Weighted Averaged One-Dependence Estimators の提案と実験的評価

黒川 茂莉 †，横山 浩之 †，櫻井 彰人 ††

† (株) KDDI 研究所　†† 慶應義塾大学理工学部

Naive Bayes (NB) は，クラス変数を条件とした変数間の条件付き独立性という仮定を置く性質上，データに含まれる変数間の依存関係が適切に表現できず分類精度が十分でない場合がある．この問題を改善するために，NB の条件付き独立性の仮定を緩和する Augmented Naive Bayes (ANB) として，比較的小さい計算コストで高い分類精度を示す Averaged One-Dependence Estimators (AODE) が提案されている．AODE は NB の変数間にリンクを追加した複数のベイジアン分類器をモデル平均するものであるが，その平均方式に改善の余地がある．本論文では，AODE におけるモデル平均方式を改善するため，複数の重み付け方式 (Weighted AODE) を提案し，分類精度の観点で従来方式と比較した結果を示す．

Experimental Evaluation of Weighted Averaged One-Dependence Estimators

Mori KUROKAWA †　Hiroyuki YOKOYAMA †　Akito SAKURAI ††

† KDDI R&D Laboratories Inc.

†† Faculty of Science and Technology, Keio University

Naive Bayes (NB) is a simple Bayesian classifier, which is nevertheless widely accepted as an efficient classifying technique in many real-world applications. However, in some cases NB underperforms due to its conditional independence assumption. Augmented Naive Bayes (ANB) models are extensions of NB by relaxing the assumption and Averaged One-Dependence Estimators (AODE) is known as an excellent ANB model with high classification accuracy. AODE utilizes a simple model averaging method of ODEs and therefore there is room for improvement in it. In this paper, we propose some weighted averaging methods of ODEs and show experimental results on classification accuracy.

1　はじめに

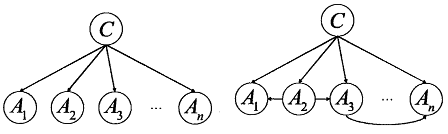
現代の情報化社会においては，大量のデータを自動的に収集・分析し，それらの特徴にもとづいて分類を行うことが，より価値の高いサービスを提供する上で重要になっている．本論文では，ベイジアン分類器の中で単純な分類器である Naive Bayes (NB) の改善を目指した Augmented Naive Bayes (ANB) に焦点をあて，ANB の中で特に高い分類精度を示す Averaged One-Dependence Estimators (AODE) の改良方式を提案し，従来方式と分類精度の観点で比較評価する．

次節では AODE を含むベイジアン分類器に関する関連研究を，3 では AODE を改善する提案方式を，4 では実世界のデータを用いた分類精度の観点での比較実験およびその結果を，最後に 5 ではまとめを示す．

2　従来方式

分類問題とは，対象の属性 $\boldsymbol{A} = (A_1, A_2, \cdots, A_n)$ の値 $\boldsymbol{a} \in \mathcal{R}^n$ をもとに対象をあらかじめ定められたクラス変数 $C = \{c_1, c_2, \cdots, c_m\}$ のいずれかに割り当てる写像を決定する問題を言う．分類器を構築するにあたり利用する事例集合を訓練データと呼び，$D = \{d^i : i = 1, 2, \cdots, N\}$ と表す．個々の訓練事例 d^i はクラス変数の値と属性値ベクトルの組み合わせ $\{c^i, \boldsymbol{a}^i\}$ で構成される．ベイジアン分類器は，Bayesian Networks (BN) モデルと呼ばれる確率モデルを分類問題に適用したものであり，単純なものから複雑なものに至るまで数多く提案されている．

Naive Bayes (NB) は，図 1(a) の通りクラス変数を条件とした属性間の条件付き独立性を仮定した単純なベイズ分類器である．Augmented Naive



