NewsImages Fusion: Bridging Textual Context and Visual Content in Media Representation

Dr.R.Priyadharsini[1], Arvind.V[1,*], Harish.J[1], P.VettriChezhian[1] and Mohanapriya.E[1]

[1]*Department of Computer Science and Engineering, Sri Sivasubramaniya Nadar College of Engineering, Chennai - 603110, Tamil Nadu, India*

Abstract

As the consumption of news content becomes increasingly visual, the evaluation of news images plays a pivotal role in media understanding and interpretation. This research addresses the challenges associated with the automated assessment of news images with the mapping of textual information using Convolutional Neural Networks (CNNs). The work leverages a comprehensive dataset of news images and proposes a CNN architecture tailored to the intricacies of media content. The research first delves into the existing landscape of news image evaluation, highlighting gaps and limitations in current methodologies. Motivated by the need for robust and efficient image assessment tools, our work focuses on the design and implementation of a CNN tailored for news media.

Keywords:NewsImages Fusion, Text-Image Relationship,image captioning

1. Introduction

In the contemporary landscape of digital media, news dissemination is increasingly characterized by the integration of visual content, with news images serving as crucial elements in shaping public perception. As society navigates an era inundated with information, the ability to assess the credibility, relevance, and impact of news images becomes paramount. This research addresses the imperative need for automated and efficient methodologies to evaluate news images, a challenge exacerbated by the sheer volume and diversity of media content. Online news articles are multimodal: the textual content of an article is often accompanied by a multimedia item such as an image. The image is important for illustrating the content of the text, but also attracting readers' attention. Research in multimedia and recommender systems generally assumes a simple relationship between images and text occurring together. For example, in image captioning [1] the caption is often assumed to describe the literally depicted content of the image. In contrast, when images accompany news articles, the relationship becomes less clear [2]. Since there are often no images available for the most recent news messages, stock images, archived photos, or even generated photos are used. An additional challenge is the wide spectrum of news domains, reaching from politics to economics to sports and to health and entertainment. The goal of this task is to investigate these intricacies in more depth, in order to understand the implications that it may have for the areas of journalism and news personalization. The task takes a large set of news articles paired with their corresponding images. The two entities have been paired but we do not know how. For instance, journalists could have selected an appropriate picture manually, generated an illustration using generative

MediaEval'23: Multimedia Evaluation Workshop, February 1–2, 2024, Amsterdam, The Netherlands and Online

*Corresponding author.

† These authors contributed equally.

priyadharsinir@ssn.edu.in (Dr.R.Priyadharsini); arvind2320028@ssn.edu.in (Arvind.V); harish2320045@ssn.edu.in (Harish.J); pvettri2320071@ssn.edu.in (P.VettriChezhian); mohanapriya2320034@ssn.edu.in (Mohanapriya.E)

AI, or a machine could have selected an image from a stock photo database. The image can have a semantic relation to the story but has not necessarily been taken directly at the reported event, nor event exist (in case of synthetic images). Automatic image captioning is insufficient to map the images to articles.

