人間の認知バイアスの実装による機械学習モデルの性能向上

防衛大学校理工学研究科後期課程

電子情報工学系専攻 情報知能メディア学教育研究分野

谷口 英貴

平成３１年３月

目次

第 1 章　序論

本論文では，機械学習モデルに認知バイアスを適用し，より優れた学習能力を持つ手法を提案する．機械学習は情報学の分野において今日特に注目される研究分野のひとつであり，様々な用途に利用されている．一例として，インターネット利用者が日々利用するサービスであるスパムメールフィルタ [Androutsopoulos et al. 2000] や，画像分類器 [Sermanet et al. 2014] がある．また，機械学習の分野で長年研究対象となっているタスクとして，病気の分類タスク [Marcano-Cedeno et al. 2011, Han et al. 2017] や画像生成タスク [Goodfellow et al. 2014] などがある．これ等のタスクにおいて，機械学習は目覚ましい成果を挙げてきた [LeCun et al. 2015, Goodfellow et al. 2016]．近年のブレークスルーとしては，ニューラルネットワークが画像分類において驚くべき成績を見せたことや [LeCun et al. 2015]，多層ニューラルネットワークによって構成された囲碁 AI がプロの棋士に勝利したこと [Silver et al. 2016]，テレビゲームでプロのプレイヤーに勝るゲーム AI が開発されたこと [Firoiu et al. 2017] などが挙げられる．機械学習がこれらのタスクにおいて優れた性能を発揮できる理由は，近年の情報技術の発展により，膨大なデータを取得できるようになったことと，そのデータをより高速に処理できるようになったことが挙げられる [Bishop 2006]．また，機械学習は，スパムメールフィルタや自動翻訳，画像検索といった，インターネット上で日々利用されるサービスの根幹となっている．

機械学習を用いたサービスでは，多数のユーザーに対する補助を行うのと同時に，ユーザーの行動から得られた情報を記録する [Mitchell 1999]．このようにして機械学習は，ユーザーのフィードバックから得られた膨大なデータを利用し，より適切に特徴の重み付けを行い，統計的に有意な結果を得ることができるようになった．一例として，スパムメールフィルタのケースを考える．例えばスパムメールと分類されたメールが，実際には非スパムメールであった場合，ユーザーはこのメールを非スパムメールであるとタグ付けしなおすことができる．また，逆に非スパムメールと分類されたメールが実際にはスパムメールであった場合，ユーザーはこのメールをスパムメールであるとタグ付けしなおすこともできる．

これらの処理を通じて，機械学習は新しいスパムメールのパターンを学習し，分類能力を向上することができる．

このユーザーの行動の記録と学習の繰り返しが，機械学習を発展させ，大きな需要を生み出してきた [Ma et al. 2009]．これ等のタスクに用いられ，中でも特に注目される手法に教師あり学習がある．教師あり学習とは，ラベル付けされたデータを学習し，この情報をもとに予測を行うものである．この時，モデルの学習に用いるデータは教師データと呼ばれる．教師データの形式は多様であり，タスクに応じて異なる形式が用いられる．一例として，スパムメール分類や病気の良性･悪性分類などに代表される，二値分類タスクについて説明する．二値分類タスクの場合，教師データには正例・負例のラベル付けが事前になされている必要がある．教師データから得られた情報は特徴ベクトルと呼ばれる．特徴ベクトルは機械学習モデルを学習するためのデータ集合であり，各特徴の出現頻度や，出現回数などの情報が含まれる [Alpaydin 2014]．そして，教師あり学習モデルは特徴ベクトルから正例におけるデータの傾向，および負例におけるデータの傾向を学習し，新規のデータが正例であるか，あるいは負例であるかの予測を行う．

1.1 研究の目的

一般的に，機械学習手法は学習時において，データに偏りの無い多量の教師データが存在するという前提のもと，優れた性能を示す [Mitchell 1997]．しかしながら，これらのモデルは教師データが少量である場合や，不均衡である場合に性能が大幅に低下する [Mitchell 1997]．こうした状況から，大規模かつバランスの良いデータではなく，むしろ小規模なデータや，不均衡なデータからでも優れた学習を行う手法の開発が望まれている [Lin et al. 2014]．その実現のため，人間の認知能力の機械学習への応用が近年研究されている [Hattori & Oaksford 2007]．

人間は少量かつスパースな情報から素早く新たな概念を学習できることが知られている [Gerken et al. 2015, Lake et al. 2017]．例えば，人間の幼児が初めて動物園でカバを見た時，一頭のカバから，その大きさや形，他の動物との違いなど，多くの情報を瞬時に学習する [Lake et al. 2015a, Lake et al. 2015b]．機械学習で同様の

学習を行う場合，数百から数千の教師データを必要とする可能性がある [Tenenbaum 1999]．また，人間は正例を学習するために，負例を学習する必要はない [Tenenbaum 1999]．例えば，幼児はカバという概念を学習するために，ゾウなどの他の動物を学習する必要はない．すなわち，人間は単一のクラスに属する少量のデータから新たな概念を獲得可能な一方で，機械学習は複数のクラスとそれに属する多量のデータを必要とする．

上記のような人間の優れた学習能力には，認知バイアス [Kahneman 2002, Tversky & Kahneman 1973, Tversky & Kahneman 1974, Feldman 2000, Goodman et al. 2008] が貢献しているとされる [Hattori & Oaksford 2007, 篠原ら 2007]．認知バイアスとは，必ずしも論理的ではないものの，素早い状況判断に貢献する，人間特有の認知の偏りである．例えば，コイントスを数回行い，その結果が「表・表・表・表・表・裏」だったとする．このような結果が得られた場合，人間はたびたびコインに細工がされているのではないか，などと推論を始める [Tversky & Kahneman 1973]．また，もしもコイントスの結果が「表・表・表・裏・裏・裏」となった場合，その結果がランダムであると感じず，ゲームそのものが不公平であると感じる．つまり，この例において，人間はわずかな試行回数のコイントスから状況判断を行う．人間はこの特有の認知能力により，わずかなサンプルからの優れた学習を実現しているとする指摘があり [Hattori & Oaksford 2007]，また，このような人間の認知能力を機械学習に応用する研究が盛んに行われている [Lake et al. 2011, Salakhutdinov et al. 2012, Lin et al. 2014].

本研究の目的は，少量データ下における機械学習の問題点を克服すべく，人間の認知バイアスを導入し，より優れた学習能力を持つ機械学習モデルを提案することである．具体的には，因果推論モデルを機械学習手法に適用し，認知バイアスによる特徴の重み付けを行う．

本研究では，スパムメール分類と病気の良性・悪性分類をテーマに，モデルの性能比較を行った．これ等のタスクは長年，機械学習の研究の題材とされてきたものであり，その性能向上が望まれている．これ等のタスクにおける代表的な手法である，後述するナイーブベイズおよびニューラルネットワークに認知バイアス

を適用した．ナイーブベイズとニューラルネットワークは，分類タスクにおいて頻繁に用いられる．ナイーブベイズはベイズの定理を基にした統計学的手法な手法であり，ニューラルネットワークは動物の脳の仕組みを基にした，必ずしも統計学的ではない手法である．本研究ではこれらの２つのモデルに認知バイアスを適用し，分類タスクにおける提案手法の妥当性を示す．

提案手法に用いた Loosely Symmetric モデル [篠原ら 2007] は条件付き確率を扱うあらゆるモデルに利用可能であり [Takahashi et al. 2011]，このことから，本研究では機械学習における分類タスク・生成タスク，統計学的手法・非統計的手法の両者に，認知バイアスを適用した．提案手法は特定のタスクにおいて有用とするものではなく，機械学習への認知科学手法の適用，並びに両研究分野の橋渡しを行うことを目的とする．次節では，機械学習への応用に用いられる因果推論，ならびに認知科学の既存研究について概観する．

1.2 因果推論

1.2.1 三段論法と演繹

因果推論の歴史は古く，初期の考えにアリストテレスの三段論法と，エウクレイデス（ユークリッド）の「原論 (Στοιχεία) 」に纏められた公理を前提とした演繹がある．アリストテレスの三段論法は，例えば「 (1) 全てのサルが霊長類である．(2) 全ての霊長類が哺乳類である．(3) このため,全てのサルが哺乳類であると帰結する.」という推論で，真とみなせるある言明と，そこから必然的に真であると帰結する言明とから構成される．アリストテレスの三段論法は疑いなく真であると認められる言明から始まらなくてはならず，このことから，あらゆる事柄に対してこの手法から帰結を得ることは不可能である．また，エウクレイデスは命題 p, q, r が存在する時，「 (1) p ならば q, (2) q ならば r, (3) ゆえに p ならば r」という仮言的三段論法を繋げて帰結の連鎖を行い，公理に基づく定理の推論を行った．つまり一見関係のなさそうな事象同士を，帰納と演繹との組み合わせから結び付けていくことが論理的推論の目的である [Katz 1993].

1.2.2 Pierce の仮説形成

先述のように, アリストテレスの三段論法は「疑いなく真である言明」から始まらなくてはならず, またそこから「必然的に真であると帰結する言明」とから構成される. つまりアリストテレスの三段論法は曖昧さを許容せず, 各事象がまぎれもなく真と見なせなくてはならないのである. 一方, Pierce は「仮説形成 (abduction)」と呼ばれる, 演繹 (deduction) とも帰納 (induction) とも異なる推論形式を提唱した [篠原 & 中野 2007]. Pierce の仮説形成は以下のように, アリストテレスの三段論法とは異なる形の三段論法として形式化できる.

