avg recall accuracy, weight constraint35
pointer_network vs transformer

- transformer_same, best val-mean: 0.78
- transformer_lr1-e4_decay05, best val-mean: 0.73
- transformer_lr1-e6, best val-mean: 0.8
- transformer_lr1-e8, best val-mean: 0.77
- pointer_network, best val-mean: 0.83