

Chapter 2.9

Combining Classifiers and Learning Mixture-of-Experts

Lei Xu

Chinese University of Hong Kong, Hong Kong & Peking University, China

Shun-ichi Amari

Brain Science Institute, Japan

INTRODUCTION

Expert combination is a classic strategy that has been widely used in various problem solving tasks. A team of individuals with diverse and complementary skills tackle a task jointly such that a performance better than any single individual can make is achieved via integrating the strengths of individuals. Started from the late 1980' in the handwritten character recognition literature, studies have been made on combining multiple classifiers. Also from the early 1990' in the fields of neural networks and machine learning, efforts have been made under the name of ensemble learning or mixture of experts on how to learn jointly a mixture of experts (parametric models) and a combining strategy for integrating them in an optimal sense.

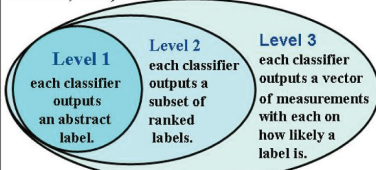
DOI: 10.4018/978-1-60960-818-7.ch2.9

The article aims at a general sketch of two streams of studies, not only with a re-elaboration of essential tasks, basic ingredients, and typical combining rules, but also with a general combination framework (especially one concise and more useful one-parameter modulated special case, called α -integration) suggested to unify a number of typical classifier combination rules and several mixture based learning models, as well as max rule and min rule used in the literature on fuzzy system. (Figure 1)

BACKGROUND

Both streams of studies are featured by two periods of developments. The first period is roughly from the late 1980s to the early 1990s. In the handwritten character recognition literature, various classifiers have been developed from different

Figure 1. Essential tasks and their implementations

Two Tasks (a quotation from Xu, Krzyzak and Suen, 1992)					
<p><i>Task 1:</i> "How many and what type of classifiers should be used for a specific application?, and for each classifier what type of features should we use?, as well as other problems that are related to the construction of those individual and complementary classifier".</p> <p><i>Task 2:</i> "How to combine the results from different existing classifiers so that a better result can be obtained?"</p>					
Two Styles of Implementations					
<p>Two Stage Implementation</p> <ul style="list-style-type: none"> • <i>Task 1</i> is completed in advance, with the resulted classifiers being diverse and complementary. • Perform <i>Task 2</i> in one of three levels (Xu, Krzyzak and Suen, 1992).  <p>Level 1: each classifier outputs an abstract label.</p> <p>Level 2: each classifier outputs a subset of ranked labels.</p> <p>Level 3: each classifier outputs a vector of measurements with each on how likely a label is.</p>	<p>Joint Implementation</p> <p>Two tasks made jointly or alternatively</p> <table border="1"> <thead> <tr> <th>under a same criterion</th> <th>others</th> </tr> </thead> <tbody> <tr> <td> <ul style="list-style-type: none"> • Mixture of experts (ME) (Jacobs, et al, 1991; Jordan & Jacobs, 1994); • Alternative ME (Xu & Jordan, 1993; Xu, Jordan & Hinton, 1994); • EM-RBF (Xu, 1998) • Three layer nets, etc. </td> <td> <p>Stacking, Boosting, ..., etc (Breiman, 1992&94; Wolpert, 1992)</p> </td> </tr> </tbody> </table>	under a same criterion	others	<ul style="list-style-type: none"> • Mixture of experts (ME) (Jacobs, et al, 1991; Jordan & Jacobs, 1994); • Alternative ME (Xu & Jordan, 1993; Xu, Jordan & Hinton, 1994); • EM-RBF (Xu, 1998) • Three layer nets, etc. 	<p>Stacking, Boosting, ..., etc (Breiman, 1992&94; Wolpert, 1992)</p>
under a same criterion	others				
<ul style="list-style-type: none"> • Mixture of experts (ME) (Jacobs, et al, 1991; Jordan & Jacobs, 1994); • Alternative ME (Xu & Jordan, 1993; Xu, Jordan & Hinton, 1994); • EM-RBF (Xu, 1998) • Three layer nets, etc. 	<p>Stacking, Boosting, ..., etc (Breiman, 1992&94; Wolpert, 1992)</p>				

