CONVERGENCE OF MULTIMEDIA WITH WEB MINING

G. Ranadheer Reddy¹, V. Pranathi², P. Pramod Kumar³

¹,²Assistant Professor, Department of CSE, SRITW, India
³Senior Assistant Professor, Department of CSE, S R University, India

https://doi.org/10.26782/jmcms.2020.08.00039

Abstract

A ubiquitous process to evoke the most needed data information from huge amount of unprocessed data to analyze the patterns is called as data mining which is also named as data through knowledge discovery. It helps the enterprises to extract the data information to gain knowledge for better. [I] Data mining usually deals with text for mining. Since we are using internet for the accessibility of data. In other words, we are making use of web to extract the data, modify and process the text using the Web Pages. Evoking the information data which is present on internet is done using data mining is called as Web Mining. [II] It is an integral part of data mining for searching and analyzing the pattern. There are various data resources to obtain the data from web which is categorized into metadata, text documents, web links and web content. A web mining also consist of images, videos and audio information data which are considered as multimedia data. As, many users are more keen towards extracting information in form of images and videos from the web pages, so there’s a need of bringing out the required multimedia data information from unused scattered multimedia data present in the web. Here, we need to coalesce mining concepts through web into the multimedia stored data. Such concept is considered as Multimedia Web Mining. [V] It reaps the hidden information of a multimedia file as metadata, represents relationship between multimedia data files. For better and efficient working performance of mining techniques, multimedia mining also index and classify the various modes of multi data such as animation, moving, still, playback and video modes. Multimedia information is divided into two halves as organized and semi organized. Similarly web mining is categorized into utilization mining, organized mining and substance mining. In this paper, we explore the integration of multimedia with web mining for better enhancement in achieving the classification of data.

Keywords: Web mining, levels of data mining

I. Introduction

The primary definition of Mining is to wring some productive kind information from the bottom level of surface or substance. [II] Mostly we are dealing with data, to acquire the needed information from the large scale of stored data; we
need to perform certain mining techniques to extract a pattern which will be helpful. Such method for extracting the information is called as data mining. It is classified into structured and unstructured data. Since last decade, there is a presence of very huge amount of data which used for retrieval, search and modification but every piece of data is relevant information to the end users who is connected through web technologies to access the data saved at central storage. Hence, we require data mining at web level which is called as web mining. Web mining is broadly divided into data stream and retrieval stream. It ascertains the functionally required data and its patterns which are needed. [III] Data Mining of Multimedia is a process to find associations within the statistical data extracted from the huge amount of data which is represented in terms of audio, video, speech and even along with text data. Multimedia mining can also used to create and search the patterns. Multimedia content mining is considered as representation of heterogeneous data. The resultant product consists of data obtained by various preprocessing at each and every stage with different techniques of mining. In this paper, we elaborate the types of mining that exists in multimedia and what different types of tools we can use to extract a pattern from multimedia through web mining.

II. Levels Of Data Mining

Every data that’s stores in the central storage requires manipulation and modification for better result. To obtain better output result, certain basic concepts of mining process need to initiate irrespective of its presence either in the form of web or text or multimedia information data. The mining process is more reciprocal and sequential in nature. Generally mining process deals with text data in the early stages of technological world, it is divided into following levels:

1. Sphere knowledge:

In this stage, a very good knowledge of sphere in terms of data domain is required which will be helpful to identify the useful data. At this level, the data is not organized in proper format and there does not exist any particular mode of relational model for extraction of data that leads to more consumption of time for mining of concerned business processes. The basic convention of this level needs to bring all relevant in-hand data together for the future implementation of mining productive outputs. Identification of relevant data can be achieved through sphere knowledge. In the next level of data selection, the end client users need to choose sub fields of data records for mining process for the selected database. Specially, for multimedia mining, the relevant data in stored in form of relational type and there are no subfields to choose.

2. Data Pre-Processing:

The embodiment of information data from various remote locations and generating possibilities for depicting few of data fields which are considered as source of knowledge for the level of the pattern discovery. Since multimedia data
consist of semi organized and unorganized information data, this level of data preprocessing is very important for mining.

3. Detection of Pattern:

This is level where the inclination of conceal patterns for information data is revealed. There are ways to discover the patterns such as regression, analysis, classification, visualization, and association. There is a computing technique model for each and every way of discovering of the patterns, Hence data mining is considered as heap-sectored terrain.

4. Analysis:

In this level, assessment of revealed pattern which are hidden and its merit to find whether the previous level can be implemented again and again for successive pattern searching or not. To fine the exact merit and demerit, one has to hold a very good understanding and execution of the fist level that is sphere knowledge.

