2023 IEEE Symposium Series on Computational Intelligence (SSCI)
Mexico City, Mexico. December 5-8, 2023

Analysis of emotions in speech acts for chatbots: an overview and a model proposal

Emmanuel Castro
Centro de Investigación en Computación
Instituto Politécnico Nacional
Mexico, City of Mexico
ccastrom2023@cic.ipn.mx

Hiram Calvo
Centro de Investigación en Computación
Instituto Politécnico Nacional
Mexico, City of Mexico
hcalvo@cic.ipn.mx

Olga Kolesnikova
Centro de Investigación en Computación
Instituto Politécnico Nacional
Mexico, City of Mexico
kolesnikova@cic.ipn.mx

Citlali Castro
Centro de Estudios Científicos y Tecnológicos N° 6
Instituto Politécnico Nacional
Mexico, City of Mexico
ccastromu@ipn.mx

***Abstract*—A chatbot is a machine with conversational capabilities that tries to resemble a person. In the 90s the A.L.I.C.E. chatbot was created, showing significant advances over its predecessors. Since then, different progress has been made until, thanks to the advancement of technology, the development of current improved models has been achieved. However, now special attention is being paid to chatbots having affective recognition capabilities to enrich the user experience, which is an understudied area. This paper overviews state-of-the-art works on recognition of speaker's emotions and intentions and proposes to design a speech acts-based model of a chatbot that can interpret human emotions in text and give a coherent response in content and the expressed feelings. A set of techniques will be included in the design to recognize both the user's emotions and her intentions when expressing herself.**

***Index Terms*—Chatbot, emotion recognition, intention recognition, speech acts**

I. Introduction

A chatbot is a machine with conversational capabilities that tries to resemble a person [1]. This is the idea originally put forward by Alan Turing (1950) in the *Imitation game* [2], now called the Turing test.

One of the early efforts to simulate human conversation was made in 1966 with the ELIZA chatbot [3] designed by Joseph Weizenbaum, being this the first time that a machine engaged in a brief conversation with a person by recognizing keywords and using a rule-based approach for generating its responses, but in the end without truly understanding [4]. This development was followed by the PARRY chatbot [5] in 1972 which showed an improvement over ELIZA. Although PARRY gives responses due to a system of *emotional responses* and some assumptions [6], at the time of this development, researchers still did not consider that a chatbot or a dialog system must have affective recognition. It took almost three decades to start work on including emotions in chatbots [7].

In the mid-90s, the A.L.I.C.E. (Artificial Linguistic Internet Computer Entity) chatbot was created by Richard Wallace [8] showing significant advances over its predecessors. This chatbot was developed using Artificial Intelligence Markup Language (AIML), it generates responses by rec-ognizing [illegible] input, winning the *Loebner Prize* several times, but still unable to pass the Turing test [9]. The work on chatbots continued giving rise to Personal Assistants (PAs) such as Apple Siri[1], Microsoft Cortana[2], Amazon Alexa[3], Google Assistant[4]. These PAs access a wide range of sources to retrieve information necessary to generate responses, they show a better understanding of the user input and allow for a more flexible interaction, however, in general, still do not meet user expectations [10].

Currently, there is a growing demand for chatbots which can be widely used in different areas, such as customer service, healthcare, and education [11]. Dialogue systems have become popular because they can reduce operating costs and handle multiple users at the same time, requiring minimal supervision. This has led to the development of new technologies, strategies for their efficient implementation [12], and their quality evaluation [13]. Also, special attention is now paid to affective recognition capabilities as there is evidence that systems capable of expressing emotions improve user experience [14], but as of today, this issue still remains understudied.

Chatbots are expected to continue to grow in use, reaching a $2 trillion market by 2024 [15], as they are a cost-effective way for businesses to serve their customers 24/7 [16]. According to Wolk [17], conversational agents must be fast enough to avoid frustration and have the ability to answer simple answers accurately. He continues to say that it would be also desirable if chatbots could provide a unique personalized experience for each client, build lasting relationships, and obtain positive feedback. Wolk suggests that this would be achieved if chatbots were capable of processing customer intents. Unfortunately, this is not yet feasible, or not without an uncertain risk of error.

