

# Convolutional Neural Network

Technical Detail

# Convolution

นางงาม	0.60	0.89	0.82	0.44
ปี	0.42	0.62	0.42	0.73
นี้	0.53	0.59	0.53	0.02
เนีย	0.11	0.60	0.77	0.27
สวย	0.22	0.42	0.64	0.91
ตรงไหน	0.65	0.81	0.52	0.64
หรือ	0.18	0.29	0.36	0.99

-0.28	-0.08	0.14	0.41
0.16	0.32	0.03	0.15
-0.32	-0.21	-0.14	0.50

0.086

# Convolution

นางงาม	0.60	0.89	0.82	0.44
ปี	0.42	0.62	0.42	0.73
นี้	0.53	0.59	0.53	0.02
เนีย	0.11	0.60	0.77	0.27
สวย	0.22	0.42	0.64	0.91
ตรงไหน	0.65	0.81	0.52	0.64
หรือ	0.18	0.29	0.36	0.99

-0.28	-0.08	0.14	0.41
0.16	0.32	0.03	0.15
-0.32	-0.21	-0.14	0.50

0.086
0.3494

# Convolution

นางงาม	0.60	0.89	0.82	0.44
ปี	0.42	0.62	0.42	0.73
นี้	0.53	0.59	0.53	0.02
เนีย	0.11	0.60	0.77	0.27
สวย	0.22	0.42	0.64	0.91
ตรงไหน	0.65	0.81	0.52	0.64
หรือ	0.18	0.29	0.36	0.99

-0.28	-0.08	0.14	0.41	0.086
0.16	0.32	0.03	0.15	0.3494
-0.32	-0.21	-0.14	0.50	0.3668

# Convolution

นางงาม	0.60	0.89	0.82	0.44
ปี	0.42	0.62	0.42	0.73
นี้	0.53	0.59	0.53	0.02
เนีย	0.11	0.60	0.77	0.27
สวย	0.22	0.42	0.64	0.91
ตรงไหน	0.65	0.81	0.52	0.64
เธอ	0.18	0.29	0.36	0.99

-0.28	-0.08	0.14	0.41	0.086
0.16	0.32	0.03	0.15	0.3494
-0.32	-0.21	-0.14	0.50	0.3668
				0.3341

# รู้ตกลงไปจนสุด

นางงาม	0.60	0.89	0.82	0.44
ปี	0.42	0.62	0.42	0.73
นี้	0.53	0.59	0.53	0.02
เนีย	0.11	0.60	0.77	0.27
สวย	0.22	0.42	0.64	0.91
ตรงไหน	0.65	0.81	0.52	0.64
หรือ	0.18	0.29	0.36	0.99

-0.28	-0.08	0.14	0.41	0.086
0.16	0.32	0.03	0.15	0.3494
-0.32	-0.21	-0.14	0.50	0.3668
filter				0.3341
คำศัพท์				1.1684

Embedding ขนาด 4 units

Filter ขนาด 3 units

1D Convolution เพราะเขยิบแค่ทีศเดียว

Stride = 1 เพราะเขยิบทีละหนึ่ง

output

# ใช้ embedding เดียวกัน

นางงาม	0.60	0.89	0.82	0.44
ปี	0.42	0.62	0.42	0.73
นี้	0.53	0.59	0.53	0.02
เนีย	0.11	0.60	0.77	0.27
สวย	0.22	0.42	0.64	0.91
ตรงไหน	0.65	0.81	0.52	0.64
หรือ	0.18	0.29	0.36	0.99

-0.28	-0.08	0.14	0.41
0.16	0.32	0.03	0.15
-0.32	-0.21	-0.14	0.50

0.086
0.3494
0.3668
0.3341
1.1684

ชา	0.12	0.23	0.41	-0.34
ร้าน	-0.45	0.44	0.91	0.12
นี้	0.53	0.59	0.53	0.02
หรุ	0.22	0.45	0.65	0.44
ตรงไหน	0.65	0.81	0.52	0.64
หรือ	0.18	0.29	0.36	0.99

Embedding เหมือนกัน

Embedding เหมือนกัน

Embedding เหมือนกัน

# ใช้ filter/kernel/feature detector เดียวกัน

นางงาม	0.60	0.89	0.82	0.44
ปี	0.42	0.62	0.42	0.73
นี้	0.53	0.59	0.53	0.02
เนีย	0.11	0.60	0.77	0.27
สวย	0.22	0.42	0.64	0.91
ตรงไหน	0.65	0.81	0.52	0.64
หรือ	0.18	0.29	0.36	0.99

