Authorship detection: Who wrote the Letter to the Hebrews?

by Kai He, Yan Xie, Zhaokun Wang

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

THE UNIVERSITY of ADELAIDE

1. Introduction

Debate over the authorship of *The letter to the Hebrews* has continued for centuries.

In the 4th century, Paul was commonly supported as the author and *The letter to the Hebrews* was identified as the fourteenth letter by Paul (Fonck, 1910). However, Milligan George commented that the author of Hebrews was unlikely to be Paul since the writing style and contents of Hebrews are inconsistent to Paul's other letters.

Conversely, Clement of Alexandria claimed that the differences in writing style is caused by Luke who translated Paul's letter, Paul could likely be the author of the original letter in Greek.

A list of potential authors had been generated over the year by Biblical scholars and they are: Barnabas (BA), Clement (CL), John (JO), Luke (LU), Mark (MA), Matthew (MT), Paul (PA) and Peter (PE).
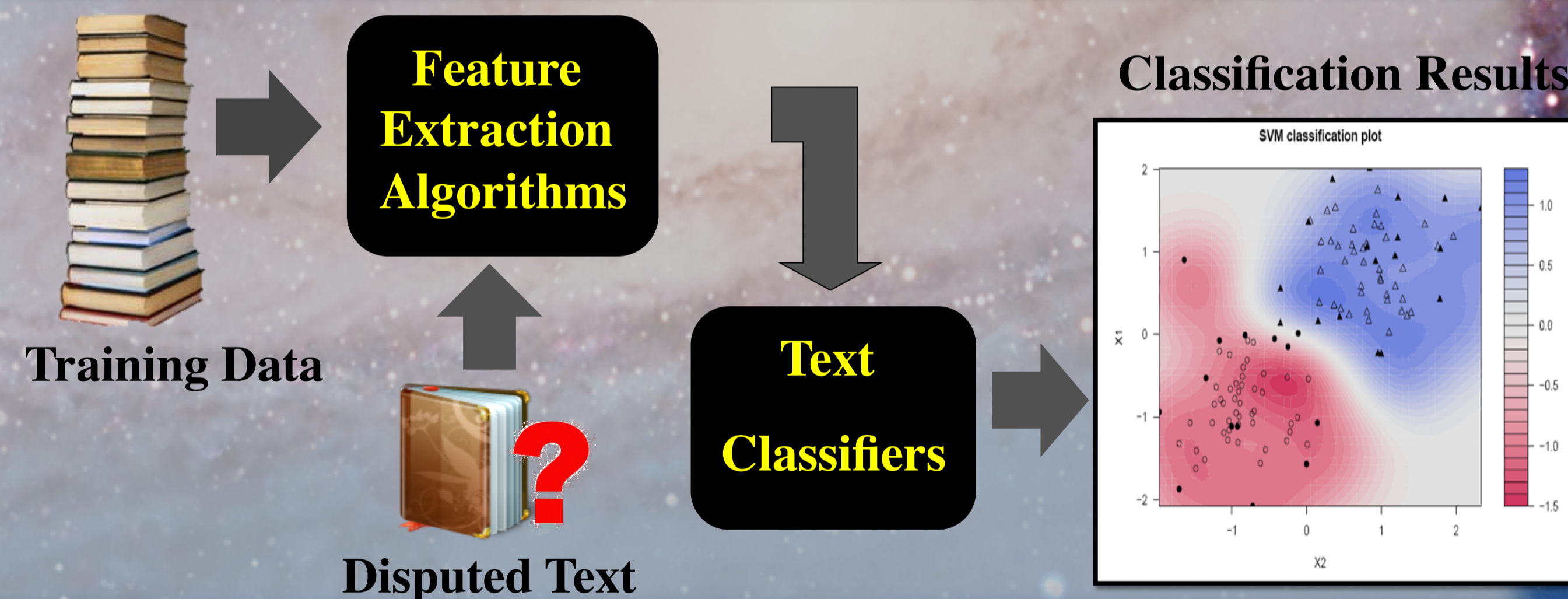
2. Project Objectives

The objective of this project is to identify the author of *The letter to the Hebrews*.

This project will use non-biased approach tin authorship attribution. Two extraction methods and three text classifiers are implemented in this project that aims to provide static evidence to help identify the author of *The letter to the Hebrews.*

3. Implementation

EXTRACTION METHOD	Maximum Frequent Word Sequence	Common N-gram
	MFWS are word patterns that frequently appear in texts.	Common N-gram breaks words into byte level. Text = Te_, _ex, xt, t_
TEXT CLASSIFIER	Naïve Bayes	• Support Vector Machine • Dissimilarity Calculation

