Research on the challenges and the possible ways to augment customer service Chatbot capability

Rong Peng
Blekinge Institute of Technology
ropeaudrey@gmail.com

Abstract
In recent years, intelligent customer service chatbot has become an important part of the customer service system, they can provide 24-hours online service. But in e-commerce area, the most common form is human-robot coordination, due to the chatbot still has many scenarios which they cannot understand and handle properly. If the problems are out of their predefined capability scope, the customer service chatbot's answers will be embarrassed or tend to take the "safe" answer, which really makes customers feel disappointed. Therefore, this field has a large improvement space. In this proposal, I will do a systematic literature review and find the current existing challenges of customer service chatbot, and then propose a possible conceptual framework, which could augment the capability of customer service chatbot, while demonstrating the corresponding underlying techniques.

Keywords
Chatbot, challenges, augment capability, screenshots, Nature Language Processing(NLP), LSTM,OCR, Encoder-Decoder Model

ACM Classification Keywords
Computing methodologies --- Artificial Intelligence --- Control methods --- Robotic planning --- Evolutionary robotics

Introduction
A chatbot is an instant messaging account that able to provide services using instant messaging frameworks with the aim of providing conversational services to users in an efficient manner. [1] And nowadays, chat bots have been widely used in e-commerce, banking, insurance and other industries. They can provide 24-hours questions and answers sessions with customers, as well as products recommendation and retrieval capabilities. These have liberated the labor to a certain extent.

Accenture, a large consulting firm, its recent research reported that, In the banking, insurance and other financial industries, 70% of consumers are willing to choose artificial intelligence customer service to provide advice for their consumption decisions. However, under some specific complex circumstances, consumers still hope to get help from customer service staff.[2] This report explains why most markets take the mode of combination of human-machine collaboration.

Hill, J., Ford, W. R., & Farreras, I. G [3] have compared the 100 instant messaging conversation to 100 exchanges with the popular chatbot Cleverbot along seven dimensions, then they found people were actually inclined to send more than twice as many messages to chatbots compared to other people, many people are willing to interact with chatbots and be more patient.

It can be seen that most of human is interested in interacting with artificial intelligence, and the customer service chatbot has a very bright prospect and broad market. However, we still need to see the defects and challenges behind the robot.

Also there are some companies have found that buying an intelligent customer service chatbot is expensive, but not as good as they expected. After all, the customer service chatbot as a program, it is impossible to have the compassion of the customer service staff, they can not handle the problem as flexible as the customer service staff in the complaint scenario. And the biggest feature of the customer service contact center is that the business knowledge is updated quickly.

How to make the internal knowledge base system achieve better docking is also a difficult problem. Also the dialogue generation requires vast logic and linguistic resources, and also should take care of the sentimental analysis. A simple chatbot is not a challenging task as compared to complex chatbots and developers should understand and consider the stability, scalability and flexibility issues along with high level of intention on human language[1]

Literature Review Summary

In this round of literature review, I found challenges such as text input, dialogue ridge problems, screenshot recognition problem and cannot better recognize emotion in emoticons, speech message in dialect and so on.

My main purpose in this literature review is to find the possible way to augment customer service chatbot. In the Background section, I will introduce the Existing Platforms, LSTM, Encoder-Decoder Model, OCR, GMVAE Model .etc. These architectures will help to improve chatbot functions more or less, but still with some limitations, like the *Text Extraction and Retrieval from Smartphone Screenshots*[6] introduces the method to extract the text from screenshot then retrieval, similar to the translation technology, which not means generate response with emotion, even no mention recognize the embedded picture in the sceenshot. Therefore, this round LR made me see the real programming challenges in improving customer service chatbot.

