Lecture 3：Advanced Word2Vec Methods

2018年7月4日 星期三　　下午11:58

SGD

- Very Sparse: #words in window = 2m+1
- Only have to update Specific vectors

1.Negative sampling in skip-gram:

随机采样，是否会采到上下文词语或与中心词很相关的词？

Loss Function:

$$J_t(\theta) = log\ \sigma(u_o^T v_c) + \sum_{j \sim P(w)} [log\ \sigma(-u_j^T v_c)]$$

- Randomly choose some words, reduce computation
- **Maximize** the co-occurrence of **real** context words
- **Minimize** the probability of **random** words (negative samples)

2.Window based co-occurrence matrix

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

- Weakness:
  - Very large dimension
  - Word meaning is not fixed
- Solution: **SVD**
  - **Advantage:**
    - Best Rank-k approximation
    - Fast training in small corpus
  - **Disadvantage:**
    - Time cost O($mn^\uparrow 2$), restricted in small data
    - Hard to incorporate new words
    - Without information about word similarity
    - Extremely large weight on large counts (the\a\is)

Interesting patterns

- (Teach - Teacher); (Swim - Swimmer) seems to have similar distance & direction
- Grammar induction
- Words with the same root cluster together.

3.GloVe:

Combine the best of **Count based** and **Direct prediction** model.

$$J(\theta) = \frac{1}{2} \sum_{i,j=1}^{W} f(P_{ij})(u_i^T v_j - \log P_{ij})^2$$

让 $<u_i, v_j>$ 逼近 $\log P_{ij}$

其核心是否仍是统计模型？

引入 J(θ) 可用SGD加速，然而它仍需计算共现矩阵，且无法解决加入新语料后重新计算 P_{ij} 的问题...

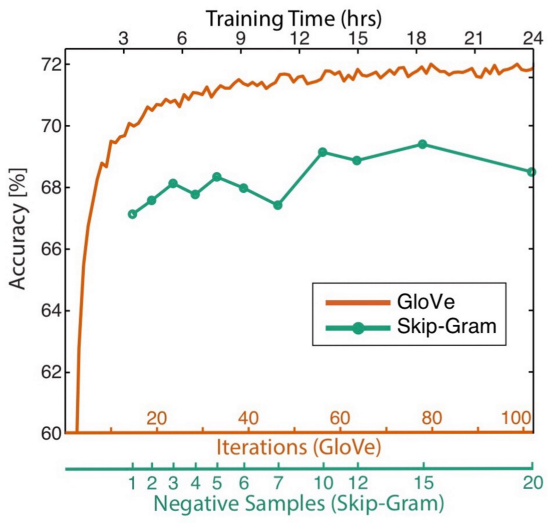
Not scalable?

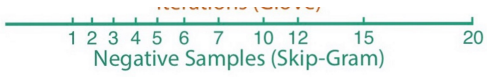
where: $P_{ij} = \frac{X_{ij}}{X_i} = \frac{X_{ij}}{\sum_k X_{ik}}$

$f(x) = \max(P_{ij}, 1)$　防止常见词权重过大.

4.Evaluation of word vector:

- Intrinsic:
  - Intermediate subtask
  - What parameter give improvement (change one at a time)
  - Better understanding the system
  - **Word Vector Analogies**
    - Semantic: Queen King : Man Woman
    - Syntactic: Short shorter shortest : large larger largest
    - Paris - France + Italy = Rome
- Extrinsic:
  - Apply in real task.
  - Time Consuming





可否举例我们在什么情形下使用何种评估方式？