The Evolution and Future Development of the Q&A Dialogue System

Wenbo Sun *

Department of Artificially Intelligence and Big Data, Qingdao Institute of Technology University, Qingdao, 266000, China

* Corresponding Author Email: 2729281582sehun@gmail.com

Abstract. The development of artificial intelligence is very fast. Early conversational systems could only deal with simple tasks. Now modern systems understand our conversation and answer with the right emotion. It has a great significance for human-machine interaction. Their goal is not only to follow a rigid response to simple commands but to have a dialogue system that can understand the context and emotion. At the initial stage, the technology was rigid and rule-based with only fixed patterns like “if-then”. If the user’s question matches the pattern, then the answer can be delivered. Otherwise, the right answer is impossible. But when the computer became clever enough to learn from various information sources, it started understanding multiple big data sources. By stacking and mining big data, the latest achievement based on deep learning makes the conversation with machines even smarter and better at understanding us. Each stage has opened a new way to promote the development. Text sentiment recognition is a great achievement. For example, online machines can understand the mood of users’ questions and generate the right answer with the proper tone and sentiment through big data emotional matching. But it is still hard to reach the next level. It is hard for machines to understand the real meaning of conversation; questions may go out of the overall context and the solution to real-life problems is not complete. There will be performance issues when facing topics that have not been trained on. Future development should focus on exploring new ways to take more multi-dimensional information like sound and picture as input and to develop machines that can self-learning without huge data. Only by solving these problems can make the computer interaction more natural and more emotional.

Keywords: Dialogue system, Affective computing, Human computer interaction, Deep learning, Multimodal fusion.

1. Introduction

Dialogue systems are an application of Natural language processing. They started as simple pattern matching. Now people can understand a more complex meaning. The first ones were very simple tools that could only do specific jobs. They followed a set path to service the requests. Then there was a change, statistical methods. People got better at understanding the context and responding. Another change was using information from different data sources. This change led to the latest achievement in deep learning. Each of these helped advance technology. A significant new capability is being able to read emotion from text. Machines can now understand the mood of the user’s question and generate the right answer with proper tone and sentiment through emotional matching based on big data. Machines attempting to get to this stage took many steps. It is still hard for machines to understand the overall meaning of a conversation. Another problem is using information about the world. These systems also have trouble when used for new topics that they were not trained on. Looking ahead, one path is to take different kinds of input, like sound and picture. Another direction is creating machines that can self-learning without needing huge data. Solving these problems could make interacting with a computer more natural and helpful.

The paper discusses the technical development of traditional systems. In this analysis, find the merits and limitations of each method. Secondly, this paper analyzes some major developments in systems that perceive emotion. This includes new methods that can detect feelings, give an appropriate response, and integrate different data sources. This paper compared these two development paths and found a clear shift. This paper found that the field is changing its focus. This

change is due to different research goals. New work aims to build systems with functional ability and emotional intelligence.

This paper describes technical developments in dialogue systems. It can help researchers see the future in a clear way. It can also help show limits in current technology. These limits show places where future work should focus. This development path can connect research and practical application. This work provides a way for companies and universities to work together. This allows new ideas to become working products. Many companies look for this kind of practical development.

2. Introduction to Dialogue System

2.1. Traditional Dialogue Systems

2.1.1. Introduction to traditional dialogue systems

Previous dialogue systems are task-specific. They can set alarms, answer questions or control appliances. Their goal is to do their job and get the right answer. These are old technologies with a new use: maybe you have seen them in customer service chats or in lamps controlled by voice. Because they are old technology with a new use, people use this technology in many products today for today. This technology improves slowly and in small steps. With each release, the system understands a little better what a user requests and makes fewer mistakes. With each release, the system understands a little better what a user requests and makes fewer mistakes. This improvement comes from continuous use in real life.

Early dialogue systems used people experts. They used rules and templates written by people to match keywords in a user's request. This method worked well when users triggered actions in limited situations, such as one specific topic. But it failed when people spoke in unexpected ways. Studies of these systems found that they were difficult to maintain. To use the system on a new topic, experts had to write many new rules. This system could not handle the natural variation in people's speech [1].

