

# 収集事例の精製を目的とする 少数学習データからのアンサンブル学習の性能評価

Performance Evaluation on Ensemble Decision-Tree Learning from Small Training Data

小野 裕美\*<sup>1</sup>  
Hiromi Ono

岩沼 宏治\*<sup>2</sup>  
Koji Iwanuma

山本 泰生\*<sup>2</sup>  
Yoshitaka Yamamoto

\*<sup>1</sup>山梨大学大学院医学工学総合教育部コンピュータ・メディア工学専攻  
Department of Computer Science and Media Engineering, Interdisciplinary Graduate School of Medicine and Engineering,  
University of Yamanashi

\*<sup>2</sup>山梨大学大学院コンピュータ・メディア工学専攻  
Department of Computer Science and Media Engineering, University of Yamanashi

As a technique for building a domain-specific WEB search, there is a query-expansion technology called the keyword spice. The keyword spice is generated by learning a decision tree, but it costs for the preparation of good training examples for a decision tree learning. It's nice if it generate good training examples semi-automatically, so we discuss ensemble decision-tree learning of decision trees that for refining training examples from small training data. In particular, we evaluate the performance of bagging and boosting each of which is known as an ensemble decision-tree learning.

## 1. はじめに

現在、インターネット上には大量の情報が存在し、その中からユーザが望む情報を含むページを見つけ出すのは時間もかかり、知識と経験、スキルなどが必要となる。こうした問題解決の方法として、“専門検索エンジン”がある。専門検索エンジンは、ドメインを限定した検索エンジンであり、ユーザの検索要求に対して、ドメインに関連のない結果の出力を抑えることができる。専門検索エンジンの構築の技術の1つに、小山ら [1] の“検索隠し味”技術がある。これは、想定ドメインに関して決定木学習により事前に拡張語  $s$  (以後、検索隠し味と呼ぶ) を求めておく。ユーザが検索質問  $q$  を与えると、 $s$  を付加した質問  $q + s$  を汎用検索エンジンに渡す手法である。検索隠し味は、決定木学習を行い生成する。小山ら [1] は、事前に収集した大量のページを人手により正例、負例に分類し決定木学習用訓練事例を作成した。人手による分類には、膨大な時間や手間が必要となり、実用上の大きな問題点となる。

人手による分類の時間や時間の削減を目的とし、宮川ら [4] は、分類器の作成と適用について“Web カテゴリを利用した訓練集合の半自動生成”及び“類似度に基づいた訓練集合の半自動生成”を提案した。両手法において、精度は手動生成のものに近いが、収集データの精錬する決定木を少数学習データから構築するため分類の信頼性や安定性が問題点となっていた。少数学習データからでもより安定した分類精錬を行えるように、平林ら [2] は、アンサンブル学習で知られるバギング法を用いた手法を提案した。宮川ら [4] よりも精度が向上したが、バギング法についてのみ検証しており、頑健性や有効性について問題が残る。本稿では、バギング法とブースティング法の2つの代表的なアンサンブル学習の各種の性能評価を行い、有効性について検証を行う。

## 2. 先行研究

### 2.1 検索隠し味を用いた専門検索エンジンの構築

小山ら [1] により提案された“検索隠し味を用いた専門検索エンジン”はユーザーの検索質問  $q$  に対して、特定ドメインを良く分類する検索隠し味  $s$  を付加した  $q + s$  を汎用検索エンジンに渡すものであり、汎用検索エンジンを再利用したものである。汎用検索エンジンから返ってくる Web ページは  $q + s$  を含むので、すべてがドメインに関連しているページと仮定でき、フィルタリング処理を省くことが可能であり、そのため汎用検索エンジンと同様な早い反応性能を期待できる。

#### 2.1.1 検索隠し味の生成

以下に検索隠し味の抽出アルゴリズムを示す。

1. ユーザーが入力すると推定される 10 件のキーワードに対して 200 件、計 2000 件の Web ページを収集する。
2. 収集した Web ページを人手によって正例・負例に分類し、訓練集合・検証集合の 2 つに共通元を持たないように分割する。
3. 訓練集合に対して、決定木学習アルゴリズム (C4.5) を適用して決定木を作成し、根から正例を示す葉までのルートすべてをとり、選言標準形で表す。これが初期の検索隠し味  $h$  となる。
4. 上で作成した検索隠し味  $h$  は一般に非常に複雑になるので、適合率と再現率の調和平均が減少しないようにリテラルや連言を削除し単純化を行う。

