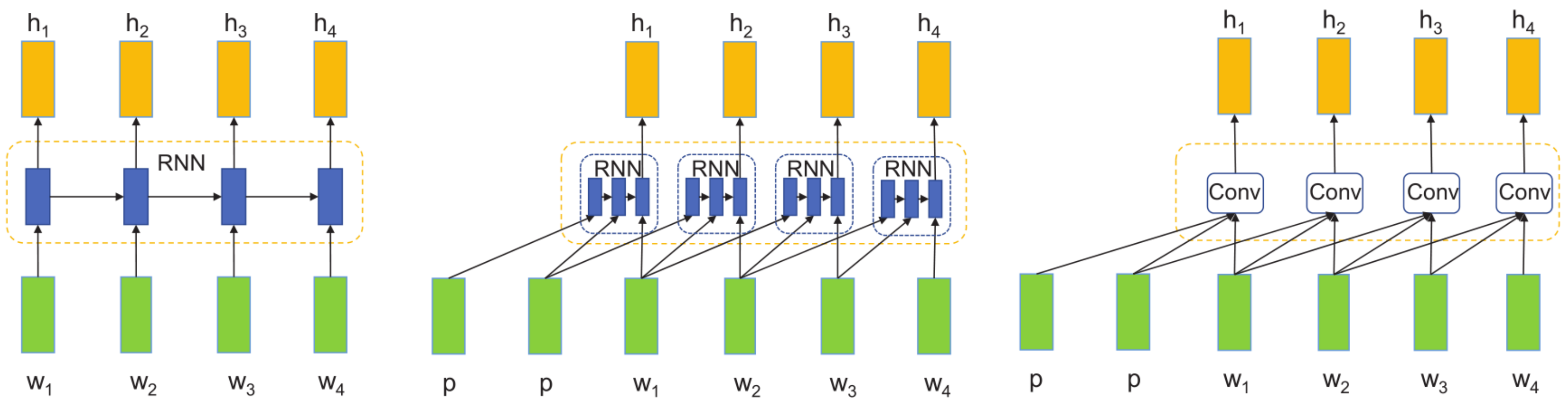
Disconnected Recurrent Neural Networks for Text Categorization

Baoxin Wang

Joint Lab of HIT and iFLYTEK, iFLYTEK Research, Beijing, China

INTRODUCTION

RNN can model the entire sequence and capture long-term dependencies, but it does not do well in extracting key patterns. In contrast, convolutional neural network (CNN) is good at Extracting local and position-invariant features. We present a novel model named disconnected recurrent neural network (DRNN), which incorporates position-invariance into RNN by constraining the distance of information flow in RNN. DRNN achieves the best performance on several benchmark datasets for text categorization.



