Advances in Intelligent Systems and Computing 712

Vikrant Bhateja
João Manuel R. S. Tavares
B. Padmaja Rani · V. Kamakshi Prasad
K. Srujan Raju *Editors*

Proceedings of the Second International Conference on Computational Intelligence and Informatics

ICCII-2017

Springer

Advances in Intelligent Systems and Computing

Volume 712

Series editor

Janusz Kacprzyk, Polish Academy of Sciences, Warsaw, Poland
e-mail: kacprzyk@ibspan.waw.pl

The series "Advances in Intelligent Systems and Computing" contains publications on theory, applications, and design methods of Intelligent Systems and Intelligent Computing. Virtually all disciplines such as engineering, natural sciences, computer and information science, ICT, economics, business, e-commerce, environment, healthcare, life science are covered. The list of topics spans all the areas of modern intelligent systems and computing such as: computational intelligence, soft computing including neural networks, fuzzy systems, evolutionary computing and the fusion of these paradigms, social intelligence, ambient intelligence, computational neuroscience, artificial life, virtual worlds and society, cognitive science and systems, Perception and Vision, DNA and immune based systems, self-organizing and adaptive systems, e-Learning and teaching, human-centered and human-centric computing, recommender systems, intelligent control, robotics and mechatronics including human-machine teaming, knowledge-based paradigms, learning paradigms, machine ethics, intelligent data analysis, knowledge management, intelligent agents, intelligent decision making and support, intelligent network security, trust management, interactive entertainment, Web intelligence and multimedia.

The publications within "Advances in Intelligent Systems and Computing" are primarily proceedings of important conferences, symposia and congresses. They cover significant recent developments in the field, both of a foundational and applicable character. An important characteristic feature of the series is the short publication time and world-wide distribution. This permits a rapid and broad dissemination of research results.

Vikrant Bhateja · João Manuel R. S. Tavares
B. Padmaja Rani · V. Kamakshi Prasad
K. Srujan Raju
Editors

Proceedings of the Second International Conference on Computational Intelligence and Informatics

ICCII-2017

Springer

Editors
Vikrant Bhateja
Department of Electronics
and Communication Engineering
Shri Ramswaroop Memorial Group
of Professional Colleges
Lucknow, Uttar Pradesh
India

João Manuel R. S. Tavares
Departamento de Engenharia Mecânica
Universidade do Porto
Porto
Portugal

B. Padmaja Rani
Department of Computer Science
and Engineering
JNTUH College of Engineering Hyderabad
(Autonomous)
Hyderabad, Telangana
India

V. Kamakshi Prasad
Department of Computer Science
and Engineering
JNTUH College of Engineering Hyderabad
(Autonomous)
Hyderabad, Telangana
India

K. Srujan Raju
CMR Technical Campus
Hyderabad, Telangana
India

ISSN 2194-5357 ISSN 2194-5365 (electronic)
Advances in Intelligent Systems and Computing
ISBN 978-981-10-8227-6 ISBN 978-981-10-8228-3 (eBook)
https://doi.org/10.1007/978-981-10-8228-3

Library of Congress Control Number: 2018931430

Printed on acid-free paper

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd.
The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

Preface

The Second International Conference on Computational Intelligence and Informatics (ICCII-2017) was hosted by the Department of Computer Science and Engineering, JNTUHCEH, Hyderabad, in association with CSI during September 25–27, 2017. It provided a great platform for researchers from across the world to report, deliberate, and review the latest progress in the cutting-edge research pertaining to computational intelligence and its applications to various engineering fields.

The response to ICCII-2017 was overwhelming with a good number of submissions from different areas relating to computational intelligence and its applications in main tracks. After a rigorous peer review process with the help of program committee members and external reviewers, 66 papers were accepted for publication in this volume of AISC series of Springer.

ICCII-2017 was organized in the honor of Prof. K. Venkateswara Rao's retirement, commemorating his services rendered in the JNTUHCEH University in general and the Department of Computer Science and Engineering, JNTUHCEH, Hyderabad, in particular.

For the benefit of authors and delegates of the conference, a preconference workshop was held on emerging technologies in the field of computer science on Information Security, Machine Learning, and Internet of Things. Professor Suresh Chandra Satapathy, PVPSIT, Vijayawada, Prof. L. Pratap Reddy, JNTUHCEH, Mr. Kiran Chandra, FSMI, Prof. C. Raghavendra Rao, HCU, Prof. R. Padmavathi, NIT Warangal, have delivered lectures that were very informative.

Dr. Hock Ann Goh from MMU, Malaysia, and Prof. Ravi from IDRBT have delivered keynote speeches on September 26–27, 2017.

