CS 7016: TOPICS IN DEEP LEARNING

CONVOLUTIONAL NEURAL NETWORKS FOR CHARACTER EMBEDDINGS

Input: “I really love this show!!"

Split it into characters and turn it into a list

I		r	e	a	l	l	y		l	o	v	e		t	h	i	s		s	h	o	w	!	!

zero pads

One-hot encoding

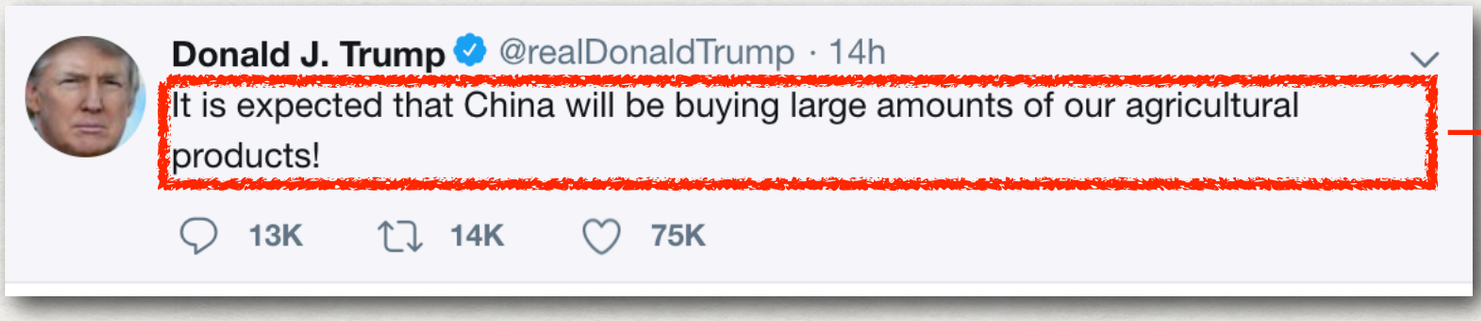
1	0	0
0	0	0
0	0	0
0	0	0
…	…	…
0	0	0
0	0	0
0	0	1
0	0	0
0	0	0

0	0	0	0
0	0	0	0
1	0	0	0
0	0	0	0
…	…	…	…
0	0	0	0
0	0	0	0
0	0		
0	1	0	0
0	0	1	1

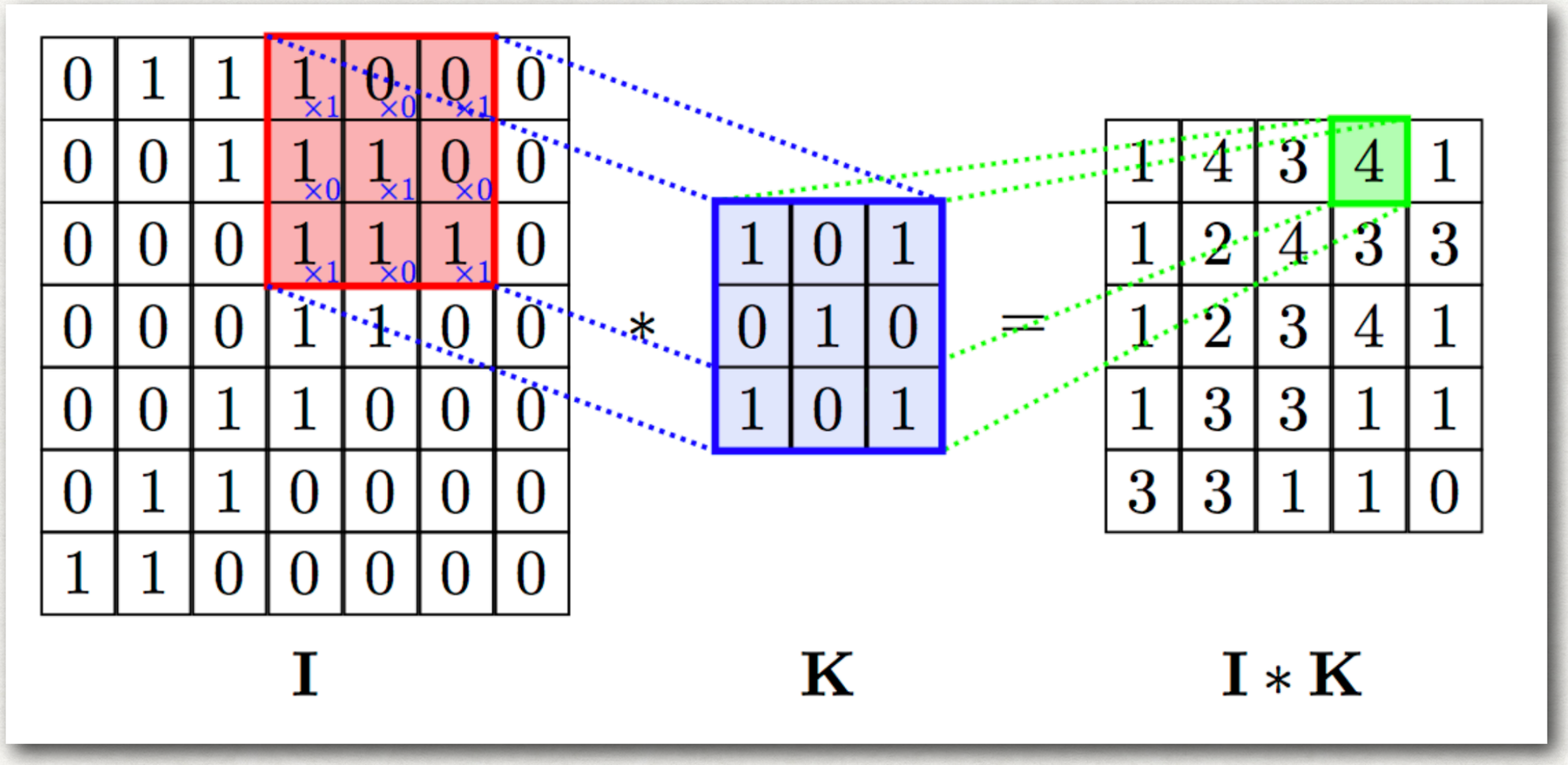
0	0	0
0	0	0
1	0	0
0	0	0
…	…	…
0	0	0
0	0	0
0	0	
0	0	0
0	0	0

