ICSEE2019 International conference on systems energy and environment
JULY 2019

A Deep Learning Approach for Polite Dialogue Response Generation

Resmi P
Department of Computer Science and Engineering
Government Engineering College
Palakkad, India
reshma.resmi81@gmail.com

Naseer C
Department of Computer Science and Engineering
Government Engineering College
Palakkad, India
naseercholakkalathu@gmail.com

***Abstract*—Dialogue generation is the act of creating response sentences that are contextually relevant, meaning full and grammatically correct. A natural language sentence is characterized by different features including linguistic features and stylistic features. In a dialogue generation system, the generated response can be polite, rude or neutral. The main objective of this proposed system is to generate polite dialogue responses by using deep learning techniques. The system mainly contains two modules: the first module implements a politeness classifier and the second module implements a response generation system. Politeness classifier calculates the politeness score for the input sentence and by using this score the response is generated. For implementing the classifier the Stanford politeness corpus is used and for implementing dialogue generation model the Cornell movie dialogue corpus is used. The advantages of Bi-directional Long Short Term Memory (BLSTM) is used for building the system. Input to the system is a sentence which will be processed by the classifier and the output of the classifier and the sentence after vectorization is given to the dialogue generation module which will generate the final output. The main advantage of this system is that it is created by a deep learning model which will perform well in large dataset.**

***Keywords*—Linguistic features, Dialogue generation, Deep learning, BLSTM(Bidirectional Long Short Term Memory, Politeness, politeness Score**

I. INTRODUCTION

In this technical world, automatic text generation is an important step in many applications. The introduction of conversational agents and personal assistant systems changed the traditional concept of interacting with machines. Nowadays this kind of interaction is more common so we are learning style from these kinds of interactive systems. Generating stylistic and human-like language is crucial for developing, engaging, and convincing trustworthy conversational agents, such that it can be used in dialogue generation systems. All such system uses Natural language processing as the basic building block. Every language has some parameters that indicate the linguistic and stylistic features. Politeness is one of such feature present in languages. A sentence or a response text can be polite, rude or neutral based on the style and words used in it.

Fig. 1. Overview of the system

Generating stylistic dialogue responses could be a considerably difficult task as a result of generated response needs to be syntactically and semantically fluent, contextually-relevant to the conversation. This is often any difficult by the very fact that content and style are only available in separate datasets. In translation type dataset we can see the parallel dataset which contain the source-to-destination data. Hence, we'd like indirectly-supervised models that may incorporate vogue into the generated response in absence of parallel data while still maintaining conversation relevance.

A politeness classifier is a model that will predict the politeness score given a sentence. There are various strategies for building a classifier, includes machine learning and deep learning. This classifier is trained to predict the politeness level present in the input so that we can classify it as either polite or rude. Dialogue generation involves the understanding of the context of the input text and analyzing its meaning to generate the corresponding reply. So this work aims to create a polite response generation system. The basic idea of the proposed system is shown in Fig 1.

The main task in this system involves understanding the context of input sentence and generating the output sentence with required politeness level. The main challenge for building a dialogue generation system is that, it must generate sentence that are contextually relevant and grammatically correct. This task can be completed by using Deep neural network which is an efficient and effective technology, that performs well

in the case of sequence prediction, image classification etc. Controlling the politeness level in natural language has so many applications such as in chatbot, conversational agents and in everything that uses automatic natural language text generation.

Rest of the paper is organized as follows. Section II describe some previous work done in this area. Section III describe the design and implementation of a politeness classifier and Generation of dialogue conditioned on required politeness level is also described in this section. Section IV discuss about the results and evaluation measures. Finally section V concludes the discussion with a description on future works and finally added various references used for the work.

