Transcending Traditional Paleography through Computational Analysis

Berat Kurar-Barakat Daria Vasyutinsky Shapira Sharva Gogawale
Mohammad Suliman Nachum Dershowitz

21.07.2024

Tel Aviv University

Abstract

This study aims to transcend the boundaries of traditional human paleography by utilizing computational algorithms to cluster medieval Hebrew script types. Focusing initially on Ashkenazi square script, we plan to apply successful methods to other types of medieval Hebrew script, such as Byzantine and Yemenite. In doing so, we hope to discover new paleographic patterns and enhance our understanding of sub-clustering within these script types.

We curated a dataset comprising four pages from each of 55 manuscripts, all written in Ashkenazi square script. This collection, which lacks specific annotations regarding the date or region of origin, is thought to include distinct French and German groups, as well as stand-alone manuscripts from England dating before 1290.

Conventional algorithms have been relatively ineffective in identifying features critical for paleographical clustering. As a result, this study employs a more deterministic approach, identifying crucial paleographical features as determined by expert paleographers. These features are then organized within a hierarchical multi-label classification problem. A convolutional neural network (CNN) is used to predict these paleographical attributes for each manuscript page. Following this, a brute-force search is performed to find the feature combinations that lead to the most cohesive clusters.

This approach addresses the challenge faced by paleographers who can easily identify individual features on a single page but struggle to simultaneously

remember and analyze these features across multiple pages to discern grouping patterns. Thus we end up with well-defined clusters plus insights into the features driving these formations, surpassing the limitations of traditional paleographic analysis.