Mining Repetitive Patterns in Multimedia Data

Junsong Yuan
Northwestern University, USA

Ying Wu
Northwestern University, USA

INTRODUCTION

One of the focused themes in data mining research is to discover frequent and repetitive patterns from the data. The success of frequent pattern mining (Han, Cheng, Xin, & Yan, 2007) in structured data (e.g., transaction data) and semi-structured data (e.g., text) has recently aroused our curiosity in applying them to multimedia data. Given a collection of unlabeled images, videos or audios, the objective of repetitive pattern discovery is to find (if there is any) similar patterns that appear repetitively in the whole dataset. Discovering such repetitive patterns in multimedia data brings in interesting new problems in data mining research. It also provides opportunities in solving traditional tasks in multimedia research, including visual similarity matching (Boiman & Irani, 2006), visual object retrieval (Sivic & Zisserman, 2004; Philbin, Chum, Isard, Sivic & Zisserman, 2007), categorization (Grauman & Darrell, 2006), recognition (Quack, Ferrari, Leibe & Gool, 2007; Amores, Sebe, & Radeva, 2007), as well as audio object search and indexing (Herley, 2006).

- In image mining, frequent or repetitive patterns can be similar image texture regions, a specific visual object, or a category of objects. These repetitive patterns appear in a sub-collection of the images (Hong & Huang, 2004; Tan & Ngo, 2005; Yuan & Wu, 2007, Yuan, Wu & Yang, 2007; Yuan, Li, Fu, Wu & Huang, 2007).
- In video mining, repetitive patterns can be repetitive short video clips (*e.g.* commercials) or temporal visual events that happen frequently in the given videos (Wang, Liu & Yang, 2005; Xie, Kennedy, Chang, Divakaran, Sun, & Lin, 2004; Yang, Xue, & Tian, 2005; Yuan, Wang, Meng, Wu & Li, 2007).
- In audio mining, repetitive patterns can be repeated structures appearing in music (Lartillot, 2005) or broadcast audio (Herley, 2006).

Repetitive pattern discovery is a challenging problem because we do not have any *a prior* knowledge of the possible repetitive patterns. For example, it is generally unknown in advance (i) what the repetitive patterns look like (e.g. shape and appearance of the repetitive object/contents of the repetitive clip); (ii) where (location) and how large (scale of the repetitive object or length of the repetitive clip) they are; (iii) how many repetitive patterns in total and how many instances each repetitive pattern has; or even (iv) whether such repetitive patterns exist at all. An exhaustive solution needs to search through all possible pattern sizes and locations, thus is extremely computationally demanding, if not impossible.

BACKGROUND

As a purely unsupervised task, repetitive pattern discovery is different from traditional pattern detection/retrieval problem, where a pattern template or a query example is provided and the task is to find its re-occurrences (Divakara, Peker, Chang, Radhakrishnan, & Xie, 2004). When mining repetitive patterns, we have no *a prior* knowledge of the pattern. Therefore, it differs from supervised learning problems (*e.g.* classification and retrieval) that have been widely studied in machine learning research.

Repetitive pattern discovery is also different from unsupervised tasks like clustering. The task of clustering is to partition a collection of data samples into several disjoint groups, where data samples belonging to the same group are similar to each other, whereas data samples from different groups are dissimilar. Instead of clustering individual data samples, pattern discovery aims at repetitive patterns that are shared by some data samples. Such a repetitive pattern describes the common characteristics among the data. For example, given an image dataset, a repetitive pattern can be a sub-image region (e.g. a visual object) that appears in

many images. Thus the repetitive pattern corresponds to an image region while not the whole image.

MAIN FOCUS

Multimedia data mining is different from transaction and text data mining mainly in two aspects. Firstly, transaction and text data are discrete data, for example, transactions are composed of items and texts are composed of vocabularies. However, multimedia data are usually characterized by continuous features and these features generally exhibit much larger variabilities and uncertainties than predefined items and vocabularies. Taking the visual pattern in image data for example, the same visual pattern (e.g. a visual object) is likely to exhibit quite different visual appearances under different lighting conditions, views, scales, not to mention partial occlusion. As a result, it is very difficult find invariant visual features that are insensitive to these variations to uniquely characterize visual patterns.

Another characteristic of multimedia data mining, besides the features uncertainties, is that multimedia patterns have more complex structure than transaction and text patterns. For example, the spatial configuration is important in characterizing an image pattern. Therefore, when discovering repetitive patterns from images, the difficulty of representing and discovering spatial patterns prevents straightforward generalization of traditional frequent pattern mining methods that are applicable to transaction data, which are orderless sets of items, or text data, which are represented by strings of words. Although there exist methods for spatial collocation pattern discovery from geo-spatial data (Huang, Shekhar, & Xiong, 2004), they cannot be directly applied to image data which are characterized by high-dimensional continuous features. Similarly, when mining audio and video data, we should take the temporal structure into consideration, as the repetitive patterns are usually temporal sequences.

The main focus of repetitive pattern discovery in multimedia data mining is to develop methods that address the aforementioned two challenges. In the following, we summarize three major components of repetitive pattern discovery in multimedia data: (1) pattern representation, (2) pattern matching and (3) pattern mining. We will discuss the specific challenges and the-state-of-the-art techniques in each component.

Pattern Representation

Pattern representation is important for general machine learning and data mining tasks. In multimedia data mining, it determines how we represent the image/video/audio data, typically in terms of features and data models. Data mining algorithms perform directly on the extracted features instead of the raw multimedia data. Due to the page limit, we do not list all possible features that can be extracted from the multimedia data. Instead, we discuss the requirements of the features for repetitive pattern discovery.

- For repetitive pattern discovery in image data, it is usually desirable that the extracted visual features are robust under different pattern variations like rotation, scale changes and even affine transformations. Recently, it is found that the "bag of features" representation is beneficial for pattern discovery because it can handle pattern variations as well as incomplete patterns. The basic visual features are local visual primitives detected from images (Philbin, Chum, Isard, Sivic & Zisserman, 2007; Quack, Ferrari, Leibe & Gool, 2007; Sivic & Zisserman, 2004; Yuan, Wu, & Yang, 2007). Each visual primitive describes an informative local image region and is expected to be invariant under the following variations. (Mikolajczyk, Tuytelaars, Schmid, Zisserman, Matas, Schaffalitzky, Kadir, & Gool, 2005).

1. **View point changes:** The same image pattern can look different under different view-points, depending on the capture angle of the camera.
2. **Scale changes:** The same image pattern can look smaller or larger, depending on the distance of the camera to the object, as well as the zoom in/out parameters of the camera.
3. **Lighting condition changes:** The appearances of a same image pattern may vary according to different lighting conditions at the capture.
4. **Partial Occlusions:** It is possible that a pattern (e.g. a repetitive object) is occluded in the image. Such an incomplete pattern, however, could still be of interests and want to be discovered.