II. RELATED WORK

In sociolinguistics, a style in a text is a set of linguistic features with specific meanings. In this context, social meanings can include group membership, personal attributes, or beliefs. Linguistic variation is the concept of linguistic style–without variation, there is no basis for distinguishing social meanings. Variation can occur syntactically, lexically, and phonologically. Many approaches are there for interpreting and defining a style which incorporates the concepts of indexicality, indexical order, stance-taking, and linguistic ideology. The style for a person is not fixed. It may vary depending on the context and some other social factors. Additionally, speakers often incorporate elements of multiple styles into their speech, either consciously or subconsciously, thereby creating a new style.

One of the main challenge involved in style transfer of text is the un-availability of the parallel dataset(parallel datasets for regular-to-stylistic pairs are usually unavailable). In the work proposed by Z. Fu et al. [1], they explore two models for style transfer without using parallel dataset and try to find a solution for the problem of lacking parallel training data and hardness to separate the style from the content. The first model is a multi-decoder seq2seq model where the encoder is used to capture the content c of the input X and the second one is the multi-decoder contains n(n 2) decoders to generate outputs in different styles. This is an exploratory evaluation of style transfer, so instead of finding the best set of parameters, a comparison between different parameters in different model is performed here.

Written prose is a method in which we impart our thoughts and feelings to one another. The same message can be conveyed in different ways. Distinct wording may pass different levels of politeness or familiarity with the reader. Style transfer, or stylistic paraphrasing, is the task of rewriting a sentence preserving the meaning of the text but alter the style. The problem of style transfer is relevant to the creation of natural language generation systems. K. Carlson et al. [2] suggests that style can be viewed as a machine translation problem where the source language and target language are in different style. They have created a corpus for style transfer which can be used for various natural language tasks. They train statistical machine translation models and encoder decoder recurrent neural networks on the created corpus and both the models performed well.

Unsupervised text style transfer models are autoencoder model, encoder-decoder model, and adversarial training model. All these methods use classifier as the discriminator. Un-supervised text style transfer requires learning disentangled representations of attributes (e.g., negative/positive sentiment, plaintext/ciphertext orthography) and underlying content. A method by Z. Yang et al. [3] uses a locally normalized language model as discriminators. This method consists of two parts: reconstruction and transfer. Training of the LM is in adversarial way, such that it will minimize the loss of language model for real sentences and maximizes the loss of language model for transferred sentences. They found that the adversarial language model outperforms the traditional binary classifiers discriminators.

Discussions in M. Aubakirova and M. Bansal, [4] presents a neural network model for identifying politeness in request without considering any hand featured engineering. It is a simple Convolutional Neural Network(CNN) based model which do not take any syntactic features that identifies politeness, but it perform better than any such feature-based models. Several network visualizations based on activation clusters, first derivative saliency, and embedding space transformations, helping us automatically to identify several subtle linguistics markers of politeness theories. Positive politeness strategies focus on making the hearer feel good through offers, promises, and jokes. Negative politeness examples include favor seeking, orders, and requests. Differentiating among politeness types is a highly nontrivial task, because it depends on factors such as a context, relative power, and culture [4]. To predict the politeness level in requests they have created a neural network model. They have discovered some novel activation clusters, potentially corresponding to new politeness strategies: indirect pronoun and punctuation.

Two kind of languages are there: honorific and non honorific. Many languages use honorifics to express politeness, social distance, or the relative social status between the speaker and their addressee(s). In machine translation if we are translating a sentence from a language without honorifics such as English, it is difficult to predict the appropriate honorific level. R. Sennrich et al. [5] propose a simple and effective method for including target-side T–V annotation in the training of a neural machine translation (NMT) system, which translate the sentence in source language into sentence in target language by allowing us to control the level of politeness at test time through side constraints.