-0.28	-0.08	0.14	0.41
0.16	0.32	0.03	0.15
-0.32	-0.21	-0.14	0.50

0.086
0.3494
0.3668
0.3341
1.1684

ชา	0.12	0.23	0.41	-0.34
ร้าน	-0.45	0.44	0.91	0.12
นี้	0.53	0.59	0.53	0.02
หรุ	0.22	0.45	0.65	0.44
ตรงไหน	0.65	0.81	0.52	0.64
หรือ	0.18	0.29	0.36	0.99

-0.3776
0.524
0.0206
0.9747



# ความยาวประโยคไม่เท่ากัน ทำไงดี

นางงาม	0.60	0.89	0.82	0.44
ปี	0.42	0.62	0.42	0.73
นี้	0.53	0.59	0.53	0.02
เนีย	0.11	0.60	0.77	0.27
สวย	0.22	0.42	0.64	0.91
ตรงไหน	0.65	0.81	0.52	0.64
หรือ	0.18	0.29	0.36	0.99

-0.28	-0.08	0.14	0.41
0.16	0.32	0.03	0.15
-0.32	-0.21	-0.14	0.50

0.086
0.3494
0.3668
0.3341
1.1684

ชา	0.12	0.23	0.41	-0.34
ร้าน	-0.45	0.44	0.91	0.12
นี้	0.53	0.59	0.53	0.02
หรุ	0.22	0.45	0.65	0.44
ตรงไหน	0.65	0.81	0.52	0.64
หรือ	0.18	0.29	0.36	0.99

-0.3776
0.524
0.0206
0.9747

# Pooling แปลงให้เหลือแค่ค่าเดียว

นางงาม	0.60	0.89	0.82	0.44
ปี	0.42	0.62	0.42	0.73
นี้	0.53	0.59	0.53	0.02
เนีย	0.11	0.60	0.77	0.27
สวย	0.22	0.42	0.64	0.91
ตรงไหน	0.65	0.81	0.52	0.64
หรือ	0.18	0.29	0.36	0.99

ชา	0.12	0.23	0.41	-0.34
ร้าน	-0.45	0.44	0.91	0.12
นี้	0.53	0.59	0.53	0.02
หรุ	0.22	0.45	0.65	0.44
ตรงไหน	0.65	0.81	0.52	0.64
หรือ	0.18	0.29	0.36	0.99

-0.28	-0.08	0.14	0.41
0.16	0.32	0.03	0.15
-0.32	-0.21	-0.14	0.50

Filter นี้ detect ว่าเจอ pattern  
positive word + ตรงไหน + หรือ  
หรือไม่

0.086
0.3494
0.3668
0.3341
1.1684

Max 1.1684

-0.3776
0.524
0.0206
0.9747

Max 0.9747

# โมเดลควรรจะมี filter เยอะๆ

นางงาม	0.60	0.89	0.82	0.44
ปี	0.42	0.62	0.42	0.73
นี้	0.53	0.59	0.53	0.02
เนีย	0.11	0.60	0.77	0.27
สวย	0.22	0.42	0.64	0.91
ตรงไหน	0.65	0.81	0.52	0.64
หรือ	0.18	0.29	0.36	0.99

ชา	0.12	0.23	0.41	-0.34
ร้าน	-0.45	0.44	0.91	0.12
นี้	0.53	0.59	0.53	0.02
หฐ	0.22	0.45	0.65	0.44
ตรงไหน	0.65	0.81	0.52	0.64
หรือ	0.18	0.29	0.36	0.99

Filter 1

-0.28	-0.08	0.14	0.41
0.16	0.32	0.03	0.15
-0.32	-0.21	-0.14	0.50

Filter 2

1	0	1	1
1	1	1	0
-1	0	2	-1

Filter 3

-2	-0.3	1	-0.4
0	0	0	1
-1	2	0	-3

0.086	3.83	0.497
0.3494	4.38	-0.598
0.3668	2.71	-2.555
0.3341	2.18	0.222
1.1684	3.3	-2.22

-0.3776	1.6	0.947
0.524	2.87	1.01
0.0206	2.15	-1.225
0.9747	2.84	-2.031

# Max pooling over time

นางงาม	0.60	0.89	0.82	0.44
ปี	0.42	0.62	0.42	0.73
นี้	0.53	0.59	0.53	0.02
เนี่ย	0.11	0.60	0.77	0.27
สวย	0.22	0.42	0.64	0.91
ตรงไหน	0.65	0.81	0.52	0.64
หรือ	0.18	0.29	0.36	0.99

ชา	0.12	0.23	0.41	-0.34
ร้าน	-0.45	0.44	0.91	0.12
นี้	0.53	0.59	0.53	0.02
หรุ	0.22	0.45	0.65	0.44
ตรงไหน	0.65	0.81	0.52	0.64
หรือ	0.18	0.29	0.36	0.99

