots, x_T)$, where (x_1, \dots, x_T) is an input sequence and $(y_1, \dots, y_{T'})$ is its corresponding output sequence whose length T' may differ from T . The *LSTM* computes this conditional probability by first obtaining the

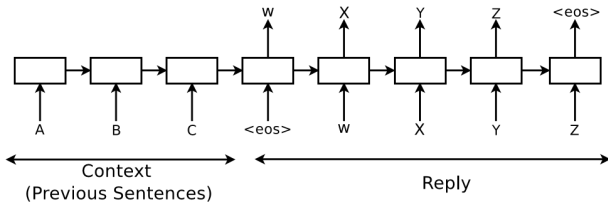


Fig. 1. seq2seq model [11].

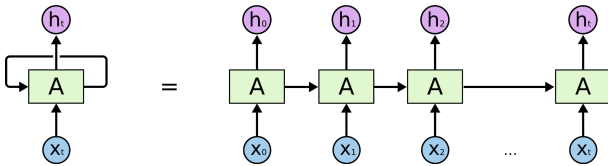


Fig. 2. Recurrent Neural Network (RNN) [14].

fixed dimensional representation v of the input sequence (x_1, \dots, x_T) given by the last hidden state of the *LSTM*, and then computing the probability of $y_1, \dots, y_{T'}$ with a standard *LSTM* language model formulation whose initial hidden state is set to the representation v of x_1, \dots, x_T :

$$p(y_1, \dots, y_{T'} | x_1, \dots, x_T) = \prod_{t=1}^{T'} p(y_t | v, y_1, \dots, y_{t-1}) \quad (1)$$

In this equation, each $p(y_t | v, y_1, \dots, y_{t-1})$ distribution is represented with a softmax over all the words in the vocabulary. The overall scheme is outlined in Figure 1, where the shown *LSTM* computes the representation of $A, B, C, <EOS>$ (sequences of the input data) and then uses this representation to compute the probability of $W, X, Y, Z, <EOS>$ (sequences of the output data) as shown in Fig. 1. Numerous of neural conversational agent based on [13] have been developed and unprecedented results have been observed.

The focus of this paper will be on open-domain, deep neural network based conversational agents. Success of Seq2Seq model applied to Neural Machine Translation and neural conversational agent [13] initiated the state-of-the-art research in the dialogue generation. This review paper presents the current state of the art open-domain generative neural conversational agents based on Seq2Seq model. Future directions in neural conversational agent research towards creation of a full AGI will also be discussed in this paper.

This paper is organized as follows: analysis of open-domain neural conversational agents is presented in the first section, neural conversational agent with reasoning and logical association is presented in the second section, sentiment analysis and reasoning networks in conversational agents are presented in the third section, neural conversational agent with real-time learning component are analyzed in the fourth section, the fifth section presents the discussion and future work.

II. CURRENT STATE OF THE ART NEURAL CONVERSATIONAL AGENTS

Progress in Text and Natural Language Processing has always been a crucial component in the pursue of Artificial General Intelligence. Despite the groundbreaking achievement

in Neural Machine Translation [12], researchers are still a distance away from achieving conversational agents that are able to conduct open-ended conversations that are indistinguishable from a human. Most open-ended neural network conversational agents research utilize Sequence to Sequence model (Seq2Seq) [11] which uses a multi-layered Long Short-Term Memory (LSTM) to map the input sequence to a vector of a fixed dimensional, and then another deep LSTM to decode the target sequence from the vector hence enabling conversational models to be trained end-to-end, thus decreasing the number of hand-crafted rules [13]. This method as proposed in [13] has enabled generation of simple conversations when is trained on a large conversational dataset. This method has not only successfully extracted knowledge from a domain specific dataset but has also demonstrated simple form of common sense reasoning when trained on a large and general domain dataset of movie subtitles [15]. The success of this research inspired many subsequent open-domain conversational agent research with the model proposed by [13] as the baseline model.

Despite its success as a dialogue generation and machine translation model, one significant weakness of RNN (Fig. 2.) based architectures such as Seq2Seq is its tendency to generate generic, non-meaningful and non-diverse responses due to its maximum likelihood estimation mechanism in dialogue generation [16]. Maximum likelihood mechanism is the procedure of finding the parameters value of given statistic which maximise likelihood distribution [17]. Another weakness would be its tendency to be caught in an infinite repetitive loop [18]. Hence, there is numerous research extensions aimed to overcome the major shortcomings of RNN based architectures.