We take this opportunity to thank all the speakers and session chairs for their excellent support in making ICCII-2017 a grand success. The quality of a refereed volume depends mainly on the expertise and dedication of the reviewers. We are indebted to the program committee members and external reviewers who not only produced excellent reviews but also reviewed in the stipulated period of time without any delay. We would also like to thank CSI, Hyderabad, for coming forward to support us in organizing this mega-convention. Our thanks are due to

Dr. Ramakrishna Murthy, ANITS, Vizag, for his valuable support in reviews. Also, we thank special session chairs Dr. Tanupriya and Dr. Praveen of Amity University, Noida.

We express our heartfelt thanks to our Chief Patrons, Prof. A. Venugopal Reddy, Vice Chancellor, JNTUHCEH, Prof. N. V. Ramana Rao, Rector, JNTUHCEH, Prof. N. Yadaiah, Registrar, JNTUHCEH, Prof. A. Govardhan, Principal, JNTUHCEH, faculty and administrative staff of JNTUHCEH for their continuous support during the course of the convention.

We would also like to thank the authors and participants of this convention, who have considered the convention above all hardships. Finally, we would like to thank all the volunteers who spent tireless efforts in meeting the deadlines and arranging every detail to make sure that the convention ran smoothly. All the efforts are worth and would please us all, if the readers of this proceedings and participants of this convention found the papers and event inspiring and enjoyable. Our sincere thanks to the press, print, and electronic media for their excellent coverage of this convention.

Lucknow, India — Vikrant Bhateja
Porto, Portugal — João Manuel R. S. Tavares
Hyderabad, India — B. Padmaja Rani
Hyderabad, India — V. Kamakshi Prasad
Hyderabad, India — K. Srujan Raju

Organizing Committee

Conference Chair

Dr. K. Venkateswara Rao, JNTUHCEH, Telangana, India

Organizing Chairs

Prof. B. Padmaja Rani, JNTUHCEH, Telangana, India
Prof. V. Kamakshi Prasad, JNTUHCEH, Telangana, India

Technical Chairs

Prof. L. Pratap Reddy, JNTUHCEH, Telangana, India
Prof. G. Vijaya Kumari, JNTUHCEH, Telangana, India
Prof. O. B. V. Ramanaiah, JNTUHCEH, Telangana, India

Technical Co-Chairs

Prof. R. Sridevi, JNTUHCEH, Telangana, India
Prof. M. Chandra Mohan, JNTUHCEH, Telangana, India
Prof. D. Vasumathi, JNTUHCEH, Telangana, India
Prof. S. Viswanadha Raju, CSE, JNTUHCEJ, Telangana, India
Prof. B. Vishnu Vardhan, CSE, JNTUHCEJ, Telangana, India

International Advisory Committee

Dr. Kun-lin Hsieh, NTU, Taiwan
Dr. Ahamad J. Rusumdar, KIT, Germany
Dr. V. R. Chirumamilla, EUT, Netherland
Dr. Halis Altun, MU, Turkey
Dr. P. N. Suganthan, Singapore
Dr. Boka Kumsa, Ethiopia

Advisory Committee

Dr. B. L. Deekshatulu, IDRBT, Telangana, India
Prof. Y. N. Narasimhulu, VC, Rayalseema University, Andhra Pradesh, India
Prof. A. Damodaram, VC, SVU, Uttar Pradesh, India
Prof. L. S. S. Reddy, VC, KLU, Andhra Pradesh, India
Prof. Allam Appa Rao, CRR Institute of Technology, Delhi, India
Prof. Hema A. Murthy, IIT Madras, Chennai, Tamil Nadu, India
Dr. P. Sateesh Kumar, IIT Roorkee, Uttarakhand, India
Prof. N. Somayajulu, NIT Warangal, Telangana, India
Prof. Arun Agarwal, HCU, Telangana, India
Prof. Atul Negi, HCU, Telangana, India
Prof. C. Raghavendra Rao, HCU, Telangana, India
Prof. V. N. Sastry, IDRBT, Hyderabad, Telangana, India
Mr. Gautam Mahapatra, DRDO, New Delhi, India
Prof. A. Ananda Rao, JNTUA, Andhra Pradesh, India
Prof. S. Satyanarayana, JNTUK, Andhra Pradesh, India
Dr. A. Ansari, JMIU, Delhi, India
Prof. M. Srinivasa Rao, SIT, JNTUHCEH, Telangana, India
Prof. S. Durga Bhavani, SIT, JNTUHCEH, Telangana, India
Prof. S. V. L. Narasimham, SIT, JNTUHCEH, Telangana, India
Prof. M. Madhavi Latha, JNTUHCEH, Telangana, India
Prof. D. Srinivasa Rao, JNTUHCEH, Telangana, India
Prof. M. Asha Rani, JNTUHCEH, Telangana, India
Prof. B. N. Bhandari, JNTUHCEH, Telangana, India

