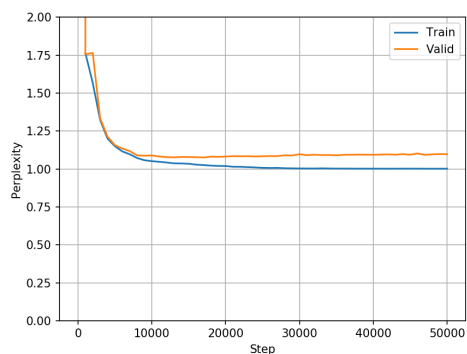
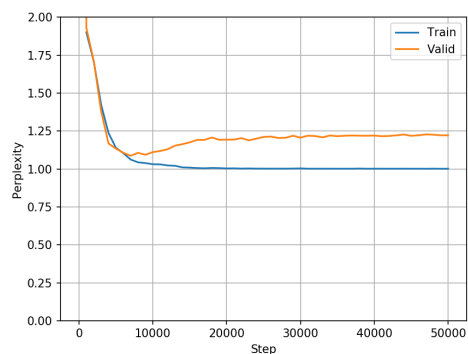


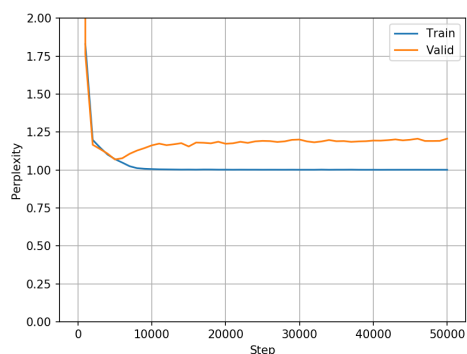
## Plots of perplexity on the training and validation set of Monument for each model



