Recitation 9
(CTC Decoding and Beam Search)

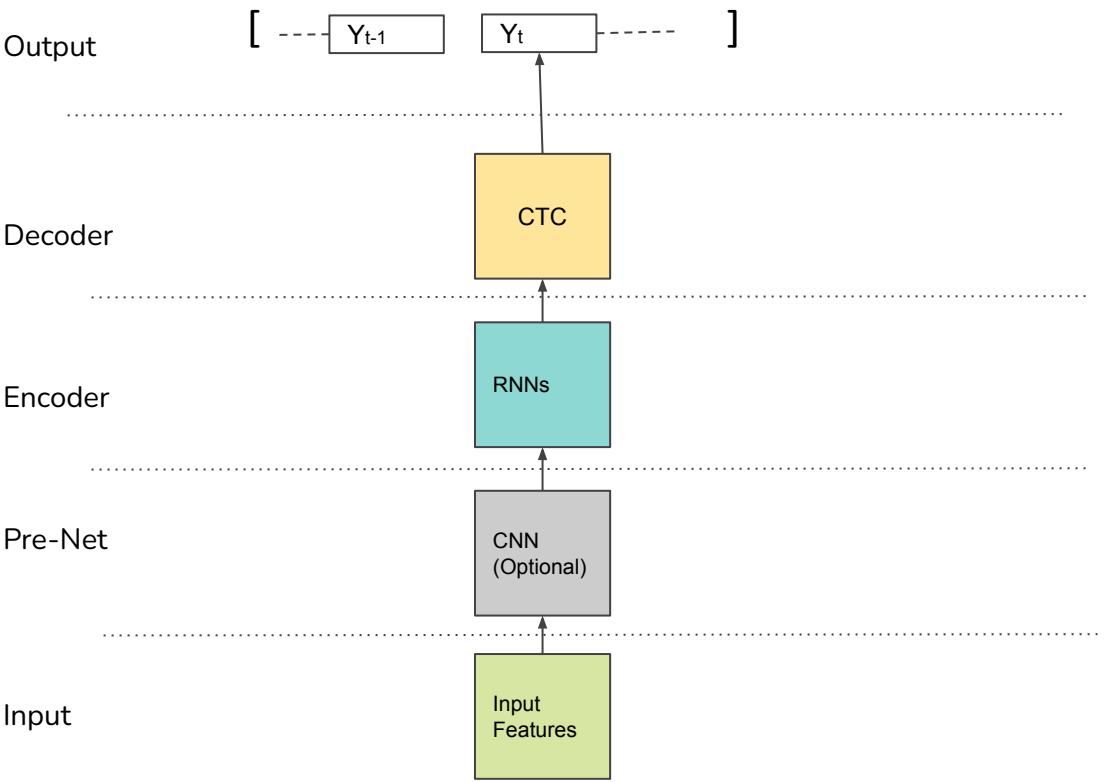
Eshani Agrawal and Shikhar Agnihotri

CTC (Connectionist Temporal Classification)

Problem Use Case : Sequence to Sequence Modelling

- Input Sequence: X1....Xn
- Output Sequence: Y1....Ym
- |X| != |Y|

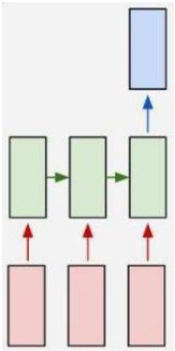
Where does it fit?



HW3P2 Problem: Sequence to Sequence with Order Synchrony

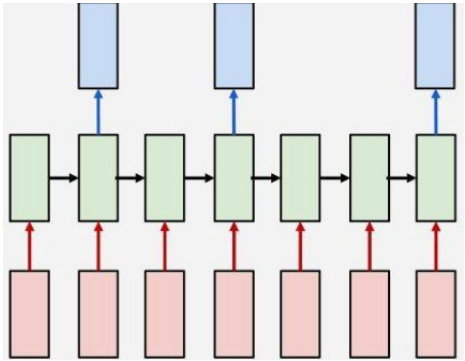
HW1P2: Sequence Classification for phoneme recognition.