Publication Chair

Dr. S. C. Satapathy, PVPSIT, Andhra Pradesh, India

Editorial Board

Prof. V. Kamakshi Prasad, JNTUHCEH, Telangana, India
Prof. B. Padmaja Rani, JNTUHCEH, Telangana, India
Dr. João Manuel R. S. Tavares, FEUP, Portugal
Prof. Vikrant Bhateja, SRMGPC, Lucknow, Uttar Pradesh, India
Prof. K. Srujan Raju, CMR Technical Campus, Telangana, India

Organizing Committee

Mr. B. Rama Mohan, JNTUHCEH, Telangana, India
Ms. K. Neeraja, JNTUHCEH, Telangana, India
Dr. M. Nagarathna, JNTUHCEH, Telangana, India
Ms. J. Ujwala Rekha, JNTUHCEH, Telangana, India
Dr. K. P. Supreethi, JNTUHCEH, Telangana, India
Ms. Kavitha, JNTUHCEH, Telangana, India
Ms. Hemalatha, JNTUHCEH, Telangana, India
Ms. I. Lakshmi Manikyamba, JNTUHCEH, Telangana, India

Technical Committee

Prof. P. V. S. Srinivas Rao, JBIET, Telangana, India
Prof. T. V. Rajinikanth, SNIST, Telangana, India
Prof. Khaleel Ur Rahman Khan, ACE, Telangana, India
Prof. C. R. K. Reddy, CBIT, Telangana, India
Prof. A. Nagesh, MGIT, Telangana, India
Prof. A. Jagan, BVRIT, Telangana, India
Dr. G. Narasimha, JNTUHCEJ, Telangana, India
Dr. P. Sammulal, JNTUHCEJ, Telangana, India
Dr. K. Shahu Chatrapathi, JNTUHCEM, Telangana, India
Dr. Kranthi Kiran, JNTUHCEJ, Telangana, India
Dr. T. Venu Gopal, JNTUHCES, Telangana, India
Dr. K. Rama Krishna, CMRCET, Telangana, India

Web and Publicity Committee

Ms. B. Kezia Rani, JNTUHCEH, Telangana, India
Ms. P. Subhashini, JNTUHCEH, Telangana, India
Mr. R. Sarath Babu, JNTUHCEH, Telangana, India

Contents

About the Editors

Prof. Vikrant Bhateja is Associate Professor, Department of Electronics and Communication Engineering, Shri Ramswaroop Memorial Group of Professional Colleges (SRMGPC), Lucknow, and also the Head (Academics and Quality Control) in the same college. His areas of research include digital image and video processing, computer vision, medical imaging, machine learning, pattern analysis and recognition, neural networks, soft computing, and bio-inspired computing techniques. He has more than 90 quality publications in various international journals and conference proceedings. He has been on TPC and chaired various sessions from the above domains in international conferences of IEEE and Springer. He has been the track chair and served in the core technical/editorial teams for international conferences: FICTA 2014, CSI 2014, and INDIA 2015 under Springer AISC series and INDIACom-2015, ICACCI-2015 under IEEE. He is Associate Editor in *International Journal of Convergence Computing (IJConvC)* and also serving in the editorial board of *International Journal of Image Mining (IJIM)* under Inderscience Publishers. At present, he is Guest Editor for two special issues floated in *International Journal of Rough Sets and Data Analysis (IJRSDA)* and *International Journal of System Dynamics Applications (IJSDA)* under IGI Global Publications.

João Manuel R. S. Tavares graduated in Mechanical Engineering from the University of Porto, Portugal (1992); M.Sc. in Electrical and Computer Engineering, in the field of Industrial Informatics, University of Porto (1995); Ph.D. in Electrical and Computer Engineering, University of Porto (2001). From 1995 to 2000, he was a researcher at the Institute of Biomedical Engineering (INEB). He is co-author of more than 350 scientific papers in national and international journals and conferences, co-editor of 18 international books and guest editor of several special issues of international journals. In addition, he is Editor-in-Chief of the *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization (CMBBE: Imaging & Visualization)*; Editor-in-Chief of the *International Journal of Biometrics and Bioinformatics (IJBB)*; Co-Editor-in-Chief of the *International Journal for Computational Vision and Biomechanics (IJCV & B)*;

Co-Editor of the Lecture Notes in *Computational Vision and Biomechanics (LNCV & B)*; Associate Editor of the EURASIP *Journal on Advances in Signal Processing (JASP)*, *Journal of Engineering*, *ISRN Machine Vision*, *Advances in Biomechanics & Applications*, and of the *Journal of Computer Science (INFOCOMP)*, and reviewer of several international scientific journals. Since 2001, he has been Supervisor and Co-Supervisor of several M.Sc. and Ph.D. theses and involved in several research projects, both as researcher and as scientific coordinator. Additionally, he is co-author of three international patents and two national patents. His main research areas include Computational Vision, Medical Imaging, Computational Mechanics, Scientific Visualization, Human-Computer Interaction and New Product Development.