Filter

-0.28	-0.08	0.14	0.41
0.16	0.32	0.03	0.15
-0.32	-0.21	-0.14	0.50

Filter

1	0	1	1
1	1	1	0
-1	0	2	-1

Filter

-2	-0.3	1	-0.4
0	0	0	1
-1	2	0	-3

Output

0.086	3.83	0.497
0.3494	4.38	-0.598
0.3668	2.71	-2.555
0.3341	2.18	0.222
1.1684	3.3	-2.22

Output

-0.3776	1.6	0.947
0.524	2.87	1.01
0.0206	2.15	-1.225
0.9747	2.84	-2.031

Feature map

1.1684	4.38	0.497
--------	------	-------

Feature map

0.9747	2.87	1.01
--------	------	------

# Dilated filter — dilation = 2

นางงาม	0.60	0.89	0.82	0.44
ปี	0.42	0.62	0.42	0.73
นี้	0.53	0.59	0.53	0.02
เนี่ย	0.11	0.60	0.77	0.27
สวย	0.22	0.42	0.64	0.91
ตรงไหน	0.65	0.81	0.52	0.64
หรือ	0.18	0.29	0.36	0.99

ชา	0.12	0.23	0.41	-0.34
ร้าน	-0.45	0.44	0.91	0.12
นี้	0.53	0.59	0.53	0.02
หรุ	0.22	0.45	0.65	0.44
ตรงไหน	0.65	0.81	0.52	0.64
หรือ	0.18	0.29	0.36	0.99

