American Journal of Software Engineering and Applications
2021; 10(1): 11-18
http://www.sciencepublishinggroup.com/j/ajsea
doi: 10.11648/j.ajsea.20211001.12
ISSN: 2327-2473 (Print); ISSN: 2327-249X (Online)

A Technical Survey on the Modeling of Topical Bot

Aishna Gupta, Anuska Rakshit

Vellore Institute of Technology, Vellore, India

Email address:
aishna.gupta2019@vitstudent.ac.in (A. Gupta), anuska.rakshit2019@vitstudent.ac.in (A. Rakshit)

To cite this article:
Aishna Gupta, Anuska Rakshit. A Technical Survey on the Modeling of Topical Bot. *American Journal of Software Engineering and Applications.* Vol. 10, No. 1, 2021, pp. 11-18. doi: 10.11648/j.ajsea.20211001.12

Received: May 26, 2021; **Accepted**: July 8, 2021; **Published**: July 16, 2021

Abstract: This paper explains the working of the conversational AI models and their characteristic features. The primary objective of this paper is to let the readers know about what topical chats are and how they work. Topical chats have huge data/knowledge stored in them for making the conversation interactive and engaging with humans. The first-generation conversational AI models were simply focused on short task-oriented dialogs, such as telling jokes, the weather of the day, or playing songs. But now advanced models can have everyday smooth conversations. These models are built to understand different languages and their different accents. These models can identify whether the user is female/male/other, detect the change in the user's emotion during the conversation and switch the topic of discussion accordingly. Building a conversational AI model has been a challenging task for researchers as well as the developers as they require deep knowledge in NLU, ASR, LM, Semantics, etc. Understanding human emotions and sentiments is a difficult task for an AI model. Recognizing the speech and giving a sensible response is challenging too. But nowadays AI models are so developed that they can even differentiate between good words and slang words.

Keywords: Topical Chat, NLU, ASR, Inappropriate Response Filtering, Conversational AI

1. Introduction

Building a conversational AI model has been a challenging task for researchers as well as developers. An AI device that could carry on the conversation with humans in their native language could understand their accent, give sensible replies, and have immense knowledge and common sense reasoning. As technology develops, advancement in deep learning happened which opened the path for the solution to these problems. But then researchers faced problems in image recognition and speech recognition.

Voice assistants like Siri, Alexa were made, to achieve the goal of having a smooth conversion with humans. They are highly advanced and successfully deployed devices. But still having a broad conversation, including many different topics and vast knowledge is impossible. And the main limiting factor is the scarcity of conversational datasets and associated knowledge sources.

Challenging things for a conversational AI model are correlating with human emotions and sentiments, understanding the gender by pitch or tone, picking up the right search result for the user and for researchers having deep knowledge of NLU (Natural Language Understanding), ASR (Conversational Automatic Speech Recognition), Inappropriate Response Filtering and NLG (Natural Language Generation) is a crucial task.

This paper elaborates on the working of the topical chat and the challenges faced during the entire process.

2. Conversational AI

A. KNOWLEDGE-BASED CREATION

Knowledge-based data is collected from Wikipedia and articles which cover adequate topics and provide enough knowledge of the topic.

B. USER TOPIC DECISION PREDICTION

To estimate the engagement of the user in a social bot conversation a task is usually done on a suggested topic. However, it was observed that due to privacy concerns, the social bot has denied access to any of the personal private information about the users. Furthermore, for the user ID, the device address is not an authentic indicator as the device are used by multiple users. So, we can say that conversational interaction is the most appropriate way to profile the user.

Conversational Data: In the system, every conversation

begins with a greeting and ends with a stop command depending upon the user. Negotiation is done during the conversation using an explicit confirmation turn and records the user's binary decision i.e., accept or reject on the topic which is further recorded. Usually, the conversations are split into training, development, and test sets by 3/1/1. It has been observed that a total of 31,862 conversations with more than 5 user turns are used to curate the dataset for the topic decision prediction task.

Topic Decision Classifier: For evaluating the quality of the dialog, a metric is proposed known as topic-based metric that will describe the conversational ability of the topical bot to continue the coherent and interacting conversation on different topics. There are two most effective methods.

