RAND-WALK: A latent variable model approach to word embeddings

Sanjeev Arora   Yuanzhi Li   Yingyu Liang   Tengyu Ma   Andrej Risteski *

Presented by

Thanawat Sornwanee

April 2026

Pointwise Mutual Information (PMI)

Church & Hanks 1990

$$\mathrm{PMI}(w, w') = \log \frac{p(w, w')}{p(w)p(w')}$$

Pointwise Mutual Information (PMI)

Church & Hanks 1990

$$\mathrm{PMI}(w, w') = \log \frac{p(w,w')}{p(w)p(w')}$$

Empirically low rank!

$$\langle v_w, v_{w'} \rangle \approx \mathrm{PMI}(w, w')$$

This Paper

- Science: Why do we have low rank?

  What is the sensible data generating process leading to this approximate low rank behavior?

Word Embeddings as Metric Recovery in Semantic Spaces

Tatsunori B. Hashimoto, David Alvarez-Melis and **Tommi S. Jaakkola**
CSAIL, Massachusetts Institute of Technology
{thashim, davidam, tommi}@csail.mit.edu

$$P(X_t = j | X_{t-1} = i) = h(\frac{1}{\sigma}||x_i - x_j||_2^2)$$

Random walk on the word embedding directly

Log-linear word production model

- This paper assumes discourse/context c_t

$$\Pr[w \text{ emitted at time } t \mid c_t] \propto \exp(\langle c_t, v_w \rangle)$$

Discourse Random Walk

- Slow random walk
- Stationary being Gaussian

Isotropy

$$v = s \cdot \hat{v}$$

Recipe

Log-linear selection → $\Pr[w \text{ emitted at time } t \mid c_t] \propto \exp(\langle c_t, v_w \rangle)$

Recipe

Log-linear selection → $\Pr[w \text{ emitted at time } t \mid c_t] \propto \exp(\langle c_t, v_w \rangle)$

Isotropy of v
Gaussian Stationary of c → $\Pr[w \text{ emitted at time } t \mid c_t] = \underbrace{\exp(\langle c_t, v_w \rangle)}_{\text{constant}}$

Recipe

Log-linear selection → $\Pr[w \text{ emitted at time } t \mid c_t] \propto \exp(\langle c_t, v_w \rangle)$

Isotropy of v

Gaussian Stationary of c

→ $\Pr[w \text{ emitted at time } t \mid c_t] = \underbrace{\exp(\langle c_t, v_w \rangle)}_{\text{constant}}$

$$p(w, w') = \mathop{\mathbb{E}}_{c,c'} \left[\Pr[w, w' | c, c'] \right]$$

$$= \mathop{\mathbb{E}}_{c,c'} \left[p(w|c) p(w'|c') \right]$$

Recipe

Log-linear selection → $\Pr[w \text{ emitted at time } t \mid c_t] \propto \exp(\langle c_t, v_w \rangle)$

Isotropy of v

Gaussian Stationary of c → $\Pr[w \text{ emitted at time } t \mid c_t] = \underline{\exp(\langle c_t, v_w \rangle)}$

constant

$$p(w, w') = \mathop{\mathbb{E}}_{c,c'} [\Pr[w, w' | c, c']]$$

$$= \mathbb{E} [p(w|c)p(w'|c')]$$

Small Walk of c → $\mathbb{E}[\exp(\langle v_w + v_{w'}, c \rangle)].$

Recipe

Log-linear selection → $\Pr[w \text{ emitted at time } t \mid c_t] \propto \exp(\langle c_t, v_w \rangle)$

Isotropy of v

Gaussian Stationary of c → $\Pr[w \text{ emitted at time } t \mid c_t] = \underbrace{\exp(\langle c_t, v_w \rangle)}_{\text{constant}}$

$$p(w, w') = \mathop{\mathbb{E}}_{c,c'} \left[\Pr[w, w' | c, c']\right]$$

$$= \mathop{\mathbb{E}}_{.} \left[p(w|c)p(w'|c')\right]$$

Small Walk of c → $\mathbb{E}[\exp(\langle v_w + v_{w'}, c \rangle)].$

Gaussian Stationary of c → $\exp\left(\frac{\|v_w + v_{w'}\|^2}{2d}\right)$

Relation

Practice

$$\text{“}man{:}woman{::}king{:}??,\text{”}$$

$$v_{king} - v_{man} + v_{woman}.$$

The paper probably explains that we want the change of measure density ratio for the two changes to be similar

Discussion

We assume

- 1. Log-linear selection
- 2. Small discourse walk
- 3. Isotropy

For relation, we assume the goal is somewhat linear(?)