How did we solve? Used Linear Layers, now you can revisit and try applying a variant of RNNs.



HW3P2: Sequence to Sequence with Order Synchrony

How to solve this??

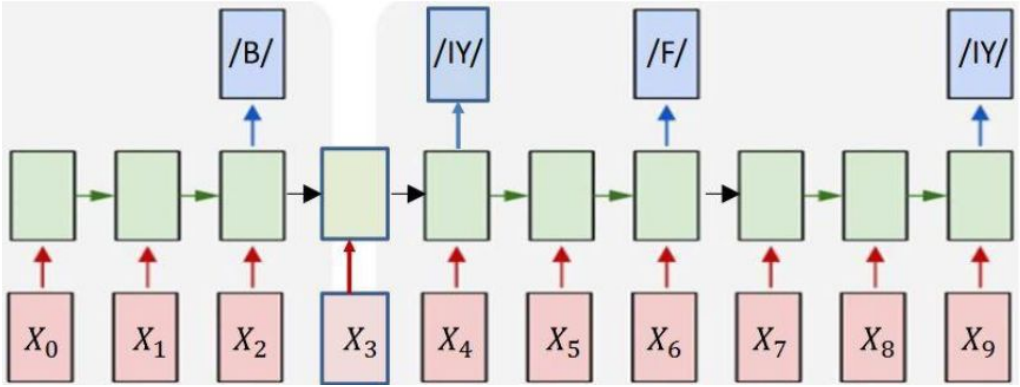


Objective:

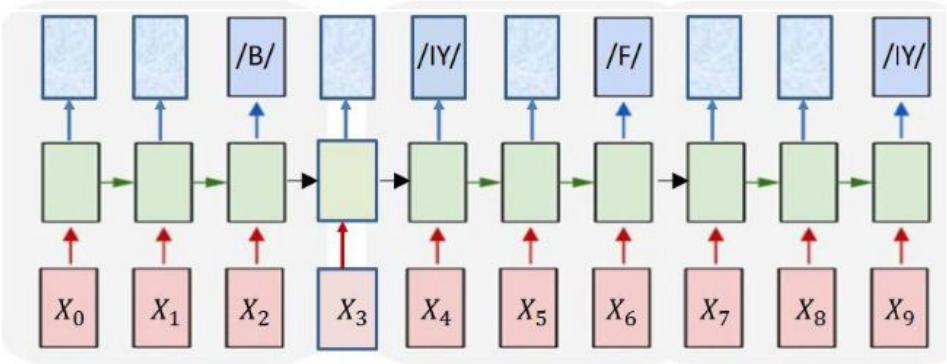
Given a sequence of inputs, asynchronously output a sequence of symbols **# Complex Problem.**

Missing Alignment!

Concatenation of many copies of the simple model



- 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 the network outputs a probability for each output symbol given all inputs until that time.
    - The most likely symbol sequence given the inputs. How?

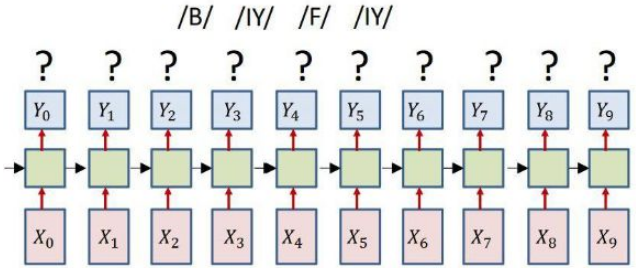
Possible Solutions

- **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

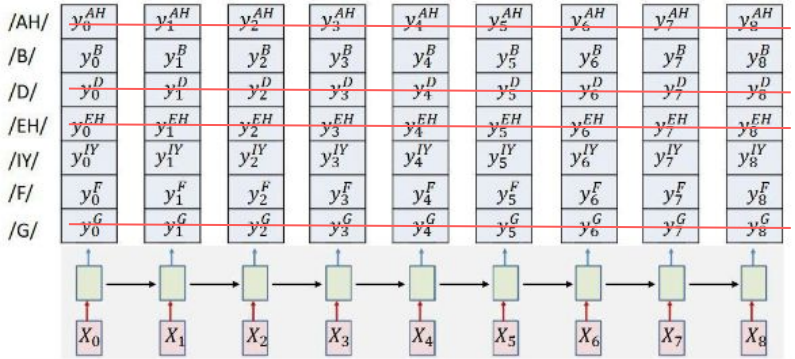
- **Solution 2:** Impose external constraints on what sequences are allowed
  - **Issue 1:** A suboptimal decode that actually finds the most likely time-synchronous output sequence.

