Multi-label Classification of Chinese Resume with Word Embedding

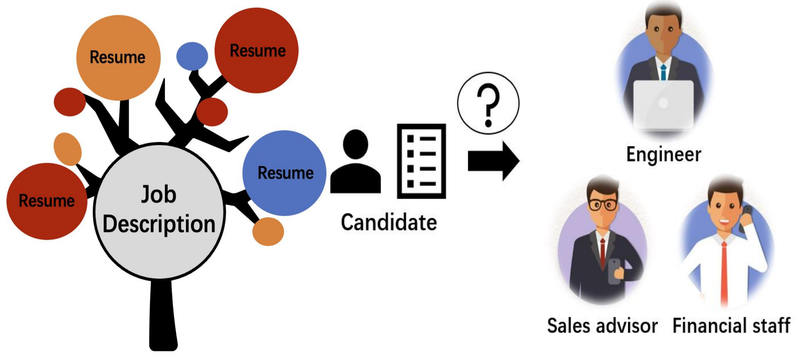
Citong Que, Jie Min and Zhengze Rong

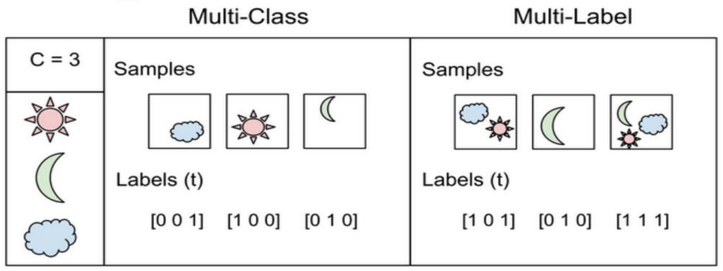
TBSI, Tsinghua University, China

ABSTRACT

- Construct a series of resume-field pairs based on the existing resume and job dataset.
- Compare different word embedding methods and dimensions with classification as the supervised task.
- Develop a bi-directional LSTM model for multi-label classification of large number of labels.

MOTIVATION



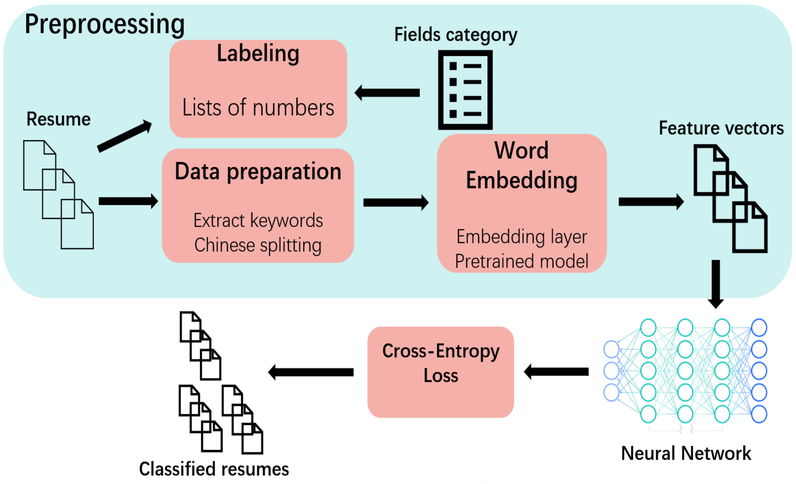


- Multi-label classification of resumes is the key to resume-job matching for employees.
- One candidate may be suitable for different fields if he/she has enough matching experiences.
- Control the dimension of word embedding to make classification more efficient.
- Traditional classifiers perform poor with the number of labels increases.

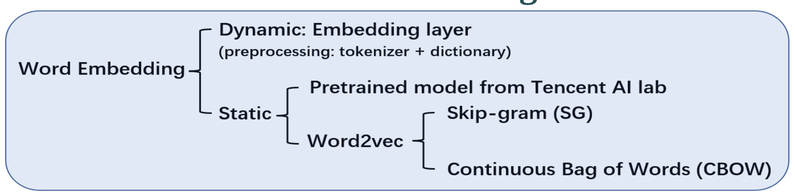
  → Deep Neural Network

METHOD

Pipeline

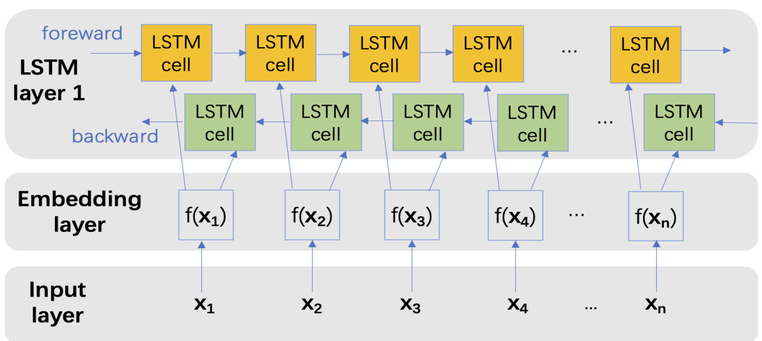


Word Embedding



- Dynamic: Word representations will be trained and updated during the training of model.
- Static: Words will be mapped to a fixed dimensional vector by pretrained model.

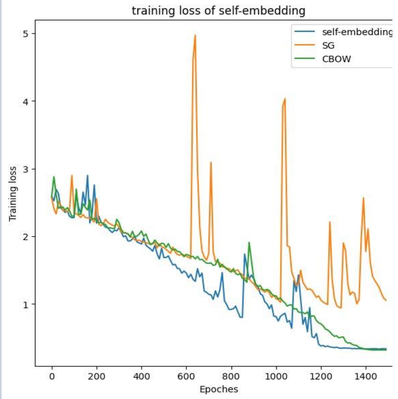
Bi-directional LSTM



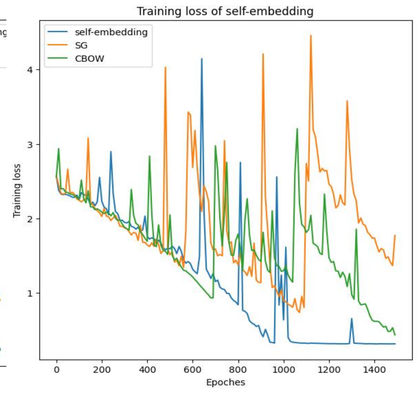
- Bi-directional LSTM will read each sentence in both positive and negative direction.

RESULT

embedding dimension=50



embedding dimension=100



Dimension	50			100		
Embedding	Self	SG	CBOW	Self	SG	CBOW
Accuracy	90.3%	68.9%	90.1%	89.7%	57.7%	86.3%

- Among word embedding methods, un-pretrained word embedding gives a more desirable training result.
- Multi-label classification achieves higher accuracy with lower embedding dimension.

CONCLUSION

- Un-pretrained word embedding is more suitable for small-sized corpus.
- Better word embeddings can be achieved while maximizing the learning objective.
- Bi-directional LSTM makes lower dimensional word embedding achieves even better result while improving efficiency.

REFERENCE

[1] Wang P , et al. A Unified Tagging Solution: Bidirectional LSTM Recurrent Neural Network with Word Embedding, 2015.

[2] Bdeir A M , Ibrahim F . A Framework for Arabic Tweets Multi-label Classification Using Word Embedding and Neural Networks Algorithms. BDE 2020: 2020 2nd International Conference on Big Data Engineering. 2020.