The method proposed by Danescu-Niculescu-Mizil et al. [6] , develops a computational framework for identifying and characterizing politeness marking in requests. They have identified different type of politeness in requests. The method by F. Mairesse and M. Walker [7] suggest a method for neural dialogue response generation that allows generating semantically reasonable responses according to the dialogue history, and explicitly controlling the sentiment of the response via sentiment labels. A Conditional Generative Adversarial

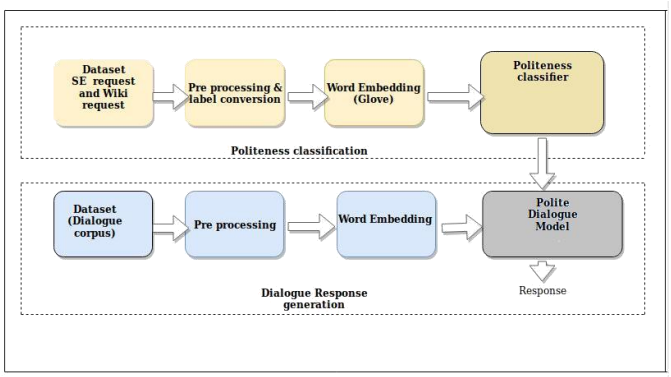

Fig. 2. System Architecture

network (CGAN)-based framework for sentiment controlled dialogue generation is used. In this framework, the desiderata of fluency and controlability are explicitly enforced by creating a model with two subcomponents: a generator and a discriminator.

III. METHODOLOGY

The proposed work implements a system to generate polite dialogue responses using deep learning technique. The system is implemented in two main phases which are listed below.

Politeness classifier

- Data collection and preprocessing
- Word embedding
- Model building
- Calculating politeness score

Dialogue generation

- Dataset collection and Annotating politeness score
- word embedding
- Encoder decoder model
- Creating responses

A. Politeness classifier

The politeness classifier mainly contains four sub modules: Dataset collection and preprocessing, word embedding using Glove[12], model building using deep neural network training and testing. Here, the bidirectional LSTM is used for model building, that capture the long term dependencies in the input sequence. Bidirectional LSTMs are used as the extension of LSTM network. Traditional LSTM can predict the target sequence by looking only future information. But BLSTMs can predict target by considering past and future information, one LSTM for each direction. In problems where all information at a particular time of the input sequence are available, Bidirectional LSTMs train two LSTMs on the input sequence. In the first LSTM the input sequence as it is trained and in the second LSTM a reversed copy of the input sequence is trained. This will give more context information to the network.

Dataset and preprocessing Stanford politeness corpus and Cornell dialogue corpus are the two dataset used for building this system. Stanford politeness corpus is used for training the politeness classifier which contain a collection of requests from Wikipedia editors talk pages and the Stack Exchange (SE) question answering communities. These requests are assigned with a politeness score by human annotators, by using this score these requests are assigned with a label which is either polite or rude. Cornell movie dialogue corpus contain short conversations from scripts, annotated with character metadata. It contain 220000 dialogues with character metadata.

Preprocessing involves sentence tokenization of the data, removal of stop words from the data and removal of the unwanted links and symbols. The unwanted symbols and links are removed first. Then the sentence tokenization is performed using tokenizer function in keras. After tokenization a dictionary like object is create which contain tokens and corresponding index value. The unwanted symbols and links are removed first. Then the sentence tokenization is performed using tokenizer function in keras. After tokenization a dictionary like object is created which contain tokens and corresponding index value.

GloVe is the abbreviation for global vectors for word representation[18]. It is an unsupervised learning algorithm developed by Stanford for generating word embeddings by aggregating global word-word co-occurrence matrix from a corpus[18]. In the proposed system a 100-dimensional word embedding model is used for building the classifier.

Deep learning model for classifier The classifier model contains an embedding layer which is used for embedding the input data, a Bidirectional LSTM layer which capture the long term dependency in the input sentence, a convolution layer that capture the features in input, a drop out layer used for avoiding overfitting , a max pooling layer and finally a dense layer which will generate the desired output. Figure 3.2 shows the architecture of BLSTM. The existing methodology only uses a convolutional layer for classifier, which is learned to predict politeness score without using any hand featured engineering. BLSTM contain a forward and backward layer so that, for predicting a target the model uses past and future information. Hence more accurate result will be obtained.