A. Latent Variable Hierarchical Recurrent Encoder-Decoder (VHRED)

VHRED uses a hierarchical generation process in order to exploit the with-in sequence structure in utterances and is trained using a variational lower bound on the log-likelihood. The proposed method generates a response as a sample of the continuous latent variable which is a representation of the high-level semantic content of the response. The model then generates the response word by word conditioned on the latent variable [19]. The study has demonstrated that humans prefer generic responses generated by LSTM for short contexts while for long contexts humans prefer the semantically richer responses generated by the proposed model. The results of VHRED model are shown in Table I.

B. Diversity Promoting Model

Diversity promoting model proposed Maximum Mutual Information (MMI) approach [20] as the objective function in neural network models. MMI approach replace traditional objective function such as the likelihood of output (response) to given input(message). The research in [20] demonstrated that the proposed MMI models produce more diverse, interesting, and appropriate responses, yielding substantive gains in BLEU scores on two conversational datasets and in human evaluations [21]. The BLEU score algorithm compares the N-grams of two text fragments and counts the number of matches, the similarity score of these texts is a function of the number of matches [22].

TABLE I. TWITTER EXAMPLES FOR THE NEURAL NETWORK MODELS. THE TOKEN INDICATES A CHANGE OF TURN [19]

Context	Response
jamaica blood!! jurlz lol. yes [name]!. [name]! what's poppin in toronto homie	VHRED: nothing much. just chillin. what's good with you? LSTM: something!!!!!! HRED: something!!!!!!
reaches for [name]! - approaches him warily - hey there buddy. - swings him into the air - how you doin? tilts his head and your wariness and reaches for your face [name].	VHRED: - looks up at you and smiles - you're so cute. - kisses your cheek - LSTM: - smiles - i'm fine thank you. how are you? HRED: - grins, looking at you as i look at you -
my math teacher over here talkn bout nickels nd i'm tryna make big money i need to be in calculus lol	VHRED: i need to go to the gym and get my hair done lol LSTM: lol HRED: how u been
i think i want an ipad2. but i already have an ibook. want vs. need. if you have an ibook. then you deserve an ipad!	VHRED: that's what i'm thinking. i'm going to buy a new one. LSTM: i want one!!!! HRED: i don't know what to do with it. i don't know what to do
ah echt? dan antwoord ik volgende keer via twitter als je me smst. haha doe maar p. goed bezig hoor myrthe d goed bezig? die snap ik even niet p	VHRED: haha, ik heb er geen zin in de something, maar ik heb er geen something meer something LSTM: haha something something something HRED: something something something something something something something something something

C. Deep Reinforcement Learning for Dialogue Generation

Researchers in [18] proposed neural network model that is able to generate coherent and interesting dialogues by applying deep reinforcement learning to model future reward in chatbot dialogue. Model is able to recognize future direction of conversation. Presented research shows that the proposed algorithm generates more interactive responses and manages to foster a more sustained conversation in dialogue simulation. Hence marking the first step towards learning a neural conversational model based on the long-term success of dialogues. The examples of generated conversation by RN and MMI model [18] are given in Table II.

As mentioned above, the maximum likelihood estimation technique utilized in RNN based architectures such as Seq2Seq often end up generating redundant and meaningless responses such as "I don't know" or "Maybe" due to the high frequencies of these generic responses in the training datasets. Technique, known as "mutual information" generate a list of possible responses based on an input source, known as the mutual information score from a pre-trained Seq2Seq model. This mutual information score will then be used as a reward and back-propagated to the encoder-decoder model, tailoring it to generate sequences with higher rewards using reinforcement learning technique. This method has demonstrated its success in suppressing undesirable generic responses and promoted diversity in responses. Longer sentences are also observed as a result of this proposed model, bridging state-of-the-art dialogue generation research a step closer towards creating conversational agents that are indistinguishable from a human speaker.

