Pattern recognition

Introduction

The term pattern recognition refers to the task of placing some object to a correct class based on the measurements about the object. Usually this task is to be performed automatically with the help of computer. Objects to be recognized, measurements about the objects, and possible classes can be almost anything in the world. For this reason, there are very different pattern recognition tasks. A system that makes measurements about certain objects and thereafter classifies these objects is called a pattern recognition system. A spam (junk-mail) filter is another example of pattern recognition systems. A spam filter recognizes automatically junk e-mails and places them in a different folder (e.g. /dev/null) than the user's inbox. The list of pattern recognition systems is almost endless. Pattern recognition has a number of applications ranging from medicine to speech recognition.

Some pattern recognition tasks are everyday tasks (e.g. speech recognition) and some pattern recognition tasks are not-so-everyday tasks.

Syntactic and Statistical pattern recognition form classes of pattern recognition methods. The basic idea of syntactic pattern recognition is that the patterns (observations about the objects to be classified) can always be represented with the help of simpler and simpler sub patterns leading eventually to atomic patterns which cannot anymore be decomposed into sub patterns. Pattern recognition is then the study of atomic patterns and the language between relations of these atomic patterns. The theory of formal languages forms the basis of syntactic pattern recognition.

Pattern Recognition Systems

2.1Examples

2.1.1 Optical Character Recognition (OCR)

Optical character recognition (OCR) means recognition of alpha-numeric letters based on image-input. The difficulty of the OCR problems depends on whether characters to be recognized are hand-written or written out by a machine (printer). The difficulty level of an OCR problem is additionally influenced by the quality of image input, if it can be assumed that the characters are within the boxes reserved for them as in machine readable forms, and how many different characters the system needs to recognize.

Basic Structure of Pattern Recognition Systems

The task of the pattern recognition system is to classify an object into a correct class based on the measurements about the object. Note that possible classes are usually well-defined already before the design of the pattern recognition system. Many pattern recognition systems can be thought to consist of five stages:

1. Sensing (measurement);
2. Pre-processing and segmentation;
3. Feature extraction;
4. Classification;
5. Post-processing;

Sensing refers to some measurement or observation about the object to be classified. For example, the data can consist of sounds or images and sensing equipment can be a microphone array or a camera. Often one measurement (e.g. image) includes information about several objects to be classified. For instance, assume that we want to recognize the address written on the envelope. We must then classify several characters to recognize the whole address. The data here is probably an image of the envelope including the sub-images of all the characters to be classified and some background that has nothing to do with the pattern recognition task.

Pre-processing refers to filtering the raw data for noise suppression and other operations performed on the raw data to improve its quality. In segmentation, the measurement data is partitioned so that each part represents exactly one object to be classified. For example in address recognition, an image of the whole address needs to be divided to images representing just one character. The result of the segmentation can be represented as a vector that is called a pattern vector.

Feature extraction. Especially when dealing with pictorial information the amount of data per one object can be huge. A high resolution facial photograph (for face recognition) can contain 1024*1024 pixels. The pattern vectors have then over a million components. The most part of this data is useless for classification. In feature extraction, we are searching for the features that best characterize the data for classification The result of the feature extraction stage is called a feature vector. The space of all possible feature vectors is called the feature space. In face recognition, a widely used technique to reduce the number features is principal component analysis (PCA). (This yields so called eigenfaces.) PCA is a statistical technique to reduce the dimensionality of a data vector while retaining most of the information that the data vector contains. In addition to mathematical/computational/statistical techniques, feature extraction can be performed heuristically by picking such features from a pattern vector that could be assumed to be useful in classification. For face recognition, such a feature can be the distance between the two eyes of a person. Sometimes, dimensionality of the input data is very limited and the pattern vector can be chosen as the feature vector. For example this is the case with the image segmentation methods that are based on the pattern recognition principles. Feature extraction is highly application specific although some general techniques for feature extraction and selection have been developed. You will hear more about these during the course ’Pattern Recognition’.

In general, the line between feature extraction and classification is fuzzy. The task of the feature extractor is to produce a representation about the data that enables an easy classification. On the other hand, the task of the classifier is to produce the best classification accuracy given the extracted features. Clearly, these two stages are interdependent from the application point of view. Also, and perhaps more importantly, what is the best representation for the data depends on the classifier applied. There is no such thing as the universally optimal features.

The classifier takes as an input the feature vector extracted from the object to be classified. It places then the feature vector (i.e. the object) to class that is the most appropriate one. In address recognition, the classifier receives the features extracted from the sub-image containing just one character and places it to one of the following classes: ’A’,’B’,’C’..., ’0’,’1’,...,’9’. The classifier can be thought as a mapping from the feature space to the set of possible classes. Note that the classifier cannot distinguish between two objects with the same feature vector.

The main content of this course is within the classifier design. There is a simple reason: The abstraction offered by the concept of the feature vector makes it possible to develop a general, application-independent theory for the design of classifiers. Therefore, it is possible to use the same underlying principles when designing classifiers for address recognition systems and bottle recycling machines.

Post-processing. A pattern recognition system rarely exists in a vacuum. The final task of the pattern recognition system is to decide upon an action based on the classification result(s). A simple example is a bottle recycling machine, which places bottles and cans to correct boxes for further processing. Different actions can have also different costs associated with them. This information can be included to the classifier design as we will see later on. Also, we can have several interdependent classification results in our hands. For example in an address recognition task, we have the classification results for multiple characters and the task is to decide upon the address that these characters form. Therefore, it is possible to use the context: In this case, the information about the other classification results to correct a possible misclassification. If the result of the classifier is ’Hollywoud Boulevard’ then the address is probably ’Hollywood Boulevard’.