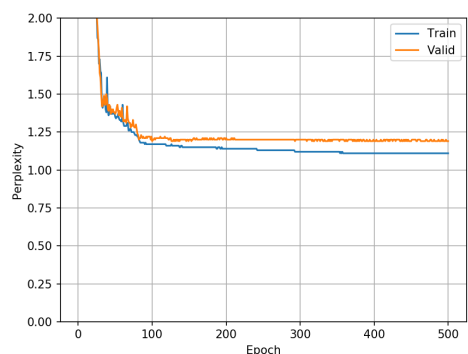
(a) NSpM



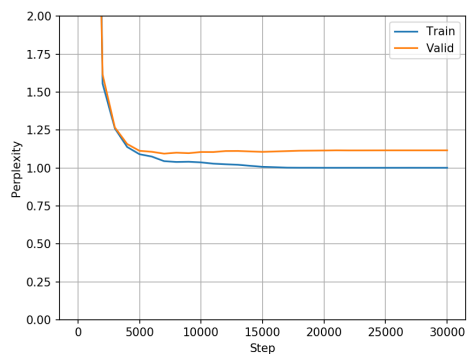
(b) NSpM+Att1



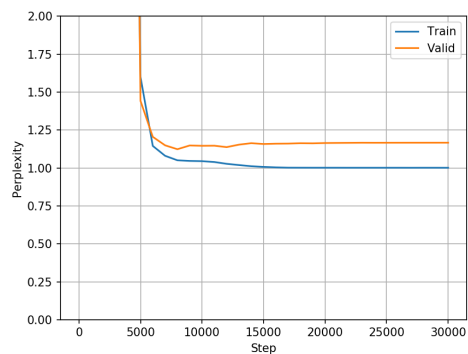
(c) NSpM+Att2



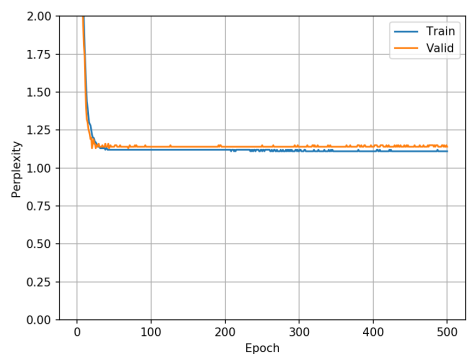
(d) LSTM\_Luong



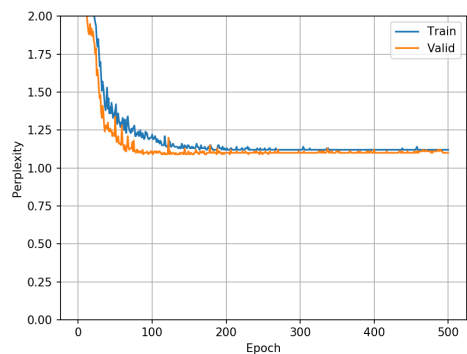
(e) GNMT-4



(f) GNMT-8

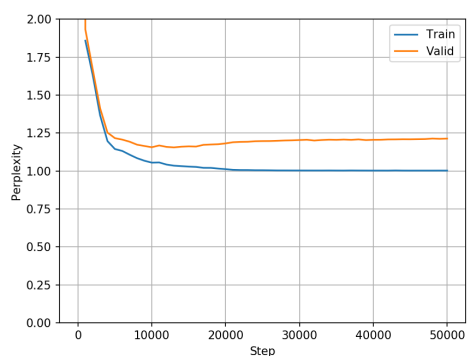


(g) ConvS2S

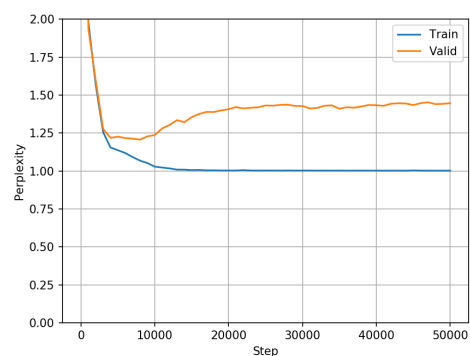


(h) Transformer

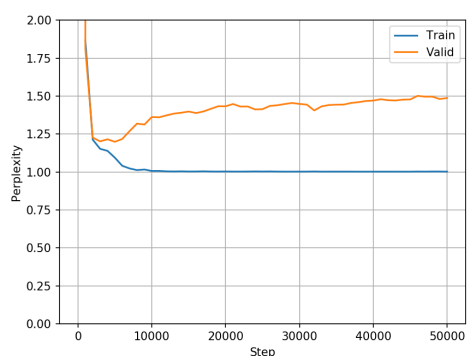
## Plots of perplexity on the training and validation set of Monument80 for each model



