Academic Journal of Science and Technology
ISSN: 2771-3032 | Vol. 6, No. 3, 2023

Large-Scale Networked Visual Data: Current Research and Future Trends in Associative Inference and Semantic Comprehension

Yu Wu[1, a]

[1]Jilin University, Changchun, China
[a]3829103891@qq.com

Abstract: Aiming at the fact that most of the existing visual data comprehension deals with individual visual objects in isolation and focuses on their inherent characteristics, and pays insufficient attention to the characteristics of network heterogeneous distribution and interconnections, which leads to difficulties in solving low computational efficiency, and understanding of low-level semantics, etc., we elaborate on the hotspots of the research of associative inference and semantic comprehension of large-scale networked visual data, analyze the current status of the research at home and abroad, and look forward to the development trend in this direction. We analyze the current research situation at home and abroad and make an outlook on the development trend of this direction.

Keywords: Visual cognitive mechanisms, Machine learning, Association understanding, Semantic structure.

1. Introduction

Digital media data are increasingly networked and socialized, and have a significant impact on all aspects of social life. Unlike traditional structured and unstructured data, digital media data are embodied in various forms such as text, images and videos. How to semantically analyze and deeply utilize the network data is an important frontier issue of common concern in the fields of network communication and information technology, facing challenges such as dimensionality disaster and semantic divide, and this trend has become a bottleneck that seriously restricts the development of digital media technology applications.

Network visual data pose new challenges to the theory of visual computing, visual data association understanding provides new research ideas for solving the challenges faced by visual media data understanding and is expected to give rise to new theories and methods of visual data semantic understanding, focusing on and utilizing such data association presentation characteristics is quietly changing the traditional visual computing focus on a single object research ideas. In addition, the exponential growth of digital information will be multidisciplinary knowledge in a variety of ways associated together, the analysis and understanding of these associations have led to a new development in the study of visual data understanding, which has gone beyond the scope of conventional data mining, and has helped to produce a new visual data processing and pattern analysis techniques that can break through the bottleneck of visual media data semantic understanding technology.

For polymorphisms existing in open, massive, complex and dynamic thematic extraction and knowledge mining of associated visual data from changing Internet platforms for content aggregation and integrated synthesis, facilitating the Deepening and advancing the semantic understanding of hot and sensitive topics and major events reflected in cyberspace, these expected research results have broad application prospects in national security and safety, development of advanced productive forces and improvement of people's living standards.

2. Literature Review

Representation and computation are two essential core issues in visual data association understanding. The current main research can be summarized into four aspects: visual knowledge expression model, feature extraction and representation, feature fusion and processing, and feature and semantic association attribute maps, which respectively realize the modeling and learning of visual objects, feature representation of high-dimensional heterogeneous data, feature dimensionality reduction and association, and feature to high-level semantic mapping. The first two belong to the representation problem and the latter two belong to the computational problem.

2.1. Relevant domestic and international government funding for this research

Currently, a large number of research works on visual data association understanding have been carried out at home and abroad. For example, the U.S. National Natural Science Foundation. In 2003, more than $12 million were funded for research in this area; In 2007, the U.S. Defense Advanced Research Projects Agency (DARPA) launched an AI research project called Bootstrapped Learning, which involves natural language processing, visual data cognition and content understanding, etc., with the aim of deeply studying the mechanisms and processes of human cognition and visual content understanding. The aim is to deeply investigate the mechanism and process of human cognition and visual content understanding, which is considered a new kind of AI research of machine learning approaches to explore.

China's 973 Program has successively supported projects such as "Image, Speech, Natural Language Understanding and Knowledge Mining", "Theories and Methods of Digital Content Understanding", "Theories and Methods of Intelligent Processing of Visual Media", "Theories and Key Technologies of Unstructured Information Processing Based on Visual Cognition" and so on. ", "Theory and Key

Technology of Unstructured Information Processing Based on Visual Cognition" and other projects, and the National Natural Science Foundation of China has also launched the "Cognitive Computing of Visual and Auditory Information" major research program. The implementation of these projects has promoted China's research in this field, and produced a number of results that have reached the international level. However, since the understanding of visual data association is a very challenging problem, the existing research still needs to conduct in-depth study on the organic combination of the characteristics of human visual cognition, the inherent characteristics of visual objects and the characteristics of visual computing.