A feed-forward neural network (FFNN) is used to make binary predictions for each topic. The FFNN takes two inputs for each topic which includes:

1) An embedding $x_{t'}$ for the suggested topic at system turn t', and
2) At every user turn t, a user embedding vector z_t. [1]
3) Then Adam optimizer is used to train the FFNN classifier with the logistic loss. (Adam is an optimization algorithm that is generally used instead of the classical stochastic gradient descent procedure; based on training data, it helps to update network weights iteratively). Finally, a user-agnostic TopicPrior baseline is used. It builds a probability lookup using its acceptance rate on the training set for individual topics.

Deep Average Networks (DAN) can be adopted to detect conversation topics per utterance which will further help in training a topic classifier in many different query data's and question. Deep neural networks (DNNs), especially convolutional neural networks (CNN), have been widely used in recent years and demonstrated excellent performance in pattern recognition fields, such as speech recognition, image classification, and face recognition. [8]

3. Science

A. NLU

NLU is one of the most important segments in topical bots. It helps in selecting and producing system responses.

Segmentation of sentences: This is performed first as punctuation marks, double quotes, commas are not available in results of ASR and it helps in obtaining the appropriate units. ASR is responsible for identifying the human voice commands which include street names, landmarks, point of interests, distances, etc. Speech recognition does the speech to text conversion where the final output is the text corresponding to the recognized speech. [11]

Sentiment and Emotion Recognition: To provide high-quality responses, recognizing the sentiments and emotions of the user is one of the most important keys. This will lead to a meaningful and interesting conversation between the user and the device. There are 3 different types of sentiment in general; those are positive, negative, and neutral. Detecting these sentiments requires the ability to understand the difference in the tone and pitch of the user. The device should have built-in features that differentiate the normal/neutral pitch from positive, negative ones. This ability of understanding and simulating prosody is one of the most important parts of topical chat.

Sentiment analysis is done by using different methods like lexicon-based methods, deep learning, or a combination of both.

Emotion plays an important role in human life. Interpersonal human communication includes not only language that is spoken, but also non-verbal cues as hand and body gestures, tone of the voice, which are used to express feeling and give feedback and most importantly through facial expression. [10] Emotion recognition includes recognizing fear, happiness, anger, sadness, disgust, surprise, and many more human emotions which can be detected through a change in tone. It is a difficult task to recognize all these emotions. And that is why the perfect topical chat device still doesn't exist. There are times when Alexa or Siri or Google assistants say that they aren't able to get us. These devices can't understand human jokes, sarcasm, double-meaning words, slang, etc. It is difficult for AI to understand the meaning of words behind sarcasm. They give results on the basis of the literal meanings of words. Understanding these deep-down human emotions, sentiment and language is still impossible for AI and so as for topical chat devices. Even after providing the data of thousands of words, accents, jokes still people face problems in communicating with these devices efficiently.

Conversation Intent: Conversation intent was introduced in the default NLU model so that it can recognize the common natural ways of initiating the conversation like "Hello", "How are you", "Let's talk", "Tell me a joke" etc

Response generation: For generating a response we have four different approaches:

1) Rule-based
2) Retrieval
3) Generative
4) Hybrid

For a topical chat to give the best possible response, generally a huge amount of data from Wikipedia, online articles, and Reddit are extracted. Knowledge of adjectives, adverbs, pronouns, and slang are added to make the conversation more fun.

MELD is a multimodal multi-party conversational emotion recognition dataset. MELD contains raw videos, audio segments, and transcripts for multimodal processing. We believe this dataset will also be useful as a training corpus for both conversational emotion recognition and multimodal empathetic response generation. [5]

A functional system can be a combination of these techniques, or follow a waterfall structure or use a hybrid approach with complementary modules retrieval-based modules that try to identify an appropriate response from the dataset of dialogs available. Retrieval can be performed using techniques such as entity matching, N-gram matching, or similarity-based on vectors such as TF-IDF, word or sentence embeddings, skip-thought vectors dual-encoder system, etc.

