CAP 5619 - Deep and Reinforcement Learning

HW3 Hua Huang

1 Problem 1

This architecture explored the relationship between channels, it performed channel-wise feature recalibration. First it did a squeeze, which is an average through its spatial dimension $H \times W$ and output z. Second, it did an excitation, in which two fully connected layers are added to the z and with a sigmoid gate, output s. Finally, output the rescaled U by $s_c \cdot u_c$.
To improve the generalization, one of the practices is use the same set of W_1 and W_2 in excitation for each and every channel. The authors did not try to increase the complexity there, say design different Ws for different channels. Also pooling is used here to improve generalization.

2 Problem 2

This work uses a CNN-BLSTM archetecture. There are two types of CNN model architectures: ResNet and LACE. A bidirectional architecture (BLSTM) for acoustic models.
To tackle accent, we can add resource-rich accent(s) data to the trainning set to improve the trainning representation.

3 Problem3

(1) Meaning are encoded in sentences by the semantic logics, or we can say by grammars.
(2) We can use deep model to find a good representation of sentences. Deep Neural Network is well known for its power of representation. Once we get the representation, we can train it to acquire reasonning skills, with which it can understand sentences.

(3) It's a recursive tree pair architecture. In which there are two phrases. At the beginning phase, a pair of tree-structured networks that share a single set of parameters are trainned to generate a vector representation of the sentences. In the second stage, the resulting vectors are fed into a separate comparison layer that is meant to generate a feature vector capturing the relation between the two phrases. The output of this layer is then given to a softmax classifier, which produces a distribution over the semantic relations.
(4) Since the authors build a recursive tree architectures, it can learn the semantic logics recursively. It is possible to understand words, sentences, paragraphs, and the whole article. I think we can build a learning system based on the architectures in this paper.

4 Problem 4

(1) It encodes the input sentence using Bidirectional RNN. In bidirectional RNN, an annotation for each word is constructed by concatenating the forward hidden state and the backward one. In this way, the annotation contains the summaries of both the preceding words and the following words.
(2) The encoder uses Bidirectional RNN, so annotation contains preceding and following words; In the decoder, context vector c is computed as a weighted sum of annotations h . Weights are from an alignment model which scores how well the inputs around position j and the output at position i match. These weights can be understood as attention mechanism, decoder can then focus on the more related vectors.
(3) The context vector is no-longer a fixed-size vector, so more information can be incorporated. At the same time, a Bidirection RNN is used, so both preceding and following words will be considered in the annotation. What's more, an attention mechanism will help the decoder to focus on the more relevant words.
(4) Attention mechanism can be used for control problem, like in solving Reinforcement Learning problem. To better reveal the cause-relation effects between action and reward, attention mechanism can be adopted to tell what action really matters among all the action sequences. Another application is the speech recogination with audio data.