

튜토리얼

# End-to-End Sentence Classification

이찬희

chanhee0222@gmail.com



**KOREA**  
UNIVERSITY



Natural Language  
Processing  
& Artificial Intelligence



# Background

Brief Introduction to  
Sentence Classification and  
Neural Networks



# Sentence Classification

- Classifying a sentence into one of pre-defined categories
  - Binary sentiment (e.g. positive / negative)
  - Fine-grained sentiment (e.g. very positive, positive, neutral, negative, very negative)
  - Type (e.g. person, location, numeric information, etc.)
  - ...



## Examples - Binary Sentiment

- no movement , no yuks , not much of anything (NEG)
- this is one of polanski 's best films (POS)
- the film would work much better as a video installation in a museum , where viewers would be free to leave (NEG)
- much of the way , though , this is a refreshingly novel ride (POS)
- ridiculous (NEG)
- a masterpiece four years in the making (POS)



## Examples - Binary Sentiment

- A computational approach to politeness
- To construct a deep learning model for automatically classifying requests according to politeness
- Binary classification - polite/impolite
- Based on Danescu-Niculescu-Mizil et al.'s (2015) *A computational approach to politeness with application to social factors*



## Examples - Binary Sentiment

- Data: a new corpus of requests annotated for politeness (From two large online communities in which members frequently make requests of other members: Wikipedia and StackExchange)

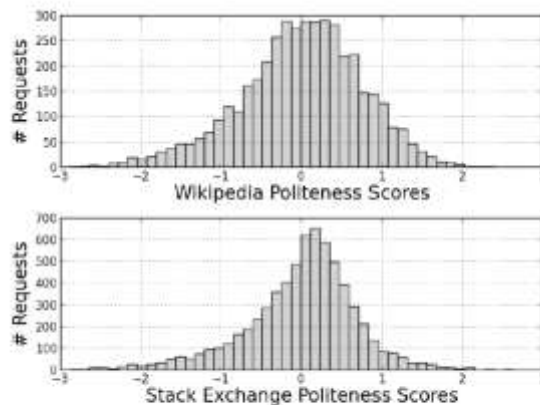
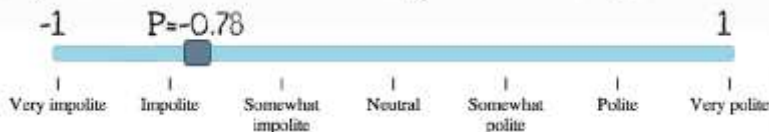


Figure 1: Distribution of politeness scores. Positive scores indicate requests perceived as polite.

"Any code? or do we have to guess at how you did it?"





## Accuracy

Train Test	In-domain		Cross-domain	
	Wiki Wiki	SE SE	Wiki SE	SE Wiki
BOW	79.84%	74.47%	64.23%	72.17%
Ling.	83.79%	78.19%	67.53%	75.43%
Human	86.72%	80.89%	80.89%	86.72%



## Interesting Results

Role	Politeness	Top quart.
Question-asker	0.65***	32%***
Answer-givers	0.52***	20%***





## Interesting Results

Reputation level	Politeness	Top quart.
Low reputation	0.68***	27%***
Middle reputation	0.66***	25%
High reputation	0.64***	23%***

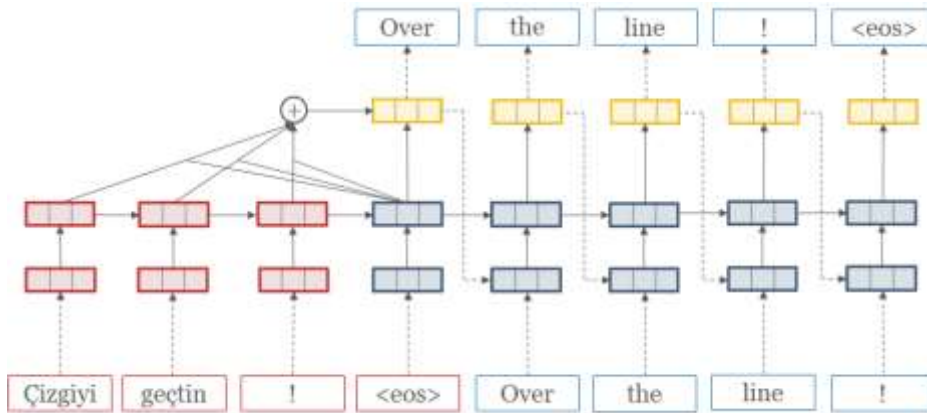


## Interesting Results

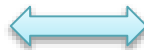
PL name	Politeness	Top quartile
Python	0.47***	23%
Perl	0.49	24%
PHP	0.51	24%
Javascript	0.53**	26%**
Ruby	0.59***	28%*



