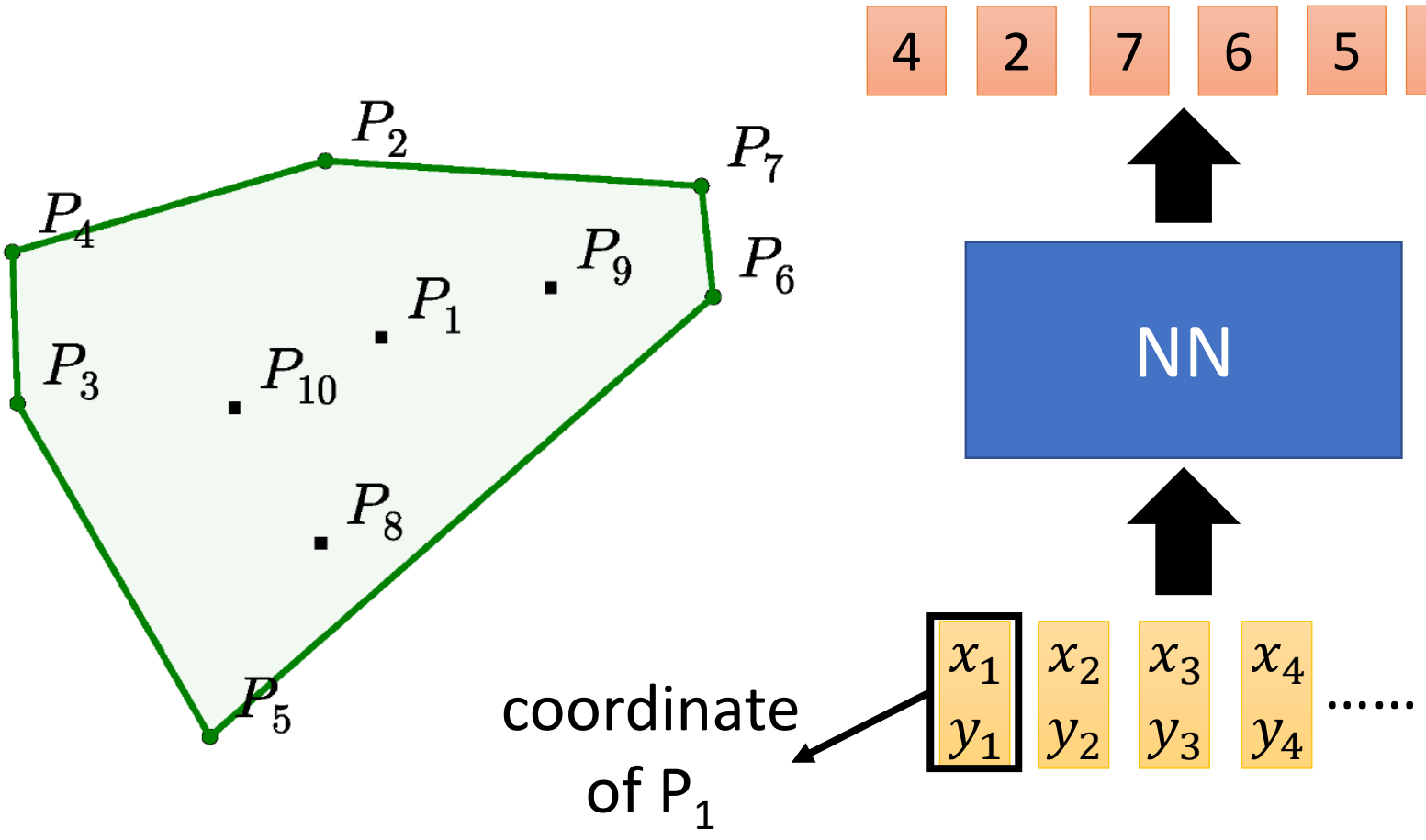
Pointer Network

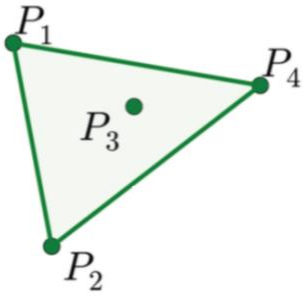
最早是想要解系列演算法的問題

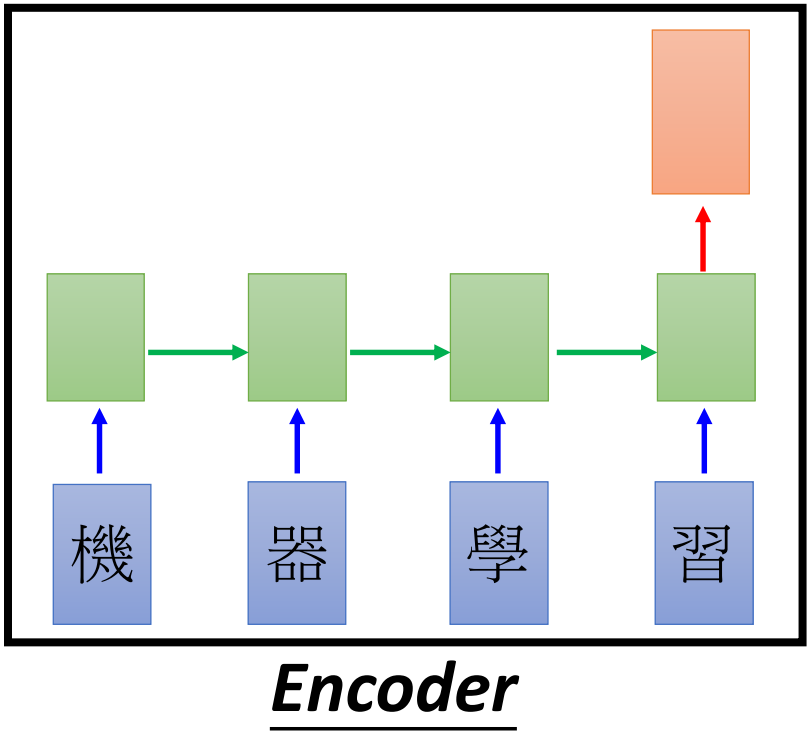