Hybrid approaches utilizing retrieval in combination with generative models are genuinely new and have shown promising outcomes in recent years, typically with sequence-to-sequence approaches with some variants. [2]

Dialog Management: This is responsible for making the conversation engaging and finding the results related to the conversation. Once response generation is done, the dialog manager works in selecting the right response and the context. It also helps in smoothly changing or transiting the topics which the user can find interesting

Dialog Management: This is responsible for making the conversation engaging and finding the results related to the conversation. Once response generation is done, the dialog manager works in selecting the right response and the context. It also helps in smoothly changing or transiting the topics which the user can find interesting.

B. SPEECH RECOGNITION BY CONVERSATIONAL AUTOMATIC

Speech recognition systems can be separated in several different classes by describing what types of utterances they have the ability to recognize, these classes are classified as the following: Isolated Words, Connected Words, Continuous Speech, Spontaneous Speech [13]. Automatic Speech Recognition (ASR) is very important for voice-based assistants like google assistant. It's the advanced technology that helps us (humans) in talking to computer-based interface and to an extent the device recognize human conversations. ASR is based on or revolves around NLP (Natural Language Processing). NLP's work is to comprehend human speech and reply accordingly. Although having an accuracy of more than 95%, it still can't have a smooth interactive conversation between the device and the human. One of the majors is, even after having thousands of words, dictionaries, informal words, slangs it still is not enough when comes to having a comparison with humans. NLP enables computers to perform a wide range of natural language related tasks at all levels, ranging from parsing and part-of-speech (POS) tagging, to machine translation and dialogue systems. [9]

C. CONVERSATIONAL DATASETS AND COMMON-SENSE REASONING

Open source and pre-processed datasets can be used to get the information about trending topics, culture, preparing long and short-term memory, and integrating them within their dialog manager to make the responses seem as natural as possible. The term "Social Bot" is a superordinate concept which summarizes different types of (semi-) automatic agents. These agents are designed to fulfill a specific purpose by means of one- or many-sided communication in online media [17]. Moreover, to complement common sense reasoning, user satisfaction modules can be included to improve engagement along with the coherence of conversations.

D. CONTEXT AND DIALOG MODELLING

The most powerful component of a conversational agent is a robust system that can handle dialogs effectively. Tasks that can be accomplished by the system include:

1) Help break down the complexity of the open domain problem to a manageable set of interaction modes, and
2) Be able to scale the topic based on its diversity and breadth differs.

A common strategy is using a hierarchical architecture dialog modeling with the main Dialog Manager (DM) and multiple smaller DMs corresponding to specific tasks, topics, or contexts. The focus should be made not just on response generation but also on customer experience, and conversational strategies to increase engagement. Social bots received an obfuscated user hash code to enable personalization for repeat users.

E. CONVERSATIONAL TOPIC TRACKER

For detecting the conversation topics, Deep Average Networks (DAN) can be adopted and train a topic classifier on interaction data categorized into multiple matters. The conversational topic tracker was identified for various purposes such as sentiment analysis, conversation evaluation, entity extraction, response generation, and many more. A novel extension was made by adding topic-word attention to formulate an attention-based DAN (ADAN) that allows the system to jointly capture topic keywords in an utterance and perform topic classification. It was observed that a user's satisfaction correlates well with long and coherent on-topic conversations, while metrics of topic breadth may provide complementary information to user ratings, as the repetitiveness of topics is hardly captured in user ratings due to the intrinsic limitations of live user data collection [12].

