Memoryless Document Vector

Dongxu Zhang

Advised by Dong Wang

2016.1.18

Introduction

- What is “Memory”?
- Why do we want “Memoryless”?
- How to achieve that goal?

Introduction

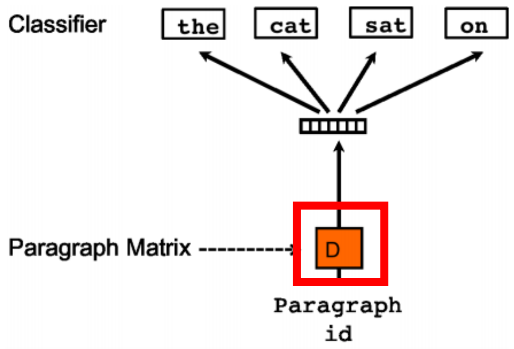
- What is “Memory”?
  - Latent Semantic Indexing(LSI) [1]

$$X = TS\boxed{D}^{T} \rightarrow d = xTS^{-1}$$

  - Probabilistic Latent Semantic Indexing(PLSI)[2]

$$P(w_j|d_i) = \sum_{k=1}^{K} P(w_j|z_k)\boxed{P(z_k|d_i)}$$

  - Paragraph Vector with Distributed Bag of Words(PV-DBOW)[3]



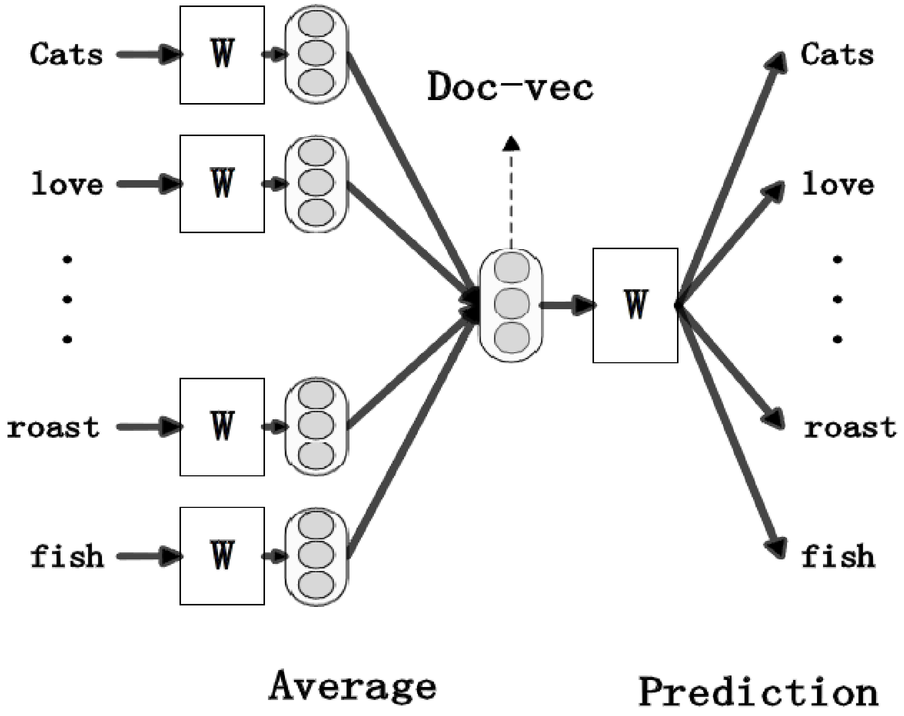
Introduction

- Why “Memoryless”?

Introduction

- How to achieve?
  - Word vector pooling[4]
    - Shortcoming: Does not involve pooling in model training, which leads to mismatch between word vector learning and document vector producing.
  - Memoryless Document Vector

