Analysis of Philosophical Texts: Insights into Language, Sentiment, and Classification

GIULIO NEBBIAI
nebbiai16@gmail.com

RICCARDO VENDRAMINETTO
riccardo.vendraminetto10@gmail.com

ANTONIO MATTESCO
antonio.mattesco01@gmail.com

NICOLA PREGNOLATO
nicola.pregnolato@outlook.it

PAOLO FABRIS
paolofab01@gmail.com

1. Introduction

Language is the medium through which philosophers convey their ideas, explore complex concepts, and stimulate critical thinking, explained by texts. A linguistic analysis of these texts allows us to uncover how philosophers use words to rationally organize their most firmly-held beliefs. While it may not seem apparent at first glance, a data-driven analysis of philosophical texts holds nearly endless potential applications. Philosophy represents the deepest convictions of individuals, and its history is a record of attempts to systematically organize these beliefs. By integrating linguistic analysis with quantitative methods, we can gain a more profound understanding of how philosophers articulate and structure their theories, opening new avenues for exploring human thought.

1.1. Dataset and Preprocessing

Starting from a dataset obtained from 51 texts of **13** different **philosophical school** belonging to different time periods: from philosophy of Plato to the more recent feminist idealism. These texts were preprocessed according to the procedure described in*data load clean.ipynb*, in order to obtain the original dataset, that had 10 columns with 360808 total sentences. The dataset used to accomplish our tasks consider only 2 original columns ($school$, $sentence\ str$) and a new one ($tokenized\ sentences$), mainting the same number of total sentences. We **preprocessed** it **removing sentences that are too short** and **too long** and converted them to lower case. So we removed the stopwords and punctuation and then we applied the tokenizen process with Nltk method.

1.2. Objectives

Our goals in this project are:

1. **Word Usage:** Use Word2Vec models to explore how different philosophical schools employ language, focusing on key terms and their synonyms and contextuale words to uncover conceptual nuances.

2. **Sentiment Analysis:** Apply sentiment analysis techniques, including the iSA algorithm and BERT, to determine the sentiment distribution within philosophical texts and identify patterns or trends across different schools of thought.

3. **Classification:** Develop and evaluate a classification model to categorize texts according to their respective philosophical schools.

In summary, this project aims to provide an in-depth computational analysis of philosophical texts, leveraging modern techniques to reveal insights into language use, sentiment, and classification within this domain.

2. Word Usage

This section proposes the use of Word2Vec models for analysing the use of words by philosophers belonging to different schools of thought. The goal of this approach is to draw attention to how crucial language is for expressing and sharing philosophical ideas. Philosophy, by its very nature, makes use of a complex and articulated language in which words take on deep and nuanced meanings. Understanding how different philosophers use keywords in the context of their works allows one to grasp the peculiarities of their thought and the conceptual subtleties that distinguish them. The main objective is to identify the synonyms of certain key words and the words most related to them in order to better understand the context of philosophical thought. The choice of synonyms makes it possible to explore how a single concept can be multifaceted through a multiplicity of linguistic expressions; it also makes it possible to assess the linguistic variety within each philosophical school, highlighting the expressive richness. The analysis of related words makes it possible to identify recurring themes and conceptual connections that characterise each school of thought.

2.1. Word Frequency Analysis

To assess the linguistic variety is possible to analyze the unique words percentage for each school. As shown by the graph 1, it's clear that the greater the number of total words, the greater the number of unique words, but this is due to the unbalanced number of sentences per philosophical current. So, also considered unique words percentage, it is possible to notice that the more contemporary schools of thought have a higher value, perhaps due to the greater variety and complexity of topics covered; the exception being Stoicism, which has the bias of having a significantly lower total word count.