methodologies and different features, which motivate studies on combining multiple classifiers for a better performance. A systematical effort on the early stage of studies was made in (Xu, Krzyzak & Suen, 1992), with an attempt of setting up a general framework for classifier combination. As re-elaborated in Figure1, not only two essential tasks were identified and a framework of three level combination was presented for the second task to cope with different types of classifier's output information, but also several rules have been investigated towards two of the three levels, especially with Bayes voting rule, product rule, and Dempster-Shaferrule proposed. Subsequently, the rest one (i.e., rank level) was soon studied in (Ho, Hull, & Srihari, 1994) via Borda count.

Interestingly and complementarily, almost in the same period the first task happens to be the focus of studies in the neural networks learning literature. Encountering the problems that there are different choices for the same type of neural net by varying its scale (e.g., the number of hidden units in a three layer net), different local optimal results on the same neural net due to different initializations, studies have been made on how to train an ensemble of diverse and complementary networks via cross-validation- partitioning, cor-

relation reduction pruning, performance guided re-sampling, etc, such that the resulted combination produces a better generalization performance (Hansen & Salamon, 1990; Xu, Krzyzak, & Suen, 1991; Wolpert, 1992; Baxt, 1992, Breiman, 1992&94; Drucker, et al, 1994). In addition to classification, this stream also handles function regression via integrating individual estimators by a linear combination (Perrone & Cooper, 1993). Furthermore, this stream progresses to consider the performance of two tasks in Figure1 jointly in help of the mixture-of-expert (ME) models (Jacobs, et al, 1991; Jordan & Jacobs, 1994; Xu & Jordan, 1993; Xu, Jordan & Hinton, 1994), which can learn either or both of the combining mechanism and individual experts in a maximum likelihood sense.

Two stream studies in the first period jointly set up a landscape of this emerging research area, together with a number of typical topics or directions. Thereafter, further studies have been further conducted on each of these typical directions. First, theoretical analyses have been made for deep insights and improved performances. For examples, convergence analysis on the EM algorithm for the mixture based learning are conducted in (Jordan & Xu, 1995; Xu & Jordan, 1996). In

8 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/combining-classifiers-learning-mixture-experts/56144

Related Content

Remote Conversation Support for People with Aphasia

Nilar Aye, Takuro Ito, Fumio Hattori, Kazuhiro Kuwabara and Kiyoshi Yasuda (2010). *International Journal of Software Science and Computational Intelligence* (pp. 72-85).

www.irma-international.org/article/remote-conversation-support-people-aphasia/39106

Feature Selection Algorithms for Classification and Clustering in Bioinformatics

Sujata Dash and Bichitrananda Patra (2014). *Global Trends in Intelligent Computing Research and Development* (pp. 111-130).

www.irma-international.org/chapter/feature-selection-algorithms-for-classification-and-clustering-in-bioinformatics/97055

Optimization of Dairy Feeding Models with C-SOMGA

Pratiksha Saxena, Dipti Singhand Neha Khanna (2016). *Problem Solving and Uncertainty Modeling through Optimization and Soft Computing Applications* (pp. 147-165).

www.irma-international.org/chapter/optimization-of-dairy-feeding-models-with-c-somga/147088

An Action Guided Constraint Satisfaction Technique for Planning Problem

Xiao Jiang, Pingyuan Cui, Rui Xu, Ai Gao and Shengying Zhu (2016). *International Journal of Software Science and Computational Intelligence* (pp. 39-53).

www.irma-international.org/article/an-action-guided-constraint-satisfaction-technique-for-planning-problem/172126

A Hybrid Approach to Missing Data: Bayesian Neural Networks, Principal Component Analysis and Genetic Algorithms

Tshilidzi Marwala (2009). *Computational Intelligence for Missing Data Imputation, Estimation, and Management: Knowledge Optimization Techniques* (pp. 45-70).

www.irma-international.org/chapter/hybrid-approach-missing-data/6795