

Chapter 5

Feature Extraction

ABSTRACT

This chapter is an illustration of feature extraction for working with large datasets. The basic definition of feature extraction, selection of effective features, and the existing problems and solutions are provided. How feature extraction maps the high dimensional space to smaller space is explained.

5.1 INTRODUCTION

Feature extraction is an attribute reduction process. Unlike feature selection, which ranks the existing attributes according to their predictive significance, feature extraction actually transforms the attributes. The transformed attributes, or features, are linear combinations of the original attributes.

Digital libraries are places for handling a vast amount of data and information. This provides the ability to access and interact to a lot of documents in the form of electronic version. Schatz (1997) has a definition for digital libraries: “A digital library enables users to interact effectively with information distributed across a network”.

A digital library works as a network information system to support some tasks such as search and display of an item in a database. Once comfortable with the new tools, they demand new materials to be available in digital libraries. Any task in the digital library requires the digital representation of the data. Extending a digital library is an easy task, because of faster and cheaper process than the physical library. Obviously, this increase in the amount of information has a strong impact on the supporting software.

DOI: 10.4018/978-1-5225-5580-3.ch005

Three are multimedia content in the electronic version of text data, which is an issue. To obtain the image in the document, one can provide the query and ask for all the electronic documents that contain similar pictures. Digital images and any other multimedia data for that matter are *complex data*. Computers have the ability to represent and manipulate the digital data. However, decoding the contents is an issue for researching (Manolescu, 2000).

5.1.1 Problem

The question is that how a software can handle the requirements related to the applications to deal with large amounts of information, similarity searching and complex data.

5.1.2 Forces

- Difficulties in information retrieval systems with large amounts of data.
- Difficulties in similarity searching.
- Difficulties in multimedia databases containing digital representations of acoustic and visual data.
- Difficulties in information retrieval systems for fast responding time.

5.1.3 Mapping the Data

The solution to overcome the difficulties is to work with a simpler presentation of the data. The representation contains the unique form for each data. There are functions for these computations. It means that the function maps the data from the problem space into a feature space. For this reason, it is also called a “feature extraction function”. Feature extraction is mapping the data from a larger problem space into a smaller feature space. Therefore, feature extraction enables scalable solutions for problems that deal with large amounts of information (Manolescu, 2000).

Feature extraction has some limitation for similarity searching problem. Any data items are similar as long as their representation in the feature space are similar. But this is not necessarily how humans perceive things. Two reasons for this are the following:

6 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/chapter/feature-extraction/205178

Related Content

Ethnic Use, Phytochemistry, and Pharmacology of *Cyperus rotundus*: A Medicinal Plant

Mohammed Rahmatullah, Khoshnur Jannat, Gerald R. Reeck, Rownak Jahan, Taufiq Rahman, Nasrin A. Shovaand Maidul Islam (2022). *Research Anthology on Recent Advancements in Ethnopharmacology and Nutraceuticals* (pp. 431-453).

www.irma-international.org/chapter/ethnic-use-phytochemistry-and-pharmacology-of-cyperus-rotundus/289495

Service Level Agreements for Smart Healthcare in Cloud

Mridul Pauland Ajanta Das (2020). *Virtual and Mobile Healthcare: Breakthroughs in Research and Practice* (pp. 1-14).

www.irma-international.org/chapter/service-level-agreements-for-smart-healthcare-in-cloud/235302

Herbal Drug Interactions

Mymoona Akhter (2018). *Complementary and Alternative Medicine and Kidney Health* (pp. 201-231).

www.irma-international.org/chapter/herbal-drug-interactions/191968

Sialadenitis and Sialadenosis

(2021). *Diagnostic Techniques and Therapeutic Strategies for Parotid Gland Disorders* (pp. 136-157).

www.irma-international.org/chapter/sialadenitis-and-sialadenosis/256614

Nutrition and Cancer

Shazia Ali, Asiya Batool, Asma Tariq, Aashiq Hussain Bhat, Abid Qureshi, Basharat Bashir Teli, Iqbal Qasim, Sumit Kumar, Abhilash Gangadharan, Raj Kumarand Barre Vijay Prasad (2021). *Handbook of Research on Advancements in Cancer Therapeutics* (pp. 570-586).

www.irma-international.org/chapter/nutrition-and-cancer/267058