

Investigating the Influence of Chinese Writing Systems and Fonts on Reading Performance Using Deep Learning

Authors: Zhang Yuxin, Liu Bin, Siqu Huang, Zhang Yuxin, Liu Bin

Date: 2022-01-05T00:00:00+00:00

Abstract

[Objective] This paper conducts an objective comparative study on the reading performance of different fonts and writing systems in Chinese publications.

[Method] Specifically, Chinese characters are rendered into images of their corresponding glyphs, and further, according to word order, the Chinese character images in a sentence are folded into a three-dimensional sentence tensor. For the same piece of Chinese text, using different fonts or simplified/traditional characters yields visually differentiated tensor representations of the sentence. By further inputting the obtained sentence tensors into our designed deep language model and testing on tasks such as text classification, the impact of fonts and writing systems on reading performance can be objectively compared.

[Results] Through testing on two Chinese text classification datasets, it was found that some special, infrequently used fonts exhibit lower machine recognition accuracy compared to commonly used fonts, and that there are also differences in reading performance among different commonly used fonts.

Conclusion Through hypothesis testing, it was concluded that datasets using KaiTi and HeiTi fonts exhibit a significant difference in accuracy on text classification tasks, with KaiTi demonstrating higher reading performance compared to HeiTi. There exists a significant difference in reading performance between Simplified Chinese and Traditional Chinese.

Full Text

Study of Chinese Writing Systems and Fonts on Reading Performance Using Deep Learning

Yuxin Zhang, Siqu Huang, Bin Liu*

Center of Statistics Research, School of Statistics

Southwestern University of Finance and Economics, Chengdu 610031

Abstract

[Objective] This paper presents an objective comparative study on the reading performance of different fonts and writing systems used in Chinese publications. [Methods] Specifically, Chinese characters are rendered into their corresponding glyph images, which are then folded into three-dimensional sentence tensors according to word order. For the same Chinese text, different fonts or simplified/traditional variants produce visually distinct tensor representations. By feeding these sentence tensors into our designed deep language model for text classification tasks, we can objectively compare the impact of fonts and writing systems on reading performance. [Results] Testing on two Chinese text classification datasets reveals that machine recognition accuracy for some uncommon fonts is lower than that for common fonts, and even among common fonts, reading performance varies. [Conclusions] Through hypothesis testing, we find significant differences in text classification accuracy between datasets using regular script (Kai) and bold script (Hei), with regular script demonstrating higher reading performance. Significant differences also exist between simplified and traditional Chinese writing systems.

Keywords: reading performance; text classification; deep neural network

Introduction

Different typographic features of Chinese characters produce distinct visual-psychological effects during reading, leading to variations in reading performance. For instance, most Chinese printed materials and web pages use regular script (Kai) or Song typeface for body text rather than bold typeface. For Chinese, beyond the visual-psychological differences induced by fonts, the coexistence of two writing systems—simplified and traditional Chinese—also creates visual-psychological disparities in textual expression efficiency. While some simplified characters may appear less ideographically transparent than their traditional counterparts, this does not necessarily mean the ideographic attributes of Chinese characters have diminished after simplification. For example, when “塵” (dust) is simplified to “尘” and “滅” (extinguish) to “灭,” the simplified forms become more comprehensible than the traditional ones. Debates over the merits of simplified versus traditional Chinese have persisted for decades, with scholars also investigating the visual communication and recognition performance of Chinese fonts.

Since the 1920s and 1930s, experimental research on Chinese characters has been a joint concern in linguistics and psychology. Linguistic studies of character recognition typically approach from a philological perspective, examining correlations between structural features and character recognition. Other researchers, drawing on Chinese informatics and cognitive psychology, simulate

the psychological processes of human character recognition to address practical issues in computer information processing systems. Additionally, some scholars have conducted experimental studies on discrimination rates of different Chinese fonts from physiological and psychological perspectives. For example, Jin et al. compared reading performance across four Chinese fonts—Song, bold, long imitation Song, and regular imitation Song—using correct judgment rates as an indicator under three lighting conditions.