III. NEURAL CONVERSATIONAL AGENTS WITH REASONING AND LOGICAL ASSOCIATION COMPONENTS

Despite the progress in generating conversations using deep neural network models, most of these work produce little or no ability to form clear and human like reasoning, hence generating responses that are less human like, adhering human users from conducting elaborated and continuous meaningful conversations. Research presented by [21] and [18] might have improved dialogue generation model presented in [13] but they are still unable to produce conversations with elaborated logical associations and reasoning ability. The work presented

in [19] has significantly improved the quality of conversations by producing semantically richer responses especially conversations in longer contexts. Despite its success in answer generation based on the context of the input, the model is still lack to conduct continuous conversations with reasoning and association ability.

A. Sentiment Analysis and Reasoning Networks in Neural Conversational Agents

Progress in sentiment analysis and reasoning ability of a conversational agent is crucial towards creating more human-like conversations. However, it is important for the agent to be able to understand the underlying basis of the input, which could be a question, statement with differing sentiment or objectives in order to perform accurate reasoning and logical association. Therefore, a good sentiment analysis component in the conversational agent model plays an important role to compensate the lack of transparencies of most complex neural models in natural language processing (NLP). The ideal complex neural conversational model should yield improved performances and offer interpretable rationales for answer predictions. Current cutting-edge approach presented in [23], incorporate rationale generation as an integral part of the learning problem. This approach limit the models to extractive rationales by limiting the rationales to be subsets of words from input text that are short and coherent or must alone suffice for prediction as a substitute of the original text. Rationale generation must be learned in an unsupervised manner, hence a model with rationales is trained on the same data as the original neural models. The model demonstrated in [23], trained in an end-to-end manner, gives rise to quality rationales in absence of any explicit rationale annotations. To minimize the ambiguity of what counts as rationales, as well as difficulty in evaluation of rationale selections, the model focuses on the two following domains: (i) concerns with multi-aspect sentiment analysis, (ii) concerns with the problem of retrieving related questions. The model obtained high performance on both tasks, and, for the sentiment prediction task, the model achieved an accuracy of 96%, a significantly higher accuracy compared to bigram support vector machine (SVM) [24] and neural attention baseline. The approach was evaluated on multi-aspect sentiment analysis against a manually annotated test case, and it outperformed attention-based baseline by a significant

TABLE II. SAMPLED RESPONSES GENERATED FROM THE MUTUAL INFORMATION MODELS AND THE RL MODEL [18]

Input Message	Mutual Information Model	RL
How old are you?	I'm 16.	I'm 16, why are you asking?
What's your full name?	I have no idea.	What's yours?
I don't want to go home tonight.	Really?	Why?
Do you have any feelings for me?	I don't know what you are talking about.	I don't want to hurt your feelings.
How much time do you have here?	Not long enough. Sorry, sir	Ten seconds
Shall we get started?	Of course!	Yes. We've got a lot of work to do here.
Do you play football?	No, I don't	Yes. I love football!
We'd have to talk to him	I mean, he's a good guy	About what?
How come you never say it?	Because I don't want to hurt you.	I don't think it's a good idea to say it

margin.

Another approach to create conversational agent that is able to perform logical associations and reasoning is by incorporating Reasoning Networks (RN) in neural conversational agent model. An example of RN is demonstrated by authors in [25], achieved state of the art performance on a challenging visual Question-Answer (QA) dataset CLEVR. This is a good example where additional intelligence emerges from the composition of simple modules. Relational reasoning is a principal component of generally intelligent behavior but has proven difficult for neural networks to learn. In the work [25], authors describe how to use Relation Networks (RNs) as a simple plug-and-play module to solve problems that fundamentally hinge on relational reasoning. Authors stated that RNs deep learning architecture can discover and learn how to reason about entities and their relations. Reasoning NNs are a potential venue for a breakthrough towards AI that is now hopeful to be found. Reasoning Networks can be used for a variety of problems that can benefit from structure learning and exploitation, such as rich scene understanding in Reinforcement Learning (RL) agents, modelling social networks, and abstract problem-solving.

