Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 199 (2022) 392–398

Procedia Computer Science

www.elsevier.com/locate/procedia

The 8th International Conference on Information Technology and Quantitative Management (ITQM 2020 & 2021)

A Review on the Construction of Business Intelligence System Based on Unstructured Image Data

Linzi Zhang[a,b,c], Zhiquan Qi[a,b,c*], Fan Meng[d]

[a]School of Economics and Management, University of Chinese Academy of Sciences, Beijing, 100190, China
[b]*Research Center on Fictitious Economy & Data Science, Chinese Academy of Sciences, Beijing 100190, China*
[c]*Key Laboratory of Big Data Mining and Knowledge Management, Chinese Academy of Sciences, Beijing 100190, China*
[d]*Department of Infomation Management, Peking University, Beijing, 100871, China*

Abstract

Business Intelligence (BI) has been an emerging significant issue in modern government and corporate functions, which attracts the attention of both researchers and merchants. Using the raw data from customers, BI provides effective decision-making knowledge for various cooperate governance operations with the method of Data Warehouse and Data Mining. As the explosive growth of data volume, it has evolved into using unstructured data including image and text in business intelligence system, from the initial descriptive analysis of computational statements. This paper will review the methods of storing and analyzing the unstructured image data used in the construction of BI system, including traditional image processing and novel deep learning methods. In each method, the specific algorithms and application scenarios will be summarized. Together, we will discuss the possible innovative progress of BI system construction and its future trends.

Keywords: Business intelligence, Unstructured image data, Big data fusion

1. Introduction

Business Intelligence (BI), transforming data into information and further into knowledge, creates conditions and environment for effective decision-making process [1]. BI has been an emerging hot spot in the field of cooperate governance, since the development and growth of a company is highly important to its managers, investors, shareholders and even governments [2]. This research concept can be traced back to 1989, when IBM seek for commercial proposals to Gartner Group, one of a top consulting company. Since 1990's, Business Intelligence has been extensively applied in companies to discover the intelligent knowledge and offer decision-making support [3], along with the enterprise unique BI system coming into being.

Using the data warehouse, online analysis and data mining technology, BI systems deal with and analyze business data in a specific application field from different industry [4], summarizing effective and tailored

10.1016/j.procs.2022.01.048

solutions from complex and scrambled raw data. This system helps enterprise to make quick response and more reasonable business decision in the face of the rapid change of business environment [5]. Based on the development of data warehouse technology, BI systems consist of data collection, data preprocessing, data warehouse building, data analysis and result display [6].

Among them, as the premier stage, data collection is of great importance. Limited by the technology of data analysis, the system used to focus on customer statements such structural data and make basic descriptive analysis [7]. As the markets become more complex, these analyses were far from sufficient. Through some data mining and coding methods, some unstructured data like texts and websites gradually made sense in the BI systems [8]. However, another important unstructured data, image, is still difficult to deal with to some extent. The image processing technology has experienced and is experiencing a tough time [9].

With the abundant and colorful information contained, unstructured image data has long been widely used for decision making [10]. Initially, the image was treated only as a whole data point for an overall descriptive statistical analysis. For instance, as the number and locations of images in an advertisement will affect consumers' decision-making psychology, BI system helps companies to optimize advertising design and improve marketing methods [11]. Also, in the meteorology area, taking pictures of the distribution of clouds over the course of a day is an effective way to predict weather conditions [12]. Moreover, the difference between various stations, fitting well with the passengers' flow data, supports the optimization of urban transportation network [13].

Similar to the texts, image certainly contains a great deal of effective information, which means it can be further analyzed by coding and feature extraction [14]. In general, the task of image processing is to extract effective and useful knowledge, which transforms the unstructured data into standard structural information [15]. Since 1960's, based on the pixel data of image, the digital image processing methods have been extensively applied as tools to support decision-making, such as contour detection, image denoising, image restoration and image classification [16]. These methods extract the decision-effective information from gray image. To deeply explore the knowledge concealed in the image, computer vision based on deep learning , has emerged and gradually developed into a powerful for a wide range of applications. It has achieved great success in image super-resolution, image registration, image identification, as well as image production, digging the image characteristics and support the decision-making [17]. Through the pixel calculation and feature extraction, we can obtain more precise and decision-effective knowledge to promote Business Intelligence. This is the basic principle of unstructured image data analysis using digital image processing and computer vision methods. In this paper, we will mainly review the image processing methods used in the construction of Business Intelligence, and summarize the specific algorithms, experiments, advantages and weaknesses by analyzing typical applications and articles [18].