F. SELECTION AND RANKING TECHNIQUES

In the case of rule-based rankers, the ranker chooses a response from the candidate responses obtained from sub-modules based on some logic. For model-based strategies, a supervised or reinforcement learning approach can be applied, trained on user ratings or on pre-defined large-scale dialog datasets such as Yahoo Answers, Reddit commentaries, OpenSubtitles, and much more. Higher scores are provided to the ranker to correct responses (e.g., follow-up comments on Quora are considered correct responses) while ignoring the incorrect or non-coherent responses obtained by sampling. To categorize responses and choose the best one, social bots need mechanisms to achieve the goal of having coherent and engaging conversations. When a reinforcement learning approach is used developed frameworks where the agent is a ranker, the actions are the candidate responses obtained from sub-modules, and the agent is trying to maximize the trade-off between satisfying the customer immediately versus taking into account the long-term reward of selecting a certain response. [3]

G. OFFENSIVE AND INAPPROPRIIATE SPEECH DTEECTION

To obtain high-quality conversational data in social bot interactions is one of the most strenuous aspects of delivering a positive experience. Several classes of inappropriate responses were identified such as 1) insulting response, 2) racially offensive responses, 3) hate speech, 4) sexual responses 5), profanity and 6) violent responses. In addition, a variety of Support Vector Machines and Bayesian classifiers were tested and trained on N-gram features using labelled ground truth data. The best accuracy results were in

profanity (>97% at 90% recall), racially offensive responses (96% at 70% recall), and insulting responses (93% at 40% recall). Dataset cleansing is also needed for online filtering of candidate social bots' responses prior to outputting them to ASK for text-to-speech conversion.

H. CONVERSATION EVALUATION

Social conversations are mostly open-ended. For example, if a user asks the question "what you think about BTS?", there are several angles of distinct, valid, and reasonable responses. This results in making training and evaluating extremely challenging. Objective metrics are also developed for conversational quality aligned with the goals of a social bot (the capacity to converse and engage in trending topics and latest events):

1) Conversational User Experience (CUX): To resolve these issues, average ratings are used from frequent users with multiple interactions who have their expectations established and evaluate the social bots in comparison to others.
2) Coherence: Using the annotations, the response error rate (RER) can be calculated for each social bot.
3) Engagement: The performance of conversations is evaluated and identified as being in alignment with social bots' goals.
4) Domain Coverage: Performance targeted on high entropy while minimizing the standard deviation of the entropy across multiple domains [3].
5) High entropy ensures that the social bots are talking about a variety of topics while a low standard deviation gives us the assurance that the metric is equally good for all the domains. Topical Diversity: Higher range of topics within each domain implies increase in topical affinity [3].
6) Conversational Depth: Evaluates social bot's ability to have multiple turn of conversations on specific topics within the five domains.

I. CHALLENGES IN NEURAL NETWORK SYSTEM

Research has shown that deep neural networks can be trained not only to perform classification but also to map complex functions. For instance, in language translation, sequence-to-sequence mapping is done. Moreover, this exact mapping can also be applied to conversational agents. Most of the neural response generation models follow the neural text generation frameworks, which include sequence-to-sequence (Seq2Seq), conditional variational autoencoder, and generative adversarial network (GAN). In spite of the various advantages, it fails to establish long-term emotional connections with users.

Semantics: Semantics can be referred to as the major component of a dialog system as a conversation is a semantic activity. Semantics can be in different levels of scope or abstraction [14]. A typical feature of a dialog system is that it suffers from a semantic issue that it often generates blank and dry responses, such as "I don't know", "OK", or simply repeats whatever the user is saying. From a technical viewpoint, it mainly involves the key techniques of natural language understanding and user understanding, including named entity recognition, entity linking, domain detection, topic and intent detection, user sentiment/emotion/opinion detection, and knowledge/ common sense reasoning [4].

Latent Semantic Analysis (LSA): Consider the following two sentences

1) This house needs a good clean.
2) He clean forgot about dropping the letters at the post office.

In the first sentence, the word 'clean' refers to an act of cleaning, and in the second sentence, it means completely.

These words can be easily distinguished by us because we have the ability to understand the context behind these words. However, as machines cannot understand such context it would fail to capture such context. In such a situation Latent Semantic Analysis (LSA) is used to capture the hidden concepts also known as topics.

