KISTI 한국과학기술정보연구원 Korea Institute of Science and Technology Information www.kisti.re.kr

과학기술사이버안전센터 Sience and Technology Cyber Security Center

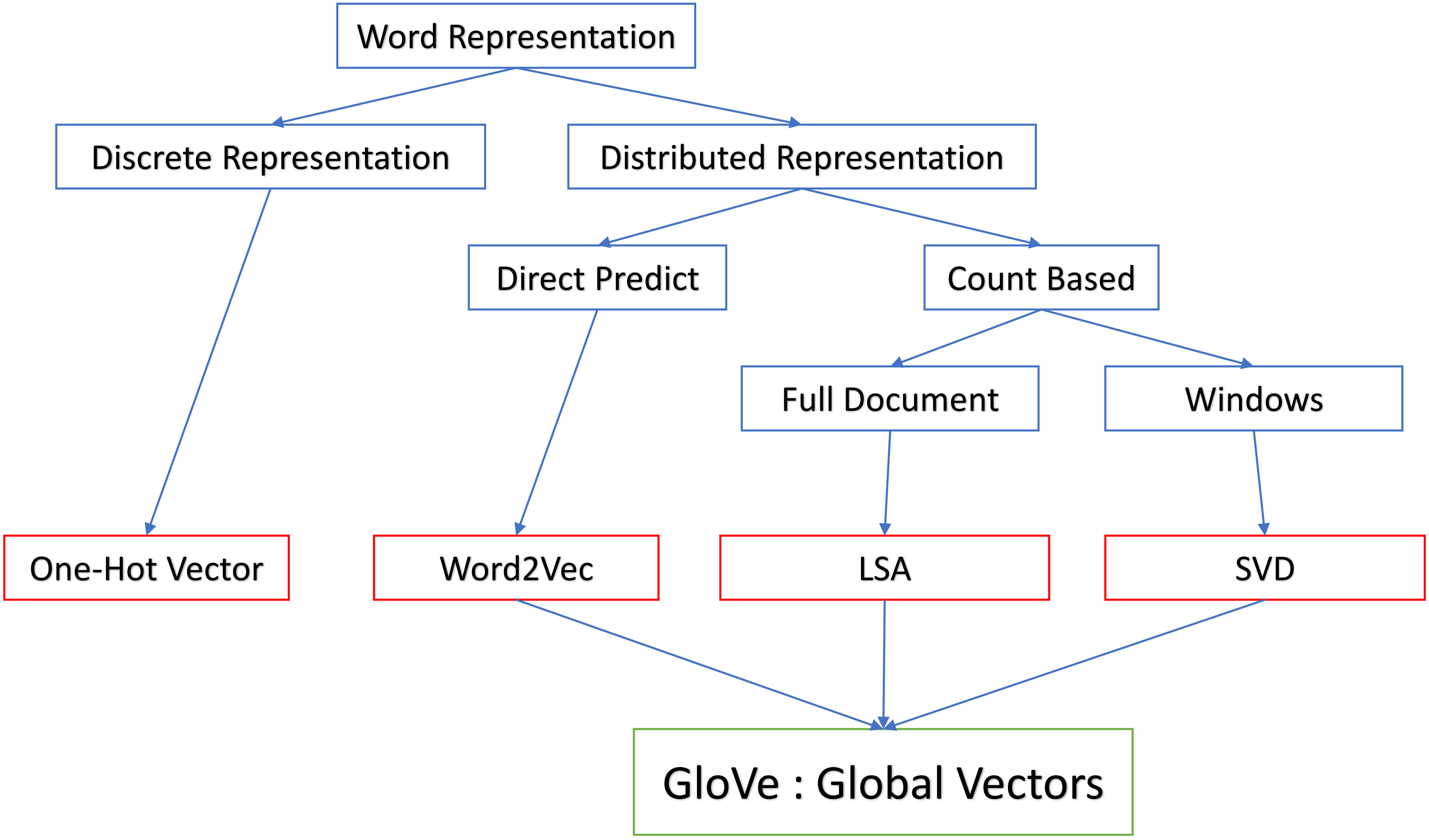
NLP Seminar

GloVe: Global Vectors for Word Representation

2022. 03. 25

KISTI - UST IKJE CHOI

Vectorizing



Vectorizing

Word2Vec

- ✓ Learning words based on where they appear (Skip-grams, CBOW)
- ✓ Advantage : Good at words’ similarity check, widely used
- ✓ Disadvantage : Not using total corpus statistical information

: Center Word

: Context Word

c=0 The cute cat jumps over the lazy dog.

c=1 The cute cat jumps over the lazy dog.

c=2 The cute cat jumps over the lazy dog.