This paper will be arranged as follows. Section 2 will summarize the digital image processing methods at the pixel-level used in the construction of business intelligence. Section 3 will review some typical BI systems using computer vision methods based on deep learning algorithms. Together, Section 4 mainly discusses the possible innovative progress of BI construction and its future trends.

2. Digital Image Processing Methods

In this section, we will review several representative digital image processing technologies used in the construction of business intelligence systems, such as image transformation, image enhancement and restoration, image segmentation and image classification. Some classical applications will be illustrated to explain the specific processing methods and its characteristics.

2.1. Image Transformation (IT)

Because the image array is very large, it expends enormous time complexity to process in spatial domain directly. Therefore, a variety of image transformation methods, such as Fourier transform (eq.1&2), are often used to transform spatial domain processing into frequency domain processing [19]. The newly studied wavelet transform has good localization characteristics in both spatial domain and frequency domain and shows promising effectiveness in image processing.

$$F(u) = \int_{-\infty}^{\infty} f(x) \cdot e^{-j2\pi ux} dx \tag{1}$$

$$f(x) = \int_{-\infty}^{\infty} F(u) \cdot e^{j2\pi ux} du \tag{2}$$

As for the applications, in the aerospace decision support system, the images collected in the air are first converted into digital signals stored on magnetic tape, then transmitted down at high speed, when the satellite passes over the ground station, and finally analyzed by the processing center. In the field of communication, the amount of image data is very large, considering the high rate of transmission of color TV signal, which is more than 100Mbps. To transmit such high volume of data in real time, coding techniques must be used to compress the bits of information [20].

2.2. Image Enhancement and Restoration (IER)

The purpose of image enhancement and restoration is to improve the quality of the image, such as removing noise and improving the clarity of the image [21]. By strengthening the high-frequency component of the image, it can make the outline of the object in the image clear and the details obvious. Enhancing the low frequency component can reduce the influence of noise in the image. Image restoration uses some filtering method to restore or reconstruct the original image. The steps of image restoration are shown as follows:

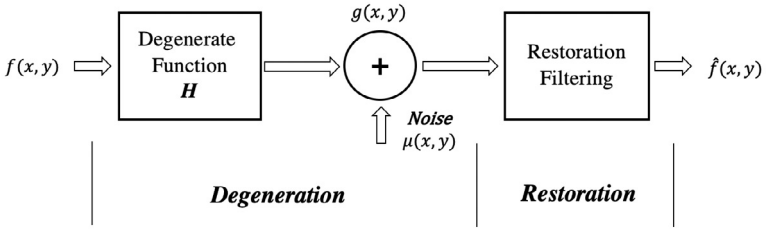

Fig. 1. The process of image restoration

And the mathematical expression is as follows:

$$g(x,y) = h(x,y) * f(x,y) + \eta(x,y) \tag{3}$$

Image restoration is widely used in biomedical engineering and is very effective. In addition to the well-known CT technology, there are other types of medical microscopic image processing and analysis, such as X-ray lung image enhancement and ultrasonic image processing. In the aspect of culture and art, image restoration also has important applications, such as the reproduction and restoration of photos of cultural relics [22].

2.3. Image Segmentation (IS)

Image segmentation is one of the key technologies in digital image processing. Image segmentation is to extract the meaningful feature parts of the image (eq.4) [23]. The meaningful features include edges and regions in the image. This is the basis for further image recognition, analysis and understanding.

$$g(i,j) = \begin{cases} 1, f(i,j) \geq T \\ 0, f(i,j) < T \end{cases} \tag{4}$$

Now many countries in the world are using the edge information of images to carry out resource investigation (such as forest investigation, water resources investigation, etc.), disaster detection (such as pest detection, environmental pollution detection, etc.), resource investigation (such as oil exploration, mining output detection, etc.), agricultural planning and urban planning, etc [24].

