IDEA GROUP PUBLISHING

701 E. Chocolate Avenue, Suite 200, Hershey PA 17033-1240, USA

Tel: 717/533-8845; Fax 717/533-8661; URL-http://www.idea-group.com

ITB10790

Chapter XIII

Similarity Learning in GIS: An Overview of Definitions, Prerequisites and Challenges

Giorgos Mountrakis, University of Maine, USA

Peggy Agouris, University of Maine, USA

Anthony Stefanidis, University of Maine, USA

Abstract

In this chapter we review similarity learning in spatial databases. Traditional exact-match queries do not conform to the exploratory nature of GIS datasets. Non-adaptable query methods fail to capture the highly diverse needs, expertise and understanding of users querying for spatial datasets. Similarity-learning algorithms provide support for user preference and should therefore be a vital part in the communication process of geospatial information. More specifically, we address machine learning as applied in the optimization of query similarity. We review appropriate definitions of *similarity and we position similarity learning within data mining and machine learning tasks. Furthermore, we outline prerequisites for similarity learning techniques based on the unique characteristics of the GIS domain. A description of specific methodologies follows based on the highly diverse attributes of GIS datasets (for example, text, images, video), and application examples are presented. We summarize previously set requirements and present future trends expected to emerge in the coming years.*

Introduction

Databases containing geographic information have been through substantial changes in the past decade. Improvements in acquisition methods coupled with increased information dissemination through the Internet have altered substantially their characteristics. Nowadays geospatial databases are increasingly incorporating temporal information in order to enable novel analysis capabilities through spatiotemporal modeling. They may also incorporate diverse information types like video, text and sound, in addition to the traditional raster, vector and thematic types. All these changes are expanding the applicability of geographic information systems (GIS), while at the same time increasing the difficulty in modeling and managing their content.

Problem Description

The increase in geospatial information availability has been matched by an expansion of the relevant user community, and an increase in the complexity of demands by such users. Thus, even though there exist increasing volumes of geospatial information, there are very good chances that a GIS query may not have an exact match in a corresponding database. To address this problem, *similarity* algorithms have been developed. The goal of a similarity algorithm is to compare each query to the available information in the database and produce a metric (or ranking). This metric expresses how close (appropriate) each of the available answers is to the query. For example, when a user requests an aerial photograph from 1956, the similarity algorithm will compare the request (Time = 1956) to the temporal footprint of the aerial photographs available in a database, and will rank these photographs based on their similarity to the user request. A popular example of such a similarity algorithm is nearest neighbor: The photograph that is closest (temporally) to 1956 would be the best choice.

Traditional similarity algorithms have been widely accepted because their simplicity allows fast answers to user queries, accompanied by preference

Related Content

An Investigation of the Impact of Organization Size on Data Quality Issues
G. Daryl Nord, Jeretta Horn Nordand Hongjiang Xu (2005). *Journal of Database Management (pp. 58-71).*
www.irma-international.org/article/investigation-impact-organization-size-data/3337

Multiple Query Optimization with Depth-First Branch-and-Bound and Dynamic Query Ordering
Ahmet Cosar, Ee-Peng Limand Jaideep Srivastava (1995). *Journal of Database Management (pp. 14-19).*
www.irma-international.org/article/multiple-query-optimization-depth-first/51143

Designing Information Systems Capabilities to Create Business Value: A Theoretical Conceptualization of the Role of Flexibility and Integration
Christoph Schlueter Langdon (2009). *Database Technologies: Concepts, Methodologies, Tools, and Applications (pp. 866-885).*
www.irma-international.org/chapter/designing-information-systems-capabilities-create/7946

TOS: A Temporal Object-Oriented System
Farshad Fotouhi, Imran Ahmad, William I. Groskyand Abad Shah (1994). *Journal of Database Management (pp. 3-15).*
www.irma-international.org/article/tos-temporal-object-oriented-system/51138

Rapid Privacy Preserving Algorithm for Large Databases
K. Anbumaniand R. Nedunchezhian (2009). *Database Technologies: Concepts, Methodologies, Tools, and Applications (pp. 1050-1061).*
www.irma-international.org/chapter/rapid-privacy-preserving-algorithm-large/7958