Neural Response Generation Model: The encoder-decoder framework used in Neural response generation models consists of four components:

1) A decoder - It assigns additional probability mass to preferable words so that it can generate candidate responses. A method generally used to gain some control as to what to generate.
2) A ranker - Beam Search is used to generate multiple candidates in order to get diverse responses which are then ranked by another model that uses information that is either unavailable in decoding or is too expensive to use in decoding to select the final response [4].
3) An intermediate representation - A method has been proposed to use more flexible intermediate representations rather than encoding $X \oplus C$ using a fixed-size vector. This enhances the representation capability to address the one-to-many issue in the dialog system and to improve the interpretability of the representation to more readily control the response generation.
4) An encoder that encodes user input and dialog context - It encodes richer information and generates more informative responses. Moreover, it can be used to extract topic words using LDA and encodes such words in a topic-aware model.

Dialog Models based on grounded knowledge: Data stored in Wikipedia or Freebase are referred to as Knowledge facts. A knowledge grounded open-domain dialog system can identify the entities and topics mentioned in user input further linking them into real-world facts, retrieve related background information, and thereby respond to users in a proactive way that is by recommending new, related topics to discuss [4].

A knowledge-grounded model assists to generate a response by integrating few retrieved posts that are relevant to the input. The generation model (one of the systems in which the knowledge-grounded model is used) can generate a word in response from the context or the knowledge base. The DAN performs competitively with more complicated neural networks that explicitly model semantic and syntactic compositionality. [7] Then decoding is done which includes a graph attention mechanism in which the model initially takes care of a knowledge graph and then the decoder chooses a word to generate from either the graph or the common lexis. The plan is to provide the model with appropriate long-form

text as a input source of external knowledge. A reading comprehension is performed on this text in response to each conversational turn, thereby resulting in a more focused integration of external knowledge than prior approaches. Most of the studies focus on two challenges:

1) Knowledge-aware generation – providing the required knowledge into a generated response.
2) Knowledge selection – selecting appropriate knowledge that can be incorporated in the next response given in the dialog context and previously-selected knowledge.

Consistency: While searching for blogs on a specific topic, it has been observed that information seekers prefer blogs that place a central focus on that topic rather than mentioning the topic in a diffused format [4]. Therefore, to be able to focus on a particular topic consistency is needed.

A topical dialog system needs to encapsulate consistent behaviours so that it can gain the user's confidence and trust. However, there are three major consistency issues which include.

Persona Consistency: Can be grouped into two categories that address the dialog models:

1) Implicit personalization - the persona is implicitly represented by a persona vector-like proposing a ranking-based approach to integrate a personal knowledge base and user interests in a dialogue system. It utilizes learned user persona features to capture user-level consistency implicitly.
2) Explicit personalization- to generate personality-coherent responses given a pre-specified user profile. This explicit persona model controls the conversation generation using explicitly defined user profiles.

The chatbot's persona is defined by a key-value table which consists of name, gender, age, occupation, and many more. During generation, firstly the model chooses a key-value from the profile and then a response is decoded from the chosen key-value pair forward and backward.

Stylistic Response Generation: It is a form of personalization in conversation. It is closely related to domain adaptation and transfer learning. There are two main challenges:

1) To construct training data having pairs of responses that are of the same content but are usually in different approach.
2) To extract content and style in representation.

In order to solve the above problems an idea was proposed to train a general conversation model on a large corpus in the source domain then to transfer the model to a new speaker or target domain using small amounts of personalized (or stylistic) data in the target domain [4]. Some of the other proposed models include:

1) A two-phase transfer learning approach i.e., initialization then adaptation which can be used to generate personalized responses. Furthermore, a quasi-Turing test method can be implemented to evaluate the performance of the generated responses.
2) A multi-task learning approach where the response generation and utterance representation are treated as two sub-tasks for speaker role adaptation.

Contextual Consistency: Earlier work focused on representing better dialog contexts using hierarchical models, this was viewed as implicit modelling of contextual consistency. Recently, contextual consistency is noticed as a natural language inference (NLI) problem.

Interactiveness: Interactiveness refers to the system's capacity to execute complex social objectives such as entertainment and conforming by optimizing its behaviours and applying dialog strategies in multi-turn conversation. From a technical viewpoint, interactiveness mainly involves sentiment and emotion detection, dialog state tracking, topic detection and recommendation, dialog policy learning, and controllable response generation. However, optimizing the behaviours and strategies of a dialog system to maximize long-term user engagement and accomplish long-term, complex goals is still a major issue. To overcome this and improve interactiveness, it is important to understand the user's emotion and optimize the system's behaviour and interaction strategy in multi-turn conversations.

