IJFANS INTERNATIONAL JOURNAL OF FOOD AND NUTRITIONAL SCIENCES
ISSN PRINT 2319 1775 Online 2320 7876
Research paper Volume 11, Iss 11, 2022

Digital Palaeography and Copyist Identification: The Case of the Latin 'Monster Bible

Venkateswarlu B[1*], Dr Somasekhar Donthu[2]

[1*,] Assistant Professor, Computer Science and Engineering, Koneru Lakshmaiah Education Foundation (KLEF), Vaddeswaram, Green fields, Guntur, Andhra Pradesh, India -522302.

[2*,] Assistant Professor,School of Business,GITAM University, Bangalore, , Andhra Pradesh, India.

[1*]bvenki289@gmail.com, somuecom@gmail.com
DOI : 10.48047/IJFANS/V11/ISS11/465

ABSTRACT

Utilizing sophisticated image processing techniques within the framework of palaeography has garnered increasing attention recently, giving rise to a new area of study sometimes referred to as "computerized palaeography". Both example acknowledgment and highlight extraction procedures, which provide quantitative arguments to support master derivations, are important in this subject. In this work, we provide an example acknowledgment structure that aims to address a common palaeographic problem: identifying the various copyists who have collaborated on the documentation of a single middle-aged book. In the case of a highly standardized book typology (the purported Latin "Monster Bible"), we wanted to see if extracting any deliberately artificial highlights related to page format would yield results that were acceptable. To this point, we have additionally played out a measurable examination of the considered highlights to describe their discriminant power. The tests, performed on an enormous dataset of advanced pictures from the alleged "Avila Bible" - a monster Latin duplicate of the entire Thus far, we have also conducted a quantifiable analysis of the highlights under consideration in order to characterize their discriminant potential. Experiments conducted using a massive dataset of sophisticated images from the so-called "Avila Bible"—a massive Latin copy of the complete Bible distributed between Italy and Spain in the 18th century—verified the optional nature of the suggested method. Bibles distributed in Italy and Spain in the 18th century attested to the elective nature of the suggested method.

Introduction:

Regarding palaeographic considerations, there has been a growing scientific interest in using PC-based research processes recently. The goal of these procedures is to provide new and more targeted approaches for characterizing

ISSN PRINT 2319 1775 Online 2320 7876
 Volume 11, Iss 11, 2022

middle-aged penmanship styles and identifying scribal hands. The application of these techniques, which were first developed in the field of criminological investigation, gave rise to a new area of study known as computerized palaeography. On a simpler level, the automated method can be used to replace subjective judgments with quantitative ones, such as when evaluating limits like the point and stroke width or the relationship between complex letter structures. Innovation is used in the situations to carry out "conventional" conceptions faster and more effectively than before. On the other hand, in the last few years, some completely new approaches have emerged, made possible by the combination of amazing PCs and excellent, cutting-edge graphics. These new approaches consider the development of frameworks to assist the decisions made by the experts when they examine old handwriting styles.

From these reflections, we suggest a sample recognition strategy for identifying the several recorders who have contributed to the documentation of a single middle-aged book. The suggested framework considers a number of highlights that palaeographers frequently use and that are directly obtained from the examination of the page design. A conventional Multi-Layer Perceptron (MLP) organization constructed using the Back Propagation computation is used for classification. MLP classifiers were used for two main reasons: From one angle, MLP classifiers are simple, highly efficient, and exhibit a respectable capacity for speculation. However, the main goal of our research is not to develop a highly performing recognition framework, but rather to verify that the use of page design highlights enables achieving good results. Finally, to characterize the separating force of each of the highlights under consideration, we have also conducted a quantifiable examination of them.