B. *Dialogue generation*

For comparison, a seq2seq encoder decoder model is used for generating the response sentences in addition to the proposed method. The Encoder-Decoder architecture with recurrent neural networks is the most effective deep learning approach for neural machine translation and sequence to sequence prediction. This method shows reasonable performance.

For dialogue generation, we use the popular Cornell dialogue corpus which contains 245K conversations. In pre-processing step, this dataset is divided into two part question set and answer set. Each set equally contains 150000 data. Each conversation is converted to a vector for inputting to the encoder decoder model.

Encoder decoder model for sequence prediction: An Encoder-Decoder design was created where an info succession was perused in sum and information arrangement is encoded

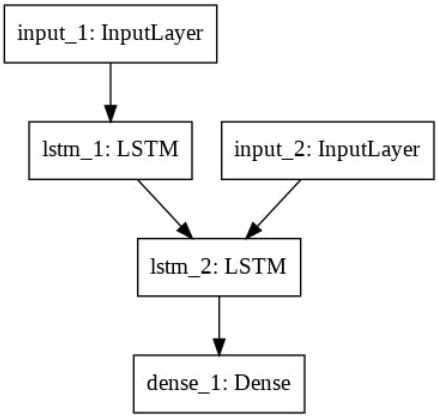

Fig. 3. Architecture

to a fixed-length encoded portrayal . A decoder system is compelled to anticipate the yield sentence until the sentence limit came to. A solitary layer of LSTM systems were utilized for both the encoder and decoder [8].

We start with input sequences from question sentences and corresponding target sequences from another set(answer sentences). An encoder LSTM turns input sequences to 2 state vectors. The decoder will use this value in teacher forcing part.

IV. EVALUATION AND RESULTS

Accuracy (percentage of correctly labeled messages for binary polite/rude labels) to evaluate our politeness classifiers generalization ability. The proposed method achieves accuracy of 87.5%. This is a good measure compared to previous SVM classification model, CNN model and LSTM-CNN model.

TABLE I
CLASSIFICATION RESULT: THE BEST RESULT IS BOLDFACED

Method	Accuracy
SVM	82.6%
CNN	85.8%
LSTM-CNN	85.0%
BLSTM-CNN	87.29%

Comparison of result between different methods used for classification is given in Table 1.

Training accuracy of the politeness classifier is 93% and accuracy of sequence to sequence model is 99% and for encoder decoder model is 89%. Table 1. represent the result of politeness classifier with some randomly-selected responses from dialogue corpus and their politeness classifier scores. We can see that the classifier provides a reasonably correct score a majority of the time. This model capture several features for politeness because we are using BLSTM–CNN network

Since there were no ground truth metrics for evaluating dialogue quality we can evaluate the proposed system manually or by calculating BLEU score. The Bilingual Evaluation Understudy Score, or BLEU for short, is a metric for evaluating a generated sentence to a target sentence [15]. A perfect match between the source and target results in a score of 1.0, whereas a perfect mismatch results in a score of 0.0. The score was

Target sequence	**Score**
Polite examples	
Hi, how are you	0.89
I know amazing	0.89
Well thanks. I appreciate that	0.99
Rude examples	
Are you a bad boy?	0.05
Then she is a liar	0.04
Oh, excuse me to all hell	0.02

Fig. 4. Result from politeness classifier

developed for evaluating the predictions made by automatic machine translation systems. It is not perfect, but does offer 5 compelling benefits: It is quick and inexpensive to calculate, It is easy to understand, It is language independent, It correlates highly with human evaluation, It has been widely adopted. The proposed system achieves BLEU score of 0.65. But in our model we can not completely rely on BLEU score for dialogue quality.

In order to compare the dialogue quality of two dialogue generation model, we need similar vocabulary set and dataset. The previous method they used MovieTriple dialogue corpus. So 300 data were randomly extracted from Cornell Dialogue corpus and tested in seq2seq model and encoder decoder model. The seq2seq model generates sentences with limited length. But encoder-decoder model generates sentences with reasonable dialogue quality and contextually relevant responses.

