GloVe

Jason Park
jason_park@korea.ac.kr
Data Science and Business Analytics lab

KOREA UNIVERSITY

GloVe Review

- $X \in \mathbb{R}^{V \times V}$: Co-occurrence matrix
- X_{ij} : Frequency of word i co-occurring with word j
- $X_i = \sum_k^V X_{ik}$: total number of occurring
- $P_{ij} = P(j|i) = \frac{X_{ij}}{X_i}$: Probability of word j occurring within the context of word i
- $w \in \mathbb{R}^d$: a word embedding of dimension d
- $\widetilde{w} \in \mathbb{R}^d$: a context word embedding of dimension d

GloVe Review

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

- Want to preserve $\frac{P(k|ice)}{P(k|steam)}$!

GloVe Review

단어 3개의 관계를 F라는 함수를 이용하여 P의 관계로 표현

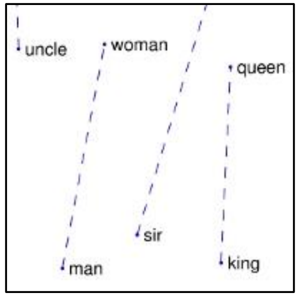
$$F(w_i, w_j, \tilde{w}_k) = \frac{P_{ik}}{P_{jk}}$$

w_i와 w_j의 관계를 차로 표현

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

w_i와 w_j의 관계에 $\tilde{w}_k$를 내적으로 엮고 싶음

$$F\left((w_i - w_j)^T \tilde{w}_k\right) = \frac{P_{ik}}{P_{jk}}$$



Issue

- w와 $\tilde{w}$를 바꾸어서도 쓰고 싶음

- 저자의 주장
  - 이런 Symmetry를 유지하기 위해 다음 2가지 trick 사용
    - $\boxed{F: \mathbb{R} \rightarrow \mathbb{R}_{>0}, \text{ homomorphism}}$ **Why?**
    - $w_i^T \tilde{w}_k = \log(P_{ik}) = \log(X_{ik}) - \log(X_i)$
      $w_i^T \tilde{w}_k = \log(P_{ik}) = \log(X_{ik}) - b_i - \tilde{b}_k$

which prevents F from mixing the vector dimensions in undesirable ways. Next, note that for word-word co-occurrence matrices, the distinction between a word and a context word is arbitrary and that we are free to exchange the two roles. To do so consistently, we must not only exchange $w \leftrightarrow \tilde{w}$ but also $X \leftrightarrow X^T$. Our final model should be invariant under this relabeling, but Eqn. (3) is not. However, the symmetry can be restored in two steps. First, we require that F be a homomorphism between the groups $(\mathbb{R},+)$ and $(\mathbb{R}_{>0},\times)$, i.e.,

$$F\left((w_i - w_j)^T \tilde{w}_k\right) = \frac{F(w_i^T \tilde{w}_k)}{F(w_j^T \tilde{w}_k)}, \quad (4)$$

which, by Eqn. (3), is solved by,

$$F(w_i^T \tilde{w}_k) = P_{ik} = \frac{X_{ik}}{X_i}. \quad (5)$$

The solution to Eqn. (4) is $F = \exp$, or,

$$w_i^T \tilde{w}_k = \log(P_{ik}) = \log(X_{ik}) - \log(X_i). \quad (6)$$

Homomorphism

- 수학 Abstract Algebra의 Group theory에서 다룸
- Group에서 정의하는 함수 homomorphism

- Group(군)
  - 다음 group axiom을 만족하는 set G와 operation $\circ$의 쌍 $(G,\circ)$
    - Closure
    - Associativity
    - Identity element
    - Inverse element

  - 쉽게 말하면 공간과 연산을 하나로 묶었다고 생각
  - Ex) $(\mathbb{R}_{>0},\times),(\mathbb{R},+)$ 등

Homomorphism

- Homomorphism F between $(G,\circ)$ and $(H,\diamond)$
  - $F : G \rightarrow H$ such that

$$\forall g_1, g_2 \in G$$
$$F(g_1 \circ g_2) = F(g_1) \diamond F(g_2)$$

  - 즉, 연산을 보존하는 함수
  - Ex)$\log(x) : (\mathbb{R}_{>0},\times) \rightarrow (\mathbb{R},+)$

- 저자는 $(\mathbb{R},+)$로부터 $(\mathbb{R}_{>0},\times)$로의 homomorphism을 찾고자 함 **Why?**

consistently, we must not only exchange $w \leftrightarrow \tilde{w}$ but also $X \leftrightarrow X^T$. Our final model should be invariant under this relabeling, but Eqn. (3) is not. However, the symmetry can be restored in two steps. First, we require that F be a homomorphism between the groups $(\mathbb{R},+)$ and $(\mathbb{R}_{>0},\times)$, i.e.,

$$F\left((w_i - w_j)^T \tilde{w}_k\right) = \frac{F(w_i^T \tilde{w}_k)}{F(w_j^T \tilde{w}_k)}, \qquad (4)$$