2.4. Image Classification or Recognition (IC/IR)

Image classification (recognition) belongs to pattern recognition. After some pre-processing, it segments the image and extract their features [25]. And the final judgement and classification are based on the comparison between images. Among them, the main method of comparison is to calculate the distance (eq.5) between two pictures.

$$d_1(I_1, I_2) = \sum_p \left| I_1^p - I_2^p \right| \tag{5}$$

The application area of image classification and recognition includes public security department realizes fingerprint recognition, face identification, incomplete picture recovery, traffic monitoring, accident analysis and so on through picture interpretation and analysis [26]. The automatic recognition of vehicle and license plate in the expressway automatic toll collection system, which has been put into operation at present, is one of the most successful application examples of image recognition technology.

3. Deep Learning based Computer Vision Method

Deep learning, which has been privileged for almost 15 years, now is fairly hot and mature in both academic research and practical applications [27]. The usage of deep learning methods in computer vision mainly can be divided into two categories. One is self-learning models, which needs no external information. The other is the generative models based on the additional knowledge.

3.1. Self-learning Methods

Object detection [28], image registration and super resolution algorithms are all implemented without additional information. Among them, object detection is to accurately find the location of the object in the given picture and mark the category of the object, as shown in Figure 2. Super-resolution aims to generate an image with higher resolution and clearer details corresponding to the original image. Super-resolution task has important applications in the restoration of important historical images [29]. Image registration [30] is the process of matching and superposition of two or more images obtained at different times, different sensors (imaging equipment) or under different conditions (weather, illumination, camera position and Angle, etc.). Image

registration, which can identify the risk of lesions and cell canceration, has been successfully applied in the intelligent decision support system of medical system.