Overall Solution

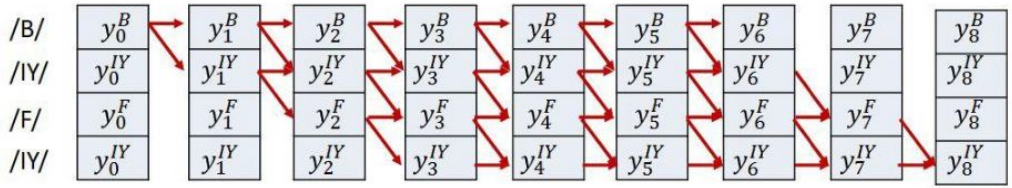
- Apply both previous solutions
  - At each time the network outputs a probability for each output symbol



  - 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 assignment.
- 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

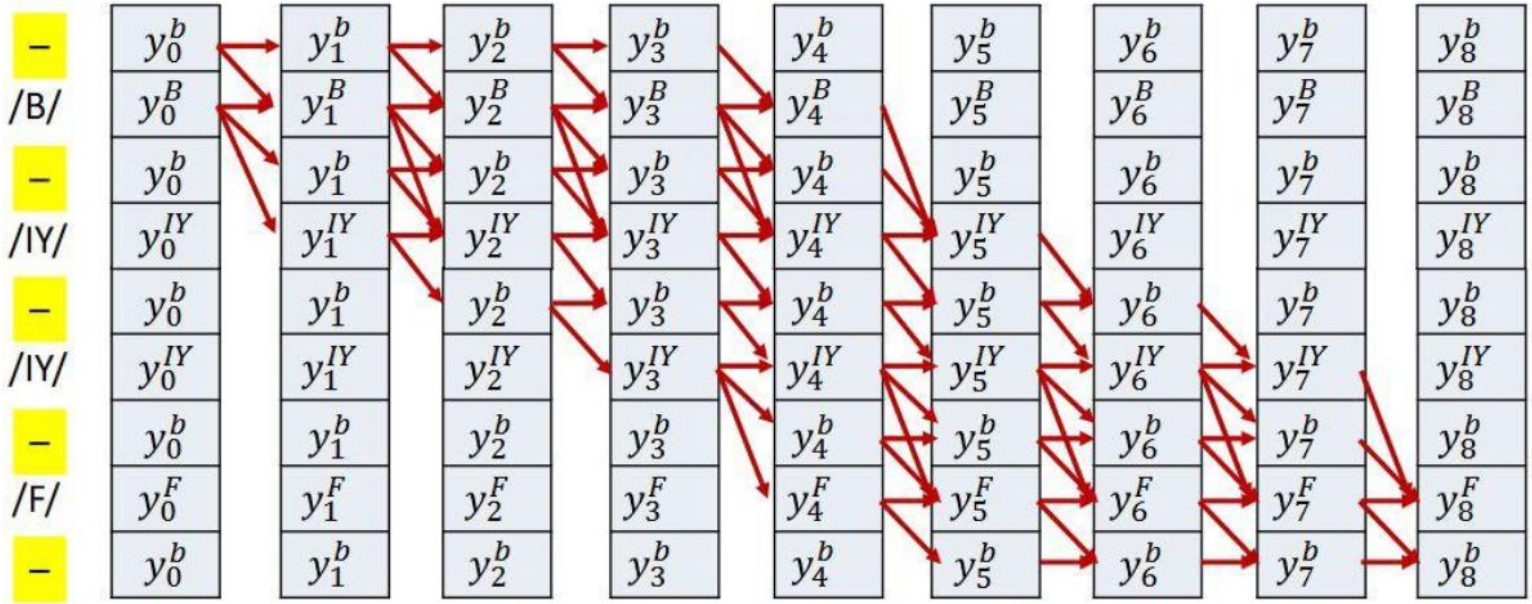
Repetition Issue and Solution

- We have a decode:
  - RRROOOOOD
  - 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:



