Image and Vision Computing 67 (2017) 29–42

Contents lists available at ScienceDirect

Image and Vision Computing

journal homepage: www.elsevier.com/locate/imavis

Review article

Feature selection in multimedia: The state-of-the-art review☆

Pui Yi Lee, Wei Ping Loh *, Jeng Feng Chin

School of Mechanical Engineering, Engineering Campus, Universiti Sains Malaysia, 14300 Nibong Tebal, Penang, Malaysia

ARTICLE INFO

Article history:
Received 22 April 2016
Received in revised form 12 July 2017
Accepted 9 September 2017
Available online 20 September 2017

Keywords:
Feature selection
Multimedia
Data mining
Search strategies

ABSTRACT

Multimedia data mining, particularly feature selection (FS), has been successfully applied in recent classification and recognition works. However, only a few studies in the contemporary literature have reviewed FS (e.g., analyses of data pre-processing prior to classification and clustering). This study aimed to fill this research gap by presenting an extensive survey on the current development of FS in multimedia. A total of 70 related papers published from 2001 to 2017 were collected from multiple databases. Breakdowns and analyses were performed on data types, methods, search strategies, performance measures, and challenges. The development trend of FS presages the increased prominence of heuristic search strategies and hybrid FS in the latest multimedia data mining.

Contents

1. Introduction

Multimedia data mining detects, classifies, and retrieves relevant features from a collection of multimedia to identify informative patterns and relationships for knowledge acquisition. The performance of multimedia data mining is directly influenced by the level of information in data features that can be recognized during data pre-processing [1].

Data pre-processing aims to reduce data dimensionality [2,3] by preserving useful features, which are variables (attributes) used as input variables for selection algorithms [4,5]. This improves the runtime (processing time), predictive accuracy, and results comprehensibility for facilitating human understanding and insight [1].

Two key steps in data pre-processing are feature extraction (FE) and feature selection (FS). FS techniques are a subset of the general field of FE [6]. FE extracts multiple features from the original data to generate a dataset [6]. Contrarily, FS selects a subset of original features to construct model [7]. According to Cantú-Paz et al. [8], identifying related features provides insights into underlying phenomenon. On the other hand, discarding irrelevant features improves the accuracy of data

☆ This paper has been recommended for acceptance by Yi Yang.

* Corresponding author.
E-mail addresses: melissa_lee_88@hotmail.com (P.Y. Lee), meloh@usm.my (W.P. Loh), chinjengfeng@usm.my (J.F. Chin).

https://doi.org/10.1016/j.imavis.2017.09.004
0262-8856/© 2017 Published by Elsevier B.V.

classifications. FE and FS are integral and mutually influential. In the work of Mohammadi and Abadeh [6], FE extracts images with hidden message through image steganalysis, and FS was adopted to determine appropriate features from the extracted images.

Numerous FS studies have underlined the challenges in selecting the best data features to obtain high classification accuracies [9–19]. Most scientific data consists of low-level data such as images or mesh data from computer simulations that could be noisy and massively large. For example, in Liu et al. [9], sufficiently discriminatory key frames has to be selected since not all frames in a sequence are relevant to the corresponding action. Besides, assigning labels to identify relevant scientific data requires subjective processes by domain knowledgeable expertise.

Since 2001, approximately 553,388 articles were obtained from the ScienceDirect database when “feature selection” was searched on July 2017. Of this number, a maximum of 70.25% of the articles were focused on multimedia. The studies applied FS on multimedia data include face recognition [20–22], pattern recognition [23], text detection and recognition in images and video frames [24,25], biomedicine [26,27], speech emotion recognition [28], image retrieval and classification [13], and image annotation [16,29].

Considerable research has extensively analysed FS, but the approaches related to multimedia data mining are yet to be widely looked into. Manjunath et al. [30] discussed the current state of multimedia data mining and knowledge discovery, including the current approaches and techniques used to mine multimedia. Pudło and Ząbkowski [31] presented an article on multimedia data mining techniques with applications highlighting information retrieval and future directions. However, both these works insufficiently discussed data pre-processing. Image classification, face recognition, and text classification have recently been outlined as popular contributions of FS applications in the context of big data [32]. Nevertheless, the issues identified in each application of Bolón-Canedo et al. [32] were not discussed in depth. Liu and Kender [33], Lin et al. [13], Tolias et al. [34], and Han et al. [35] correspondingly expressed their concerns over the inefficiencies of existing FS methods in catering to the continuous database growth over time.