Dr. B. Padmaja Rani is a Professor in Computer Science and Engineering Department at JNTUH College of Engineering, Hyderabad. Her interest area is information retrieval embedded systems. She has published more than 25 papers in reputed journals and conferences in the areas of agile modeling, Web services and mining, etc. She was the former Head of Department of CSE, JNTUHCEH. She is a professional member of CSI.

Dr. V. Kamakshi Prasad is a Professor of Computer Science and Engineering Department at JNTUH College of Engineering, Hyderabad. He completed his Ph.D. in speech recognition from IIT Madras, India. He did his M.Tech. from Andhra University and B.Tech. from K. L. College of Engineering. He has completed over 12 years in JNTU on various positions. He has 21 years of teaching and 11 years of research experience. He has been teaching subjects like speech processing, pattern recognition, computer networks, digital image processing, artificial neural, artificial intelligence and expert systems, computer graphics, object-oriented analysis and design through UML, and soft computing. He has supervised 12 Ph.D. and 2 MS students. His research areas are speech recognition and processing, image processing, neural networks, data mining, and ad hoc networks. He has authored two books published by Lambert Academic Publishing and over 50 papers in national- and international-level journals.

Dr. K. Srujan Raju is the Professor and Head, Department of CSE, CMR Technical Campus, Hyderabad, India. He earned his Ph.D. in the field of network security, and his current research includes computer networks, information security, data mining, image processing, intrusion detection, and cognitive radio networks. He has published several papers in refereed international conferences and peer-reviewed journals and also he was in the editorial board of CSI 2014 Springer AISC series; 337 and 338 volumes. In addition to this, he has served as reviewer for many indexed journals. He is also awarded with Significant Contributor, Active Member Awards by Computer Society of India (CSI).

Efficient Video Indexing and Retrieval Using Hierarchical Clustering Technique

D. Saravanan

"We take the full responsibility of the ethical issue of the work. We have taken permission to use the information in our work. In case of any issue we are only responsible."

Abstract Technology has brought the use of image-based information on the World Wide Web. Many professionals, academicians, researchers, and many other users use images for their work. An image retrieval system based on the image content brings the image based on the semantic similarity between the input and stored image. For this, it uses many image properties such as color, shape, and texture of the image. In such situations, user needs extra care to bring the preferred output in large and complex database. The process is based on ranking according to the similarity measure which is computed from the low-level image. The proposed technique is based on hierarchical clustering algorithm, and it consists of two steps. In the first step, images are trained and stored in the database. In step two, input query image properties are extracted using this property and preferred images are extracted based on image property matching technique. This method is less time-consuming and is more user-friendly than text-based technique.

Keywords Image retrieval · Hierarchical cluster · Content retrieval Data mining · Image comparison · Image feature

1 Introduction

Today, most of the users spent their time for browsing, searching, and retrieving information from the large database. Traditional technique is supported for key-based information retrieval but technology made this process simple. Extracting the preferred output is difficult for the user [1], as retrieving the required information is comparitively more tough than storing images or videos in the web. Retrieving the preferred output required by users is a challenge due to the increased collection of image database on web. For this, efficient methods are needed to

D. Saravanan (✉)
Faculty of Operations & IT, The ICFAI Foundation for Higher Education (Deemed-to-be-University), IBS Hyderabad, Hyderabad, Telangana, India
e-mail: sa_roin@yahoo.com

V. Bhateja et al. (eds.), *Proceedings of the Second International Conference on Computational Intelligence and Informatics*, Advances in Intelligent Systems and Computing 712, https://doi.org/10.1007/978-981-10-8228-3_1

retrieve the needed information from this collection. For searching the image, user needs to specify query in terms of text, key frame image, or feed; any image information for this system will return similar type of image information contents. This similarity is used for searching the image content using the image low-level feature such as color, image attributes, etc. [2]. Due to increasing demand, retrieve the preferred images that are hot spot for many researchers [3]. This content-based searching is used in many areas such as artificial intelligence, in the field of medicine, security, and education training. Most of this searching technique is based on text-based today; user enters the query by a set of keywords, and system finds the match based on keyword available in the query [4]. This method has some drawbacks such as keyword search; considering the huge collection of image, it is not reasonable to automatically interpret them and images are not defined by keyword; this drawback brings the new technology in the information retrieval.

1.1 Information Retrieval Based on the Content

Technology brings digital images based on the visual content [5]. This technique is based on low-level image features such as an image color, texture, shape, pixel value, motion, etc.; and these features help to find the similarity between two images. Input image is classified based on this property and helps the user for increasing image retrieval in the complex database that increases the efficiency when it performs the image retrieval operation. In the proposed technique, hierarchical clustering technique is used based on the hierarchical decomposition of image database. Most clustering technique used is distance measure for finding cluster similarity [6]. This technique is applied for various video data files and identify the best clustering technique.