The results presented in Section 4 validated that the suggested approach enabled us to select the element subset that enhances classification outcomes. The so-called "Monster Bibles"—at least 100 successively delivered Latin works that each include the full sacred text in a single volume of huge size (up to 600 × 400 mm and beyond)—offer an exceptionally favourable setting for testing the efficacy of this methodology. The Bible was created in Central Italy (first in Rome) in the eleventh century as a component of the "Gregorian Reform" political program, which managed the church's autonomy, moral rectitude, and relationship to the Holy Roman Emperor. The Bibles, which had virtually the same fit, material highlights, design, and content, were produced by groups of copyists who assembled their typical work according to models that had to be deeply understood. Palaeographical examinations that are quite lengthy and patient are often necessary for the qualification of their hands.

In this case, we have used the "Avila Bible" example for our research. Written in Italy in the third decade of the 12th century by at least nine recorders, it was eventually sent (for unclear reasons) to Spain, where local recorders completed its embellishment and content. At a later stage (in the fifteenth century), additional copyists added augmentations to the literary arrangement to conform to new ceremonial requirements. The Bible provides a "treasury" of scribal hands that are both contemporary and not contemporary, and it then poses a serious test to determine the viability and effectiveness of our current approach to scribal hand

ISSN PRINT 2319 1775 Online 2320 7876
 Volume 11, Iss 11, 2022

qualification. The remainder of the document is organized as follows: The framework's engineering is shown in Section 2, the highlight investigation method is described in Section 3, and the exploratory results are discussed and analysed in Section 4. Finally, certain goals are discussed in Section 5.

LITERATURE REVIEW

3.1 PAPER-1

AIM: Connection based Feature Selection for Machine Learning

ABSTRACT:

Finding an agent's set of highlights from which to construct a categorization model for a given task is a central problem in artificial intelligence. This proposal uses a relationship-based mechanism to address the problem of AI highlight choice. According to the focus idea, acceptable capabilities have highlights that are deeply aligned with the class but not with each other. An operational definition of this theory is provided by a component assessment recipe, considering ideas from the test hypothesis. A computation known as CFS (Correlation based Feature Selection) combines this evaluation formula with an appropriate relationship measure and a heuristic inquiry approach.

TECHNIQUES USED:

FEATURE SELECTION FOR MACHINE LEARNING

Many factors affect how well AI performs on a particular task. First and foremost is the way the model information is portrayed and its type. Theoretically, greater highlights ought to result in significantly more separation power. However, practical experience with AI computations has demonstrated that this isn't always the case. Many learning computations can be viewed as producing a (biased) estimate of the probability of receiving a passing grade based on a number of highlights. This is a high dimensional, confusing dissemination. Lamentably, acceptance is regularly performed on restricted information. This complicates evaluating the several probability limits. Many computations use the Occam's Razor inclination to create a simple model that really achieves a respectable degree of execution on the preparation data in order to avoid overfitting it.. This tendency often leads to a computation that favors a small number of predictive characteristics over innumerable highlights that, when used in the right combination, are predictive of the class mark in full. Getting the swing of things during the preparation stage becomes more challenging if there is a lot of unnecessary and surplus data present, or if the information is erratic and noisy. Finding the highlight subset is the first step in separating and removing as much redundant and irrelevant data as is reasonable. As a result, the information is less dimensional, which could speed up and improve the accuracy of learning computations. Sometimes future classification accuracy can be increased; other times, the result is a more condensed, easily understood representation of the

FEATURE SELECTION IN STATSTICS AND PATTREN RECOGNITION

Determining the highlight subset has long been a research area within measurements and example acknowledgment [DK82, Mil90]. Given that both domains share the common task of classification, it is not surprising that include choice poses as big of a challenge for AI as it does for design recognition. Include determination in design acknowledgment can have an impact on the accuracy and complexity of the classifier as well as the financial aspects of information security.