User Emotion Model: Emotion perception and expression is an important factor in building a human-like machine. To generate emotional responses, we have Emotional Chatting Machine (ECM) ECM consists of three components:

1) While decoding the internal emotional state gradually decays and finally reach zero.
2) Each decoding position is fed with an emotion category embedding.
3) An external memory that permits the model to select emotional or generic words.

Patterns have been observed in human-human conversations such as empathy and comfort, which would inspire a more delicate but firm design of emotional communication between humans and machines.

Another method of affective response generation was developed which consists of three components:

1) For searching effective responses - the effective beam search algorithm.
2) To increase or decrease the affective consistency between a post and a response - the effective loss functions.
3) To supplement word vectors - the effective vectors based on Dominance dimensions.

To generate the reviews of a particular polarity, a multi-class generative adversarial network was proposed which consists of generators for multiple polarities and discriminators. A challenging issue in this is emotion representation.

With the massive increase in social interactions on online social networks, there has also been an increase of hateful activities that exploit such infrastructure. Detecting such hateful speech is important for analyzing public sentiment of a group of users towards another group, and for discouraging associated wrongful activities. [6]

Emotional representation is also a very major issue in the existing model. One of the simple approaches is to project implicit subtle emotion label to a vector. However, this approach failed to explicitly model the user's emotional changeover during a conversation. it would try to cheer the user up through e.g., shifting to new discussion that are more comfortable for both parties.

This model is crucial for a dialog system to establish a

long-term connection with a user because the user is more interested to engage with the system if the system can always detect a negative transformation in her emotion during the conversation it would try to cheer the user up through e.g., shifting to new discussion that are more comfortable for both parties.

Strategy of Conversation Behaviour: A framework was built to capture the abstract emotions such as politeness strategies freedom concepts that are used to start a conversation and examined their relation. When is framework is applied in a controlled environment it is possible to detect early warning signs of antisocial behaviour in online conversation. Firstly, to detect signs of deadlock a retrieval-based method is proposed and then the response is received containing the entities related to the input.

A proactive suggestion method was proposed for the user that would provide a look-ahead post in addition to the system response, circumstances, and prior generated response. The user can use the generated post directly or type a new one during the conversation. However, asking upright questions in conversation can be shown as an important proactive behaviour. Thus, a typed decoder is proposed to generate meaningful questions by predicting a type distribution over topic words at each decoding position. The final output distribution was modelled by the type distribution, leading to a strong control over the question to be generated. In addition, a dataset of clarification questions was evaluated and a neural network model was built for ranking clarification questions. The most important drift for future research on this field are:

1) The comprehensive investigation of conversation behaviours in the human-human dialog.
2) The second is to create a more sophisticated real-world dialog setting for system development and evaluation [4].

J. SPEAKER GENDER ANALYSIS

Gender detection systems based on Gaussian Mixture Models, i-vectors and Convolutional Neural Networks (CNN) were trained using an internal database of 2,284 French speakers and evaluated using REPERE challenge corpus out of which the CNN system obtained the best performance with a frame-level gender detection F-measure of 96.52 and a hourly women speaking time percentage error below 0.6% [15]. The data is extracted from SwDA and analysis is done on the speaker gender information to check whether latent modes of unsupervised learning in the dynamic speaker model could pick up some gender language variations. The process involves gathering the latent mode association scores for every 32 modes i.e. the no of modes differs according to the input data, which was computed in Latent Model Analyzer.

Then to test the associate score distributions of male vs female utterances group mean tests for individual modes are carried out. The strength of the difference is measured using the most effective way to measure effect size - The cohen-d score.

Cohen's D Formula is computed as:

$$D = (M1 - M2) / SP$$

Where,

M1 and M2 denotes the sample means for groups 1 and 2

SP denotes the pooled estimated population standard deviation.