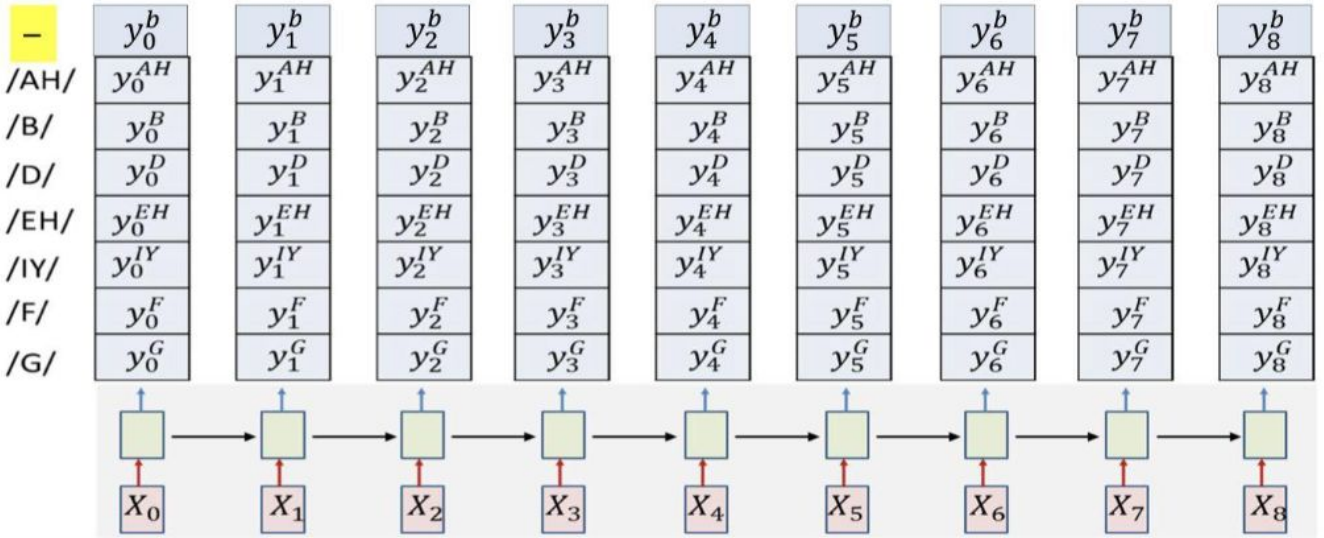
Training Procedure

Connectionist Temporal Classification(CTC)

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

Inference

The Forward Output



Inference

Greedy Search

- Greedy Search is an easy-to-implement option for CTC decoding at inference time
- Greedy Search simply selects the most probable time step at each time-step
- Although this method is easy to implement and fast, it has the disadvantage of missing out on high-probability (score) overall paths due to it’s greedy search

Greedy Search

Y_probs

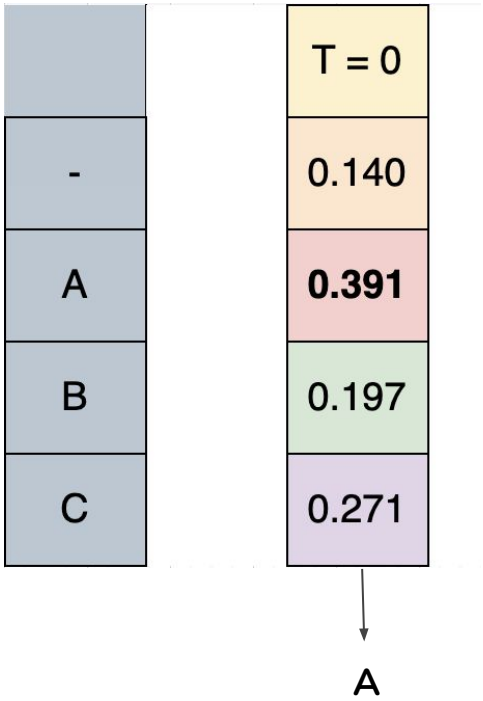
	T = 0	T = 1	T = 2	T = 3
-	0.140	0.257	0.248	0.149
A	0.391	0.096	0.402	0.336
B	0.197	0.341	0.267	0.358
C	0.271	0.305	0.083	0.157

