Simple Embedings(GloVe) by Review Length and Label Frequency

1.Experiment Explanation

Using Glove and Simple Word Embedding-based Models finish 3 experiments:

1. the basic classification results
2. results by user generated text length (20% long & 80% short)
3. results by label frequency in the train set

2.Train/dev/test explanation

Train set is all the amazon review data

Validation/ Test sets are splited 0.5/0.5 from user generated needs data.

User generated needs dataset is shuffled before splitting, this gives better results.

3.Traning details

- Use MLP Classifier with 2 hidden layers, each hidden layer has 256 neurons (Many combination have been tried and this gives good enough results)
- Trained on the whole train set, take the weights out and finetuned by validation set, test on the test set
- optimizer:'Adam'
- Use early stopping techniques to find best model
- Save best model to disk

4.Results

1. **Numeric Results in 2 doc files:**

   A. simple embs(glove) by review length+frequency.doc：

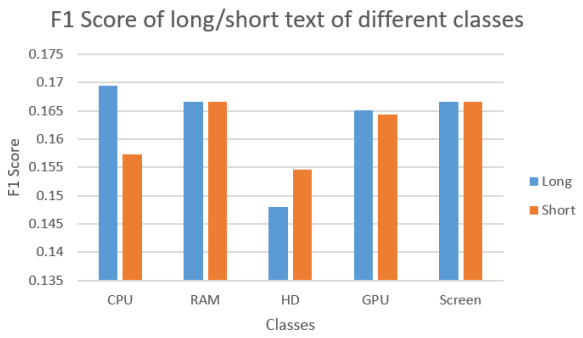
      - a. the basic classification results
      - b. results by user generated text length (20% long & 80% short)
      - c. results by label frequency in the train set (use top-5)

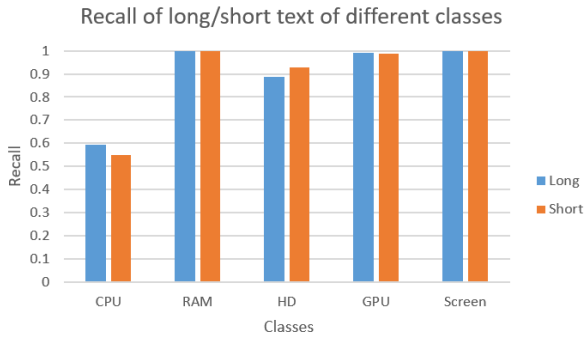
   B. new label frequency.doc：

- rerun the label freuqency experiment the use **top-3** Recall and F1
- change the group number of GPU,RAM,Sreen into **3 groups**

2. **Image results:**

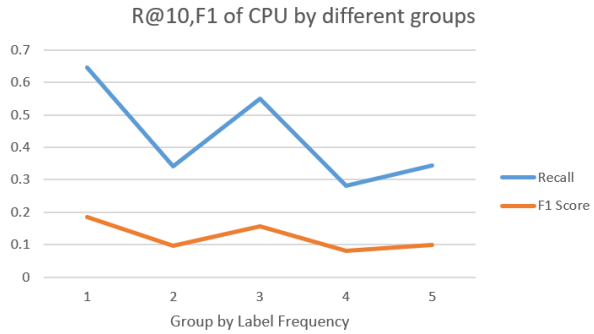
- Long/Short(by concat embedding method):