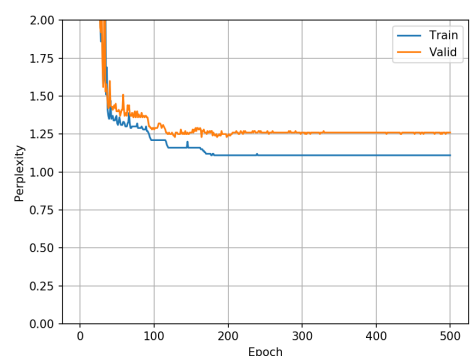
(a) NSpM



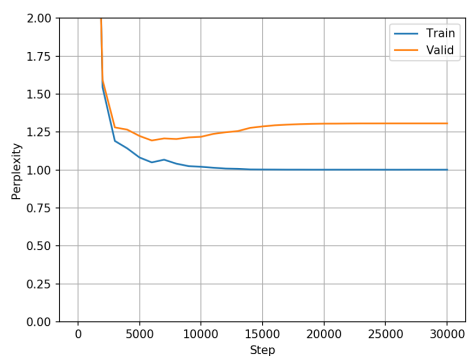
(b) NSpM+Att1



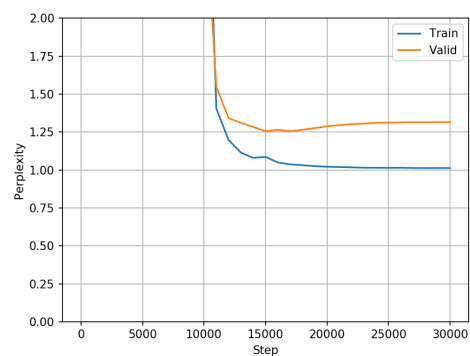
(c) NSpM+Att2



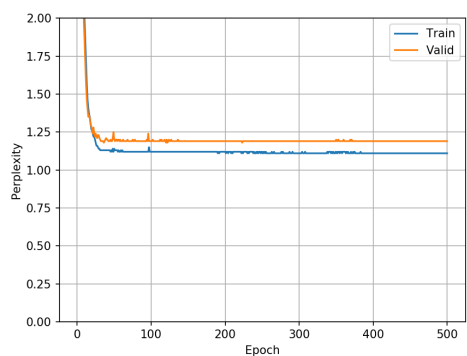
(d) LSTM\_Luong



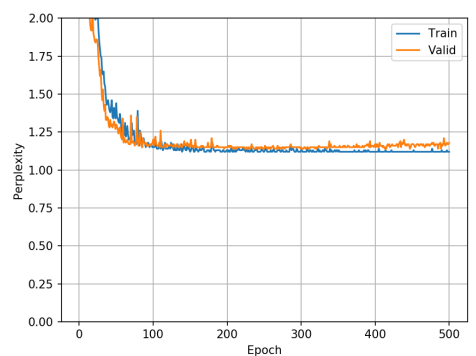
(e) GNMT-4



(f) GNMT-8

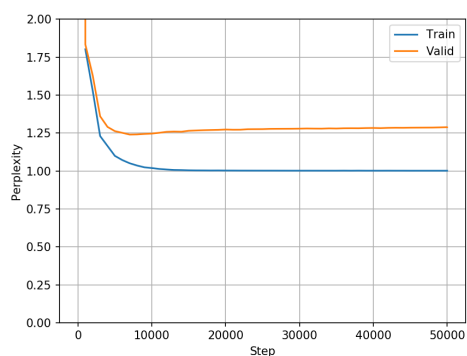


(g) ConvS2S

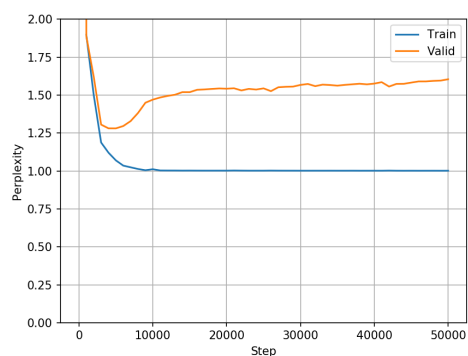


(h) Transformer

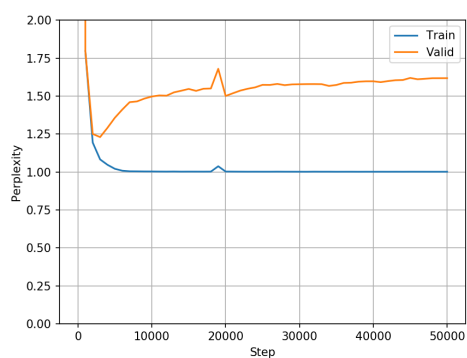
## Plots of perplexity on the training and validation set of Monument50 for each model