Existing research on Chinese font recognition primarily originates from sociology fields related to writing systems, adopting largely subjective methodologies from linguistics and psychology. This paper attempts to objectively study the relationship between Chinese fonts, writing systems, and reading performance through machine reading. Natural reading performance can be approximated by machine text classification accuracy, which serves as the machine-reading equivalent of human reading performance.

Specifically, we utilize character glyphs as text representation, then input them into natural language models for text classification testing. Sentence words or characters are rendered as images and folded into three-dimensional sentence tensors $\chi \in \mathbb{R}^{w \times h \times l}$, where w represents the character image size and l is sentence length, as shown in [Figure 1: see original paper]. Each slice $\chi_i \in \mathbb{R}^{w \times h}$ corresponds to a word or character image, enabling sentence representation through a three-dimensional tensor. We then render the same text segment using different fonts or writing systems to obtain corresponding sentence tensor representations and test their recognition performance in machine language models. Based on this tensor representation, we employ the three-dimensional convolutional language model proposed by Liu et al. to approximate reading performance.

The main contributions of this paper are twofold: (1) From a machine-reading perspective, we propose an objective method to evaluate reading performance across different Chinese writing systems (simplified/traditional) and fonts; (2) Using hypothesis testing, we validate the impact of different writing systems and fonts on reading performance on the Toutiao and THUCNews datasets. Testing on two Chinese text classification datasets reveals performance differences between common and uncommon fonts. Among common fonts, significant differences exist between regular script and bold script. Simplified and traditional writing systems also show significant performance differences.

1.1 Overview

We first render a sentence S into a three-dimensional tensor as shown in [Figure 1: see original paper], where each slice corresponds to a Chinese character. Specifically, character v_i in S is rendered as image $\chi_i \in \mathbb{R}^{w \times h}$. Characters are then sequentially folded into a three-dimensional sentence tensor $\chi \in \mathbb{R}^{w \times h \times l}$, where w is the character image size and l is sentence length. We apply three-

dimensional convolution kernels of size $w \times h \times n$ to this “text tensor,” where w and h are character image width and height, and n is the number of characters. In other words, 3D convolution slides across n characters at once, equivalent to n -gram feature detection. By varying n , we obtain different n -gram detectors, and multiple n -gram convolutions can extract textual features. In our experiments, n takes values from $\{2, 3, 4, 5\}$. Within our framework, integrating multiple n -grams can be implemented easily and efficiently.

In neural language modeling, forward and reverse text information represent two distinct input types. We employ bidirectional convolution to extract textual features.

[Figure 1: see original paper]

1.2 Network Architecture

The neural network architecture is described as follows:

1. 3D Convolutional Layers: - Kernel size = (20, 20, 2), stride = (1, 1, 1), number of kernels = 50, padding = 0 - Kernel size = (20, 20, 3), stride = (1, 1, 1), number of kernels = 50, padding = 0 - Kernel size = (20, 20, 4), stride = (1, 1, 1), number of kernels = 50, padding = 0 - Kernel size = (20, 20, 5), stride = (1, 1, 1), number of kernels = 50, padding = 0

2. Experiments

2.1 Experimental Configuration

Based on available datasets, we test classification accuracy across different fonts and simplified/traditional Chinese variants, where classification accuracy approximates reading performance. The datasets are summarized in . The THUCNews dataset is derived from historical data obtained from Sina News RSS feeds between 2005 and 2011. The original THUCNews contains 14 categories, but we removed four categories with insufficient samples to avoid imbalance issues, resulting in a 10-class dataset. The Toutiao news dataset was collected from the Toutiao App, with each item containing news titles and keywords. Data processing follows the methodology in reference [13]. provides sample sizes for training, validation, and testing for both datasets.

2.2 Hypothesis Testing Across Fonts

We categorize frequently used fonts in popular publications as common fonts, sourced from the Windows default font library, while uncommon fonts are sourced from Founder Type. Common fonts include: Song (SIMSUN),

regular script (STKAITI), bold (SIMHEI), Dengxian (DENG), and imitation Song (STFANGSO). Uncommon fonts include: FZShuTi (FZSTK), Founder Handwriting-Yangtze Running Script Simplified (FZZJ-CJXSJW), Founder DaCao Simplified (FZDCJW), Founder LuXun Simplified (FZLUXTJW), and Founder Handwriting-Ouyang Changdi Running Regular (FZZJ-OYCDXKJW).

