Chapter 28
A Highly Scalable and Adaptable Co-Learning Framework on Multimodal Data Mining in a Multimedia Database

Zhongfei (Mark) Zhang
SUNY Binghamton, USA

Zhen Guo
SUNY Binghamton, USA

Christos Faloutsos
Carnegie Mellon University, USA

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

ABSTRACT

This chapter presents a highly scalable and adaptable co-learning framework on multimodal data mining in a multimedia database. The co-learning framework is based on the multiple instance learning theory. The framework enjoys a strong scalability in the sense that the query time complexity is a constant, independent of the database scale, and the mining effectiveness is also independent of the database scale, allowing facilitating a multimodal querying to a very large scale multimedia database. At the same time, this framework also enjoys a strong adaptability in the sense that it allows incrementally updating the database indexing with a constant operation when the database is dynamically updated with new information. Hence, this framework excels many of the existing multimodal data mining methods in the literature that are neither scalable nor adaptable at all. Theoretic analysis and empirical evaluations are provided to demonstrate the advantage of the strong scalability and adaptability. While this framework is general for multimodal data mining in any specific domains, to evaluate this framework, the authors apply it to the Berkeley Drosophila ISH embryo image database for the evaluations of the mining performance. They have compared the framework with a state-of-the-art multimodal data mining method to demonstrate the effectiveness and the promise of the framework.

DOI: 10.4018/978-1-4666-2455-9.ch028

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 chapter, 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 chapter, 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 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 chapter, 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 chapter 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 co-learning framework. This co-learning framework 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 database[1]. 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.

This chapter is organized as follows. This Introduction section is followed by a brief literature review for the most related work. Then the co-learning framework based on the MIL theory is presented. A theoretic analysis on the scalability of the framework when the database is updated is reported. Finally, the extensive evaluations on the retrieval and across-stage inferencing based on the Berkeley Drosophila ISH data are documented in comparison with a state-of-the-art multimodal retrieval method. We also report the empirical evaluations using the same data set to demonstrate the scalability advantage of the framework. The chapter is concluded at the end.

RELATED WORK

In the machine learning community, MIL has become a focused topic in recent years and has