Background

A. Existing Platforms

The existing chatbots platforms can be divided into three major types, they are Nonprogramming chatbots, Conversation-Oriented chatbots and Platforms by tech giant's chatbots.[1]
Nonprogramming chatbots doesn't require strong programming technique, the Conversational-Oriented chatbots use the AIML(Artificial Intelligence Markup Language) to program, and the Tech giants can be considered as Google Api.ai, Facebook Wit.ai, Microsoft LUIS, Amazon Lex and IBM Waston, Alibaba Groups' AliMe.
AliMe is an advanced and strong customer service chatbot, we can easily interact with it when we shopping online in TaoBao application. AliMe has already realized the text input dialogue, voice recognition dialogue, picture recognition and retrieval, products recommendation functions and so on.
However, this customer service chatbot also faces challenges. Showing in the Figure1, in the dialogue, I tried to chatted with AliMe in a scenario when we may send messages to our friends in more than two message bubbles, but AliMe cannot understand the dialogue if I send the message in more than one message bubble. AliMe also cannot understand the sending screenshot, at the same time the multi-turn questions and answers showed limitation, after I sending a picture, I even had a no chance to enter any word. If the AliMe can understand screenshots, then can apply in the complaint scenario.

B. Nature Language Processing

Natural language processing, which is to achieve natural language communication between people and computers, or to achieve natural language understanding and natural language generation is very difficult. The root cause of the difficulties is the variety of vagueness or ambiguity in natural language texts and dialogues at all levels.

C. Long-Short Term Memory (LSTM)

The LSTM network is a RNN, which has three special gate structures: Input Gate, Forget Gate and Output Gate. The input gate adds new information to the neuron, the forget gate is used to control whether the neuron will discard the previously saved information, and output gate will output the final saving information.

Zi, Zf, Zo respectively through an activation function f to obtain an output value to control the Input Gate, Forget Gate and the Output Gate, the activation function f is generally a sigmoid function will map the input value to the probability value between 0 ~1, indicating the degree of opening the gate structure, 0 means the gate structure is completely closed, 1 means the gate structure is fully open. [4]

The input for the neuron Z passed through a function g and output is g(Z) through the Input Gate to decide whether the input information can pass this gate.The Input Gate output is g(Z)f(Zi). The previously stored information in the neuron is C. The Forget Gate decides whether or not to forget the past stored information. The Forget Gate output is Cf(Zf) and the information of neuron updated is C'= g(Z)f(Zi) +Cf(Zf).The Output Gate determine whether to output the currently saved information through the output gate.The output value of Output Gate is h(C')f(Zo).[4]

D. Encoder-Decoder Model

Encoder-Decoder model can be used to generate the dialogue response. When we input sentence into a sequence E(a1,a2,…aT), then transfer E to vector X(x1,x2,…xT), the Encoder will converts the input sequence to hidden layer H(h1,h2,…,hT).

Until the last input, Encoder sends to the semantic vector, for the decoder part, the inverse of the Encoder is actually performed, and the obtained semantic vector C is gradually decoded with the input of the semantic vector C, each step output a vector yi and at each time-step forms a sequence Y(y1,y2,…yT).[4]

E. Knowledge Graph

The knowledge graph can be overall divided into two parts: entities and relations.
Entity: Refers to something that is distinguishable and independent.
Semantic class: A collection of entities with the same characteristics
Attribute: Point to an attribute value from an entity
Relationship: Formalized as a function that maps kk points to a Boolean value
A triple tuple is a general representation of a knowledge map, it can help capture semantic similarity. Compared with traditional information retrieval(IR) model, knowledge graph approach increase accuracy by 10%.[5]

F. Optical Character Recognition

The process of checking the characters printed on the paper by an electronic device (such as a scanner or a digital camera) is to determine the shape by detecting dark and light patterns, and then converting the shape into computer text using character recognition; that is, for printing characters. Optical text is used to convert text in a paper document into a black and white bitmap image file. The text in the image is converted to text format by the recognition software, and the text processing software further edits and processes the text.

After pre-processing screenshots, each segmented region was fed to the OCR engine, using the Python wrapper for Tesseract. Tesseract recognizes text in a "two-pass process" that integrates character segmentation with the recognition module, and uses backtracking to improve the quality of the output.[6] This paper uses such way to recognize the text on screenshots.