DECODED STRING

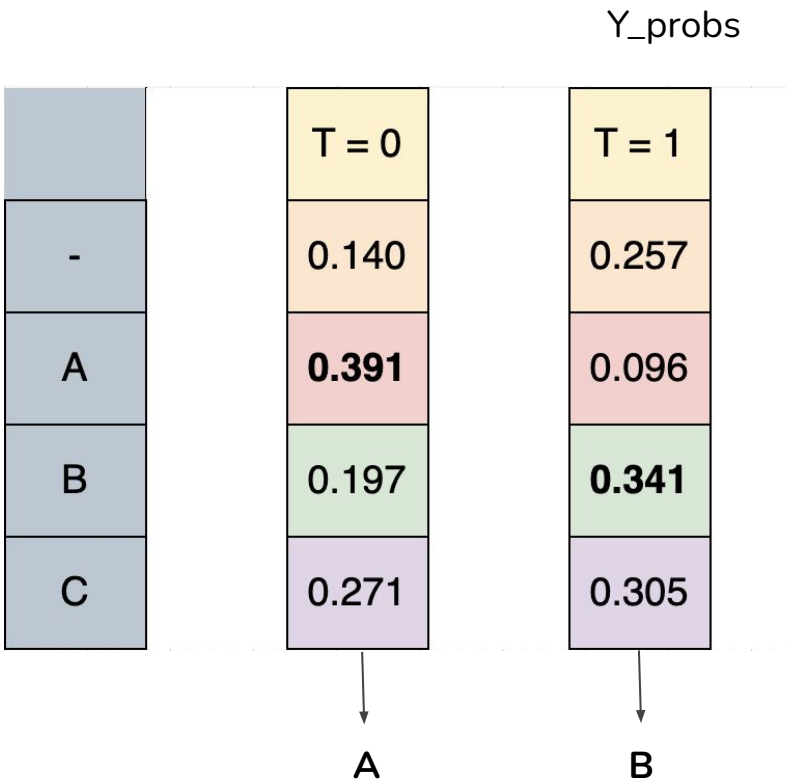
?

