Chapter VII
Visual Data Mining Based on Partial Similarity Concepts

Juliusz L. Kulikowski
Polish Academy of Sciences, Poland

ABSTRACT

Visual data mining is a procedure aimed at a selection from a document’s repository subsets of documents presenting certain classes of objects; the last may be characterized as classes of objects’ similarity or, more generally, as classes of objects satisfying certain relationships. In this chapter attention will be focused on selection of visual documents representing objects belonging to similarity classes.

INTRODUCTION

In numerous scientific research, medical examination, technological applications, education, artistic activity, and other areas visual data play a substantial, sometimes irreplaceable role. Intensive progress in image acquisition, recording, enhancement and storage technology in the last decades caused a dramatic increase in total volume of visual documents stored in archives and repositories of various types over the world. However, because of legal or organizational difficulties and low effectiveness of available remote data access and visual information retrieval methods only a small part of useful information from visual documents can be mined and effectively used in new investigations, decision making and/or in application tasks solution. This is why the interest to new methods of data mining from visual documents is still actual.

A visual document is here considered as an electronic digital file consisting of a textual and a graphical part, the last being, in fact, a digital representation of a graph or an image. The textual part of the document may contain some formal characteristics of the document as a whole (numerical identifier, emission date, source, accessibility conditions, etc.) as well as of its graphical part: image contents char-
characteristics (string of key words, a semantic classification code, etc.), code of image modality, technical parameters, quality characteristics, and so forth. The graphical part of a visual document is assumed to present graphs or images in one of admissible standard forms that makes it possible to be displayed and visualized in order to be manually or automatically processed.

Image processing means here image selection, enhancement, compression/decompression, transmission, visualization as well as any type of image analysis: functional transformation, parameters extraction, recognition, contents’ interpretation, application value assessment, and so forth.

Visual data mining is a procedure aimed at a selection from a documents’ repository subsets of documents presenting certain classes of objects; the last may be characterized as classes of objects’ similarity or, more generally, as classes of objects satisfying certain relationships. In this chapter attention will be focused on selection of visual documents representing objects belonging to similarity classes.

GENERAL CONCEPT OF SIMILARITY

Observing the surrounding world as an extremely sophisticated variety of objects, relationships between them and processes we try to put order on it for making it more comprehensible. One of ways to reach this is establishment of similarities between objects or processes which make it possible to apply some statements about individual objects or processes to the members of their similarity classes. This way of action is not only a human discovery. A wild animal first-time testing an apple and finding it good fixes in its mind a pattern of similar size, form, color and scent objects. In its later experiences the pattern may be corrected by exclusion the small size and dark-green color objects which finally leads to a fixation in the mind a pattern (a concept) of “good for being eaten” objects. Establishment of similarity concepts and detection of similarities in individual cases in many living beings is thus a capability of fundamental importance. Non less important seems explaining what, in general, does the similarity mean.

Strong similarity. In the forthcoming, the notion object will be assigned to any formal representative (geometrical point, vector, string of symbols, etc.) of a real or abstract being, characterizing its features and identifying it among other beings. Similarity is a property of a class \( C \) of objects that in the simplest case can be described by a formal similarity relation [Rasiowa H., Sikorski R., 1968]:

1. reciprocal (each object \( a \) is similar to itself),
2. symmetrical (if object \( a \) is similar to object \( b \) then \( b \) is similar to \( a \)), and
3. transitive (if \( a \) is similar to \( b \) and \( b \) is similar to \( c \) then \( a \) is similar to \( c \)).

For the reasons that will be explained below, the above-defined similarity can be called a strong similarity concept. Similarity of triangles is a well known example of similarity in the above-given sense: two triangles on an Euclidean plane are similar if angular measures of the pairs of respective angles are equal. This property can easily be extended on any set of triangles. Similarity of polygons on a plane can also be based on the concept of pair-wise equality of angular measures of the pairs of respective angles.

Strong similarity can be based on various similarity criteria assigning different sense of similarity to the objects under consideration. For example, it is possible to assume that in the set \( C \) of triangles “similar” are the triangles of the same surface measure, of the same sum of lengths of edges, etc.
Related Content

Assuring Database Integrity
[www.igi-global.com/article/assuring-database-integrity/51075?camid=4v1a](www.igi-global.com/article/assuring-database-integrity/51075?camid=4v1a)

Vertical Database Design for Scalable Data Mining
[www.igi-global.com/chapter/vertical-database-design-scalable-data/11232?camid=4v1a](www.igi-global.com/chapter/vertical-database-design-scalable-data/11232?camid=4v1a)

Data Model Versioning and Database Evolution
[www.igi-global.com/chapter/data-model-versioning-database-evolution/11131?camid=4v1a](www.igi-global.com/chapter/data-model-versioning-database-evolution/11131?camid=4v1a)

Assigning Ontological Meaning to Workflow Nets
Pnina Soffer, Maya Kaner and Yair Wand (2012). *Cross-Disciplinary Models and Applications of Database Management: Advancing Approaches* (pp. 209-244).
[www.igi-global.com/chapter/assigning-ontological-meaning-workflow-nets/63668?camid=4v1a](www.igi-global.com/chapter/assigning-ontological-meaning-workflow-nets/63668?camid=4v1a)