(a) NB　　　(b) TAN

図 1　NB および TAN のグラフ構造

Bayes (ANB) は，NB の条件付き独立性の仮定を緩和したベイジアン分類器の総称であり，条件付き独立性を緩和することにより分類精度が向上することが知られている．Sahami により導入された $\mathcal{N}$-dependence[10] の定義によれば，各属性がクラス変数に加え最大 $\mathcal{N}$ 個の親属性を持つベイジアン分類器は $\mathcal{N}$-dependence となり，計算量の問題から One-Dependence (1-dependence) が妥当であるとされている [11]．One-Dependence ベイジアン分類器の代表的なものとして，Tree Augmented Naive Bayes (TAN)[6] がある．TAN のグラフ構造は，図 1(b) の通り NB のグラフ構造を除いた属性のリンクが木となるものである．TAN は，そのグラフ構造を決めるために構造を探索する必要があり，構造探索の方式として Chow & Liu アルゴリズム [3] と呼ばれる Greedy Search を適用したものが有名である．Averaged One-Dependence Estimators (AODE)[11] は，図 2 の One-Dependence Estimators (ODE) と呼ばれる複数の TAN をモデル平均した分類器であり，高い分類精度を示す．各 ODE は，各属性と対応付けられ，属性 A_j に対応する ODE m_j は，クラス変数 C とある属性 A_j (Super Parent と呼ぶ) をそれ以外のすべて属性の親属性とするような TAN である．

3　提案方式

我々は，AODE の改良として，ODE の単純なモデル平均ではなく，ODE m_j の重み w_j を適切に計算し，ODE の重み付け平均する分類器 Weighted Averaged One-Dependence Estimators (Weighted AODE) を提案する．提案方式は，必ずしも真のモデルではない複数の ODE を用いて，真のモデルに近い分布を再現しようとする方式であり，重み付けの方式が重要である．提案方式と対比すると，AODE は重みを一様な重みとしており，一様な重みがよい場合は確率分布の空間において真のモデルが全 ODE の重心に一番近い特殊な場合に限られる．また，クラス変数 C と Super Parent A_j の間の相互情報量により重み付けモデル平均を行う Weightily AODE[7] も提案されているが，相互情報量はクラス変数に対する説明変数としての属性の順序づけには適している [5] が，複数の分類器の重み付けへの拡張は自然ではない．

表 1　重み付け方式の提案

weighting method	w_j: weight value for ODE m_j
重み付け方式 (cBMA)	$\alpha \exp\left\{\frac{1}{N}\sum_{i=1}^{N}\log P_{\bar{\theta}_j}(d^i)\right\}$
重み付け方式 (EM)	更新式 ($\bar{w}_j \to w_j$): $w_j = \frac{1}{N}\sum_{i=1}^{N} P_{\bar{\theta}_j}(j \mid d^i)$ $P_{\bar{\theta}_j}(j \mid d^i) = \frac{\bar{w}_j P_{\bar{\theta}_j}(d^i)}{\sum_j \bar{w}_j P_{\bar{\theta}_j}(d^i)}$

そこで，我々は表 1 で示す 2 種類の重み付け方式を提案する．ここで,「重み付け方式 (cBMA)」は Bayesian Model Averaging (BMA) の改良方式であり,「重み付け方式 (EM)」は潜在変数を含むモデルの最尤推定を行う Expectation-Maximization (EM) アルゴリズムを適用したものである．

BMA[8] は，ベイズ定理に従い混合する分類器の事後確率に基づき重み付けモデル平均を行うが，混合する分類器の集合を限定した場合においてうまく機能しないことが指摘されている．例えば Minka は，同程度の精度をもつ複数の分類器がありそれらの単純な投票方式 (uniform vote) によるモデル混合で正確に分類できる場合にも，訓練事例数が増えるに従って分類精度が悪くなる場合があることを指摘している [9]．そこで，我々は，BMA の重みの訓練事例数への依存性をなくすため，分類器 m_j の事後確率 $P(j|D)$ を $1/N$ 乗することにより，データ 1 個についての確率値に変換する[1]．すなわち，次式で ODE m_j の重み w_j を定める．

$$w_j = \alpha P(j|D)^{1/N} = \alpha \exp\left\{\frac{1}{N}\log P(j|D)\right\} \quad (1)$$

ここで，α は正規化のための項である．分類器の事後確率 $P(j|D)$ は，事例の独立性と事前確率 $P(j)$ が一様分布であることを仮定して $P(j|D) \propto \prod_{i=1}^{N} P_j(d^i)$ とする．さらに，事例毎の $P_j(d^i)$ を ODE m_j のパラメータの推定値 $\bar{\theta}_j$ により計算す

