ABSTRACT

Natural language generation (NLG) plays a critical role in spoken dialogue systems (SDSs) and aims at converting a meaning representation, *i.e.*, a dialogue act (DA), into natural language utterances. NLG process in SDSs can typically be split up into two stages: sentence planning and surface realization. Sentence planning decides the order and structure of sentence representation, followed by a surface realization that converts the sentence structure into appropriate utterances. Sentence planning is to map input semantic symbols onto a linguistic structure, *e.g.*, a tree-like or a template structure. Surface realization is then to convert the structure into an appropriate sentence. Conventional methods to NLG rely heavily on extensive hand-crafted rules and templates that are time-consuming, expensive and do not generalize well. The resulting NLG systems, thus, tend to generate stiff responses, lacking several factors: adequacy, fluency and naturalness. Recently, taking advantages of advances in data-driven and deep neural networks (DNNs), NLG has received much attention in the study. Recent advances in data-driven and deep neural networks (DNNs) methods have facilitated investigation of NLG in the study. DNN methods to NLG for SDS have demonstrated to generate better responses than conventional methods concerning factors as mentioned above. Nevertheless, when dealing with the NLG problems, such DNN-based NLG models still suffer from some severe drawbacks, namely completeness, adaptability and low-resource setting data. Thus, the primary goal of this dissertation is to propose DNN-based generators to tackle the problems of the existing DNN-based NLG models.

Firstly, we present gating generators based on a recurrent neural network language model (RNNLM) to overcome the NLG problems of completeness. The proposed gates are intuitively similar to those in the Long short-term memory (LSTM) or Gated recurrent unit (GRU) to restrain the gradient vanishing and exploding. In our models, the proposed gates are in charge of sentence planning to decide ``*How to say it?*", whereas the RNNLM forms a surface realization to generate surface texts. More specifically, we introduce three additional semantic cells based on the gating mechanism, into a traditional RNN cell. While a refinement cell is to filter the sequential inputs before RNN computations, an adjustment cell and an output cell are to select semantic elements and to gate a feature vector DA during generation, respectively. The proposed models further obtain state-of-the-art results over previous models regarding BLEU and slot error rate ERR scores.

Secondly, we propose a novel hybrid NLG framework to address the first two NLG problems, which is an extension of an RNN Encoder-Decoder incorporating with an attention mechanism. The idea of attention mechanism is to automatically learn alignments between features from source and target sentence during decoding. Our hybrid framework consists of three components: an encoder, an aligner, and a decoder, from which we propose two novel generators to leverage gating and attention mechanisms. In the first model, an additional cell is introduced into aligner cell by utilizing another attention or gating mechanisms to align and control the semantic elements produced by the encoder with a conventional attention mechanism over the input elements. In the first model, we introduce an additional cell into aligner cell by utilizing another attention or gating mechanisms to align and control the semantic elements produced by the encoder with a conventional attention mechanism over the input elements. In the second model, we develop a refinement adjustment LSTM (RALSTM) decoder to select, aggregate semantic elements and to form the required utterances. The hybrid generators not only tackle the NLG problems of completeness, achieving state-of-the-art performances over previous methods, but also deal with adaptability issue by showing an ability to adapt faster to a new, unseen domain and to control feature vector DA effectively.

Thirdly, we propose a novel approach dealing with the problem of low-resource setting data in a domain adaptation scenario. The proposed models demonstrate an ability to perform acceptably well in a new, unseen domain by using only *10%* amount of the target domain data. More precisely, we first present a variational generator by integrating a variational autoencoder into the hybrid generator. We then propose two critics, namely domain and text similarity, in an adversarial training algorithm to train the variational generator via multiple adaptation steps. The ablation experiments demonstrated that while the variational generator contributes to learning the underlying semantic of DA-utterance pairs effectively, the critics play a crucial role in guiding the model to adapt to a new domain in the adversarial training procedure.

Fourthly, we propose another approach dealing with the problem of having low-resource setting of in-domain training data. The proposed generators, which combines two variational autoencoders, can learn more efficiently when the training data is in short supply. In particularly, we present a combination of a variational generator with a variational CNN-DCNN, resulting in a generator which can perform acceptably well using only *10%* to *30%* amount of in-domain training data. More importantly, the proposed model demonstrates state-of-the-art performance regarding BLEU and ERR scores when training with all of the in-domain data. The ablation experiments further showed that while the variational generator makes a positive contribution to learning the global semantic information of pairs of DA-utterance, the variational CNN-DCNN play a critical role of encoding useful information into the latent variables.

Finally, all the proposed generators in this study can learn from unaligned data by jointly training both sentence planning and surface realization to generate natural language utterances. Experiments further demonstrate that the proposed models achieved significant improvements over previous generators concerning two evaluation metrics across four primary NLG domains and variants in a variety of training scenarios. Moreover, the variational-based generators showed a positive sign in unsupervised and semi-supervised learning, which would be a worthwhile study in the future.

Keywords: natural language generation, spoken dialogue system, domain adaptation, gating mechanism, attention mechanism, encoder-decoder, low-resource setting, RNN, GRU, LSTM, CNN, Deconvolutional CNN, VAE.