CSE881 PROJECT PRESENTATION

Zhongzheng Chen and Chang Liu

MODELS

- TFIDF and PCA/LDA
- Naïve Bayes Classifier
- DAN (Deep Averaging Networks)
- Bidirectional LSTM Classifier
- Ensemble method
  - Bagging

TFIDF AND PCA/LDA

Model

PCA/LDA and KNN classification

Training

5 folds random cross-validation set
1 fold for validation

Validation Result

Highest accuracy is about 20%

$$\mathrm{tf}(t,d) = 0.5 + 0.5 \cdot \frac{f_{t,d}}{\max\{f_{t',d} : t' \in d\}}$$

$$\mathrm{idf}(t,D) = \log \frac{N}{|\{d \in D : t \in d\}|}$$

$$\mathrm{tfidf}(t,d,D) = \mathrm{tf}(t,d) \cdot \mathrm{idf}(t,D)$$

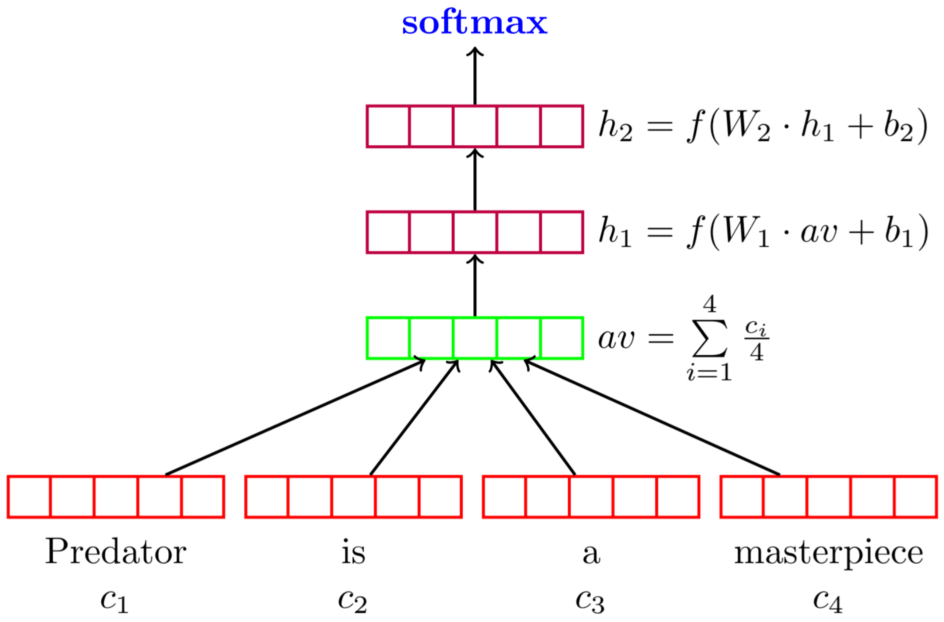
TFIDF Calculation Equation

NAÏVE BAYES CLASSIFIER

- Model
  - $P(class|document) = \frac{P(document|class)\,P(class)}{P(document)}$
  - Bag of Words assumption:
    - $P(document|class) = P(word_1, word_2, \cdots, word_n|class) = P(class) \prod_{i=1}^{n} P(word_i|class)$
  - $class = arg \max_{class} P(class|document)$
  - Laplace smoothing for unknown word probability allocation
- Training
  - 5 folds random partition of the training data
  - 1 of the 5 folds (2634 samples) is used for validation
- Result
  - Validation accuracy of 0.86

DAN (DEEP AVERAGING NETWORKS)

- A simple feed forward network
- Training
  - 5 folds random partition of the training data
  - 1 of the 5 folds (2634 samples) is used for validation
- Training setup:
  - Epoch: 20
  - Batch size: 50
  - Learning rate: 0.01
  - Optimizer: Adam
  - Activation function: ReLU
  - Embedding size: 150
  - Hidden layer size: 100
  - Number of hidden fully connected layer: 3



DAN (DEEP AVERAGING NETWORKS)

- Training results:
- [01/20] train: loss : 2.87, acc: 19.73% | Val: mean_loss : 2.75, acc: 33.81%
- [02/20] train: loss : 2.71, acc: 36.95% | Val: mean_loss : 2.76, acc: 31.53%
- [03/20] train: loss : 2.66, acc: 41.33% | Val: mean_loss : 2.66, acc: 40.93%
- [04/20] train: loss : 2.63, acc: 45.08% | Val: mean_loss : 2.65, acc: 43.35%
- [05/20] train: loss : 2.63, acc: 46.15% | Val: mean_loss : 2.62, acc: 46.51%
- [06/20] train: loss : 2.66, acc: 42.43% | Val: mean_loss : 2.69, acc: 40.28%
- [07/20] train: loss : 2.64, acc: 45.78% | Val: mean_loss : 2.66, acc: 43.58%
- [08/20] train: loss : 2.64, acc: 44.84% | Val: mean_loss : 2.69, acc: 40.19%
- [09/20] train: loss : 2.65, acc: 44.19% | Val: mean_loss : 2.65, acc: 44.47%
- [10/20] train: loss : 2.62, acc: 47.86% | Val: mean_loss : 2.74, acc: 35.35%
- [11/20] train: loss : 2.70, acc: 39.66% | Val: mean_loss : 2.68, acc: 41.49%
- [12/20] train: loss : 2.66, acc: 43.74% | Val: mean_loss : 2.65, acc: 44.74%
- [13/20] train: loss : 2.68, acc: 41.40% | Val: mean_loss : 2.68, acc: 40.70%
- [14/20] train: loss : 2.68, acc: 41.04% | Val: mean_loss : 2.68, acc: 41.49%
- [15/20] train: loss : 2.68, acc: 40.76% | Val: mean_loss : 2.79, acc: 29.95%
- [16/20] train: loss : 2.71, acc: 37.98% | Val: mean_loss : 2.81, acc: 28.09%
- [17/20] train: loss : 2.73, acc: 36.57% | Val: mean_loss : 2.72, acc: 37.21%
- [18/20] train: loss : 2.71, acc: 37.81% | Val: mean_loss : 2.66, acc: 42.93%
- [19/20] train: loss : 2.64, acc: 44.83% | Val: mean_loss : 2.69, acc: 39.53%
- [20/20] train: loss : 2.71, acc: 38.53% | Val: mean_loss : 2.76, acc: 33.44%