前提 1	雨が降ったら,地面がぬれる	$(A \to B)$
前提 2	地面がぬれている	(B)
結論	∴雨が降ったのだろう	(A)

パースの仮説形成の大きな特徴は, アリストテレスの三段論法とは異なり, 憶測や曖昧さを許容するという点にある. 上記の例においてこの三段論法は「地面がぬれている」という事象に対し「雨が降ったら, 地面がぬれる」という根拠から「雨が降ったのだ」という結論を導き出している. 当然ながら地面がぬれる理由は雨だけに限らず「バケツの水をこぼした」や「スプリンクラーが散水した」など様々な理由が考えられ, 地面がぬれていることを理由に雨が降ったと推測するのは誤りである恐れがある. しかしながら, こうした曖昧さを含んだ推論は人間にたびたび見受けられ [篠原ら 2007, Sidman et al. 1982], それを引き起こす要因のひとつに認知バイアスがある.

1.2.3 認知バイアス

認知バイアスとは，必ずしも論理的ではない，人間特有の認知的傾向であり [Kahneman 2002, Tversky & Kahneman 1973, Tversky & Kahneman 1974, Feldman 2000, Goodman et al. 2008]，概念学習に有利に働くとされている [篠原ら 2007, Takahashi et al. 2010]．人間の認知に含まれるとされるバイアスの中で，最も注目されている物の例に，対称性バイアス [篠原ら 2007, Sidman et al. 1982] と相互排他性バイアス [Markman & Wachtel 1988, Merriman et al. 1989] がある．両バイアスは近年どの程度人間の認知に寄与しているか，実験をもとに定量化されており [郡司 2008]，対称性バイアスと相互排他性バイアスが素早い学習に貢献しているという主張がある [Hattori & Oaksford 2007]．次項以降に，対称性バイアスと相互排他性バイアスについて概観する．

1.2.3.1 対称性バイアス

対称性バイアスとは，原因 p から結果 q が発生することを繰り返し観測した後に，結果 q を目撃した際，その原因が p であると考える傾向である [篠原ら 2007, Sidman et al. 1982]．例えば原因 p が「雨が降った」，結果 q が「地面がぬれた」とした場合，「雨が降ったら，地面がぬれた」という事象を複数回観測した後に「地面がぬれたのは，雨が降ったからだ」と考える現象である．対称性バイアスの例示を Fig. 1 に示す．「地面がぬれた」理由には「雨が降った」以外にも，「スピリンクラーが散水した」など他の原因も考えられるため，このような判断を直ちに下すのは論理的に誤りを含む恐れがある．しかし，このような推論は私たちの生活において度々見受けられ，日常生活において十分実用的な働きをすると考えられている[篠原 & 中野 2007]．

1.2.3.2 相互排他性バイアス

相互排他性バイアスは「事象 A が満たされれば B が起こる」という事実から

「事象 A が満たされなければ B は起こらない」と考える傾向である [Markman & Wachtel 1988, Merriman et al. 1989]. 例えば, 母親が子供に「部屋を片付けないとおもちゃを買ってあげません」と言ったとする. 事象 A が「部屋を片付ける」, B が「おもちゃを買ってあげる」とした場合, 子供は「部屋を片付けないとおもちゃを買ってあげません」という母親の発言を「部屋を片付けるとおもちゃを買ってもらえる」と解釈し, 一生懸命部屋を片付ける. 相互排他性バイアスの例示を Fig. 2 に示す. この時, 母親は子供に「部屋を片付けないとおもちゃを買ってあげません」と ”罰” として伝えているが, 子供は母親の発言を「部屋を片付けるとおもちゃを買ってもらえる」と “報酬” として捉えている. 子供は事象間の因果関係を混同しているが, 結果的に円滑な意思の疎通には成功している [Matoba et al. 2011]. 相互排他性バイアスは幼児の語彙学習に有用に働くとされている [Markman & Wachtel 1988, Merriman et al. 1989, Markman 1990, Diesendruck & Markson 2001, Halberda 2003, Birch et al. 2008, Takahashi et al. 2010]. 例えば幼児がリンゴを初めて見たとき「これの名前は, リンゴである」と学んだ後, 次に幼児が初めてオレンジを見たとき「これ（オレンジ）の名前は, リンゴではない」と考えることがある. これにより幼児は, リンゴ・オレンジという名前を混同することなく, 正しく覚えることができる. つまり, 語彙学習における相互排他性バイアスとは, 「別の対象には別の名称が存在する」という仮定をもたらし, この仮定のもとで幼児は学習を進めていくという効果を与える [Markman 1989, Imai et al. 1994].

1.2.3.3 因果推論モデル

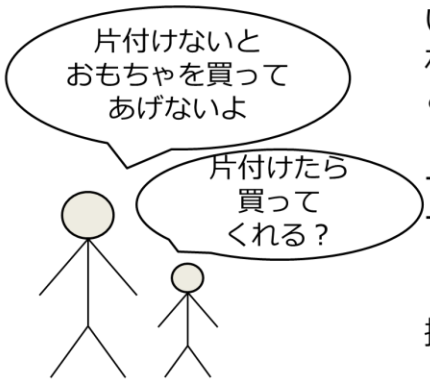
これ等の認知バイアスは, 原因と結果の関係性を調べる因果推論の研究において広く研究されてきた [篠原ら 2007]. 原因 p と結果 q が存在する時, それらの関係性は Table 1 に示す 2×2 分割表によって表すことができる.

・原因 p から結果 q が発生することを観察した後に, 結果 q を目撃した際, その原因が p であると考える傾向
- 原因 p:雨が降った
- 結果 q:地面が濡れている

・$p \to q$ から $q \to p$ を対称的に導く

・これは論理的には誤りだが, 素早い判断を可能にする

Figure 1. 対称性バイアスの例示



・「原因 $\neg p$ が満たされなければ結果 $\neg q$ は起こらない」 を 「原因 p が満たされれば結果 q が起こる」と考える傾向のこと

-原因 p: 部屋を片付けない
-結果 q: 買ってもらえない

・$p \to q$ から $\neg p \to \neg q$ を排他的に導く

・論理的には誤りだが, 左の図の例では円滑な意思疎通の役に立っている

Figure 2. 相互排他性バイアスの例示

Table 1. Contingency table of the LS model.

	q	$\bar{q}$
p	a	b
$\bar{p}$	c	d

この時, a は原因 p があった時に結果 q が起きた確率 $P(q|p)$, b は原因 p があった時に結果 q が起きなかった確率 $P(\bar{q}|p)$, c は原因 p がなかった時に結果 q が起きた確率 $P(q|\bar{p})$, d は原因 p がなかった時に結果 q が起きなかった確率 $P(\bar{q}|\bar{p})$ である. $\bar{p}$ および $\bar{q}$ はそれぞれ p と q の否定である.

これまで考案された因果推論モデルの一例に [Jenkins & Ward 1965] の ΔP モデルや, [Hattori & Oaksford 2007] の Dual Factor Heuristics (DFH) モデルがある. ΔP モデルは式 (1)-(6) で求められ, 原因 p があった時に結果 q が起きた確率と, 原因 p が起きなかった時に結果 q が起きた確率の差を考え, その差が 0 よりも十分に大きい場合は因果関係があると考え, そうでない場合は因果関係がないとするものである.

$$P(q|p) = \frac{a}{a+b} \tag{1}$$

$$P(q|\bar{p}) = \frac{c}{c+d} \tag{2}$$

$$P(p|q) = \frac{a}{a+c} \tag{3}$$

$$P(\bar{q}|\bar{p}) = \frac{d}{c+d} \tag{4}$$

これ等の共変動情報から, 原因と結果の依存度 ΔP は式 (5), (6) のように求まる.

$$\Delta P(q|p) = P(q|p) - P(p|q) \tag{5}$$

$$\Delta P(q|\bar{p}) = P(q|\bar{p}) - P(q|p) \tag{6}$$

また，この時, 原因と結果が $\bar{p}$ と $\bar{q}$ であるとすると式 (7) に示す関係が得られ，この式が相互排他性を常に満たすことがわかるが，対称性は満足しない [篠原 & 中野 2007].

$$\begin{aligned} \Delta P(\bar{q}|\bar{p}) &= \Delta P(\bar{q}|\bar{p}) - P(\bar{q}|p) \\ &= \Delta P(\bar{q}|\bar{p}) - (1 - P(q|p)) \\ &= \Delta P(q|p) \end{aligned} \tag{7}$$

一方, DFH モデルは，原因から結果の予測可能性 $P(q|p)$ および原因の結果に対する適合性 $P(p|q)$ の 2 つの確率が共に強い時は原因と結果の関係性が強く認知され，いずれも低い時は因果関係がないと推論される．DFH モデルは対称性を満足するが，相互排他性を満足しない [篠原 & 中野 2007, 篠原ら 2007]. DFH モデルにおける原因と結果の因果推論を式 (8)-(9) に示す.

$$H(q|p) = \sqrt{P(q|p)P(p|q)} \tag{8}$$

$$H(q|\bar{p}) = \sqrt{P(q|\bar{p})P(\bar{p}|q)} \tag{9}$$

Rigidly Symmetric (RS) モデルは，篠原らによって提案されたモデルであり，対称性 $RS(p|q) = RS(q|p)$ および相互排他性 $RS(\bar{p}|q) = RS(\bar{q}|p)$ が常に成立する [篠原 & 中野 2007]. このモデルは，極めて強い対称性を持つため，完全対称性モデルと呼ばれるが，実験から RS モデルは人間の評価値と弱い相関しか見られず，これは相互排他性バイアスが一貫して働くためとされている [Ohmura et al. 2012]. 人間には，これ程までに強くこれ等の認知バイアスがかかっているとは考えにくいとされる [篠原 & 中野 2007]. RS モデルにおける原因と結果の因果推論を式 (10) に示す.