Lastly, the p-value is computed using the Mann-Whitney U test. The Mann-Whitney U test is generally used to work around the underlying assumption of normality in parametric tests. However, this method is used when the validity of the assumptions of the t-test is not certain thus this test has wider applicability. Cohen's d relies on the pooled standard deviation (the denominator of equation) to standardize the measure of the ES; it assumes the groups having (roughly) equal size and variance [16].

Therefore, the group mean tests are carried out on the following three sets:

1) All conversations,
2) Conversations involving either males or females, and
3) Conversations involving both genders.

The Cohen's D scores for the group mean tests are shown below graphs.

Previously in Telephone Conversations, numerous researches were done to analyse the difference between genders. The main motive in analysing the gender linguistic differences in genders was of two folds.

1) From the scientific perspective, it can increase the understanding of language production.
2) From the engineering perspective, the performance of a number of natural language processing tasks, such as text classification, machine translation, or automatic speech recognition by training better language models can be improved.

The methods which are used for characterizing the differences between genders and gender pairs are similar to the processes used for text classification. Studies have shown that the most effective ways for characterizing the differences between gender categories. Firstly, the transcript of each speaker has to be classified i.e., based on the appropriate gender category. And the second approach involves the application of feature selection methods, this would reveal the most characteristic features for each gender.

From each set, the most female-like mode (with the most positive Cohen-d score) and the most male-like mode (with the most negative Cohen's-d score) are identified. The significant advantage of using gender information for automatic speech recognition is that it can be robustly detected using acoustic features.

4. Framework of Social Bot Management

A. INNOVATION AND FEEDBACK FRAMEWORK:

The welcome and feedback systems are implemented individually coordinated through a framework called Links that can be used to combine multi-stage conversational experiences. To improve the accuracy of rating capture, support can be added for fractional numbers. Ratings and feedback on each conversation should be shared daily to drive incremental model improvements.

B. UPTIME MEASUREMENT AND AVAILABILITY MONITORING

This is the major key to deliver a successful experience to millions of users that help to maintain reliability even in

cases when one or more of the social bots is offline or malfunctioning. Thus, a monitoring system was developed that collects availability metrics on each social-bot and removes social bots that are not responding properly to the user's input. This was achieved with a combination of passive monitoring for failure modes in user traffic, as well as active monitoring via simulated traffic to each social bot [3]. A notification and reminder system, with self-service reactivation, can be installed to track any kind of response to any issues in their social bot and bring them back online as soon as possible. The system also enabled us to deactivate social bots that produced inappropriate responses.

C. MANAGING CUSTOMER EXPERIENCE (CX)

To minimize the risks of social bots trained on potentially problematic data sets with offensive contents, such as some unverified datasets which are publicly available on the Internet, a system was developed to monitor conversations for offensive responses from social bots. If a user-initiated a conversation about an inappropriate topic, such as sex, social bots will redirect the conversation by suggesting topics they could chat about. In such situations, the social-bot will be deactivated and the notification and reminder system mentioned will notify the team and allow them to reactivate it after they have addressed the issue [3].

A more sensitive and contextually aware classifier that would identify the following is recommended to be installed:

1) Racially inflammatory content.
2) Other hate speech.
3) Violent content.
4) Profane content, and
5) Sex content.

5. Conclusion

A future empathetic machine can distinguish a user's emotional state and change and deliver emotionally influential conversations. A robust NLU system leads to high coherence when supported by strong domain coverage. All different response generation techniques in particular retrieval, generative and hybrid mechanisms are required in a system to achieve various conversation goals from different angles. A detailed discussion is done on the challenges faced by neural networks that are semantics, consistency, and interactiveness. To gain human trust a coherent personality is important for a social bot. However, the most challenging problem in the artificial intelligence field is to ensure personality-coherent behaviors in conversations and evaluate such behaviors from the perspectives of multi-disciplines. Two modes that are female speakers and male speakers have been identified on the basis of the Cohen-d score from this observation is made those female modes are mostly agreement, acknowledgment, and backchannel whereas male modes have several filled pauses.

