

A Multiple-Instance Learning Based Approach to Multimodal Data Mining

Zhongfei (Mark) Zhang, SUNY Binghamton, USA

Zhen Guo, SUNY Binghamton, USA

Jia-Yu (Tim) Pan, Google Inc., USA

ABSTRACT

This paper presents multiple-instance learning based approach to multimodal data mining in a multimedia database. This approach is a highly scalable and adaptable framework that the authors call co-learning. Theoretic analysis and empirical evaluations demonstrate the advantage of the strong scalability and adaptability. Although this framework is general for multimodal data mining in any specific domain, to evaluate this framework, the authors apply it to the Berkeley Drosophila ISH embryo image database for the evaluations of the mining performance in comparison with a state-of-the-art multimodal data mining method to showcase the promise of the co-learning framework.

Keywords: Adaptability, Co-Learning, Multimodal Data Mining, Multiple Instance Learning, Scalability

INTRODUCTION

Multimodal data mining in a multimedia database is a challenging topic in data mining research (Zhang et al., 2006). In this context, a multimedia database refers to a data collection in which there are multiple modalities of unstructured data such as text and imagery. By multimodal data mining in a multimedia database it is meant that the knowledge discovery to the multimedia database is initiated by a query that may consist of multiple modalities of unstructured data such as text and imagery. In

this paper, we focus on a multimedia database as an image database in which each image has a few textual words given as annotation. We then address the problem of multimodal data mining in such an image database as the problem of retrieving similar data and/or inferencing new patterns to a multimodal query from the database.

Specifically, in the context of this paper, multimodal data mining refers to two aspects of activities. The first is the multimodal retrieval. This is the scenario where a multimodal query consisting of either textual words alone, or imagery alone, or in any combination is entered and an expected retrieval data modality is specified

DOI: 10.4018/jdls.2010040102

that can also be text alone, or imagery alone, or in any combination; the retrieved data based on a pre-defined similarity criterion are returned back to the user. The second is the multimodal inferencing. While the retrieval based multimodal data mining has its standard definition in terms of the semantic similarity between the query and the retrieved data from the database, the inferencing based mining depends on the specific applications. In this paper, we focus on the application of the fruit-fly image database mining. Consequently, the inferencing based multimodal data mining may include many different scenarios. A typical scenario is the across-stage multimodal inferencing. There are many interesting questions a biologist may want to ask in the fruit fly research given such a multimodal mining capability. For example, given an embryo image in stage 5, what is the corresponding image in stage 7 for an image-to-image three-stage inferencing? What is the corresponding annotation for this image in stage 7 for an image-to-word three-stage inferencing? The multimodal mining technique we have developed in this paper also addresses this type of across-stage inferencing capability, in addition to the multimodal retrieval capability.

Based on the motivation to develop such a technique for the multimodal data mining in a multimedia database, we propose a multimodal data mining approach called co-learning framework that is based on the Multiple-Instance Learning (MIL) theory (Dietterich et al., 1997; Maron & Lozano-Perez, 1998; Auer, 1997). While this co-learning framework is general for any specific domains, to demonstrate the effectiveness of this framework, we apply this framework to the Berkeley *Drosophila* (fruit-fly) ISH embryo image database¹. In addition, we have also compared this co-learning framework on this database with a state-of-the-art multimodal data mining method to demonstrate the effectiveness and the promise of the framework.

RELATED WORK

In the machine learning community, MIL has become a focused topic in recent years and has received extensive attention in the literature ever since the classic work of (Dietterich et al., 1997; Auer, 1997; Maron & Lozano-Perez, 1998). Recent developments on MIL include (Andrews et al., 2003; Andrews & Hofmann; 2004; Rahmani & Goldman, 2006). (Yang & Lozano-Perez, 2000) and (Zhang et al., 2002) were among the first to apply MIL to image retrieval, which led to more subsequent work on this topic (Zhang et al., 2006; Zhu et al., 2006).

(Chen et al., 2006) recently added the embedded instance selection principle into the classic MIL algorithm resulting in a better learning performance, and also applied this method to image retrieval.

On the other hand, image data mining, and in particular image retrieval, has been studied for over a decade. One of the notorious bottleneck of image retrieval is the semantic gap (Smeulders et al., 2000). Recently, it is reported that this bottleneck may be effectively reduced using multimodal approaches (Barnard et al., 2003; Duygulu et al., 2002; Feng et al., 2004; Zhang et al., 2005) by taking the advantage that in many applications imagery data do not exist in isolation but typically co-exist with other modalities of information such as text. It is demonstrated in the literature

(Jeon & Manmatha, 2004; Feng et al., 2004; Zhang et al., 2005) that given such a presence of the multimodal data, there are effective methods to reduce the semantic gap by exploiting the synergy among the different modalities of the data.

Following this line of reasoning, multimodal data mining, and in particular, multimodal data retrieval, has recently received substantial attention since Barnard and Duygulu et al. started their pioneering work on image annotation (Barnard et al., 2003, Duygulu et al., 2002). Recent work includes (Jeon et al., 2003; Jeon & Manmatha, 2004; Chang et al., 2003;

17 more pages are available in the full version of this document, which may be purchased using the "Purchase" button on the product's webpage:

www.irma-international.org/article/multiple-instance-learning-based-approach/42970/

Related Content

Development of Indonesia's National Digital Library Network

Ismail Fahmi (2005). *Design and Usability of Digital Libraries: Case Studies in the Asia Pacific* (pp. 38-54).

www.irma-international.org/chapter/development-indonesia-national-digital-library/8131/

Logical Structure Recovery in Scholarly Articles with Rich Document Features

Minh-Thang Luong, Thuy Dung Nguyen, and Min-Yen Kan (2010). *International Journal of Digital Library Systems* (pp. 1-23).

www.irma-international.org/article/logical-structure-recovery-scholarly-articles/48200/

A Content-Driven System Architecture for Tackling Automatic Cataloging of Animated Movie Databases

Bogdan Ionescu, Alexandru Marin, Patrick Lambert, Didier Coquin, and Constantin Vertan (2010). *International Journal of Digital Library Systems* (pp. 1-23).

www.irma-international.org/article/content-driven-system-architecture-tackling/42969/

Comparing Open Source Digital Library Software

George Pyrounakis, and Mara Nikolaidou (2009). *Handbook of Research on Digital Libraries: Design, Development, and Impact* (pp. 51-60).

www.irma-international.org/chapter/comparing-open-source-digital-library/19869/

Collaborative Digital Library Development in India: A Network Analysis

Anup Kumar Das, B. K. Sen, and Chaitali Dutta (2010). *Developing Sustainable Digital Libraries: Socio-Technical Perspectives* (pp. 206-222).

www.irma-international.org/chapter/collaborative-digital-library-development-india/42744/