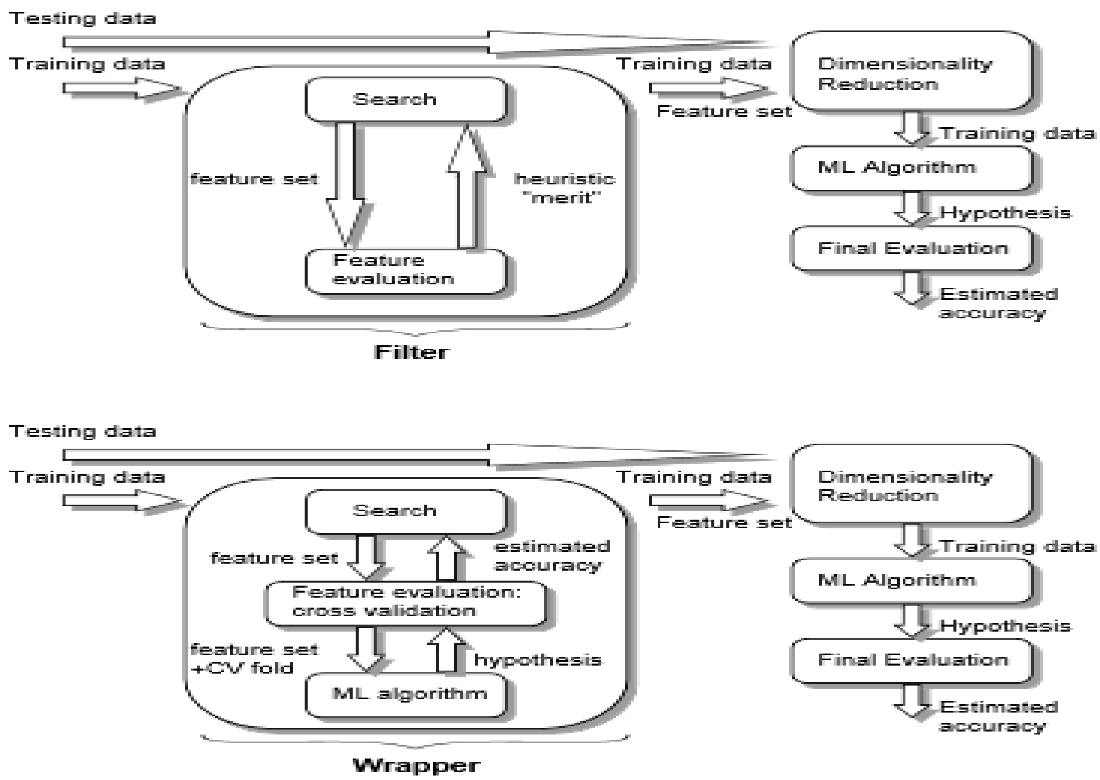
This also applies to AI, which has the extra concern of extracting useful information from data. Fortunately, it appears that inclusion choice enhances the comprehensibility of material that has been eliminated. AI has absorbed inspiration and learned from insights as well as example recognition. Allen [All74] suggested using 26cross-approval to evaluate a component subset's precision, which has since become the cornerstone of AI's covering method. This was done in relation to the problem of selecting indicators in straight relapse. Numerical highlights are simply ideal for a wide range of quantifiable methods1 that evaluate the value of highlight subsets based on characteristics of the preparation information. Furthermore, these behaviors are often monotonic—a requirement that is not met by sensible AI algorithms2. Therefore, search algorithms that rely on monotonicity to reduce the search space, like dynamic programming and branch and bound [NF77], are not pertinent to emphasize determination algorithms that use or try to align with the general bias of AI algorithms.

HURATIONAL SEARCH

If a component choice computation has to operate on data with a large number of highlights, it is imperative to search across the space of highlight subsets within reasonable time constraints. A simple search strategy, known as covetous slope climbing, considers local modifications to the current component subset. Often, a close modification is simply an addition to or deletion of one single element from the subset. Forward choice occurs when the

IJFANS INTERNATIONAL JOURNAL OF FOOD AND NUTRITIONAL SCIENCES
ISSN PRINT 2319 1775 Online 2320 7876
Research paper Volume 11, Iss 11, 2022

computation only takes into account additions to the component subset; considering just cancellations is known as in reverse end.