[1] このような変換を行った BMA を，本論文では corrected BMA (cBMA) と呼ぶ．

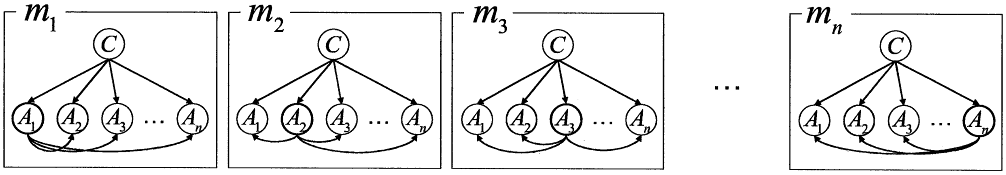


図 2　ODE のグラフ構造

る．すなわち，次式で計算する．

$$P_j(d^i) \sim P_{\bar{\theta}_j}(d^i) = \bar{P}(c^i, a^i_j) \prod_{k \neq j} \bar{P}(a^i_k | c^i, a^i_j) \quad (2)$$

EM アルゴリズム[4] は，混合する分類器 m_j に属する事例 d^i の確率の期待値を求める E ステップと重みの再推定を行う M ステップを交互に繰り返すことにより，尤度極大化を行う．初期値と反復回数により，得られる重みの推定結果が異なる．

以上の重みの推定の前に，各 ODE のパラメータ $\bar{P}(c, a_j)$ および $\bar{P}(a_k|c, a_j)$ の推定を行うが，ここでは最大事後確率 (MAP; Maximum a Posteriori) 法による推定を用いる．MAP 法による推定方式として一般的に知られている方式に，Laplace correction と M-estimation がある．Laplace correction は相対頻度の分子に 1.0 を加え分母に属性の値の数を加えるスムージングを行ったものであり，M-estimation[2] は相対頻度の分子に $M\cdot$(一様分布を仮定した場合の確率値) を加え分母に M を加えるスムージングを行ったものである．

4　提案方式の検証実験

4.1　実験方法

我々は，提案方式と従来方式を複数の実世界のデータを用いて比較検証した．提案方式としては，表 1 の 2 種類の重み付け方式を検証する．提案方式のパラメータ推定方式は，M の値を変えながら比較実験した結果，M-estimation (M=0.125) とした．また,「重み付け方式 (EM)」では，初期値への依存性をなくすため，初期値を一様乱数により 10 通り生成し，それらによる推定値を平均する方式を用いた．それぞれの EM アルゴリズムによる推定では，反復回数を変えながら比較実験した結果，反復回数を 3 回とした．比較対象の従来方式は，NB，TAN，LBR[13]，HNB[12]，AODE (Laplace correction, M-estimation (M=1.0))，Weightily AODE (M-estimation (M=1.0)) の 6 種類とした．TAN の構造探索では，モデル選択規準として Cooper & Herskovitz の K2 規準を用いた．

実験には，University of California at Irvine (UCI) の Machine Learning Repository[1] の 36 種類の分類用データを用いた．これらのデータは Jiang らの実験で用いられているものと同じであり，全データに対し Jiang らの実験と同様に，欠損値除去，離散化，値の分散が小さすぎる属性と値の分散が大きすぎる属性の除去を行った．[7, 12]

4.2　実験結果

表 2 は，分類精度の比較表を表す．比較表の各エントリーの値 $W/D/L$ は，36 種類のデータについて対応する行の方式の対応する列の方式に対する $Win/Draw/Loss$ を判定した場合の Win の数/$Draw$ の数/$Loss$ の数を表す．手法間の $Win/Draw/Loss$ の判定は両側 t 検定により行い，平均値の差に統計的な有意差がない場合を $Draw$ とし，有意差がある場合は平均値の差の大小判定にもとづいて $Win/Loss$ を判定する．

表 2 によれば，Weighted AODE の「重み付け方式 (cBMA)」は，AODE に対して $W/D/L$ = 9/27/0 であり，Weightily AODE に対して $W/D/L$ = 2/33/1 であり，その他の従来方式にも優った．「重み付け方式 (EM)」も同様に，AODE, Weightily AODE, その他の従来方式に優ったが，初期値，反復回数の指定の仕方には課題が残る．

5　おわりに

本論文では，AODE のモデル平均において重み付けを行う方式 Weighted AODE を提案した．Weighted AODE を従来方式である AODE および Weightily AODE と分類精度において比較し，AODE を上回る分類精度を得，Weightily AODE に対して同等以上の分類精度を得た.「重み付け方式 (cBMA)」の場合の Weighted AODE の学習の時