Filter

-0.28	-0.08	0.14	0.41
0.16	0.32	0.03	0.15
-0.32	-0.21	-0.14	0.50

Filter

1	0	1	1
1	1	1	0
-1	0	2	-1

Filter

-2	-0.3	1	-0.4
0	0	0	1
-1	2	0	-3

Output

0.5553		
0.3332		
0.5382		

Feature map

0.5553		
--------	--	--

Output

0.0276		
0.8582		

Feature map

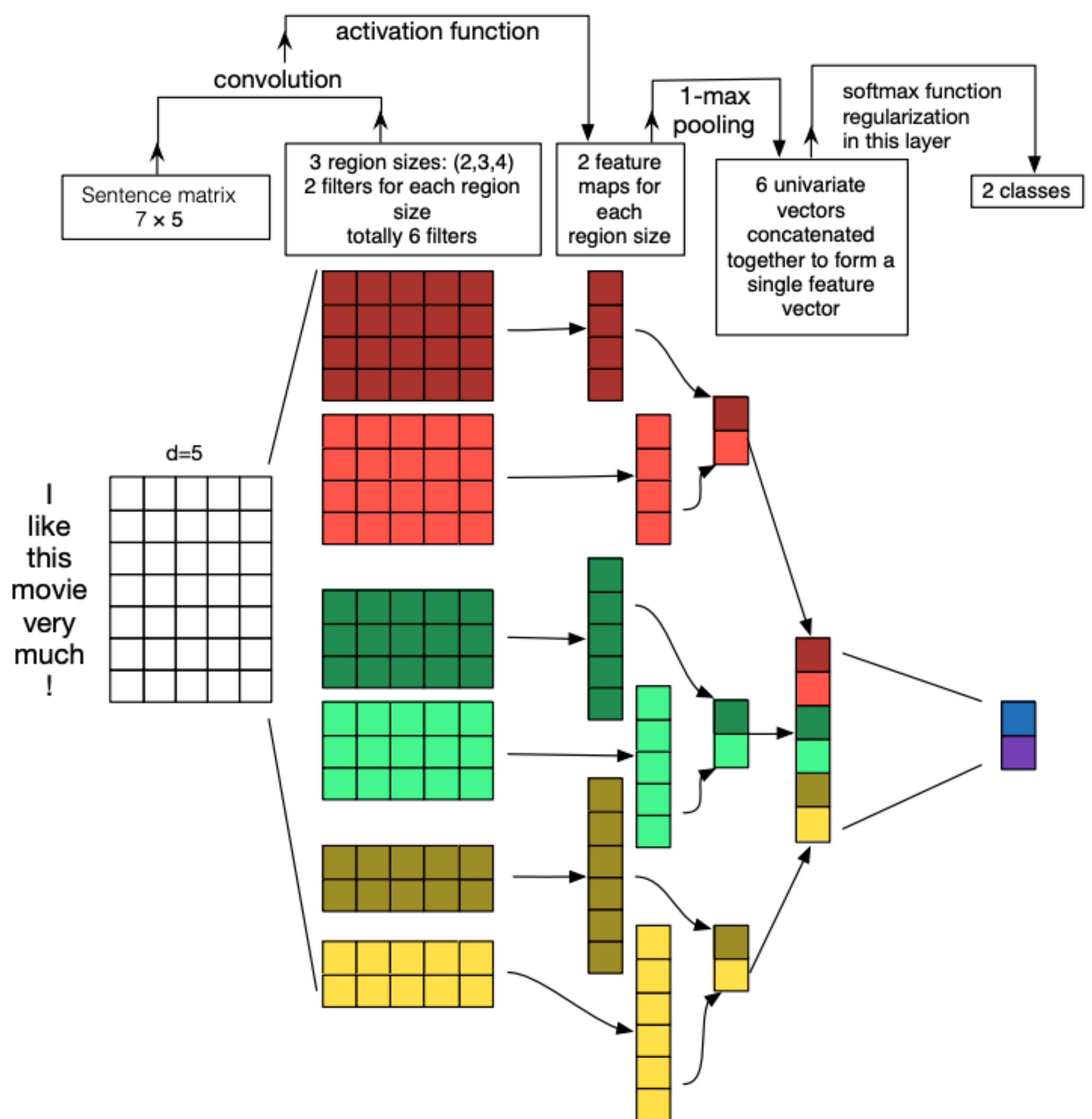
0.8582		
--------	--	--

# Convolutional Network

- Word Embedding : word  $\rightarrow$  embedding
- Filter ขนาดต่าง ๆ : embeddings  $\rightarrow$  output
- (Max) pooling operation: output  $\rightarrow$  feature map

# เวลาใช้จริง

- หลาย ๆ ขนาด
- หลาย ๆ filter





# CNN for Sentence Classification

- Yoon Kim (2014): Convolutional Neural Networks for Sentence Classification. EMNLP 2014.
- Hyperparameters:
  - Word embedding pretrained with word2vec ขนาด 300 units
  - ReLu Activation Function
  - Filter sizes = 3, 4, 5
  - แต่ละขนาดมี 100 filters = 3 x 100 feature maps
  - Dropout  $p = 0.5$
  - L2 regularization
  - Minibatch size = 50



