Analysing Seq-to-seq Models in Goal-oriented Dialogue: Generalising to Disfluencies.
S.G.J. Bouwmeester

Abstract

Data-driven dialogue systems are still far from understanding natural dialogue. Several aspects of natural language make it hard to capture in a system, such as unpredictability, mistakes and the width of the domain.
In this thesis we take a step towards more natural data by examining disfluencies (i.e. mistakes).
We test sequence to sequence models with attention on goal-oriented dialogue.
Sequence to sequence models were chosen to overcome the unknown aspect of the mistakes, since they are known for their ability to generalise to unseen examples.
The models are tested on disfluent dialogue data, the bAbI+ task, in addition to normal goal-oriented dialogue data, the bAbI task.
In contrast to previous findings with memory networks, we find that the sequence to sequence model performs both the bAbI tasks as the bAbI+ task well achieving near perfect scores on both tasks.
A slight decrease in performance is noticed when introducing disfluencies only to test data, only 80% accuracy is measured in this condition.
This is surprising because memory networks are very similar to sequence to sequence models with attention.
The results of the main experiment suggest that sequence to sequence models learn to parse disfluencies.
Attention visualisation results suggest that the bAbI+ model does indeed learn to pay attention to disfluencies in a meaningful way.
Even though attention shows that the model is aware of disfluencies, further analyses using diagnostic classifiers and diverse inputs suggest that the encoder is not learning to parse disfluencies, as we expected, but functions more as a memory.
The decoder in turn appears to access the encoder as this memory using the attention mechanism, which proved crucial to learning the bAbI tasks.