(a) RNN (b) DRNN (c) CNN

MOTIVATION

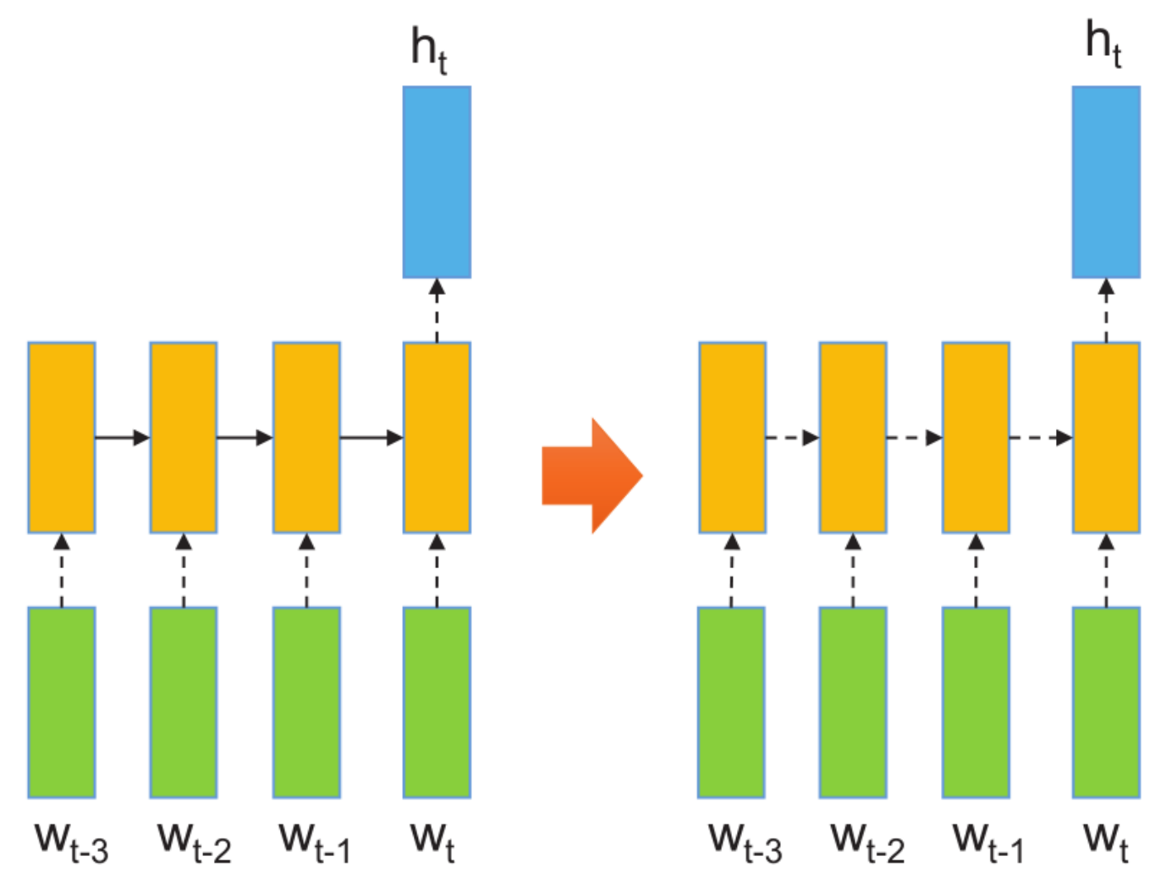
> **Case1:** *One of the seven great* ***unsolved mysteries of mathematics*** *may have been cracked by a reclusive Russian.*
>
> **Case2***: A reclusive Russian may have cracked one of the seven great* ***unsolved mysteries of mathematics****.*

The key phrase that determines the category is ***unsolved mysteries of mathematics***, which can be well extracted by CNN due to position-invariance. RNN can model the whole sequence and capture long-term dependencies, yet it doesn't address the above issue well because the representation of the key phrase relies on all the previous terms and it changes as the key phrase moves.