Upon our literature review, it was found that few studies tried to address the problems described in the previous paragraphs concerning the development of empathetic chatbots, and researchers agree that there is still much to be

[1]https://www.apple.com/siri/
[2]https://www.microsoft.com /en-us/cortana
[3]https://developer.amazon.com/en-GB/alexa
[4]https://assistant.google.com/

978-9-7281-4406-5/23/$31.00 ©2023 IEEE

done [18]. This imposes great challenges involving various areas of computer science, one of such challenges is the difficulty to classify informal or short chat messages with a high degree of language creativity, also misspellings and expressions of feelings without real intentions [19]. Another challenge is fluctuations and ambiguities in human mood which makes it difficult to create a single standard to encode emotions [20].

A good start in recognizing user intentions in a dialogue is to analyze them with respect to their classification in the Speech Act Theory (SAT) [21]. The authors of [22] state that "SAT conceptualizes all forms of speech as acts and suggests interpretations of communicated words require recognition of a higher-order linguistic context", so, in addition to continuing with the solid line of research on emotion recognition techniques, the advances in the field of linguistics should be taken into consideration, particularly, speech acts classification, thus providing a better analysis tool, hopefully leading to better results.

This paper presents an overview of chatbot technologies designed along the lines of two approaches, i.e., recognition of emotions and recognition of intentions, and further proposes a solution for a novel modular model of a chatbot system.

The rest of this paper is organized as follows: section II offers an overview of related works giving the background of the employed technologies as well as explaining some important concepts related to this research. Subsection II-A highlights theories on emotions and mentions diverse chatbot technologies to incorporate emotions and their implementation techniques. Subsection II-B illustrates the issue of intention recognition. Section III provides an analysis and the proposed solution for the design of a new modular model of a chatbot system. Finally, section IV concludes the paper.

II. Related Works & Background

A. Emotion Recognition

Psychology has dedicated an arduous effort to try to clarify the multidimensional phenomenon of human emotions, this is reflected in the development of various theories to explain it [23]. Emotions are reactions of the body to certain stimuli, such reactions are impulsive and unconscious, unlike feelings, which occur after perceiving physical changes resulting from emotions [24]. A limited number of basic emotions is distinguished, but their inventory is different in different theoretical models. Among them, two models are most used: Ekman's basic emotions [25] and Plutchik's wheel of emotions [26].

In the Ekman's model, six basic emotions are contemplated: anger, disgust, fear, joy, sadness, and surprise [27]. Sometimes another emotion is added to the model, which is contempt [28]. However, in our study, only the first six will be considered as consistent and coinciding to the most with other authors [29]. The simplicity of the Ekman's model makes it ideal to work with, this is one of its main strengths.

The Plutchik's model is a result of a multidimensional approach, where there are four pairs of opposite axes and emotions are defined as points along these axes. Each emotion is determined by an emotional axis and its intensity; the pairs of axes are joy-sadness, anger-fear, trust-disgust, and surprise-anticipation. For Plutchik, those are the primary emotions, all the other emotions are their combinations.

Emotion recognition in text is a difficult and challenging task. Some researchers tried to address it by defining it as an emotion identification problem approaching it from different perspectives. The work of Zhou et al. [30] was the first attempt to include the emotional factor in conversation generation at a large scale, where the authors proposed an emotional chatting machine (ECM) capable of generating responses relevant with respect to the user's utterance content as well as her emotions. The ECM was based on a sequence-to-sequence (seq2seq) model implementing it with gated recurrent units.

After Zhou's work, Asghar et al. [31] proposed an improvement over it using a Long Short-Term Memory (LSTM) model. The model recognized the user's emotion in the affective words of the input statements by means of various inference criteria and objective functions designed to this effect.

Song et al. [14] highlights the desire for dialogue systems to have the ability to express emotions during conversations for greater user satisfaction. They proposed an emotional dialogue system using a seq2seq model based on LSTM as well as on a *lexical-based attention* mechanism to determine the emotion with which to respond in an explicit or implicit way.