which, by Eqn. (3), is solved by,

Homomorphism in GloVe

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

- Want to preserve $\frac{P(k|ice)}{P(k|steam)}$ using $F\left((w_i - w_j)^T \tilde{w}_k\right) = \frac{P_{ik}}{P_{jk}}$

$$\frac{P(solid|ice)}{P(solid|steam)} \approx F\big((ice - steam)solid\big)$$

$$\frac{P(solid|steam)}{P(solid|ice)} \approx F\big((steam - ice)solid\big)$$

Homomorphism in GloVe

$$\frac{P(solid|ice)}{P(solid|steam)} \approx F\big((ice - steam)solid\big)$$

$$\frac{P(solid|steam)}{P(solid|ice)} \approx F\big((steam - ice)solid\big)$$

$$F\big((ice - steam)solid\big) = \frac{P(solid|ice)}{P(solid|steam)} = \frac{1}{F\big((steam - ice)solid\big)}$$

- $(ice - steam)solid = -(steam - ice)solid$: 덧셈의 역원 관계
- $F\big((ice - steam)solid\big) = \frac{1}{F\big((steam - ice)solid\big)}$: 곱셈의 역원 관계

Homomorphism in GloVe

- $(ice - steam)solid = -(steam - ice)solid$: 덧셈의 역원 관계
- $F\big((ice - steam)solid\big) = \frac{1}{F\big((steam-ice)solid\big)}$: 곱셈의 역원 관계

- Homomorphism은 연산을 보존하고, 이에 따라 역원을 보존함에 착안

- $(\mathbb{R}, +)$를 $(\mathbb{R}_{>0}, \times)$로 사상하는 homomorphism 고려

Homomorphism in GloVe

Equation (4) 유도

- F : $(\mathbb{R},+)$를 $(\mathbb{R}_{>0},\times)$로 사상하는 homomorphism

$$w_i^T \widetilde{w}_k = \left(w_i - w_j\right)^T \widetilde{w}_k + w_j^T \widetilde{w}_k$$

$$F\left(w_i^T \widetilde{w}_k\right) = F\left(\left(w_i - w_j\right)^T \widetilde{w}_k + w_j^T \widetilde{w}_k\right)$$

$$= F\left(\left(w_i - w_j\right)^T \widetilde{w}_k\right) \times F\left(w_j^T \widetilde{w}_k\right)$$

$$F\left(\left(w_i - w_j\right)^T \widetilde{w}_k\right) = \frac{F\left(w_i^T \widetilde{w}_k\right)}{F\left(w_j^T \widetilde{w}_k\right)}$$

consistently, we must not only exchange $w \leftrightarrow \tilde{w}$ but also $X \leftrightarrow X^T$. Our final model should be invariant under this relabeling, but Eqn. (3) is not. However, the symmetry can be restored in two steps. First, we require that F be a homomorphism between the groups $(\mathbb{R},+)$ and $(\mathbb{R}_{>0},\times)$, i.e.,

$$F\left((w_i - w_j)^T \tilde{w}_k\right) = \frac{F(w_i^T \tilde{w}_k)}{F(w_j^T \tilde{w}_k)}, \tag{4}$$

which, by Eqn. (3), is solved by,

$$F(w_i^T \tilde{w}_k) = P_{ik} = \frac{X_{ik}}{X_i}. \tag{5}$$

Homomorphism in GloVe

$$F\left(\left(w_i - w_j\right)^T \widetilde{w}_k\right) = F\left(w_i^T \widetilde{w}_k - w_j^T \widetilde{w}_k\right)$$
$$= \frac{F\left(w_i^T \widetilde{w}_k\right)}{F\left(w_j^T \widetilde{w}_k\right)}$$
$$= \frac{P_{ik}}{P_{jk}}$$

- 이 식으로부터 $F = exp$ 유도

consistently, we must not only exchange $w \leftrightarrow \tilde{w}$ but also $X \leftrightarrow X^T$. Our final model should be invariant under this relabeling, but Eqn. (3) is not. However, the symmetry can be restored in two steps. First, we require that F be a homomorphism between the groups $(\mathbb{R},+)$ and $(\mathbb{R}_{>0},\times)$, i.e.,

$$F\left((w_i - w_j)^T \tilde{w}_k\right) = \frac{F(w_i^T \tilde{w}_k)}{F(w_j^T \tilde{w}_k)}, \qquad (4)$$

which, by Eqn. (3), is solved by,

$$F(w_i^T \tilde{w}_k) = P_{ik} = \frac{X_{ik}}{X_i}. \qquad (5)$$

GloVe

- 저자의 주장
  - 이런 Symmetry를 유지하기 위해 다음 2가지 trick 사용
    - $F: \mathbb{R} \rightarrow \mathbb{R}_{>0}$, homomorphism **Why?**
    - $w_i^T \tilde{w}_k = \log(P_{ik}) = \log(X_{ik}) - \log(X_i)$