2 Existing System

- Traditionally, image retrieval is based on text-based query and images are stored in unordered way; it brings the searching more difficult.
- No proper image indexing techniques with this collection of images increases the searching time.
- Retrieval of the needed image from this complex database increases user burden.
- Performance of the existing system is very slow and time-consuming.

3 Proposed System

The proposed system works as follows: the input video is converted into frames, using frame extract action process; the unwanted frames, i.e., duplicate frames, are removed. Using hierarchical clustering, technique frames are clustered. Finally, user retrieves relevant frames using image query input; the proposed model is shown in the figure.

3.1 Advantage of Proposed Technique

- Proposed technique is based on hierarchical clustering algorithms.
- Cluster formation is based on the input illustration of pixels value.
- It is based on clustering to identify a needed image from small relevant group that increases the operation efficiency.
- Information retrieval then can be stored for later user reference.
- Proposed technique can be easily automated.

3.2 Proposed Architecture

See Fig. 1.

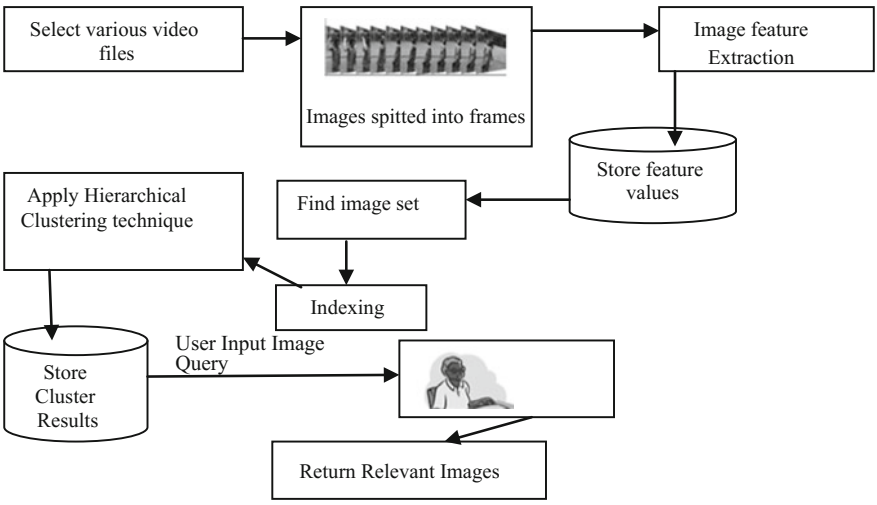

Fig. 1 Proposed architecture

3.3 Proposed Algorithm

Step 1 Get the various video files.
Step 2 Input the image and perform hierarchical clustering.
Step 3 Identify pixel as its own cluster.
Step 4 Find similar values and group of cluster.
Step 5 Find similarity between groups of cluster and merge most similar groups into one cluster.
Step 6 Process gets continued until all pixels are combined into one group.
Step 7 Apply proposed technique and find number of cluster k.
Step 8 Each cluster group identifies which point center it is closest to.
Step 9 Generate centric points with the help of Step 8.
Step 10 Repeat steps 7–10 up to process get terminated

4 Experimental Setup

4.1 Image Elements Extraction

Compared to the traditional text-based technique, the proposed technique works differently. Proposed technique works based on stored images from a huge complex image collection by comparing image features [7]. The common features used here are color and pixel value. In the proposed technique, find the histogram value of the given input image. These values are used in future for image processing. Image extraction process finds the equivalent values which are stored in the database; corresponding images are returned as output of the input.

4.1.1 Pseudocode for Color Histogram Calculation

```
double A = Image.getHeight();
 double B = Iimage.getWidth();
ColorHistogram      =      ColorStuctureExtraction((int)A,
(int)B1);
        ColorHistogram = reQuantization(ColorHistogram);
if(Image descriptor of Colorformationexecution)
        {                Colorformationexecution        =
(Colorformationexecution) image descriptor;
            if(Colorformationexecution.quantizationLevels
== quantizationLevels)
            {f   =   0.0F;   for(int   i   =   0;   i   <
ColorHistogram.length;   i++)f+=Math.abs(ColorHistogram[i]
- Colorformationexecution.ColorHistogram[i]);
```

4.2 Indexing of Images

Video is an effective tool to exchange the information, instead of typing big text-based information, and it brings easy communication between the user groups. For almost all digital image processing, the RGB color space is utilized in normal way for color monitors. Image indexing reduces the user searching time and also improves the performance of image retrieval system. Getting a better result avoids numerous measurement of images. Proposed technique uses image histogram value. Experimental output verifies that proposed techniques work well for different videos types.