2. Related Work

The evolving landscape of multimedia content in news articles has spurred significant research efforts to understand and enhance the interaction between text and images. This section provides a comprehensive overview of the background and related work in this domain. Recent work by Lommatzsch et al. [3]has made substantial strides in bridging the "Depiction Gap" with the introduction of NewsImages. This online news dataset focuses on text-image rematching, offering valuable insights into the intricate relationship between news articles and their associated images. The authors highlight the challenges in accurately pairing textual and visual content, setting the stage for a deeper exploration. Garcin et al. [4] contribute to the discourse on recommendation systems, emphasizing the limitations of offline evaluations in predicting the performance of diverse recommendation techniques. Their study underscores the need for sophisticated models that incorporate novelty into recommendations and questions the reliability of Click-Through Rate (CTR) as a sole metric, especially for popular items. These findings resonate with the challenges encountered in multimedia recommendation tasks. Ge and Persia [5] provide a comprehensive survey of multimedia recommender systems, shedding light on challenges and opportunities in this domain. Their work spans across research communities, delving into the intersection of multimedia information systems and recommender systems. Categorizing papers based on recommender algorithm, multimedia object, and application domain, the survey identifies key features that pave the way for potential research opportunities. Continuous evaluation in large-scale information access systems is explored by Hopfgartner et al.[6]. They advocate for the adoption of living labs, presenting a case for ongoing evaluation. The relevance of their approach extends to the evaluation of multimedia recommendation systems, providing a framework for refining algorithms and adapting to evolving user preferences. Hossain et al.[1] contribute to the landscape of multimedia understanding with a comprehensive survey of deep learning for image captioning. The survey encompasses the evolving techniques used to bridge the semantic gap between textual descriptions and visual content, a challenge inherent in the news domain explored by our work. The stream-based recommender task overview presented by Lommatzsch et al. [7] at CLEF 2017 is particularly relevant to our study. It emphasizes the need for ongoing evaluation and education in the field of recommender systems, aligning with our goal of refining algorithms based on insights gained from continuous assessments. Oostdijk et al. [2] contribute insights into the connection between text and images in news articles. Their work offers new perspectives for multimedia analysis, which resonates with our exploration of the impact of image content on consumer engagement in the context of social media posts related to major U.S. airlines and compact SUV models. Lops et al. [8] provide a comprehensive survey of content-based recommender systems, addressing fundamental aspects characterizing this category of systems. Their exploration of techniques for representing items to be recommended aligns with the challenges posed by diverse multimedia content in news articles. Li and Xie [9] leverage observational data to explore the impact of image content on consumer engagement with social media posts. The study introduces pathways through which image content influences engagement, aligning with our investigation into the interaction between text and images in the realm of news articles. Finally, Liu, Han, and Chilton [10] present a significant contribution to the field with their work

on multimodal image generation for news illustration. Their exploration of generating images for news articles aligns with the overarching theme of our study, emphasizing the importance of understanding the relationship between textual and visual content.

3. Objective

Develop a comprehensive dataset of news images representative of diverse media contexts. Design and implement a CNN architecture tailored to the specific characteristics of news images. Evaluate the performance of the proposed CNN against benchmark methods using carefully selected metrics. Provide insights into the potential applications and limitations of CNNs in the realm of news image evaluation.This task explores the relationship between text and images in news articles. A dataset includes paired news articles and images, with undisclosed pairing methods—whether manual selection, generative AI, or automatic machine choice. The images may have semantic ties to the story but need not depict the reported event. Conventional image captioning falls short in accurately mapping images to articles in this diverse context.This dataset is curated from web news articles, providing crucial details for each article, including URL, Title, and initial news text. Paired with each article is a corresponding image, and the dataset covers both English and German articles, with machine-translated versions for the latter.With a 1:1 relationship, the dataset follows a structure akin to NewsImages 2022 data structures.

4. Methodology

A Convolutional Neural Network (CNN) model is employed for the task. The model consists of convolutional layers (conv1 and conv2) with ReLU activation functions and max-pooling layers (pool1 and pool2) for feature extraction. Fully connected layers (fc1 and fc2) follow the convolutional layers for classification. The model is designed for binary classification with two output classes. A Single Channel Wrapper class is implemented to enable the use of the CNN model with single channel images. It takes an existing CNN model and duplicates the single channel to create a three channel image. The NewsDataset class is defined to load and preprocess the dataset. It reads textual data from a CSV file (rt-train-textImg.csv) containing hash values and image paths. For each entry, it loads the corresponding image, or a default image if the path is missing or invalid. The transforms.ToTensor() transformation is applied to convert PIL images to PyTorch tensors. This transformed image is used during the training loop for the CNN model. The CNN model is trained using a binary cross-entropy loss(nn.CrossEntropyLoss) and Adam optimizer (optim.Adam).The training loop runs for a specified number of epochs. For each epoch, the model is trained on batches of data from the train loader. Accuracy is computed, and Precision@K is calculated for different values of K. A test dataset (rt-test-text-1.csv) is loaded using the NewsDataset class. The test dataset is processed using a DataLoader (test loader) with a batch size of 1. Predictions are generated for each test sample using the get predictions function. Predictions are then written to an output file (output.txt), following a specified format.