V. CONCLUSIONS AND FUTURE WORK

A polite dialogue response generation system using deep learning technique is implemented here. In dialogue generation system we can create responses based on particular style. Two deep learning models are presented here that can generate polite dialogue responses. First one is a sequence to sequence generation model which is used for comparing with the second one, and an encoder decoder based text generation model. The politeness classifier is a Bidirectional LSTM model that achieves accuracy of about 82.6% which is a good measure when comparing with the previous models. The classifier will predict the probability of the sentence for being polite. This score is used for generating the responses. Training data for dialogue generation is annotated with politeness score by using this classifier. In encoder decoder model, input is given to the encoder LSTM using the politeness classifier and output is obtained from the decoder LSTM. This dialogue generation model is compared with

seq2seq model and LSTM model. Proposed method achieves best BLEU score.

In future work, by using more advanced neural networks such as variational, adversarial, and decoder-regulation techniques, we can improve the dialogue quality. Here the encoder-decoder model is used with movie dialogue corpus as training data. Only the politeness is considered for the scope of this project, dialogues can be generated by constraining more stylistic parameters [9].

ACKNOWLEDGMENT

We extend our sincere gratitude to all the teaching staff, Department of computer science engineering, Government Engineering College Palakkad, for their valuable guidance and support at each stage of the project.

We are also thankful to our parents for the support given in connection with the project. Gratitude may be extended to all well-wishers and friends who supported us to complete the project in time.

REFERENCES

[1] Z. Fu, X. Tan, N. Peng, D. Zhao, and R. Yan, “Style transfer in text: Exploration and evaluation,” in Thirty-Second AAAI Conference on Artificial Intelligence, 2018.

[2] K. Carlson, A. Riddell, and D. Rockmore, “Evaluating prose style transfer with the bible,” Royal Society open science, vol. 5, no. 10, p. 171920, 2018.

[3] Z. Yang, Z. Hu, C. Dyer, E. P. Xing, and T. Berg-Kirkpatrick, “Unsu-pervised text style transfer using language models as discriminators,” in Advances in Neural Information Processing Systems 31 (S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Gar-nett, eds.), pp. 7287–7298, Curran Associates, Inc., 2018.

[4] M. Aubakirova and M. Bansal, “Interpreting neural networks to improve politeness comprehension,” arXiv preprint arXiv:1610.02683, 2016.

[5] R. Sennrich, B. Haddow, and A. Birch, “Controlling politeness in neural machine translation via side constraints,” in Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pp. 35–40, 2016.

[6] C. Danescu-Niculescu-Mizil, M. Sudhof, D. Jurafsky, J. Leskovec, and C. Potts, “A computational approach to politeness with application to social factors,” arXiv preprint arXiv:1306.6078, 2013.

[7] X. Kong, B. Li, G. Neubig, E. Hovy, and Y. Yang, “An adversarial ap-proach to high-quality, sentiment-controlled neural dialogue generation,” arXiv preprint arXiv:1901.07129, 2019.

[8] K. Cho, B. Van Merrienboer,¨ D. Bahdanau, and Y. Bengio, “On the properties of neural machine translation: Encoder-decoder approaches,” arXiv preprint arXiv:1409.1259, 2014.

[9] D. Bahdanau, K. Cho, and Y. Bengio, “Neural machine translation by jointly learning to align and translate,” arXiv preprint arXiv:1409.0473, 2014.

[10] J. Lehmann, R. Isele, M. Jakob, A. Jentzsch, D. Kontokostas, P. N. Mendes, S. Hellmann, M. Morsey, P. Van Kleef, S. Auer, et al., “Dbpedia–a large-scale, multilingual knowledge base extracted from wikipedia,” Semantic Web, vol. 6, no. 2, pp. 167–195, 2015.

[11] E. Pavlick and J. Tetreault, “An empirical analysis of formality in online communication,” Transactions of the Association for Computational Linguistics, vol. 4, pp. 61–74, 2016.