4.3 Image Retrieval Using Clustering Technique

Clustering algorithms can follow a hierarchical or a partitioning model. Both use some specific points within the cluster to represent it in order to find out the similarities [8]. Clustering not only faces the time complexity and quality but also frequency estimation, order statistics, and compute data stream [9].

4.3.1 Psecudocode for Cluster Calculation

```
int l = 0;      double A = Cluster(closest, lk, f, r);
          while(d3 > EPSGLOB)
          { Centroids(closest, lk, f, c);
              B = Cluster(closest, lk, i, r);
              if(A > 0.0D) C = (A - B) / A; else    C =
0.0D;
              A = B;     if(l == 0 || C < pixelspliet&&
m_CurrSize < m_MaxSize)
              { Split(closest, af, f, c);
                  B = Cluster(closest, lk, f, r);
                  C= 1.0D; }for(j = 0; j < m_CurrSize; j++)
{                  for(int k2 = 0; k2 < j; k2++)}}
```

5 Experimental Outcomes

See Figs. 2, 3, 4, 5, and 6 and Table 1.

Fig. 2 Based on user image input query 1 output

Fig. 3 Based on user image input query 3 output

Fig. 4 Based on user image input query 4 output

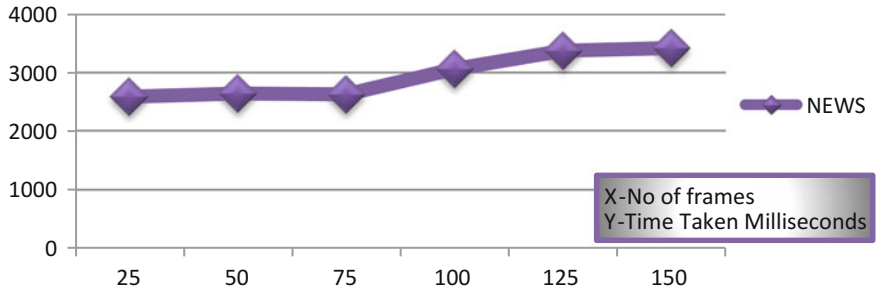

Fig. 5 Performance graph of news video file

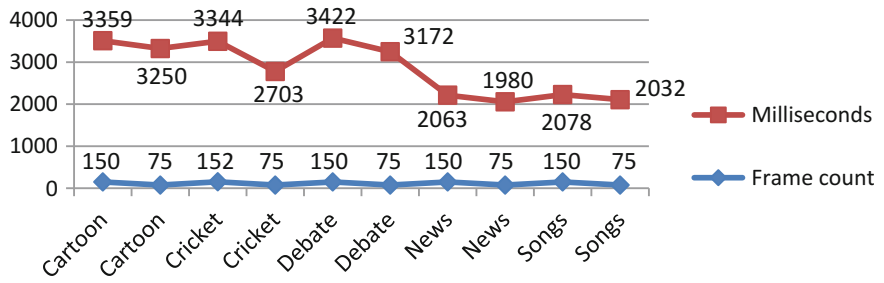

Fig. 6 Comparison performance graph of various video files

Table 1 Frame count versus time for different video files

Frame count	Milliseconds	Category of video file
150	3359	Cartoon
75	3250	Cartoon
152	3344	Cricket
75	2703	Cricket
150	3422	Debate
75	3172	Debate

6 Conclusion

In this work, an efficient framework for image retrieval from the given video is proposed. Hierarchical clustering technique for image retrieval is presented; images are initially clustered into groups based on image color; experimental output verified the proposed technique which brings more image outputs for given input image query, and it also works well in all types of videos. From the observations made, it is found that the existing techniques can be used only for limited sets of videos. Proposed technique fetches accurate image results for all types of video files and the outcomes have verified this. The technique may be further combined with other clustering algorithms to design other cross-algorithms.