GREEDY HILL CLIMBING SEARCH ALGORITHM

1. Let $s \leftarrow$ start state.
2. Expand s by making each possible local change.
3. Evaluate each child t of s.
4. Let $s' \leftarrow$ child t with highest evaluation $e(t)$.
5. If $e(s') \geq e(s)$ then $s \leftarrow s'$, goto 2.
6. Return s.

Table 3.1: Greedy hill climbing search algorithm

IJFANS
International Journal of Food And Nutritional Sciences
Official Publication of International Association of Food and Nutrition Scientists

ISSN PRINT 2319 1775 Online 2320 7876
 Volume 11, Iss 11, 2022

BESTFIRSTSEARCHALGORITHM

1.	Begin with the OPEN list containing the start state, the CLOSED list empty, and BEST←start state.
2.	Let $s = \arg\max e(x)$ (get the state from OPEN with the highest evaluation).
3.	Remove s from OPEN and add to CLOSED.
4.	If $e(s) \geq e(\text{BEST})$, then BEST $\leftarrow s$.
5.	For each child t of s that is not in the OPEN or CLOSED list, evaluate and add to OPEN.
6.	If BEST changed in the last set of expansions, goto 2.
7.	Return BEST.

Table 3.2: Best first search algorithm

SIMPLEGENETICSEARCHSTRATEGY

1.	Begin by randomly generating an initial population P.
2.	Calculate $e(x)$ for each member $x \in P$.
3.	Define a probability distribution p over the members of P where $p(x) \propto e(x)$.
4.	Select two population members x and y with respect to p.
5.	Apply crossover to x and y to produce new population members x' and y'.
6.	Apply mutation to x' and y'.
7.	Insert x' and y' into P' (the next generation).
8.	If $\|P'\| < \|P\|$, goto 4.
9.	Let $P \leftarrow P'$.
10.	If there are more generations to process, goto 2.
11.	Return $x \in P$ for which $e(x)$ is highest.

Table 3.3: Simple genetic search strategy.

CONCLUSION:

There is no one learning calculation that works better than the others for every problem. Research in AI aims to provide insight into the capabilities and constraints of different computations. Equipped with this knowledge, as well as the fundamental data for a particular problem, experts are able to choose which computations to use. This is the case with CFS; on the whole, it can improve (or not corrupt) the way AI computations are displayed while also

ISSN PRINT 2319 1775 Online 2320 7876
 Volume 11, Iss 11, 2022

achieving a reduction in the number of highlights used for learning. However, when information comprises clearly interacting highlights or highlights with values predictive of a small region of the occasion space, CFS may fail to choose noteworthy highlights.

FUTURE SCOPE

Refinement of Techniques: Further refinement and development of image processing techniques to enhance the accuracy and efficiency of identifying copyists. This may involve exploring advanced algorithms, machine learning, or deep learning methods.

Expanding the Dataset: Expanding the dataset beyond the "Avila Bible" to include a more diverse range of medieval texts and documents. This can improve the generalizability and robustness of the proposed method.

Interdisciplinary Collaboration: Collaborating with experts in other fields such as history, art history, and linguistics to gain deeper insights into the historical and cultural context of the texts. This interdisciplinary approach can lead to a more comprehensive understanding of the copyists' work.

Automation and Efficiency: Investigating ways to automate the process of identifying copyists, reducing the need for manual intervention and increasing the efficiency of analysis.

Integration with Preservation Efforts: Exploring how the findings of computerized palaeography can contribute to the preservation and restoration of medieval texts and manuscripts.

Application to Other Languages and Scripts: Adapting the method to work with texts in languages and scripts beyond Latin, expanding the applicability of the technique to a wider range of historical documents.

User-Friendly Tools: Developing user-friendly software tools or platforms that can be used by researchers and archivists to apply these techniques to their own collections.

