Author Prediction for Poetry

Katrin Schmidt
Immatriculation-no
@ims.uni-stuttgart.de

Carlotta Quensel
3546286
@ims.uni-stuttgart.de

Abstract

Same structure as the whole paper, but in short

1 Introduction

Short motivation and explanation of relevancy of your task, research questions/hypothesis

- what is author classification
- why poetry
- research question
  - Goal: find features inherent to poetry
    - is our goal possible
    - which features are good
  - Problem: style features might depend on medium (e.g. Limmerick) more than on the author
- Motivation: ??

2 Method

Description of your method (e.g. perceptron) without talking about the specific task too much. Explain features used (with or without being task specific), but do not judge them.

- Maximum Entropy classifier (explain why not other approaches)
  - short program description (training, classification)
- Features: MaxEnt/Bag-of-Words(/poetry specific?), learnFeatures (PMI)
- subsection with data/corpus creation from Poetry Foundation
  - which information is included
  - preprocessing steps (tokenizing)
  - statistics (number of poems/poets/poems per poet with graph)

3 Experiments

3.1 Experimental Design

Explain how you perform your experiments, which data is used, statistics of data.

- data statistics (decision for number of authors as hyperparameter)
- Train/Test split
- Program:
  - Hyperparameters: accuracy threshold, track loss & accuracy, #features/author
- Baseline (bag of words), Advanced: #verses, #stanzas, rhyme scheme

3.2 Results

Explain how your model performs, different models or configurations of your models.

- Feature combinations: baseline=BoW, advanced=all, other=?
- recall/precision/f_1 for all combinations (table)

3.3 Error Analysis

Given the configurations in the Results section, what are frequent sources of errors

- specifics and numbers about errors?
- overprediction of alphabetically first author
- many authors not predicted (uneven data distribution or bad features)
- feature weights converge similarly (no real weighting)

4 Summary & Conclusion

Explain and summarize your results on a more abstract level. What is good, what is not so good. What are the main contributions in your experiments?

5 Future Work

What did you have in mind what else your would have liked to experiment with? Other ideas?

- other models (e.g. Neural Net)
- other features (Topics from Poetry Foundation website)
- feature interdependencies/more data analysis
- genre interaction with author classification (multitask learning?)

A Contributions

Who implemented what? Who participated in the design of which components? Who wrote which part of the review?

B Declaration of Originality