# Experiment Results

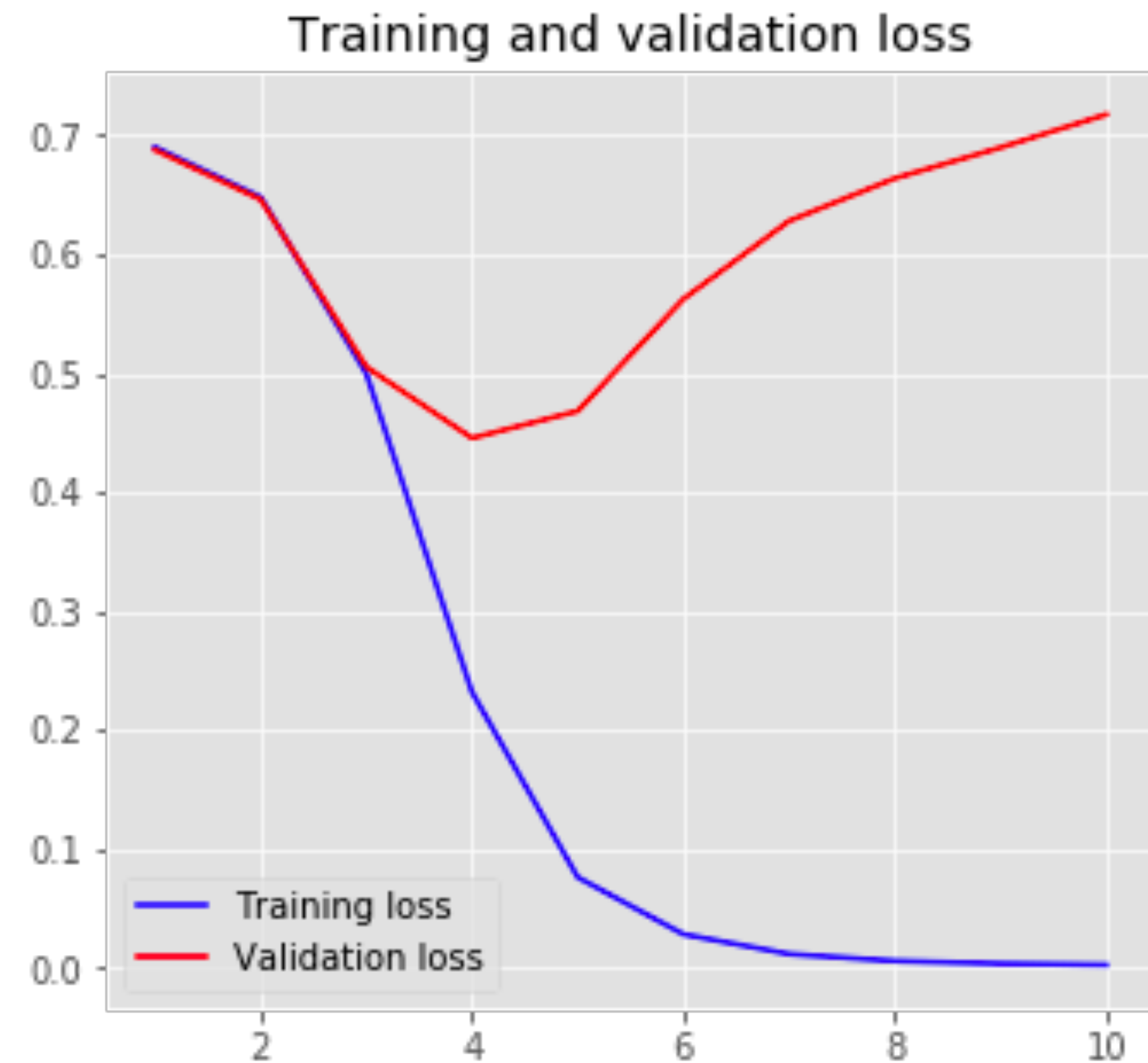
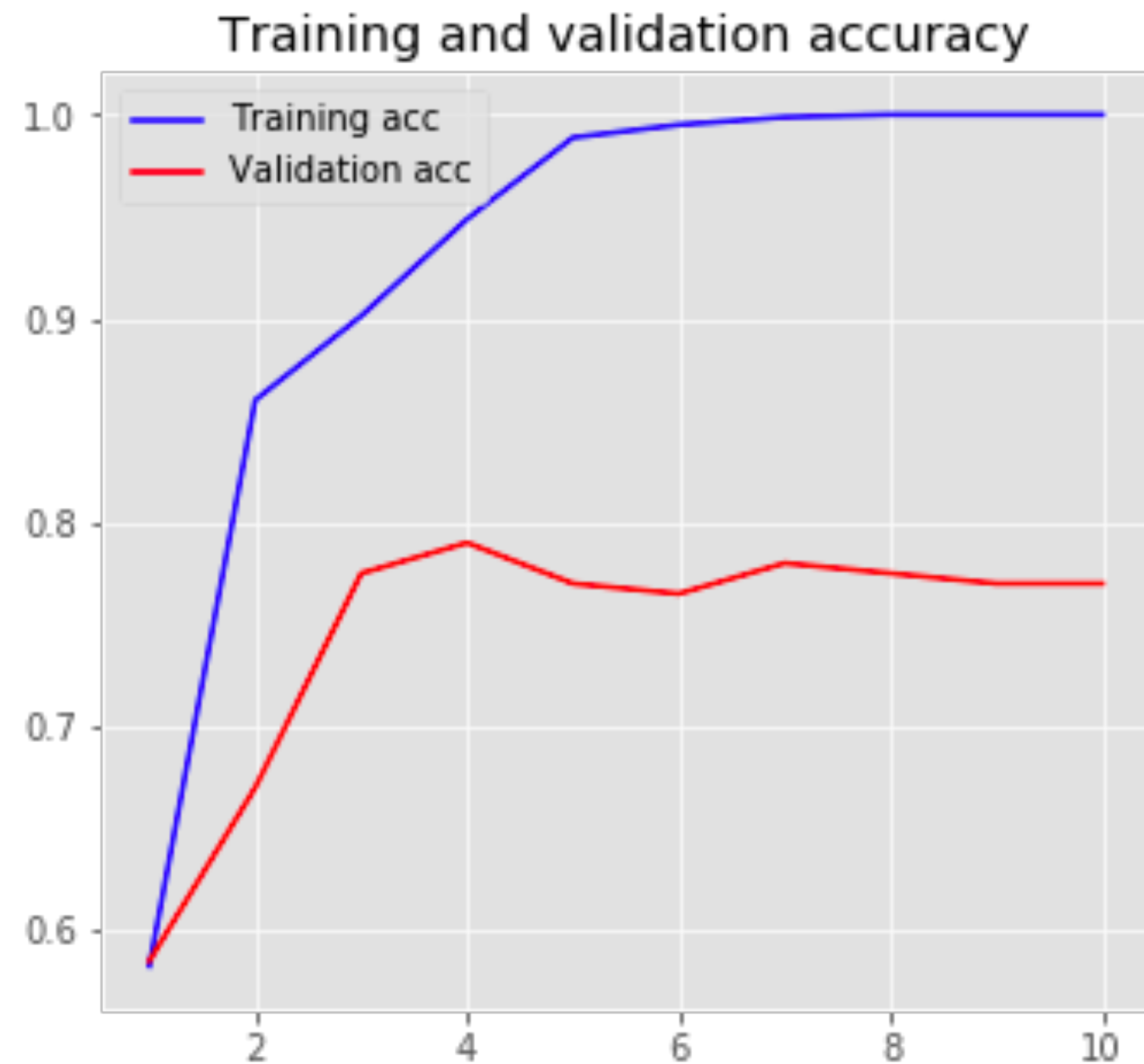
Model	MR	SST-1	SST-2	Subj	TREC	CR	MPQA
CNN-rand	76.1	45.0	82.7	89.6	91.2	79.8	83.4
CNN-static	81.0	45.5	86.8	93.0	92.8	84.7	<b>89.6</b>
CNN-non-static	<b>81.5</b>	48.0	87.2	93.4	93.6	84.3	89.5
CNN-multichannel	81.1	47.4	<b>88.1</b>	93.2	92.2	<b>85.0</b>	89.4
RAE (Socher et al., 2011)	77.7	43.2	82.4	—	—	—	86.4
MV-RNN (Socher et al., 2012)	79.0	44.4	82.9	—	—	—	—
RNTN (Socher et al., 2013)	—	45.7	85.4	—	—	—	—
DCNN (Kalchbrenner et al., 2014)	—	48.5	86.8	—	93.0	—	—
Paragraph-Vec (Le and Mikolov, 2014)	—	<b>48.7</b>	87.8	—	—	—	—
CCAE (Hermann and Blunsom, 2013)	77.8	—	—	—	—	—	87.2
Sent-Parser (Dong et al., 2014)	79.5	—	—	—	—	—	86.3
NBSVM (Wang and Manning, 2012)	79.4	—	—	93.2	—	81.8	86.3
MNB (Wang and Manning, 2012)	79.0	—	—	<b>93.6</b>	—	80.0	86.3
G-Dropout (Wang and Manning, 2013)	79.0	—	—	93.4	—	82.1	86.1
F-Dropout (Wang and Manning, 2013)	79.1	—	—	<b>93.6</b>	—	81.9	86.3
Tree-CRF (Nakagawa et al., 2010)	77.3	—	—	—	—	81.4	86.1
CRF-PR (Yang and Cardie, 2014)	—	—	—	—	—	82.7	—
SVM <sub>S</sub> (Silva et al., 2011)	—	—	—	—	<b>95.0</b>	—	—