原始paper解了三個，以下介紹convex hole的問題

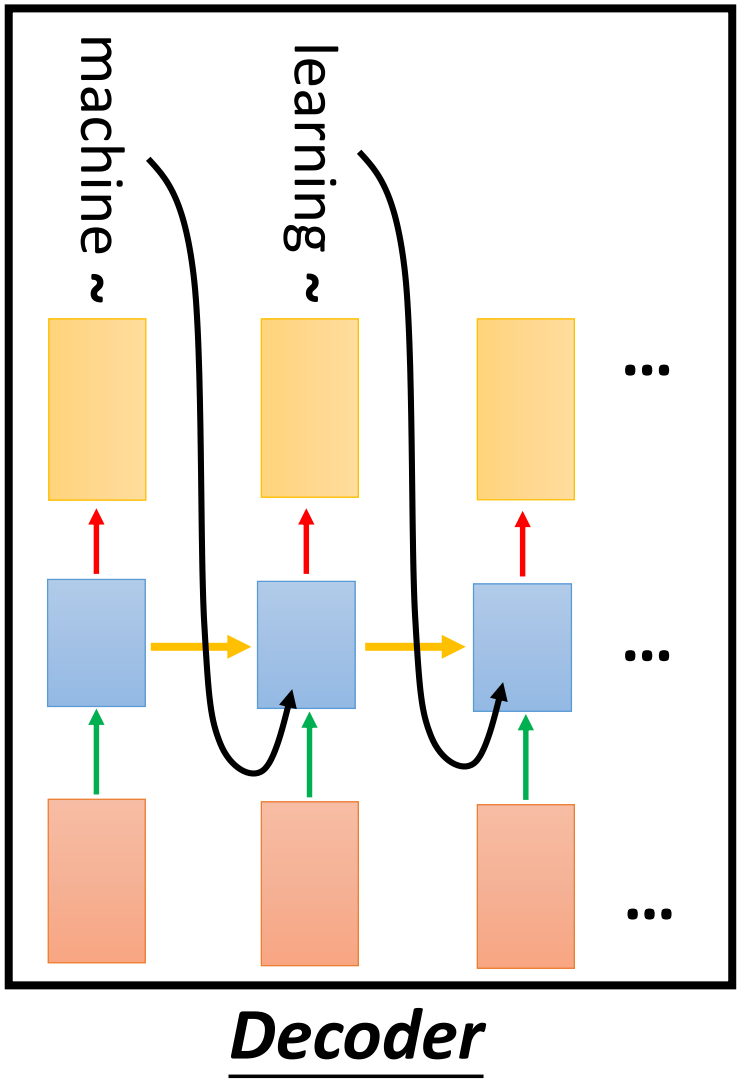
Pointer Network



Sequence-to-sequence?





