CTC Decoding

Greedy Search, Exhaustive Search, Beam Search

Ameya Mahabaleshwarkar

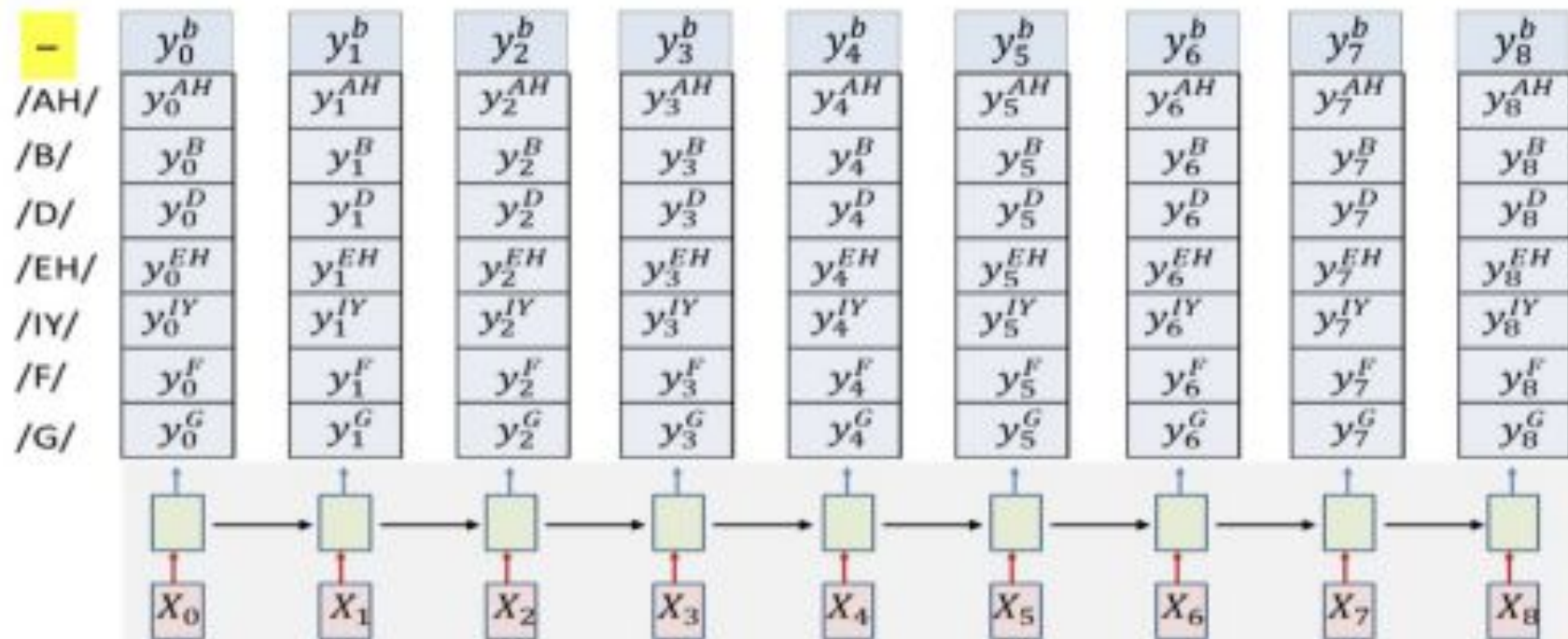IDL – Fall 2022

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

Exhaustive Search

- An alternative to the "short-sighted" Greedy Search, we can conduct a search over all possible paths, and then select the best possible output
- This method will guarantee decoding an optimal path / sequence
- The disadvantage is that Exhaustive Search will be exponential in output symbol sequences, and hence is not a feasible option

Beam Search - The Middle Ground

- 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

	y_0^b	y_1^b	y_2^b	y_3^b	y_4^b	y_5^b	y_6^b	y_7^b	y_8^b
−									
/AH/	y_0^{AH}	y_1^{AH}	y_2^{AH}	y_3^{AH}	y_4^{AH}	y_5^{AH}	y_6^{AH}	y_7^{AH}	y_8^{AH}
/B/	y_0^{B}	y_1^{B}	y_2^{B}	y_3^{B}	y_4^{B}	y_5^{B}	y_6^{B}	y_7^{B}	y_8^{B}
/D/	y_0^{D}	y_1^{D}	y_2^{D}	y_3^{D}	y_4^{D}	y_5^{D}	y_6^{D}	y_7^{D}	y_8^{D}
/EH/	y_0^{EH}	y_1^{EH}	y_2^{EH}	y_3^{EH}	y_4^{EH}	y_5^{EH}	y_6^{EH}	y_7^{EH}	y_8^{EH}
/IY/	y_0^{IY}	y_1^{IY}	y_2^{IY}	y_3^{IY}	y_4^{IY}	y_5^{IY}	y_6^{IY}	y_7^{IY}	y_8^{IY}
/F/	y_0^{F}	y_1^{F}	y_2^{F}	y_3^{F}	y_4^{F}	y_5^{F}	y_6^{F}	y_7^{F}	y_8^{F}
/G/	y_0^{G}	y_1^{G}	y_2^{G}	y_3^{G}	y_4^{G}	y_5^{G}	y_6^{G}	y_7^{G}	y_8^{G}

X_0 X_1 X_2 X_3 X_4 X_5 X_6 X_7 X_8

Let's use some actual values

	T=1	T=2	T=3
B	0.47	0.18	0.58
A	0.03	0.44	0.40
--	0.49	0.38	0.02

X1   X2   X3