ISSN PRINT 2319 1775 Online 2320 7876
 Volume 11, Iss 11, 2022

Ethical Considerations: Addressing ethical considerations related to the digitization and analysis of historical documents, including issues of cultural sensitivity and data privacy.

Educational Initiatives: Creating educational initiatives to train the next generation of scholars and researchers in the field of computerized palaeography.

Integration with Archives and Libraries: Collaborating with archives and libraries to integrate these techniques into their digitization and cataloging processes, making historical documents more accessible to researchers and the public.

These future directions aim to advance the field of computerized palaeography, making it a valuable tool for historical research, preservation, and cultural understanding.

References:

1. M. Kestemont and D. Stutzmann, “Script identification in medieval Latin manuscripts using convolutional neural networks,” in Digital humanities 2017. Book of abstracts (Montreal, August 10, 2017), 2017, pp. 283–285.
2. D. Stutzmann, “Digital Palaeography: New Machines and Old Texts (Dagstuhl, 20-24 July 2014),” vol. 4, no. 7, pp. 112–134, 2014.
3. T. Hassner, M. Rehbein, P. A. Stokes, and L. Wolf,“Computation and Palaeography : Potentials and Limits,” Manif. from Dagstuhl Perspect. Work., vol. 2,no. 1, pp. 14–35, 2013.
4. C. Tensmeyer, D. Saunders, and T. Martinez,“Convolutional Neural Networks for Font Classification,” 2017. https://doi.org/10.1109/ICDAR.2017.164
5. N. Abbott, “Review: Arabic Paleography,” vol. 8, pp.65–104, 2011.
6. Wan Ali Wan Mat, “Paleografi Jawi : Satu Pengenalan,” J. Filol. Melayu, vol. 14, pp. 65–69, 2006.
7. F. M. H. Al-shareefi, “A Haar Wavelet-Based Zoning For Offline Arabic Handwritten Character Recognition,” no. 2, pp. 575–585, 2015.

8. R. S. Cook, “Unicode Chinese paleography : making the evolutionary leap from bone , bronze , silk , and paper , to electronic bits,” in Early Chinese Workshop Series, 2005.
9. M. S. Azmi, K. Omar, M. F. Nasrudin, B. Idrus, and K. Wan Mohd Ghazali, “Digit Recognition for Arabic/Jawi and Roman using Features from Triangle Geometry,” AIP Conf. Proc., vol. 1522, pp. 526–537, 2013.
10. Borji, M. Hamidi, and F. Mahmoudi, “Robust handwritten character recognition with features inspired by visual ventral stream,” Neural Process. Lett., vol. 28, no. 2, pp. 97–111, 2008. https://doi.org/10.1007/s11063-008-9084-y
11. a. Brink, J. Smit, M. L. Bulacu, and L. R. B.Schomaker, “Writer identification using directional ink-trace width measurements,” Pattern Recognit.,vol. 45, no. 1, pp. 162–171, 2012.https://doi.org/10.1016/j.patcog.2011.07.005 26. K. M. Bin Abdl and S. Z. M. Hashim, “Swarm-based
12. feature selection for handwriting identification,” J.Comput. Sci., vol. 6, no. 1, pp. 80–86, 2010. https://doi.org/10.3844/jcssp.2010.80.86
13. L. Schomaker, Louis Vuurpijl, and Lambertus Schomaker, “Forensic writer identification: A benchmark data set and a comparison of two systems,” in NICI, 2000.
14. U.-V. Marti and H. Bunke, “A full English sentence database for off-line handwriting recognition,” Proc. Fifth Int. Conf. Doc. Anal. Recognition. ICDAR ’99 (Cat. No.PR00318), no. November, pp. 705–708, 1999. https://doi.org/10.1109/ICDAR.1999.791885
15. B. B. Chaudhuri, “A Complete Handwritten Numeral Database of Bangla – A Major India Script,” 2006. 30. U. Bhattacharya and B. B. Chaudhuri, “Handwritten numeral databases of Indian scripts and multistage