

Singularity Systems Inc

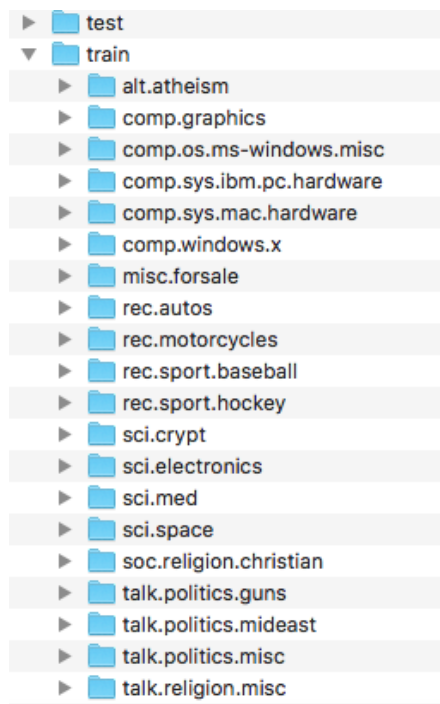
INSTRUCTION

Given the input data write a classifier to train the model on all the data in `train`, and evaluate the model's result with all the data under `test`. During the training, the model should NOT see any of the test data.

There are 20 classes for this dataset. Each class has many text documents. The goal is to use the training data to learn a text document classifier and apply this classifier on all the test data to have them properly labeled. Compared the model label of each test doc to the test gold label and compute the evaluation. Each sub dir's name is the gold label for all the text documents under it.

Input Data:

<https://drive.google.com/open?id=1WnvlSjnAbWkGfvPAPe-9GxIvCQnCvC2c>



You can use any tool/programming language to finish the classifier.



100 Overlook Center
2nd Floor
Princeton, NJ 08540,
USA

PHONE 732-343-0199
FAX
EMAIL info@singularitysystems.com
WEBSITE <http://www.singularitysystems.com>

We will evaluate your results based on:

- 1) How long does it take you to finish the project
- 2) The Evaluation Results (based on macro-average of all classes' F1 scores).
https://drive.google.com/file/d/1FSWE438DY1td5rLQIjGadg-m2fCZ0HS_/view?usp=sharing
- 3) The Source Code Review
- 4) Please write a report about what you find and how did you solve the problem.
And we will review your report's quality



100 Overlook Center
2nd Floor
Princeton, NJ 08540
USA

PHONE	732-343-0199
FAX	
EMAIL	info@singularitysystems.com
WEBSITE	http://www.singularitysystems.com