      $w_i^T \tilde{w}_k = \log(P_{ik}) = \log(X_{ik}) - b_i - \tilde{b}_k$

which prevents F from mixing the vector dimensions in undesirable ways. Next, note that for word-word co-occurrence matrices, the distinction between a word and a context word is arbitrary and that we are free to exchange the two roles. To do so consistently, we must not only exchange $w \leftrightarrow \tilde{w}$ but also $X \leftrightarrow X^T$. Our final model should be invariant under this relabeling, but Eqn. (3) is not. However, the symmetry can be restored in two steps. First, we require that F be a homomorphism between the groups $(\mathbb{R},+)$ and $(\mathbb{R}_{>0},\times)$, i.e.,

$$F\left((w_i - w_j)^T \tilde{w}_k\right) = \frac{F(w_i^T \tilde{w}_k)}{F(w_j^T \tilde{w}_k)}, \tag{4}$$

which, by Eqn. (3), is solved by,

$$F(w_i^T \tilde{w}_k) = P_{ik} = \frac{X_{ik}}{X_i}. \tag{5}$$

The solution to Eqn. (4) is $F = \exp$, or,

$$w_i^T \tilde{w}_k = \log(P_{ik}) = \log(X_{ik}) - \log(X_i). \tag{6}$$

GloVe

- 저자의 주장
  - 이런 Symmetry를 유지하기 위해 다음 2가지 trick 사용
    - $\boxed{F: \mathbb{R} \rightarrow \mathbb{R}_{>0}, \text{ homomorphism}}$ **Why?**
    - $w_i^T \tilde{w}_k = \log(P_{ik}) = \log(X_{ik}) - \log(X_i)$

      $w_i^T \tilde{w}_k = \log(P_{ik}) = \log(X_{ik}) - b_i - \tilde{b}_k$

$$\frac{P(solid|ice)}{P(solid|steam)} \approx F\big((ice - steam)solid\big)$$

$$\frac{P(solid|steam)}{P(solid|ice)} \approx F\big((steam - ice)solid\big)$$

- 위 식의 필요로 Homomorphism 고려

which prevents F from mixing the vector dimensions in undesirable ways. Next, note that for word-word co-occurrence matrices, the distinction between a word and a context word is arbitrary and that we are free to exchange the two roles. To do so consistently, we must not only exchange $w \leftrightarrow \tilde{w}$ but also $X \leftrightarrow X^T$. Our final model should be invariant under this relabeling, but Eqn. (3) is not. However, the symmetry can be restored in two steps. First, we require that F be a homomorphism between the groups $(\mathbb{R},+)$ and $(\mathbb{R}_{>0},\times)$, i.e.,

$$F\left((w_i - w_j)^T \tilde{w}_k\right) = \frac{F(w_i^T \tilde{w}_k)}{F(w_j^T \tilde{w}_k)}, \quad (4)$$

which, by Eqn. (3), is solved by,

$$F(w_i^T \tilde{w}_k) = P_{ik} = \frac{X_{ik}}{X_i}. \quad (5)$$

The solution to Eqn. (4) is $F = \exp$, or,

$$w_i^T \tilde{w}_k = \log(P_{ik}) = \log(X_{ik}) - \log(X_i). \quad (6)$$

GloVe

- 저자의 주장
  - 이런 Symmetry를 유지하기 위해 다음 2가지 trick 사용
    - $\boxed{F: \mathbb{R} \rightarrow \mathbb{R}_{>0}, \text{ homomorphism}}$ **Why?**
    - $w_i^T \tilde{w}_k = \log(P_{ik}) = \log(X_{ik}) - \log(X_i)$
      $w_i^T \tilde{w}_k = \log(P_{ik}) = \log(X_{ik}) - b_i - \tilde{b}_k$

$$\frac{P(solid|ice)}{P(solid|steam)} \approx F\big((ice - steam)solid\big)$$
$$\frac{P(solid|steam)}{P(solid|ice)} \approx F\big((steam - ice)solid\big)$$

- Homomorphism의 필요와
  Symmetry의 필요를 분리하여 서술했으면
  읽기 더 편하지 않았을까..

which prevents F from mixing the vector dimensions in undesirable ways. Next, note that for word-word co-occurrence matrices, the distinction between a word and a context word is arbitrary and that we are free to exchange the two roles. To do so consistently, we must not only exchange $w \leftrightarrow \tilde{w}$ but also $X \leftrightarrow X^T$. Our final model should be invariant under this relabeling, but Eqn. (3) is not. However, the symmetry can be restored in two steps. First, we require that F be a homomorphism between the groups $(\mathbb{R},+)$ and $(\mathbb{R}_{>0},\times)$, i.e.,

$$F\left((w_i - w_j)^T \tilde{w}_k\right) = \frac{F(w_i^T \tilde{w}_k)}{F(w_j^T \tilde{w}_k)}, \tag{4}$$

which, by Eqn. (3), is solved by,

$$F(w_i^T \tilde{w}_k) = P_{ik} = \frac{X_{ik}}{X_i}. \tag{5}$$

The solution to Eqn. (4) is $F = \exp$, or,

$$w_i^T \tilde{w}_k = \log(P_{ik}) = \log(X_{ik}) - \log(X_i). \tag{6}$$

Q & A