5. Brief:

After obtaining the needed information in the form of pattern through various levels of mining. It is implemented to generate actions for better output results.

III. Flow Sequence For Web Mining

1. Gather: The information is received and collected from various sources of web in form of resource discovery.

2. Pre–process: Segregate the needed information data for the web pages with the help of generalization.

3. Examine: the evaluation of information data for pattern discovery is filter based on the particular token through clustering and classification for a better output results.

4. View: Reports are generated converting the output results into a better beneficial information data.

Classification of Web Mining:

[IV]The data information available on the web represents unkempt information data in form of structures and un-structured which are shown in the format of text documents, HTML Web Pages. Blistering of data information is the main purpose of web mining, since a web content consist of all sorts of information such as text, audio, video and graphical representation, it is called a multimedia data mining with the web. Hence, Web mining is classified into three types namely:

1. Web content:

The main aim of web content mining is to exert the needed information from the contents within the Web Pages of a same website. The contents of webpage consist of
multiple objects varying from text to picture and animation with video data, which can produce better patterns for the usage of end users

2. **Web Usage:**

   It helps to collect data that represents the usage patterns based on log details, dynamic content and multimedia data. The pattern provided information of reference with regard to internet protocol, accession time

3. **Web Structure:**

   Web structure is used to classify and produce details dealing to association and resemblance between two different web pages on a same website or on two different websites

IV. **Multimedia Web Mining Architecture**

1. Insertion of data. The stage of entering the raw data into the web which consist of all type of information for identify and searching the pattern for the purpose of mining the raw data can be also stored in multimedia database.

2. Multimedia data information: It is the stage where the selection of data which is needed by the end users though database that’s holds subfields of data.

3. Segmentation: Object segmentation of data which is represented in animation or videos are spilt into images so that pattern finding can be easy

4. Quality descent: Incorporating the data base don various resources by means of pre processing and generating possibilities for the next stage of pattern finding by implicating code of few subfields of data.

5. Pattern similarity. It is central part of complete mining process, in this stage the patterns which are hidden patterns and conceptual store will be revealed in form of association, visualization

6. Assessment: to check the hidden pattern which is revealed can be used for better output results and can the earlier stages be used for reevaluation of other data elements.
V. Models of Multimedia Web Mining

1. Classification:
   Analysis of data through multimedia content from each and every property of a subfield data information. It is a method to build the data into groups of sub field of labels for a better output result. A model is generated for training a dataset with the remaining domain values of the field. It has a perceptive of no loss for end users who are working on multimedia data.

2. Association:
   It is a process to build a relationship among the content of multimedia data which is categorized into image and non-image data. It also creates various evaluation information on the same data set of multimedia content.

3. Clustering:
   It divides the data into groups named as clusters. These processes work on grid-based, density-based, and hierarchical-based. Various methods are hierarchical methods, density-based methods, grid-based methods, and they can be implied on the same type of images which are grouped by their attributes.

4. Statistical Modeling:
   The evaluation variants are regulated for statistical aspects of knowledge with correlation for finding the pattern. It establishes a link between the text and the image.
VI. Application Of Multimedia Mining

1. Traffic Management:
   The flow of traffic is maintained through various images and videos. We can detect the wrong doers by matching the pattern available with the data storage. We can identify the person who doesn’t follow the rules of traffic.

2. Clinic Management:
   Now a days, it is mostly in hospitals to detect the diseases by inserting the cam in to the body or taking images of human body and comparing it with the pattern that exist within the central data base.

VII. Conclusion and Challenges

These paper summaries concepts of data mining for multimedia contents through web technologies, we have provided the fundamental concepts that relate to multimedia data mining which can eb solved in long run. There’s a need of storing the data at client level and prompt the user with its predication and product implication.

References

I. http://airccse.org/journal/ijcga/papers/5115ijcga05.pdf
II. https://www.researchgate.net/publication/319404075_A_Survey_on_Web_Mining_Techniques_and_Applications
III. https://www.researchgate.net/publication/230639907_A_Survey_on_Multimedia_Data_Mining_and_Its_Relevance_Today
IV. http://www.ijcstjournal.org/volume-5/issue-3/IJCST-V5I3P21.pdf
V. http://www.academia.edu/Documents/in/Web_Data_Mining
VI. https://ieeexplore.ieee.org/abstract/document/5992597
VII. https://arxiv.org/pdf/1109.1145
VIII. Kiran Kumar S V N Madupu, "Opportunities and Challenges Towards Data Mining with Big Data", International Journal of Scientific Research in Science and Technology (IJSRST), Online ISSN : 2395-602X. Print ISSN : 2395-6011, Volume 1 Issue 3, pp. 207-214, July-August 2015. Available at doi : https://doi.org/10.32628/IJSRST207255
IX. Kiran Kumar S V N Madupu, "A Survey on Cloud Computing Service Models and Big Data Driven Networking", International Journal of Scientific Research in Science and Technology (IJSRST), Online ISSN : 2395-602X, Print ISSN : 2395-6011, Volume 4 Issue 10, pp. 451-458, September-October 2018. Available at doi : https://doi.org/10.32628/IJSRST207257