#### 2.1.2 検索隠し味手動生成の問題点

小山ら [1] による検索隠し味の手動生成では、収集した 2000 件の Web ページを人手で正例・負例に分類し、検索隠し味の抽出を行っている。この分類作業は人手によるものであるため、膨大な時間と手間がかかってしまう。

正例または負例を人手で分類する作業のコスト削減を行うことを目的とし、宮川ら [4] は、少数データからの分類器の作

連絡先: 小野裕美, 山梨大学大学院医学工学総合教育部コンピュータ・メディア工学専攻, g11mk010@yamanashi.ac.jp

成と適用について研究しているが、分類器の安定性が問題点となっていた。平林ら [2] は、少数の学習データからも、安定で信頼性のある分類器の生成が期待できるアンサンブル学習を用いた手法を提案している。アンサンブル学習とは、個々の分類器では精度に限界があるため、複数の分類器を組み合わせることで予測性能を向上させる方法である。平林ら [2] はバギング法と呼ばれるアンサンブル学習法を用いて検証を行った。一定の精度向上が確認され、手動生成の精度に近づくことが示されている。しかし、用いたデータの多様性、他のドメインに関しての有効性に対して課題があった。本研究では、有効性の検証とともに、アンサンブル学習で知られているブースティング法を用いて検証実験を行い、より安定に精度向上を試みる。

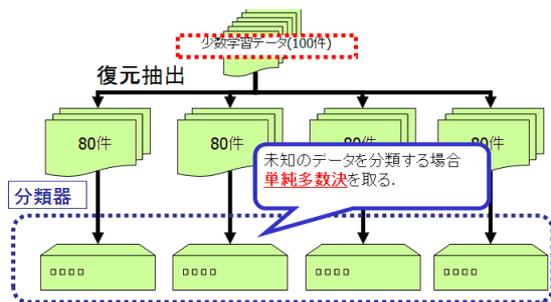


図 1: バギング法の概要

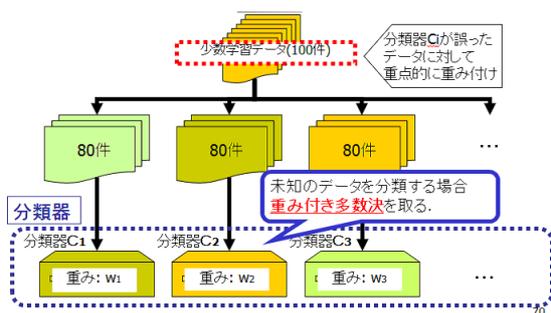


図 2: ブースティング法の概要

ブースティング法とは、逐次学習により互いに相補的な複数の分類器の生成を行い、未知のデータを分類するとき、分類器の重要性による重み付き多数決を取る方法である。一般に、バギング法は簡便で雑音に強く、ブースティング法は学習データの質によってはバギング法よりもブースティング法の方が良い結果を示すことが知られているため、ブースティング法とバギング法の比較評価を行う。収集した Web ページを正例・負例に分類する手順を以下に示す。

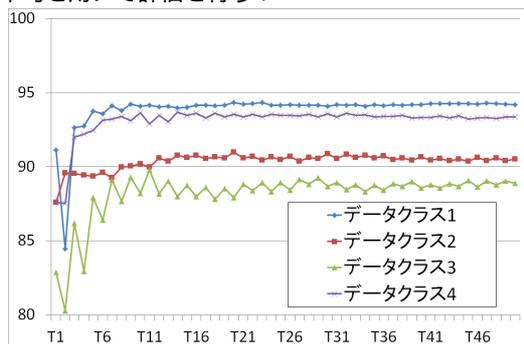
1. 多数の Web ページのデータを収集し、各キーワードから収集したページ毎にランダムで 10 件、計 100 件 Web ページの抽出を行う。それらを人手で正例・負例に分類し、少数学習データを作成する。
2. 少数学習データを用いて、アンサンブル学習を行い訓練事例精練器を作成する。
3. 訓練事例精練器を用いて収集したページの分類を行う。