Memoryless Document Vector



$$v^{(d)} = \Big(\sum_{i \in L^{(d)}} W_{d_i}\Big)/L^{(d)}$$

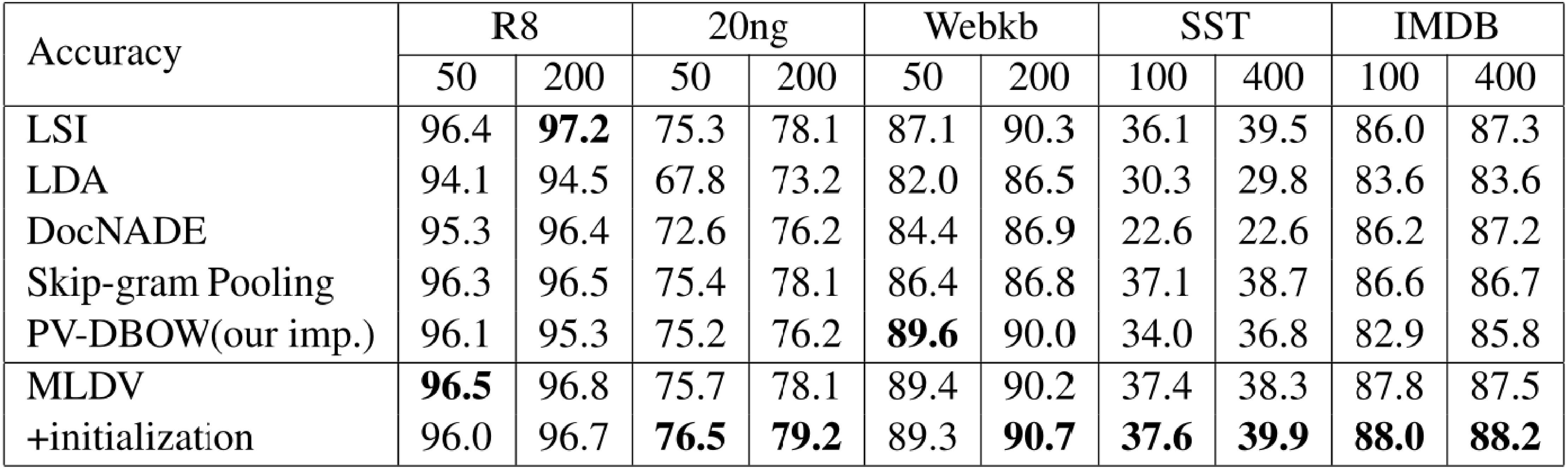
$$\mathcal{L}(D; W) = \sum_{d \in D} \sum_{i \in L^{(d)}} -\log \mathcal{P}(d_i | v^{(d)})$$

Experiments

- Datasets
    - Webkb, reuters 8, 20 newsgroup
    - SST, IMDB

- Setup
    - Topic classification tasks with 50 and 200 dimensions.
    - Sentiment classification tasks with 100 and 400 dimensions.
    - Logistic regression classifier.

Results



Accuracy	R8		20ng		Webkb		SST		IMDB	
	50	200	50	200	50	200	100	400	100	400
LSI	96.4	**97.2**	75.3	78.1	87.1	90.3	36.1	39.5	86.0	87.3
LDA	94.1	94.5	67.8	73.2	82.0	86.5	30.3	29.8	83.6	83.6
DocNADE	95.3	96.4	72.6	76.2	84.4	86.9	22.6	22.6	86.2	87.2
Skip-gram Pooling	96.3	96.5	75.4	78.1	86.4	86.8	37.1	38.7	86.6	86.7
PV-DBOW(our imp.)	96.1	95.3	75.2	76.2	**89.6**	90.0	34.0	36.8	82.9	85.8
MLDV	**96.5**	96.8	75.7	78.1	89.4	90.2	37.4	38.3	87.8	87.5
+initialization	96.0	96.7	**76.5**	**79.2**	89.3	**90.7**	**37.6**	**39.9**	**88.0**	**88.2**

Conclusion

Raise up the memory issue of conventional document representation methods and then propose a simple yet effective method for document representation to nail it.

Reference

- [1] S.C. Deerwester, S.T. Dumais, T.K. Landauer, G.W. Furnas, and R.A. Harshman. 1990. Indexing by latent semantic analysis. Journal of the American Society of Information Science, 41(6):391–407.
- [2] Hofmann, Thomas. "Probabilistic latent semantic indexing." *Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval*. ACM, 1999.
- [3] Quoc Le and Tomas Mikolov. 2014. Distributed representations of sentences and documents. In Proceedings of the 31st International Conference on Machine Learning (ICML-14), pages 1188–1196.
- [4] Chao Xing, Dong Wang, Xuewei Zhang, and Chao Liu. 2014. Document classification based on i-vector distributions. In APSIPA 2014.

Thank you.