G. Nature Language Inference(NLI) Corpus

Understanding entailment and contradiction are very important. They are the basis of understanding the natural language. Inference about entailment and contradiction is a valuable testing ground for the development of semantic representations [7]. NLI has been addressed using a variety of techniques, including those based on symbolic logic, knowledge bases, and neural networks.[7]

But machine learning research in this area requires large resources. Introduce Standford Natural Language (SNLI) corpus, a new,

freely available collection of labeled sentence pairs.

H. Guassian Mixture Variational Auto Encoder (GMVAE) Tacotron

GMVAE-Tacotron, a TTS model which learns an interpretable and disentangled latent representation to enable fine-grained control of latent attributes and provides a systematic sampling scheme for them. If speaker labels are available, we demonstrate an extension of the model that learns a continuous space that captures speaker attributes, along with an inference model which enables one-shot learning of speaker attributes from unseen reference utterances.[8]

A text-to-speech(TTS) model can control latent attributes in the generated speech. And GMVAE-Tactron can evaluate a widge degree of variations.

Research Questions

RQ1.What are the state-of-arts and challenges of existing customer service chatbot?

RQ1.1What are the state-of-arts?

RQ1.2What are the limitations and challenges?

RQ2. Are there any possible ways to augment the customer service chatbot?

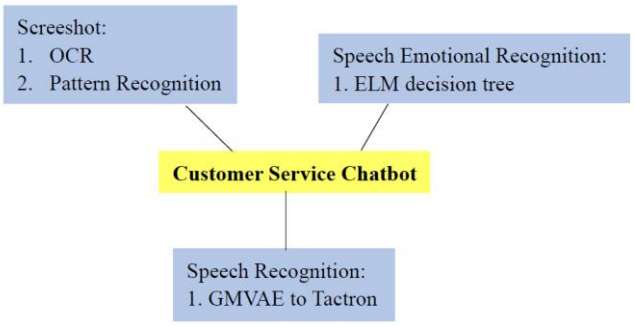
RQ2.1 If true, what are the possible ways?

RQ2.2 How to augment? Is it work?

RQ2.3 How to evaluate the metrics?

RQ2.4What about comparing with other models?

Conceptual framework:



H1: Combining the OCR and Pattern Recognition methods chatbot can recognize, extract and understand the key information.

H2:Adding ELM decision tree method, the customer service chatbot can understand better customer emotion.

H3:Adding GMVAE to Tractron can let the chatbot understand better the customer voice input.

Research Method

I will take the Empirical Research method. And the research design is divided into the following seven steps:

1. Establishing the research theme;
2. Literature review;
3. Defining the research question;
4. Collecting the data and conduct experiment;
5. Analyzing the data;
6. Survey
7. Result.

The following I will explain the details.

1. Establishing the research theme

The research theme should surround with the challenges and augmentation ways of the customer service chatbot.

2. Literature review

In this systematic literature review, I found the current existing challenges and possible ways to augment the customer service chatbot.

Challenges:

A. Text Input

Segmented sentences composed of multiple message bubbles cannot understand.

B. Dialogue

The capture of the interlocutor's intentions is not perfect, and the dialogue is limited to multiple rounds of questions and answers.

C. Emoticons

Can't understand the emoticons input by the interlocutor perfectly.

D. Screenshots

Sometimes we need to use screenshots to apply in the complaint scenario, the chatbot directly use the screenshot as a picture to process image retrieval.

E. Speech recognition

Can't understand dialect well.

F. Lack of rich training data

When applying solutions to paraphrase identification and NLI in chatbot systems in the E-commerce industry, all solutions rely on a large amount of labeled data, however it is time-consuming and costly to manually annotate sufficient labeled data for each domain.[9]

G. Hard to reach a high Queries Per Second(QPS)

For real industry applications, when real-time responses are expected and a large number of customers are being served simultaneously, we need an efficient method to support a high QPS.[9]

Possible ways to augment:

1. In small business domain, as Singh, R., Paste, M., Shinde, N., Patel, H., & Mishra, N. [10] experiment, chatbot can add some intent files that will yield a certain level of accuracy.
2. To improve chatbot dialogue generation, can consider the Second Response Generation[11], it combines the first response generation and second response generation to improve the emotional and informative dialogue. To make the conversation become human-like emotion, using SeqGAN is a good idea. Sun, X., Chen, X., Pei, Z., & Ren, F.[4] argue that generator can use seq2seq model which is used to generate a sentence response. And the discriminator designed to distinguish between the human-generated dialogues and machine-generated ones.
3. For multi-turn response selection, C. Tao, W. Wu, C. Xu, Y. Feng, D. Zhao, and R. Yan[12] used the pre-training a sentence-level and a session-level contextualized word vectors by learning a dialogue generation model from a large-scale human-human conversations with a hierarchical encoder-decoder architecture.
4. Regarding screenshot, as Figure 2 showing, turning the screenshot from RGB to grayscale, then binarization method transform grayscale to binary format(i.e.,black and white), then segmentation step identified rectangular bounding boxes wrapping the textual content of the images, then using the OCR to obtain Unicode text file.[6]

   But this paper just realized the screenshot text extraction, what about the nested pictures? Such part can deep research.
5. For speech recognition, the latent attributes like speaking style, accent, background noise, and recording condition should be considered. There are some ways like GMVAE to Tactron model [8] can independently control latent attributes, and is able to cluster them without supervision.
6. Also for the speech emotion recognition, Liu, Z. T., Wu, M., Cao, W. H., Mao, J. W., Xu, J. P., & Tan, G. Z.[13] also gave their idea, using the extreme learning machine(ELM) decision tree to select features.
7. Meanwhile I found Alibaba Group paper[5][9], which supporting an overview on AliMe system architecture, I can better understand the overall processing flow.

3. Defining the research question

The research question refers to the *Research Question* part.

4. Collecting the data and conduct experiments

Find some institutions or companies to provide data. There are also many resources online can utilize.

Dataset should be divided into Training and Testing dataset. For the three different hypotheses, try to program on the accessible platform. Using the training dataset to run the algorithms, then can gain the new data result. Using the testing data to

run the model again, calculate the precision and accuracy.

During the experiment may use the knowledge graph, encoder-decoder model, LSTM model and so on.

5. Analyzing the data

When I obtain the result, I can do some variance analysis. Also known as variance analysis or F test, it was invented by R.A. Fisher for the significance test of the difference in the mean of two or more samples. Due to various factors, the data obtained from the study showed fluctuations. The causes of fluctuations can be divided into two categories, one is the uncontrollable random factor, and the other is the controllable factor exerted in the study on the outcome. Variance analysis starts with the variance of the observed variables and studies which of the many control variables are variables that have a significant impact on the observed variables. At the same time, I can combine with the factor analysis. Factor analysis is to find intrinsic connections from a large amount of data, reducing the difficulty of decision-making.

6. Survey

In this step have already known which models will augment the customer service chatbot. Therefore, I will survey 50 people, and random sampling into two groups, each group has 25 people, group one chatting with the original chatbot, and the rest group testing with my augmented chatbot. The evaluation will be given one to three scores, one means *not satisfied*, two means *generally*, and three means *very satisfied*. Then compared the survey feedbacks and gain the customer experience.

7. Result

Conclude the final result, and reflect the pitfall, and discuss about the future work.