Another work also developed a model with LSTM, but at this time taking advantage of the benefits of the Linguistic Inquiry and Word Count (LIWC) dictionary [32]. Ghosh et al. [33] used LIWC to recognize affective features according to keyword spotting; their model generates a sequence of words corresponding to the emotion as well as to the context.

On occasions, chatbots' responses are repetitive, mindless, and lack human touch [34]. The research of [35] tried to avoid this drawback by employing a Conditional Variational Autoencoder-based (CVAE-based) model. The model generated more meaningful sentences with emotional diversity improving the performance shown by a seq2seq model alone. CVAE achieved this but not without a price: often correct syntax and grammar were compromised [35]. The model built in [36] tried to select the most adequate responses considering grammar, context, and emotion, and the model reported in [37] was capable of generating different and consistent responses, using a latent space variable.

Zhou et al. [38] assured that in the development of conversational agents, it was necessary to consider emotions within human-agent interactions. They presented a neural model to generate responses in a supervised fashion, which produced sentences with affective diversity of two types: specified and unspecified emotion.

In another study [18], under the premise that empathic responses are to be generated by imitating the user's emotions, a chatbot based on an encoder-decoder transformer was created including some randomness in its emotional responses to obtain a greater variety. The authors argued that there was still much to improve.

Lastly, concerning techniques for emotion recognition,

the authors of [39] mention that companion robots would be expected to show more empathy, which can assist them in detecting different emotions. Their approach was based on an analysis of three aspects to determine human emotion: 1) images (facial expression), 2) audio, and 3) text. Using deep learning, they implemented a prototype system of empathetic robots, believing that one day robots will be able to understand our emotions and even become our friends.

B. Intention Recognition

To recognize intentions in a dialogue, let's begin with their classification under SAT, which states that the speaker uses language to convey meaning as well as intentions, i.e., what she wants to communicate through words, sentences, and their relationships [40].

A *speech act* is a statement that the speaker produces to achieve an intended effect, and according to [41] speech acts can be of three types: 1) locutives acts, which are current declarative acts, that is, utterances to simply express facts (e.g., "It is time to do homework"), 2) illocutionary acts: ones incorporating a social function, linking actions and consequences in the environment where they were produced by participants of a conversation (e.g., "Do your homework before you go to sleep"), and 3) perlocutionary acts, which result from what was said previously, being a reaction to illocutionary acts, and eventually, leading to fulfillment of an action (e.g., "OK, I will do it").

Illocutionary acts in their turn are divided into five subcategories [40]:

- Declarative. In a declarative speech act, the speaker introduces a change into the external situation through affirmation, with consequences in the immediate surrounding reality (e.g., "You have a rating of 10").
- Commissive. In a commissive speech act, the speaker commits to doing something in future, thus creating an obligation with some purpose in the context of the situation (e.g., "I will be right back!").
- Expressive. In an act of expressive speech, the speaker exposes her feelings, state of mind, or emotional reactions with respect to the situation being experienced (e.g., "To be honest, today I feel very bad").
- Directive. In a directive speech act, the speaker tries to incite a desired action, with different degrees of assertiveness (e.g., "You must finish your chores today").
- Representative. In a representative speech act, the speaker expresses her belief about the veracity of a proposition, accepting, denying, or simply expressing her opinion about something, trying to convince the interlocutor of what is being said (e.g., "The math test was easy, really").

The second line of research on affective chatbots includes works addressing the issue of intention recognition using speech act classification. In [42] the authors proposed a hierarchical Recurrent Neural Network (RNN) for learning sequences of Dialogue Acts (DAs). The input to the network was a sequence of expressions and the output was a sequence of labels. The model relied on the hierarchical structure of dialog data by using two nested RNNs, thus making it possible to capture long-range relationships at both levels: dialog and expression. The model enhanced its responses with an attention mechanism that focused on salient cues in utterances. To evaluate the authors' proposal, two sets of data were used, Switchboard and MapTask, and the subsequent experiments showed a good performance of the model.

The work in [43] explored different techniques or methods of context representation using neural networks for dialog acts classification in such a way that, given sentences in a conversation, sufficient information for classification was captured by combining RNN architectures with attention mechanisms at different context levels. The results obtained in this work showed that the use of RNN architecture is relevant for an adequate representation of the context.