Greedy Search

Y_probs

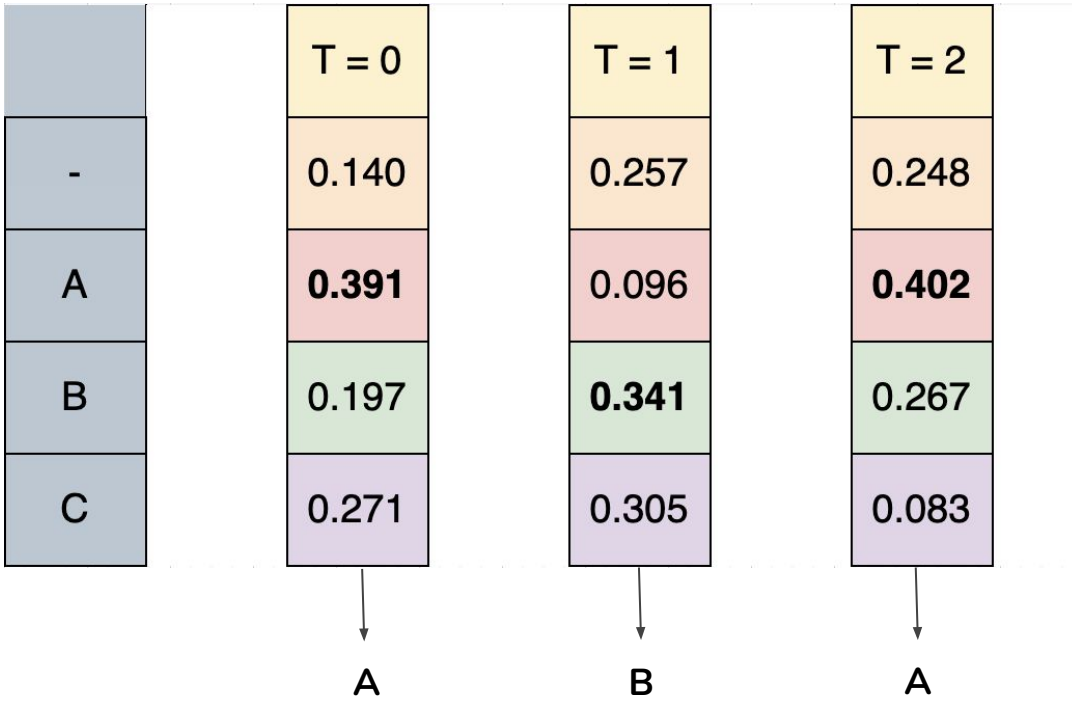


Greedy Search



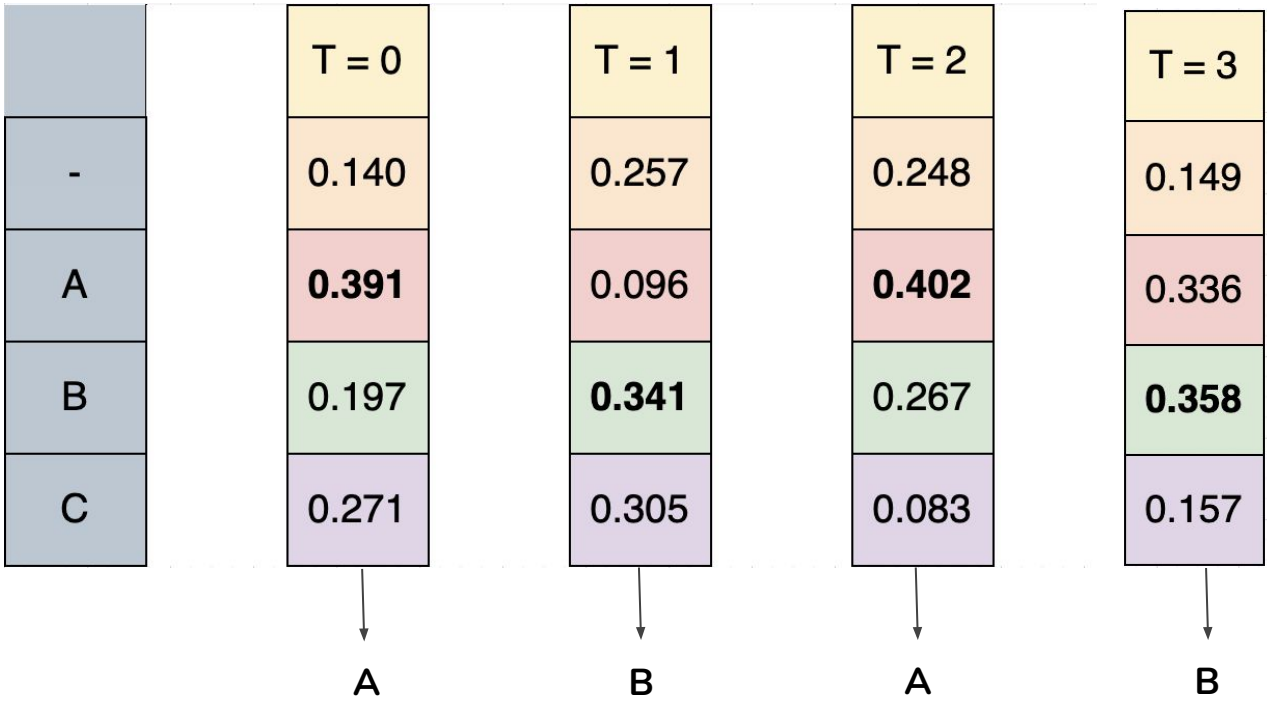
Greedy Search

Y_probs



Greedy Search

Y_probs



Greedy Search

Y_probs

	T = 0	T = 1	T = 2	T = 3
-	0.140	0.257	0.248	0.149
A	0.391	0.096	0.402	0.336
B	0.197	0.341	0.267	0.358
C	0.271	0.305	0.083	0.157