2.2.1 Common vs. Uncommon Fonts We render both datasets in five common and five uncommon fonts, then run text classification tasks m times. Results in show that common and uncommon fonts exhibit different forms and reading performance differences.

[Figure 2: see original paper] displays the visual effects of Song, regular script, bold, Founder Handwriting-Yangtze Running Script Simplified, Founder LuXun Simplified, and Founder Handwriting-Ouyang Changdi Running Regular. [Figure 3: see original paper] shows the classification accuracy distributions for common and uncommon fonts on both datasets. presents the average classification accuracies.

Hypothesis testing results comparing reading performance between common and uncommon fonts are shown in :

Hypothesis: H_0 : No significant difference in machine reading performance between common and uncommon fonts.

For the Toutiao dataset: test statistic $t = 8.036$, degrees of freedom $df = 362.275$, two-tailed $p = 1.3266 \times 10^{-14}$. At $\alpha = 0.05$, $p < \alpha$, so we reject H_0 . The difference is statistically significant, indicating significant reading performance differences between common and uncommon fonts.

For the THUCNews dataset: test statistic $t = 2.300$, $df = 397.633$, two-tailed $p = 0.0220$. At $\alpha = 0.05$, $p < \alpha$, so we reject H_0 , confirming significant reading performance differences.

These results demonstrate significant reading performance differences between common and uncommon fonts, explaining why printed books typically select from a limited range of fonts. High-performance fonts facilitate reading and provide better experiences, while unsuitable fonts hinder reading and may affect book sales.

2.2.2 Among Common Fonts Although Song, bold, regular script, imitation Song, and Dengxian are all common fonts, reading performance differences persist. ranks their performance: on Toutiao, regular script achieves the highest performance and bold the lowest; on THUCNews, regular script and imitation Song tie for highest, with bold lowest. We further test whether significant differences exist between regular script and bold.

[Figure 4: see original paper] shows the fonts used for hypothesis testing. presents text classification results across different common fonts. [Figure 5: see original paper] displays accuracy distributions for bold and regular script.

Hypothesis testing results are shown in :

Hypothesis: H_0 : No significant difference in machine reading performance between regular script and bold.

For Toutiao: test statistic $t = -4.4227$, $df = 66.837$, two-tailed $p = 3.6756 \times 10^{-5}$. At $\alpha = 0.05$, $p < \alpha$, so we reject H_0 , confirming significant differences.

For THUCNews: test statistic $t = -3.7310$, $df = 48.971$, two-tailed $p = 0.0005$. At $\alpha = 0.05$, $p < \alpha$, so we reject H_0 .

These experiments confirm significant reading performance differences between regular script and bold, with regular script outperforming bold. Visually, regular script characters have disconnected structural components and clear forms, while bold characters have uniform stroke thickness. Bold text throughout an article makes every character prominent, causing visual fatigue. Regular script's less heavy strokes provide clearer visual perception, aligning with typographic conventions that use bold for headlines and regular script for body text to enhance readability.

2.3 Hypothesis Testing: Simplified vs. Traditional Chinese

Chinese employs two writing systems: simplified and traditional Chinese, with traditional characters generally having more strokes. Debates over their differences and relative efficiency have persisted for years. We compare their reading performance differences using our model framework, running text classification tasks 40 times on both datasets.

[Figure 6: see original paper] shows simplified and traditional Chinese fonts. [Figure 7: see original paper] displays accuracy distributions. presents reading performance results.

Hypothesis testing results are shown in :

Hypothesis: H_0 : No significant difference in machine reading performance between simplified and traditional Chinese.

For Toutiao: test statistic $t = 5.152$, $df = 53.304$, two-tailed $p = 3.8302 \times 10^{-6}$. At $\alpha = 0.05$, $p < \alpha$, so we reject H_0 , indicating significant differences.