4. Approach

Training Data → Feature Extraction Algorithms → Text Classifiers → Classification Results
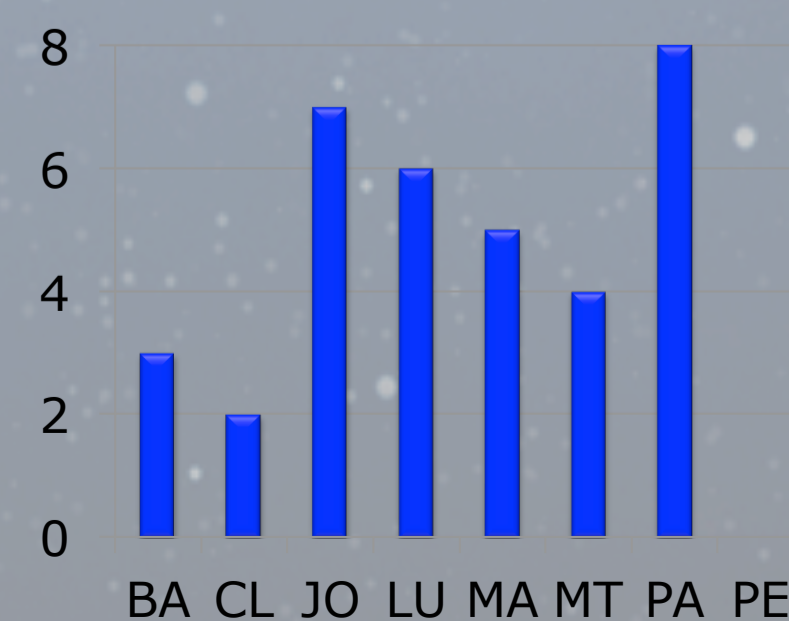
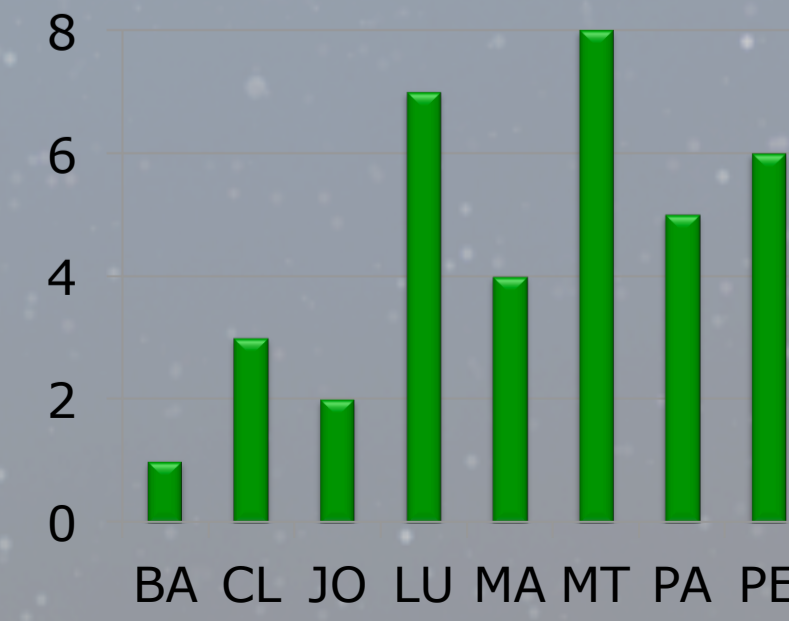Disputed Text

5. Result

Maximum Frequent Word Sequence & Naïve Bayes

Common N-gram & Support Vector Machine

Common N-gram & Dissimilarity Calculation

6. Conclusion

The two feature extraction algorithms with the three classifiers together achieved reasonable accuracy in classification tests. Final results from these methods also agree that the most likely author of *the letter to the Hebrews* is Paul. Note there are inconsistency in the rest results. Since New Testaments only has small collection of training data to use, the classfication models are highly limited and the results can not be further verified. However, based on colloected testing data, the methods used in this project are competent as classification tools and can be further developed to have various of appliacations.

7. Future Applications

References

[1] J. Vernon McGe, The Authorship of Hebrews or Did Paul Write Hebrews?,Thru the Bible RadioNetwork<http://www.thruthebible.org/atf/cf/%7B91e2424c-636c-40c2-9c55-890588e90ece%7D/AUTHORSHIP%20OF%20HEBREWS.PDF> viewed August 2011.
[2]Jie. D., Leng. Y. T. & Tien-en. J. P., Who Wrote the Letter to the Hebrews? – Data Mining for Detection of Text Authorship, University of Adelaide, 2010.
[3] Talis J. Putnins, Domenic J. Signoriello, Samant Jain, Matthew J. Berryman and Derek Abbott, "Advanced text authorship detection methods and their application to biblical texts", Proc. SPIE: Complex Systems 6039 ed. Axel Bender, Brisbane, Qld., Australia, December 11-14, 2005.
[4]Rosa M C, Luis V.P, Manuel M.G, Paolo R, Authorship Attribution using Word Sequences, Universidad Politécnica de Valencia.
[5] Noble, S. William, What is support Vector machine, available: http://www.nature.com/naturebiotechnology, Nature Publishing Group, 2006