2.2. Modeling and Learning of Visual Semantic Objects

The core of the representation problem is the modeling and learning methods for visual objects. Existing research ideas can be summarized as obtaining a hierarchical model for representing holistic semantics by improving the representation ability of feature and generative visual object models, the discriminative ability of discriminative models, and the performance of learning algorithms.

Visual data has properties that increase the difficulty of visual object representation. In terms of visual knowledge representation, visual words (visual words) and visual grammar [1] have been proposed for representing image and video data. However, the room for performance improvement in designing visual object models using only the underlying cues with a bottom-up approach or a top-down approach based on limited a priori information is extremely limited, and there is a need to integrate different cues under the framework of a unified generative model integrating a priori knowledge and data-driven learning generalization theory in order to improve the descriptive capability of visual objects and the efficiency of learning algorithms.

In terms of feature extraction and representation, heterogeneous features extracted from traditional text, image and video representations are often collapsed into high-dimensional vectors, which not only leads to dimensionality disaster, but also loses the structural constraints present in the original data. Secondly, there is not a simple and clear correspondence between the underlying description and the higher-level semantics, e.g., one cannot understand the theme of a video just from the underlying features such as color and texture, which is also an important reason for the semantic divide.

Recently, researchers have begun to gradually realize that machine learning may be one of the ways to solve the semantic divide problem, and have begun to work on visual object modeling from the perspective of visual object categorization and learning research. Initially, the typical approach is to consider the high-level semantic keywords or concepts as mutually independent classifications, and then based on the extracted underlying features, classifiers for these concepts are investigated by algorithms such as Support Vector Machine (SVM), Naive Bayes, and so on. Since then, some researchers have proposed the idea of using machine learning methods to learn and establish the mapping relationship between the underlying features and the semantics, and estimate the joint probability distribution information between the underlying features and the semantic keywords to establish the connection between them by learning the training sample set.

In summary, the current research on visual object modeling and learning is facing prominent problems: i) whether the feature representation hierarchy is reasonable and sufficient; ii) how to achieve a global semantic description of visual data. The lack of research on these two issues directly affects the progress of visual data understanding research.

2.3. Computational problems for intrinsic properties of visual data

After establishing the visual object representation model and learning algorithm, we need to solve the complex computational problems such as feature dimensionality reduction and correlation, and feature-to-higher-level semantic mapping due to the high-dimensional, heterogeneous, and multimodal characteristics of visual data.

The first is the dimensionality reduction of high dimensional features. Too high dimensionality can lead to the problem of Curse of Dimensionality. Professor Donoho of Stanford University has listed high-dimensional data analysis and processing as a hot topic. Recently, it has been widely recognized that the intrinsic dimensionality of visual data is usually much lower than the dimensionality of the feature vectors, so it is hoped that efficient computation can be achieved by dimensionality reduction of high-dimensional data. In addition to the classical DC), which is a method for analyzing visual data.

Factorization (NMF) method for face recognition and image compression on matrix computation [4]. "Element rearrangement (Element Rearrangement)" study also shows that maintaining the topology of the complex object itself during the computation process will achieve better results [5].

These data structures have become a hot research topic because they can describe the topology between the components of complex visual objects and effectively improve the computational efficiency.

Finally, it is the mapping of features to high-level semantics. Context-associated multimodal features help to understand the semantics embedded in media data, so how to realize the effective fusion of different modal features through collaborative computation is also a key issue. Traditional visual data fusion processing mainly adopts two approaches: feature layer fusion and decision layer fusion. These methods have their limitations, and there is a need to study the theory and methods of multimodal information fusion suitable for visual data understanding in networked environments.

In summary, there have been many studies on computational methods for inherently high-dimensional, heterogeneous and multimodal visual data, such as high-dimensional data dimensionality reduction, heterogeneous data processing, multimodal information synthesis, etc., but there are still no synergistic and efficient computational techniques and methods that can be adapted to the understanding of visual data.

3. Analysis of Development Trends

The great demand for large-scale visual data understanding in the network environment presents two distinct and intertwined development trends: (1) combining new theories and methods of visual cognitive mechanisms, considering the two essential core issues of visual

data understanding, representation and computation, as a whole, and promoting the mutual development of visual data understanding and cognitive mechanisms; (2) combining network data, deeply exploring and utilizing different levels and rich correlations in high-dimensional, heterogeneous network visual data, and promoting the development of feature extraction, cognitive representation, machine learning, and fusion analysis in the process of visual data computing from relatively independent to collaborative and complementary. (ii) Combined with network data, we deeply explore and utilize different levels and rich correlations in high-dimensional and heterogeneous networked visual data, so as to promote the development of feature extraction, cognitive representation, machine learning and fusion analysis in the process of visual data computation from relative independence to the direction of mutual synergy and mutual complementation.

