ORIGINAL PAPER

Cover-based multiple book genre recognition using an improved multimodal network

Assad Rasheed[1] · Arif Iqbal Umar[1] · Syed Hamad Shirazi[1] · Zakir Khan[1] · Muhammad Shahzad[1]

Abstract

Despite the idiom not to prejudge something by its outward appearance, we consider deep learning to learn whether we can judge a book by its cover or, more precisely, by its text and design. The classification was accomplished using three strategies, i.e., text only, image only, and both text and image. State-of-the-art CNNs (convolutional neural networks) models were used to classify books through cover images. The Gram and SE layers (squeeze and excitation) were used as an attention unit in them to learn the optimal features and identify characteristics from the cover image. The Gram layer enabled more accurate multi-genre classification than the SE layer. The text-based classification was done using word-based, character-based, and feature engineering-based models. We designed EXplicit interActive Network (EXAN) composed of context-relevant layers and multi-level attention layers to learn features from books title. We designed an improved multimodal fusion architecture for multimodal classification that uses an attention mechanism between modalities. The disparity in modalities convergence speed is addressed by pre-training each sub-network independently prior to end-to-end training of the model. Two book cover datasets were used in this study. Results demonstrated that text-based classifiers are superior to image-based classifiers. The proposed multimodal network outperformed all models for this task with the highest accuracy of **69.09%** and **38.12%** for Latin and Arabic book cover datasets. Similarly, the proposed EXAN surpassed the extant text classification models by scoring the highest prediction rates of **65.20%** and **33.8%** for Latin and Arabic book cover datasets.

Keywords Book covers classification · CNN · Image classifiers · Multimodal learning · Text classifiers

1 Introduction

English idiom says, "don't judge a book by the cover" but at some point, we all have been guilty of it when encountering a book. A typical book cover is composed of two components, i.e., image and text. The text of the cover page includes the author's name and title, subtitle in some cases, and publisher's information. The other part is a graphic illustration or background image to create anticipation, expectation, or mood. It attracts the potential reader and communicates that the pages within are worth a reader's time and attention. It introduces the book to a potential reader without turning the first page. In addition to giving potential readers a subtle preview of the book, the cover also impart the book's genre. In this article, we focused on automatic book classification into genres using information from cover images.

The book industry is becoming digital very rapidly. According to an estimate, EBooks make almost 30 percent of the total sale of books. Instead of taking several minutes to find a book by casually scrolling through the libraries pursuing book titles, readers now scroll through the web pages in a second. The automatic classification of books without human interference in a short period of time would be utterly beneficial for retrieval, placement, promotion, and management systems. The questions then emerge as to which information is appropriate to automate the system. Books come with a wide variety of graphic designs even for the same genre, including miss leading background, similar color, nondescript, shared styles, and textual information. Unlike other classification and detection problems, the genres are not concretely defined.

In the presence of these challenges, we have suggested a classification model based on deep machine learning. Deep learning has been used widely in a variety of tasks

✉ Assad Rasheed
assad.mscs@gmail.com

1 Department of Information Technology, Hazara University, Mansehra, Pakistan