Visual Attribute Learning and Ontology Engineering

Yang Long

Electronic and Electrical Engineering

University of Sheffield

This dissertation is submitted for the degree of

Doctor of Philosophy

University of Sheffield April 2015

Abstract

Apart from discovering data structure and make it high-level, it is also worthy to concern how to decompose our tasks so that we humans can approach the machine from our side. To this end, ontology engineering aims at refining the human knowledge into the nature, and conversely build complex high-level tasks onto it. However, the knowledge based ontology is not necessary visually related. Therefore, my target is to design ontology structure based on visual attributes learning, which bridges the gap between human taxonomy and pixel-level pattern classification. This involves the discussion about not only "how to design vision" but also the essential question of "what is vision."

So far, I have implemented several initial works that introduced the ideal of spacial-related attributes. Substantial experiments have shown promising results. Based on these achievements, my future works may focus on more challenging and complex attributes. Meanwhile, a large-scaled ontology project is on proposal as well.

Table of contents

List of figures

List of tables

Chapter 1

Introduction

1.1 Learning Visual Attributes

Vision is an important perceptive approach. Computer vision community has dedicated to improve the visual ability of machines so as to implement high-quality human-machine interaction. My research dedicates to machines to see and to understand. The first challenge is how to make the pixel-level digital information understandable in order to achieve high-level tasks. Thankfully, machine learning community has provided sufficient technical supports. However, as the digital technology is explosively increasing, large-scaled visual information is produced from various types of camera sensors, i.e. photos, films, surveillance, etc. As a result, billions of visual sources in various modalities requires uncountable computing power. The traditional machine learning algorithms tend to become incompetent when dealing with such large-scaled fine-grained learning tasks. Recently, it is proposed that a high-level task can be substituted by several subtasks of learning semantic concepts, which is known as Attribute Learning. Because attributes are human-defined properties that abstracted from the reality, it is not restricted to a certain information source and can help to design generative modal for different modalities based on human requirement. In addition, the knowledge-based visual attributes can be shared among bottom-up tasks. Such properties of attribute learning provides a possible solution to abstract the massive data into semantically understandable representations.

1.2 Ontology Engineering in Computer Vision

1.3 Extensive Applications

Chapter 2

Literature Review

The remarkable progresses on computer vision and machine learning technologies have substantially extended their applicable range. Among these applications, visual attribute learning has attracted increasingly attentions. Meanwhile, ontology engineering has been introduced to provide fundamental knowledge for learning attributes. In this chapter, the most classical machine leaning algorithms are reviewed firstly. In the second section, the state-of-the-art attribute learning methods are reviewed. Finally, existing ontology-based datasets are summarized.

2.1 Classical Algorithms

A typical visual task can be formalised as following: given a sample of x, e.g. image, video or any other information source, the goal is to find a function so that x can be project to a target domain $\mathscr{Y}$. Normally, the final goal of a task has a concrete physical meaning. For instance, x is an image of a human face, and $y \in \mathscr{Y}$ is a label value that indicating the identity of x. Such a task is known as Face Recognition. However, most of the tasks can hardly achieved by a single function: $f : x \rightarrow y$. In the next, the typical framework for such task is reviewed by three steps: feature extraction, data mining and classification.

Base Features

Data Mining

Classification

2.2 Attribute Learning Methodology

2.3 Ontology-based Datasets

Chapter 3

Undertaken Research

3.1 Learning Spatial Attributes

3.1.1 Multi-view Human Action Recognition

3.1.2 Spatial Pooling for Face Recognition

3.2 Implicit Attributes

Chapter 4

Summery and Future Works

4.1 Attribute Taxonomy and Ontology

4.2 Ensemble Learning and Graph Theory

References