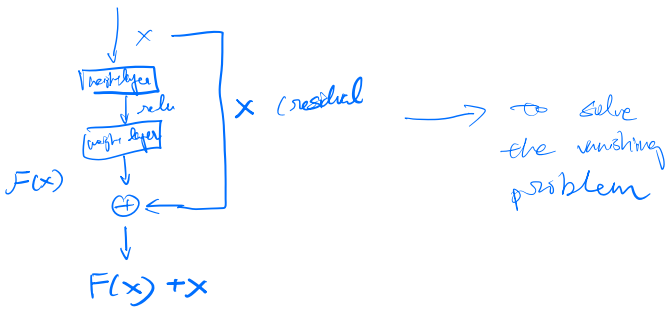
solve the problem

- deep convnet → gradient vanishing when back prop
- Residual Networks



→ to solve the vanishing problem

Natural Language Processing :

Glove

Word representation :

- Word Net
- one-hot vectors
  - vocab size fixed
  - sparsity
  - no natural notion of similarity.

---

word-coding matrix?

PPMI matrix → very sparse.

for sparsity : do SVD

minimize $\| M^{PPMI} - WV^T \|_F^2$ (Forbenious element-wise summation)

W · V

$w \times d$ $w \times r$ $d \times r$

→ reduudent info within the sparse matrix.

SVD : compression

Word2vec

---

on page 25

difference between two group of words, if similar relations, vector looks similar

---

problem: Bank of China v. river bank

↳ solution: contextual word embedding.

mission: text classification.

aRV → simplified of LSTM