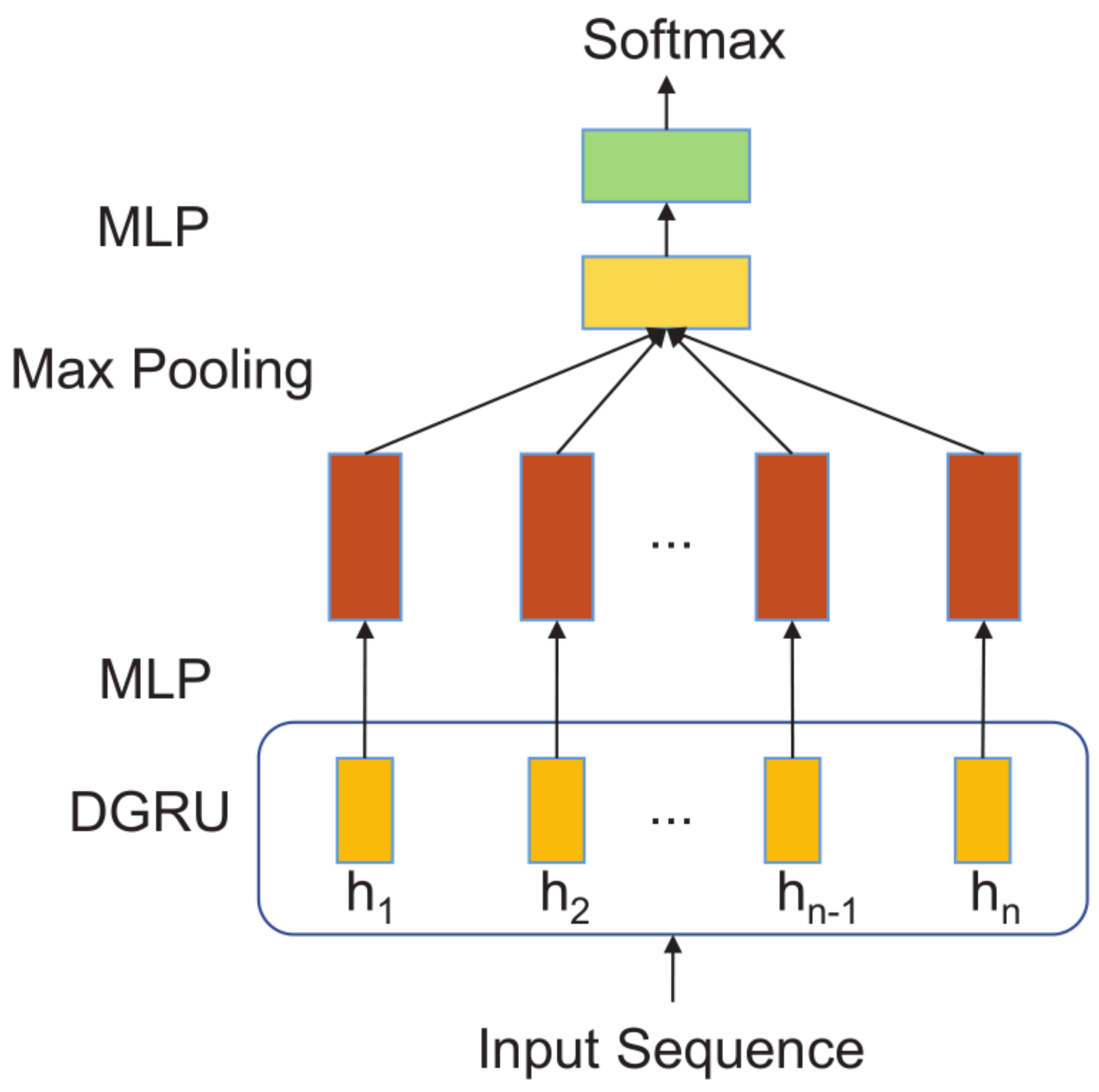
DRNN is proposed to deal with such issues while eliminating the burden of modeling the entire sentence.

THE MODEL

DRNN limits the distance of information flow in RNN. An important difference from RNN is that the state of our model at each step is only related to the previous k-1 words rather than all the previous words. DRNN can be considered as a special 1D CNN which replaces the convolution filters with recurrent units.



Figure： Dropout applied on hidden states as well.