# Neural Network and Language



Human  
-> Characters



Neural Network  
-> Numbers



# Converting Characters to Numbers

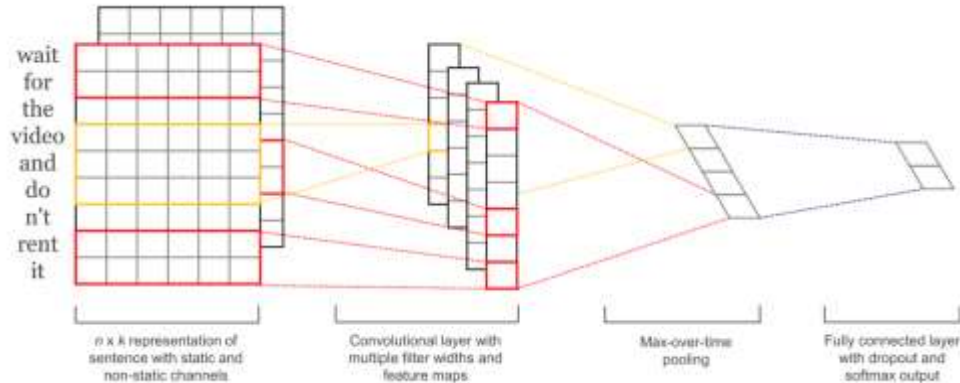
“... .. The fat cat sat  
on the mat. ... ..”



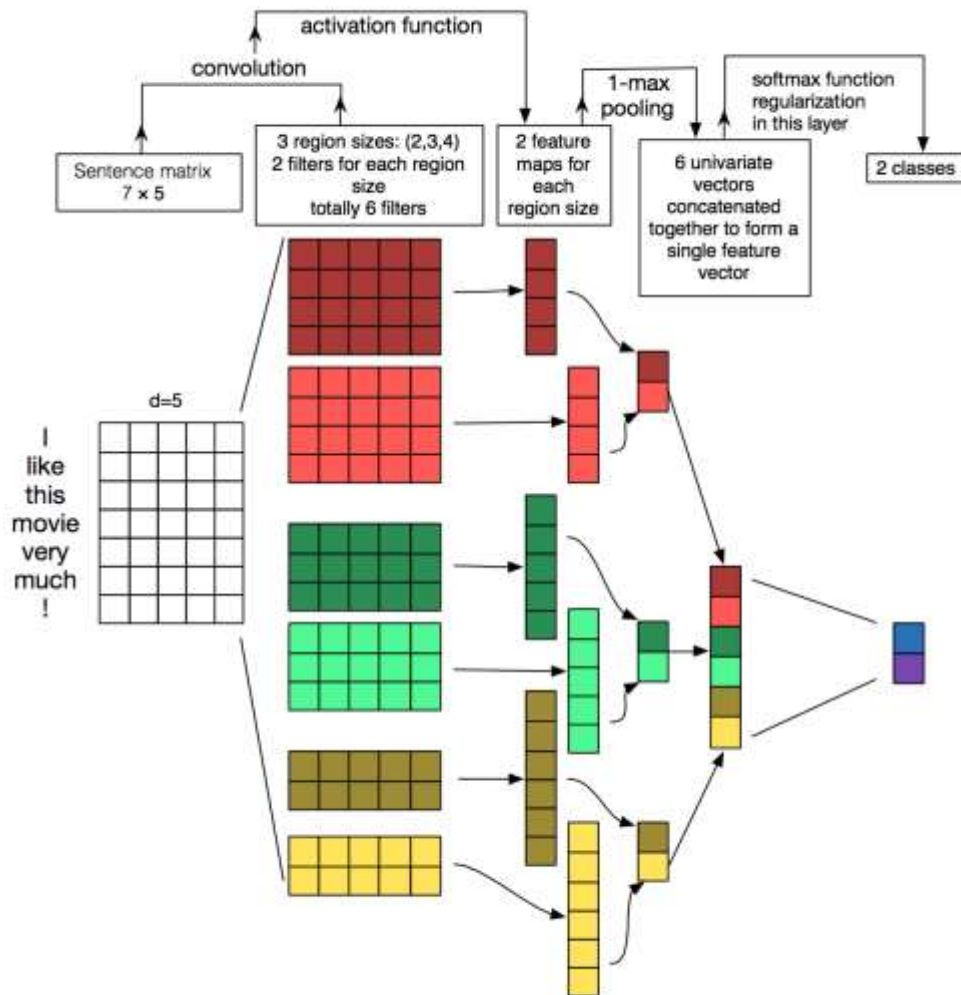
“... .. 32 832 561 634  
6132 32 565 ... ..”



# CNN for Sentence Classification



- Concatenate the vectors of the words constituting the sentence, and apply CNN
- Unlike RNN, only the number of words in the same size as the filter size can be reflected





# THANKS!

Any questions?