Original Research Article

Research on deep learning-based intelligent classification and tagging technology for media content

Hao Liu

Communication University of China, Nanjing, Jiangsu, 211172, China

Abstract: Focus on the media content intelligent classification and tag technology technology frontier, our country is facing a series of technical challenges, the application of deep learning in the field of the application status and the development trend, the basic principle of deep learning of the system, the characteristics of media content classification, media type adaptation and classification development of the comprehensive analysis, the application of multimodal feature extraction and fusion technology principle, method and its application in information processing for a comprehensive analysis.

Keywords: Deep learning; Media content classification; Multi-modality; Feature extraction; Data fusion

Foreword

In today's rapid Internet technology, the media content is facing explosive expansion pressure, including text, images, audio and video diversified information presentation, analysis of how to accurate classification and label classification, realize the content management refinement and efficiency, has become the key to break through the current, in the transfer of intelligence flow, many technical problems scattered in the forefront of science and technology research area, caused significant restrictions to the progress in the field.

1. Deep-learning concepts

(1) Concept

Deep learning constitutes a core branch discipline in the field of machine learning. This system uses multilayer neural network to simulate the operation mode of human brain neurons. See **Figure 1**; in image multi-label classification and new class detection scenarios, it starts with training image data with known labels, e. g

$$\{X_0 = [X_{-n+1}, X_{-n+2}, \dots, X_0] \in X\},\$$

Mining feature mapping and its internal laws. In the face of unknown streaming media images, the system can conduct prediction and detection analysis according to the learning mode. The model continues to learn and strengthen the newly collected data, such as updating the cognition of labels when dealing with new examples

$$\{X_t = [X_{t-1}, X_t] \in X\},\$$

and laying a solid foundation for tasks related to media content.



Figure 1. System structure.

(2) Principle

The research of deep learning technology is based on the artificial neural network technology, The system consists of three levels: input layer, hidden layer and output layer, Data import link is imported by the input layer^[1], Activation function is the core mechanism by which neurons achieve data transformation, Analysis of the hierarchical information transfer and processing mechanism, Continue to dig and extract higher level abstract indicators^[2], A back-propagation algorithm was used to adjust the connection weights of the middle layer neurons, To optimize its weight structure, Efforts to narrow the gap between the model prediction results and the actual labeling error, Thus significantly improving the expressiveness of the model, Implement the detailed analysis and appropriate treatment of media materials.

2. Accurate classification of media content based on the deep learning model

(1) Classification basis based on characteristics

In the field of text communication media, the feature space constructed based on word vector is the basic support for the implementation of classification tasks. Semantic spacing and context correlation features can reveal the core tendency of text, similar to TF-IDF features reveal the significant attributes of text keywords and highlight their core role, providing support for the third classification. Image processing technology uses features such as color histograms, texture attributes and shape descriptors for data analysis and characterization, and image classification and recognition rely on different combinations of these features. In the media research of video ^[4], the motion characteristics between frames are important, and the optical flow field characteristics are the key in the video compression technology. Deep learning system builds a mapping model between media content and category by learning relevant characteristics.

(2) Adjustment of the model to the media type

In text classification tasks, the recurrent neural network and its variant models are favored for their ability

to process sequence data, read and capture text vocabulary, and analyze long-distance semantic dependencies, among which the long-term short-term memory network (LSTM) is particularly efficient in processing text semantic consistency. In the field of image recognition, convolutional neural network (CNN), as the mainstream model, automatically extracts local and global features of images^[5] through the convolution layer and pooling layer. Models with different depths such as VGG and ResNet can meet the image classification requirements of different complexity and be used for target tracking. In the field of video content classification, the fusion technology of CNN and RNN or LSTM is widely applied to realize the comprehensive classification of spatial and temporal dimension of video content^[6], and various models play a key role in the classification of corresponding media types.

(3) Classification and refinement and expansion

In the field of news communication, it not only classifies traditional categories such as politics, economy and culture, but also focuses on policy interpretation and industry segment discussion. In the category of social media, entertainment and life sharing occupy the core, and the research vision extends to emerging hot topic tracking and content classification of niche interest community^[7]. Image recognition technology has expanded from basic fields to medical and industrial image analysis, and video content classification has extended from traditional entertainment education videos to professional monitoring and tactical analysis of sports events. Deep learning models have continued to broaden the boundary of media content classification with strong learning ability.