Softmax

MLP

Max Pooling

MLP

DGRU

h_1 h_2 … h_{n-1} h_n

Input Sequence

Figure：Model Architecture

EXPERIMENTS

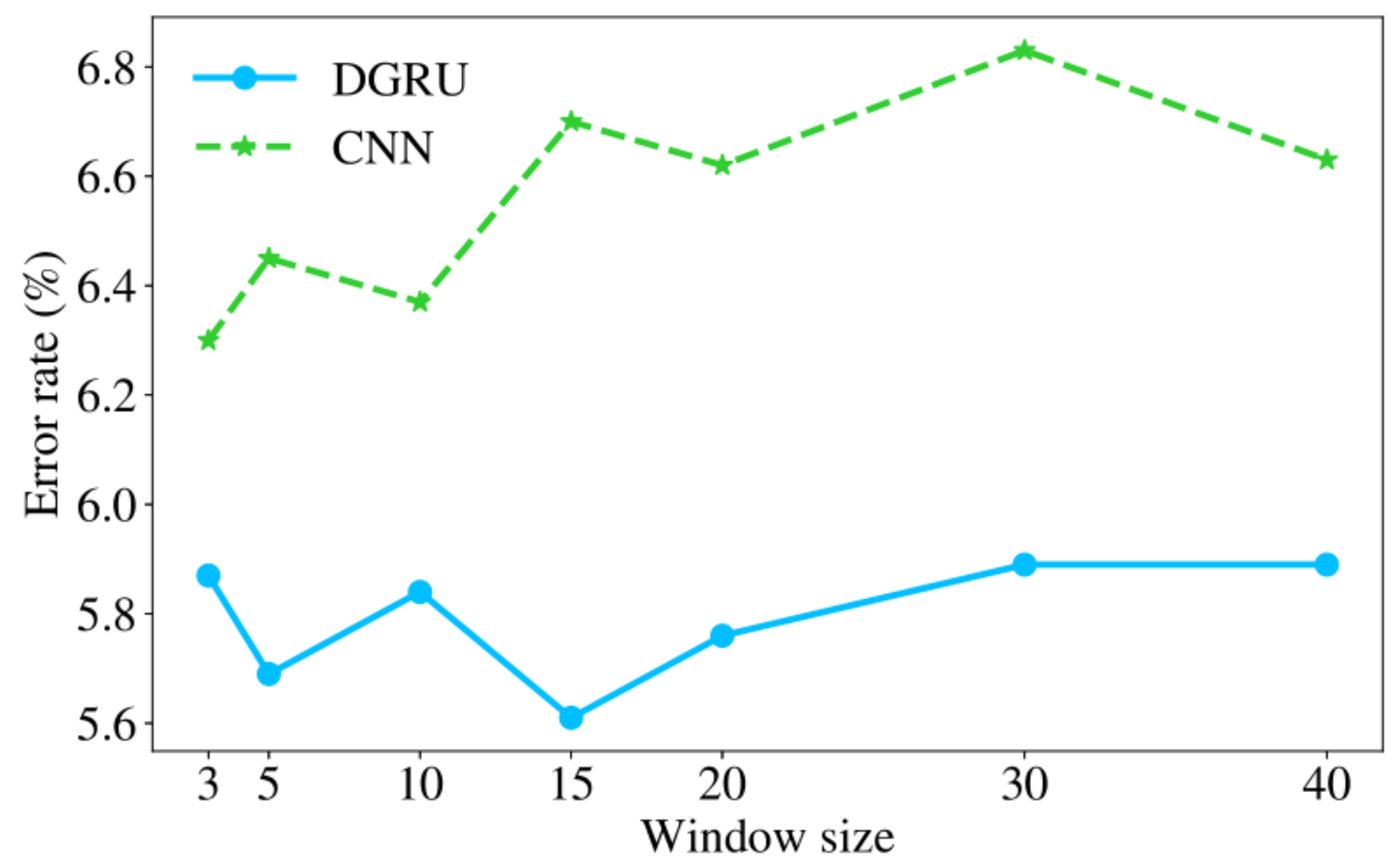
We utilize DRNN in text categorization and conduct experiments on several datasets, the table lists the error rates (%) on seven datasets:

Models	AG	DBP.	Yelp P.	Yelp F.	Yah. A.	Amz. F.	Amz. P.
Linear model (Zhang et al., 2015)	7.64	1.31	4.36	40.14	28.96	44.74	7.98
FastText (Joulin et al., 2017)	7.5	1.4	4.3	36.1	27.7	39.8	5.4
Region.emb (Qiao et al., 2018)	7.2	1.1	4.7	35.1	26.3	39.1	4.7
D-LSTM (Yogatama et al., 2017)	7.9	1.3	7.4	40.4	26.3	-	-
HAN (Yang et al., 2016)	-	-	-	-	24.2	36.4	-
char-CNN (Zhang et al., 2015)	9.51	1.55	4.88	37.95	28.80	40.43	4.93
word-CNN (Zhang et al., 2015)	8.55	1.37	4.60	39.58	28.84	42.39	5.51
VDCNN (Conneau et al., 2017)	8.67	1.29	4.28	35.28	26.57	37.00	4.28
char-CRNN (Xiao and Cho, 2016)	8.64	1.43	5.51	38.18	28.26	40.77	5.87
DRNN	**5.53**	**0.81**	**2.73**	**30.85**	**23.74**	**35.57**	**3.51**

The comparison among models:

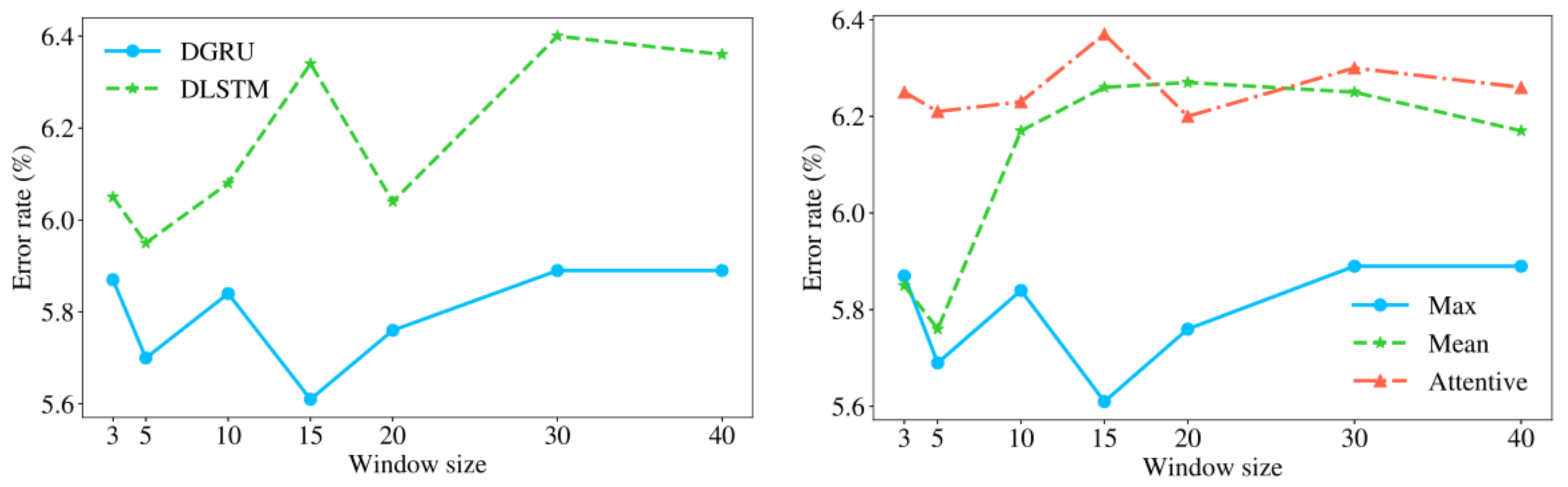
Models	AG	DBP.	Yelp P.
CNN	6.30	1.13	4.08
GRU	6.25	0.96	3.41
LSTM	6.20	0.90	3.20
DRNN	**5.53**	**0.81**	**2.73**