# CNN in Keras

```
embedding_dim = 100
```

```
model = Sequential()  
model.add(layers.Embedding(vocab_size, embedding_dim, input_length=maxlen))  
model.add(layers.Conv1D(num_filters, filter_size, activation='relu'))  
model.add(layers.GlobalMaxPooling1D())  
model.add(layers.Dense(num_hidden_units, activation='relu'))  
model.add(layers.Dense(num_labels, activation='sigmoid'))  
model.compile(optimizer='adam',  
              loss='binary_crossentropy',  
              metrics=['accuracy'])  
model.summary()
```

# Typical Learning Curves



# Word Segmentation with CNN

# Word Segmentation as Classification

- Label: ตัวอักษรแรกของคำ (1) หรือ ไม่ใช่ (0)
- Classify แต่ละตัวอักษรแยกกัน

label	1	0	0	1	0	0	0
char	ก	า	ล	เ	ว	ล	า

# Idea: หน่วยต่างๆ ทางภาษาศาสตร์

- ตัวอักษรไทยมีการแยกเป็นหมวดหมู่ เช่น พยัญชนะ วรรณยุกต์ สระที่อยู่ข้างบน ล่าง ซ้าย ขวา ของพยัญชนะ

label	1	0	0	1	0	0	0
char	ก	า	ล	เ	ว	ล	า
Type	C	BV	C	FV	C	C	BV

# Idea: หน่วยต่างๆ ทางภาษาศาสตร์

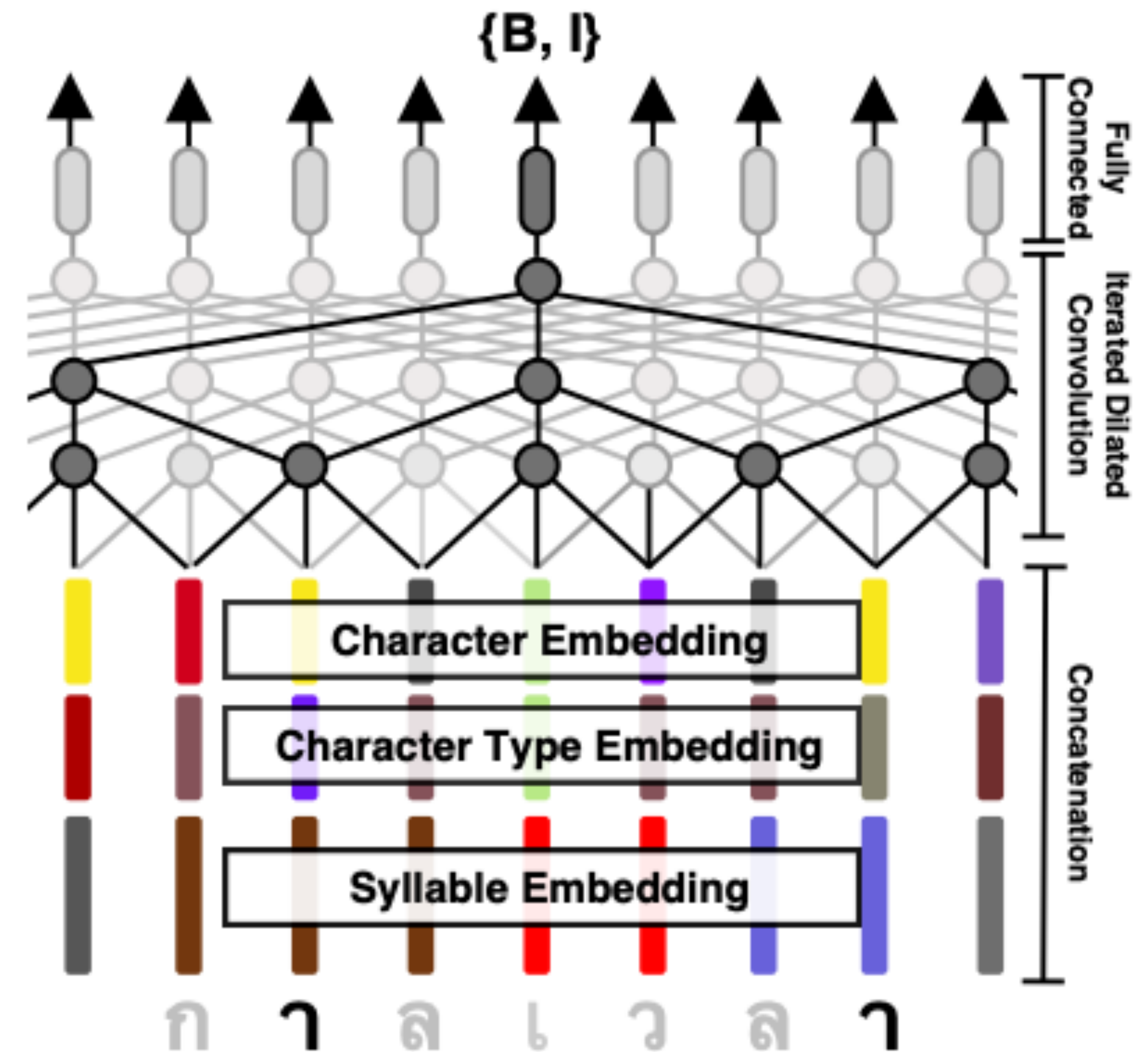
- ตัดเป็นพยางค์ก็น่าจะพอช่วยได้อยู่แล้ว เพราะว่าตัดง่ายกว่ามาก
- คำหนึ่งพยางค์มีเยอะ

label	1	0	0	1	0	0	0
char	ก	า	ล	เ	ว	ล	า
Type	C	BV	C	FV	C	C	BV
syl	กาล	กาล	กาล	เว	เว	ลา	ลา