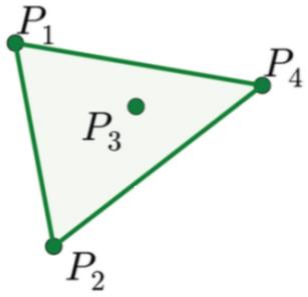


training的時候如果只有p1~p50，但是在testing的時候要是有p1~p100，則他只會output p1~p50

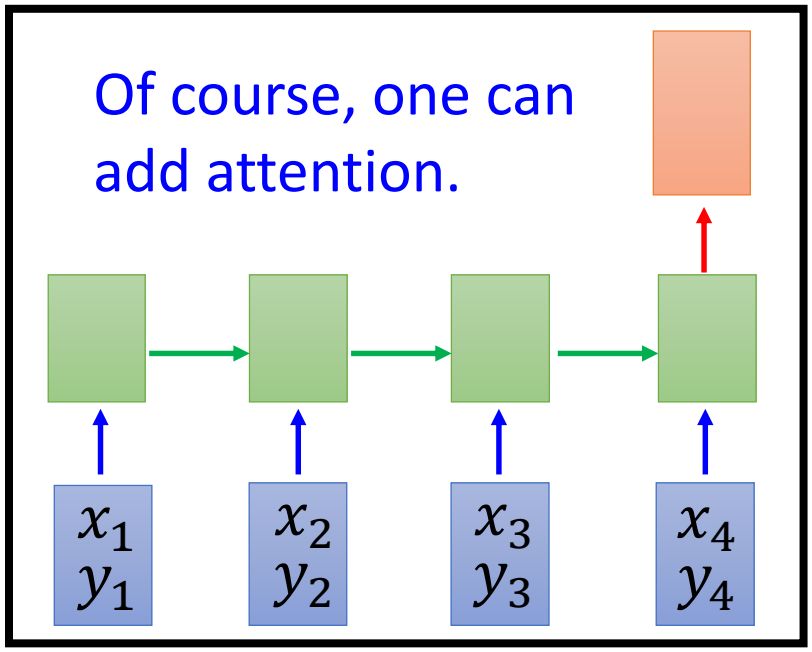
為了解決這種問題，採用attention based model

Problem?

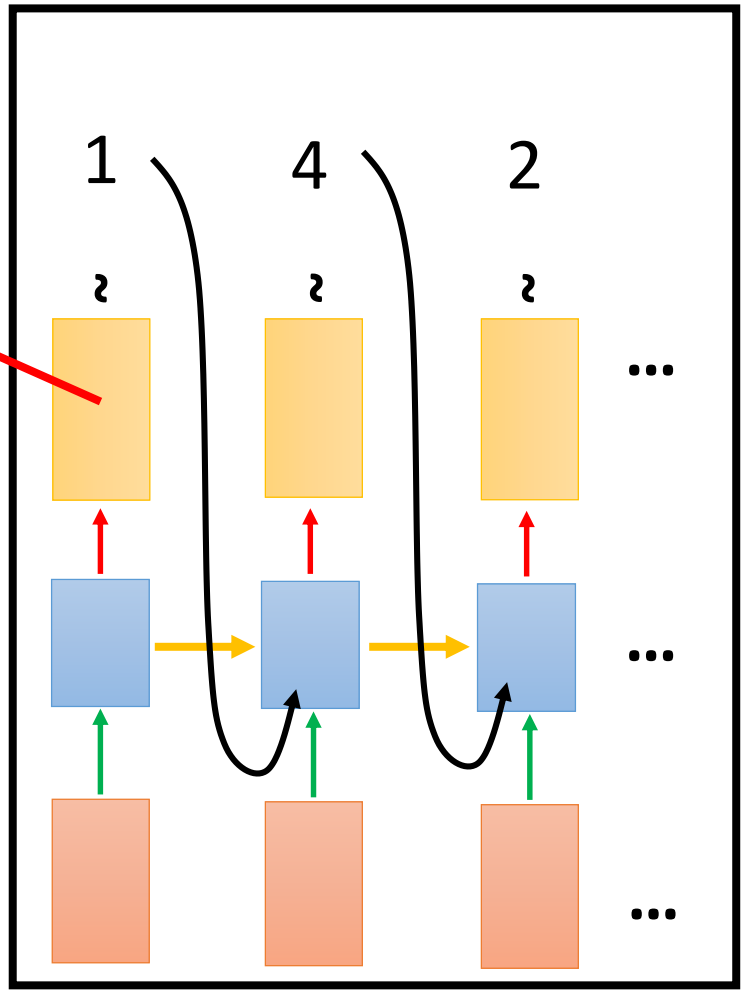
Sequence-to-sequence?