References

1. Saravanan. D, Segment Based Indexing Technique for Data file. Procedia of computer Science, 87(2016), 12–17. (2016).
2. Saravanan. D, Video Substance Extraction Using image Future Population based Techniques. ARPN Journal of Engg., and applied science, Vol. 11, No(11), 7041–7045, (2016).
3. D. Saravanan and S. Srinivasan, "Data mining framework for video data," Recent Advances in Space Technology Services and Climate Change (RSTSCC), pp. 167–170, (2010).
4. Y. Yang, F. Nie, D. Xu, J. Luo, Y. Zhuang, Y. Pan. A multimedia retrieval framework based on semisupervised ranking and relevance feedback, IEEE Trans. Pattern Anal. Mach, 2012, Intell. 34: 723–742, (2012).
5. Martin Ester, Hans-Peter Kriegel, and Xiaowei Xu, A database interface for clustering in large spatial databases In Int'l Conference on Knowledge Discovery in Databases and Data Mining (KDD-95), Montreal, Canada, (1995).
6. D. Saravanan, "Various Assorted Cluster Performance Examination using Vide Data Mining Technique", Global journal of pure and applied Mathematics, Volume 11, No. 6 (2015), ISSN 0973-1768, Pages 4457–4467, (2015).
7. D. Saravanan, "Video Substance Extraction Using image Future Population based Techniques", ARPN Journal of Engg and applied science, Vol. 11, No(11), Pages 7041–7045, (2016).
8. D. Saravanan, "Design and implementation of feature matching procedure for video frame retrieval", International journal of control theory and applications, 9(7), Pages 3283–3293, (2016).
9. Tian Zhang, Raghu Ramakrishnan, and Miron Livny, (1996) Birch: An efficient data clustering method for very large databases In Proceedings of the ACM SIGMOD Conference on Management of Data, pages 103–114, Montreal, Canada, (1996).

A Data Perturbation Method to Preserve Privacy Using Fuzzy Rules

Thanveer Jahan, K. Pavani, G. Narsimha and C. V. Guru Rao

Abstract Data mining methods analyze the patterns found in data, irrespective of the confidential information of an individual. It has led to raise privacy concerns about confidential data. Different methods are inhibited in data mining to protect these data. Privacy preserving data mining plays a major role in protecting confidential data. The paper focuses on data perturbation method to preserve confidential data present in the real-world datasets. These identified confidential data are perturbed using fuzzy membership function (FMF) and obtains fuzzy data. The mining utility such as classification and clustering methods are used. The accuracy is determined and compared between an original data and fuzzy data. The results shown in the paper proves the proposed method is efficient in preserving confidential data.

Keywords Privacy · Fuzzy logic · Fuzzy membership function
Data mining

1 Introduction

Privacy has become a major concern in data mining applications, cloud computing, and service computing. Due to increase in era of internet, data is shared and is available publicly. The valuable information is extracted from large amounts of

T. Jahan (✉) · K. Pavani
Vaagdevi College of Engineering, Warangal, India
e-mail: thanvijahan@gmail.com

K. Pavani
e-mail: bandaripavani@gmail.com

G. Narsimha
JNTU Jagityal, Karimnagar, Telangana, India
e-mail: narsimha06@gmail.com

C. V. Guru Rao
Department of CSE, S R Engineering College, Warangal, Telangana, India
e-mail: guru_cv_rao@hotmail.com

V. Bhateja et al. (eds.), *Proceedings of the Second International Conference on Computational Intelligence and Informatics*, Advances in Intelligent Systems and Computing 712, https://doi.org/10.1007/978-981-10-8228-3_2

data, for business analysis and scientific computing using data mining methods. A major problem occurred while unauthorized users have access on private data. The private data is well preserved along with accuracy in data mining through privacy preserving data mining. PPDM protects sensitive information present in individuals before releasing. In the field of research, privacy preserving data mining has become novel [1]. The problem is challenging, primarily focusing to protect crucial data and secondary on data usage, to meet specific requirements of an organization. Violation occurred when confidential data is known from the patterns extracted using data mining. Thus, developing efficient data mining methods is needed to preserve privacy, in order to making data available publicly. Only by ensuring sensitive private data has been protected. Data perturbation is well known to preserve privacy in data mining techniques. Sensitive information of an individual is not efficiently protected by data mining, this can lead violation. Data mining extracts information hidden without disclosure of private data of individuals. The major challenge of privacy preserving is to concentrate on factors to achieve privacy guarantee and data utility.

1.1 Privacy Preserving Data Mining

Privacy is extensively used in two areas, they are centralized and distributed environments. In centralized environment, the data is located at a single place. Privacy preserving data mining algorithms are used to protect the sensitive information in data. In distributed database environment, data is distributed on various different sites either by horizontal or vertical partition. In this environment, the privacy preserving techniques are applied to protect individual information by integrated data from multiple sites. Privacy preserving is a technique to study mining patterns which mask the private information and preserves data. Privacy preserving algorithms are used to protect the sensitive information in data. It is classified into reconstruction methods, heuristics methods, and cryptographic methods (secure multiparty computations). Reconstruction is classified into data perturbation, data swapping and data swapping and randomization. The main research aspects of PPDM are applicability, privacy metrics, mining accuracy, and computation [2]. The disadvantages of the existing methods in PPDM methods are mostly concerned about protection metric and accuracy of mining. Data perturbation is the popular model for privacy preserving data mining.