Based on previous work on DA classification by sequence labeling with hierarchical deep neural networks, the research in [44] added the power of a context-aware self-serving mechanism to their conversational model. The authors performed extensive evaluations on relevant datasets. The results showed a significant improvement on the Switchboard corpus. Also, the model performed well to capture semantic text representation at the expression level, while maintaining high precision.

III. Analysis & Proposed Solution

The review of studies in Section II, also summarized in Table I, shows that there still does not exist a single method or technique sufficient to correctly recognize the user's true intentions and emotions and at the same time to satisfy her high expectations of the conversational skills of a chatbot. This is the reason why it is necessary to propose a novel and hybrid approach to the issue. Our proposal is shown in the diagram of Fig. 1, which consists of three independent modules and a Dialog Manager (DM). The first and second modules will recognize the intentions and emotions of the user in a conversation, respectively. The first step in these modules is preprocessing the structured input to make it easier for the model to manage. At this step such Natural Language Processing (NLP) techniques as tokenization, lemmatization, and stemming will be used. In the second step, each module will independently use AI-driven strategies to obtain the intentions and emotions of the user, then passing them to the DM, where both results will be taken into account to get a better understanding of the user request. The flow then goes to the third module which is expected to generate a high-quality response using machine learning techniques trained on a big corpus.

IV. Conclusion

The architecture of the chatbot in this work is in development, but it is expected that once it is completed, it will generate appropriate responses in content and emotion due to our novel approach elaborated upon a careful revision of the existing studies summarized in Table I. The chatbot will use a set of techniques and independent functions of the modules of its model, where the first module will obtain the user intentions under the Speech Acts Theory (SAT) and the second module will recognize emotions. The output of both modules will be integrated within the system to produce relevant responses. In future, we will implement and evaluate the chatbot proposed in this paper.

TABLE I
PURPOSE AND CATEGORIZATION OF STUDIES IN RELATED WORKS

Study	Purpose of study	Categorization by recognition of	
		emotions	intentions
Song et al. [14]	Propose an emotional dialogue system using a seq2seq model to determine emotions with which to respond explicitly or implicitly	X	
Majumder et al. [18]	Propose a model based on an encoder-decoder transformer with some randomness included in emotional responses to obtain a variety of responses	X	
Zhou et al. [30]	Propose a model to generate relevant responses with sense and emotion employing internal and external memory	X	
Asghar et al. [31]	Develop three novel methods to generate responses with affective content using an LSTM conversational model	X	
Ghosh et al. [33]	Propose an LSTM language model using LIWC to recognize affective features and generate conversational text	X	
Liu et al. [36]	Use an affective lexicon to understand user emotions embedding them in word vectors, then employing a CVAE-based model to enhance emotion diversity in generated responses	X	
Yao et al. [37]	Propose a model to generate different and consistent emotional responses from the same input	X	
Zhou et al. [38]	Develop a neural model to generate conversational responses with a supervised approach for emotion recognition producing emotional diversity	X	
Fung et al. [39]	Propose a prototype system of empathetic robots using deep learning to recognize emotions and humor	X	
Tran et al. [42]	Propose a hierarchical RNN for learning sequences of DAs, taking advantage of the nature of dialogue information to improve results		X
Ortega and Vu [43]	Explore different techniques or methods of context representation using neural networks for the classification of DAs		X
Raheja and Tetreault [44]	Propose a method for DA classification, labeling the utterances using hierarchical deep neural networks and adding the power of a context-aware self-serving mechanism		X

