

# Is There Any Similarity between Philosophy Schools? Can We Roughly Understand Philosophy by Algorithms?

Nichole Zhang uni:qz2446

## Part 1 Introduction

Philosophy is hard to understand for most of us who are not majoring in Philosophy. However, is it feasible for us to learn more about Philosophy with the help of algorithms? In this project, I use the dataset from Kaggle to find the similarities from different schools through different dimensions and try to extract main ideas from some schools' texts.

## Part 2 Data Processing for this Project

### 2.1 Import Data

For this project, I use the dataset called “**philosophy data**”, created by Kourosh Alizadeh. This dataset contains over 300,000 sentences from over 50 texts spanning 13 major schools of philosophy.

```
philo_data <- read.csv("../data/philosophy_data.csv")
```

### 2.2 The Packages Used In This Project

For this project, I use the following nine packages:

**ggplot2** - a system for declaratively creating graphics, based on The Grammar of Graphics

**gplots** - a collection of R programming tools for plotting data, including heatmap.2

**tidyr** - tools to help to create tidy data, where each column is a variable, each row is an observation, and each cell contains a single value

**syuzhet** - a package comes with four sentiment dictionaries and provides a method for accessing the robust, but computationally expensive, sentiment extraction tool developed in the NLP group at Stanford

**tm** - A framework for text mining applications within R

**wordcloud** - Functionality to create pretty word clouds, visualize differences and similarity between documents, and avoid over-plotting in scatter plots with text

**RColorBrewer** - a tool to choose sensible colour schemes for figures in R

**dplyr** - a grammar of data manipulation

**factoextra** - a package provides some easy-to-use functions to extract and visualize the output of multivariate data analyses

## 2.3 Some Basic Data Summaries for the data

The dataset contains 59 titles, 36 authors, 13 schools and 360780 sentences. The average length of sentences is 151.

```
length(unique(philo_data$title))
```

```
## [1] 59
```

```
length(unique(philo_data$author))
```

```
## [1] 36
```

```
length(unique(philo_data$school))
```

```
## [1] 13
```

```
length(unique(philo_data$sentence_lowered))
```

```
## [1] 360780
```

```
summary(philo_data$sentence_length)
```

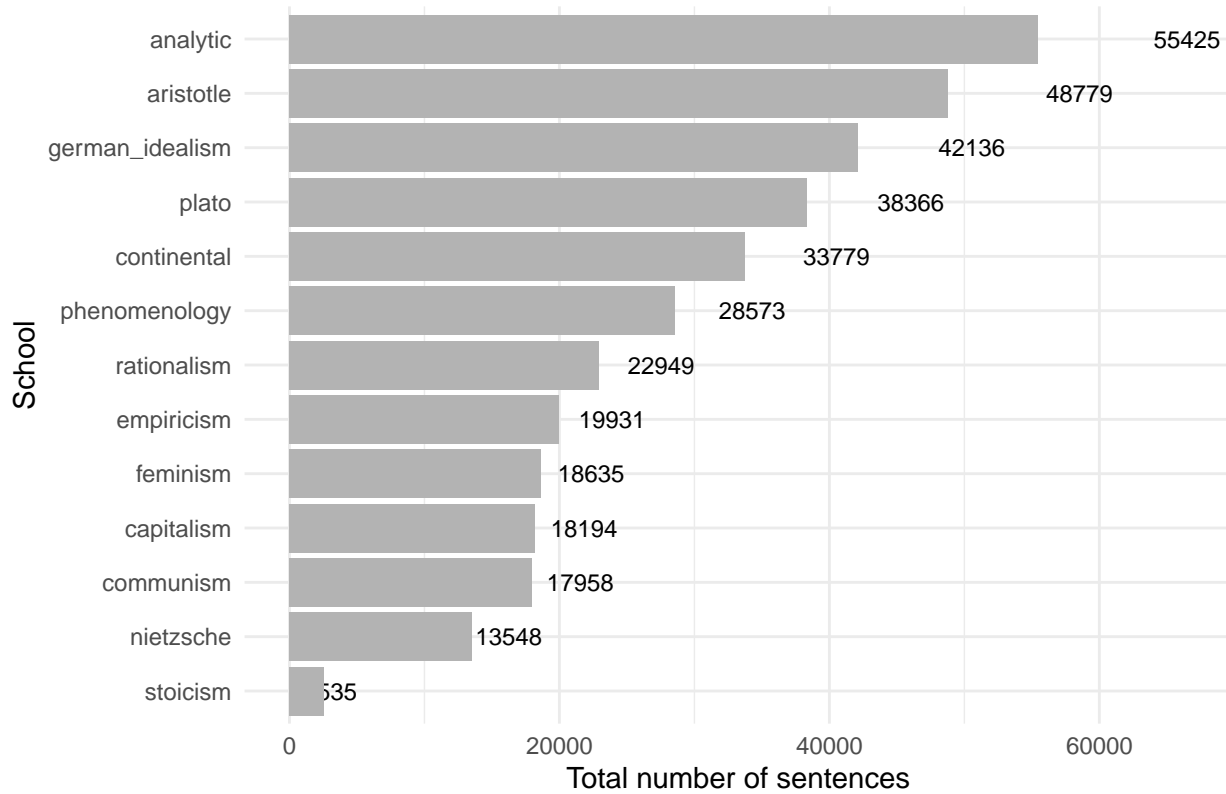
```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      20.0   75.0   127.0   150.8   199.0  2649.0
```