1.2 Data Perturbation in PPDM

Data perturbation methods are used to modify data or add noise to data, data mining techniques have proved that original and perturbed data are relatively same and accuracy is measured by different classifiers. Data perturbation is the popular model

for privacy preserving data mining. Data perturbation is mainly categorized as probability distribution and value distortion approaches. The value distortion approaches perturb confidential data using noise. Recent methods on perturbation focus on random noise applied to the datasets, but not considering various privacy requirements of the different users. Mainly, there are two types of data perturbation methods used on continuous data: additive and matrix multiplicative methods such as, SVD, SSVD [3], fuzzy logic, and other detailed multiplicative data perturbation techniques. Data perturbation methods are used by the data owners before data is outsourced. Perturbation means changing of an attribute value by a new value. The perturbed data meet the two conditions: first an attacker cannot discover the original data from the perturbed or distorted data and second, the distorted data maintains the statistical properties of original data and can be derived. Data perturbation is extensively used in two areas, they are centralized and distributed environments. In centralized environment, the data is located at a single place. Distributed database environment relies on the data that is located/distributed on different sites or places. In this environment, the privacy preserving techniques are applied to protect individual information by integrated data from multiple sites.

The paper is mainly focused on the centralized environment, where data is placed under an authorized user (Data owner). Data owner needs to respect these privacy concerns when sharing, publishing, or otherwise releasing the data. The data owner mainly primarily aims to protect data before publishing it to outsource. Data perturbation methods are applied on the original set of data. These perturbed data are then analyzed to check the accuracy along with the original data. Since, there is no trade-off difference between perturbed and original data, the perturbed data can therefore be released. Second, data owner should also assume the role of attackers and develop techniques for breaching by estimating the original data from the perturbed data and any available additional prior knowledge. The attackers work can offers into vulnerabilities on different type's data perturbation methods.

2 Background and Related Work

Huge volumes of data are collected during data publishing as there is rapid growth of internet technology [4]. Many applications such as defense, medical care, transactions based on finance rely on issues of data privacy. Data is collected and gathered during the process of data mining. There are many data mining techniques to analyze raw data and is prone to threat for individuals privacy [5]. Many PPDM techniques are seen in literature such as data swapping, aggregation, Fourier and signal transformation, data anonymization including generalization and suppression. Among all these methods, data perturbation plays a major role in PPDM. In data perturbation, original dimension of data matrix was reduced by transforming using feature selection and SSVD for analysis purpose. Small distorted values are discarded as features for classification purpose [6, 7]. A fuzzy membership function (FMF) was used on original data to provide increased data privacy and decreased

number of passes to perform clustering. The advantages of fuzzy sets are imprecision and uncertainty that rely on real-world knowledge [8, 9]. Different fuzzy membership functions are used to modify or distort data having confidential attributes [10]. These confidential attributes are distorted and analyzed using data mining techniques such as classification that has decreased complexity and processing time [11]. In privacy preserving data mining, different methods are used in the recent years, based on noise addition methods [12]. Rotation perturbation is used as well as condensation-based perturbation [13]. Projection-based perturbation and geometric based perturbation as a perturbation and large body of literatures on k-anonymity model [9]. We mainly focus on perturbation methods.

3 Data Perturbation Using Fuzzy Logic

Fuzzy logic is used to solve imprecise problem with better solutions. To decrease complexity of a system, we need to understand a system well. Fuzzy sets are the extensions of generic set theory. The fuzzy sets are crisp set and progressive transition. It deals with a fuzzy set of a pair (D, μ_d) where D is a set and μ_d: D $\rightarrow$ [0, 1]. For all x ϵ D, μ_d(x) is named as the grade of membership of x. Different linguistic variables are defined by shape such as triangular, S-shaped, Z-shaped fuzzy membership functions [14, 15]. Different FMF are defined such as S-shaped membership function is represented as

$$f(x; a, b) = \left\{ \begin{array}{cc} 0, & x \le a \\ 2\left(\frac{x-a}{b-a}\right)^2, & a \le x \le \frac{a+b}{2} \\ 1-2\left(\frac{x-b}{b-a}\right)^2, & \frac{a+b}{2} \le x \le b \\ 1, & x \ge b \end{array} \right\}, \tag{1}$$

where x is the confidential attribute or column of the dataset, a is calculated as maximum value of the attribute, and b is calculated as the minimum value of the attribute for a dataset. The FMF such as Z-shaped membership function is represented as:

$$f(x; a, b) = \left\{ \begin{array}{cc} 1, & x \le a \\ 1-2\left(\frac{x-a}{b-a}\right)^2, & a \le x \le \frac{a+b}{2} \\ 2\left(\frac{x-b}{b-a}\right)^2, & \frac{a+b}{2} \le x \le b \\ 0, & x \ge b \end{array} \right\} \tag{2}$$

3.1 Fuzzy Rule Based Systems

Fuzzy rules are based on linguistic that constructs If-Then in the form of "IF X THEN Y" where X and Y are the linguistic variables. X is named as premise whereas Y is named as consequence of the rule. Fuzzy classification is developed