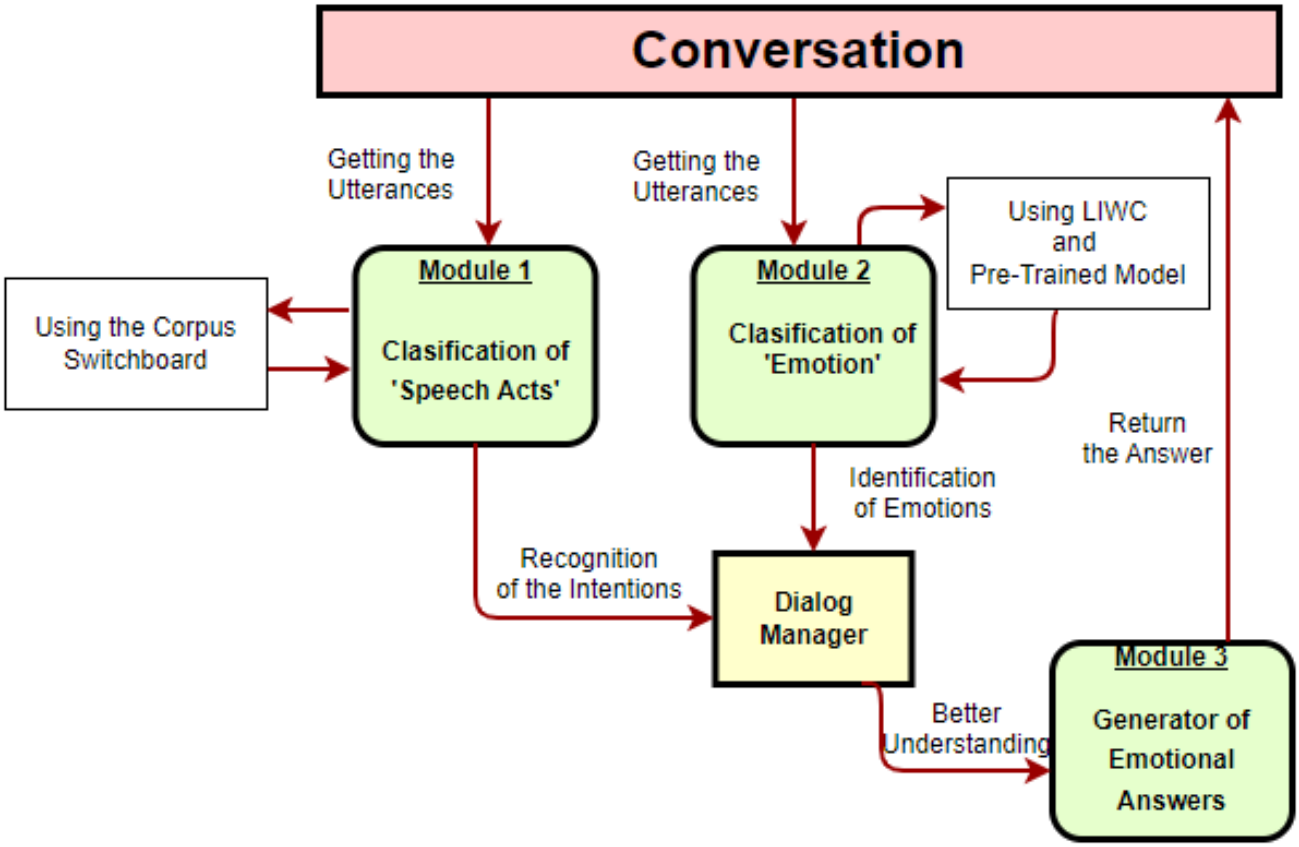

Fig. 1. System diagram of a chatbot showing the interaction between *Dialog Manager* and three modules

ACKNOWLEDGMENTS

The work was done under partial support of Mexican Government: SNI, CONAHCYT and SIP-IPN grants 20231567 and 20230140.

REFERENCES

[1] S. Nithuna and C. A. Laseena, "Review on Implementation Techniques of Chatbot," 2020 International Conference on Communication and Signal Processing (ICCSP), Chennai, India, pp. 0157-0161, 2020.

[2] A. Turing, "Computing machinery and intelligence," Springer, 2009.