$$RS(q|p) = \frac{a+d}{a+b+c+d} \tag{10}$$

以上の歴史的経緯を土台として，対称性バイアスと相互排他性バイアスを含み，なおかつその効き具合の調整を可能とするよう開発された Loosely Symmetric モデルについて，次節で紹介する．

1.3 Loosely Symmetric モデル

篠原らは，対称性バイアスと相互排他性バイアスを緩やかに満たすモデルが少量データからの概念獲得に貢献するとして，Loosely Symmetric (LS) モデルを考案した [篠原 & 中野 2007]. LS を用いた因果推論は，式 (11)-(14) のように示される．

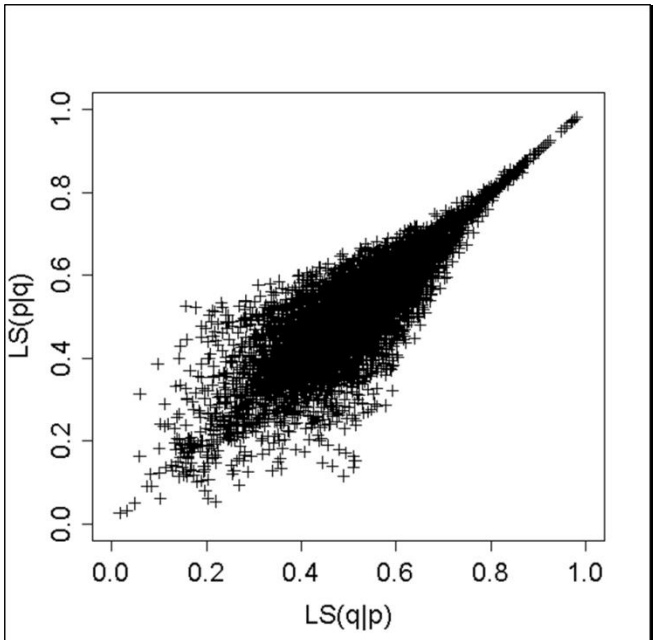
$$LS(q|p) = \frac{a + \frac{bd}{b+d}}{a + b + \frac{ac}{a+c} + \frac{bd}{b+d}} \tag{11}$$

$$LS(\bar{q}|p) = \frac{b + \frac{ac}{a+c}}{a + b + \frac{ac}{a+c} + \frac{bd}{b+d}} \tag{12}$$

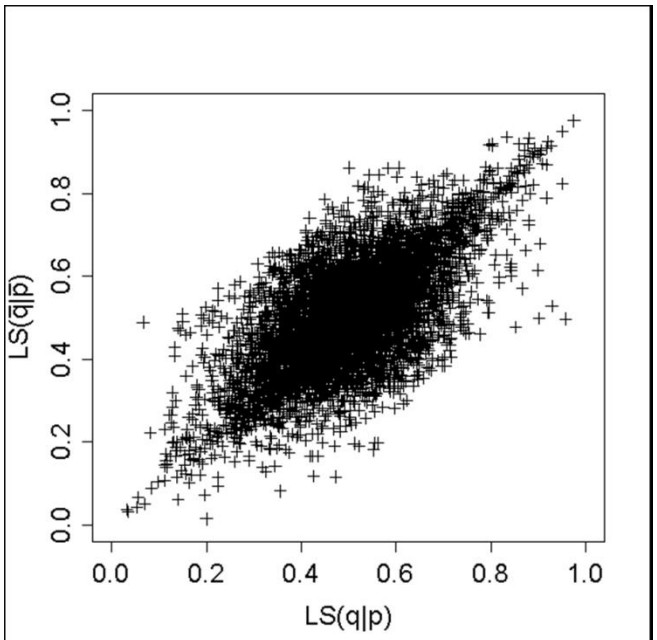
$$LS(p|q) = \frac{a + \frac{cd}{c+d}}{a + c + \frac{ab}{a+b} + \frac{cd}{c+d}} \tag{13}$$

$$LS(\bar{q}|\bar{p}) = \frac{d + \frac{ac}{a+c}}{c + d + \frac{ac}{a+c} + \frac{bd}{b+d}} \tag{14}$$

これらの式は一見複雑に見えるが，実際には条件付き確率にわずかな変更を加えたものである．例えば式 (11) は条件付き確率に $ac/(a+c)$, $bd/(b+d)$ という二つの項を加えたものである．これ等の二項の値が 0 の時，LS は条件付き確率と等価であり，対称性バイアスも相互排他性バイアスも含まれない．また $b = c$ が満たされるとき，式 (11) と式 (13) は同値となり，完全に対称性バイアスを満たす．また，$b = c$ と $a = d$ の両方が満たされるとき，式 (11) と式 (13), 式 (14) は同値となり，対称性バイアスと相互排他性バイアスを完全に満たす．Fig. 3 (a)-(b) に LS における $p \to q$ と $q \to p, p \to q$ と $\bar{p} \to \bar{q}$ の関係を示す．



(a) Relation between $LS(q|p)$ and $LS(p|q)$



(b) Relation between $LS(q|p)$ and $LS(\bar{q}|\bar{p})$

Figures 3. Symmetric and mutually exclusive relationships represented in LS.

図中の各点は区間 $[0,1]$ の乱数値を a,b,c,d のそれぞれに与えることにより生成されたもので，ここでは 10000 個の点を生成した．Fig. 3 (a) は $LS(q|p)$ と $LS(p|q)$ の関係を，Fig. 3 (b) は $LS(q|p)$ と $LS(\bar{q}|\bar{p})$ の関係をそれぞれ示す．もしも各バイアスが完全に満たされるならば，$LS(q|p) = LS(p|q)$ と $LS(q|p) = LS(\bar{q}|\bar{p})$ は常に満たされ，グラフは比例関係となる．逆に，これ等が満たされなければランダムプロットとなる．しかし，Fig. 3 (a)-(b) のように LS モデルにおいてこれらは比例関係とランダムプロットの中間となり，$ac/(a+c)$ および $bd/(b+d)$ の二つの項が対称性バイアスと相互排他性バイアスの利き具合の柔軟な調整を行っていることがわかる．このことから，LS は対称性バイアスと相互排他性バイアスを緩く (loosely に) 満たす [Takahashi et al. 2011]．次に，本研究で比較対象のために用いる，これまで良く普及している機械学習モデルについて概観する．

1.4 機械学習モデル

教師あり学習に属する手法の中でも代表的なものに，パーセプトロン [Rosenblatt 1958] やロジスティック回帰 (LR) [Cox 1958] がある．また，後にパーセプトロンを改良したニューラルネットワーク (NN) [Werbos 1975]，サポートベクターマシン (SVM) [Vapnik 1963] および，ベイズの定理 [Bayes & Price 1763] を基にしたナイーブベイズ (NB)，Breiman による Random Forest (RF) [Breiman 2001] が考案され，様々なタスクにおいて優れた性能を示してきた．以下にこれらの手法の説明を行う．

1.4.1 ベイズの定理とナイーブベイズ

ベイズの定理は，複数の事象間の条件付き確立に関して成り立つ式であり，観測された事象の頻度情報を基に確率分布を求めることを目的としている [Bayes

& Price 1763]. 事象 X, Y が存在し，それ等がお互いに独立でない時，以下の式 (15) が成り立つ.

$$P(X|Y) = \frac{P(X)P(Y|X)}{P(X)P(Y|X) + P(\bar{X})P(Y|\bar{X})} = \frac{P(X)P(Y|X)}{P(Y)} \tag{15}$$

ベイズの定理の考えの基礎となるのは，発生確率が未知の事象が存在し，それが起きた回数と起こらなかった回数が与えられているとした時，1 回の試行においてその事象が発生する確率と発生しない確率とを求めることにある．事象X，Yが存在し，それ等がお互いに独立でない時，結合確率 $P(X \cap Y) = P(X)P(Y)$ のように求まる．一方で，ベイズの定理は $P(X|Y) = \frac{P(X \cap Y)}{P(Y)}$ を求めることで，両事象が発生する確率 $P(X \cap Y)$ を Y が起きる確率で割り，商として導き出し，これを X が既に発生している確率として導く．ベイズの定理が今日大変注目されている理由は，後から起きた事象を調べ，その結果から先に起きた事象を時間を遡って推定することが可能であることにある．こうした背景から，ベイズの定理は確率論のみならず，統計学や経済学，情報学など様々な分野で幅広く利用がされてきた [McGrayne 2011]．また，ベイズの定理の通時的解釈として，仮説 H，データ D から式 (16) が成り立つ.

$$P(H|D) = \frac{P(H)P(D|H)}{P(D)} \tag{16}$$

この時 $P(H)$ は事前確率を表し，データ D が観測される前の時点の仮説 H に関する情報である．事前確率は，データを観測する前の情報である．一方，尤度 $P(D|H)$ は，データそのものに関する情報であると言える．尤度とは，事象 H を仮定した時のデータ D に関する情報であり，$P(D)$ は証拠と呼ばれ定数を常に取る.