### 3. 性能評価実験

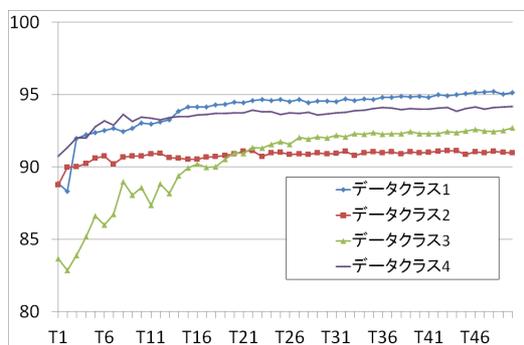
本章では、2 つのアンサンブル学習を用いて訓練事例の半自動生成の実験的性能評価を行い、バギング法とブースティング法の比較評価を行う。

#### 3.1 実験環境

実験環境について述べる。アンサンブル学習に用いた機械学習ツールは Weka[3] であり、利用する決定木学習アルゴリズムは C4.5 である。実験データは、宮川ら [4] が収集したデータの特定推定ドメインが“星の伝説・伝承”のデータと今回新たに特定推定ドメイン“日本地域特有の風習・習慣”についてのデータを収集し利用した。以下では、紙面の都合上、“星の伝説・伝承”に関する実験結果について述べるが、“日本地域特有の風習・習慣”でもほぼ同様の結果を得ている。“星の伝説・伝承”のデータは、10 個のキーワードに対して 200 ページを収集している。重複ページを除いた総収集ページ数は 1989 件であり、人手による分類結果正例数は 253 ページ、負例数は 1736 ページとなっている。本研究では、平林ら [2] で利用された 4 つのクラスの学習データを使用した。1 番目のクラスは正例 24 件、負例 76 件であり収集データに対して正例の割合が多いもの、2 番目は正例 7 件、負例 93 件であり正例の割合がかなり少ないもの、3 番目は正例 50 件、負例 50 件であり正例の割合がかなり多いもの、4 番目は、正例 13 件、負例 87 件であり収集データと同じ割合の正例の数を含むデータクラスである。各クラスに対して 5 個のデータを用意し実験を行い、平均を用いて評価を行う。



(a) バギング法を用いた場合



(b) ブースティング法を用いた場合

図 3: 実験データ 1

### 3.2 分類精度

訓練事例の分類器の生成にバギング法とブースティング法を用いた場合の比較を行う。精度評価には、全体正答率を用いる。全体正答率とは、収集データを正しく正例もしくは負例に分類した割合である。図3の横軸は決定木数、縦軸は全体正答率を示す。図3から、学習データクラス1,4が、両者の手法において他のデータクラスよりも最終的な精度が高い。データクラス2に関しては、どちらの手法においても決定木数を増加させて十分な学習を行っても、精度が向上しないことがわかる。一方、学習データクラス3では、バギング法では精度がすぐに頭打ちになるが、ブースティング法では決定木数50まで精度が向上していることがわかる。データクラス2は、正例の数が少ないため正例に対する精度が向上しないと考えられる。また、データクラス3では、正例の数が多いため決定木数が増加するにつれて精度が向上すると思われる。

### 3.3 収束速度

ここでは、収束速度は決定木数40から50の間の最小値、最大値の±1の範囲に精度が到達した決定木数とする。各学習データクラスに関する収束速度を表1に示す。学習データ1,3については、バギング法の方が速いことがわかる。学習データクラス1,3は収集データに対して正例の割合が多くなっているクラスであり、正例が多い学習データクラスに関してはバギング法の方が速いことがわかる。収束が遅いブースティング法を用いて分類器を生成する場合には、決定木の数を吟味する必要があると思われる。

表1: 収束する速度

	バギング法	ブースティング法
データクラス1	決定木数 5	決定木数 14
データクラス2	決定木数 2	決定木数 2
データクラス3	決定木数 5	決定木数 22
データクラス4	決定木数 4	決定木数 5

