Improving Bagging Performance through Multi-algorithm Ensembles

New Frontiers in Applied Data Mining
Kuo-Wei Hsu and Jaideep Srivastava

統碩一 112354008 洪芓宣

Contents

Introduction

Bagging, bootstrap aggregating :

- better ability to handle small data sets
- less sensitivity
- parallel structrue for efficient implementations
- **less accurate**

Introduction

In this paper

- The use of heterogeneous classification alogorithms in ensembles
- Develop a framework that supports the use of different classification algorithms in bagging

Diversity in Combinantions of Heterogeneous

$$T\text{-}Diversity = \mathbb{E}_S \left[\mathbb{I} \left(\hat{y}_i^{(A,S_1)} \neq \hat{y}_i^{(A,S_2)} \right) \right]$$

$$= \frac{1}{|S|} \sum_{i=1}^{|S|} \mathbb{I} \left(\hat{y}_i^{(A,S_1)} \neq \hat{y}_i^{(A,S_2)} \right)$$

(1)

Diversity in Combinantions of Heterogeneous

$$A - Diversity = \mathbb{E}_{S'} \left[\mathbb{I} \left(\hat{y}_i^{(A_1, S)} \neq \hat{y}_i^{(A_2, S)} \right) \right]$$

$$= \frac{1}{|S'|} \sum_{i=1}^{|S'|} \mathbb{I} \left(\hat{y}_i^{(A_1, S)} \neq \hat{y}_i^{(A_2, S)} \right).$$

(3)

Diversity in Combinantions of Heterogeneous

Diversity gain:

$$\Delta Div = \mathbb{E}\left[\mathbb{I}\left(\hat{y}_i^{(A_1,S_1)} \neq \hat{y}_i^{(A_2,S_2)}\right)\right] - \mathbb{E}\left[\mathbb{I}\left(\hat{y}_i^{(A_1,S_1)} \neq \hat{y}_i^{(A_1,S_2)}\right)\right] \quad (5)$$

Heterogeneous   Homogeneous

The underlying alogorithms
are different.

Relationship between Diversity and Correlation in Ensembles

Non-linear relationship between diversity (DIS) and Q statistic.

$$Q \leq \frac{(1 - DIS)^2 \cdot N^2 - 4 \cdot DIS \cdot N}{(1 - DIS)^2 \cdot N^2 + 4 \cdot DIS \cdot N} \tag{7}$$

Higher diversity would cause reduce correlation.

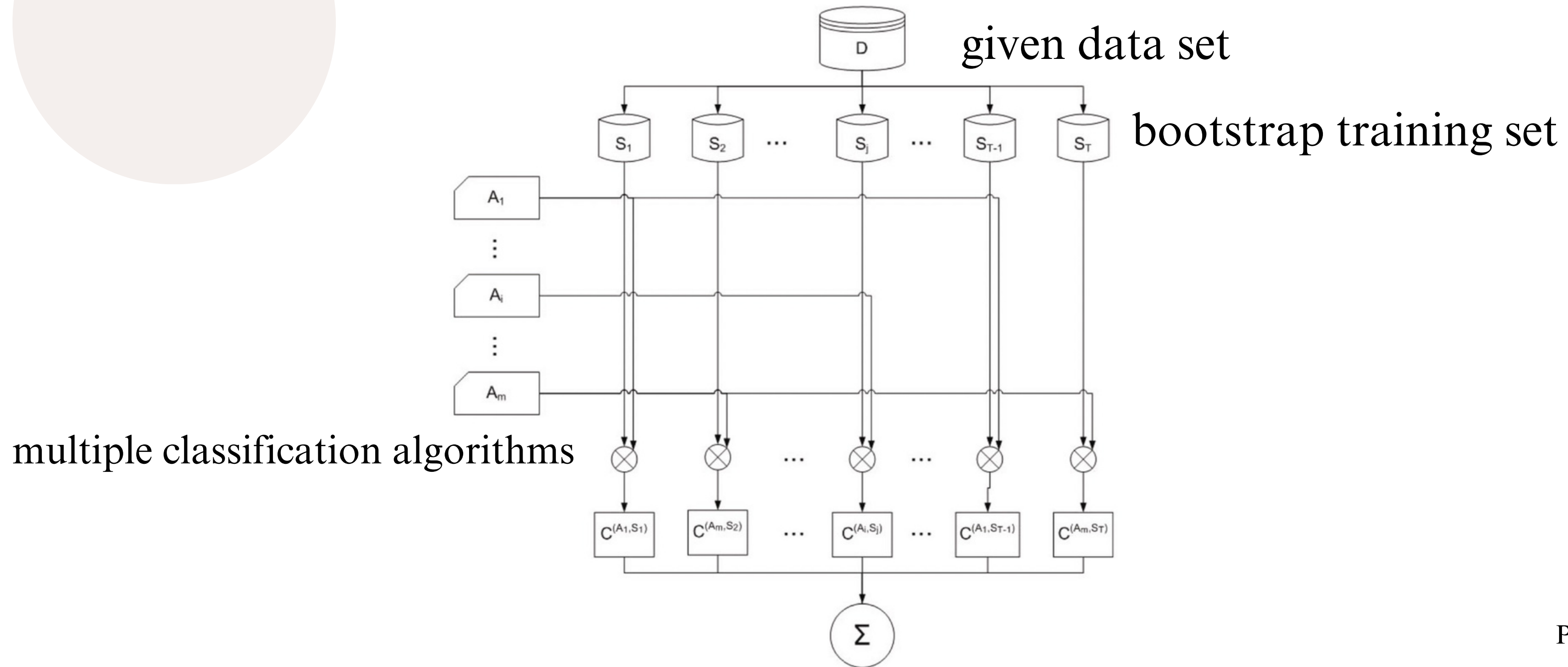$$x_c = \left(1 + \frac{2}{N}\right) - 2 \cdot \sqrt{\frac{1}{N} + \frac{1}{N^2}} \tag{8}$$

Bagging with Multi-algorithm Ensembles

- Goal:

In order to improve classificaiton performance of bagging, we **introduce another source of diversity into bagging by intoducing heterogeneity into bagging**.

Bagging with Multi-algorithm Ensembles

given data set

bootstrap training set

multiple classification algorithms

Experimental Results and Findings

- Impact of using heterogeneous algorithms on diversity
- Simulations for diversity and correlation in ensembles
- Comparison of bagging with multi-algorithm ensembles to other ensemble methods

Conclusions and Future Reserch Directions

- Study the employment of heterogeneous classification algorithms in other machine learning problems
- Efficient implementations of this approach
- Apply this approach to emerging application domains

Thanks for listening.