Media type	Classification is based on the key characteristics	Adaptation model
Text media	Word vectors, semantic distances, and TF-IDF features	RNN/LSTM
Image media	Color histogram, texture features	CNN (VGG, ResNet, etc.)
Video media	Static image features and optical flow field characteristics	CNN+RNN/LSTM

Table 1. Comparison table of the classification situation of different media types.

3. Deep learning media content labeling technology integrating multimodal information

(1) Multimodal feature extraction technology

When interpreting the connotation and extension of text mode, using the vocabulary embedding techniques such as Word2Vec and GloVe, text map to low-dimensional vector space to explore the vocabulary semantic properties; BERT pre-training algorithm system based on transformer architecture can dynamically extract text features in a specific context, covering multifaceted language information such as grammar, semantic and pragmatic. During image processing, convolutional neural network (CNN) relies on the combination of convolution layer and pooling layer to screen and summarize local and global features such as color, texture and shape from image pixel data, and different structure models can extract image features of different complexity; audio processing uses MFCC technology, and describes them with frequency and color dimension.

(2) Multimodal data fusion technology

In the early technology, the fusion technology based on feature splicing is widely used. It directly connects the multiple data feature vectors such as text, image and audio to form the extension vector for subsequent model

processing. It is simple in operation but insufficient to consider the internal connection of mode^[9]. Subsequently, the technical path driven by the fusion model appeared, and the confidence network architecture and multilayer neural network were built with the help of deep learning, and the modal feature correlation weight was independently acquired during the system training, realizing the intelligent fusion breakthrough^[10]. At present, the fusion technology based on the principle of attention attracts attention. The model dynamically adjusts the modal feature weight according to the task, strengthens the extraction of key information, and improves the fusion level to facilitate accurate label prediction.

4. Epilogue

In the field of media content classification and labeling technology, deep learning has brought about technological innovation, but there are many obstacles in data processing, model building and content understanding. However, with the continuous progress of technology, it is expected to achieve efficient and accurate system prototypes, upgrading the intelligent processing technology of media content to a new stage of development.

About the author

Hao Liu, born in April 1985, female, native of Fuyang City, Anhui Province, Han nationality. She is a graduate student, a senior engineer, and her research interests are: image processing, artificial intelligence applications.

References

- Raheel A .Emotion analysis and recognition in 3D space using classifier-dependent feature selection in response to tactile enhanced audio-visual content using EEG[J].Computers in Biology and Medicine,2024,179108807-108807.
- [2] Zhang Z ,Liew K ,Kuijer R , et al.Differing Content and Language Based on Poster-Patient Relationships on the Chinese Social Media Platform Weibo: Text Classification, Sentiment Analysis, and Topic Modeling of Posts on Breast Cancer.[J].JMIR cancer,2024,10e51332-e51332.
- [3] Hwai T T ,Dewi K V ,Fabio C .A comprehensive review of cyberbullying-related content classification in online social media[J].Expert Systems With Applications,2024,244122644-.
- [4] Okyaz E ,Jiyong T L ,Jessie G, et al.Conceptual Framework and Documentation Standards of Cystoscopic Media Content for Artificial Intelligence.[J].ArXiv,2023,
- [5] Mohsan A ,Mehdi H ,Kashif K , et al.Social media content classification and community detection using deep learning and graph analytics[J].Technological Forecasting & amp; Social Change,2023,188
- [6] Sajad D ,Mehdi N .A Novel Hybrid Multi-Modal Deep Learning for Detecting Hashtag Incongruity on Social Media[J].Sensors,2022,22(24):9870-9870.
- [7] A. C J ,K. D .Content-Based Image Recognition and Tagging by Deep Learning Methods[J].Wireless Personal Communications,2021,123(1):813-838.

- [8] Yang Z ,Lin Z .Interpretable video tag recommendation with multimedia deep learning framework[J]. Internet Research,2021,32(2):518-535.
- [9] Google LLC; Patent Application Titled "Sentiment-Based Classification Of Media Content" Published Online (USPTO 20200327157)[J].Computer Weekly News,2020,2648-.
- [10] Sambhaji D R .Social Media Content Analysis and Classification Using Data Mining and ML[J]. International Journal of Data Analytics (IJDA),2021,2(2):75-84.