Recitation 9

CTC Decoding and Beam Search

Soumya Empran & Ameya Mahabaleshwarkar

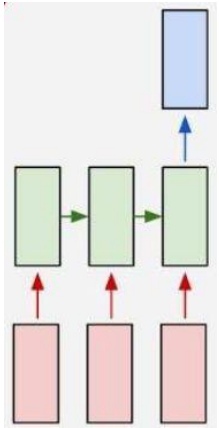
Slides by - Sean Pereira and Tony Qin

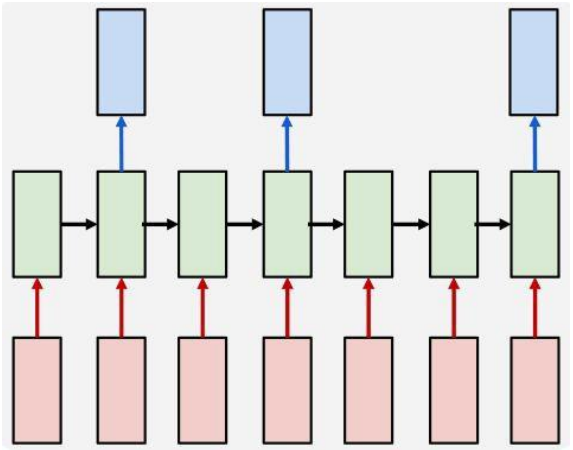
Sequence to Sequence Modeling

- Problem:
  - Input Sequence: $X_1 \dots X_n$
  - Output Sequence: $Y_1 \dots Y_m$
- $|X| \neq |Y|$

HW3P2 Problem: Sequence to Sequence with Order Synchrony

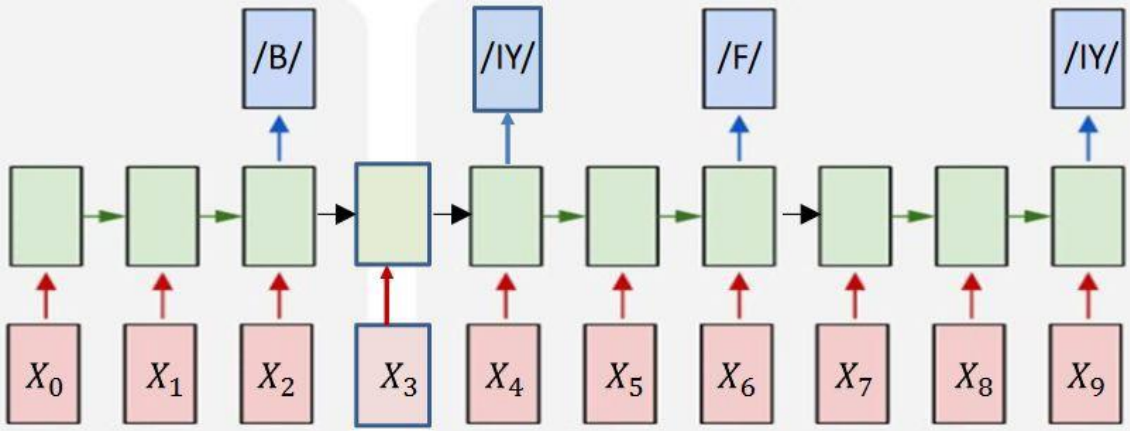
- In HW1P2, we utilized sequence classification for phoneme recognition. We can manage this problem by applying a variant using recurrent nets.
- Left: Sequence of inputs produces a single output; Right: How???



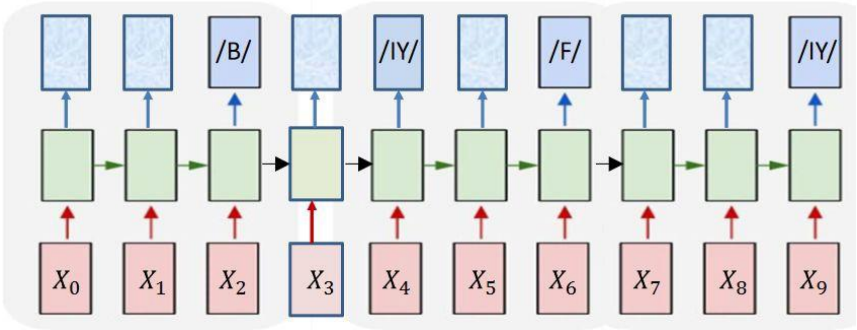


New: Complex Problem - Training

- Objective: Given a sequence of inputs, asynchronously output a sequence of symbols
  - Concatenation of many copies of the simple model in the previous slide



