

Análise de Sentimento das Principais Obras da Filosofia Ocidental ao Longo da História

Equipe

Nome: Pensadores de Dados

Membros: Breno Moura de Abreu [1561286]

Repositório GitLab: <https://gitlab.com/breno-abreu/IntroCD>

Perguntas de Pesquisa:

1. Quais as palavras mais comuns para cada escola de pensamento?
2. Quais as categorias léxicas mais presentes em cada escola de pensamento?
3. Qual o sentimento predominante em cada escola de pensamento?
4. Quais são as conexões mais fortes entre cada escola e as demais?
5. Se separarmos as escolas em N grupos, como os grupos estariam separados?

Hipóteses

1. Seguintes hipóteses:
 - a. Para o feminismo a palavra mais comum será “mulher”.
 - b. Para o capitalismo as palavras mais comuns são relativas a dinheiro e trabalho.
 - c. Para Nietzsche os termos “Deus” e “Poder” estarão na lista de palavras mais comuns
2. Seguintes hipóteses:
 - a. Para Nietzsche, categorias relacionadas às artes estarão fortemente presentes na lista.
 - b. Para o idealismo alemão, categorias com uma natureza mais dogmática e retratando rigidez de pensamento estarão presentes.
3. Seguintes hipóteses:
 - a. A escola continental possui em geral sentimentos mais positivos

- b. A escola analítica possui em geral sentimentos mais negativos
4. As conexões mais fortes serão com as escolas de pensamento imediatamente anteriores cronologicamente à escola em análise.
5. Se agruparmos as escolas de acordo com o número de séculos presente na base de dados, os grupos formados serão separados por século.

Dados e Modelos

- Dados do corpus [History of Philosophy](#) (escola de pensamento, sentenças, data de publicação)
- NLTK: utilizado para a contagem das palavras mais utilizadas
- NLTK: utilizado para gerar um score para a análise de sentimentos
- Empath: escrever a lista de categorias que uma sentença pertence através da categorização léxica
- K Means: gerar N agrupamentos utilizando os dados anteriormente encontrados e vetorizados.

Cronograma:

1. Para cada escola de pensamento fazer a lista de palavras mais comuns;
2. Para cada escola de pensamento encontrar um valor que resuma o sentimento predominante;
3. Para cada escola de pensamento encontrar as principais categorias lexicais encontradas nas sentenças;
4. Encontrar uma forma de vetorizar os dados encontrados anteriormente. Uma possível solução é criar novas colunas, cada uma para cada uma das palavras, categorias léxicas e o sentimento geral; feito isso para cada escola é necessário inserir um valor numérico que representa o score encontrado para cada uma das novas colunas;
5. Para gerar as conexões, calcular para cada escola a distância para as demais. Utilizar as N escolas mais próximas;
6. Gerar um gráfico que demonstre as conexões;
7. Utilizando os dados vetorizados, encontrar os N grupos utilizando o K Means para as escolas. Utilizar vários valores de N;
8. Explorar os grupos resultantes.