X. Kiran Kumar S V N Madupu, “Data Mining Model for Visualization as a Process of Knowledge Discovery”, International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering, ISSN: 2278 – 8875, Vol. 1, Issue 4, October 2012.

XI. Kiran Kumar S V N Madupu, “Advanced Database Systems and Technology Progress of Data Mining”, International Journal of Innovative Research in Science, Engineering and Technology, ISSN: 2319 – 8753, Vol. 2, Issue 3, March 2013

XII. Kiran Kumar S V N Madupu, “Functionalities, Applications, Issues and Types of Data Mining System”, International Journal of Innovative Research in Computer and Communication Engineering, Vol. 5, Issue 8, August 2017

XIII. M. M. Alabbadi, “Cloud Computing for Education and Learning: Education and Learning as a Service (ELaaS).” 2011 14th International Conference on Interactive Collaborative Learning (ICL), pp. 589 – 594, DOI=21-23 Sept.2011.

XIV. Naresh Kumar, S., Pramod Kumar, P., Sandeep, C.H. & Shwetha, S. 2018, "Opportunities for applying deep learning networks to tumour classification", Indian Journal of Public Health Research and Development, vol. 9, no. 11, pp. 742-747.

XV. Pramod Kumar, P., Sandeep, C.H. & Naresh Kumar, S. 2018, "An overview of the factors affecting handovers and effective highlights of handover techniques for next generation wireless networks", Indian Journal of Public Health Research and Development, vol. 9, no. 11, pp. 722-725.

XVI. Pramod Kumar P, Thirupathi V, Monica D, “Enhancements in Mobility Management for Future Wireless Networks”, International Journal of Advanced Research in Computer and Communication Engineering, Vol. 2, Issue 2, February 2013

XVII. Pramod Kumar P, CH Sandeep, Naresh Kumar S, “An Overview of the Factors Affecting Handovers and Effective Highlights of Handover Techniques for Next Generation Wireless Networks”, Indian Journal of Public Health Research & Development, November 2018, Vol.9, No. 11
XVIII. P. Pramod Kumar, K. Sagar, “Vertical Handover Decision Algorithm Based On Several Specifications in Heterogeneous Wireless Networks”, International Journal of Innovative Technology and Exploring Engineering (IJITEE), ISSN: 2278-3075, Volume-8 Issue-9, July 2019

XIX. P. Pramod Kumar ,Dr. K. Sagar, “A proficient and smart electricity billing management system ” ,International Conference on Emerging Trends in Engineering and published in Springer Nature as a Part of the Learning and Analytics in Intelligent Systems book series (LAIS, volume 3), July 2019.

XX. Pushpavathi Mannava, "An Overview of Cloud Computing and Deployment of Big Data Analytics in the Cloud", International Journal of Scientific Research in Science, Engineering and Technology (IJSRSET), Online ISSN: 2394-4099, Print ISSN: 2395-1990, Volume 1 Issue 1, pp. 209-215, 2014. Available at doi : https://doi.org/10.32628/IJSRSET207278

XXI. Pushpa Mannava, "Research Challenges and Technology Progress of Data Mining with Bigdata", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN: 2456-3307, Volume5 Issue 4, pp. 08-315, July-August 2019. Available at doi : https://doi.org/10.32628/CSEIT20627

XXII. Pushpa Mannava, "Role of Big Data Analytics in Cellular Network Design", International Journal of Scientific Research in Science and Technology (IJSRST), Online ISSN: 2395-602X, Print ISSN : 2395-6011, Volume 1 Issue 1, pp. 110-116, March-April 2015. Available at doi : https://doi.org/10.32628/IJSRST207254

XXIII. Pushpa Mannava, “A Study on the Challenges and Types of Big Data”, “International Journal of Innovative Research in Science, Engineering and Technology”, ISSN(Online) : 2319-8753, Vol. 2, Issue 8, August 2013

XXIV. Pushpa Mannava, “Data Mining Challenges with Bigdata for Global pulse development”, International Journal of Innovative Research in Computer and Communication Engineering, ISSN(Online): 2320-9801, vol 5, issue 6, june 2017

XXV. Pushpa Mannava, "Big Data Analytics in Intra-Data Center Networks and Components Of Data Mining", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN: 2456-3307, Volume 1 Issue 3, pp. 82-89, November-December 2016. Available at doi : https://doi.org/10.32628/CSEIT206272
XXVI. P. Kalagiakos “Cloud Computing Learning,” 2011 5th International Conference on Application of Information and Communication Technologies (AICT), Baku pp. 1 - 4, DOI=12-14 Oct.2011.

