

Design and Application of Clerical Style Recognition System Based on Data Mining Algorithm

Feifei Jiang

Shaanxi University of Science and Technology, China

Chenghu Ke

Xi'an University, China

Chenchen Zhong

Shaanxi University of Science and Technology, China

Xiaoling Zhang

City University of Hong Kong, China

ABSTRACT

With the advancements in high-definition imaging and parallel computing hardware, the analysis of massive visual data has become a key focus in pattern recognition and artificial intelligence. Chinese calligraphy, an integral part of traditional culture, has seen digitization of numerous works stored in digital libraries. However, current automatic calligraphy character recognition technology is limited, necessitating the development of efficient computer vision methods for recognizing calligraphy styles. Data mining, crucial in artificial intelligence, involves extracting valuable knowledge from vast and noisy datasets. Recent simulation results show promising recognition rates for Chinese text images, with an average recognition time of 5 seconds per 100 words. This system significantly improves handwriting recognition accuracy compared to existing algorithms, though further refinement and expansion are needed for optimal functionality.

KEYWORDS

Clerical Script, Calligraphy Recognition, Official Script, Data Mining, Clustering, Neural Network-Based ML Algorithms for Visual Recognition

INTRODUCTION

With the improvement of high-definition imaging technology and parallel computing hardware capability, the analysis and understanding of technology based on massive visual data has become one of the research hotspots in the field of pattern recognition and AI (Yang & Ko, 2022). Text recognition has always been an important research direction in the field of computer vision, which is not only valuable in letter and parcel sorting, manuscript editing and proofreading, bank check processing, etc., but also promotes other areas of computer vision. However, compared with other languages, the semantic level of Chinese characters is more complex, with each character having its own fixed meaning, similar to words composed of multiple letters in other languages (Rupannagudi et al., 2015). The structure of Chinese characters is also extraordinarily complex, which makes it more difficult

DOI: 10.4018/IJISMD.365344

This article published as an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>) which permits unrestricted use, distribution, and production in any medium, provided the author of the original work and original publication source are properly credited.

to recognize them, especially in handwritten Chinese characters, where similar characters may be distorted to varying degrees due to the writing styles of different people (Cheng et al., 2022a or b).

In the course of the development of Chinese characters, the LiShu stage is an important period in the change of Chinese characters' fonts. During this stage, Chinese characters were gradually transformed from pictograms to symbols, strokes were changed from flexed to straight, and the font structure was constantly standardized, which had an important impact on the simplification of Chinese characters (Li et al., 2020). However, it is difficult to determine the rules of arrangement and combination of Chinese characters, and with the development of different historical periods, the form of Chinese characters underwent many changes. Therefore, a detailed description of the evolutionary path of the clerical script not only helps to scientifically explain the origin of the simplification of Chinese characters but also provides theoretical guidance for the study and creation of modern calligraphic art (Diez-Olivan et al., 2017).

Data mining, as an interdisciplinary research field, integrates the research results of several disciplinary fields, such as database technology, ML, knowledge engineering, and so on. After years of research, data mining techniques have become clearer and developed in a deeper direction (Wang et al., 2021). With the wide application of large-scale data systems, the application of data mining techniques in the business field has been gaining more and more attention, which not only improves the application of data but also helps decision makers to make better business decisions through statistical analysis and reasoning of data (Yang et al., 2021). However, with the continuous development of information technology, database systems are faced with the challenge of dealing with massive amounts of data, and traditional search query mechanisms and statistical methods can no longer meet the current needs (Cheng, 2022a or b).

The aim of this study is to design an efficient and accurate system for recognizing calligraphic Chinese characters, and to realize the automated recognition and classification of calligraphic images by integrating data mining techniques and deep learning algorithms. The authors select samples from a large number of calligraphy image datasets and undergo meticulous preprocessing to improve image quality. Subsequently, the samples are feature extracted and classified using data mining methods, especially techniques such as cluster analysis. In model training, the authors experimentally adjusted parameters such as different batch sizes and input image sizes to enhance the accuracy of the model. Ultimately, the advantages of the authors' algorithm in terms of recognition rate and number of recognized categories are verified by comparing with other algorithms. This research is of great significance to the development of the field of calligraphic Chinese character recognition and provides strong support for the realization of automated calligraphic image recognition and classification.

In this paper, by applying the mining method to practical applications, the authors designed a data mining algorithm based on clerical style recognition system to solve the problems encountered in the development, implementation, and maintenance of the project of the international company. In it, the basic principles, steps, related techniques, and main applications of cluster analysis are introduced, as well as the advantages and disadvantages of commonly used clustering algorithms and their applications in specific fields. This research is not only methodologically innovative, but also provides technical support to quickly and accurately find suitable clustering algorithms for specific domains. Combined with the introduction, literature review, methodology, experimental analysis, and conclusion outlook sections, this paper comprehensively discusses the design and application issues of the data mining algorithm-based scribe recognition system, which provides an important reference for further research in this field.

RELATED WORK

In recent years, with the vigorous development of social information construction, the increasing network coverage and the popularization of large-scale database technology, a large amount of data generated in the application process has potential value that can't be ignored. Mining potential

15 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/article/design-and-application-of-clerical-style-recognition-system-based-on-data-mining-algorithm/365344

Related Content

Creating Applications for Real-Time Collaboration with XMPP and Android on Mobile Devices

Daniel Schuster, István Koren, Thomas Springer, Dirk Hering, Benjamin Söllner, Markus Endler and Alexander Schill (2012). *Handbook of Research on Mobile Software Engineering: Design, Implementation, and Emergent Applications* (pp. 824-844).

www.irma-international.org/chapter/creating-applications-real-time-collaboration/66501

Separation of Concerns in Model-Based Development of Distributed Real-Time Systems

Dionisio de Niz, Gaurav Bhatia and Raj Rajkumar (2010). *Behavioral Modeling for Embedded Systems and Technologies: Applications for Design and Implementation* (pp. 147-170).

www.irma-international.org/chapter/separation-concerns-model-based-development/36341

Validating Security Design Pattern Applications by Testing Design Models

Takanori Kobashi, Nobukazu Yoshioka, Haruhiko Kaiya, Hironori Washizaki, Takano Okubo and Yoshiaki Fukazawa (2014). *International Journal of Secure Software Engineering* (pp. 1-30).

www.irma-international.org/article/validating-security-design-pattern-applications-by-testing-design-models/121680

Preventing the Increasing Resistance to Change through a Multi-Model Environment as a Reference Model in Software Process Improvement

Mirna Muñoz and Jezreel Mejia (2014). *Agile Estimation Techniques and Innovative Approaches to Software Process Improvement* (pp. 97-118).

www.irma-international.org/chapter/preventing-the-increasing-resistance-to-change-through-a-multi-model-environment-as-a-reference-model-in-software-process-improvement/100274

Analysis of Psychological Distress During COVID-19 Among Professionals

Supriya Raheja (2022). *International Journal of Software Innovation* (pp. 1-17).

www.irma-international.org/article/analysis-of-psychological-distress-during-covid-19-among-professionals/309109