Embeddings (1)

Neural Machine Translation

Traditional

- One-Hot-Encoding

sen1 = ‘Deep Learning is interesting for science and economy.‘

sen2 = ‘The media claims Deep Learning is interesting.‘

11 words -> 11 dimensional vectors

(and, claims, Deep, ecomomy, for, interesting, is, Learning, media, science, The)

‘Deep‘ = (0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0)

Traditional

- Bag-of-Words

sen1 = ‘Deep Learning is interesting for science and economy.‘

sen2 = ‘The media claims Deep Learning is interesting.‘

BOW(sen1) = (1, 0, 1, 1, 1, 1, 1, 1, 0, 1, 0)

BOW(sen2) = (0, 1, 1, 0, 0, 1, 1, 1, 1, 0, 1)

D = sen1 + sen2

BOW(D) = (1, 1, 2, 1, 1, 2, 2, 2, 1, 1, 1)

Traditional

- Counter-Based
  - TF $tf(t,d) = f_{t,d}/N$ $N: Number\ of\ words\ in\ d$
  - TF-IDF $idf(t,D) = \log(\#D/\#d_t)$ $\#D: Number\ of\ Documents$

    $\#d_t: Number\ of\ Docs$
    $including\ term\ t$

sen1 = ‘Deep Learning is interesting for science and economy.‘

sen2 = ‘The media claims Deep Learning is interesting.‘

TF(‘Deep‘, sen1) = 1/8

TF(‘Deep‘, sen2) = 1/7

IDF(‘Deep‘,D) = log(2/2) = 0

Traditional

- Counter-Based
  - TF $tf(t,d) = f_{t,d}/N$ $N\colon Number\ of\ words\ in\ d$
  - TF-IDF $idf(t,D) = \log(\#D/\#d_t)$ $\#D\colon Number\ of\ Documents$

$\#d_t\colon Number\ of\ Docs$
$including\ term\ t$

sen1 = ‘Deep Learning is interesting for science and economy.‘

sen2 = ‘The media claims Deep Learning is interesting.‘

TF(‘science‘, sen1) = 1/8

TF(‘science‘, sen2) = 0

IDF(‘science‘,D) = log(2/1) = 0.301 -> TF-IDF(‘science‘, sen2, D)=0.037

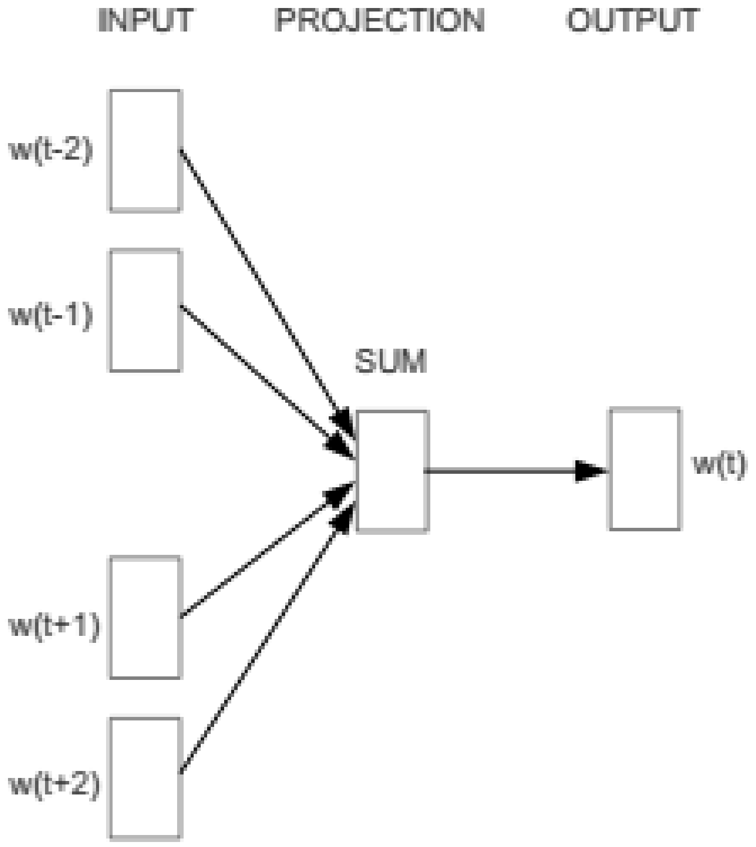
Modern

- Word2Vec
  - CBOW:

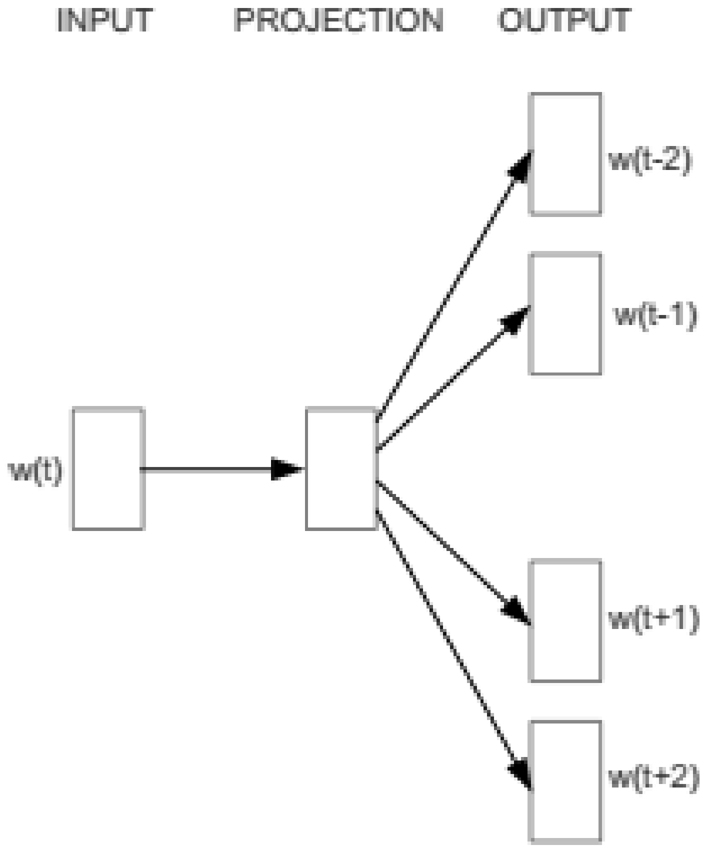
    Predict current word from bag-of-words of surrounding words
  - Skip-Gram:

    Predict context words given the current word

Modern



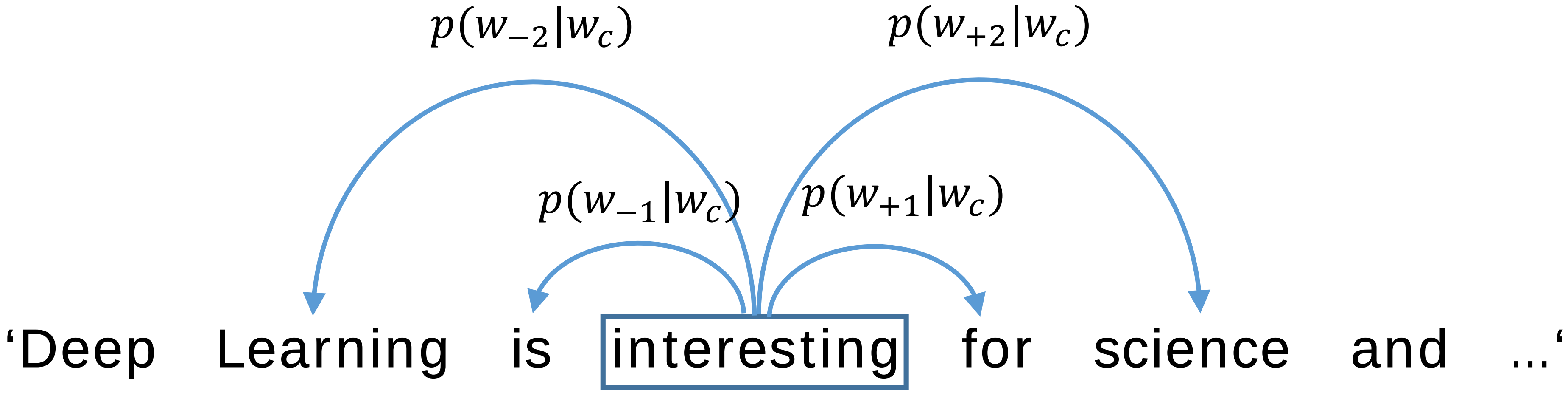
CBOW



Skip-gram

Modern

- Word2Vec
  - Minimize: $\log\sigma(u_o^T v_c) + \sum\left[\log\sigma\left(-u_j^T v_c\right)\right]$



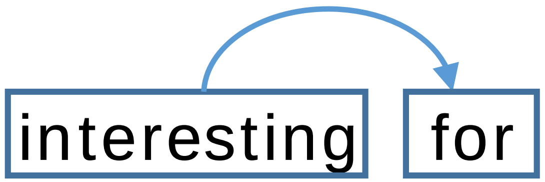
$p(w_{-2}|w_c)$ $p(w_{+2}|w_c)$

$p(w_{-1}|w_c)$ $p(w_{+1}|w_c)$

‘Deep Learning is interesting for science and ...‘

Modern

- Glove
  - Minimize: $\sum_{i,j=1}^{W} f\big(count(i,j)\big)\big(u_i^T v_j - \log count(i,j)\big)^2$



Count cooccurance

‘Deep Learning is interesting for science and …‘