Week 9 Reflections

Course	CS 598 - Deep Learning for Healthcare
Submitted By	Rohit Bhagwat (rohitb2@illinois.edu)

Questions

What are the main messages you learned from this chapter?

This week we focused on learning Attention models. The attention mechanism can help neural network models to have greater accuracy and interpretability.

Beyond DNN, CNN, and RNN, the attention mechanism is another fundamental breakthrough in deep learning. It is a general strategy that can lead to many modeling variations.

Attention based on Encoder-Decoder RNN Models

The attention mechanism was originally proposed for improving sequence-to-sequence models with machine translation application. In particular, a sequence of words in a source language is mapped into a sequence of words in a target language, usually of different lengths.

The design of an attention model involves the following components:

1. **Annotation vectors** refer to the encoding of the input sequence,
2. **Alignment model** measures the relevance between pairs of input and output positions
3. **Attention weight** is how much attention to the input words should be given when producing

   each output word
4. **Context vector** is the dynamic summary of the input depending on the current output position

RETAIN: Interpretable Deep Learning Model

Accuracy and interpretability are two dominant features of successful predictive models. Typically, a choice must be made in favor of complex black box models such as recurrent neural networks (RNN) for accuracy versus less accurate but more interpretable traditional models such as logistic regression. This tradeoff poses challenges in medicine where both accuracy and interpretability are important. This challenge was addressed by developing the REverse Time AttentIoN model (RETAIN) for application to Electronic Health Records (EHR) data. RETAIN achieves high accuracy while remaining clinically interpretable and is based on a **two-level neural attention model** that detects influential past visits and significant clinical variables within those visits (e.g. key diagnoses).

GRAM: Graph-based Attention Model for Healthcare Representation Learning

GRAM leverages the parent-child relationship of a graph to learn representation from data (e.g. EHRE data). It generates data representation vector by combining the representation vectors of its ancestors using the attention mechanism.

CAML: Convolutional Attention for Multi-Label classification

Clinical notes are text documents created by clinicians for each patient encounter. They are typically accompanied by diagnosis codes and procedure codes, mainly used for reimbursement from insurance companies. Those diagnosis codes are manually assigned by human coders, which are labor-intensive and error prone. CAML addresses this problem by combining a convolutional architecture and an attention mechanism to model this clinical text.

MINA: Multi-level knowledge-guided attention model

Electrocardiography (ECG) measures electrical activities of a heart and is a commonly used non-invasive diagnostic tool for heart diseases. The existing practice of ECG diagnosis is based on expert-defined patterns from ECG such as P-wave, QRS complex, and RR interval. However, deep learning provides a new set of powerful tools for analyzing ECG in a data-driven way. For each level (*of*

ECG signal pattern, 1) beat level, 2) rhythm level, and 3) frequency level), MINA extracts level-specific domain knowledge features. It uses them to guide the attention models, including beat morphology knowledge that guides attentive CNN and rhythm knowledge that guides attentive RNN. MINA also performs attention fusion across time- and frequency domains.

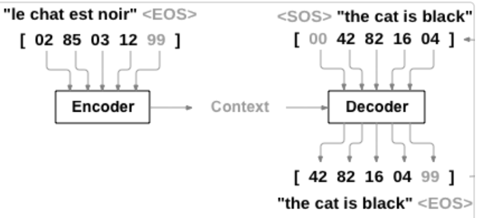
What related resources (book, paper, blog, link) do you recommend your classmates to checkout?

Some great references, that also proved useful when working on the Homeworks -

Translation with a Sequence to Sequence Network and Attention - PyTorch Tutorials 0.2.0_4 documentation

Author: Sean Robertson In this project we will be teaching a neural network to translate from French to English. [KEY: > input, = target, il est en train de peindre un tableau . = he is painting a picture . pourquoi ne pas essayer ce vin delicieux ?

http://seba1511.net/tutorials/intermediate/seq2seq_translation_tutorial.html#sphx-glr-intermediate-seq2seq-translation-tutorial-py



An effective spatial-temporal attention based neural network for traffic flow prediction

Due to its importance in Intelligent Transport Systems (ITS), traffic flow prediction has been the focus of many studies in the last few decades. Existing traffic flow prediction models mainly extract static spatial-temporal correlations, although these correlations are known to be dynamic in traffic networks.

https://www.sciencedirect.com/science/article/abs/pii/S0968090X19301330

https://arxiv.org/pdf/1608.05745v4.pdf

https://arxiv.org/pdf/1802.05695v2.pdf

Which part do you want to improve in this chapter?

In the lectures as well as in the chapter, we are introduced to the concepts very well. The author and the professor walked through the idea of Attention models..

I did struggle with completing assignments, perhaps a walkthrough of an example of PyTorch would prove to be useful for students. *Although I think the exercises will help students get that understanding, it just needs more research.*

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

- The attention model extends the sequence-to-sequence (seq2seq) models using two RNNs
- In the original seq2seq model, the encoder RNN will map the input sequence into a fixed-length context variable c
- This “static” context variable c will be passed into each step of the decoder RNN to generate the output sequence
- However, such seq2seq models have limited capacity in modeling long sequences as the information of the entire input has to be compressed into a single static context variable c
- Attention mechanism allows a “dynamic” context variable c, which provides alignment and translation between source and target sequences
- By optimizing alignment, the model aims to identify which parts of the source sequence are relevant to generate a given the word in the target sequence
- The translation optimization then combines the relevant information from the input sequence into a dynamic context variable for each output word