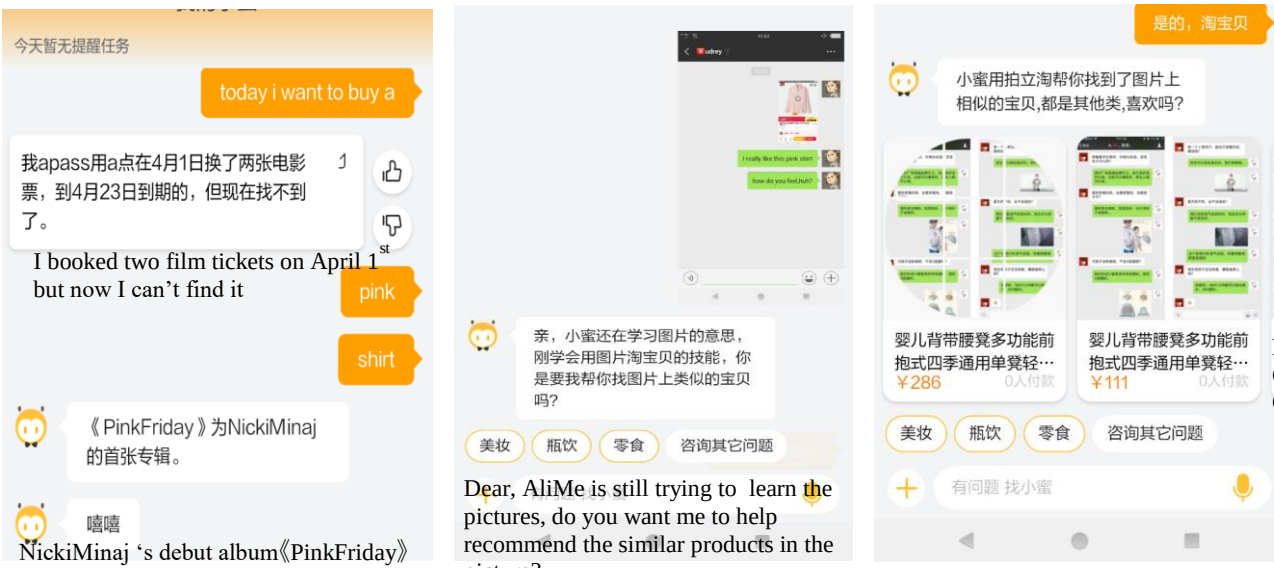


Figure 1. Sample of AliMe customer service chatbot dialogue scene

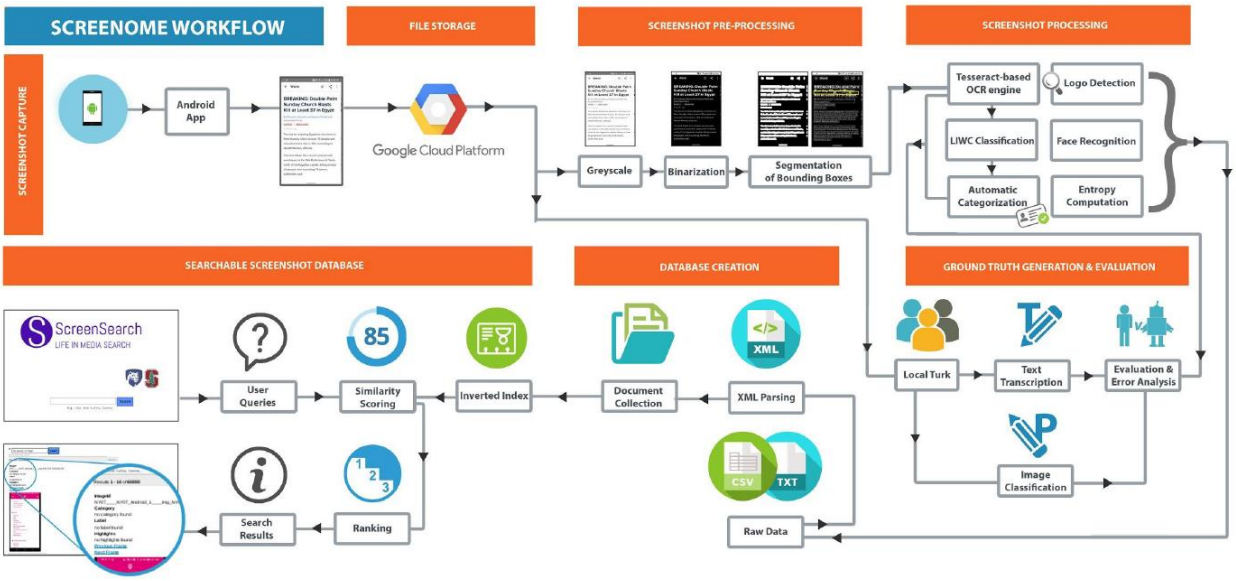


Figure 2. Overall Architecture for Smartphone Screenshot Processing, Indexing and Retrieval [6]

