

# 2022학년도 1학기 컴퓨터언어학

## 제12강 합성곱 신경망 (2)

박수지

서울대학교 인문대학 언어학과

2022년 4월 13일 수요일

## 오늘의 목표

- 1 Kim (2014)에서 CNN 모델을 사용하여 문장 분류를 훈련한 방법을 설명할 수 있다.

# 모형 개괄

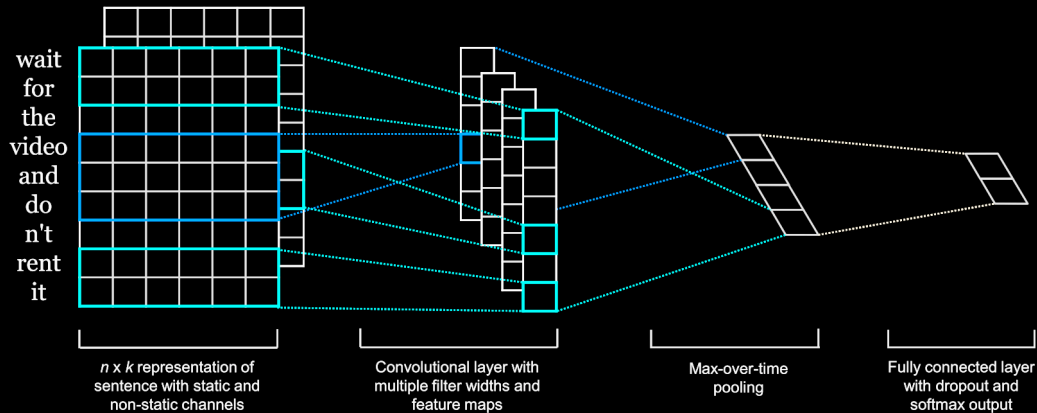


Figure 1: Model architecture with two channels for an example sentence.

██

## 자료

- 단어  $n$ 개로 이루어진 문장
- 각 단어의  $k$ -차원짜리 벡터

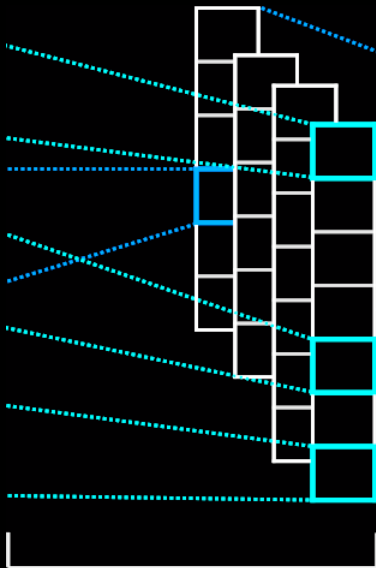
## 문장의 “이미지”

“높이” n, “너비” k

## “차원”의 두 가지 의미

[1, 2, 3, 4, 5]는 몇 차원인가?

- 5차원 벡터
- 크기가 5인 1차원 배열

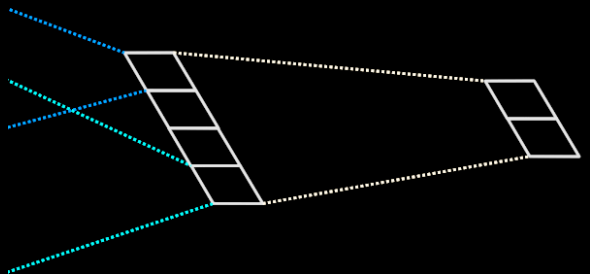


## 합성곱의 차원

**2차원** 필터가 두 방향으로 움직인다.

**1차원** 필터가 한 방향으로 움직인다.

사람의 시선은 단어를 따라 이동한다.



Fully connected layer  
with dropout and  
softmax output

- 드롭아웃
- $l_2$  정규화: 가중치 패러미터 벡터  $W$ 의 크기  $\|W\|_2$ 가 일정 값  $s$ 을 넘지 않도록 한다.

일반적으로 자주 쓰이는 정규화 방식은 아니다.

## 문장 분류 관련 데이터셋

<b>Data</b>	$c$	$l$	$N$	$ V $	$ V_{pre} $	<i>Test</i>
MR	2	20	10662	18765	16448	CV
SST-1	5	18	11855	17836	16262	2210
SST-2	2	19	9613	16185	14838	1821
Subj	2	23	10000	21323	17913	CV
TREC	6	10	5952	9592	9125	500
CR	2	19	3775	5340	5046	CV
MPQA	2	3	10606	6246	6083	CV

Table 1: Summary statistics for the datasets after tokenization.  $c$ : Number of target classes.  $l$ : Average sentence length.  $N$ : Dataset size.  $|V|$ : Vocabulary size.  $|V_{pre}|$ : Number of words present in the set of pre-trained word vectors. *Test*: Test set size (CV means there was no standard train/test split and thus 10-fold CV was used).