### 3.4 安定性

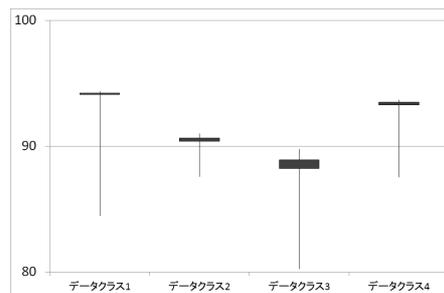
図4は、各学習データの全体正答率のデータ分布を示す。バギング法はブースティング法よりデータの値の変動が小さく安定的である。ノイズに強いという長所をもつバギング法の方が、どの決定木数においても値の変動が小さく、バギング法の方がこの意味では安定性があると思われる。

一方で、図3の各学習データの最終精度を見ると、バギング法の学習データによる差よりもブースティング法の学習データの方が狭いので、ブースティング法の方が学習データの変動に対して安定性があるといえる。ブースティング法の方が最終精度も高いことから、十分な学習を行う場合ブースティング法の方が安定していると考えられる。

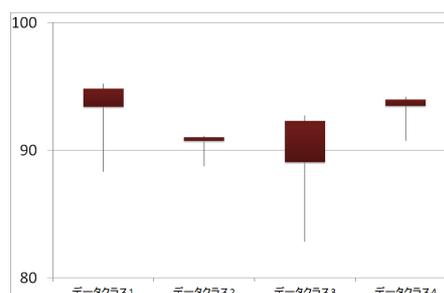
### 3.5 統計的な精度の差の検定

バギング法とブースティング法を用いた結果の精度の差について統計的な検定を有意水準0.05で母比率の差の検定を行った。その一覧を表2に示す。○はバギング法の方が統計的に良い、□はブースティング法の方が統計的に良いといえ、×は統計的に差があるとはいえない場合を示す。表2から、十分な学習を行えば両手法において統計的な精度の差がなくなる。一方、データクラス3では決定木数が増加すると精度に差があるという結果になる。学習データによっては手法の違いにより差が生じている。全体正答率ではなく、違う尺度であれば、また異なった結果となる可能性があり今後の課題となる。

以上、収束する速度、安定性、精度の差について考察を行った。バギング法は、収束する速度が速く、最終精度もデータクラス3を除いては、ブースティング法と遜色が無い。データ



(a) バギング法を用いた場合



(b) ブースティング法を用いた場合

図4: 実験結果のデータ分布

表2: 精度の差を統計的に検定

少数学習データ	決定木数					
	1	5	10	15	20	25
データクラス1	○	○	○	×	×	×
データクラス2	□	□	×	×	×	×
データクラス3	×	○	×	□	□	□
データクラス4	□	×	×	×	×	×

クラス3は、正例が極端に多い少数学習データであり、現実には、そのような少数データで精錬器を作ることは殆ど無いと考えられる。この意味で、バギング法は有力な学習法と考えられる。ブースティング法は、収束する速度が遅く、各学習データで十分な学習を行うまで値の変動が大きい、最終精度は高くどの学習データについてもバギング法よりも安定している。次に、収集データの精錬器の違いが検索隠し味の精度に影響するのか検証を行ったので次節で述べる。

## 4. 検索隠し味の抽出とその性能評価

生成した分類器を用いて、収集データを正例と負例に分類を行い、更にランダムに訓練事例と検証集合に分割した。訓練事例を利用して、検索隠し味の抽出を行った。具体的には、訓練事例から決定木アルゴリズム(C4.5)を適用し決定木を作成し検索隠し味を抽出した。更に検証集合を用いて、検索隠し味の単純化を行った。その結果について、次節にて説明を行う。

### 4.1 抽出された検索隠し味

決定木学習アルゴリズム(C4.5)を適用して検索隠し味を抽出した。その結果を表3と4に示す。今回用いた分類器は、統計的にバギング法を用いた方が良いという結果になった決定木数9のときのデータクラス1,3の分類器と統計的に差があ

表 3: バギング法で生成した分類器を用いた場合

	検索隠し味									
データクラス 1	-神	-はず	-一	-中	-こと	-星				
データクラス 2	ゼウス	(神話	-掲載	-記事	-今回	-山羊	-女性	-カス	-内	-父 座)
データクラス 3	-大神	-室	-個人							
データクラス 4	大神	(神話	-今年	-掲載	-検索	-使用	-クリック	-どこ	-万)	