Time Plan

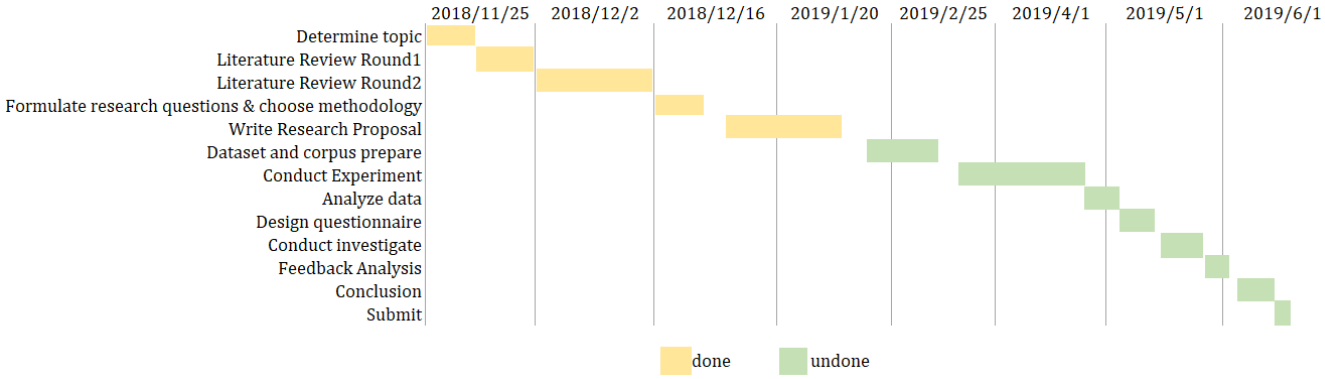


Figure3. Research schedule

Expected Outcomes

Through the relatively comprehensive literature review, I hope to understand the state-of-arts and can figure out the challenges of existing customer service chatbots. Then hope to reproduce the relative underlying technologies.

After implementing the conceptual framework and conducting experiments, I hope I can obtain the ground truth, and finally I realize the augmentation ways. For example, regarding to the screenshot recognizing, I hope the algorithms could help recognize the corresponding patterns, as the following ideal sample Figure 3, also can use some machine learning methods to improve the tracking effectiveness, even the sentences sending in more than one message bubbles.

At the same time, after realizing the models, the new generation customer service chatbot can understand the speech messages which having dialect or accent, and can understand better of the emotion in the customer sending voice message.

At last, I expect that the survey feedback of the group who using my augmented generation chatbot will be more positive than the group who not using.

Figure 3. Ideal screenshot recognition

Risk Management

1. DataSet &Corpus gaining, and the reproduce the underlying technologies are the two main risks. To manage these two risks, the best way to find an Authoritative organization or an enterprise platform, such public institutes have abundant resources to realize the experiment. The second way is to review similar theme papers in detail and discuss with professional experts, learning and practicing to understand deeply in the architecture. The dataset can search online, there are many corpus on website, and manually label data, using scientific methods to calibrate.
2. Regarding to the survey, the participants may worry about their dialogue privacy leak, might exist casual and invalid answers. During this process should protect participants from being anonymous, and improve the investigation work to let the respondent trust and willingness to answer the questionnaire seriously.

Discussion

In this paper, I have learned about the latest technologies, and found the technical challenges of current customer service chatbots, and formed a possible framework to enhance the capability of customer service chatbots. I can extract a model concept from research papers and reproduce the underlying technologies, then realizing the improving algorithms in the future.

But the acquisition of large-scale data and the accurate comparison of models will be difficult. Insufficient data or limitations in the corpus itself will affect the final result.

The final experimental results, if the results of the two groups of respondents found that the results of using my model are more optimistic, then the experiment is considered successful.