# What Embedding?

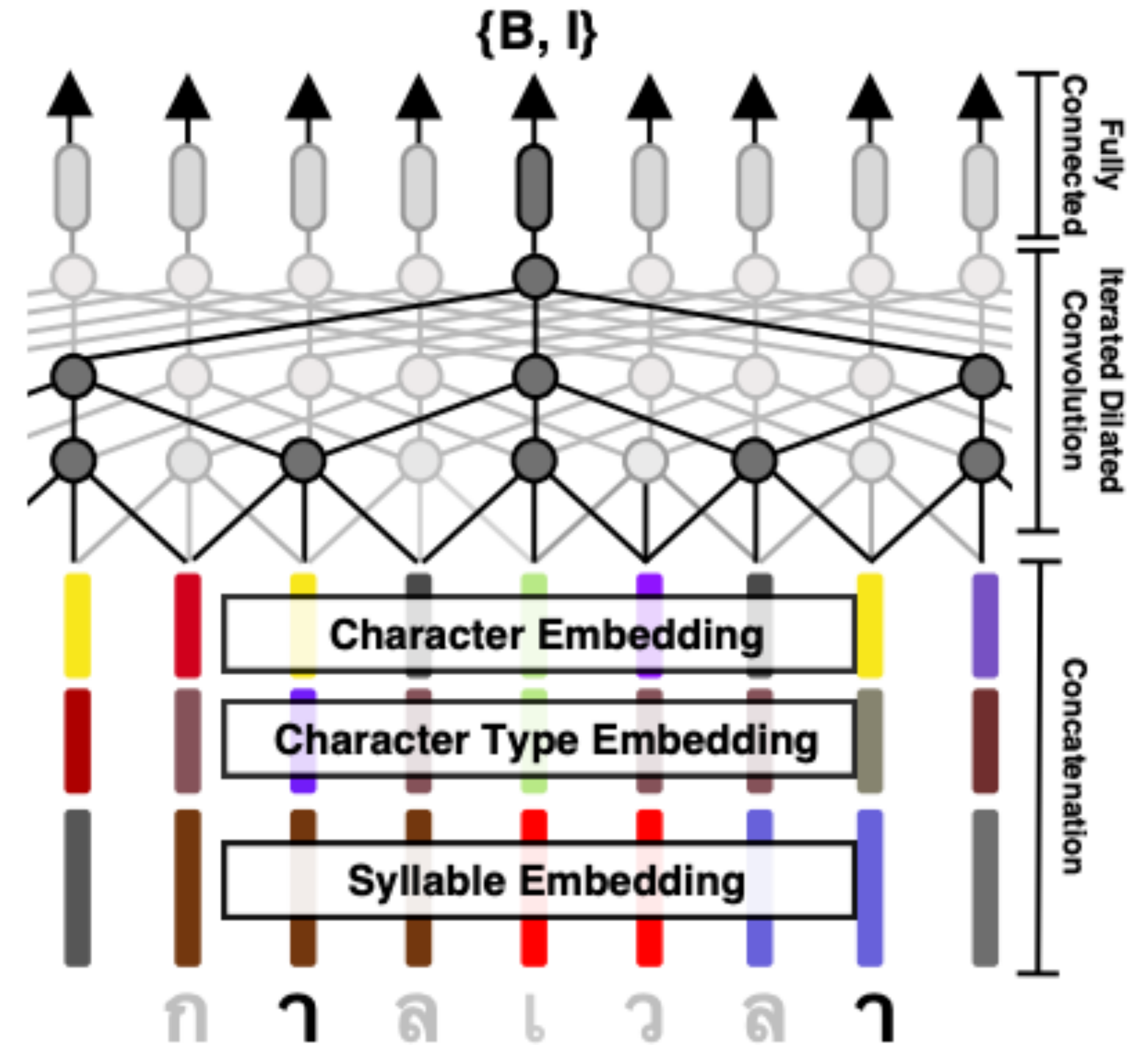
- สระอา ถูก mapped เป็น:
- Character emb. ๑
- Character type emb. สระหลัง
- Syllable emb: กาล
- เอาสาม embedding มาต่อกัน





# Iterated Dilated CNN

- รัน dilated filter หลายๆ ครั้ง โดยที่เพิ่ม dilation ขึ้นเรื่อยๆ เพื่อครอบคลุม context เยอะๆ



# Experimental Results

Method	Dataset ( $WL_{F_1}$ )		
	In-Domain	Out-Domain	
	BEST-2010	Wisesight	TNHC
<u>Previous work</u>			
Dictionary-based	71.18%	78.97%	72.70%
DeepCut	94.46%	84.45%	78.17%
<u>Ours</u>			
BiLSTM			
(CH)-BI	95.05%	85.85%	79.31%
(CH+SY)-BI	<b>95.59%</b>	<b>86.15%</b>	78.70%
-CRF(SY)-BI	95.51%	86.10%	<b>79.89%</b>
ID-CNN			
(CH)-BI	94.31%	85.80%	79.22%
(CH+SY)-BI	95.45%	<b>86.43%</b>	<b>79.87%</b>
-CRF(SY)-SchA*	<b>95.60%</b>	86.15%	79.64%

Method	OOV Recall
Dictionary-based	21.31% (24.70%)
DeepCut	52.09% (60.47%)
BiLSTM	
(CH)-BI	49.51% (59.58%)
(CH+SY)-BI	58.62% (67.42%)
-CRF(SY)-BI	58.37% (65.09%)
ID-CNN	
(CH)-BI	44.83% (51.92%)
(CH+SY)-BI	48.52% (64.09%)
-CRF(SY)-SchemeA	56.53% (65.35%)

# I-Dilated CNN Hyperparameters

- Adam Learning rate = 0.00092
- Adam Weight decay =  $1.17\text{e-}05$
- Number of filters = 200
- Dense layer size = 47
- Dropout  $p$  = 0.21