IV. JOINT MODEL BASED TRAINING TO IMPROVE PERFORMANCE OF NEURAL CONVERSATIONAL AGENTS

In order to improve question answering systems researchers analyze how communication between humans rely on both, question asking as well as question answering and the intrinsic connections between the two tasks [26]. We will present a few important incentives towards question generation as well. Questions are usually asked to access knowledge of others or direct one's own information-seeking behavior. According to authors in [27], the incentives to teach machines to ask questions are:

- improving the acquisitions of information in intelligent systems by asking appropriate questions based on the situation;
- improving the ability of machines to answer questions by teaching machines how to ask questions;
- create a system that is able to solve abstractive tasks such as question asking instead of extractive tasks such as question answering;
- providing practical solution on many possible applications on questions asking mechanisms about documents.

Current state of the art question answering systems based on deep neural network rely on the two main information retrieval (IR) models; generative model and discriminative

model. While generative models are theoretically sound and successful in modelling features, they suffer from difficulty in leveraging relevancy of the signals from other channels like links and clicks, whereas discriminative models lack a principled way of obtaining useful features or gathering helpful signals from massive unlabeled data. Efforts have been invested in solving these shortcomings via information retrieval generative adversarial networks (IRGAN) model. IRGAN model takes advantage and characteristics of both models where the generative model acts as an attacker to the current discrimination model, generating difficult examples for the discriminative model in an adversarial way by minimizing its discrimination objective. Authors in [28], achieve state of the art results using a IRGAN model.

Several research have utilized question generation as a tool to improve the efficacy of question answering models [29], [30], [26]. Authors in [30] simultaneously train the model by alternating input data between question answering (QA) and question generation (QG), both in the same model. The trained model is then used to generate questions and answers with the hypothesis that a good question generation helps models to improve QA performance. Authors in [26] incorporates the probabilistic correlation of QA and QG by leveraging the correlation to guide the training process of both models. Authors in [26] randomly initialize the parameters in both QA and QG models with a combination of fan-in and fan-out. Parameters of words of word embedding matrices are shared in QA and QG model. However, both QA and QG models use two different embedding matrices for question words and answer words in order to learn question and answer specific word meanings. This training framework successfully shown that by exploiting the "duality" of QA and QG improves both QA and QG. Another joint approach for QA and QG is presented in [29]. Authors uses approach presented in [30] to prove hypothesis that good question generation can improve QA performance. Authors in [29] leverages convolutional neural network (CNN) and Recurrent Neural Network (RNN) for question generation in order to cover both question generation approaches; retrieval based and generation based. Authors demonstrated that the question generation method successfully improve existing question answering systems. Both results for [30] and [26] are shown in Tables 3 and 4, respectively.

The methodology proposed by [30] might have not achieve state of the art results achieved by selected QA models, it nonetheless demonstrated the effectiveness of joint training between QA and QG, especially in efficacy of abstractive tasks like QG. Whereas [26] demonstrated the "duality" of both QG and QA models are able to improve both tasks, authors in [29] manage to significantly improve QA using a good question generation approach and feed QA system

TABLE III. EXAMPLES OF QA BEHAVIOUR CHANGES POSSIBLY INDUCED BY JOINT TRAINING. GOLD ANSWERS CORRESPOND TO TEXT SPANS IN GREEN. IN BOTH THE POSITIVE AND THE NEGATIVE CASES, THE ANSWERS PRODUCED BY THE JOINT MODEL ARE HIGHLY RELATED (AND THUS PRESUMABLY INFLUENCED) BY THE GENERATED QUESTIONS. [30]

Positive	Document	<i>in the 1960 election to choose his successor, eisenhower endorsed his own vice president , republican richard nixon against democrat john f. kennedy.</i>
	Q gold	<i>who did eisenhower endorse for president in 1960 ?</i>
	Q gen	<i>what was the name of eisenhower own vice president ?</i>
	Answer	A-gen: john f. kennedy JointQA: richard nixon
Negative	Document	<i>in 1870 , tesla moved to karlovac , to attend school at the higher real gymnasium , where he was profoundly influenced by a math teacher martin sekulic</i>
	Q gold	<i>why did tesla go to karlovac ?</i>
	Q gen	<i>what did tesla do at the higher real gymnasium ?</i>
	Answer	A-gen: to attend school at the higher real gymnasium JointQA: he was profoundly influenced by a math teacher martin sekulic

