Week 9- Reflection

Question 1- What are the main messages you learned from this chapter?

In this chapter mainly I learnt about Attention models and its application in healthcare applications. I also understood attention model over longitudinal model (RETAIN), attention model over medical oncology (GRAM), Attention model over clinical text (CAML), attention model electrocardiography (MINA).
In this lecture we also understood RETAIN algorithm (Reverse Time Attention Model), its application and its characteristics and easy interpretation of RETAIN model.
Apart from RETAIN, we understood GRAM (Graph based attention model) algorithm, its healthcare applications in understanding medical knowledge ontology.
We also read about Clinical coding problem, understood MIMC dataset and combination of Convolution and Attention (CAML) model.
At last in the chapter, we saw explainable prediction of medical codes from Clinical Text and multilevel neural network (MINA) which includes Frequency attention, rhythm attention, beat attention.
In short, we got a very good understanding of CAML, RETAIN and MINA models.

Question 2. What related resources (book, paper, blog, link) do you recommend your classmates to checkout?

- https://www.deeplearningbook.org/contents/part_research.html
- Neural Networks and Deep Learning: A Textbook – Springr
- http://ufldl.stanford.edu/tutorial
- https://towardsdatascience.com/intuitive-understanding-of-attention-mechanism-in-deep-learning-6c9482aecf4f
- Book - Advanced Deep Learning with Keras
- https://blog.floydhub.com/attention-mechanism/
- https://www.youtube.com/watch?v=IxQtK2SjWWM
- https://www.youtube.com/watch?v=XXtpJxZBa2c
-

Question 3- Which part do you want to improve in this chapter?

This chapter provides great overview on attention model and different algorithms such as RETAIN, CAML and MIMA. However, I want if professor could have discussed in more detail about these different types of algorithms and their differences and which algorithm will be more useful in which condition, it would have been great. I would have also liked to see some python implementation of all these algorithms and its related deep learning application.

Question 4. What is the difference between sequence-to-sequence model with RNN and that with attention?

A sequence-to-sequence (sometimes called seq-to-seq) model is a particular type of recurrent network that takes one sequence as input, and outputs a different sequence as output. This is is a common task; it is e.g. what you do when you want to train a machine translation algorithm, which takes a list of tokens in one language as input, and provides a list of tokens in a different language as output. This use case in fact is what motivated much of the research on this particular architecture.

Seq-to-seq models are composed of two pieces: an encoder module that reads in and builds an intermediate representation of the token stream, and a decoder module that reads in this intermediate representation and turns it into a token stream in a different language.

The invention of LSTMs revolutionized RNNs because they solved fundamental problems that vanilla RNNs had with utilizing information embedded in the sequence that is contained in tokens that are far away from one another. Attention was the next big "thing" after LSTMs, and caused the next largest jump in RNN performance on real-world problems.

The bottleneck that attention addresses is the fact that each time-step in an LSTM only gets to work with the output of the previous time-step. Attention adds an additional intermediary step: each LSTM steps additionally gets input from each previous LSTM time-step as an input.

There are several different ways of implementing this additional data flow. The most conceptually intuitive implementation is a weighted sum of the previous layers' data. This is what is meant by attention: the layer learns what previous states to pay attention to, and which previous states to ignore. With attention, it's even possible for a layer to completely ignore the layer immediately previous to it in favor of layers earlier in the time sequence. Hence the name "attention"