Authorship Detection: Who Wrote the Letter to the Hebrews?

Project by Jie Dong, Leng Yang Tan, Tien-en Joel Phua

Supervised by Prof. Derek Abbott Co-supervised by Dr. Brian Ng

1. Controversy

“It was unlikely Paul was the author of Hebrews because the anonymity of the letter was not consistent with Paul’s pattern”

“Paul was likely the author of Hebrews originally and not the Greek Version. The latter is the work of Luke who translated Paul’s letter”

2. Introduction

The New Testament of the Bible contains a number of text that have disputed or unknown authorship. One of the texts that has been widely debated would be the Letter to the Hebrews. Three feature extraction algorithms with Support Vector Machine were implemented to develop a classification model. The results obtained would help in identifying the rightful author of Letter to the Hebrews.

3. Project Objective

The objective of this project is to identify the author of the Letter to the Hebrews. This project will provide a non-biased approach in authorship attribution. It will also implement the use of the Support Vector Machine to develop a classification model and analyse the use of different kernel functions in authorship attribution.

4. Approach

Text Processor

Training Texts

Disputed Texts

Feature extraction algorithms:
- Function Words Analysis
- Word Recurrence Interval
- Trigram Markov Model

Support Vector Machine

5. Result

<table>
<thead>
<tr>
<th>Authors</th>
<th>Abbreviations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Barnabas</td>
<td>BA</td>
</tr>
<tr>
<td>Clement</td>
<td>CL</td>
</tr>
<tr>
<td>John</td>
<td>JO</td>
</tr>
<tr>
<td>Luke</td>
<td>LU</td>
</tr>
<tr>
<td>Mark</td>
<td>MA</td>
</tr>
<tr>
<td>Matthew</td>
<td>MT</td>
</tr>
<tr>
<td>Paul</td>
<td>PA</td>
</tr>
<tr>
<td>Peter</td>
<td>PE</td>
</tr>
</tbody>
</table>

6. Conclusion

The feature extraction algorithms together with the Support Vector Machine were able to accurately attribute a disputed text to its rightful author. However, the algorithms require a sufficiently large number of training data in order to make the classification accurate. In the case of the New Testament texts, only a small set of training data was available. This limits the accuracy of the classification model. In this case, the algorithms were only able to accurately eliminate unlikely authors from the list.

7. Future Applications

- Plagiarism Analysis
- Search Engine
- Biological Analysis

References