BLSTM

- Training
  - 5 folds random partition of the training data
  - 1 of the 5 folds (2634 samples) is used for validation
- Training setup:
  - Epoch: 20
  - Batch size: 50
  - Learning rate: 0.01
  - Optimizer: Adam
  - Activation function: ReLU
  - Embedding size: 100
  - LSTM hidden layer size: 100

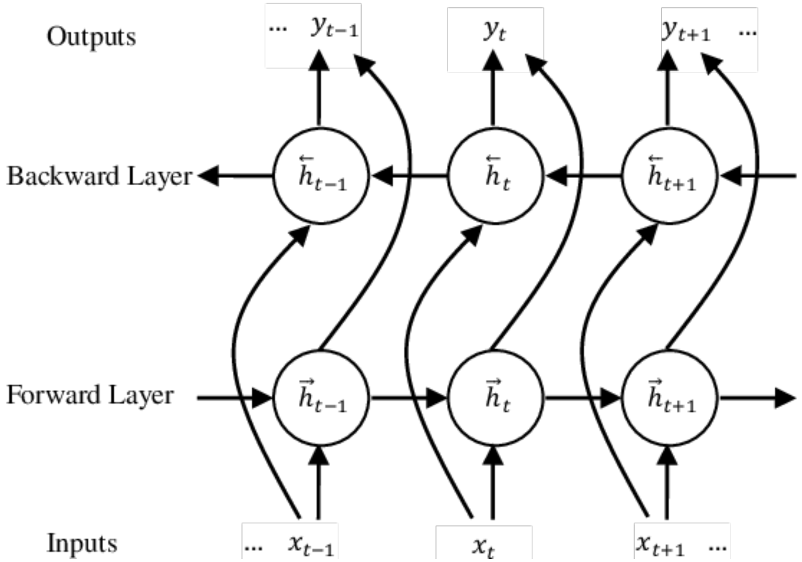


Image taken from https://www.researchgate.net/figure/Architecture-of-BLSTM_fig1_307889752

BLSTM

- Result

[01/20] train: loss : 2.86, acc: 21.38% | Validation: mean_loss : 2.72, acc: 36.37%

[02/20] train: loss : 2.54, acc: 56.60% | Validation: mean_loss : 2.45, acc: 63.86%

[03/20] train: loss : 2.32, acc: 77.74% | Validation: mean_loss : 2.35, acc: 74.51%

[04/20] train: loss : 2.19, acc: 89.75% | Validation: mean_loss : 2.28, acc: 80.09%

[05/20] train: loss : 2.14, acc: 94.73% | Validation: mean_loss : 2.25, acc: 83.21%

⋮

[16/20] train: loss : 2.09, acc: 99.10% | Validation: mean_loss : 2.23, acc: 85.16%

[17/20] train: loss : 2.09, acc: 99.17% | Validation: mean_loss : 2.23, acc: 85.21%

[18/20] train: loss : 2.09, acc: 99.17% | Validation: mean_loss : 2.23, acc: 85.21%

[19/20] train: loss : 2.09, acc: 99.24% | Validation: mean_loss : 2.22, acc: 85.44%

[20/20] train: loss : 2.09, acc: 99.18% | Validation: mean_loss : 2.23, acc: 85.16%

ENSEMBLE

- Bagging
  - 5 folds random partition of the training data
  - 1 of the 5 folds (2634 samples) is used for ensemble validation
  - Another 5 folds random partition of the remaining 4 folds
    - 4-fold (8429 samples) for model training
    - 1-fold (2107 samples) for training validation.

  - Naïve Bayes Classifier
    - No significant improvement
  - BLSTM
    - Around 6-7% accuracy improvement
  - Combination of both
    - Around 2-3% accuracy improvement with 33 BLSTM models and 20 Naïve Bayes models
    - Around 6-7% accuracy improvement with 48 BLSTM models and 3 Naïve Bayes models (our current model)

LESSONS LEARNED

- Naïve Bayes Classifier
  - Pro:
    - provides a solid baseline performance for document classification
  - Con:
    - the learned models are very consistent despite the randomness of training data partitioning and can't be used for bagging ensemble method.
- BLSTM
  - Pro:
    - good neural network model for document classification
    - learn different aspect from different initialization and data, which enable bagging ensemble for further improvement
  - Con:
    - requires more data for larger label size

REFERENCE

- Iyyer, H. (2015). Deep Unordered Composition Rivals Syntactic Methods for Text Classification. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)* (pp. 1681–1691). Association for Computational Linguistics.
- BLSTM image: https://www.researchgate.net/figure/Architecture-of-BLSTM_fig1_307889752

THANK YOU