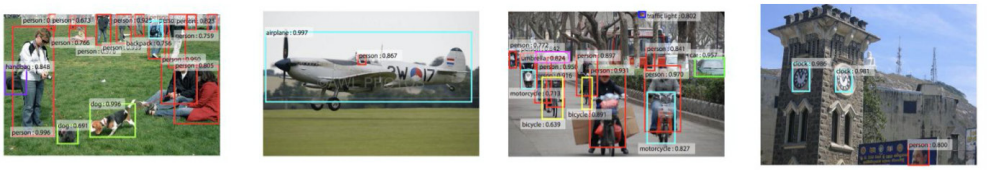

Fig. 2. Object Detection

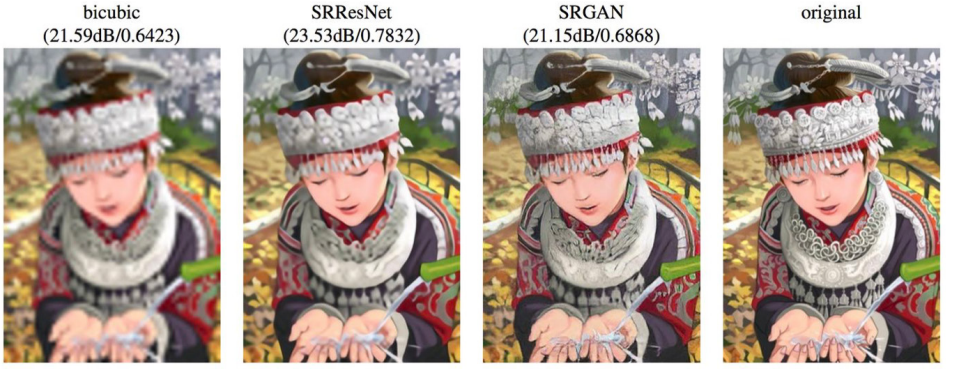

Fig. 3. Super-Resolution

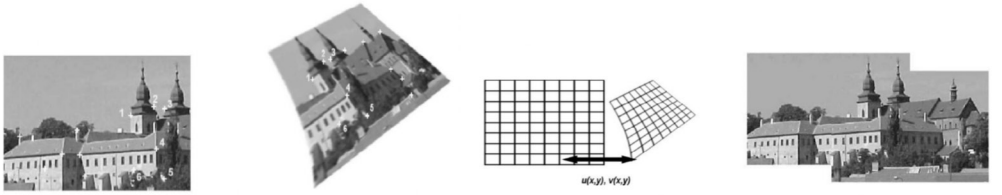

Fig. 4. Image Registration

3.2. Generative models

Generative adversarial networks (GANS) [31] are an interesting way of "teaching" computers to do human jobs. A GAN contains two competing neural network models. A network, called the generator, takes noise as input and generates a sample; The other network, called a discriminator, receives both generated data provided by generator and actual training data, and trains a classifier that correctly distinguishes between data types. The two networks compete against each other to improve accuracy respectively. The main applications are generating new images, style transfer, picture painting and so on. The most basic image generation model is a fantasy image generated from an ordinary photograph. Style transfer refers to the application of the style of one field or several pictures to other fields or pictures, such as the application of abstract style to realistic pictures [32]. Picture painting tries to convert a grayscale image into a colourful image.

Fig. 5. Image Painting

4. Discussion and Future Trend

After the above review and illustration, now we are trying to summarize the possible innovative progress and future trend in the construction of business intelligence systems using the unstructured image data.

One particular trend is the integration of big data, including both structural and unstructured data. Early business intelligence systems usually use only one type of data, which is tough to meet the current complex needs. Now the texts, images and also basic statements are all used to construct the BI systems. It generates a new concept called big-data fusion, which arises more knowledge for decision-making.

Another possible innovative trend is the increasingly important pre-processing stage, that is, taking the interpretability into consideration. With the explosive growth of data volume, more and more data are imported into BI systems while not all of them are effective. Some scholars [33] have noticed this computational crisis, and tried to find the balance between accuracy and complexity. However, it's hard to estimate the effectiveness of a piece of data before the data analysis is implemented. This means most of the data still cannot be analysed directly after pre-processing. Therefore, it's necessary to think about how to improve the effectiveness of the data imported into the BI systems.

Acknowledgements

This work was supported in part by National Natural Science Foundation of China (#71932008, #71801232). And we'd like to express our sincere gratitude to all the editors and reviewers.

References

[1] Chaudhuri S, Dayal U, Narasayya V. An overview of business intelligence technology[J]. Communications of the Acm, 2011, 54(8):88-98.

[2] Turban E, Aronson JE, Liang TP, et al. Decision support and business intelligence systems[M].