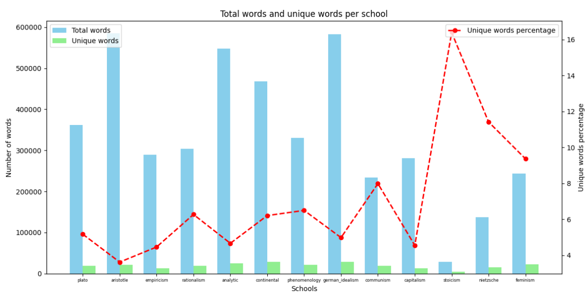

Figure 1. Analysis of unique words distribution in the schools

2.2. Models Description

In our study, we adopted two models of semantic embedding, **Word2Vec** and **GloVe** (Mohammed [1]), to address different text analysis needs within our research domain. Both models rely on the vector representation of words but differ in their implementation and the underlying theory guiding their functioning. The **Word2Vec model** is an embedding technique that captures semantic relationships between words by learning dense vectors for each word in the training corpus. We used the Word2Vec model to perform context analysis on our dataset of philosophical sentences, aiming to identify words most correlated with a given input term. The model was trained with a vector size of 50, a context window size of 5 (for **skip-gram**), and with a minimum word count value set to 1. This configuration was chosen to maximize the information captured from our limited dataset without overburdening the model with excessive detail. On the other hand, the GloVe (Global Vectors for Word Representation) model is based on the theory of word **co-occurrences** in a text corpus. GloVe combines both global and local information from the training corpus to generate vector representations of words. By utilizing the technique of factorizing a word co-occurrence matrix, GloVe learns dense vectors that capture both semantic and syntactic information. The optimization formulation of GloVe minimizes the difference between the dot product of word vectors and the logarithm of their co-occurrence frequency. This approach allows GloVe to capture both global and local information, resulting in word representations that are both semantic and syntactically informative. The model we used was trained on the extensive Wikipedia and Gigaword corpora, providing a rich dataset for learning word relationships. We employed the GloVe model to find synonyms of a given input word, first searching within the training corpus and then verifying the presence of synonyms in our dataset of philosophical sentences. The vector size was set to 50, considering the vastness of the training corpus compared to our specific data, in order to obtain sufficiently informative representations without excessive complexity. Additionally, to detect context words, we used three different methods to see different results: the first is "$most\ similar$" implemented in Word2Vec of Gensim, the second is a scalar product between the input word vector and other words' vectors, and the third is the use of cosine similarity. We experimented with the three methods to understand if there could be substantial differences in the results.

2.3. Interpretation of Results

In this subsections we show some results obtained from the model application to the real dataset and them interpretations.

School	Synonyms	Context words
Capitalism	City, Town, Central	Employ, Replace, Productive, Consisting, Maintainig
Communism	City, Town, Central, Region, Province	Possessor, Relative, Impoverishment, Millowners, Revolutionary

Table 1. Meaning of "**Capital**" for **capitalism school** and **communism school**

In synonym search, the model does not consider the context, whereas in context word search it does. In the latter, both Word2Vec's "$most\ similar$" method and cosine similarity are more precise and relevant compared to the dot product. This is because both rely on cosine similarity, a normalized and directional measure that better captures semantic relationships between words. The dot product, being non-normalized, can yield more general results influenced by vector lengths, limiting precision in semantic analysis. Thus, cosine similarity is more reliable for capturing semantic similarity between words, returning values between -1 and 1. In conclusion, adopting normalized measures like cosine similarity is crucial for achieving more accurate and meaningful results in analyzing semantic relationships between words.

3. Sentiment Analysis

The main objective of this study is to understand how sentiment varies across different schools of thought and to identify any significant patterns or trends.

3.1. iSA Algorithm

The iSA algorithm presented by Ceron *et al.* [2] is designed to discern the aggregated sentiment distribution within a corpus of texts. It operates on a set of predefined categories, denoted as $D = \{D_0, D_1, \ldots, D_{M-1}\}$, with D_0 representing the "off-topic" category. The primary objective is to estimate $P(D)$, which signifies the distribution of opinions across the corpus. Initially, we establish some key notations: the corpus comprises N distinct texts, each containing L stems. Textual content is represented through vectors S_i in the Document Term Matrix, where k observed combinations are far fewer than the 2^L possible combinations. The algorithm's strategy hinges upon the theorem of total probability:

$$P(D) = P(D|S)P(S)$$

Here, $P(D|S)$ represents the conditional frequency of sentiment D given a specific S, while $P(S)$ denotes the distribution of S_i across the corpus. Challenges arise due to sparse conditional probabilities resulting from limited expressions of opinions and potential bias stemming from the dominance of the "off-topic" category. To mitigate these challenges, a shift in perspective is employed, leveraging the relationship:

$$P(S) = P(S|D)P(D)$$

By focusing on $P(S|D)$, the frequency distribution of stems in classified texts is evaluated. The least squares formula is then utilized to estimate $P(D)$:

$$P(D) = \left[P(S|D)^T P(S|D)\right]^{-1} P(S|D)^T P(S)$$

This approach facilitates the estimation of the aggregated distribution of opinions without the need for individual text classification, thereby enhancing the overall accuracy of the estimation process. The iSA algorithm presents a robust methodology for analyzing sentiment distribution in textual corpora, offering valuable insights into the underlying opinions prevalent within the dataset.

3.2. iSA Algorithm Integration and Comparison with BERT

In the context of sentiment analysis, we integrated the iSA algorithm, as described in the previous section, into our analytical framework. We added an operation to the preoprocessing described in 1.1 which consists of removing sentences that, once the stopwords were removed, remained with less than 5 or with more than 70 words, this is because after a few attempts we saw that it improved the results obtained. Given the supervised nature of iSA, a portion of the dataset was pre-tagged using the VADER lexicon-based tool. It leverages a pre-defined list of lexical features along with grammatical rules to estimate the sentiment polarity of textual content.

After pre-tagging, we observed that the data was unbalanced. To address this imbalance, we used the RandomOverSampler technique. RandomOverSampler oversamples the minority class (negative sentiment) to match the number of instances in the majority classes, thereby balancing the dataset.

To validate the effectiveness of iSA in capturing sentiment nuances within philosophical texts, we conducted a comparative analysis with BERT, a state-of-the-art deep learning model described in 4.1. Two histograms were shown in Figure 2 and 3, each depicting the sentiment distributions derived from iSA and BERT, respectively.

From the two graphs, it is evident that both models agree on the predominance of a positive sentiment over a negative one for all schools. However, BERT exhibits more "confident" results in the sense that the bars of the different schools have more distant values from each other compared to the graph with iSA

Regarding computational efficiency, iSA demonstrates significantly shorter training and prediction times compared to BERT. iSA requires approximately 10 seconds for both training and prediction tasks, while BERT necessitates approximately 92 minutes, highlighting the trade-off between computational complexity and performance between the two models.

4. Classification

Classification in Natural Language Processing (NLP) is a fundamental technique used to assign predefined categories to text segments. This process is crucial for numerous applications such as entity recognition, sentiment analysis and, as in this specific case, the categorisation of philosophical texts according to the philosophical current they belong to.

To achieve this, we decided to use BERT (Bidirectional Encoder Representations from Transformers) as a model.

4.1. BERT

BERT (Alammary [3]) represents one of the most advanced technologies in the field of NLP. Due to its transformer-based architecture, BERT is able to understand the context of words within a sentence, allowing for a more accurate representation of the meaning of the text. This