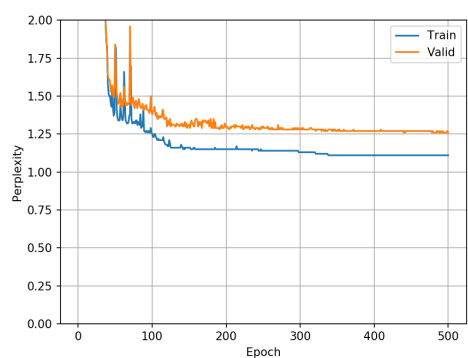
(a) NSpM



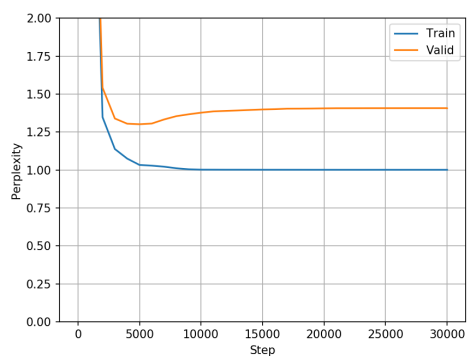
(b) NSpM+Att1



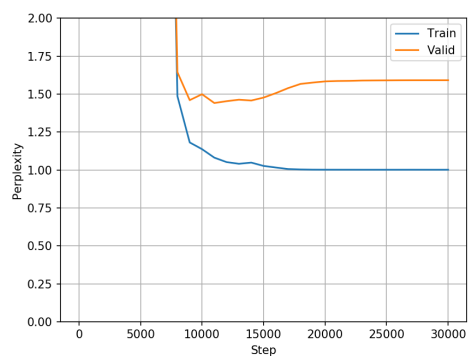
(c) NSpM+Att2



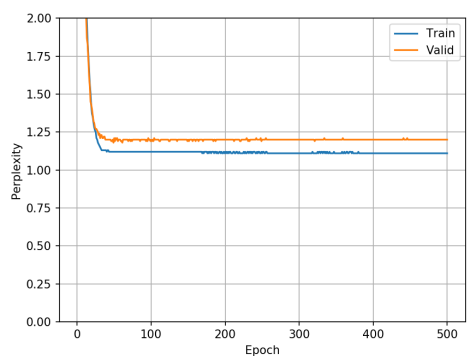
(d) LSTM\_Luong



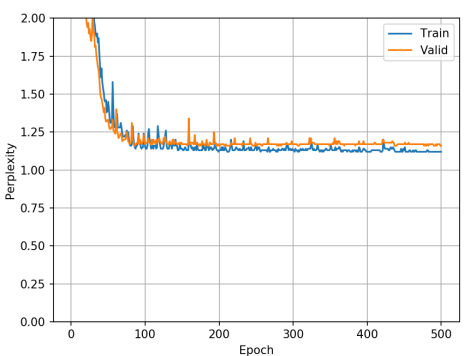
(e) GNMT-4



(f) GNMT-8

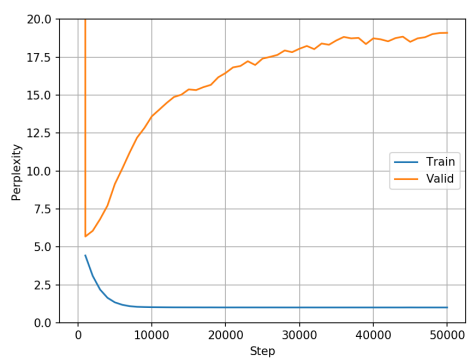


(g) ConvS2S

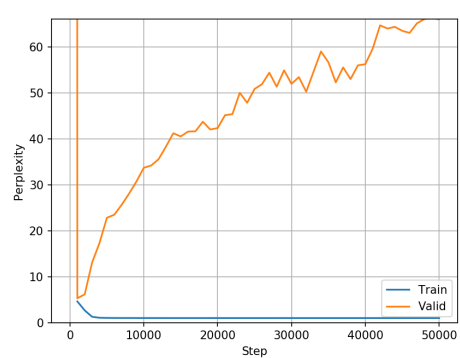


(h) Transformer

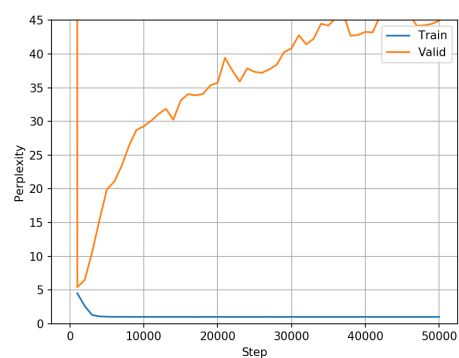
## Plots of perplexity on the training and validation set of LC-QUAD for each model



