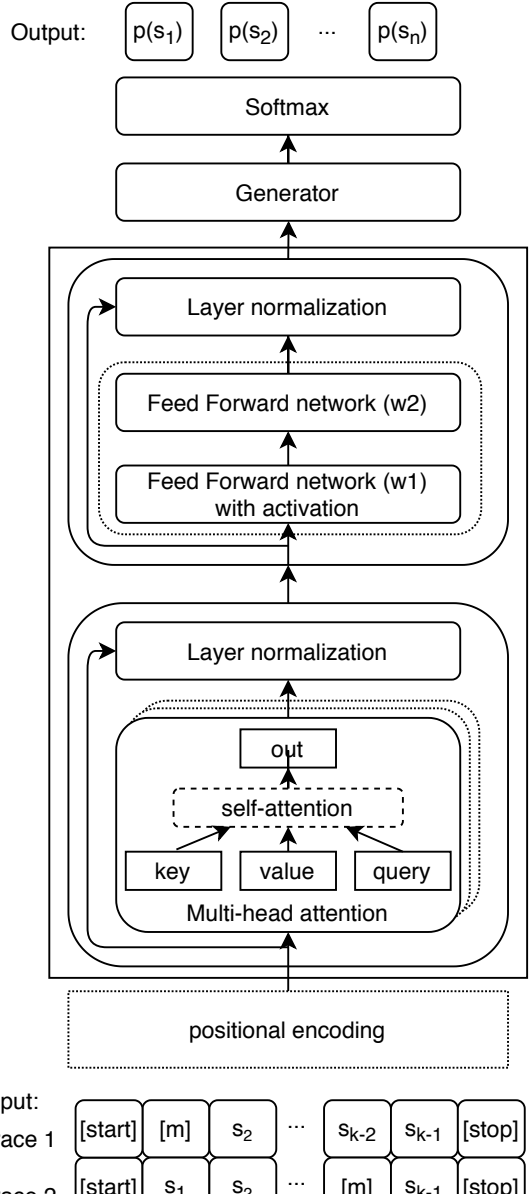
Output:
p(s1)
p(s2)
···
p(sn)
Softmax
Generator
Layer normalization
Feed Forward network (w2)
Feed Forward network (w1)
with activation
Layer normalization
out
self-attention
key
value
query
Multi-head attention
positional encoding
Input:
trace 1
[start]
[m]
s2
···
sk-2
sk-1
[stop]
trace 2
[start]
s1
s2
···
[m]
sk-1
[stop]