## 2.3 Limitation

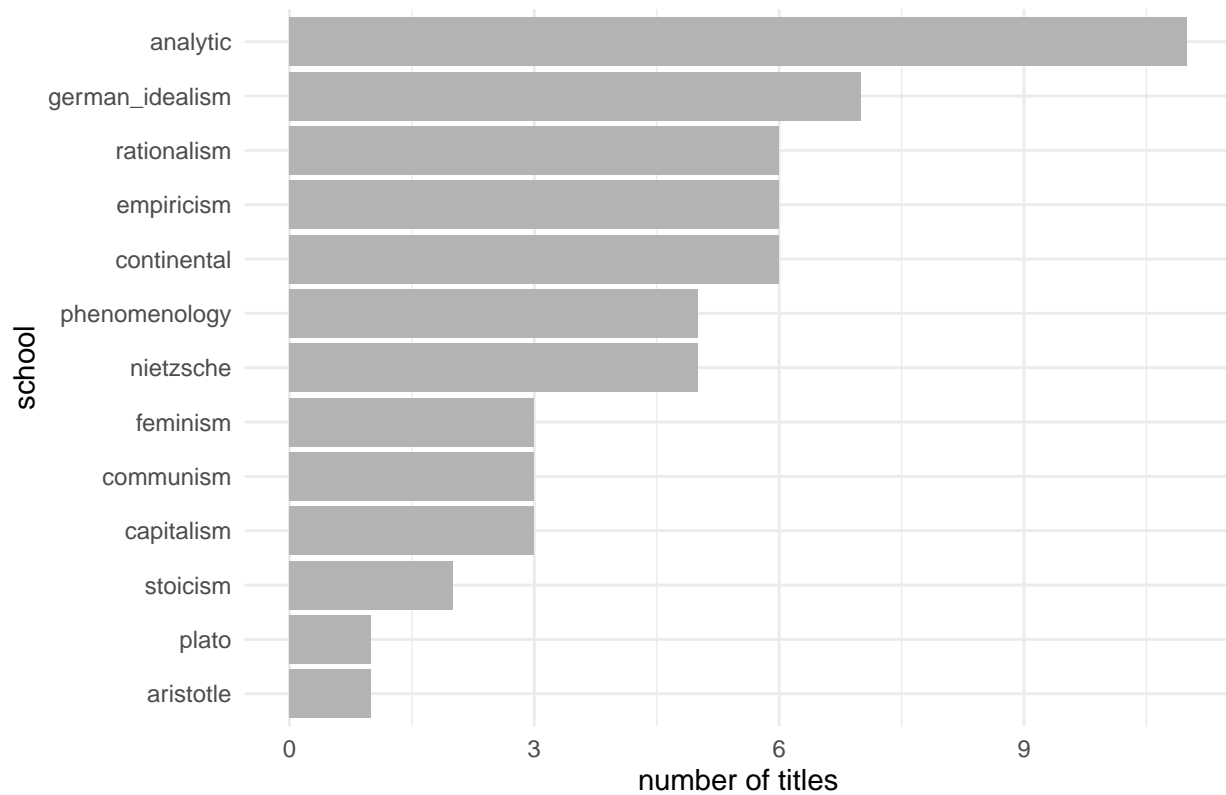
Most of the schools have more than 10,000 sentences in the dataset, however, Stoicism only has 535 sentences. It might be biased for concluding the main idea of stoicism by such a few sentences.

