Global Vectors for Word Representation (GloVe)

Haeri Kim

February 19, 2025

Outline

Introduction

- There are two main ways to learn word embeddings:
  - **Global matrix factorization methods** (e.g., LSA) that use all corpus statistics
  - **Local context window methods** (e.g., Word2Vec) that look at nearby words
- Both approaches have strengths but also some weaknesses.
- **GloVe** combines global co-occurrence information with local context ideas.
- It produces word vectors that capture useful linear patterns (e.g., analogies).

Related Work: LSA

- **Latent Semantic Analysis (LSA)** uses singular value decomposition (SVD) on a term-document matrix.
- It captures general (global) patterns in text.
- **Limitations:**
  - Sometimes the vectors are not fine-grained enough for certain tasks.
  - It does not explicitly use word order or local context.

Related Work: Word2Vec

- **Word2Vec** (skip-gram, CBOW) uses a simple neural network to predict words or contexts.
- **Advantages:**
  - Fast training on large data
  - Learns useful linear relationships
- **Limitations:**
  - Uses only local context
  - May miss some global statistics

The GloVe Model: Co-occurrence Probabilities

- GloVe focuses on the idea that **ratios** of co-occurrence probabilities can reveal word meaning.
- Example: comparing *ice* and *steam* with words like *solid* or *gas*.
- These ratios help the model learn meaningful dimensions for word vectors.

Table 1: Co-occurrence Probabilities

	Context Word			
Target Word	solid	gas	water	fashion
$P(k\|\text{ice})$	1.9×10^{-4}	6.6×10^{-5}	3.0×10^{-3}	1.7×10^{-5}
$P(k\|\text{steam})$	2.2×10^{-5}	7.8×10^{-4}	2.2×10^{-3}	1.8×10^{-5}
Ratio $P(k\|\text{ice})/P(k\|\text{steam})$	8.9	0.085	1.36	0.96

tableSample co-occurrence probabilities for *ice* and *steam* with different context words.

GloVe Model: Loss Function

- GloVe aims to learn vectors $w_i, \tilde{w}_j$ such that:

$$w_i^T \tilde{w}_j + b_i + \tilde{b}_j \approx \log(X_{ij}),$$

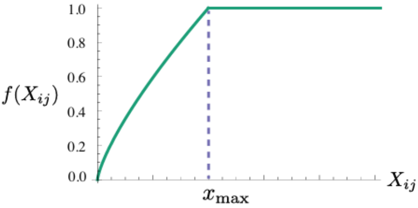
  where X_{ij} is how often words i and j co-occur.
- We use a weighted least squares approach to handle large differences in X_{ij}.

GloVe Model: Loss Function

- The overall cost function is:

$$J = \sum_{i,j} f(X_{ij}) \left(w_i^T \tilde{w}_j + b_i + \tilde{b}_j - \log(X_{ij}) \right)^2 .$$

- $f(X_{ij})$ is a weighting function that:
    - Is zero if $X_{ij} = 0$
    - Grows slowly for very frequent pairs



Experiments

- **Tasks:**
  - Word analogy tests
  - Word similarity benchmarks
  - Named Entity Recognition (NER)
- **Main results:**
  - GloVe often beats Word2Vec and SVD methods on analogy tasks
  - GloVe does well on word similarity
  - Adding GloVe vectors improves NER scores
- Different **corpus sizes**, **vector dimensions**, and **context windows** all affect performance.

Model	Dim.	Size	Sem.	Syn.	Tot.
ivLBL	100	1.5B	55.9	50.1	53.2
HPCA	100	1.6B	4.2	16.4	10.8
GloVe	100	1.6B	67.5	54.3	60.3
SG	300	1B	61	61	61
CBOW	300	1.6B	16.1	52.6	36.1
vLBL	300	1.5B	54.2	64.8	60.0
ivLBL	300	1.5B	65.2	63.0	64.0
GloVe	300	1.6B	80.8	61.5	70.3
SVD	300	6B	6.3	8.1	7.3
SVD-S	300	6B	36.7	46.6	42.1
SVD-L	300	6B	56.6	63.0	60.1
CBOW†	300	6B	63.6	67.4	65.7
SG†	300	6B	73.0	66.0	69.1
GloVe	300	6B	77.4	67.0	71.7
CBOW	1000	6B	57.3	68.9	63.7
SG	1000	6B	66.1	65.1	65.6
SVD-L	300	42B	38.4	58.2	49.2
GloVe	300	42B	**81.9**	**69.3**	**75.0**

Model	Size	WS353	MC	RG	SCWS	RW
SVD	6B	35.3	35.1	42.5	38.3	25.6
SVD-S	6B	56.5	71.5	71.0	53.6	34.7
SVD-L	6B	65.7	72.7	75.1	56.5	37.0
CBOW†	6B	57.2	65.6	68.2	57.0	32.5
SG†	6B	62.8	65.2	69.7	58.1	37.2
GloVe	6B	65.8	72.7	77.8	53.9	38.1
SVD-L	42B	74.0	76.4	74.1	58.3	39.9
GloVe	42B	**75.9**	**83.6**	**82.9**	**59.6**	**47.8**
CBOW*	100B	68.4	79.6	75.4	59.4	45.5

Model	Dev	Test	ACE	MUC7
Discrete	91.0	85.4	77.4	73.4
SVD	90.8	85.7	77.3	73.7
SVD-S	91.0	85.5	77.6	74.3
SVD-L	90.5	84.8	73.6	71.5
HPCA	92.6	**88.7**	81.7	80.7
HSMN	90.5	85.7	78.7	74.7
CW	92.2	87.4	81.7	80.2
CBOW	93.1	88.2	82.2	81.1
GloVe	**93.2**	88.3	**82.9**	**82.2**

Conclusion

- **GloVe** is a global log-bilinear model that uses word co-occurrence statistics.
- It balances both global counts and local context to learn better word vectors.
- The vectors show strong performance on analogies, similarity, and NER.
- GloVe helps unify count-based and prediction-based approaches in NLP.

Thank You!