Later, people began to treat dialogue management as a statistical problem. They used conditional random fields to understand meaning. They also used a partially observable Markov decision process to understand the conversation state. This method made dialogue systems more robust to problems. For example, it made them less sensitive to problems in understanding, such as mistakes in speech recognition. Studies of these systems found that they still needed a large amount of labeled data to work well. The performance depended greatly on the amount and quality of training data [2].

Recent breakthroughs in deep learning changed how systems work. End-to-end systems learn from input to output. Methods such as sequence-to-sequence models and attention mechanisms improved replies. Pre-trained language models, if tuned, also helped a lot. Now these systems produce language that flows more naturally. But they are hard to control, and may say incorrect facts as true. Making these systems more reliable and accurate is an important problem for researchers now [3] [4].

2.1.2. Limitations and difficulties

Nowadays, research in the field of dialogue systems encounters some long-standing technical limitations and these limitations in turn restrict the practical effect and overall effect of these systems.

There is a basic limitation: now systems cannot understand deep context. There is a very clear limitation in research regarding multi-turn conversations. If dialogue goes across two turns, systems lose track of earlier mentions, and because of this fact, they give answers that have nothing to do with the question or the point [5].

Another limitation that dialogue systems face is knowledge. They cannot represent or use factual information and for this reason, they are not very reliable. A key problem with connecting a dialogue system to external knowledge resources is that systems should access a database and documents in order to give the right answers. But methods that integrate systems and knowledge often fail. This hinders developing intelligent conversations [6].

The above conversations lack user adaptation. Most of the dialogue systems aim at completing some tasks. They do not attempt to recognize feeling. They use the same strategy for all users. This is a problem. If a system cannot recognize the user's feeling or any other preference, then the user will feel that the system does not 'know' him/her. The user will not find the system helpful and enjoyable. It is still difficult to develop systems that adapt to individuals [7].

The above conversations lack user adaptation. Most of the dialogue systems aim at completing some tasks. They do not attempt to recognize feeling. They use the same strategy for all users. This is a problem. If a system cannot recognize the user's feeling or any other preference, then the user will feel that the system does not 'know' him/her. The user will not find the system helpful and enjoyable. It is still difficult to develop systems that adapt to individuals.

The above conversations are another limitation. They perform badly on new topics. A system that is trained on restaurant queries will not work on medical queries. If one wants to use it for medicine, engineers will have to reengineer it and collect a huge amount of new data. This limitation causes an increase in development cost. Companies find it expensive to apply these systems in different fields. Finally, this limits the overall effect of this system in practice [1].

These limitations are an important part of the subsequent research and development focused on the application of this technology.

2.2. Emotion Aware Conversation Systems

Emotion-aware conversation systems constitute another part of conversation technology research [8]. While old systems had only one goal, that is to do certain tasks, these systems have another goal. They try to achieve interactions that are more human and understanding [9]. These models work by detecting feelings from the user's message. Then they return a message that is aware of that feeling. This model is similar to the way that people understand each other when talking. Because of that, they fit well in some specific scenarios. Now, researchers use them in mental health help, companion robots for the old, and personalized tutoring systems [10].

2.2.1. Emotion recognition and classification

Emotion recognition is an important component of emotion emotion-aware system. Its goal is to recognize users' feelings from conversation text [11]. There are three kinds of methods that researchers usually adopt. Some systems directly sort emotion by using a sentiment dictionary and hand-written rules [12]. Some systems use traditional machine learning. The model, like Support Vector Machines, will learn to sort emotion from extracted features [13]. Recently, deep learning architecture has become popular. The models, like CNNs, LSTMs and BERT, will sort them directly from text [14]. Recent models are obviously better than before. They not only sort with higher accuracy, but also sort varied inputs better than the previous system [15]. But there are two problems still existing. They can still sort fine-grained or mixed emotions easily. And performance is still very sensitive to large and high-quality data sets [16].

2.2.2. Emotional response generation

Emotion response is still a challenging problem for dialogue systems [17]. The goal is that the response generated is suitable for dialogue, and emotion can be well understood. Researchers have tried several approaches toward this end. One way is to use an emotion label to direct the text generation so that the text will match the emotion label [18]. Another way is to use reinforcement learning so that the system can get a reward according to how appropriate the emotion is [19]. A third way is to use adversarial training so that the discriminator will output 1 if the generated response sounds natural and appropriate [20]. The new approaches are still a little better. Both automatic metrics and human annotation are better than baseline on both emotion expression and dialogue smoothness [21]. But there is still a clear limitation. The system cannot remember the emotion over long, multiple-turn dialogues and loses track of the emotional context [22].