TABLE IV. SAMPLED EXAMPLES FROM THE SQUAD DATASET [26]

question	correct answer	question generated by Dual QG	question generated by Basic GQ
<i>what's the name of the green space north of the center of newcastle ?</i>	<i>Another green space in Newcastle is the Town Moor , lying immediately north of the city centre.</i>	<i>what is the name of the green building in the city ?</i>	<i>what is the name of the city of new haven ?</i>
<i>for what purpose do organisms make peroxide and superoxide ?</i>	<i>Parts of the immune system of higher organisms create peroxide, superoxide, and singlet oxygen to destroy invading microbes.</i>	<i>what is the purpose of the immune system?</i>	<i>what is the main function of the immune system ?</i>
<i>how much money was spent on other festivities in the bay area to help celebrate the coming super bowl 50?</i>	<i>In addition, there are 2 million worth of other ancillary events, including a week - long event at the Santa Clara Convention Center, a beer, wine and food festival at Bellomy Field at Santa Clara University, and a pep rally.</i>	<i>how much of the beer is in the santa monica convention center?</i>	<i>what is the name of the beer in the santa monica center ?</i>

with generated questions. These state of the art research paves an optimistic direction for an alternative training method to further improve the current state of the art open domain neural conversational agents.

V. DISCUSSION AND FUTURE WORK

Current state-of-the-art open-domain neural conversational agents have largely utilized Seq2Seq architecture along with several optimization methodologies which have produced unprecedented results. Most of these results have not been achieved by other open-domain dialogue generation models. Despite the accomplishments of these state of the art research, these conversational agents are lacking in several aspects which hinder these agents from creating conversational experiences that are indistinguishable from humans, which is the key in creating a system that is able to pass the Turing Test. Aspects that are lacking in current state of the art conversational agents are:

- Agents that are able to learn new information in real time and “remember” information in future conversation sessions.
- Agents that are able to conduct vigorous reasoning and logical associations.
- Agents that are able to select appropriate emotion in responses.
- Agents with personalities and individuality.
- Agents that are able to respond with appropriate timing.
- Multi-modal agents that are able to connect computer vision (sense of sight), audio processing (sense of hearing) as an input for a dialog generation.

In order to achieve full Artificial General Intelligence (AGI), a combination of approaches need to be considered for future research. Most likely an multi-modal solution will be the direction of future research in creating conversational agents

closest to full AGI, such is presented in [31]. This research has shown that a single multi-modal deep learning model is able to jointly learn a number of large-scale tasks from multiple domains, allowing as many parameters as possible to be shared. This allows transfer learning from tasks with large amount of data to the ones that are limited with data, hence benefiting tasks with less data by joint training.

While there are several approaches which tackle each of above listed problems, full AGI will most likely be achieved with the combination of different approaches with different aims in a single multi-modal architecture. There are a few researches that specifically address above mentioned problems, which has the potential in being part of a multi-modal model that will be able to create the most human-like conversational agent.

Authors in [32] presents an conversational agent with unforgettable characters who exhibit various salient emotions in conversations. The research has achieved the goal by focusing on humanizing artificial character of conversational agents. Often times, despite how human-like responses are able to be generated by conversational agents, the neutral mono-personality answers at certain situations which are inappropriate due to the lack of the emotional consideration generated easily gave away non-human like traits.

Authors in [23] utilize encoder-generator framework which is trained in an end-to-end manner to rise rationales quality in absence of explicit rationale annotations. The results from this research could be used to build a conversational agent that is able to reason and rationalize without supervision.

A. Train an Agent Based on Pre-defined Personality

Another approach would be personality and individual goal based conversational agents. Current approach rely largely on available data-sets composed of information from multiple personalities, mood and goals. Hence most conversational agents trained this way produce an amalgamation of all possible responses given one input. Despite plenty of implementations aimed to improve this aspect, Seq2Seq based architectures still

generating response with an assumption that there is only one correct response for each input, when in fact, is not the case for open-domain conversations. In open-domain conversations, many temporal or permanent factors influence the outcome of response. Temporal factors are such as the agent's mood, temporal interests, goals and desires due to external exposure and experience. Whereas permanent factors are: agent's basic personality and temperament, intelligence maturity such as the agent's knowledge base and reasoning and association ability. Personality based conversational agents are trained with pre-defined temporal conditioning such as personality, goals and desires. Hence the agent will be able to learn interests, personality and desire, hence reducing the high dimensional space of context variables which in turn alleviating the curse of dimensionality by lowering latent semantic space.