Table： Error rates on three datasets compared to RNN and CNN



Figure： The influence of window size on DRNN compared to CNN

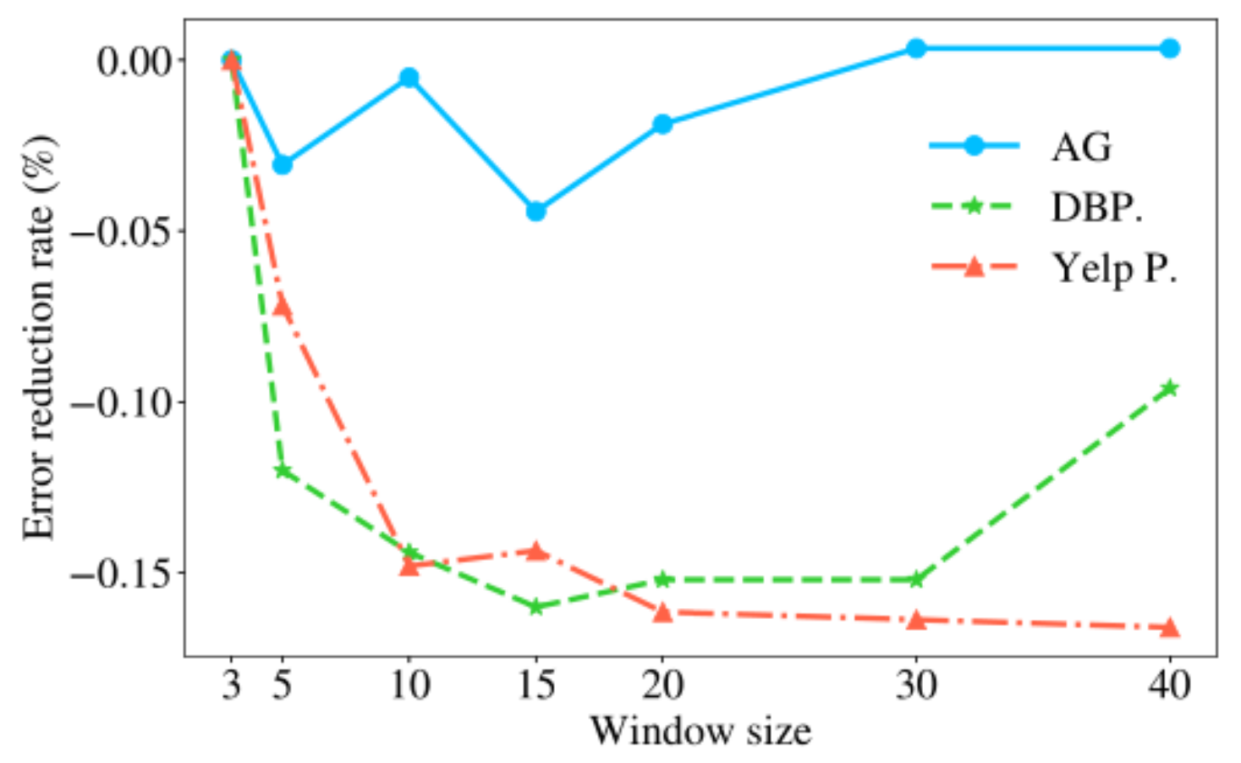
Effects of recurrent units and pooling methods:



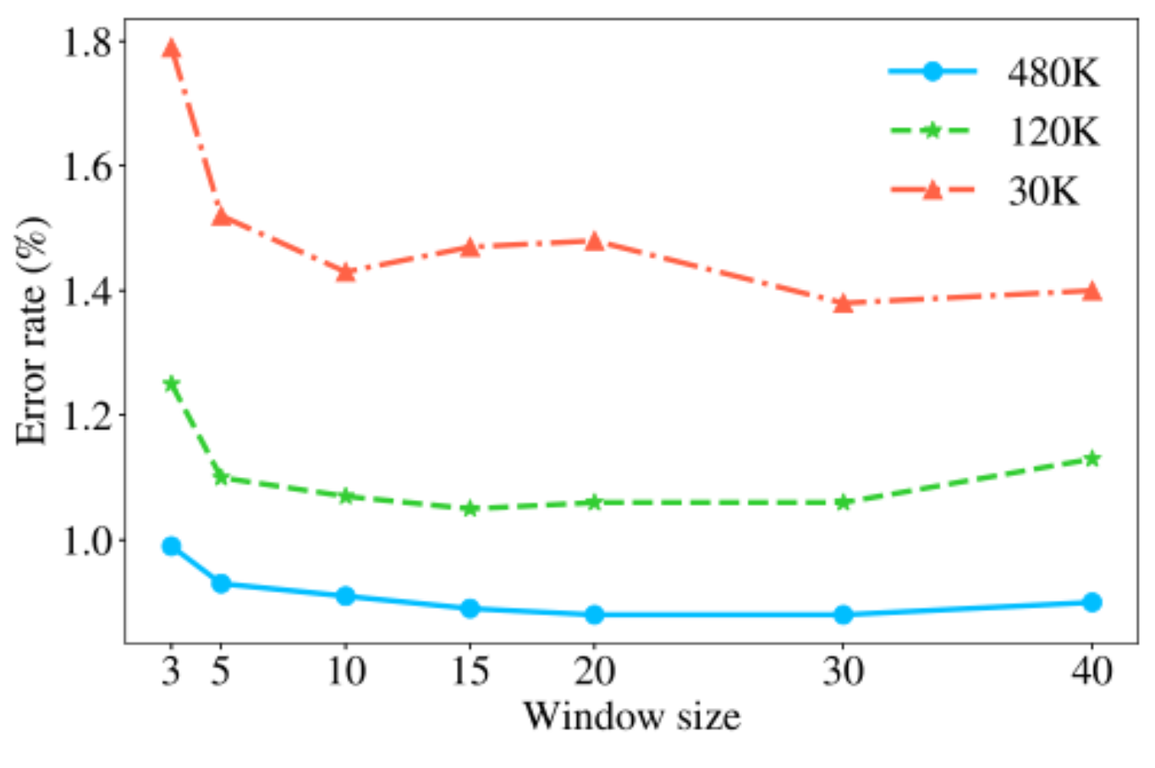
(a) Comparison of recurrent units (b) Comparison of pooling methods