Latent Semantic Analysis (LSA)

- ✓ Counting the frequency of words to derive latent meanings
- ✓ Advantage : Using total corpus statistical information
- ✓ Disadvantage : Frequent words should be considered, low performance

Example corpus

- ✓ I like deep learning.
- ✓ I like NLP
- ✓ I enjoy flying.

counts	I	like	enjoy	deep	learning	NLP	flying	.
I	0	2	1	0	0	0	0	0
like	2	0	0	1	0	1	0	0
enjoy	1	0	0	0	0	0	1	0
deep	0	1	0	0	1	0	0	0
learning	0	0	0	1	0	0	0	1
NLP	0	1	0	0	0	0	0	1
flying	0	0	1	0	0	0	0	1
.	0	0	0	0	1	1	1	0

Global Vectors for Word Representation (GloVe)

- ✓ Method of word representation based on matrix decomposition using co-occurrence probability
- ✓ Model that takes advantages of LSA & Word2Vec
  - ➢ LSA : Using total corpus statistical information
  - ➢ Word2Vec : Good performance at Distributed Representation

counts	I	like	enjoy	deep	learning	NLP	flying	.
I	0	2	1	0	0	0	0	0

Example corpus

- ✓ I like deep learning.
- ✓ I like NLP
- ✓ I enjoy flying.

- ✓ Co-occurrence probability
  - ➢ number of appearances X_{ij}

    $X_{I\ like} = 2$
  - ➢ Total number of appearances X_i

    $X_I = \sum_n X_{In} = 2 + 1 = 3$
  - ➢ Co-occurrence probability

$$P_{ij} = P(j|i) = X_{ij} / X_i = \frac{2}{3}$$

Motivation

✓ Co-occurrence probability

Probability and Ratio	$k = solid$	$k = gas$	$k = water$	$k = fashion$
$P(k\|ice)$	1.9×10^{-4}	6.6×10^{-5}	3.0×10^{-3}	1.7×10^{-5}
$P(k\|steam)$	2.2×10^{-5}	7.8×10^{-4}	2.2×10^{-3}	1.8×10^{-5}
$P(k\|ice)/P(k\|steam)$	8.9	8.5×10^{-2}	1.36	0.96

✓ ”Central word, i” appears frequently with “the specific word, k” the probability value is far from 1

✓ Both “Central word, i" and “the specific word, k” appear frequently or rarely, the probability value is close to 1.

Global Vectors for Word Representation (GloVe)

- ✓ Loss Function
  - ➢ Hypothesis: In context, the rate of Co-occurrence of two words is related to the meaning of the two words.
  - ➢ Objective Function : Dot product central word vector and the specific word vector and make it similar to Co-occurrence probability of the words
  - ➢ Loss Function : Minimize (dot product of two vectors minus Co-occurrence probability of the words)

$$W_I\, V_I \cdot W_{like} V_{like} \doteqdot \log\frac{2}{3}$$

- ✓ Co-occurrence probability
  - ➢ number of appearances X_{ij}
    $X_{I\,like} = 2$
  - ➢ Total number of appearances X_i
    $X_I = \sum_n X_{In} = 2 + 1 = 3$
  - ➢ Co-occurrence probability
    $P_{ij} = P(j|i) = X_{ij} / X_i = \frac{2}{3}$