2.2.3. Personalized emotional interaction

Emotional interaction tries to remember the user's emotional style and improve the dialogue system by enhancing the user's satisfaction [23]. People can find several interesting strategies in this area. One method is that construct a user profile to record the user's individual emotional tendency so that the system can remember better [24]. Another method is to analyze past conversation history so that it can influence the future response [25]. A third method is to use transfer learning so that the general emotion model can be adapted for specific user groups [26]. These methods can let the user be more satisfied with the dialogue and stay longer [27]. But there are also big challenges. If there comes a new user comes, he or she does not have any record and thus meets the so-called "cold-start" problem, so the system does not know him/her yet [28].

3. Current Limitations and Future Work

3.1. Current Limitations

3.1.1. Technical limitations currently

Emotion-aware systems have the following technical limitations. In emotion recognition, systems often mislead on complex/mixed feelings [11]. They also mislead on fine-grained language such as sarcasm/metaphor since they often show up in real conversation [29]. Another limitation is response generation. Expression of response generated by the system often appears unnatural and even unreasonable. All replies appear stiff and repetitive [30] Lack of warmth. In addition, the system also meets the challenge of cultural differences in emotional expression. If this response is suitable in this culture, it is not suitable in another culture. So, the system cannot be used for all users around the world [31].

3.1.2. Limitations of data and resources

Currently, there is a big problem for developing emotion-aware systems: there is not enough good training data [32]. Current collections of conversations are too small. Labels of emotion applied in these conversations are not uniform, and they are the viewpoint of only one culture [33]. If people consider a sensitive application such as support of mental health, this problem becomes much more serious. There are strict privacy rules and many ethical issues, so it is hard to collect real dialogue between patients and therapists [34]. Another problem lies in the standards of research. There is no uniform rule in labeling emotion in text. Without such a common basis, other studies can't influence each other directly, and this hinders the overall progress [35].

3.2. Future Prospects

3.2.1. Technical research directions

The future research can focus on the following technical problems. Further deepen the level of emotion understanding. The future research can focus on the systems that can know why an emotion happens and what the context of the emotion is. The models can use commonsense reasoning to understand feelings in certain situations; Integrate multiple data types. The future research can focus on the systems that can integrate the text and other visual/auditory information, such as facial expression. The integrated information is helpful to improve the accuracy of emotion recognition; Learn with limited data. The future research can focus on the methods that can learn with limited data. Few-shot learning, or better transfer learning, can decrease the amount of labeled data that is needed to train the models. It is important in many specialized areas where that is hard to collect labeled data.

3.2.2. Ethical guidelines and safety standards

The technology development must go with the ethical guidelines and safety standards. The data privacy must have strict protection. The system must have strict rules to prohibit them from leaking or misusing user information. The system must be transparent to users. The users must know why the system behaves in certain ways. The models must have an explanation for their emotional

understanding. People must also have clear standards that can define who is responsible in certain ways that the systems or technologies behave in different situations, from customer service to medical care.

4. Conclusion

This paper presents the history of dialogue systems. This review presents the change from the very simple dialogue systems which are used in early task-oriented systems to recent systems that can recognize emotion. The research has clearly shown that the field is changing from very simple functionality to more emotional intelligence.

The change is not only about technical improvements. It is also about how we think about human-computer interaction. The goal is no longer just being efficient. The new goal is to create more meaningful exchanges and understanding.

This paper presents two types of dialogue systems. This paper review presents some important findings as follows. Deep learning can significantly improve how well a system can understand basic meaning; however, it cannot get much fine-grained emotion information in conversation intuitively. Integrating different data types, that is, combining text and other visual or browse information, that as facial expression, can offer new opportunities to interact with more emotion. However, it is hard to get this information quickly without a lot of computing resources in practice. Personalizing response is now key to better experiences.

