CRIEPI REPORT (SUMMARY)

Trends in Boosting Research and Applications

(和文: ブースティング手法の研究とその適用動向)

Background

An alternative approach to just selecting a single classifier with the best estimated generalization performance is to generate and combine multiple classifiers. This approach is applicable to a wide variety of learning algorithms and can lead to improved classifier performance. One such method is called "bootstrap aggregating" or *Bagging* introduced by Breiman and another is *Boosting* introduced by Freund and Schapire . These methods deal with creating and combining multiple classifiers, but differ in how the classifiers are trained and in how their outputs are combined. Especially, many researchers pay attention to boosting, because this method achieve a very good generalization performance in many benchmark data sets. However, this method has some unclear parts about the learning character. Therefore, it is important to analyze some boosting algorithms and investigate the potential capability of this method based on some applications.

Object

From practical point of view, this work investigates in trends of boosting research by analysis of some boosting algorithms, collects some applications based on boosting methods, and presents the potential capability of boosting methods is explored.

Major result

This report analyzes the research trend of boosting methods from a practical point of view and presents some applications based on these methods. We investigate in the following three features of boosting methods as results of the analysis.

1. The original AdaBoost, which was proposed by Freund and Schapire, is a hard margin classifier like a Support Vector Machine. AdaBoost is very useful and can achieve good generalization performance in the low noise regime (few outliers, low input noise).

2. In the noisy case (lots of outliers, high input noise), the generalization performance of AdaBoost becomes less comparative to other algorithms: AdaBoost concentrates too much to some outliers while generating a hard margin.
3. Now, some improved boosting methods are researched to avoid overfitting by using a soft margin concept. In this report, AdaBoost$_{reg}$ and ν-Arc are introduced as examples of improved boosting methods that can deal with noisy data and avoid overfitting. AdaBoost$_{reg}$ is briefly introduced as the first improved boosting method.

This report presents the following applications based on Boosting:

- Empirical evaluations of boosting methods on benchmark data sets
- Application A: Handwritten digit recognition
- Application B: A non-intrusive monitoring system for household electric appliances

The boosting methods are one of the new learning strategies. There are only a few applications based on these methods yet. However, boosting methods have high potential for many application fields and it is undoubted that the number of applications based on the boosting methods will be increasing as in the case of Support Vector Machines. Therefore, it is very important to collect and spread information about new strategies and applications of boosting.

Future Work

Support Vector Machine community made a great success for collecting and spreading research information by building Support Vector Machine homepage. Therefore, in order to collect and spread information about new strategy of boosting, it is one of the useful methods to build a representative web-site about *boosting* and we will plan to open and operate this representative web-site.

注) 和文報告書は作成しておりませんので、予めご了承ください。

CRIEPI REPORT : ER99001
キーワード : Machine Learning, Ensemble Learning, AdaBoost, Classification Problem, Margin

関連研究報告書

主 担 当 者 Takashi Onoda (Communication & Information Research Laboratory)
小野田 崇（情報研究所）

連 絡 先 Central Research Institute of Electric Power Industry
Communication & Information Research Laboratory, Administrative Section
(財)電力中央研究所 情報研究所 事務担当
Tel 03-3480-2111（代）, e-mail cil-rr-ml @criepi.denken.or.jp