DECODED STRING

A B A B

Inference

Beam Search

- To have better decoding than Greedy Search, but keep the method feasible at the same time, we can choose to “explore” top-k paths at each time-step
- By exploring more than one most-probable output sequences at each time-step, we will reach a sub-optimal path that is likely to be better than the Greedy Search strategy
- By limiting our exploration options to a specific Beam Width - k, we also ensure that the computation is tractable, as opposed to the Exhaustive Search strategy

Beam Search

Y_probs

	T = 0	T = 1	T = 2	T = 3
-	0.140	0.257	0.248	0.149
A	0.391	0.096	0.402	0.336
B	0.197	0.341	0.267	0.358
C	0.271	0.305	0.083	0.157

DECODED STRING

?

Parameters

Seq Len	4
Symbol set	{‘-’, ‘A’, ‘B’, ‘C’}
Beam Width	3

Beam Search

BEAM WIDTH = 3

	T = 0
-	0.140
A	0.391
B	0.197
C	0.271

Scores

Possible Paths	Calculate Score	Score
-	1*0.140	0.140
A	1*0.391	0.391
B	1*0.197	0.197
C	1*0.271	0.271

Old

Best Paths	Score
–	–
–	–
–	–

Beam Search

BEAM WIDTH = 3

	T = 0
-	0.140
A	0.391
B	0.197
C	0.271

Scores

Possible Paths	Calculate Score	Score
-	1*0.140	0.140
A	1*0.391	0.391
B	1*0.197	0.197
C	1*0.271	0.271

Old

Best Paths	Score
–	–
–	–
–	–

New

Best Paths	Score
A	0.391
B	0.197
C	0.271

Beam Search

BEAM WIDTH = 3

	T = 1
-	0.257
A	0.096
B	0.341
C	0.305

Scores

Possible Paths	Calculate Score	Score
A-	0.391*0.257	0.10048700
B-	0.197*0.257	0.05062900
C-	0.271*0.257	0.06964700
AA -> A	0.391*0.096	0.03753600
AB	0.391*0.341	0.13333100
AC	0.391*0.305	0.11925500
BA	0.197*0.096	0.01891200
BB -> B	0.197*0.341	0.06717700
BC	0.197*0.305	0.06008500
CA	0.271*0.096	0.02601600
CB	0.271*0.341	0.09241100
CC -> C	0.271*0.305	0.08265500

Old

Best Paths	Score
A	0.391
B	0.197
C	0.271

Beam Search

BEAM WIDTH = 3

	T = 1
-	0.257
A	0.096
B	0.341
C	0.305

Scores

Possible Paths	Calculate Score	Score
A-	0.391*0.257	0.10048700
B-	0.197*0.257	0.05062900
C-	0.271*0.257	0.06964700
AA -> A	0.391*0.096	0.03753600
AB	0.391*0.341	0.13333100
AC	0.391*0.305	0.11925500
BA	0.197*0.096	0.01891200
BB -> B	0.197*0.341	0.06717700
BC	0.197*0.305	0.06008500
CA	0.271*0.096	0.02601600
CB	0.271*0.341	0.09241100
CC -> C	0.271*0.305	0.08265500

