Word embedding	512
GRU dimension	1024
Optimizer	adam
Initial learning rate	0.0001
Beam size	5