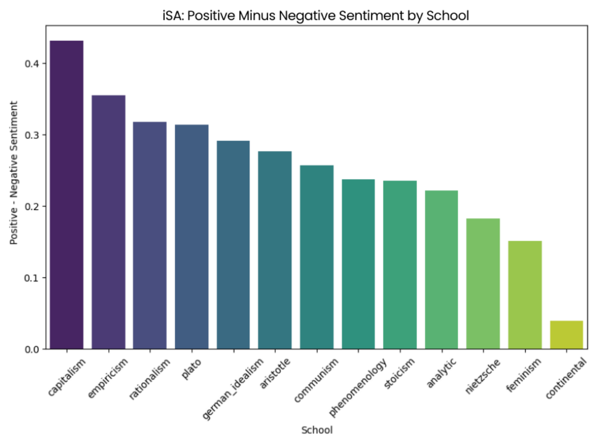

Figure 2. Sentiment distributions derived from iSA

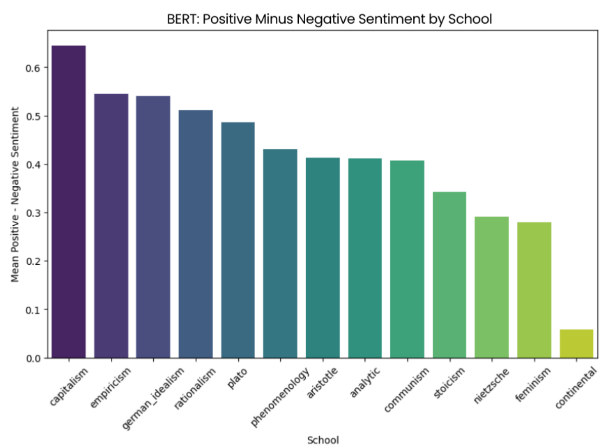

Figure 3. Sentiment distributions derived from BERT

is particularly useful for the classification of philosophical texts, where context and subtlety of language are essential.

Unlike traditional models that process text in a unidirectional manner (left-to-right or right-to-left), BERT reads text in both directions simultaneously. This allows BERT to understand the context of a word based on both its preceding and following words, leading to a more nuanced and accurate understanding of language.

BERT is built on the Transformer architecture, which uses self-attention mechanisms to weigh the importance of each word in a sentence relative to all other words. This enables the model to capture long-range dependencies and relationships within the text more effectively than previous models like RNNs or LSTMs.

Pre-training: BERT is pre-trained on a massive corpus of text from sources like Wikipedia and BooksCorpus.

Fine-tuning: After pre-training, BERT can be fine-tuned on specific tasks (e.g., classification, question answering) with relatively smaller labeled datasets, allowing it to adapt its general language understanding to particular applications.

4.2. Why BERT for classification?

BERT's bidirectional approach and self-attention mechanism (Alammary [3]) enable it to capture the meaning of words in context, making it particularly effective for text classification tasks where the meaning of a word or phrase depends heavily on its surrounding context.

BERT's pre-trained models serve as a strong foundation that can be fine-tuned for specific classification tasks with relatively small datasets. This transfer learning capability means that even with limited labeled data, BERT can achieve excellent results, reducing the need for extensive labeled datasets.

The fine-tuning process for BERT is relatively straightforward and computationally efficient compared to training a model from scratch. Fine-tuning adapts the pre-trained model to the specific nuances of the classification task, leading to faster development cycles and deployment.

4.3. Parameters Selection

Still using the dataset used for other tasks, we initially decided to exploit only part of it to find the right values for the hyper-parameters due to the long training times and limited resources available. In particular, we used 20% of the entire dataset.

We first divided the dataset into train-validation and test into 60% and 40% respectively, and then divided train and validation again into 80% and 20% respectively.

Then we started to modify the regularisation parameters such as weight decay and dropout probability to prevent overfitting. In fact, the first trained models tended to perform very well in the training set (95% of F1 score) without generalising well in the validation set and test set (under 65% of F1 score). We decided to use F1 score rather than accuracy as an evaluation metric since accuracy, which measures the percentage of correct classifications, can be misleading in the presence of an unbalanced dataset. The F1 score, on the other hand, provides a more balanced measure of model performance by combining precision and recall.

We then decided to set the number of epochs at 3, as the loss function already achieved convergence, and a batch size equal to 12 as it was a fair compromise between good model performance and speed of learning.

4.4. Performance Evaluation

After implementing the classification model using BERT, we evaluated the results obtained using, as mentioned above, the F1 score as a reference metric.

From the confusion matrix (Figure 4) and the individual F1 scores (Figure 5) for each class, we can confirm the fact that the most represented classes within the dataset are the same as those with the highest F1 score, while for the less represented currents, the F1 score was lower, but still adequate, showing that despite the smaller amount of data, the model is still able to generalise fairly well.

In conclusion, the use of BERT proved effective for the classification of philosophical texts. The adoption of the F1 score as a metric provided a balanced assessment of the model's performance, highlighting the fact that the chosen model performs well and that despite the presence of classes with few observations (such as the stoicism class with only 0.67%) it performs well. In general, we obtain a 0.86% of F1 score in the training set and 0.79% of F1 score in the test set.