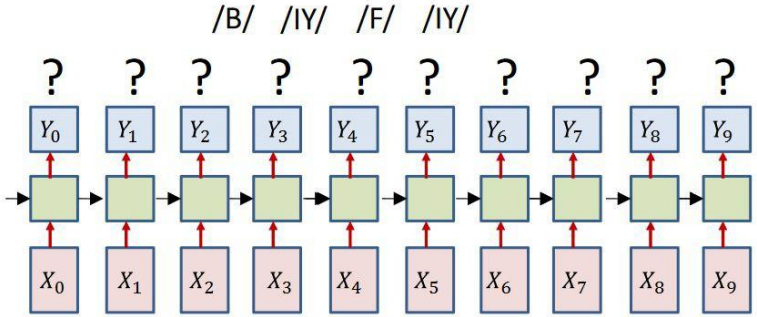
- In the previous model, we ignored intermediate steps. However, we can exploit the untagged inputs and assume the same output.

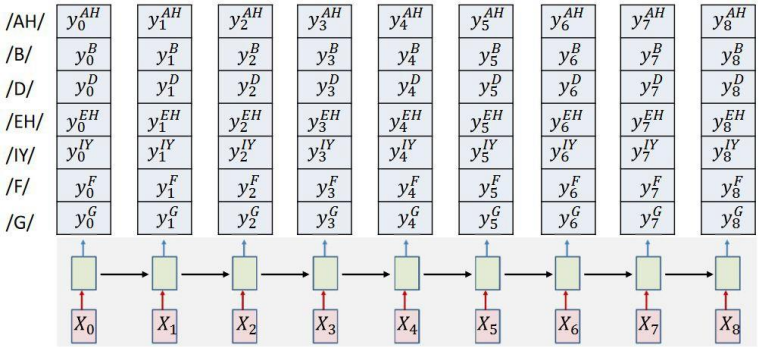


- How do we know when to output symbols?
  - Apply our ideas from HW1P2:
    - At each time in the network outputs a probability for each output symbol given all inputs until that time.
    - The most likely symbol sequence given the inputs. **How?**

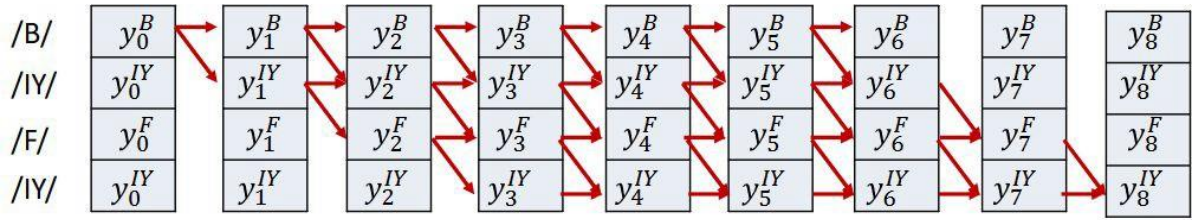
Lecture will discuss computing Divergence

- Possible Solutions
  - **Solution 1**: Simply select the most probable symbol at each time. Merge adjacent repeated symbols, and place the actual emission of the symbol in the final instant.
    - **Issue 1**: This isn't the most probable sequence of symbols
    - **Issue 2**: Cannot distinguish between an extended symbol and repetitions of the symbol
  - **Solution 2**: Impose external constraints on what sequences are allowed
    - **Issue 1:** A suboptimal decode that actually finds the most likely time-synchronous output sequence. Will be discussed in lecture.