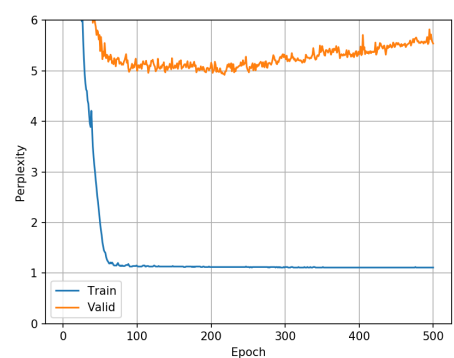
(a) NSpM



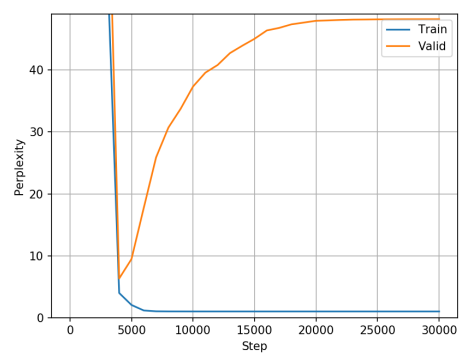
(b) NSpM+Att1



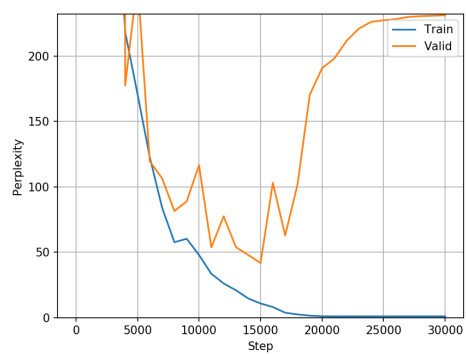
(c) NSpM+Att2



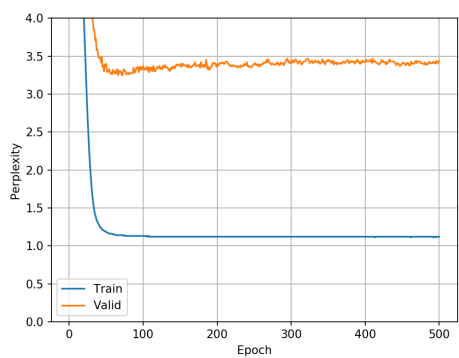
(d) LSTM\_Luong



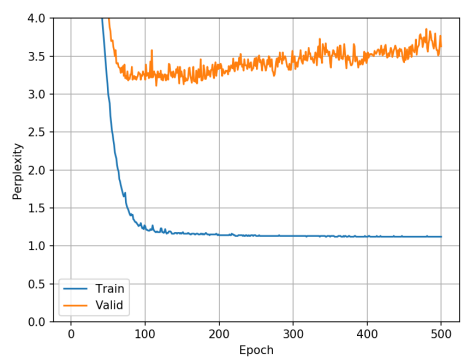
(e) GNMT-4



(f) GNMT-8

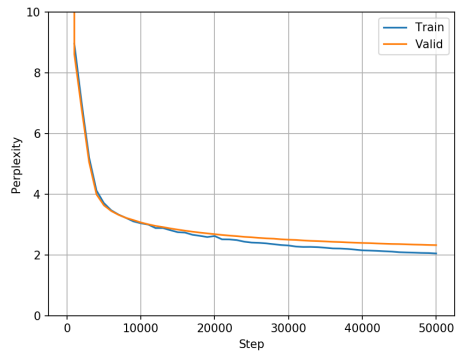


(g) ConvS2S

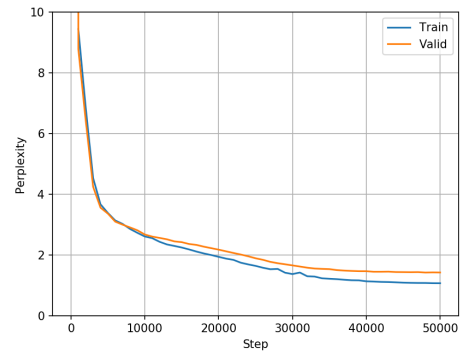


(h) Transformer

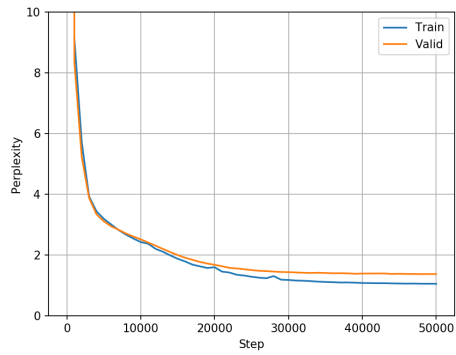