表 4: ブースティング法で生成した分類器を用いた場合

	検索隠し味									
データクラス 1	女神	(神話	-カス	-カリ	-その他	-金)	(大神	-個人)		
データクラス 2	神話	-コメント	-仕事	-案内	-どこ	-問題	-赤	-気		
データクラス 3	-宇宙	-カメラ								
データクラス 4	大神	(神話	-今年	-掲載	-ローマ	-山羊	-研究	-仕事	-まま)	
		(-神話	-大神	-勇者	-オルフェウス	-ゼウス	-息子	じゅう	その後	-うし)

表 5: 検索隠し味を付加し検索した結果の適合率

	オリオン (キーワードのみ:0.04)		カシオペア (キーワードのみ:0.04)	
	バギング	ブースティング法	バギング	ブースティング法
データクラス 1	0.24	0.22	0.08	0.06
データクラス 2	0.72	0.56	0.84	0.66
データクラス 3	0.04	0.06	0.26	0.36
データクラス 4	0.64	0.56	0.80	0.60

るとはいえないという結果になった決定木数 9 のときのデータクラス 2,4 の分類器を使用した。抽出された検索隠し味をみると、ブースティング法よりもバギング法の方が単純化された検索隠し味となっている。4 種類の学習データクラスに対して検証を行い、ブースティング法では不適切な名詞が含まれており、バギング法の方が有用な単語が抽出されていることがわかった。次に、統計的にブースティング法が良いといえるデータクラス 2 の決定木数 7, 9 のときの分類器を用いた場合についても検証を行った。具体的な実験結果については、紙面の都合上省くが、抽出された検索隠し味は、ブースティング法の方が単純化され有用な単語が抽出されていた。精錬器での統計的な精度の差は、抽出した検索隠し味にも影響するといえる。得られた検索隠し味の性能について評価を行うために、検索隠し味を用いた特定推定ドメインのページ検索に対する適合率について検証を行う。

#### 4.2 適合率

表 5 に表 3 と 4 の検索隠し味を用いて性能評価を行った結果を示す。表 5 は、適合率をまとめた表であり、キーワード  $q$  に対してそれぞれの検索隠し味  $s$  を付加した  $q \cup s$  を Google に入力したときの検索結果上位 50 件に対するものである。キーワード  $q$  のみで検索を行った結果をキーワード  $q$  の右記に明記する。

表 5 をみると、データクラス 3 以外に関しては、バギング法の方が良い結果となっている。ブースティング法よりもバギング法の方が安定して良い検索隠し味を抽出すると考えられる。一方、ブースティング法の方が統計的に良いとなった分類器を用いて抽出した検索隠し味の結果は紙面の都合上省くが、適合率においてもブースティング法の方が良い結果となった。分類器での精度の差は、検索隠し味の抽出にも影響するといえる。

## 5. まとめと今後の課題

今回、アンサンブル学習で知られるバギング法とブースティング法を用いて実験を行った。訓練事例の精製に使用する分類器においては、統計的に性能の差がある場合、実際にその分類器を用いて訓練事例の精製を行い、検索隠し味を抽出し評価実験を行った結果、分類器の精度の差が影響することがわかった。今後の課題として、アンサンブル学習で知られるランダムフォレスト法を適用し検証する。

### 謝辞

本研究は一部、文科省科学研究費補助金 (基盤 C: No.22500127) の援助を受けている。

### 参考文献

- [1] Satoshi Oyama, Takashi Kokubo, and Toru Ishida: Domain Specific Web Search with Keyword Spice. IEEE Transaction on Knowledge and Data Engineering, vol.16, No.1, pp.17-27, 2004.
- [2] 平林 広喜, 岩沼 宏治, 山本 泰生, 鍋島 英知: WEB 検索高度化のためのアンサンブル学習に基づく訓練事例の精練. 電子情報通信学会技術研究報告, 人工知能と知識処理研究会 IEICE-AI2010-5, Vol.110, No.105, pp.25-30, 2010.
- [3] W. H. Ian and F. Eibe: Data Mining: Practical Machine Learning Tools and Techniques. (Elsevier, second edition, 2005)
- [4] 宮川 礼子, 岩沼 宏治, 鍋島 英知: 専門検索エンジンの半自動生成を目的とした類似度に基づく WEB データの精練. 信学技報, vol. 105, no. 640, AI2005-52, pp. 31-36, 2006 年 3 月.