[3] J. Weizenbaum, "ELIZA—a computer program for the study of natural language communication between man and machine," Communications of the ACM, vol. 26, no. 1, pp. 23–28, 1983.
[4] A. Przegalinska, L. Ciechanowski, A. Stroz, Gloor, P. and G. Mazurek, "In bot we trust: A new methodology of chatbot performance measures," Business Horizons, vol. 62, no. 6, pp. 785-797, 2019.
[5] S. A. Thorat and V. D. Jadhav, "A review on implementation issues of rule-based chatbot systems," Proceedings of the international conference on innovative computing & communications (ICICC). 2020.
[6] E. Adamopoulou and L. Moussiades, "Chatbots: History, technology, and applications," Machine Learning with Applications, vol. 2, 2020.
[7] T. S. Polzin and A. Waibel, "Emotion-sensitive human-computer interfaces," ISCA tutorial and research workshop (ITRW) on speech and emotion, 2000.
[8] R. Wallace, "The elements of AIML style," Alice AI Foundation, New York, NY, USA: vol. 139, 2003.
[9] V. Sharma, M. Goyal and D. Malik, "An intelligent behaviour shown by chatbot system," International Journal of New Technology and Research, vol. 3 no. 4, 2017.
[10] J. Grudin and R. Jacques, "Chatbots, humbots, and the quest for artificial general intelligence," in Proceedings of the 2019 CHI conference on human factors in computing systems, pp. 1–11, 2019.
[11] G. Caldarini, S. Jaf and K. Mcgarry, "A literature survey of recent advances in chatbots," Information, vol. 13, no 1, p. 41, 2022.
[12] E. H. Almansor, and F. K. Hussain, "Survey on intelligent chatbots: State-of-the-art and future research directions‘," In Complex, Intelligent, and Software Intensive Systems: Proceedings of the 13th International Conference on Complex, Intelligent, and Software Intensive Systems (CISIS-2019) pp. 534–543. Springer International Publishing. 2020.
[13] R. Lowe, M. Noseworthy, I. Serban, N. Angelard-Gontier, Y. Bengio, and J. Pineau, "Towards an automatic turing test: Learning to evaluate dialogue responses," arXiv preprint arXiv:1708.07149, 2017.
[14] Z. Song, X. Zheng, L. Liu, M. Xu, and X. Huang, "Generating responses with a specific emotion in dialog," Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, pp. 3685–3695, 2019.
[15] G. Daniel and J. Cabot, "The software challenges of building smart chatbots," in 2021 IEEE/ACM 43rd International Conference on Software Engineering: Companion Proceedings (ICSE-Companion), IEEE, pp. 324–325, 2021.
[16] T. Okuda and S. Shoda. "AI-based chatbot service for financial industry," Fujitsu Scientific and Technical Journal vol. 54, no. 2, pp. 4-8, 2018.
[17] K. Wolk, "Real-time sentiment analysis for polish dialog systems using mt as pivot," Electronics, vol. 10, no. 15, p. 1813, 2021.
[18] N. Majumder, P. Hong, S. Peng, J. Lu, D. Ghosal, A. Gelbukh, R. Mihalcea, and S. Poria, "MIME: MIMicking emotions for empathetic response generation," arXiv, 2020.
[19] M. Brooks, K. Kuksenok, M. K. Torkildson, D. Perry, J. Robinson, T. Scott and C. Aragon, "Statistical affect detection in collaborative chat," In Proceedings of the 2013 conference on Computer supported cooperative work, pp. 317-328, 2013.
[20] J. Feine, and S. Morana, and U. Gnewuch, "Measuring service encounter satisfaction with customer service chatbots using sentiment analysis," 2019.
[21] A. Stolcke, K. Ries, N. Coccaro, E. Shriberg, R. Bates, D. Jurafsky, P. Taylor, R. Martin, C. Ess-Dykema, and M. Meteer, "Dialogue act modeling for automatic tagging and recognition of conversational speech," Computational linguistics, vol. 26, no. 3, pp. 339–373, 2000.
[22] S. Ludwig, and K. de Ruyter, "Decoding social media speak: developing a speech act theory research agenda," Journal of Consumer Marketing, vol. 33, no. 2, pp. 124-134, 2016.