There is still a major problem that these methods require very large amounts of carefully labeled data. These data need a lot of human labor and time to produce.

Emotional intelligence is now becoming the new focus in dialogue system research. To solve these problems, the future work can focus on the following technical problems. The future technical research can focus on learn with limited data. Better evaluation system can help us evaluate the system in different real-world situations. Stronger safety mechanisms must be learned to address ethical issues. It is also promising to integrate knowledge from different areas. This kind of cross-field collaboration may help us get systems that can interact in more intelligent and natural ways.

References

[1] Chen H, Liu X, Yin D, Tang J. A survey on dialogue systems: recent advances and new frontiers. ACM SIGKDD Explorations Newsletter, 2017, 19 (2): 25 – 35.

[2] Wen T-H, Vandyke D, Mrkšić N, Gašić M, Rojas-Barahona L M, Su P H, Young S. A network-based end-to-end trainable task-oriented dialogue system. In: Proceedings of the 2017 SIGDIAL Conference on Discourse and Dialogue, 2017: 438 – 449.

[3] Bao S, He H, Wang F, Wu H, Wang H. PLATO: pre-trained dialogue generation model with discrete latent variable. arXiv, 2019. https://arxiv.org/abs/1910.07931.

[4] Ham D, Lee J G, Jang Y, Kim K E. End-to-end neural pipeline for goal-oriented dialogue systems using GPT-2. In: Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34 (05): 78 – 85.

[5] Zhou L, Gao J, Li D, Shum H Y. The design and implementation of XiaoIce, an empathetic social chatbot. Journal of Artificial Intelligence Research, 2020, 67: 1 – 42. https://doi.org/10.1613/jair.1.12169.

[6] Budzianowski P, Wen T-H, Tseng B H, Casanueva I, Ultes S, Ramadan O, Gašić M. MultiWOZ — a large-scale multi-domain wizard-of-Oz dataset for task-oriented dialogue modelling. In: Proceedings of EMNLP 2018, 2018: 5016 – 5026.

[7] Zhou X, Wang Y. Emotion-aware chatbots: a survey. ACM Computing Surveys, 2021, 54 (6): 133. https://doi.org/10.1145/3472728.

[8] Zhang Z, Takanobu R, Huang X, Zhu Q. Recent advances and challenges in empathetic dialogue systems. AI Open, 2022, 3: 14 – 25. https://doi.org/10.1016/j.aiopen.2022.01.002.

[9] Rashkin H, Smith E M, Li M, Boureau Y-L. Towards empathetic open-domain conversation models: a new benchmark and dataset. In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, 2019: 5370 – 5381.

[10] Cheng H, Fang H, Ostendorf M. A dynamic speaker model for conversational interactions. In: Proceedings of NAACL-HLT 2019 (Long Papers), 2019: 2772 – 2785.

[11] Poria S, Hazarika D, Majumder N, Mihalcea R. The future of emotion recognition in conversation. In: Proceedings of EMNLP-IJCNLP 2019 (Short Papers), 2019: 1 – 5.

[12] Maas A L, Daly R E, Pham P T, Huang D, Ng A Y, Potts C. Learning word vectors for sentiment analysis. In: Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, 2011: 142 – 150.

[13] Pang B, Lee L. Opinion mining and sentiment analysis. Foundations and Trends® in Information Retrieval, 2008, 2 (1 – 2): 1 – 135. https://doi.org/10.1561/1500000011.

[14] Devlin J, Chang M-W, Lee K, Toutanova K. BERT: pre-training of deep bidirectional transformers for language understanding. In: Proceedings of NAACL-HLT 2019, 2019: 4171 – 4186.

[15] Sun C, Qiu X, Xu Y, Huang X. How to fine-tune BERT for text classification? In: Chinese Computational Linguistics, 2019: 194 – 206. Springer, Cham. https://doi.org/10.1007/978 - 3 - 030 - 33424 - 7_15.

[16] Wang Y, Huang M, Zhu X. Attention-based LSTM for aspect-level sentiment classification. In: Proceedings of EMNLP 2016, 2016: 606 – 615.

[17] Asghar N, Poupart P, Hoey J, Jiang X, Mou L. Affective neural response generation. In: Advances in Information Retrieval, 2018: 154 – 166. Springer.