[3] Baars H, Kemper HG. Management Support with Structured and Unstructured Data - An Integrated Business Intelligence Framework[J]. Journal of Information Systems Management, 2008, 25(2):132-148.
[4] Rooke W A . Business intelligence system[J]. US, 2002.
[5] Park L . Intelligent profitable customers segmentation system based on business intelligence tools[J]. Expert Systems with Applications, 2005.
[6] Gilad B , Gilad T , Association A M . The business intelligence system : a new tool for competitive advantage[J]. 1988.
[7] Wu D , Zhang G , Lu J . A Fuzzy Preference Tree-Based Recommender System for Personalized Business-to-Business E-Services[J]. IEEE Transactions on Fuzzy Systems, 2015, 23(1):29-43.
[8] Yeoh W , A Popovič. Extending the understanding of critical success factors for implementing business intelligence systems[J]. Journal of the Association for Information Science & Technology, 2016, 67.
[9] Inmon W , Nesavich A . Tapping into Unstructured Data: Integrating Unstructured Data and Textual Analytics into Business Intelligence. Prentice Hall Press, 2008.
[10] Abdullah M F , Ahm Ad K . Business intelligence model for unstructured data management[C]// 2015 International Conference on Electrical Engineering and Informatics (ICEEI). IEEE, 2015.
[11] Hill J , Scott T . A consideration of the roles of business intelligence and e-business in management and marketing decision making in knowledge-based and high-tech start-ups[J]. Qualitative Market Research, 2004, 7(1):48-57.
[12] Raza M , Hussain O K , Chang H E . Maturity, distance and density (MD2) metrics for optimizing trust prediction for business intelligence[J]. Journal of Global Optimization, 2011.
[13] Sampaio P , Pereira G , Carvalho M S , et al. A business intelligence solution for public transportation sector. 2011.
[14] Gonzalez R C , Woods R E . Digital image processing[J]. Prentice Hall International, 2008, 28(4):484 - 486.
[15] Sonka M , V Hlavác, Boyle R . Image processing, analysis, and machine vision[J]. Journal of Electronic Imaging, 2014, xix.
[16] Bo Y Y , Elbuken C , Ren C L , et al. Image processing and classification algorithm for yeast cell morphology in a microfluidic chip[J]. Journal of Biomedical Optics, 2011, 16(6):066008.
[17] Schwartzman A , Kagan M , Mackey L , et al. Image Processing, Computer Vision, and Deep Learning: new approaches to the analysis and physics interpretation of LHC events[J]. Journal of Physics Conference, 2016, 762:012035.
[18] Sun Z , Bebis G , Miller. Monocular precrash vehicle detection: Features and classifiers. IEEE Transactions on Image Processing, 15(7), 2019-2034[J]. 2006.
[19] Tang Y Y , Suen C Y . Image Transformation Approach to Nonlinear Shape Restoration[J]. IEEE Transactions on Systems Man & Cybernetics, 1993, 23(1):155-172.
[20] Acharya T , Ray A K . Image Processing (Principles and Applications) || Image Transformation[J]. 2005.
[21] Aponso A C , Krishnarajah N . Review on state of art image enhancement and restoration methods for a vision based driver assistance system with De-weathering. Soft Computing & Pattern Recognition, 2011:135-140.
[22] Alruwaili M , Gupta L . A statistical adaptive algorithm for dust image enhancement and restoration[C]// IEEE International Conference on Electro/information Technology. IEEE, 2015:286-289.
[23] Kaut H , Singh R . A Review on Image Segmentation Techniques for Future Research Study[J]. International Journal of Engineering Trends and Technology, 2016, 35(11):504-505.
[24] P Arbeláez, Maire M , Fowlkes C , et al. Contour Detection and Hierarchical Image Segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2011, 33(5):898-916.
[25] Bosch A , Zisserman A , X Muñoz. Image Classification using Random Forests and Ferns[C]// 2007 IEEE 11th International Conference on Computer Vision. IEEE, 2007.
[26] Vailaya A , Jain A , Zhang H J . On Image Classification: City Images vs. Landscapes[J]. Pattern Recognition, 1998, 31(12):1921-1935.
[27] Zhang Q , Yang L T , Chen Z , et al. A survey on deep learning for big data[J]. Information Fusion, 2018, 42:146-157.
[28] Cheng G , Han J . A Survey on Object Detection in Optical Remote Sensing Images[J]. Isprs Journal of Photogrammetry & Remote Sensing, 2016, 117:11-28.
[29] Li B , Wang B , Liu J , et al. s-LWSR: Super Lightweight Super-Resolution Network[J]. IEEE Transactions on Image Processing, 2020, PP(99):1-1.
[30] B Zitová, Flusser J . Image Registration Methods: A Survey[J]. Image and Vision Computing, 2003, 21(11):977-1000.
[31] Goodfellow I J , Pouget-Abadie J , Mirza M , et al. Generative Adversarial Networks[J]. Advances in Neural Information Processing Systems, 2014, 3:2672-2680.
[32] Gao X , Tian Y , Qi Z . RPD-GAN: Learning to Draw Realistic Paintings with Generative Adversarial Network[J]. IEEE Transactions on Image Processing, 2020, PP(99):1-1.
[33] Yaroslavsky L P , Kober V . Redundancy of signals and transformations and computational complexity of signal and image processing[J]. IEEE, 1994.