Old

Best Paths	Score
A	0.391
B	0.197
C	0.271

New

Best Paths	Score
A-	0.10048700
AB	0.13333100
AC	0.11925500

Beam Search

BEAM WIDTH = 3

	T = 2
-	0.248
A	0.402
B	0.267
C	0.083

Scores

Possible Paths	Calculate Score	Score
A--	0.100487*0.248	0.0249207760
AB-	0.133331*0.248	0.0330660880
AC-	0.119255*0.248	0.0295752400
A-A -> AA	0.100487*0.402	0.0403957740
ABA	0.133331*0.402	0.0535990620
ACA	0.119255*0.402	0.0479405100
A-B -> AB	0.100487*0.267	0.0268300290
ABB -> AB	0.133331*0.267	0.0355993770
ACB	0.119255*0.267	0.0318410850
A-C -> AC	0.100487*0.083	0.0083404210
ABC	0.133331*0.083	0.0110664730
ACC -> AC	0.119255*0.083	0.0098981650

Old

Best Paths	Score
A-	0.10048700
AB	0.13333100
AC	0.11925500

Beam Search

BEAM WIDTH = 3

	T = 2
-	0.248
A	0.402
B	0.267
C	0.083

Scores

Possible Paths	Calculate Score	Score
A--	0.100487*0.248	0.0249207760
AB-	0.133331*0.248	0.0330660880
AC-	0.119255*0.248	0.0295752400
A-A -> AA	0.100487*0.402	0.0403957740
ABA	0.133331*0.402	0.0535990620
ACA	0.119255*0.402	0.0479405100
A-B -> AB	0.100487*0.267	0.0268300290
ABB -> AB	0.133331*0.267	0.0355993770
ACB	0.119255*0.267	0.0318410850
A-C -> AC	0.100487*0.083	0.0083404210
ABC	0.133331*0.083	0.0110664730
ACC -> AC	0.119255*0.083	0.0098981650

Old

Best Paths	Score
A-	0.10048700
AB	0.13333100
AC	0.11925500

New

Best Paths	Score
ABA	0.0535990620
ACA	0.0479405100
AB	0.0624294060

Beam Search

BEAM WIDTH = 3

	T = 3
-	0.149
A	0.336
B	0.358
C	0.157

Scores

Possible Paths	Calculate Score	Score
ABA- -> ABA	0.0535990620*0.149	0.0079862602380
ACA- -> ACA	0.0479405100*0.149	0.0071431359900
AB- -> AB	0.0624294060*0.149	0.0093019814940
ABAA -> ABA	0.0535990620*0.336	0.0180092848320
ACAA -> ACA	0.0479405100*0.336	0.0161080113600
ABA	0.0624294060*0.336	0.0209762804160
ABAB	0.0535990620*0.358	0.0191884641960
ACAB	0.0479405100*0.358	0.0171627025800
ABB -> AB	0.0624294060*0.358	0.0223497273480
ABAC	0.0535990620*0.157	0.0084150527340
ACAC	0.0479405100*0.157	0.0075266600700
ABC	0.0624294060*0.157	0.0098014167420

Old

Best Paths	Score
ABA	0.0535990620
ACA	0.0479405100
AB	0.0624294060

Beam Search

BEAM WIDTH = 3

	T = 3
-	0.149
A	0.336
B	0.358
C	0.157

Scores

Possible Paths	Calculate Score	Score
ABA- -> ABA	0.0535990620*0.149	0.0079862602380
ACA- -> ACA	0.0479405100*0.149	0.0071431359900
AB- -> AB	0.0624294060*0.149	0.0093019814940
ABAA -> ABA	0.0535990620*0.336	0.0180092848320
ACAA -> ACA	0.0479405100*0.336	0.0161080113600
ABA	0.0624294060*0.336	0.0209762804160
ABAB	0.0535990620*0.358	0.0191884641960
ACAB	0.0479405100*0.358	0.0171627025800
ABB -> AB	0.0624294060*0.358	0.0223497273480
ABAC	0.0535990620*0.157	0.0084150527340
ACAC	0.0479405100*0.157	0.0075266600700
ABC	0.0624294060*0.157	0.0098014167420

Old

Best Paths	Score
ABA	0.0535990620
ACA	0.0479405100
AB	0.0624294060

New

Best Paths	Score
ABA	0.04697182548
AB	0.031651708842
ACA	0.02325114735

Beam Search

BEAM WIDTH = 3