klassify

Text classification with Redis

Fatih Erikli
Software Developer
@Adphorus

fatiherikli@gmail.com
http://fatiherikli.com

Text Classification

The task is to assign a document to one or more classes or categories.

Use cases

- Spam filters
- Web page classification
- News and and topic categorization
- Sentiment Analysis

Techniques

- Neural Networks

- K-nearest neighbour algorithms

- Decision Trees

- Naive Bayes Classification

Training

Just count the words under a label.

Politics	
democrat	4
socialism	20
democrat	30
communism	20
politician	10
holacaust	20

Drugs	
smoke	30
weed	22
lsd	34
heroin	54
cannabis	23
marijuana	52

Classification

Stem, singularize and eliminate given text

Marijuana should be **legal**ized **national**ly in the **United States** just as it is already in **Colorado.**

Classification

Calculate scores for each labels

drugs: 4

drugs: 2

Marijuana should be **legal**ized **national**ly in the **United States** just as it is already in

politics: 3

Colorado.

Drugs = 4 + 2 = 6 / TotalWordCountOfDrugs
Politics = 3 / TotalWordCountOfPolitics

Redis Implementation

Initial labels

initial labels on a Set which is called `labels`

```
SADD labels Politics
SADD labels Drugs
SADD labels Game
SADD labels Programming
```

Querying for training

"Brainfuck! Brainfuck is awesome" as Programming

```
HINCRBY Programming Brainfuck 2

`awesome` and `is` should not be
trained. they are fuzzy words for
training logic.
```

Queries for classification

```
HVALS Programming
# 3 3 2 3 5

HGET Programming Brainfuck
# 5
```

A protoype:

http://github.com/fatiherikli/klassify

klassify

Overview
Train
Classify
Browse

21 trained labels

37 classifications

1368 words

Endpoints:

http://127.0.0.1:8888/train
http://127.0.0.1:8888/classify

Training

```
HTTP POST :8888/train

{
  'text': 'Brainfuck is awesome',
  'label': 'Programming'
}
```

Classification

request

```
HTTP POST :8888/classify

{
  'text': 'Brainfuck is awesome'
}
```

response

```
{
    "label": "Programming",
    "scores": {
        "Aliens": 1.428,
        "Animals": 2.380,
        "Society": 6.4935,
        "Technology": 9.523
    }
}
```

Response

```json
{
    "label": "Programming",
    "scores": {
        "Aliens": 1.4285714285714287e-05,
        "Animals": 2.380952380952381e-06,
        "Society": 6.493506493506494e-07,
        "Technology": 9.523809523809525e-07
    }
}
```

DEMO