ベイズの定理を基にした機械学習手法に，ナイーブベイズがあり，たびたびス

パムメール分類をはじめとするテキスト分類に用いられる [Ng & Jordan 2002]. スパムメール分類タスクにおいて，テキストの属するクラスは $C = \langle spam, ham \rangle$，$c_i \in C$ と表され，テキストから得られる語特徴ベクトルは $W = \langle w_1, w_2, \ldots, w_{|W|} \rangle$ と置かれる．この時 $spam$ はスパムメール，ham は非スパムメールを意味する．W が c_i に所属する確率 $P(c_i|W)$ は式 (17) のように求められる．

$$P(c_i|W) = \frac{P(c_i)P(W|c_i)}{P(W)} = \frac{P(c_i)\prod_{j=1}^{|W|} P(w_j|c_i)}{P(W)} \tag{17}$$

この時，証拠 $P(W)$ は全てのクラスにおいて同値を取るため，式 (18) のように省略可能である [Domingos and Pazzani 1997] .

$$P(c_i|W) \propto P(c_i)\prod_{j=1}^{|W|} P(w_j|c_i) \tag{18}$$

NB 分類器は，テキスト中の語句同士が条件付き独立であるという仮定を置き，テキストのクラスへの所属確率を算出する [Alpaydin 2014]. この仮定は，独立性仮定と呼ばれ,クラスと特徴間,および特徴同士の発生の確率は独立であると仮定をするものである．現実のスパムメール分類においては，“money” や “casino” といった単語は同一のテキストに共起しやすく，また ham メールよりも $spam$ メールから観測されやすい [Conway & White 2012] . しかしながら，この仮定により NB 分類器は n 次元の特徴ベクトルを 1 次元の分布として見積もることができ [Dougherty 2013] , アルゴリズムの簡易化と計算速度の向上を実現している [Domingos & Pazzani 1997] . NB は，最もシンプルな形式のベイジアン・ネットワーク [Friedman et al. 1997] と呼ぶことができ [Zhang 2004] , Fig. 4 のような構造を持つ.

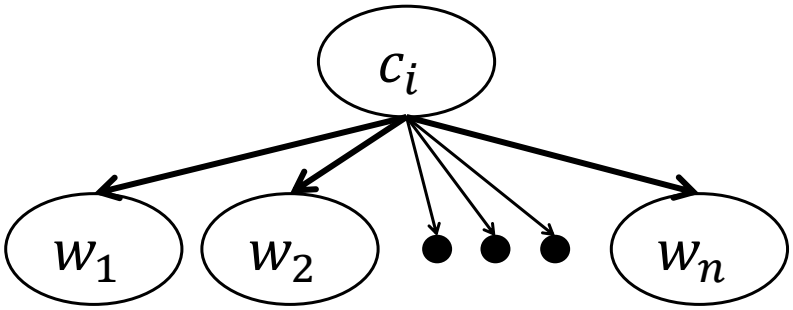


Figure 4. ナイーブベイズの木構造

1.4.2 ニューラルネットワーク

1.4.2.1 形式ニューロン

NN の基となった考えに，McCulloch & Pitts の形式ニューロンがある [McCulloch & Pitts 1943]. 人間など動物の神経系は，ニューロン（神経細胞）で構成された巨大なネットワークであり,それぞれのニューロンはソーマ（神経細胞体）とアクソン(軸索)を持つ. このネットワークは，ニューロン間のシナプスの結合によって構築され,シナプスはあるニューロンのアクソンと別のニューロンのソーマとの間に存在する. 全てのニューロンは閾値 (threshold) を持っており，この閾値を超えた電気信号がシナプスから送られた時,ニューロンは発火し活動を開始する. ニューロンのイメージを Fig. 5 に示す. 形式ニューロンはこの神経系の働きに触発をされたもので，x_i を i 番目のニューロンの出力信号，w_i を x_i に対するシナプスの荷重値，u を内部情報，θ を閾値とした時，ニューロンの出力は式 (19)-(21) のように表される.

$$u = \sum_{i=1}^{n} w_i x_i \tag{19}$$

$$f(u) = \begin{cases} 0 \text{ if } u \leq \theta \\ 1 \text{ if } u > \theta \end{cases} \tag{20}$$

$$y = f(u) \tag{21}$$

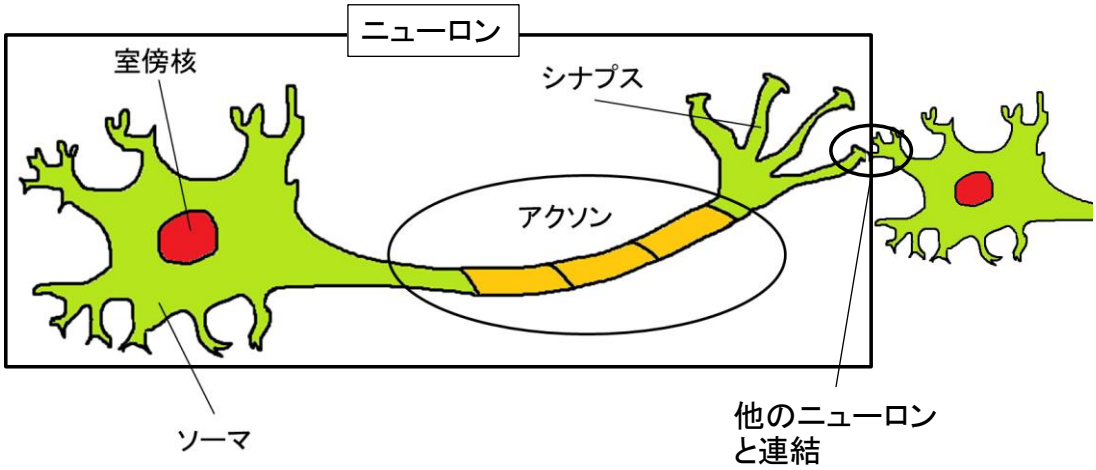


Figure 5. ニューロンとその連結

1.4.2.2 ヘッブの法則

Hebb は，ある細胞が他の細胞を発火させた時，その二つの細胞の結合が強まるという考えのもと，ヘッブの法則 (Hebb's rule) を提案した [Hebb 1949]. 細胞 A のアクソンが細胞 B が発火するのに十分近い距離にあり，また両者が繰り返し発火したとする．この時，片方の細胞，あるいは両者に変化が生じ，細胞 B を発火させる細胞の１つとして細胞 A の効率が増加する．W_{ij} をニューロン i とニューロン j の結合荷重，x_i と x_j をニューロンの出力信号，η を学習率とした時，ヘッブの学習法則は式 (22) で表される．

$$\Delta W_{ij} = \eta * x_i * x_j \tag{22}$$

1.4.2.3 パーセプトロン

形式ニューロンに触発されたモデルとして，Rosenblatt のパーセプトロンがある [Rosenblatt 1958]. パーセプトロンは，形式ニューロンの学習則にヘッブの学習法則を用いたもので，S 層(感覚層，Sensory Layer)，A 層(連合層，Associative

Layer), R 層(出力層, Response Layer)から構成される３層のネットワークである [Rosenblatt 1958]. 入力ベクトル $X = (x_1, \ldots, x_{|X|})^T$ が存在し, w_i を x_i に対するシナプスの荷重値 , w_0 をバイアス, z を出力とした時, パーセプトロンは式 (23)-(24) のように表される.

$$\eta = f(\sum_{i=1}^{/X/} w_i x_i + w_0) \tag{23}$$

$$f(\eta) = \begin{cases} 1 & \text{if } \eta \geq 0 \\ 0 & \text{otherwise} \end{cases} \tag{24}$$

1.4.2.4 ニューラルネットワークとバックプロパゲーション

NN は人間の脳の働きに触発された学習アルゴリズムであり, 入力層, 隠れ層, 出力層と呼ばれる３種類の層を持つ [Rumelhart et al. 1986]. これ等の層はそれぞれ１つ以上のノードを含む. ３層ニューラルネットワークのイメージを Fig. 6 に示す. 入力層のノード数は, 特徴ベクトルの次元数に等しく, 出力層のノード数は, クラスの数に応じ変化する. 隠れ層に含まれるノードの数は任意である. m 個の層を持つ NN があり, k 層目の j 番目のノードへの総入力を x_j^k, このノードの出力を y_j^k, $k-1$ 層目の i 番目のノードから k 層目の j 番目のノードへの結合荷重を $w_{i,j}^{k-1,k}$ とする. ノードの活性化関数をロジスティックシグモイド関数とした時, 各ノードの出力は式 (25)-(26) のように表される.

$$y_j^k = \frac{1}{1 + e^{-x_j^k}} \tag{25}$$

$$x_j^k = \sum_i^n w_{i,j}^{k-1,k} y_i^{k-1} \tag{26}$$

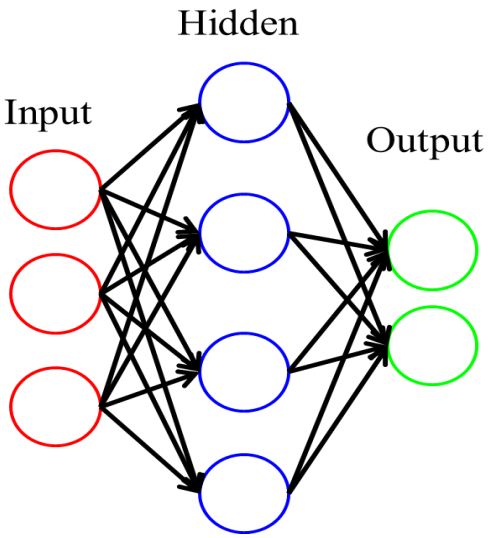


Figure 6. ３層ニューラルネットワーク

この時，出力層のノード y_i^m と真値 t_i の誤差関数を 2 乗和誤差関数とした場合，誤差関数 E は式 (27) のように表される．δ_i^m はネットワークの出力と真値との差である．

$$E = \frac{1}{2}\sum_i \left(y_j^m - t_j\right)^2 = \frac{1}{2}\sum_i (\delta_i^m)^2 \tag{27}$$

この誤差関数 E は，教師信号と出力との差の 2 乗に比例して大きくなり，これが減少するように結合荷重 $w_{i,j}^{k-1,k}$ の値を更新していくのがバックプロパゲーションである [Werbos 1975]．学習係数 α がある時，結合荷重 $w_{i,j}^{k-1,k}$ の変化量 $\Delta w_{i,j}^{k-1,k}$ は，式 (28) のように計算される．

$$\Delta w_{i,j}^{k-1,k} = -\alpha \delta_j^k y_j^k \left(1 - y_j^k\right) y_i^{k-1} \tag{28}$$

1.4.3 サポートベクターマシン

サポートベクターマシン (Support Vector Machine, SVM) は，二値分類に利用される機械学習アルゴリズムである [Vapnik 1963]. SVM は，特徴空間に存在する各データ点から，マージンを最大化する超平面を発見する．Fig. 7 に SVM のイメージを示す.