For THUCNews: test statistic $t = 1.747$, $df = 70.683$, two-tailed $p = 0.0850$. At $\alpha = 0.05$, $p > \alpha$, so we fail to reject H_0 ; insufficient evidence exists for significant differences on this dataset.

Traditional Chinese emphasizes ideographic expression, with each character having etymological origins and detailed evolutionary processes where form and meaning align. Simplified Chinese reduces strokes and sometimes merges multiple meanings into one character, potentially affecting recognition efficiency. The Toutiao dataset shows significant differences, while THUCNews does not. We

hypothesize this occurs because Toutiao samples are shorter, yielding more precise accuracy calculations (and lower overall accuracies), whereas THUCNews samples are longer, diminishing performance differences. Since Toutiao results provide sufficient evidence to reject the null hypothesis, we conclude that significant reading performance differences exist between simplified and traditional Chinese.

Conclusion

Through sentence tensors, textual features can be extracted via multiple n -grams in both forward and reverse sequences. To investigate recognition differences across Chinese fonts and writing systems, we conducted text classification tasks on several datasets. Hypothesis testing based on classification accuracies reveals performance differences between common and uncommon fonts, significant differences between regular script and bold among common fonts, and significant differences between simplified and traditional writing systems.

These findings confirm that font selection in printed publications affects reading performance. However, our uncommon fonts were sourced only from Founder Type; the expression efficiency of other uncommon fonts requires further investigation.

References

- [1] Xiaoqing Ding. Review of Chinese character recognition research[J]. Acta Electronica Sinica, 2002, 30(9): 1364-1368.
- [2] Yonghe Fu. Analysis of four printing fonts[J]. Language Construction, 1993(5): 24-26.
- [3] Lijia Wang. Visual psychological effects of book typography design[J]. Journal of Heilongjiang Vocational Institute of Ecological Engineering, 2007(01): 124-125.
- [4] Mincong Huang. Quantitative analysis of simplified and traditional Chinese character evolution based on humanities computing[J]. Science & Technology Vision, 2012, 18: 64-67.
- [5] Jinliang Xie. Reflections on traditional and simplified Chinese characters[J]. Journal of Minjiang University, 2009(30): 45-49.
- [6] Zhiming Hu. Traditional and simplified characters in the history of Chinese character development[J]. Strait Educational Research, 2016(02): 43-47.
- [7] Shuangyi Wu. The imagistic relationship between Chinese character construction methods and visual communication design[J]. Contemporary Art, 2011, 000(002): 95-96.
- [8] Yexin Wan. Comprehensive research on Chinese characters and cognition from character recognition perspective[J]. Language Teaching and Linguistic

Studies, 2003(2): 72-79.

[9] Yijie Guan, Fuxi Fang. New progress in Chinese character recognition research in China[J]. Advances in Psychological Science, 2000, 8(2): 1-6.

[10] Wei Xin, Daodian Tao, Tao Huang, et al. Differences in transient myopia and recovery time between reading simplified and traditional Chinese characters[J]. Chinese Journal of Optometry Ophthalmology and Visual Science, 2012, 14(10): 581-583.

[11] Wei Xin, Pincheng Shen, Chenchen Zheng, et al. Differences in accommodative microfluctuation between emmetropic and myopic eyes reading simplified and traditional Chinese characters[J]. Chinese Journal of Ophthalmology, 2018(4): 288-

[12] Wenxiong Jin, Zuxiang Zhu, Mowei Shen. Effects of Chinese font types on judgment performance[J]. Applied Psychology, 1992, 7(3): 8-11.

[13] Bin Liu, Liang Wang, Guosheng Yin. Learning distributed sentence vectors with bi-directional 3D convolutions[C]// Proceedings of the 28th International Conference on Computational Linguistics, Barcelona, Spain (Online): International Committee on Computational Linguistics, 2020: 6820-6830.

[14] Jingyang Li, Maosong Sun. Scalable term selection for text categorization[C]//Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL), Prague, Czech Republic: Association for Computational Linguistics, 2007: 774-782.

Note: Figure translations are in progress. See original paper for figures.

Source: ChinaXiv –Machine translation. Verify with original.