Acknowledgements

We would to thanks our respected professor and colleagues, who helped us in every possible way, whenever we needed in doing this wonderful research paper. We learnt and came to know about so many new things related to Conversational AI, NLU and many more.

References

[1] Hao Cheng Hao Fang Mari Ostendorf, 2019. A Dynamic Speaker Model for Conversational Interactions, Proceedings of NAACL-HLT 2019, pages 2772–2785.

[2] Jurgita Kapočiūtė-Dzikienė, 2020. A Domain-Specific Generative Chatbot Trained from Little Data. Appl. Sci. 2020, 10, 2221; doi: 10.3390/app10072221.

[3] Ashwin Ram, Rohit Prasad, Chandra Khatri, Anu Venkatesh, Raefer Gabriel, Qing Liu, Jeff Nunn, Behnam Hedayatnia, Ming Cheng, Ashish Nagar, Eric King, Kate Bland, Amanda Wartick, Yi Pan, Han Song, Sk Jayadevan, Gene Hwang, Art Pettigrue, 2018. Conversational AI: The Science Behind the Alexa Prize. arXiv: 1801.03604.

[4] Minlie Huang, Xiaoyan Zhu, Jianfeng Gao, 2020. Challenges in Building Intelligent Open-domain Dialog Systems, arXiv: 1905.05709.

[5] Soujanya Poria, Devamanyu Hazarika, Navonil Majumder, Gautam Naik, Erik Cambria, Rada Mihalcea, 2019. MELD: A Multimodal Multi-Party Dataset for Emotion Recognition in Conversations arXiv: 1810.02508v6.

[6] Pinkesh Badjatiya, Shashank Gupta, Manish Gupta, Vasudeva Varma, 2017 Deep Learning for Hate Speech Detection in Tweets, arXiv: 1706.00188v1.

[7] Mohit Iyyer, Varun Manjunatha, Jordan Boyd-Graber, Hal Daume, 2015. Deep Unordered Composition Rivals Syntactic Methods for Text Classification.

[8] Dawei Dai, Weimin Tan, Hong Zhan, 2017. Understanding the Feedforward Artificial Neural Network Model From the Perspective of Network Flow.

[9] Tom Young, Devamanyu Hazarika, Soujanya Poria, Erik Cambria, 2018. Recent Trends in Deep Learning Based Natural Language Processing, arXiv: 1708.02709v8.

[10] Ashwini Ann Varghese, Jacob P Cherian, Jubilant J Kizhakkethottam, 2015. Overview on emotionrecognition system, DOI: 10.1109/ICSNS.2015.7292443.

[11] Pooja Withanage; Tharaka Liyanage; Naditha Deeyakaduwe; Eshan Dias; Samantha Thelijjagoda, 2018. Road Navigation System Using Automatic Speech Recognition (ASR) And Natural Language Processing (NLP), DOI: 10.1109/R10-HTC.2018.8629859.

[12] Fenfei Guo, Angeliki Metallinou, Chandra Khatri, Anirudh Raju, Anu Venkatesh, Ashwin Ram, 2018. Topic-based Evaluation for Conversational Bots, arXiv: 1801.03622v1.

[13] M. A. Anusuya, S. K. Katti, 2009. Speech Recognition by Machine: A Review, http://sites.google.com/site/ijcsis/ ISSN 1947-5500.

[14] Xiaoping Sun, Xiangfeng Luo, Jin Liu, Xiaorui Jiang, Junsheng Zhang, 2015. Semantics in Deep Neural-Network Computing, doi: 10.1109/skg.2015.42.

[15] David Doukhan; Jean Carrive; Felicien Vallet; Anthony Larcher; Sylvain Meignier, 2018. An Open-Source Speaker Gender Detection Framework for Monitoring Gender Equality, DOI: 10.1109/ICASSP.2018.8461471.

[16] Cristiano Ialongo, 2016. Understanding the effect size and its measures, doi: 10.11613/BM.2016.015.

[17] Christian Grimme, Mike Preuss, Lena Adam, and Heike Trautmann, 2017. Social Bots: Human-Like by Means of Human Control?, arXiv: 1706.07624v1.