3.1. Visual Cognition Mechanisms and Their Processing Models

The research on the mechanism of holistic grasping and hierarchical organization in the cognitive process has made great progress, and the results of these researches provide an important biological foundation and scientific basis for visual computing technology based on the mechanism of visual cognition [6], e.g., in the process of face recognition, the object is expressed as an undecomposable whole rather than a simple local combination. Therefore, visual data for semantic analysis and understanding requires a holistic semantic description of its content. Although the combination of local features or global-scale features can describe and differentiate visual objects to a certain extent, target occlusion, background interference to and changes in light, angle, and affine transformation are more sensitive, and their stability and distinguishability are not high.

Accordingly, deep learning methods such as Boltzmann machine, Deep Belief Nets (Deep Belief Nets) and Convolutional Neural Networks (CNNs) have successively appeared in the simulation of layer-by-layer gradual learning process of hierarchical organization mechanism, and these methods have achieved better recognition results than shallow learning. The theory of topological perception, which has been systematically developed in recent years by the laboratory led by Academician Chen Lin of the Chinese Academy of Sciences, holds that the visual process starts from a large range of properties, which can be described by topological properties, and that the geometrically invariant properties of each level, with topological properties as the most basic level, are the basic units of the perceptual information characterization of the figure. In addition, J. B. Tenenbaum et al. explored the relationship between the neural system and low-dimensional cognitive concepts embedded in high-dimensional datasets from a cognitive point of view, pointing out that high-dimensional information is generally embedded in a low-dimensional manifold during human cognitive processes, and that cognitive processes to a large extent recognize a variety of things by means of this low-dimensional manifold.

In addition, to cope with the wide range of visual objects in the networked environment

sex, the mechanism of selective attention in the biological visual system has been a hot research topic in visual cognitive modeling, which is currently focused on bottom-up computational models of salience, and many computational models have been proposed in recent years [7]. American scientists analyzed the rat cortex and found that its brain perceives external information in an internet-like structure, rather than the top-down theory that dominates experimental neuroscience [8]. Corresponding studies have also shown that the human brain is good at using the attention model in the process of processing external data, selecting differentiated attribute information from redundant information, i.e., taking into account the structural a priori knowledge of the classification model itself, and applying different constraints on regularization factors during the learning process to obtain sparse representations of the data itself and its embedded manifold topology.

In summary, cognitive science has made a more in-depth overview of the global characteristics of the information processing process, but the detailed mechanisms of realization are not yet sufficiently researched; neuroscience has carried out in-depth research on cognitive micro-mechanisms centered on neurons and neural circuits, and it has been able to partially explain hierarchical representations, but it is difficult to comprehensively and systematically expound on the holistic issues; information science, although it can link hierarchical and holistic with specific applications, is often limited by specific application goals, but the integration of the results of different levels of content comprehension is insufficient. correlate hierarchy and wholeness with specific applications, but are often limited by specific application goals, incomplete research on visual description and understanding of different levels, and insufficient integration of the results of understanding different levels of content.

3.2. Semantic Understanding of Visual Data Combined with Web Information and Machine Learning

Visual data is increasingly socialized and networked, and a large number of digital media websites, represented by Flickr and LabelMe, now allow online users to label images to improve the understanding of visual data. These labeling information is a powerful complement to the visual feature information, and also causes a research boom in semantic understanding of visual data combined with web information [9].

In the face of the heterogeneous format, unstructured content, and huge data volume of online media data, the existing visual computing for feature processing is extremely lacking in multilevel feature association and cooperative and efficient computational techniques and methods. Therefore, the research on visual data understanding in the network environment shows the following development trends: to construct higher-level feature operators to extract middle and high-level features, to establish multi-level features and high-level semantic mapping and correlation of heterogeneous visual data, to form an efficient computation method for high-dimensional heterogeneous multimodal data, and to reduce the semantic gap in visual data understanding; to realize the extraction of multi-level feature information of content, to establish content structure and to establish content structure based on content's spatial and temporal correlation, and to realize the extraction of content information based on content's temporal and spatial correlation. We can extract multi-level feature information of content, establish content structure based on spatio-temporal correlation, and correlate bottom-level features, middle-level features and high-level semantics, so as to achieve the purpose of understanding user's intention and realizing more accurate and fast content analysis, retrieval and browsing.