On average, there are only 5 books and 3 authors for each school, which means our conclusion for these schools may not reflect the true idea of these schools. We need more lines, if we want to understand their philosophy idea more accurately.

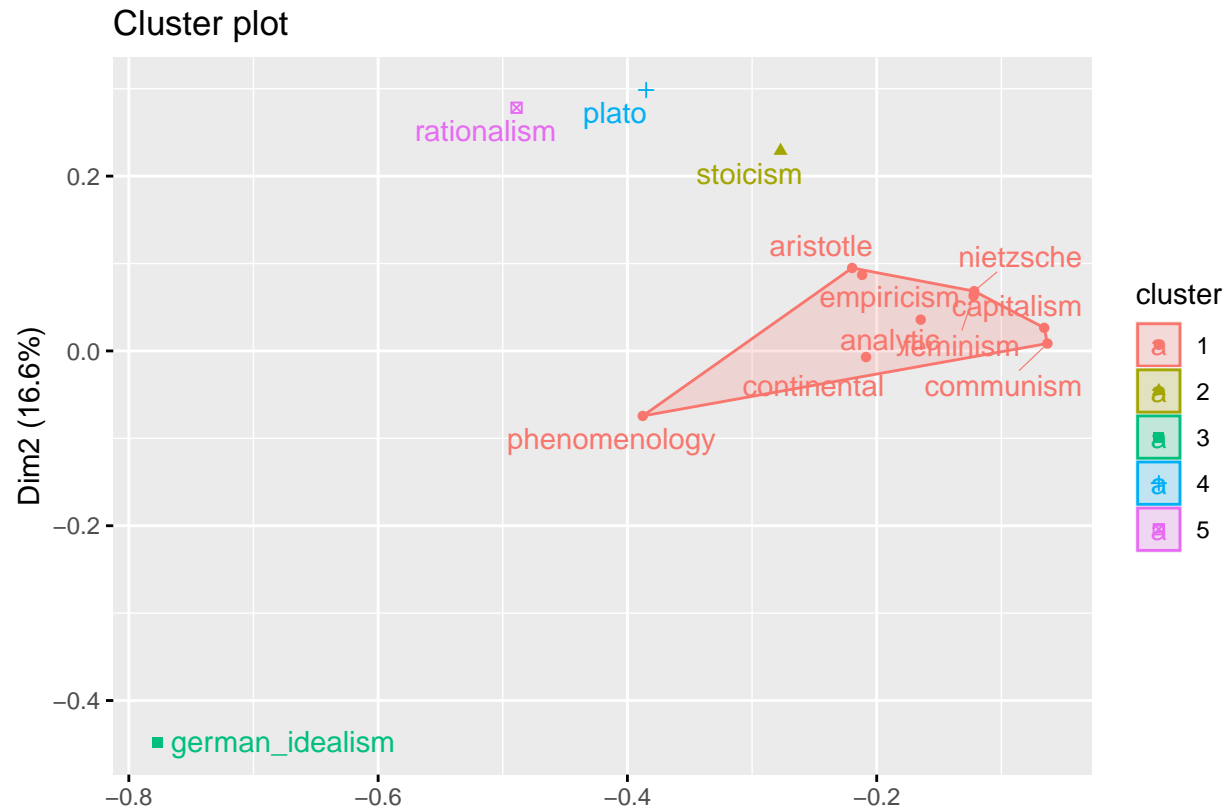
the number of sentences for each school



the number of titles for each school