## Plots of perplexity on the training and validation set of DBNQA for each model



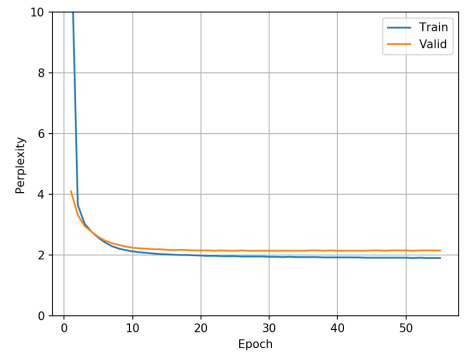
(a) NSpM



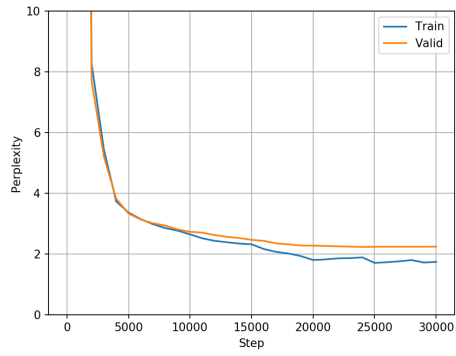
(b) NSpM+Att1



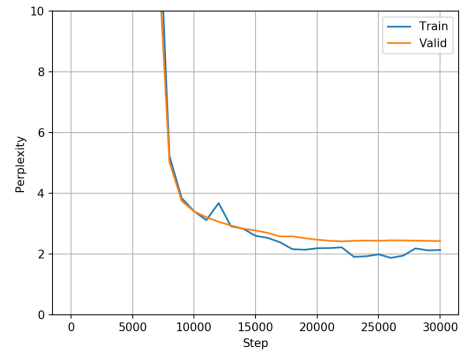
(c) NSpM+Att2



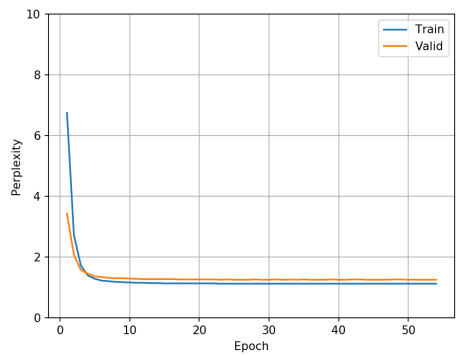
(d) LSTM\_Luong



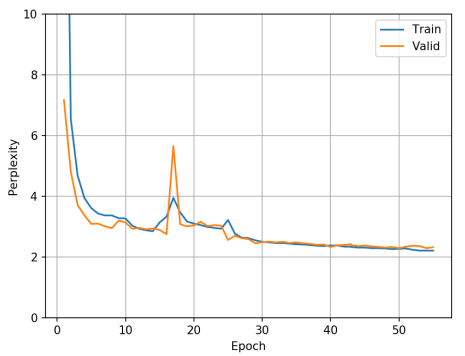
(e) GNMT-4



(f) GNMT-8



(g) ConvS2S



(h) Transformer