Realizing these deficiencies in the literature, our article aims to deliver a state-of-the-art review to (i) present an extensive survey on the current development of multimedia FS, including the methods, common issues, and challenges regarding data type, applications, FS role, and performance measures and (ii) provide relevant insights into the issues encountered in multimedia FS techniques.

The rest of the paper is organized into seven sections. Following the Introduction section, Section 2 discusses the research methodology. Section 3 provides information on multimedia. Section 4 discusses the data sources for multimedia data mining studies. Section 5 reviews the FS methods and search strategies used in state-of-the-art papers. Section 6 cites the challenges in multimedia data mining faced by previous research. Finally, Section 7 concludes the paper with some discussions on the future of FS in multimedia data studies.

2. Research methodology

A total of 70 studies published beginning from 2001 to the second quarter of 2017 were retrieved from multiple databases. Since year 2001 significant development of the computer industry with enhanced graphical user interface allows an advanced platform to investigate multimedia data processing and data mining was observed [36–40].

The selected papers were classified according to multimedia data category and data pre-processing methods. Multimedia data category includes video, image, audio, image-video, images extracted from video, and text recognition from video. Image-video includes past research works that investigated both image and video. The images extracted from video only include studies on images taken from video sequences. Data pre-processing is divided into i) FS and ii) both FS and FE. These papers were further analysed in terms of FS methods and FS search strategies used. FS methods are categorized as hybrid and non-hybrid (filter, wrapper, and embedded), whereas FS search strategies are classified as optimal, heuristic, and randomized. An interval graph represented yearly was formed based on the FS search strategies to illustrate the prospective trends and the preference of forecast researchers of the near future. The issues encountered, intended applications, database used, and performance measures were also analysed. The collected papers included studies that applied both novel and existing FS methods on multimedia data.

3. Multimedia

Multimedia data are semi-structured or unstructured data entities whose primary content is individually or collectively used for communication [1,41]. The term “multimedia” covers more than one media object such as a combination of text, image, video, audio, numeric, sound, animation, graphical, and categorical data on a computer display terminal [42,43]. Researchers commonly use the term “multimedia” interchangeably with media objects [44]. In the advent of software technology, multimedia is defined as “a computer program consisted of texts, graphic, sound and images and animations” [45].

Text is composed of a large number of unstructured and amorphous characters of natural language text [46,47]. Texts can be embedded in videos in two forms, namely, caption and scene texts [48]. Caption text refers to the texts superimposed during video editing, whereas scene text is the texts that naturally exist during video captures. To deal with the embedded text for a formal exchange of information (e.g., title, date of an event, name of the speakers), caption text can be extracted directly from videos and images because it is simply placed over these visual representations [49]. However, scene text in video is usually extracted in an image (frame-by-frame) basis (e.g., Ye et al. [25] and Lee and Kim [24]). Hence, a video should first be converted into an image before the scene text can be extracted given that it cannot be laid over videos or images.

Audio is a signal resulting from oscillations or pressure variations generated by a vibrating body or a turbulent fluid flow in an elastic medium such as air, water, and solids [50]. Uzun and Sencar [51] classified audio into speech, music, non-speech and non-music signals, and complex sound mixtures that arose from several distinct acoustic sources. Audio content can be automatically analysed and searched within itself (e.g., pathological detection reported by Muhammad and Melhem [26]). For example, speech within audio detection is commonly used for identification or recognition such as aiding evidence collection [51] and diagnosing Alzheimer's disease [52]. Huang et al. [53] dealt with music retrieval with melody matching a hummed query. Query refers to a request for information from a database. Research has been carried out by Liu et al. [54] to complement the need to analyse audio in wearable devices such as smartwatches, smart clothing, and virtual reality headsets due to limited graphical user interface (GUI) or no GUI as well as limited energy and computational capacity available on the wearable devices.

Image is a form of static pictorial data that conveys implicit or explicit information, and static images are made up of bitmaps (grids of pixels) that can be gradually transformed into a digital format for editing, storage, and sharing [55]. Large amount of the current multimedia data is images for its valuable representation and can be processed into meaningful information for knowledge acquisition [56]. Yu [12] applied cartoon images, whereas Elguebaly and Bouguila [14] used painting and photo captured images as representations of computer-generated images. An image can be analysed within itself such as in image steganalysis [6], which screens hidden messages from multimedia data, and image browsing [11]. Content-based image retrieval (CBIR) is a popular process of searching for similar images based on their visual contents [56–58].