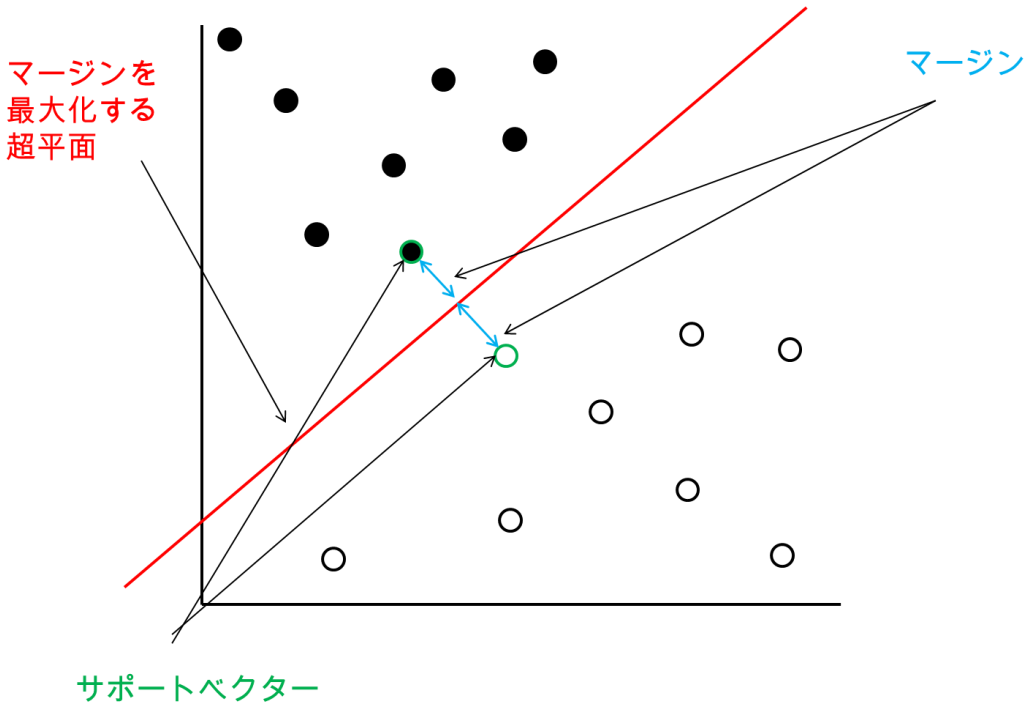


Figure 7. SVM のイメージ

二値のクラス情報 $y_i \in \{-1,1\}$ および，教師データから得られた N 次元の特徴ベクトル $(x_1, y_1), \ldots, (x_N, y_N)$ がある時， SVM は $y_i(w \cdot x_i + b) \geq 1$ を満たす超平面および $y_i(w \cdot x_i + b) = 1$ を満たすサポートベクター x_i を発見することを目的とする．これらのサポートベクターは，各クラスにおける超平面を決定し，それらの距離はマージンとして定義される．このマージンは，重みベクトル $||w^*||$ を最小化するよう式 (29) のように計算され，この時 a_i は係数，" ・ " はドット積，および a_j は制約係数を表す.

$$W(\alpha) = \sum_{i=1}^{N} a_i - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} a_i\, a_j (x_i \cdot x_j) y_i y_j \tag{29}$$

SVM の評価関数は，式 (30) のように求められる．なお，この時 b は，バイアス項である．

$$F(x_j) = sign\{w \cdot x_j - b\} \tag{30}$$

1.4.4 ロジスティック回帰

ロジスティック回帰 (Logistic Regression, LR) は，統計的回帰モデルの一種であり，ベルヌーイ分布に基づいた二値分類を行う [Cox 1958]. LR はベルヌーイ分布に基づき，確率 π_i で 1 を出力し，確率 $1-\pi_i$ で 0 を出力する [King & Zeng 2001] . 特徴 Y_i がある時，ロジスティック回帰は，式 (31)-(32) のように求められる．

$$Y_i \sim \text{Bernoulli}(Y_i|\pi_i) \tag{31}$$

$$\pi_i = \frac{1}{1+e^{-X_i\beta}} \tag{32}$$

この時，ベルヌーイ分布は $\mathrm{P}(Y_i|\pi_i) = \pi_i^{Y_i}(1-\pi_i)^{1-Y_i}$ となり，パラメータ $\beta = (\beta_0, \beta_1')'$ である．

1.4.5 ランダムフォレスト

ランダムフォレスト (RandomForests, RF) は，Breiman によって考案されたアンサンブル学習手法である [Breiman 2001]. RF は木構造を持つ弱教師あり学習器を複数生成し，それらの多数決によってクラスの予測を行う．この時， RF は N 個の決定木を，ブートストラップ法によってサンプリングし、決定木を構成す

る特徴の重複を許容する．決定木として表される弱教師あり学習器 $h_1(x), h_2(x), \ldots, h_n(x)$ が教師データのサンプリングに用いる乱数ベクトル X, Y がある時, 以下のマージン関数が定義される．

$$mg(X,Y) = a\,v_k I(h_k(X) = Y) - \max_{j \neq Y} a v_k I(h_k(X) = j) \tag{33}$$

ここで $I(\cdot)$ は指示関数であり，マージン関数は弱教師あり学習器の多数決の結果によって、クラス間の距離を推定する．この時, RF の損失 は以下の式 (34) によって求められる．

$$E = P_{X,Y}(mg(X,Y) < 0) \tag{34}$$

1.5 データの偏りと少量データに関する機械学習手法

本節では教師データの偏り，並びにその数が少量である場合における，機械学習の手法を俯瞰する．こうした手法の代表的なものにオーバーサンプリング (over sampling), アンダーサンプリング (under sampling), data augumentaiton があり，これ等の手法は機械学習モデルを問わず利用可能なものである．

1.5.1 オーバーサンプリングとアンダーサンプリング

オーバーサンプリングとアンダーサンプリング [He et al. 2008] は，教師データの分布を調整する手法であり，.オーバーサンプリングはあるクラスにラベル付けされた教師データが，他のクラスに属するデータと比べ少量であった場合，その数を補完する手法である．一方，アンダーサンプリングは，あるクラスにラベル

付けされた教師データが，他のクラスに属するデータと比べ多量であった場合，少量のデータから得られた分布へと補完する手法である．これ等の手法は．特定のクラスに対するオーバーフィッティング，およびアンダーフィッティングを防ぐために用いられる．

1.5.2 Data Augumentation

Data Augumentation とは，教師データに何かしらの補正を加えることで，そのデータ量を増加させる手法である．一例として，画像データに対する Data Augumentation は，画像データの場合，画像の反転を行ったり，サイズの拡大・縮小を行うことで，データ量の増加を試みる．つまり，データの母集団に何らかのノイズを加え，増加分のデータを教師データとして加えることが Data Augumentation の動作と言える [Tanner & Wong 1987].

1.6 論文の構成

本論文では，人間の認知バイアスを機械学習に応用することにより，少量かつ不均衡なデータから優れた学習を行う機械学習モデルを提案し，スパムメール分類タスク，病気の良性･悪性分類タスク，データの生成タスクの評価についての議論を行う．本論文は４つの章から構成される．続く第２章において機械学習を用いたスパムメール分類器への認知バイアスの適用，第３章にて医療データ分類タスクにおける認知バイアスの適用を行う．第２章のスパムメール分類タスクにおける課題として，スパムメールの数の増加，およびその構成の変化が挙げられる．このため，教師データが少量の場合においても，より正確に分類を行うことができる手法の開発が望まれている．その実現のため，NB をベースとしたスパムメール分類器に人間の認知バイアスを適用し，少量かつ偏りのあるデータからでも優れた学習を行うモデルを提案する．また，その評価のため，既存の機械学習アルゴリズムとの性能評価を (1) 教師データが少量である場合，(2) 教師データが不均衡である場合，(3) 教師データが少量かつ不均衡である場合の３つの実験設

定のもとで行った．また，人間の認知バイアスの機械学習への適用のもう一つのアプローチとして，第３章では医療データをもとに，病気の良性･悪性分類を行った．医療データはプライバシーの理由などから大規模なデータが手に入りにくく，また良性・悪性のデータがバランスよく手に入るとも限らない．こうしたことから，本研究では病気の良性・悪性分類にたびたび用いられるニューラルネットワークに認知バイアスを適用し，不均衡なデータからでも優れた学習を行うモデルを提案する．第４章において，本論文の統括および，結論を示す．

第 2 章　スパムメール分類タスクにおけるナイーブベイズへの認知バイアスの適用

本章では, Loosely Symmetric (LS) モデルの機械学習への応用として, ナイーブベイズ (NB)に LS モデルを実装し, 少量の教師データからでも素早い学習を行う手法を提案する. スパムメール分類を題材とした実験の結果とその考察を通じて, 提案手法の妥当性を示す.