## 관찰

- 문장이 짧다.
- 데이터셋이 작다.
- 단어 대부분이 사전학습된 벡터가 있다.

비교적 쉬운 편...

## 주요 데이터셋: 사용자 생성

### MR Movie reviews (Pang and Lee 2005)

- <https://www.cs.cornell.edu/people/pabo/movie-review-data>
- 영어 1문장짜리 영화평 10662개로 구성
  - 긍정 5331개, 부정 5331개
  - Rotten Tomatoes fresh/rotten → 긍정/부정 분류
- 감정분석 연구의 선구: Pang and Lee (2002) “Thumbs up? Sentiment Classification using Machine Learning Techniques”

**SST-1** Stanford Sentiment Treebank

**SST-2** Stanford Sentiment Treebank



## 주요 데이터셋: 전문가의 가공·주석

### TREC Text REtrieval Conference (1992-)

- <https://trec.nist.gov/>
- 미국 NIST에서 주최하는 정보 검색 시스템 경진대회
- 트랙별 테스트셋 제공

### MPQA Multi-perspective Question Answering (2005)

- <http://www.cs.pitt.edu/mpqa>
- 감정 표현들의 의미 주석

## 최근의 동향

클라우드소싱: 불특정 다수의 사람들에게 정답 레이블링을 맡김

- Amazon Mechanical Turk 등

## 참조: 한국어 데이터셋: 사용자 생성

### NSMC Naver Sentiment Movie Corpus (2016)

- <https://github.com/e9t/nsmc>
- 한국어 140자 이내의 네이버 영화평 20만 개로 구성
  - 띄어쓰기, 철자 변형 등 노이즈가 많음
    - ▶ “괜찮네요오랜만포켓몬스터잼있어요”
    - ▶ “한번본적은업지만재미있을것같다”
    - ▶ “완전잼없음보지마삼요후회함.”
- 작성자가 부여한 평점에 따라 긍정/부정 분류
  - 긍정: 9-10점, 부정: 1-4점
- 한국어 문장 분류 연구에서 자주 활용됨

# 구체적 훈련 과정

## 하이퍼패러미터

활성화 함수 ReLU

필터 크기 3, 4, 5

필터 개수 100

드롭아웃 비율 0.5

$l_2$  제약 3

미니배치 크기 50

## 초매개변수

- Word2Vec (CBOW)
- 구글 뉴스 1000억 개 단어에서 학습
- 300차원 벡터

학습된 목록에 없는 단어는 랜덤으로 초기화

## CNN 모형들

**rand** 단어 벡터의 값들을 랜덤으로 초기화 & 업데이트

- Baseline: 아래의 모형들이 이것보다는 잘해야 모형을 만드는 의미가 있다!

**static** 단어 벡터로 Word2Vec 임베딩 값 사용

**non-static** 단어 벡터로 Word2Vec 임베딩 값 사용 & 업데이트

**multichannel** static과 non-static을 별개의 채널로 모두 사용

## 실험 결과

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>	—	—

Table 2: Results of our CNN models against other methods. **RAE**: Recursive Autoencoders with pre-trained word vectors from Wikipedia (Socher et al., 2011). **MV-RNN**: Matrix-Vector Recursive Neural Network with parse trees (Socher et al., 2012).

## 관찰

- 1 채널을 두 개 사용하면 성능이 더 좋아지는가?  
⇒ 그렇지만은 않다(표 2 참조).
- 2 훈련 과정에서 단어 벡터의 값들을 업데이트하면 어떻게 되는가?  
⇒ 과제(예: 감정 분류)의 목적에 맞게 학습된다(표 3 참조).

	Most Similar Words for	
	Static Channel	Non-static Channel
<i>bad</i>	<i>good</i> <i>terrible</i> <i>horrible</i> <i>lousy</i>	<i>terrible</i> <i>horrible</i> <i>lousy</i> <i>stupid</i>
<i>good</i>	<i>great</i> <i>bad</i> <i>terrific</i> <i>decent</i>	<i>nice</i> <i>decent</i> <i>solid</i> <i>terrific</i>
<i>n't</i>	<i>os</i> <i>ca</i> <i>ireland</i> <i>wo</i>	<i>not</i> <i>never</i> <i>nothing</i> <i>neither</i>
<i>!</i>	<i>2,500</i> <i>entire</i> <i>jez</i> <i>changer</i>	<i>2,500</i> <i>lush</i> <i>beautiful</i> <i>terrific</i>
<i>,</i>	<i>decasia</i> <i>abysmally</i> <i>demise</i> <i>valiant</i>	<i>but</i> <i>dragon</i> <i>a</i> <i>and</i>