表 2 $W/D/L$ 表: NB，従来の ANB 方式，および Weighted AODE の分類精度 (mean accuracy) の比較

		NB	TAN	LBR	HNB	AODE		WlyAODE	WedAODE
						LC	ME1.0		cBMA
TAN		13/21/2							
LBR		12/24/0	3/31/2						
HNB		16/18/2	9/25/2	8/25/3					
AODE	LC	14/21/1	5/27/4	3/30/3	1/27/8				
AODE	ME1.0	15/20/1	8/27/1	11/23/2	2/30/4	7/29/0			
WlyAODE	ME1.0	17/18/1	10/25/1	10/24/2	3/32/1	9/27/0	3/32/1		
WedAODE	cBMA	16/19/1	9/26/1	9/25/2	5/30/1	9/27/0	4/32/0	2/33/1	
WedAODE	EM	16/20/0	9/27/0	10/24/2	5/31/0	10/26/0	5/31/0	2/33/1	2/34/0

ここでは，Weightily AODE，Weighted AODE をそれぞれ WlyAODE，WedAODE と略記する．また，パラメータ推定方式を Laplace correction を LC，M-estimation ($M = 1.0$) を ME1.0 とする．

間計算量のオーダーは AODE の学習の時間計算量のオーダーと変わらないが，「重み付け方式 (EM)」の時間計算量は反復回数が多くなるほど追加の時間計算量が必要となる．また，Weighted AODE の場合，パラメータ推定の方式としては M-estimation を用い，M の値は $M < 1$ の値，とくに $M = 0.125$ とする場合がよい結果を示すことが分かった．

今後は，各重み付け方式の適切な設定条件を詳細に分析するとともに，提案方式が適するデータの傾向をつかむためにデータを追加して実験を行う予定である．

参考文献

1) A. Asuncion and D.J. Newman, "UCI Machine Learning Repository," Irvine, CA: University of California, School of Information and Computer Science, (2007), http://www.ics.uci.edu/~mlearn/MLRepository.html.
2) B. Cestnik and I. Bratko, "On Estimating Probabilities in Tree Pruning," Proc. European Working Session on Learning (EWSL) 1991, pp.138-150, (1991).
3) C.K. Chow and C.N. Liu, "Approximating Discrete Probability Distributions with Dependence Trees," IEEE Transactions on Information Theory, Vol.14, pp.462-467, (1968).
4) A.P. Dempster, N.M. Laird, and D.B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," Journal of the Royal Statistical Society, Series B, Vol.39, No.1, pp.1-38, (1977).
5) I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," Journal of Machine Learning Research, Vol.3, pp.1157-1182, (2003).
6) L. Jiang, H. Zhang, Z. Cai, and J. Su, "Learning Tree Augmented Naive Bayes for Ranking," Proc. Database Systems for Advanced Applications (DASFAA) 2005, pp.688-698, (2005).
7) L. Jiang and H. Zhang, "Weightily Averaged One-Dependence Estimators," Proc. Pacific Rim International Conferences on Artificial Intelligence (PRICAI) 2006, pp.970-974, (2006).
8) D. Madigan, A.E. Raftery, C. Volinsky, and J. Hoeting, "Bayesian Model Averaging," AAAI Workshop on Integrating Multiple Learned Models, pp.77-83, (1996).
9) T.P. Minka, "Bayesian model averaging is not model combination," MIT Media Lab note, (2000), http://research.microsoft.com/~minka/papers/bma.html.
10) M. Sahami, "Learning Limited Dependence Bayesian Classifiers," Proc. Knowledge Discovery and Data Mining (KDD) 1996, pp.335-338, (1996).
11) G.I. Webb, J.R. Boughton, and Z. Wang, "Not So Naive Bayes: Aggregating One-Dependence Estimators," Machine Learning, Vol.58, No.1, pp.5-24, (2005).
12) H. Zhang, L. Jiang, and J. Su, "Hidden Naive Bayes," Proc. Association for the Advancement of Artificial Intelligence (AAAI) 2005, pp.919-924, (2005).
13) Z. Zheng, G.I. Webb, and K.M. Ting, "Lazy Bayesian Rules: A Lazy Semi-Naive Bayesian Learning Technique Competitive to Boosting Decision Trees," Proc. International Conference on Machine Learning (ICML) 1999, pp.493-502, (1999).