{1, 2, 3, 4, END}



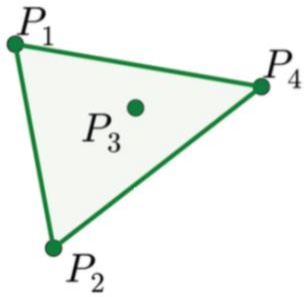
Encoder



Decoder

Pointer Network

x_0
y_0 : END



Attention Weight

key

0.5

z^0

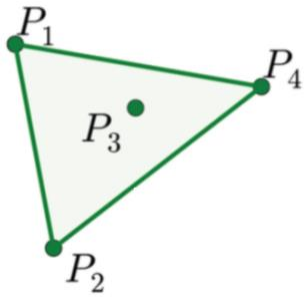
h^0 h^1 h^2 h^3 h^4

x_0 y_0 | x_1 y_1 | x_2 y_2 | x_3 y_3 | x_4 y_4

Pointer Network

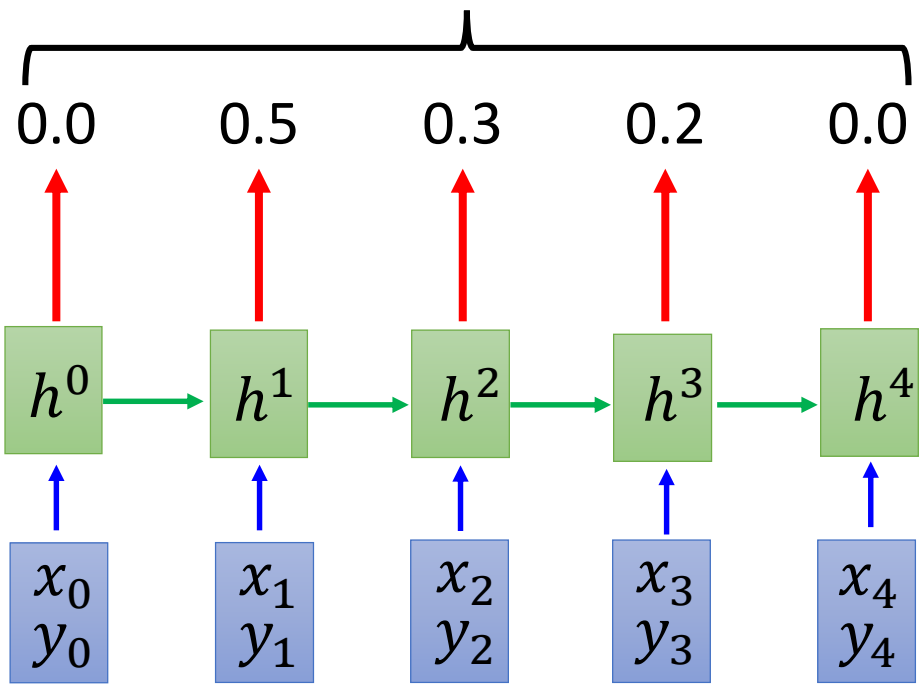
原本的attention model是做weighted sum，現在這邊改成直接argmax

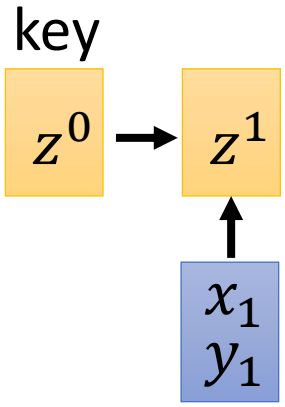
x_0 y_0 : END



Output: 1

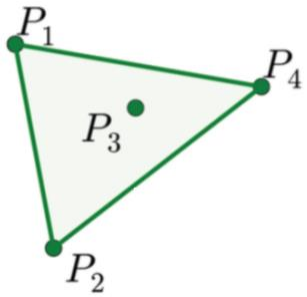
What decoder can output depends on the input.