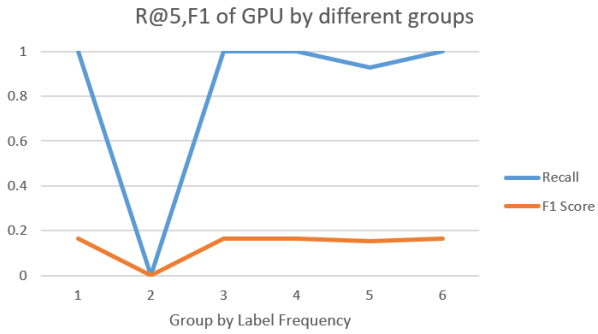


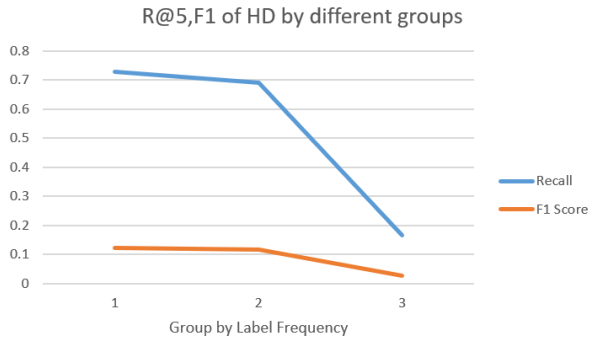


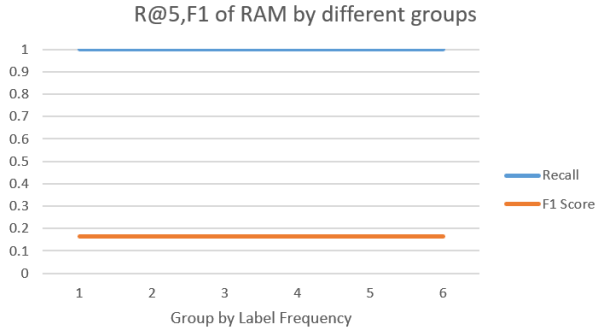
- Label Frequency(by concat embedding method):

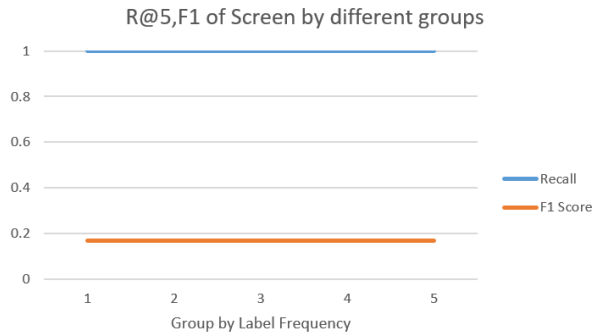
Labels are splitted into different groups based on ***number of labels*** and ***top 5*** in each class:





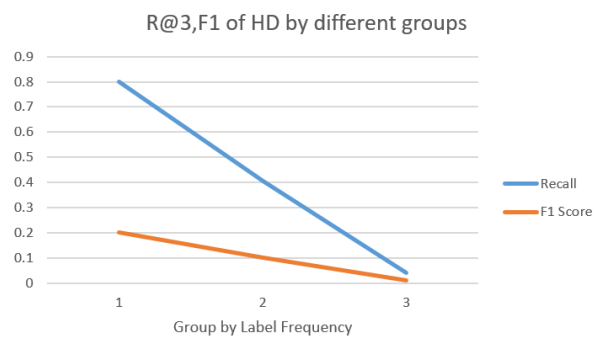


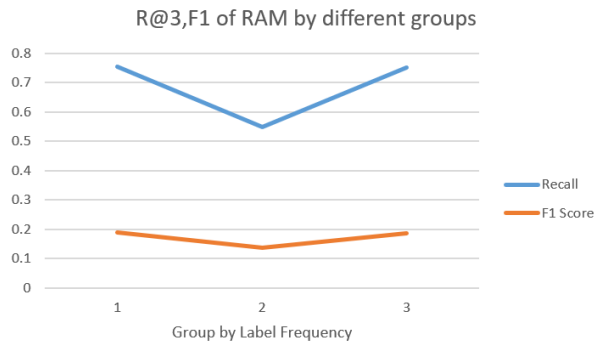


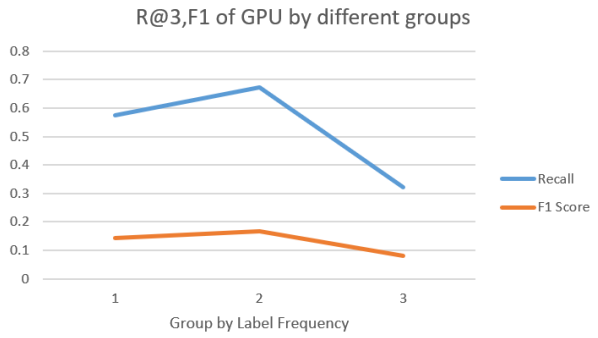


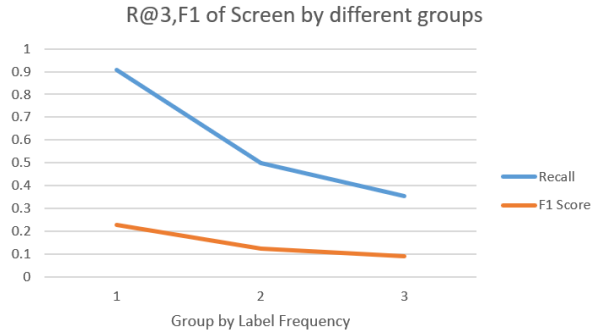
Label Frequency (***top3*** and ***group number=3*** for GPU,RAM,Sreen,HD)

- use hier embedding method here because it contains more non-zero value):









5. Reference

The Simple Word Embedding-based Models Refer to the paper [link (https://arxiv.org/abs/1805.09843)](https://arxiv.org/abs/1805.09843)