2.1 はじめに

スパムメールは 1978 年に最初に観測されて以来, インターネット技術の発展に伴って爆発的に増加した [Pitsillidis et al. 2010]. 有害メールをスパム (Spam) と呼称するようになったのは 1993 年ごろであり, この名称は米国の肉製品, および英国のテレビ番組に由来する [Rao & Reiley 2012]. スパムメールの数は年々増加傾向にあり, 2012 年には全てのメールの 90% にあたるメールがスパムに属するという報告がなされた [Rao & Reiley 2012]. スパムメールはその数の膨大さから, データ量や通信量の膨大な浪費が懸念されている [Kanaris et al. 2006]. また, スパムメールの形態は近年多様化しており, その危険性が近年ますます高まっていることと, 検出の困難化が指摘されている [Rao & Reiley 2012] . こうしたことから, スパムメールを自動で取り除くことができるスパムメール分類器の研究が盛んに行われている [Rao & Reiley 2012, Androutsopoulos et al. 2000]. 機械学習によるスパムメールの検出は1990年代ごろから行われており, この技術は電子メールをスパムメールか, ハムメールのいずれかに分類するスパムメール分類タスクと呼ばれている [Rao & Reiley 2012]. 機械学習モデルは, スパムメールとハムメールを教師データとして学習を行う [Rao & Reiley 2012, Androutsopoulos et al. 2000, Conway and White 2012]. この際に教師データとして扱う電子メールにはラベル付けが施されている必要があるが, この作業は多くの場合, 手動で行われる [Rao & Reiley 2012]. このため, 機械学習を用いたスパムメール分類には時間的コ

ストがかかる傾向にある．また，機械学習の研究で利用されるスパムメールコーパスと，現実のスパムメールの構成には相違点がある．スパムメール分類タスクにおいてたびたび用いられるデータセットに，Ling-Spam コーパス [Androutsopoulos et al. 2000] と SpamAssassin コーパス [Mason 2003] がある．これ等のコーパスが含むスパムメールの割合は約 20-30% である．一方で，今日インターネット上でやり取りされるメールは，90% 以上がスパムである．このことから，現実においてやり取りされるデータは，機械学習の研究に用いられるデータセットよりも大幅に不均衡であると言える．また，スパムメールの形態は年々多様化しており [Goodman et al. 2007]，特定の単語のスペルを記号で置き換えたものや，ランダムな単語を頻出させ，分類器の判断を狂わせるテクニックが出現している [Eberhardt 2015]．こうしたことから，不均衡なデータからの優れた学習，並びに未知語が頻出する状況において，適切な分類を行うスパムメール分類器の開発が望まれる．

本章では，この問題の解決のため人間の認知バイアスを導入した，教師あり学習によるスパムメール分類器の提案を行う．実験では，上記の問題の再現のため，少量かつ不均衡な教師データのみを機械学習モデルに与え，性能比較を行った．また，少量かつ不均衡な教師データが与えられた状況下において，認知バイアスがどのように機械学習モデルの学習を補助するのかを議論する．本章で提案する手法はスパムメール分類のみを対象としている訳ではなく，むしろ人間の認知的特性を機械学習に導入することで，少量かつ不均衡なデータからの優れた学習を実現し，それを通じて機械学習の人間レベルの学習を目的とする．本章では，人間の持つ「曖昧性を含む推論」が優れた学習を実現するという仮定のもと，ごく少量かつ偏りを含むデータから優れた学習を行う分類モデルを提案する．提案手法は，人間の認知バイアスを用いることで，統計的な有意さを得ることが難しい状況においても，より柔軟に重み付けを行う．第 1 章でも述べたように，人間は統計的手法，あるいは確率論に基づく手法とは異なる判断基準で事象間の関係性の推論を行う．その例として [Tversky & Kahneman 1973] の研究があり，人間は発生確率が等しい複数の事象に遭遇しても，事象ごとに独自の重み付けを行い，

それを基に意思決定を行う．このため，認知バイアスは必ずしも真の解を導くとは限らないものの，限られた情報からの意思決定に貢献する．本章におけるアプローチは，既存の機械学習モデルが不均衡データからの偏った分布に適合してしまう現象を回避することを目標としており，これは人類が進化の過程で得た「生存のために手元の情報に賭ける」ことを機械学習に導入する試みである．つまり，機械学習を始めとする人工知能の手法が「データから得た情報」から学習フレームを形成するのに対し，提案手法は複数の特徴ベクトル間で対称性，相互排他性を導き出し，「データから得た情報」と「データから得られなかった情報」の双方を学習する．本研究では機械学習と認知科学の双方の観点から，より人間に近い学習能力を持つ機械学習手法を提案する．実験結果から，提案モデルは不均衡データからのより優れた概念学習に成功した．

2.2 スパムメール分類における認知バイアス

本節では，人間の認知バイアスである対称性バイアスと相互排他性バイアスが，スパムメール特有の傾向から，スパムメール分類タスクにおいて有効に働くと仮定し，機械学習モデルへの適用を行った．スパムメールは，たびたび “Casino” や “Slot” といったスパムらしい単語 (spam-likely words) を含む [Conway & White 2012]．ここで p を ”あるメールがスパムらしい単語を含んでいる”, q を ”あるメールはスパムである” と見なすとする．この時，対称性バイアスは “もしもメールがスパムであれば (q)，このメールはスパムらしい単語を含んでいる (p)” という事象から “もしもメールがスパムらしい単語を含んでいるのであれば (p)，このメールはスパムである (q)” を対称的に導くことができる．また，相互排他性バイアスは “もしもメールがスパムらしい単語を含んでいないのであれば $(\bar{p})$，このメールはハムである $(\bar{q})$” を相互排他的に導く．これ等の推論形式において，本章では [Edgington 1995] の知見から，$P(if\ p\ then\ q)$ を条件付き主観確率 $P(q|p)$ と等しいものとみなす．これまでの研究から，“もしも p ならば q である” という推論と，条件付き確率との間の関連性が示されてきた [Over & Evans 2003, Over et al. 2007]．一方で, Barrouillet & Gauffroy は，特定の状況下においての

み $P(if\ p\ then\ q) = P(q|p)$ が満たされるとした [Barrouillet & Gauffroy 2015]. しかしながら, [Edgington 1995] における $P(if\ p\ then\ q)$ と $P(q|p)$ は等しいとする考えは本章のタスクにおいて有用に働くと考えられる. 例えば, “もしもメールがスパムであれば (q), このメールはスパムらしい単語を含んでいる (p)” という事象と, 条件付き主観確率 P (このメールはスパムらしい単語を含んでいる | このメールはスパムである) は満たされる. このため, スパムメール分類タスクは本章での手法を検証するための, 最も優れたテストベッドと考える.

2.3 提案手法

本節では, 人間の認知バイアスを機械学習に利用したモデルである loosely symmetric naive Bayes (LSNB) および enhanced LSNB (eLSNB) を提案する.

2.3.1 Loosely Symmetric Naïve Bayes

まず篠原らの LS モデルを NB に適用した LSNB を開発した結果を示す. LSNB のスパムメール分類における因果推論として, クラス c_i およびテキストから得られた特徴ベクトルに含まれる単語 $w_j \in W$ が存在する時, c_i と w_j の共起頻度の 2×2 分割表を Table 2 のように示す.

Table 2. Contingency table used in the LSNB model.

	w_j	$\overline{w_j}$
c_i	a	b
$\overline{c_i}$	c	d

$a = P(w_j|c_i)$ はクラス c_i から単語 w_j が観測される確率であり, $b = P(\overline{w_j}|c_i)$ は c_i から w_j が観測されない確率を表す. また, $c = P(w_j|\overline{c_i})$ はクラス $\overline{c_i}$ から

w_j が観測される確率であり，$d = P(\overline{w_j}|\overline{c_i})$ は $\overline{c_i}$ から w_j が観測されない確率を表す．LSNB は対称性バイアスと相互排他性バイアスに基づいて，式 (35)-(38) のように重みの調整を行う．

$$a = P(w_j|c_i) \tag{35}$$

$$b = P(\overline{w_j}|c_i) \tag{36}$$

$$c = P(w_j|\overline{c_i}) \tag{37}$$

$$d = P(\overline{w_j}|\overline{c_i}) \tag{38}$$

そして，w_j の c_i における尤度は LS によって重み付され，NB に従い式 (39)-(41) のように変更される．

$$P_{LS}(w_j|c_i) = \frac{a + \frac{bd}{b+d}}{a + b + \frac{ac}{a+c} + \frac{bd}{b+d}} \tag{39}$$

$$P_{LS}(W|c_i) = \prod_{j=1}^{n} P_{LS}(w_j|c_i) \tag{40}$$

$$P_{LS}(c_i|W) = P(c_i)P_{LS}(W|c_i) \tag{41}$$

例えば，もしも “money” や “casino” といった単語が，ham クラスのテキストよりも $spam$ クラスのテキストから高い頻度で観測された場合，これ等の単語は $spam$ に属するテキストと関連性が高いと考えられる．本モデルはこうした学習状況を考慮したモデルである．NB と LSNB の大きな違いは，尤度の計算方法である．NB において，尤度 $P(W|c_i)$ は $P(w_j|c_i) = \frac{a}{a+b}$ の総積によって求められる．一方で，LSNB では尤度 $P_{LS}(W|c_i)$ を $P_{LS}(w_j|c_i) = \frac{a+\frac{bd}{b+d}}{a+b+\frac{ac}{a+c}+\frac{bd}{b+d}}$ の総積によって求め，a, b, c, d の全ての項を使う．

