Introduction to Machine Learning

IWIS Seminar Report

Yujian Lin

Sep.19,2008

Content

➢ What is machine learning
➢ Machine Learning Issues
   algorithms
   applications
➢ Automatic Webpage Classifier

What is machine learning (ML)

- Any computer program that improves its performance at some task through experience, we say it has the learning ability.

- Various definitions about machine learning.

- Tom M. Mitchell:

  A computer program is said to *learn* from experience E with respect to some tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E.

ML examples

- A checkers learning problem

  T: playing checkers

  P: percent of games won against opponents

  E: playing practice games against itself

- A handwriting recognition learning problem

- A robot driving learning problem

- A human face recognition problem

- and so on

ML issues

- Algorithms

  to develop algorithms that can efficient solve
   practical problems
    Decision Tree, ANN, Bayesian Learning, SVM
  and so on

- Application

  Data Mining, Pattern Recognition and so on

Automatic Webpage Classifier (AWC)

- Internship Project at Poson Technology Ltd. of China Telecom, Guangzhou

- Project Description

- ✓ too much information over the Internet, people hope that only WebPages they are interested in are provided

- ✓ Firstly, a certain number of WebPages (200) are chosen as training samples. They are divided into several categories, such as History, Sports, Financial, Cooking, and Others, and the number are counted for each category

Automatic Webpage Classifier (Cont.)

- Second, 1000 words are abstracted from the sample pages as features, i.e., each page is represented by 1000 attributes.
- Third, for each category, words frequency is computed.

- For a new coming page, calculate the posteriori probability of each category, and the category with the maximum probability is assumed to be the category which the new coming page should be put into.

Details about AWC

- Candidate words selection
- Prior Probability of the candidate words
- Naïve Bayes classifier
- An other approach: SVM

- Discussion