B. Future Direction of Research in Neural Conversational Agents with Reasoning Component

The approach proposed by [25] could be incorporated into future dialogue generation models combined with sentiment analysis and attention mechanisms (to filter unimportant relations, thus bound the otherwise quadratic complexity of the number of considered pairwise relation), to create a truly powerful open-ended conversational agent which truly understand input sources at each sentence input level as well as the sentiment of the entire conversation as a whole. Connecting the conversational agent that is able to perform robust sentiment analysis and reasoning to a real-time scalable ontology could be an important direction in the field of research in open-ended dialogue generation models.

C. Neural Conversational Agent with Real-Time Learning Component

Modelling the concept of a malleable knowledge database, conversational agent will able to learn from conversations. With the ability to learn from conversation, neural conversational agents will able to conduct more human-to-human like conversations and improve logical associations and reasoning ability. Implementation of an ontology to an conversational agent architecture will enable logical associations, reasoning and knowledge saving. These knowledge can be saved as data and updated in real-time. Conversational Agents with Real-Time Learning component model will be able to conduct more realistic human like conversations.

VI. CONCLUSION

This review paper explored current state-of-the-art in open domain neural conversational agent, as an effort in evaluating current stage in research towards achieving Artificial General Intelligence. Ever since the success of Seq2Seq model in creating state of the art generative neural conversational agent, numerous of research efforts has been invested in this direction with the aim of improving open domain neural conversational agents. Despite the great leap in breakthroughs of research in this area, current cutting edge open domain neural conversational agents are still far from being indistinguishable from a human speaker. In this paper latest research in the field of Neural Network reasoning and logical association, sentiment analysis and real-time learning approaches applied to open domain neural conversational agents are presented. Future

research directions in the domain of neural conversational agents and present state of the art research directions in alternative training methods are also explored at the end of this paper as an effort to discuss the directions needed to be addressed in order to move open domain neural conversational agent research one step closer towards full AGI.

REFERENCES

- [1] B. Goertzel and P. Wang, "A foundational architecture for artificial general intelligence," *Advances in Artificial General Intelligence: Concepts, Architectures and Algorithms*, vol. 6, p. 36, 2007.
- [2] o. Weizenbaum, "Eliza-a computer program for the study of natural language communication between man and machine," *Communications of the ACM*, vol. 9, no. 1, pp. 36–45, 1966.
- [3] R. Carpenter and J. Freeman, "Computing machinery and the individual: the personal turing test," *Computing*, Accessed September, vol. 22, p. 2009, 2005.
- [4] R. S. Wallace, "The anatomy of alice," in *Parsing the Turing Test*. Springer, 2009, pp. 181–210.
- [5] R. Khan and A. Das, "Introduction to chatbots," in *Build Better Chatbots*. Springer, 2018, pp. 1–11.
- [6] L. Chris, "Why cortana assistant can help microsoft in the smartphone market. the street," 2014.
- [7] R. Crist, "Amazon alexa: Device compatibility, how-tos and much more," *CNET.com*, 2016.
- [8] K. Crockett, O. James, and Z. Bandar, "Goal orientated conversational agents: applications to benefit society," in *KES International Symposium on Agent and Multi-Agent Systems: Technologies and Applications*. Springer, 2011, pp. 16–25.
- [9] R. Higashinaka, K. Imamura, T. Meguro, C. Miyazaki, N. Kobayashi, H. Sugiyama, T. Hirano, T. Makino, and Y. Matsuo, "Towards an open-domain conversational system fully based on natural language processing," in *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, 2014, pp. 928–939.
- [10] J. Wakefield, "Microsoft chatbot is taught to swear on twitter," Mar 2016. [Online]. Available: <http://www.bbc.com/news/technology-35890188>
- [11] 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.
- [12] Y. Wu, M. Schuster, Z. Chen, Q. V. Le, M. Norouzi, W. Macherey, M. Krikun, Y. Cao, Q. Gao, K. Macherey *et al.*, "Google's neural machine translation system: Bridging the gap between human and machine translation," *arXiv preprint arXiv:1609.08144*, 2016.
- [13] O. Vinyals and Q. Le, "A neural conversational model," *arXiv preprint arXiv:1506.05869*, 2015.
- [14] "Understanding lstm networks." [Online]. Available: <http://colah.github.io/posts/2015-08-Understanding-LSTMs/>
- [15] C. Danescu-Niculescu-Mizil and L. Lee, "Chameleons in imagined conversations: A new approach to understanding coordination of linguistic style in dialogs." in *Proceedings of the Workshop on Cognitive Modeling and Computational Linguistics, ACL 2011*, 2011.
- [16] I. V. Serban, A. Sordoni, Y. Bengio, A. C. Courville, and J. Pineau, "Building end-to-end dialogue systems using generative hierarchical neural network models." in *AAAI*, vol. 16, 2016, pp. 3776–3784.
- [17] J. W. Harris and H. Stöcker, *Handbook of mathematics and computational science*. Springer Science & Business Media, 1998.
- [18] J. Li, W. Monroe, A. Ritter, M. Galley, J. Gao, and D. Jurafsky, "Deep reinforcement learning for dialogue generation," *arXiv preprint arXiv:1606.01541*, 2016.
- [19] I. V. Serban, A. Sordoni, R. Lowe, L. Charlin, J. Pineau, A. C. Courville, and Y. Bengio, "A hierarchical latent variable encoder-decoder model for generating dialogues." in *AAAI*, 2017, pp. 3295–3301.
- [20] L. Bahl, P. Brown, P. De Souza, and R. Mercer, "Maximum mutual information estimation of hidden markov model parameters for speech recognition," in *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP'86.*, vol. 11. IEEE, 1986, pp. 49–52.