In addition, the increasing number of online media resources and the demand for online social computing [10] make it a great challenge to directly apply traditional machine learning theories and methods for semantic understanding. Prof. Tom Mitchell of Carnegie Mellon University pointed out in an article in Science that, unlike traditional research on mining historical data, research on a variety of media data closely related to people's daily lives is not the same as traditional research on mining historical data. Analyzing and processing data is an important trend in the future development of machine learning and search engines. Massive data means more difficult challenges for scientific research, such as Science in February 2011 published a special issue of Dealing with Data that: the collection, maintenance and use of massive data has become a major aspect of scientific research, and better organization and use of these data will help us to turn the great opportunities into reality.

Accordingly, the computability of massive data has become an issue that deserves attention. In the ICML08 workshop, experts concluded that: large-scale data = complex learning models. Recently, algorithmic complexity smoothing analysis establishes a framework for algorithms to run complexity computation in real time by perturbing the input data, instead of evaluating the algorithm performance in the worst-case scenario in the traditional way, and this research not only won the Gödel Prize of Theoretical Computer Science 2008, but also was selected as the NSF 3 Highlights of the Year in Computing and Information Science 2003, which aroused great interest from related scholars. This research not only won the 2008 Gödel Prize in Theoretical Computer Science, but also was selected as one of the 2003 Highlights in Computer and Information Science by the NSF, attracting great interest from scholars.

4. Conclusion

On the one hand, the study of semantic association and migration of visual data in network environment starts from the core problem of visual data understanding and focuses on the association relationship between visual data in network environment, which will help to break through the bottleneck of semantic understanding technology of visual perceptual data, improve the level of semantic understanding of network media data, and enrich and develop the traditional visual computing theory; on the other hand, it will also promote the development of data mining, machine learning, artificial intelligence, multimedia technology, and provide theoretical and technical support for the efficient processing, utilization, management and storage of massive and heterogeneous network media data. On the other hand, it will also promote the development of data mining, machine learning and artificial intelligence, multimedia technology, and provide theoretical and technical support for the efficient processing, utilization, management and storage of massive and heterogeneous online media data.

References

[1] Chen, S., Li, K., Fu, H., Wu, Y. C., & Huang, Y. (2023). Sea Ice Extent Prediction with Machine Learning Methods and Subregional Analysis in the Arctic. Atmosphere, 14(6), 1023.

[2] Surden, H. (2014). Machine learning and law. Wash. L. Rev., 89, 87.

[3] Zhang, X., Li, H., Jian, Y., Fu, H., Wang, Z., & Xu, M. (2022). Vernacular or modern: Transitional preferences of residents living in varied stages of urbanisation regarding rural landscape features. Journal of Rural Studies, 95, 95-108.

[4] Carbonell, J. G., Michalski, R. S., & Mitchell, T. M. (1983). An overview of machine learning. Machine learning, 3-23.

[5] Fillo, M., Keckler, S. W., Dally, W. J., Carter, N. P., Chang, A., Gurevich, Y., & Lee, W. S. (1997). The m-machine multicomputer. International Journal of Parallel Programming, 25, 183-212.

[6] Avanzo, M., Wei, L., Stancanello, J., Vallieres, M., Rao, A., Morin, O., ... & El Naqa, I. (2020). Machine and deep learning methods for radiomics. Medical physics, 47(5), e185-e202.

[7] Zantalis, F., Koulouras, G., Karabetsos, S., & Kandris, D. (2019). A review of machine learning and IoT in smart transportation. Future Internet, 11(4), 94.

[8] Meyer, P., Noblet, V., Mazzara, C., & Lallement, A. (2018). Survey on deep learning for radiotherapy. Computers in biology and medicine, 98, 126-146.

[9] Tang, B., Fu, H., & Wu, Y. C. On reservation prices in the all-pay contest with correlated information. Managerial and Decision Economics.

[10] Tang, J., Tang, L., Wu, H., Wu, Y., Sun, H., Zhang, H., ... & Xia, K. (2021). Towards on-demand heralded single-photon sources via photon blockade. Physical Review Applied, 15(6), 064020.