[18] Zhou H, Huang M, Zhang T, Zhu X, Liu B. Emotional chatting machine: emotional conversation generation with internal and external memory. In: Proceedings of the AAAI Conference on Artificial Intelligence, 2018, 32.

[19] Jaques N, Taylor S, Sano A, Picard R. Predicting tomorrow's mood, health, and stress level using personalized multitask learning and domain adaptation. In: IJCAI 2017 Workshop on Artificial Intelligence in Affective Computing, 2017: 17 – 33. PMLR.

[20] Li J, Galley M, Brockett C, Gao J, Dolan B. A diversity-promoting objective function for neural conversation models. In: Proceedings of NAACL-HLT 2016, 2016: 110 – 119.

[21] Bostan L A, Klinger R. An analysis of annotated corpora for emotion classification in text. In: Proceedings of the 27th International Conference on Computational Linguistics (COLING 2018), 2018: 2104 – 2119.

[22] Sabour S, Zheng C, Huang M. CEM: commonsense-aware empathetic response generation. In: Proceedings of the AAAI Conference on Artificial Intelligence (AAAI 2022), 2022, 36 (10): 11229 – 11237.

[23] Zhang S, Dinan E, Urbanek J, Szlam A, Kiela D, Weston J. Personalizing dialogue agents: I have a dog, do you have pets too? In: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, 2018: 2204 – 2213.

[24] Li Y, Su H, Shen X, Li W, Cao Z, Niu S. DailyDialog: a manually labelled multi-turn dialogue dataset. In: Proceedings of the 8th International Joint Conference on Natural Language Processing (IJCNLP 2017), 2017: 986 – 995.

[25] Madotto A, Lin Z, Wu C S, Fung P. Personalizing dialogue systems via meta-learning. In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, 2019: 5454 – 5459.

[26] Pan S J, Yang Q. A survey on transfer learning. IEEE Transactions on Knowledge and Data Engineering, 2009, 22 (10): 1345 – 1359. https://doi.org/10.1109/TKDE.2009.191.

[27] Gao J, Galley M, Li L. Neural approaches to conversational AI. In: Proceedings of SIGIR 2018, 2018: 1371 – 1374.

[28] Wu Y, Wu W, Xing C, Xu C, Li Z, Zhou M. A sequential matching framework for multi-turn response selection in retrieval-based chatbots. Computational Linguistics, 2019, 45 (1): 163 – 197. https://doi.org/10.1162/coli_a_00320.

[29] Zadeh A, Liang P, Mazumder N, Poria S, Cambria E, Morency L-P. Memory fusion network for multi-view sequential learning. In: Proceedings of the AAAI Conference on Artificial Intelligence, 2018, 32.

[30] Tsai Y H, Bai S, Liang P, Kolter J Z, Morency L-P, Salakhutdinov R. Multimodal transformer for unaligned multimodal language sequences. In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, 2019: 6558 – 6569.

[31] Li Y, Kazemeini A, Mehta Y, Cambria E. Multitask learning for emotion and personality traits detection. Neurocomputing, 2022, 493: 340 – 350. https://doi.org/10.1016/j.neucom.2022.03.023.

[32] Majumder N, Poria S, Hazarika D, Mihalcea R, Gelbukh A, Cambria E. DialogueRNN: an attentive RNN for emotion detection in conversations. In: Proceedings of the AAAI Conference on Artificial Intelligence, 2019, 33 (01): 6818 – 6825.

[33] Deriu J, Rodrigo A, Otegi A, Echegoyen G, Rosset S, Agirre E, Cieliebak M. Survey on evaluation methods for dialogue systems. Artificial Intelligence Review, 2021, 54 (1): 755 – 810. https://doi.org/10.1007/s10462 - 020 - 09881 - 0.

[34] Liu C W, Lowe R, Serban I V, Noseworthy M, Charlin L, Pineau J. How NOT to evaluate your dialogue system: an empirical study of unsupervised evaluation metrics for dialogue response generation. In: Proceedings of EMNLP 2016, 2016: 2122 – 2132.

[35] Bostan L A, Klinger R. An analysis of annotated corpora for emotion classification. In: Proceedings of EMNLP 2020, 2020: 1 – 5.