Pointer Network

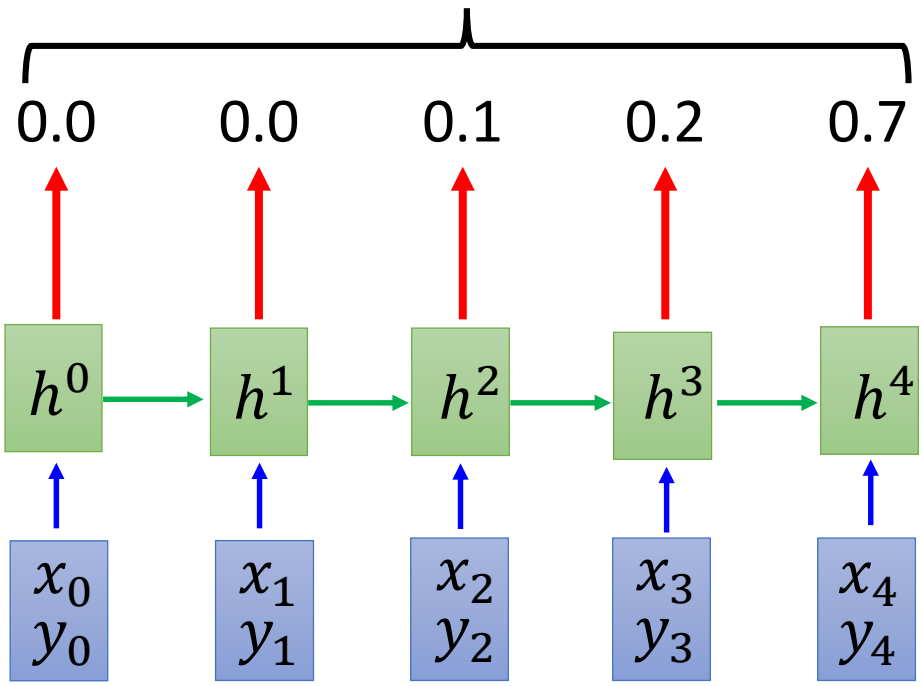
x_0 y_0 : END



Output: 4

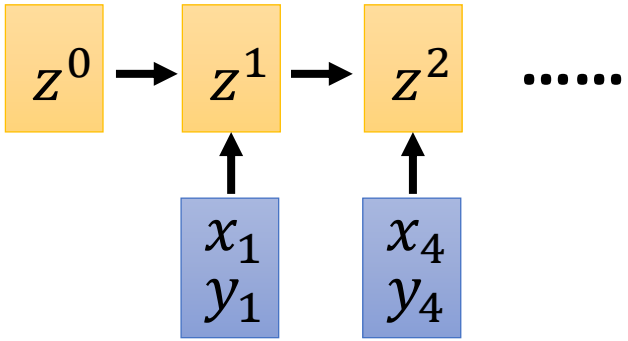
≈

argmax from this distribution



What decoder can output depends on the input.

key



The process stops when "END" has the largest attention weights.