Results of different window sizes:

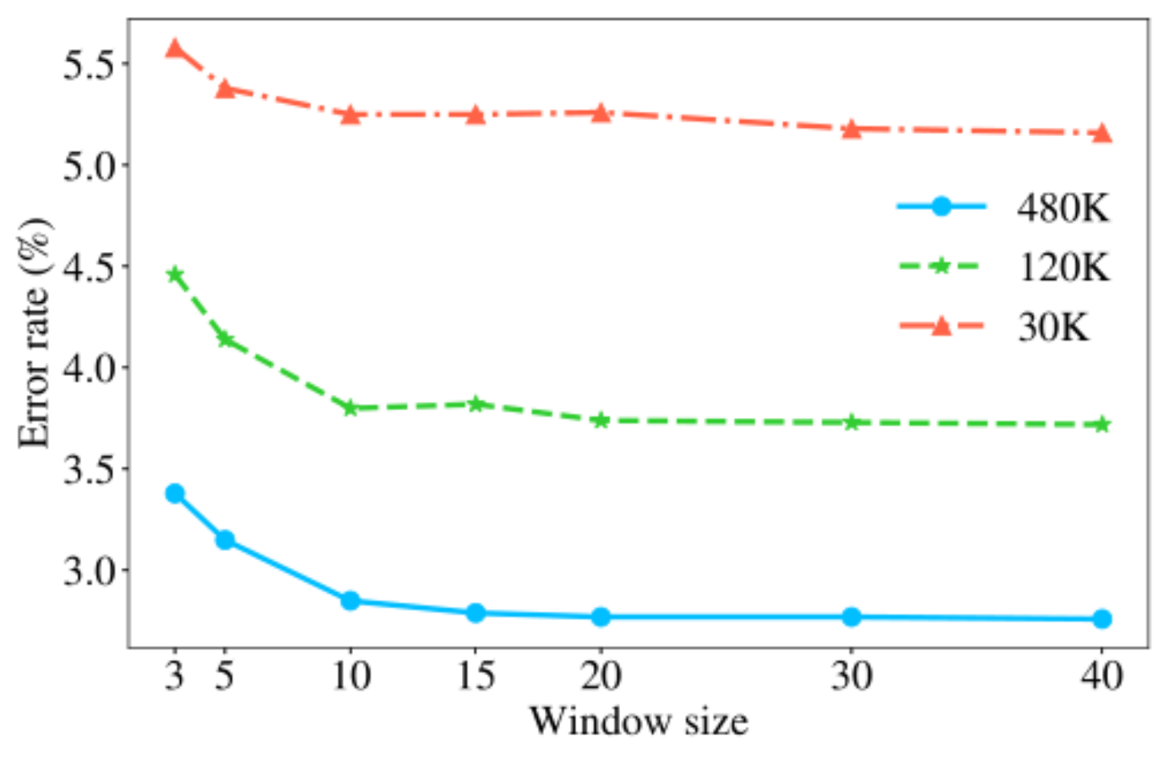
The type of task has a great impact on the optimal window size, while the effect of different training data sizes on the optimal window size seems little.



(a) Different tasks



(b) Different training sets of DBP.



(c) Different training sets of Yelp P.

ACKNOWLEDGMENTS & CONTACT

This work was supported by the National Key Research and Development Program of China (No.2016YFC0800806).

To get in touch with our lab (HFL), please scan the QR-code on the right.

E-mail: bxwang2@iflytek.com

RESEARCH POSTER PRESENTATION DESIGN © 2015
www.PosterPresentations.com