GloVe

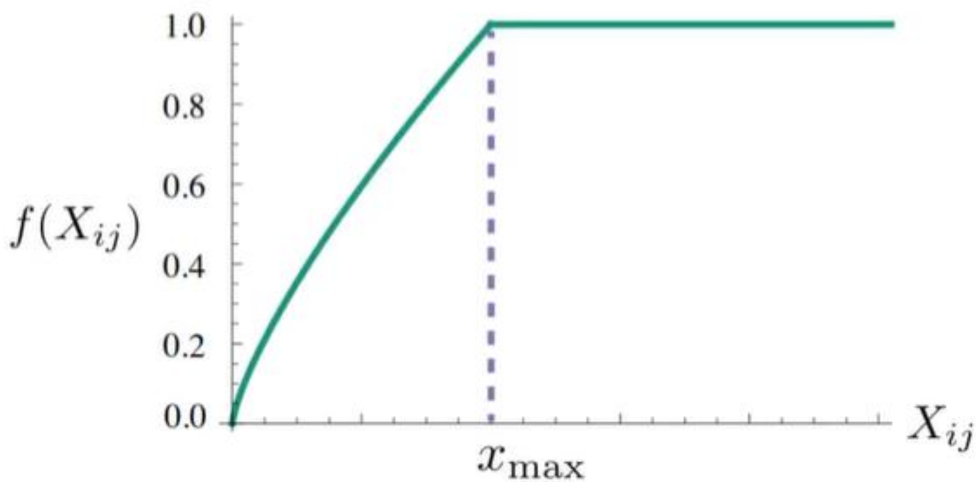
Objective Function

$$F(w_i, w_j, \tilde{w}_k) = \frac{P_{ik}}{P_{jk}}$$

✓ A least squared objective function

$$J = \sum_{i,j=1}^{V} f\left(X_{ij}\right)\left(w_i^T \tilde{w}_j + b_i + \tilde{b}_j - \log X_{ij}\right)^2$$

where $f(x) = \begin{cases} \left(\frac{x}{x_{\max}}\right)^{\alpha} & \text{if } x < x_{\max} \\ 1 & \text{otherwise} \end{cases}$



Glove results

Nearest words to **frog**:

1. frogs
2. toad
3. litoria
4. leptodactylidae
5. rana
6. lizard
7. eleutherodactylus

litoria

leptodactylidae

rana

eleutherodactylus

GloVe

Word analogy task

Ex)

- ✓ Korea : Seoul = Mongolia : ?
- = India : ?
- ✓ Study : Studied = go : ?

Model	Dim.	Size	Sem.	Syn.	Tot.
ivLBL	100	1.5B	55.9	50.1	53.2
HPCA	100	1.6B	4.2	16.4	10.8
GloVe	100	1.6B	67.5	54.3	60.3
SG	300	1B	61	61	61
CBOW	300	1.6B	16.1	52.6	36.1
vLBL	300	1.5B	54.2	64.8	60.0
ivLBL	300	1.5B	65.2	63.0	64.0
GloVe	300	1.6B	80.8	61.5	70.3
SVD	300	6B	6.3	8.1	7.3
SVD-S	300	6B	36.7	46.6	42.1
SVD-L	300	6B	56.6	63.0	60.1
CBOW†	300	6B	63.6	67.4	65.7
SG†	300	6B	73.0	66.0	69.1
GloVe	300	6B	77.4	67.0	71.7
CBOW	1000	6B	57.3	68.9	63.7
SG	1000	6B	66.1	65.1	65.6
SVD-L	300	42B	38.4	58.2	49.2
GloVe	300	42B	**81.9**	**69.3**	**75.0**

Word Similarity

- ✓ Two words given and similarity check

Table 3: Spearman rank correlation on word similarity tasks. All vectors are 300-dimensional. The CBOW* vectors are from the word2vec website and differ in that they contain phrase vectors.

Model	Size	WS353	MC	RG	SCWS	RW
SVD	6B	35.3	35.1	42.5	38.3	25.6
SVD-S	6B	56.5	71.5	71.0	53.6	34.7
SVD-L	6B	65.7	72.7	75.1	56.5	37.0
CBOW†	6B	57.2	65.6	68.2	57.0	32.5
SG†	6B	62.8	65.2	69.7	58.1	37.2
GloVe	6B	65.8	72.7	77.8	53.9	38.1
SVD-L	42B	74.0	76.4	74.1	58.3	39.9
GloVe	42B	**75.9**	**83.6**	**82.9**	**59.6**	**47.8**
CBOW*	100B	68.4	79.6	75.4	59.4	45.5

GloVe

Name Entity Recognition

- ✓ Human Name
- ✓ Place Name
- ✓ Time
- ✓ Extra…

Table 4: F1 score on NER task with 50d vectors. *Discrete* is the baseline without word vectors. We use publicly-available vectors for HPCA, HSMN, and CW. See text for details.

Model	Dev	Test	ACE	MUC7
Discrete	91.0	85.4	77.4	73.4
SVD	90.8	85.7	77.3	73.7
SVD-S	91.0	85.5	77.6	74.3
SVD-L	90.5	84.8	73.6	71.5
HPCA	92.6	**88.7**	81.7	80.7
HSMN	90.5	85.7	78.7	74.7
CW	92.2	87.4	81.7	80.2
CBOW	93.1	88.2	82.2	81.1
GloVe	**93.2**	88.3	**82.9**	**82.2**

Reference

GloVe: Global Vectors for Word Representation

- ✓ (2014) J.Pennington et al

THANK YOU