- [21] J. Li, M. Galley, C. Brockett, J. Gao, and B. Dolan, "A diversity-promoting objective function for neural conversation models," *arXiv preprint arXiv:1510.03055*, 2015.
- [22] P. Koehn, J. Martin, R. Mihalcea, C. Monz, and T. Pedersen, "Proceedings of the acl workshop on building and using parallel texts," in *Proceedings of the ACL Workshop on Building and Using Parallel Texts*, 2005.
- [23] T. Lei, R. Barzilay, and T. Jaakkola, "Rationalizing neural predictions," *arXiv preprint arXiv:1606.04155*, 2016.
- [24] B. E. Boser, I. M. Guyon, and V. N. Vapnik, "A training algorithm for optimal margin classifiers," in *Proceedings of the fifth annual workshop on Computational learning theory*. ACM, 1992, pp. 144–152.
- [25] A. Santoro, D. Raposo, D. G. Barrett, M. Malinowski, R. Pascanu, P. Battaglia, and T. Lillicrap, "A simple neural network module for relational reasoning," in *Advances in neural information processing systems*, 2017, pp. 4974–4983.
- [26] D. Tang, N. Duan, T. Qin, and M. Zhou, "Question answering and question generation as dual tasks," *arXiv preprint arXiv:1706.02027*, 2017.
- [27] X. Yuan, T. Wang, C. Gulcehre, A. Sordoni, P. Bachman, S. Subramanian, S. Zhang, and A. Trischler, "Machine comprehension by text-to-text neural question generation," *arXiv preprint arXiv:1705.02012*, 2017.
- [28] J. Wang, L. Yu, W. Zhang, Y. Gong, Y. Xu, B. Wang, P. Zhang, and D. Zhang, "Irgan: A minimax game for unifying generative and discriminative information retrieval models," in *Proceedings of the 40th International ACM SIGIR conference on Research and Development in Information Retrieval*. ACM, 2017, pp. 515–524.
- [29] N. Duan, D. Tang, P. Chen, and M. Zhou, "Question generation for question answering," in *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, 2017, pp. 866–874.
- [30] T. Wang, X. Yuan, and A. Trischler, "A joint model for question answering and question generation," *arXiv preprint arXiv:1706.01450*, 2017.
- [31] L. Kaiser, A. N. Gomez, N. Shazeer, A. Vaswani, N. Parmar, L. Jones, and J. Uszkoreit, "One model to learn them all," *arXiv preprint arXiv:1706.05137*, 2017.
- [32] J.-C. Heudin, "Emotion selection in a multi-personality conversational agent." in *ICAART (2)*, 2017, pp. 34–41.