2.3.2 Enhanced Loosely Symmetric Naïve Bayes

次に LSNB を拡張し，単語密度情報 [Kwan et al. 2002] を付与した eLSNB モデルを提案する．単語密度情報は式 (42)のように示され，$N(c_i \cap w_j)$ はクラス c_i における単語 w_j を含む文章数，$WD(c_i \cap w_j)$ はクラス c_i における単語 w_j の出現回数を，クラス c_i に属する全ての単語の出現回数で割ったものである [Kwan et al. 2002].

$$WD(c_i \cap w_j) = \frac{N(c_i \cap w_j)}{\sum_{k=1}^{|W|} N(c_i \cap w_k)} \tag{42}$$

本モデルの目的は，各クラスに属する各特徴の重みを単語密度情報によって補正し，より優れたバイアス調整を行うことである．例えば単語 w_j が $spam$ クラスのみから頻繁に観測され，なおかつ，ham クラスからは観測されなかった場合，単語 w_j は $spam$ クラスと関係のある単語であると考えられる．eLSNB モデルは単語の出現確率を単語密度で補正することで，より強いバイアスを付与し，クラス間の特徴の差を大きくすることが可能であると予想される．eLSNB モデルにおける共起頻度の計算は式 (43)-(46) のように行われる． $a - d$ はクラスと単語の共起頻度に，単語密度情報をかけ合わせたものである．

$$a = P(w_j|c_i)WD(c_i \cap w_j) \tag{43}$$

$$b = P(\overline{w_j}|c_i)WD(\overline{c_i} \cap w_j) \tag{44}$$

$$c = P(w_j|\overline{c_i})\, WD(\overline{c_i} \cap w_j) \tag{45}$$

$$d = P(\overline{w_j}|\overline{c_i})\, WD(c_i \cap w_j) \tag{46}$$

式 (39)-(41) に示した LSNB における尤度計算と同様，w_j の c_i における尤度は LS によって重み付され，式 (47)-(48) のように計算される．

$$P_{eLS}(W|c_i) = \prod_{j=1}^{n} P_{eLS}(w_j|c_i) \tag{47}$$

$$P_{eLS}(c_i|W) = P(c_i)P_{eLS}(W|c_i) \tag{48}$$

提案モデルである LSNB モデルおよび eLSNB モデルには,３つの特徴がある．１つ目は，LS モデルの利用により優れたバランスで特徴の重み付けを行うことにより，ノイズや未知データに柔軟に対応可能なことである．２つ目は NB 分類器がテキストのクラスへの所属確率の計算にそのクラスに属する N 次元の特徴ベクトルを用いるのに対し，提案モデルはあるクラスにテキストが属する確率を計算するのに，複数の特徴ベクトルを用いることである．そして３つ目は NB 分類器の持つ実装の簡易さと処理速度の速さを保ちつつ，より優れた識別が可能なことである．

2.4 不均衡かつ少量の教師データを用いたスパムメール分類実験

2.4.1 実験設定

本節の実験では，機械学習の代表的な手法であり，スパムメール分類タスクにたびたび利用される NB，ニューラルネットワーク (NN)，ロジスティック回帰 (LR)，サポートベクターマシン (SVM)，ランダムフォレスト (RF) を比較対象とし，提案モデルである LSNB と eLSNB を加えた計 7 モデルを用い，実験 1 で３種，実験 2 で２種，実験 3 で１種の計６種の実験を行った．

- 実験 1, 2 には，スパムメール分類における代表的なデータセットである SpamAssassin, Ling-Spam の２つのメールコーパスを利用した．
- 実験 3 には, Ling-Spam コーパスのみを利用した．

実験 1 と 2 で利用した SpamAssassin コーパス [Mason 2003] は，1897 通の $spam$ メールと 3900 通の ham メールを含み，コーパス内の $spam$ メールの割

合は 33% である．また, Ling-Spam [Androutsopoulos et al. 2000, Androutsopoulos 2004] コーパスは, Linguist List [Androutsopoulos et al. 2000] から抽出された 481 通の $spam$ メールと 2412 通の ham メールを含み, コーパス内の $spam$ メールの割合は 17% である．実験では, メール本文中の名詞や動詞を原形に戻され, ストップワード [Salton 1989] が取り除かれたバージョンである *lemm* バージョンを用いた．ストップワードとは, 英語における頻出語を纏めたデータである．例えば "a" や "the" などと言った単語はクラスに関係なくあらゆる英文で頻出するものである．こうしたストップワードは, 文章分類の結果に影響を与えないため, 分類の前に特徴ベクトルから取り除いた．

特徴選択として, 事前にメール本文から句読点や数字を取り除いた.また, Burstiness の観点から,一度しか観測されなかった単語も同様に取り除いた．Burstiness とは, 文章中で一度観測された単語は,再度観測されやすいという現象である [Katz 1996, Sarkar et al.2005]. Burstiness の考えは, たびたびスパムメール分類タスクに有用に働くとされ [Schneider 2005], 本実験においても採用した．つまり, 文章の内容を全て確認し終えた後, 一度しか観測されなかった特徴は分類に影響を与えないと判断し, 特徴ベクトルから削除した．

比較対象に用いた各アルゴリズムのパラメータ設定を以下に示す．

- SVM のカーネル法には文章分類にたびたび利用されるガウシアンカーネルを用い, コストパラメータは $C = 1$ とした．
- NN は 3 層のフィードフォワード NN を用い, 活性化関数は二値分類に頻繁に用いられるシグモイド関数とした．中間層の数およびエポック数は試行錯誤の結果, 中間層 10 ノード, エポック数 100 を最良と判断し設定した．
- また, この実験は二値分類であるため, LR は二項ロジスティック回帰分析を用い, $\alpha = 1$ とした．
- RF はツリーの数を 300 とし, [Breiman 2001] に従い, ランダムサンプルする変数の数は特徴数の平方根とした．
- NB, LSNB, eLSNB の実験設定として, 事前確率はクラス毎に均一に設定し, 事前確率が分類に影響しないようにした．つまり $P(spam) = 0.5, P(ham) =$

0.5 とした.

- 各機械学習モデルのハイパーパラメータの選択には, グリッドサーチを行い, 最も良い成績となったものを設定した.
- 分類精度の指標には, 再現率 (recall) [Alpaydin 2014] を用いた. 各条件における試行回数は 50 回とし, その平均を成績とした. また, $spam$ クラスを陽性, ham クラスを陰性として扱った. 各グラフのエラーバーは標準誤差である.

2.4.2 不均衡な教師データを用いた実験 1

実験 1 の目的は, 機械学習モデルの性能が不均衡データによって, どのような影響を受けるのかを調査することである. この実験では, クラス間でデータに偏りがある状況下で, $spam$ 判別性能, ham 判別性能および F-measure の比較を行った. 教師データ量が $spam:ham = 1:1\ (spam = 50\%)$ の場合と $spam:ham = 3:5\ (spam = 38\%)$ の場合, $spam:ham = 1:5\ (spam = 17\%)$ の場合の 3 種の評価を行った. 教師データの数は, 実験 1-1 では $spam, ham$ データ共に 20 刻みで 240 個まで増加させた. 実験 1-2 では $spam$ データは 12 刻みで 240 個まで, ham データは 20 刻みで 400 個まで増加させた. また, 実験 1-3 では $spam$ データは 4 刻みで 240 個まで, ham データは 20 刻みで 1200 個まで増加させた. 実験 1-1, 1-2, 1-3 では, 教師データの割合をそれぞれ $spam:ham = 1:1, spam:ham = 3:5, spam:ham = 1:5$ で維持し, 各実験設定における機械学習モデルの性能を比較した.

2.4.2.1 実験 1-1 の結果と考察

実験 1-1 の SpamAssasin コーパスを用いた際の $spam$ 分類精度, ham 分類精度および F-measure を Fig. 8 (a)-(c) に, Ling-Spam コーパスを用いた際の実験結果を Fig. 8 (d)-(f) にそれぞれ示す.

実験全体を通じて, eLSNB, LSNB, NN および RF は $spam$ 分類において高い

性能を見せた．NB は実験を通じて *spam* 分類の成績が向上せず，他のモデルと比較して成績が低迷した．また *ham* 分類において，eLSNB，LSNB，NB および NN は高い判別成績を見せた．LR は実験を通じて *ham* 分類精度が向上せず，全モデル中最も低い成績を示した．この実験において教師データ数の最大値は 500 未満と少量であった．このため，各分類モデルは適切に特徴の重み付けを行うことが困難だったと考えられる．しかしながら，eLSNB，LSNB，NN および RF は比較的高い F-measure の値を示した．また，eLSNB と LSNB は，*spam* 分類の判別性能と *ham* 分類の判別性能の両方が，基となった NB よりも向上した．この実験において，eLSNB は F-measure の結果について最も高い成績を示し，この結果は LS モデルおよび単語密度情報が，より適した尤度の推定に貢献したと考えられる．

全モデルが教師データの増加に伴い *spam* 判別性能と *ham* 判別性能の向上を見せ，高い水準の F-measure スコアを示した．しかしながら，SVM と LR は他のモデルと比較して低い *spam* 判別成績と *ham* 判別成績を見せた．この実験における教師データの総数は 500 未満であり，データの少なさがこれ等のモデルの性能低迷に繋がったと考えられる．これとは対称的に，LSNB と eLSNB はベースとなった NB よりも両分類結果において高い性能を見せ，全モデル中最良の成績となった．

2.4.2.2 実験 1-2 の結果と考察

実験 1-2 の SpamAssasin コーパスを用いた際の *spam* 分類精度，*ham* 分類精度および F-measure を Fig. 9 (a)-(c) に，Ling-Spam コーパスを用いた際の実験結果を Fig. 9 (d)-(f) にそれぞれ示す．

この実験における *spam* 教師データの割合は，実験 1-1 における割合よりも高いものの，全モデルに大きな性能差は見受けられなかった．例えば NB，LSNB，eLSNB，NN は，実験 1 全体において近い成績を見せた．一方で RF，LR，SVM は，*spam* 分類の判別性能と *ham* 分類の判別性能の間でトレードオフを見せた．トレードオフとは，あるクラスの分類精度が向上する一方で，その他のクラスの分

類精度は低下する現象である [Alpaydin 2014]. これ等のモデルは, 実験 1-1 の結果と比較して *ham* 分類精度が向上し, *spam* 分類精度は低下した. このため, RF, LR, SVM は *spam* 教師データの割合の変化に, 他のモデルと比べ強い影響を受けた.