- Overall Solution:
  - Apply both previous solutions
    - At each time the network outputs a probability for each output symbol
    - Block out all rows that do not include symbols from the target sequence
    - Compose a graph such that every path in the graph from source to sink represents a valid alignment



- Find the most probable sequence of symbols using the graph above
  - Edge scores have a probability of 1
  - Nodes scores are probabilities resulting from the neural network

Lecture will discuss how to find the most probable sequence given the graph and how to compute the divergence once we get the most probable sequence

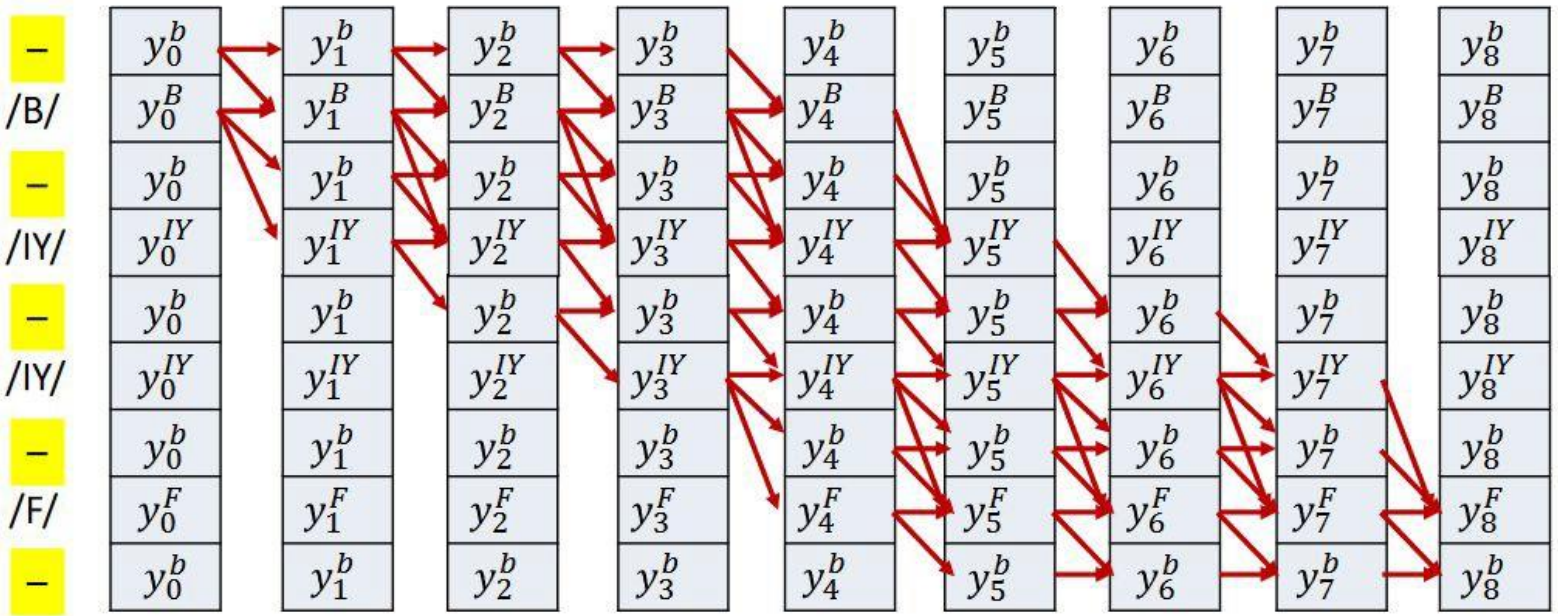
Repetition Issue and Solution

- We have a decode:
  - R R R O O O O O D
  - Is this the symbol sequence ROD or ROOD?

- Introduce an explicit extra symbol which serves to separate discrete versions of a symbol (Blank)
  - RRR---OO---DDD = ROD
  - –RR-R---OO---D-DD = RRODD

- The label recognized by the network must now include the extra blank symbol that will need to be trained

Final Graph



CTC - Training Procedure

1. Setup Network
   a. Many LSTM
2. Initialize network with a Blank Symbol
3. Pass training instances through network to obtain probabilities for all labels/symbols
4. Construct graph on previous page
5. Forward and Backward Algorithm - Lecture
6. Compute Divergence - Lecture
7. Update Parameters

Connectionist Temporal Classification

The forward output