

word embedding.

$V = [a, aaron, \dots, zulu, [UNK]]$ $|V| = 1000$

1-hot representation

Man Woman

$$\begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix} \quad \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

one-hot representation each word as itself.

not able to generalize structures.

The inner product of any two one-hot encoded entities are zero.

Objective: featurized representation, instead of one-hot.

Featurized representation

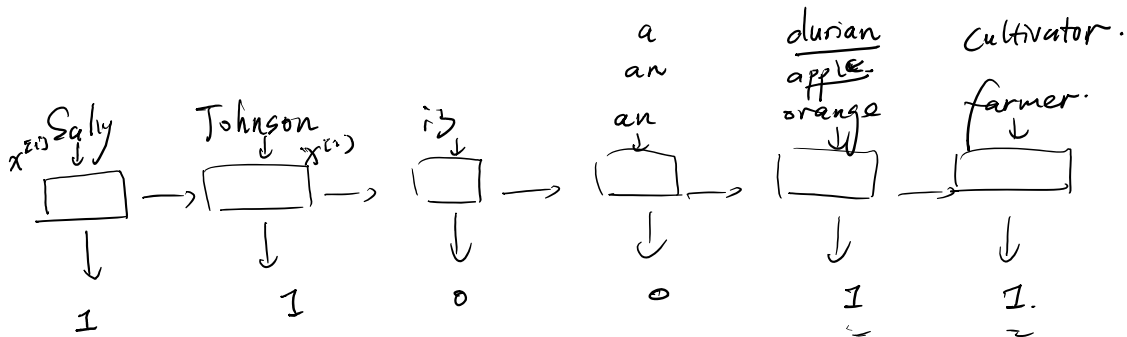
	word 1	w2	w3		
	man	women	king	queen	apple
gender	-1	1	-0.95	0.97	0
royal					
age					
size					

一些 dimension 但实际上 word embedding 是隐含的 dimension 对吧。

WORD EMBEDDING.

Using word embedding

Named Entity Recognition. example



if you use embedding and your dataset is small, you can use word embedding to generalize from orange farmer to apple farmer to durian cultivator.

- Allows you to carry out transfer learning.
learn embedding from a larger corpus / download embedded words ~~and~~ online

A ton of data from A \rightarrow small data from B.

less useful for machine translation

- Similar to face encoding

different - face: can input face pic you've never seen
- language: have a fixed vocabulary.

Property of word embedding

Analogy

Man \rightarrow Woman as King \rightarrow ?

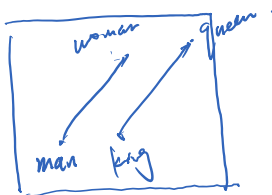
$$\boxed{e_{\text{man}} - e_{\text{woman}}} \approx \begin{bmatrix} \cdot \\ \cdot \\ \cdot \\ \cdot \end{bmatrix}$$

$$\boxed{e_{\text{king}} - e_{\text{queen}}} \approx \begin{bmatrix} \cdot \\ \cdot \\ \cdot \\ \cdot \end{bmatrix}$$

Analogy:

remarkable and surprising results about word embedding.

$e_{\text{man}} - e_{\text{woman}} \approx e_{\text{king}} - e_{\text{?}}$ want to know one word of the analogy.



Find word w .

$$\arg \max_w \text{sim}(e_w, e_{\text{king}} - (e_{\text{man}} - e_{\text{woman}}))$$

t-SNE mapping

是把高维 word vector 映射到低维.

300D $\xrightarrow{\text{non-linear}}$ 2D complicated non-linear mapping.

can be.

Cosine similarity

$$\text{sim}(u, v) = \frac{u^T v}{\|u\|_2 \|v\|_2}$$

used often

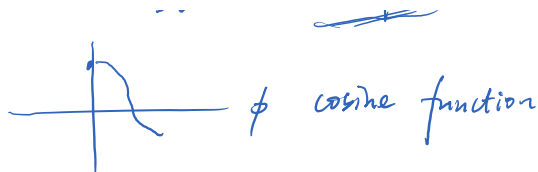


Cosine distance

较常用

$1 - \text{cosine function}$

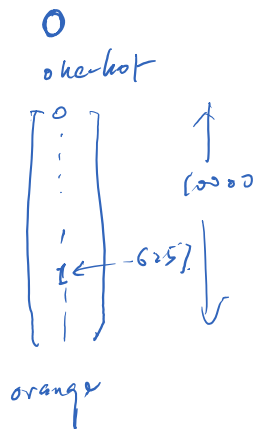
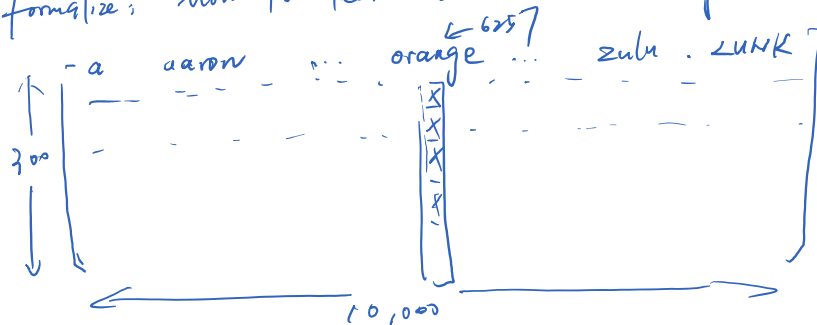
$$\text{sim}(u, v) = \frac{u^T v}{\|u\|_2 \|v\|_2}$$



Can also use square distance / distance.

Next: How to learn word embedding

Formalize: how to learn word embedding.



把第 j 个
word embedding
vector

$$E \cdot \begin{matrix} 0 \\ \vdots \\ 1 \end{matrix} = e_{625}$$

$$E \cdot \begin{matrix} 0 \\ \vdots \\ 1 \end{matrix} = e_j$$

embedding of the j -th word in the vocab.

In practice use a look up function, instead of doing matrix multiplication (slow)