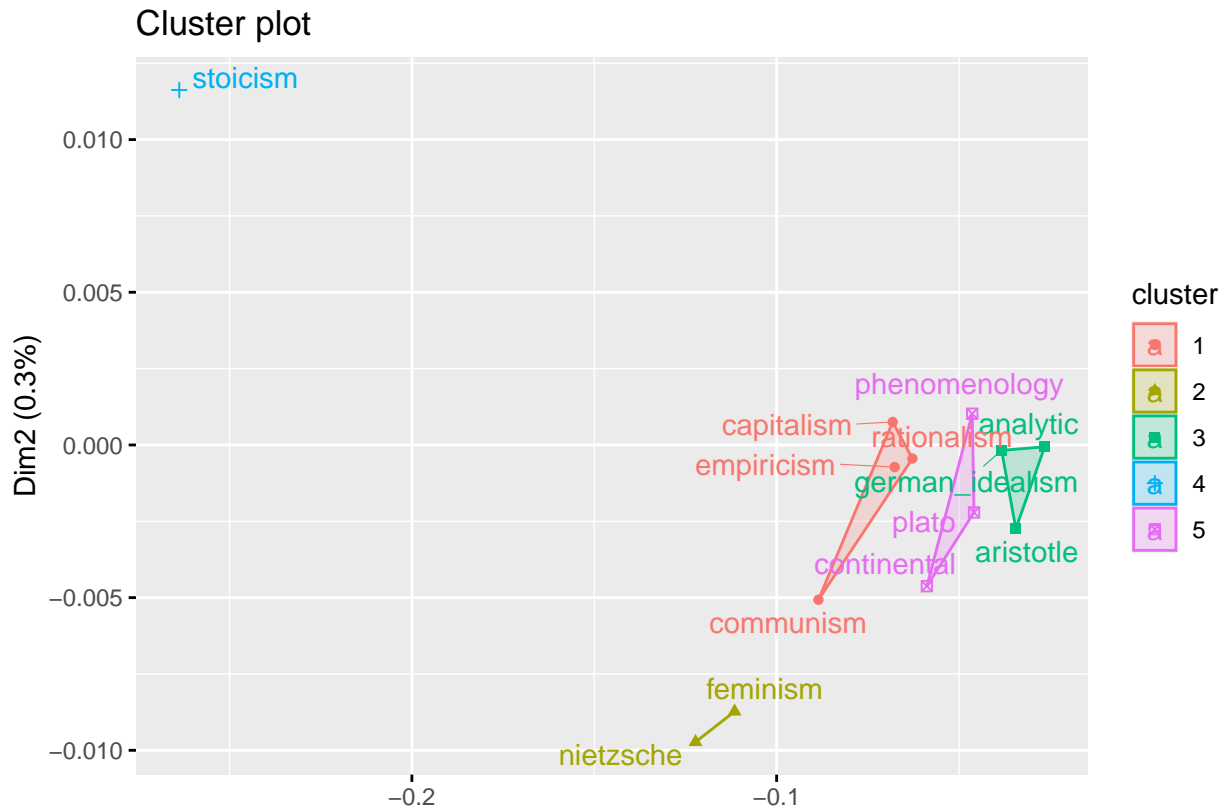




By k-mean clustering, we can figure out that those schools, **Aristotle**, **Nietzsche**, **Empiricism**, **Capitalism**, **Analytic**, **Feminism**, **Continental** and **Communism**, are more likely to discuss similar topics in texts. **Plato** and **rationalism** also focus on similar topics.