XXVII. Ramesh Gadde, Namavaram Vijay, “A SURVEY ON EVOLUTION OF BIG DATA WITH HADOOP” in “International Journal of Research In Science & Engineering”, Volume: 3 Issue: 6 Nov-Dec 2017.

XXVIII. Sandeep, C.H., Naresh Kumar, S. & Pramod Kumar, P. 2018, "Security challenges and issues of the IoT system", Indian Journal of Public Health Research and Development, vol. 9, no. 11, pp. 748-753.

XXIX. Seena Naik, K. & Sudarshan, E. 2019, "Smart healthcare monitoring system using raspberry Pi on IoT platform", ARPN Journal of Engineering and Applied Sciences, vol. 14, no. 4, pp. 872-876.

XXX. Sheshikala, M., Kothandaraman, D., Vijaya Prakash, R. & Roopa, G. 2019, "Natural language processing and machine learning classifier used for detecting the author of the sentence", International Journal of Recent Technology and Engineering, vol. 8, no. 3, pp. 936-939.

XXXI. Shailaja, P., Guru Rao, C.V. & Nagaraju, A. 2019, "A parametric oriented research on routing algorithms in mobile adhoc networks", International Journal of Innovative Technology and Exploring Engineering, vol. 9, no. 1, pp. 4116-4126.

XXXII. Sivakumar, M., Ramakrishna, M.S., Subrahmanyam, K.B.V. & Prabhandini, V. 2017, "Model Order Reduction of Higher Order Continuous Time Systems Using Intelligent Search Evolution Algorithm", Proceedings - 2017 International Conference on Recent Trends in Electrical, Electronics and Computing Technologies, ICRTEECT 2017, pp. 70.

XXXIII. Shailaja, G.K. & Rao, C.V.G, 2019, "Robust and lossless data privacy preservation: optimal key based data sanitization", Evolutionary Intelligence.

XXXIV. Siripuri Kiran, Shoban Babu Sriramouju, “A Study on the Applications of IOT”, Indian Journal of Public Health Research & Development, November 2018, Vol.9, No. 11, DOI Number: 10.5958/0976-5506.2018.01616.9

XXXV. Sriramouju Ajay Babu, Namavaram Vijay and Ramesh Gadde, “An Overview of Big Data Challenges, Tools and Techniques” in “International Journal of Research and Applications”, Oct - Dec, 2017 Transactions 4(16): 596-601
XXXVI. Srinivas, Chintakindi & Rao, Chakunta & Radhakrishna, Vangipuram. (2018). Feature Vector Based Component Clustering for Software Reuse. 1-6. 10.1145/3234698.3234737.

XXXVII. Subba Rao, A. & Ganguly, P. 2018, "Implementation of Efficient Cache Architecture for Performance Improvement in Communication based Systems", International Conference on Current Trends in Computer, Electrical, Electronics and Communication, CTCEEC 2017, pp. 1192.

XXXVIII. Venkatramulu, S. & Rao, Chakunta. (2018). CSES: Cuckoo Search Based Exploratory Scale to Defend Input-Type Validation Vulnerabilities of HTTP Requests. 10.1007/978-981-10-8228-3_23.Venkatramulu, S. & Guru Rao, C.V. 2017, "RPAD: Rule based pattern discovery for input type validation vulnerabilities detection & prevention of HTTP requests", International Journal of Applied Engineering Research, vol. 12, no. 24, pp. 14033-14039

XXXIX. W. Dawoud, I. Takouna, and C. Meinel, “Infrastructure as a Service Security: Challenges and Solutions,” 2010 7th International Conference on Informatics and System, pp. 1-8, March2010.

XL. W. Itani, A. Kayssi, and A. Chehab, “Privacy as a Service: Privacy-Aware Data Storage and Processing in Cloud Computing Architectures,” 2009 8th IEEE International Conference on Dependable, Autonomic and Secure Computing, 2009, pp.711-716.