end出現代表x0, y0 attention最大

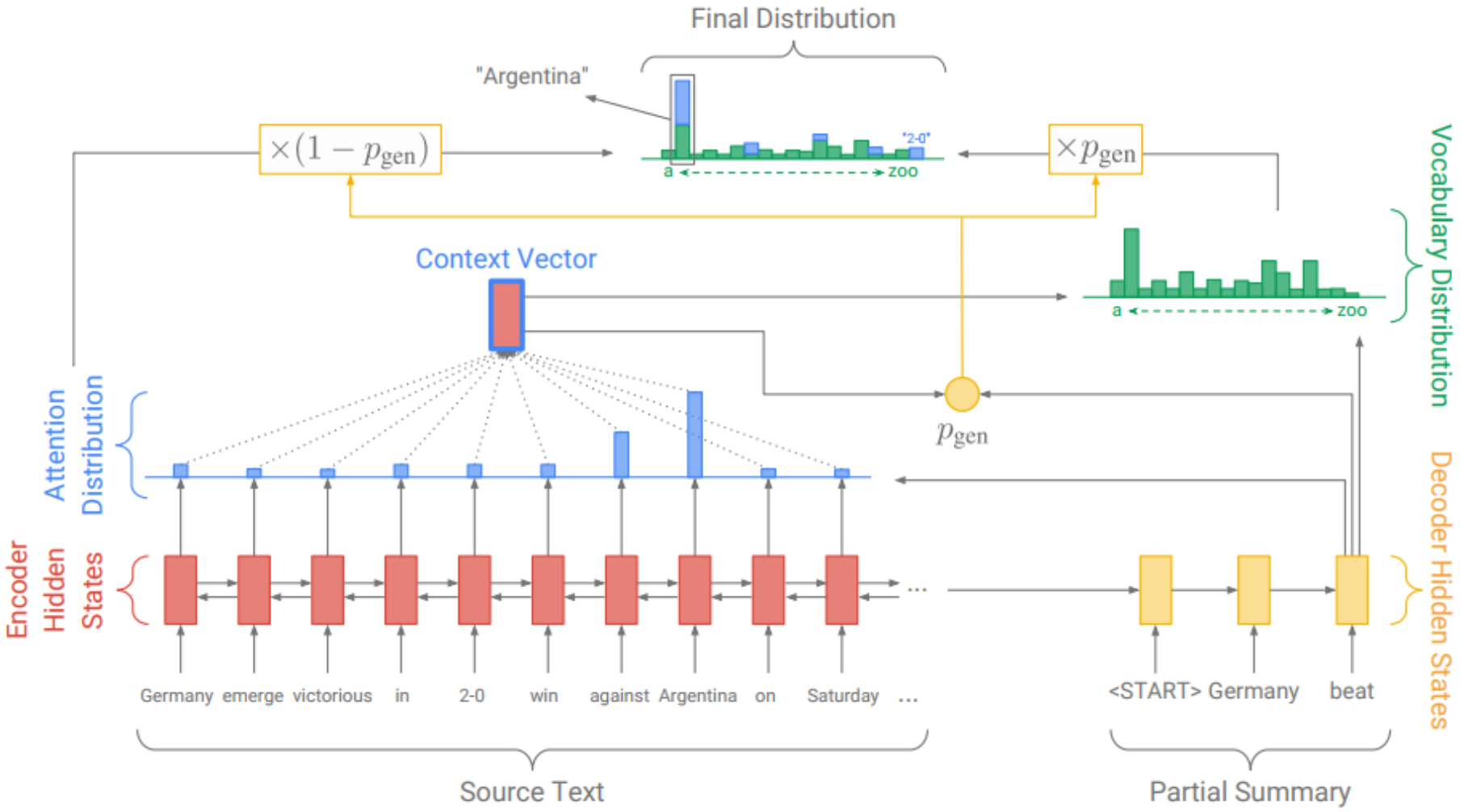
seq2seq – input document output summarization

https://arxiv.org/abs/1704.04368

然而一般的seq2seq在testing時的output只能局限於training data中的word

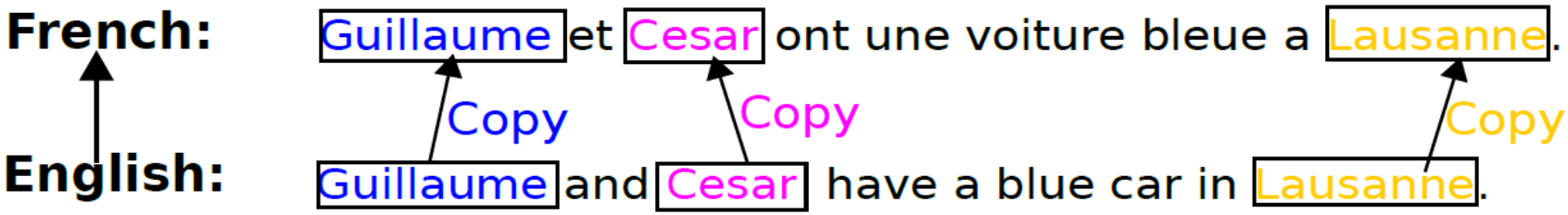
用了pointer network後的output可以based on input word而非單純依靠lexicon

Applications - Summarization



More Applications

Machine Translation



Chat-bot

User: X寶你好，我是庫洛洛

希望他直接copy input word，一般的seq2seq可能無法做到因為lexicon沒有這個字，採用pointer network可以直接copy

Machine:庫洛洛你好，很高興認識你