### 3.2.2 Which schools are writing with similar sentiments?

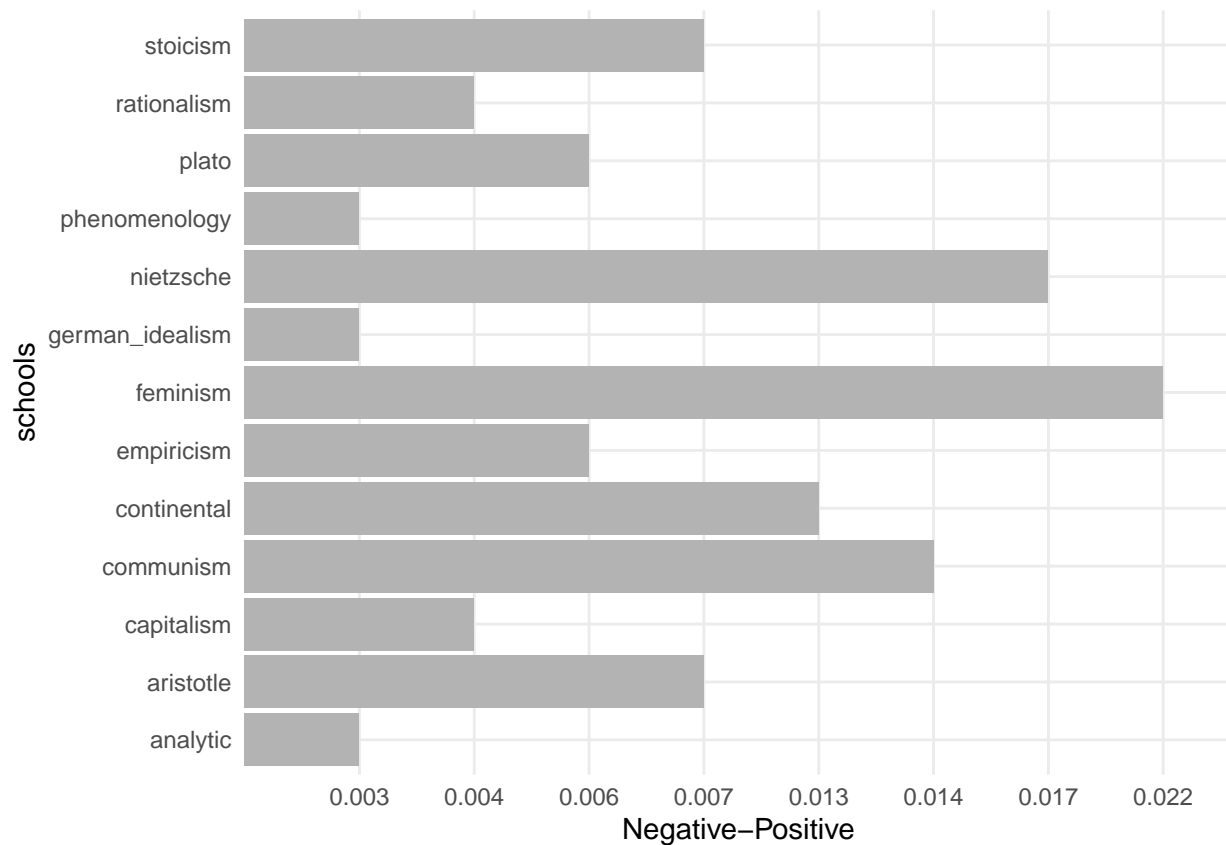
By analyzing sentiments for each word of all schools and clustering algorithm, we can find those schools writing with similar sentiments.



By the graph above, we can point out that four groups: {**Feminism & Nietzsche**}, {**Plato & Phenomenology & Continental**}, {**Capitalism & Rationalism & Communism & Empiricism**}, {**Aristotle & German Idealism**}. Each group's texts are written by similar sentiments.

### 3.3 Sentiment of each school is positive or negative?

We can use the sentiment dtm built before to do the analysis.



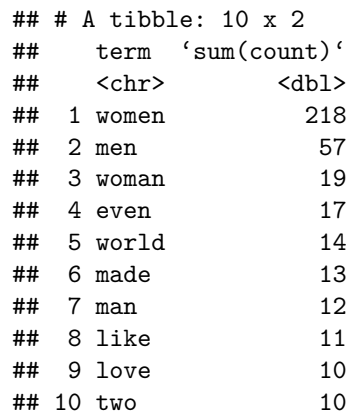
All schools' sentiments are negative. Feminism is the most negative one.

## Part 4 Extract Main Ideas for Some Schools from Texts

For most people, philosophy texts are obscure. Analyzing the dataset can help us understand the basic ideas about some schools. Two examples are given as follows.

### 4.1 Feminism

The main topic that Feminism is always discussing is **Woman**. Therefore, I extract all sentences containing **women are** from Feminism texts, and build a word cloud.



## 4.2 Rationalism and Empiricism

8



```
## # A tibble: 15 x 2
##   term                'sum(count)'  
##   <chr>                <dbl>  
## 1 knowledge            735  
## 2 god                  240  
## 3 mind                 135  
## 4 nature                89  
## 5 body                 83  
## 6 idea                 82  
## 7 know                 74  
## 8 far                  71  
## 9 good                 69  
## 10 true                69  
## 11 ideas               64  
## 12 love                59  
## 13 kind                58  
## 14 truth               58  
## 15 order               57
```

The high frequency words in **Rationalism** are “god”, “nature”, “love” and etc, which are more related to virtual side. In contrast, the high frequency words in **Empiricism** are “truth”, “real”, and etc, which are more focus on reality.

10

## Part 5 Reference

Kaggle - History of Philosophy: <https://www.kaggle.com/kouroshalizadeh/history-of-philosophy>

Wikipedia - Feminism: <https://en.wikipedia.org/wiki/Feminism>

Difference Between Rationalism and Empiricism: <https://askanydifference.com/difference-between-rationalism-and-empiricism-with-table/>