[23] M. Lewis, J. Haviland-Jones, and L. Feldman Barrett, "Handbook of emotion. chapter 31," 2008.
[24] M. Lenzen, "Feeling our emotions," Scientific American Mind, vol. 16, no. 1, pp. 14–15, 2005.
[25] P. Ekman et al, "Basic emotions," Handbook of cognition and emotion, vol. 98, no. 45-60, p. 16, 1999.
[26] R. Plutchik, "Emotions: A general psychoevolutionary theory," Approaches to emotion, vol. 1984, no. 197-219, pp. 2–4, 1984.
[27] P. Ekman, "Facial expressions of emotion: New findings, new questions," 1992.
[28] J. L. Tracy and D. Randles, "Four models of basic emotions: A review of Ekman and Cordaro, Izard, Levenson, and Panksepp and Watt," Emotion review, vol. 3, no. 4, pp. 397–405, 2011.
[29] C. E. Izard, "Forms and Functions of Emotions: Matters of Emotion–Cognition Interactions," Emotion Review, vol. 3, no. 4, pp. 371–378, 2011.
[30] H. Zhou, M. Huang, T. Zhang, X. Zhu, and B. Liu, "Emotional chatting machine: Emotional conversation generation with internal and external memory," in Proceedings of the AAAI Conference on Artificial Intelligence, vol. 32, 2018.
[31] N. Asghar, P. Poupart, J. Hoey, X. Jiang, and L. Mou, "Affective neural response generation," in Advances in Information Retrieval: 40th European Conference on IR Research, ECIR 2018, Grenoble, France, March 26-29, 2018, Proceedings 40, pp. 154–166, 2018.
[32] J. W. Pennebaker, M. E. Francis, and R. J. Booth, "Linguistic inquiry and word count: Liwc 2001," Mahway: Lawrence Erlbaum Associates, vol. 71, no. 2001, p. 2001, 2001.
[33] S. Ghosh, M. Chollet, E. Laksana, L.-P. Morency, and S. Scherer, "Affect-lm: A neural language model for customizable affective text generation," arXiv preprint arXiv:1704.06851, 2017.
[34] E. Adamopoulou and L. Moussiades, "An overview of chatbot technology," in BT - Artificial Intelligence Applications and Innovations, I. Maglogiannis, L. Iliadis, and E. Pimenidis, Eds., p. 373, Springer International Publishing, 2020.
[35] G. Bilquise, S. Ibrahim and K. Shaalan, "Emotionally Intelligent Chatbots: A Systematic Literature Review," Human Behavior and Emerging Technologies, vol. 2022, 2022.
[36] M. Liu, X. Bao, J. Liu, P. Zhao, and Y. Shen, "Generating emotional response by conditional variational auto-encoder in open-domain dialogue system," Neurocomputing, vol. 460, pp. 106–116, 2021.
[37] K. Yao, L. Zhang, T. Luo, D. Du, and Y. Wu, "Non-deterministic and emotional chatting machine: learning emotional conversation generation using conditional variational autoencoders," Neural Computing and Applications, vol. 33, no. 11, pp. 5581–5589, 2021.
[38] G. Zhou, Y. Fang, Y. Peng, and J. Lu, "Neural conversation generation with auxiliary emotional supervised models," ACM Trans. Asian LowResource Lang. Inf. Process., vol. 19, no. 2, pp. 1–17, Mar. 2020.
[39] P. Fung, D. Bertero, Y. Wan, A. Dey, R. H. Y. Chan, F. Bin Siddique, Y. Yang, C.-S. Wu, and R. Lin, "Towards empathetic human-robot interactions," in Computational Linguistics and Intelligent Text Processing: 17th International Conference, CICLing 2016, Konya, Turkey, April 3–9, 2016, Revised Selected Papers, Part II 17, pp. 173–193, Springer, 2018.
[40] J. Searle, "A classification of illocutionary acts1," Language in society, vol. 5, no. 1, pp. 1-23, 1976.
[41] J. L. Austin, How to do things with words. Oxford university press, 1975.
[42] Q. H. Tran, I. Zukerman, and G. Haffari, "A hierarchical neural model for learning sequences of dialogue acts," in Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers, pp. 428–437, 2017.
[43] D. Ortega and N. T. Vu, "Neural-based context representation learning for dialog act classification," arXiv preprint arXiv:1708.02561, 2017.
[44] V. Raheja and J. Tetreault, "Dialogue act classification with context-aware self-attention," arXiv preprint arXiv:1904.02594, 2019.