5. Evaluation Methodology

The computation involves the Mean Reciprocal Rank (MRR) as the official metric and a series of Precision@K scores, where K takes values from 1, 5, 10, 20, 50, 100. The primary metric for the task is the average MRR, providing insights into the average position at which the linked

image appears.Additionally, the average precision scores offer a comprehensive evaluation of performance across various ranks within the list.

6. Results and Analysis

A series of experiments was conducted, The proposed system was evaluated using precision,MRR metrics and with overall accuracy. The results are depicted in the below table 1.The accuracy was found out to be 50.37%.The accuracy(Using MRR) was found out to be 100.00%.

Precision@K Values	Values
1	0.1505
5	0.1504
10	0.1541
15	0.1550
20	0.0994
100	0.0497
Accuracy	0.5037
MRR	1.0000

Table 1
Precision@K-values and Accuracy

from 1,we observe that the NewsImages Fusion model has effectively paired news images with their corresponding text, achieving a 100% accuracy using Mean Reciprocal Rank (MRR). For smaller image sets, precision@K closely aligns with accuracy.Nevertheless, with a growing number of images, precision values exhibit a notable decline.

7. Discussion And Outlook

The insights gathered from the referenced works pave the way for a comprehensive discussion on the intricate relationship between text and images in news articles. The diverse perspectives offered by researchers in multimedia recommender systems, continuous evaluation, image captioning, and content-based recommendation systems provide a rich foundation for our analysis.

8. Acknowlegment

We would like to express our gratitude to our guide Dr.R.Priyadharshini,for providing the guidance in doing this project and also for providing insites for our project.We also like to extend our gratitude to Marc Gallofre from Bergen, Norway, for their invaluable support during the dataset creation process.

References

[1] M. Z. Hossain, F. Sohel, M. F. Shiratuddin, H. Laga, A comprehensive survey of deep learning for image captioning, ACM Computing Surveys (CSUR) 51 (2019) 1–36.
[2] N. Oostdijk, H. van Halteren, E. Baṣar, M. Larson, The connection between the text and images of news articles: New insights for multimedia analysis (2020) 4343–4351.

[3] A. Lommatzsch, B. Kille, Y. Zhou, J. Tesic, C. Bartolomeu, D. Semedo, L. Pivovarova, M. Liang, M. Larson, Newsimages: Addressing the depiction gap with an online news dataset for text-image rematching (2022) 227–233.
[4] F. Garcin, B. Faltings, O. Donatsch, A. Alazzawi, C. Bruttin, A. Huber, Offline and online evaluation of news recommender systems at swissinfo.ch (2014) 169–176.
[5] M. Ge, F. Persia, A survey of multimedia recommender systems: Challenges and opportunities, International Journal of Semantic Computing 11 (2017) 411–428.
[6] F. Hopfgartner, K. Balog, A. Lommatzsch, L. Kelly, B. Kille, A. Schuth, M. Larson, Continuous evaluation of large-scale information access systems: a case for living labs (2019) 511–543.
[7] A. Lommatzsch, B. Kille, F. Hopfgartner, M. Larson, T. Brodt, J. Seiler, Ö. Özgöbek, Clef 2017 newsreel overview: A stream-based recommender task for evaluation and education (2017) 239–254.
[8] P. Lops, M. De Gemmis, G. Semeraro, Content-based recommender systems: State of the art and trends (2011) 73–105.
[9] Y. Li, Y. Xie, Is a picture worth a thousand words? an empirical study of image content and social media engagement, Journal of Marketing Research 57 (2020) 1–19.
[10] V. Liu, H. Qiao, L. Chilton, Multimodal image generation for news illustration (2022). doi:10.1145/3526113.3545621.