It is also necessary to reflect on the fact that I can also use the method of qualitative research, which is to measure social facts through the measurement of social facts, to determine the relationship between them and to explain the reasons for changes to guide social practice. Qualitative research focuses on the perspectives of participants, aiming to understand the phenomena of society, focusing on how different people understand the meaning of their lives, to reveal the internal dynamics of various social situations and the characteristics of human experience that are ignored or abandoned by quantitative research level.

In researching human-computer interaction and dialogue challenges, I can look for some participants to conduct a questionnaire survey and summarize some of the problems they use when using chatbot to conduct a qualitative analysis.Then I can get a more comprehensive results.

Overall, the literature review is really helpful for knowing this field, but the experiment is the actual challenges, not only the programming challenges but also the data and corpus obtaining.

Reference (IEEE format)

[1] A. M. Rahman, A. A. Mamun, and A. Islam, “Programming challenges of chatbot: Current and future prospective,” in 2017 IEEE Region 10 Humanitarian Technology Conference (R10-HTC), 2017, pp. 75–78, Dec,2017

[2] Accenture report: 70% of consumers like AI customer service [EB/OL], CTI Forum, http://www.ctiforum.com/news/world/503185.html, January 24th,2017

[3] Hill, J., Ford, W. R., & Farreras, I. G. (2015). "Real conversations with artificial intelligence: A comparison between human–human online conversations and human–chatbot conversations". Computers in Human Behavior, 49, 245-250.

[4] Sun, X., Chen, X., Pei, Z., & Ren, F. (2018, May)."Emotional Human Machine Conversation Generation Based on SeqGAN". In 2018 First Asian Conference on Affective Computing and Intelligent Interaction (ACII Asia) (pp. 1-6). IEEE.

[5] Li, F. L., Qiu, M., Chen, H., Wang, X., Gao, X., Huang, J., ... & Jin, G. (2017, November). "AliMe assist: An intelligent assistant for creating an innovative e-commerce experience". In Proceedings of the 2017 ACM on Conference on Information and Knowledge Management (pp. 2495-2498). ACM.

[6] Chiatti, A., Cho, M. J., Gagneja, A., Yang, X., Brinberg, M., Roehrick, K., ... & Giles, C. L. (2018). Text Extraction and Retrieval from Smartphone Screenshots: Building a Repository for Life in Media. arXiv preprint arXiv:1801.01316.

[7] S. R. Bowman, G. Angeli, C. Potts, and C. D. Manning, “A large annotated corpus for learning natural language inference,” arXiv:1508.05326 [cs], Aug. 2015.

[8] Hsu, W. N., Zhang, Y., Weiss, R. J., Zen, H., Wu, Y.,Wang,Y.,...&Nguyen,P.(Oct.018.)."Hierarchical Generative Modeling for Controllable Speech Synthesis". arXiv preprint arXiv:1810.07217.

[9] Yu, J., Qiu, M., Jiang, J., Huang, J., Song, S., Chu, W., & Chen, H. (2018, February). "Modelling domain relationships for transfer learning on retrieval-based question answering systems in e-commerce". In Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining(pp. 682-690). ACM.

[10] Singh, R., Paste, M., Shinde, N., Patel, H., & Mishra, N. (2018, April)."Chatbot using TensorFlow for small Businesses". In 2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT)(pp. 1614-1619). IEEE.

[11] F.Wei, “Improv Chat: Second Response Generation for Chatbot,” arXiv:1805.03900 [cs], May 2018.

[12] C. Tao, W. Wu, C. Xu, Y. Feng, D. Zhao, and R. Yan, “Improving Matching Models with Contextualized Word Representations for Multi-turn Response Selection in Retrieval-based Chatbots,” arXiv:1808.07244 [cs], Aug. 2018.

[13] Liu, Z. T., Wu, M., Cao, W. H., Mao, J. W., Xu, J. P.,&Tan, G. Z. (2018). "Speech emotion recognition based on feature selection and extreme learning machine decision tree" Neuro computing, 273, 271-280.