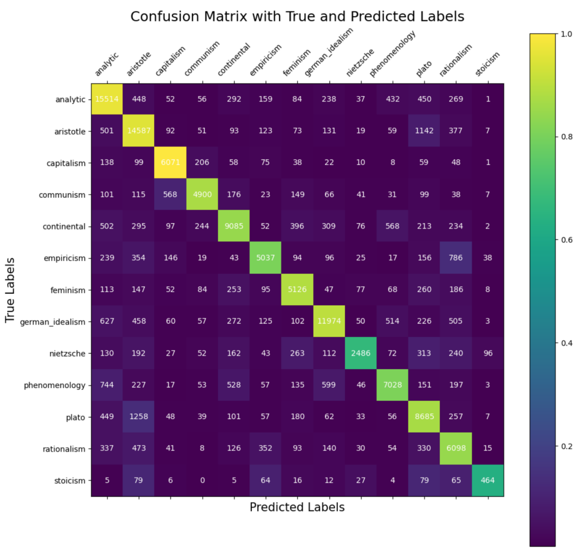

Figure 4. Confusion Matrix with True and Predicted Labels with BERT classification

```
Classification Report:
                 precision    recall  f1-score   support

       analytic       0.80      0.86      0.83     18032
      aristotle       0.78      0.85      0.81     17255
     capitalism       0.83      0.89      0.86      6833
      communism       0.85      0.78      0.81      6314
    continental       0.81      0.75      0.78     12073
     empiricism       0.80      0.71      0.76      7050
       feminism       0.76      0.79      0.77      6516
german_idealism       0.87      0.80      0.83     14973
      nietzsche       0.84      0.59      0.70      4188
  phenomenology       0.79      0.72      0.75      9785
          plato       0.71      0.77      0.74     11232
    rationalism       0.66      0.75      0.70      8097
       stoicism       0.71      0.56      0.63       826

       accuracy                           0.79    123174
      macro avg       0.79      0.76      0.77    123174
   weighted avg       0.79      0.79      0.79    123174
```

Figure 5. F1 score for each class

5. Conclusions

In this report, we presented a comprehensive analysis of philosophical texts using modern computational linguistics techniques. Here are some possible conclusions:

5.1. Key Findings

Our key findings include:

1. **Importance of Computational Linguistics in Philosophy**: The use of modern language analysis techniques has revealed the potential of computational linguistics in exploring the structure and content of philosophical texts. This intersection between philosophy and natural language processing offers new insights for understanding philosophical thought.

2. **Revelations about Philosophical Schools**: Analyzing words and sentiments in the texts has highlighted the differences and similarities between various schools of thought. From the synonyms used to sentiment distributions, we were able to identify distinctive traits of each philosophical current.

3. **Effectiveness of Analysis Models**: The use of models such as Word2Vec, GloVe, and BERT has demonstrated their utility in analyzing philosophical texts.

5.2. Limitations and Future Work

However, our study has limitations such as:

- **Limited Dataset**: Our analysis was based on a specific dataset, which may not fully represent the diversity of philosophical texts.

- **Computational Resources**: Due to constraints, we could only explore a subset of techniques and models.

- **Interpretation Challenges**: Analyzing philosophical texts requires nuanced understanding, which may pose challenges for automated methods.

Future research could focus on:

- **Expanding the Dataset**: Including more diverse texts from different philosophical traditions to enhance the breadth of analysis.

- **Incorporating Domain Knowledge**: Integrating domain-specific knowledge to enhance the interpretation of results and address nuances in philosophical discourse.

In conclusion, our study provides valuable insights into philosophical texts through the lens of computational linguistics. By leveraging modern techniques, we have shed light on the language, sentiment, and classification within this domain.

References

[1] N. Mohammed, "Extracting word synonyms from text using neural approaches," *Int. Arab J. Inf. Technol.*, vol. 17, pp. 45–51, 2019.

[2] A. Ceron, L. Curini, and S. M. Iacus, "Isa: A fast, scalable and accurate algorithm for sentiment analysis of social media content," *Information Sciences*, vol. 367, pp. 105–124, 2016.

[3] A. S. Alammary, “Bert models for arabic text classification: A systematic review,” *Applied Sciences*, vol. 12, no. 11, 2022, ISSN: 2076-3417.

[4] D. J. Hopkins and G. King, “A method of automated nonparametric content analysis for social science,” *American Journal of Political Science*, vol. 54, no. 1, pp. 229–247, 2010.