Max Length: 140 characters

Vocabulary = 70(characters + digits + special characters)



2D CONVOLUTIONS NOT SUITED FOR TEXT!



Input: “I really love this show!!"

Split it into characters and turn it into a list

1	0	0	0	0	0	0		0	0	0	0	0	0	0
0	0	0	0	0	0	0		0	0	0	0	0	0	0
0	0	0	1	0	0	0		0	0	0	1	0	0	0
0	0	0	0	0	0	0		0	0	0	0	0	0	0
…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
0	0	0	0	0	0	0		0	0	0	0	0	0	0
0	0	1	0	0	0	0		0	0	0	0	0	0	0
0	0	0	0	1	0	0		1	1	0	0	0	0	0
0	0	0	0	0	1	1		0		1	0	0	0	0

140

Input Channel
=
Alphabet size
=
70

1 1-D Kernel

Input: “I really love this show!!"

Split it into characters and turn it into a list

1	0	0	0	0	0	0		0	0	0	0	0	0	0
0	0	0	0	0	0	0		0	0	0	0	0	0	0
0	0	0	1	0	0	0		0	0	0	1	0	0	0
0	0	0	0	0	0	0		0	0	0	0	0	0	0
…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
0	0	0	0	0	0	0		0	0	0	0	0	0	0
0	0	1	0	0	0	0		0	0	0	0	0	0	0
0	0	0	0	1	0	0		1	1	0	0	0	0	0
0	0	0	0	0	1	1		0		1	0	0	0	0

140

Input Channel
=
Alphabet size
=
70

1 1-D Kernel

Input: “I really love this show!!"

Split it into characters and turn it into a list

1	0	0	0	0	0	0		0	0	0	0	0	0	0
0	0	0	0	0	0	0		0	0	0	0	0	0	0
0	0	0	1	0	0	0		0	0	0	1	0	0	0
0	0	0	0	0	0	0		0	0	0	0	0	0	0
…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
0	0	0	0	0	0	0		0	0	0	0	0	0	0
0	0	1	0	0	0	0		0	0	0	0	0	0	0
0	0	0	0	1	0	0		1	1	0	0	0	0	0
0	0	0	0	0	1	1		0		1	0	0	0	0

140

Input Channel
=
Alphabet size
=
70

1 1-D Kernel

Input: “I really love this show!!"

Split it into characters and turn it into a list

1	0	0	0	0	0	0		0	0	0	0	0	0	0
0	0	0	0	0	0	0		0	0	0	0	0	0	0
0	0	0	1	0	0	0		0	0	0	1	0	0	0
0	0	0	0	0	0	0		0	0	0	0	0	0	0
…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
0	0	0	0	0	0	0		0	0	0	0	0	0	0
0	0	1	0	0	0	0		0	0	0	0	0	0	0
0	0	0	0	1	0	0		1	1	0	0	0	0	0
0	0	0	0	0	1	1		0		1	0	0	0	0

140

Input Channel
=
Alphabet size
=
70

1 1-D Kernel

Input: “I really love this show!!"

Split it into characters and turn it into a list

1	0	0	0	0	0	0		0	0	0	0	0	0	0
0	0	0	0	0	0	0		0	0	0	0	0	0	0
0	0	0	1	0	0	0		0	0	0	1	0	0	0
0	0	0	0	0	0	0		0	0	0	0	0	0	0
…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
0	0	0	0	0	0	0		0	0	0	0	0	0	0
0	0	1	0	0	0	0		0	0	0	0	0	0	0
0	0	0	0	1	0	0		1	1	0	0	0	0	0
0	0	0	0	0	1	1		0		1	0	0	0	0

140

Input Channel
=
Alphabet size
=
70

1 1-D Kernel

1	0	1

Batch: 1 “image”
Input Channel = 70
Output Channel = 1
Output Shape =
(1,1,138)

200 1-D Kernels

1	0	1
1	0	1
1	0	1
1	0	1
1	0	1
1	0	1

Batch: 1 “image”
Input Channel = 70
Output Channel = 200
Output Shape =
(1,200,138)

Max pooling operation will result in a 200 dimensional embedding for each “image”

Note that this kernel covers tri-gram, there could be other which can cover bi-gram, five-grams etc.

ACKNOWLEDGMENTS

- The material for these slides has been taken from https://www.youtube.com/watch?v=CNY8VjJt-iQ