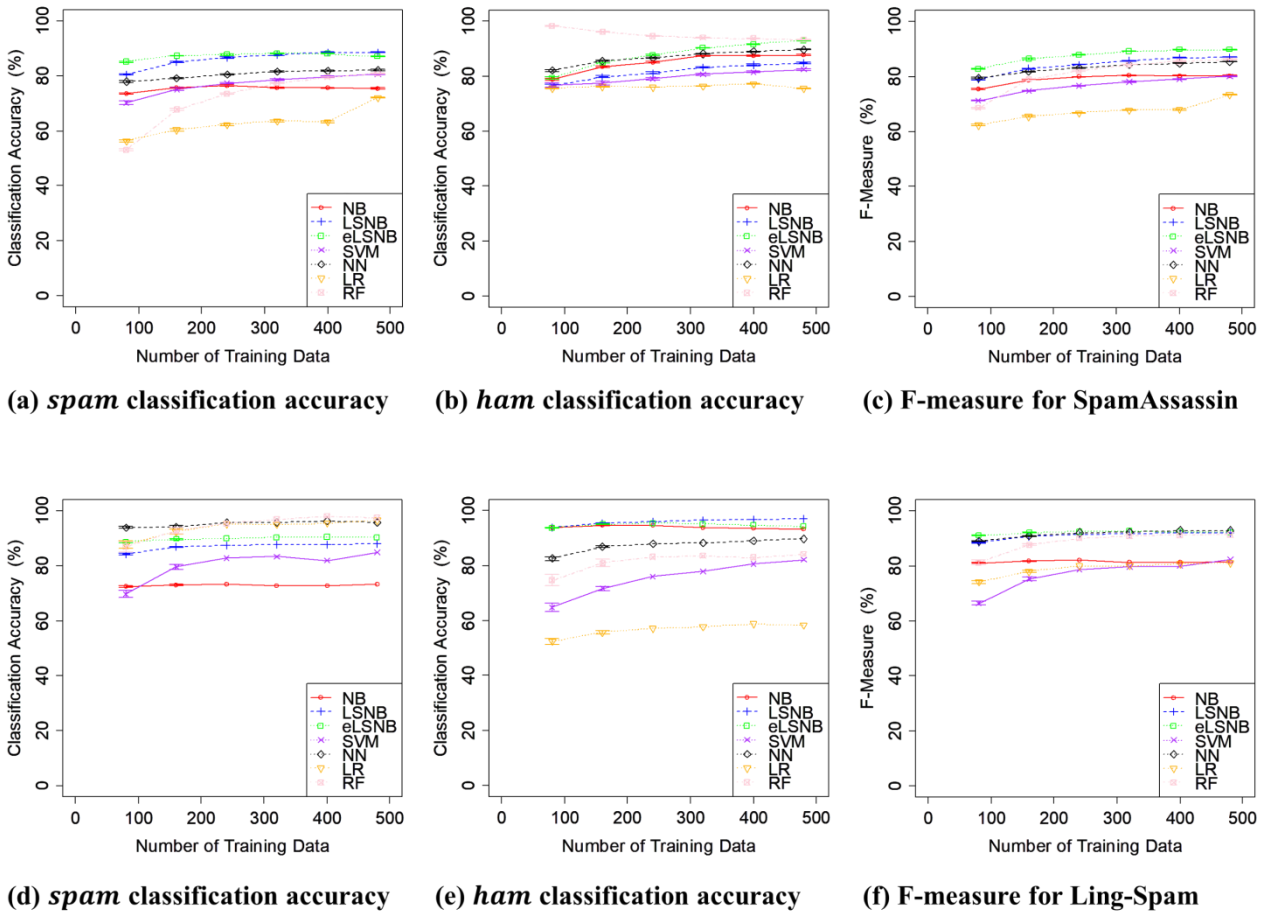
SVM, LR, RF の F-measure の値は実験 1-1 の結果と近しい水準となった. RF は他の実験と比べ高い F-measure の値を示したものの, データのバランスに対する鋭敏性を見せた. この実験設定では少量の教師データのみを用いて機械学習モデルのトレーニングを行ったため, 特徴に適切な重み付けを行うことが困難であった可能性がある. 一方で, eLSNB, LSNB および NN はトレードオフを見せず, 高い水準の *spam* 分類精度, *ham* 分類精度, F-measure を保った.

2.4.2.3 実験 1-3 の結果と考察

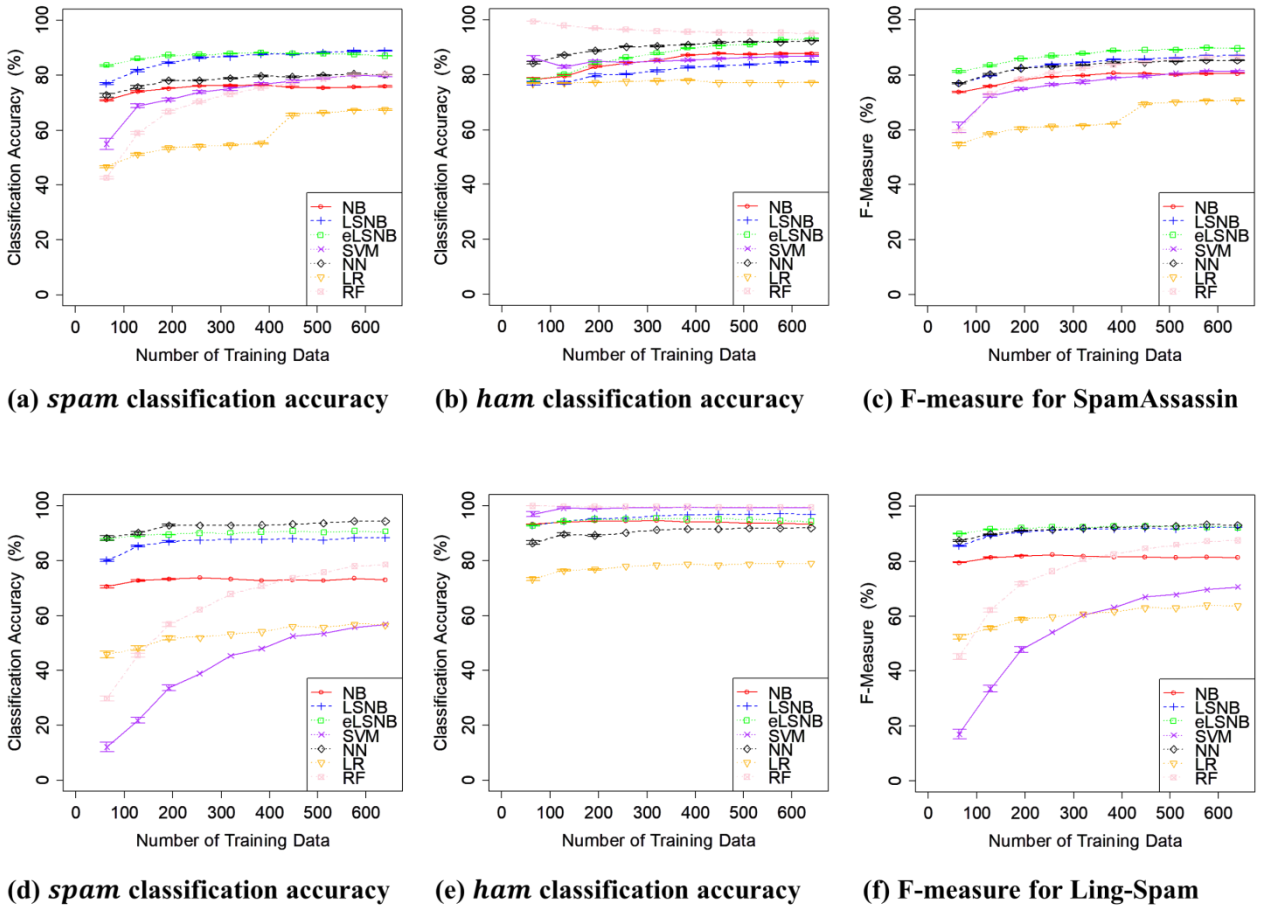
実験 1-3 の SpamAssasin コーパスを用いた際の *spam* 分類精度, *ham* 分類精度および F-measure を Fig. 10 (a)-(c) に, Ling-Spam コーパスを用いた際の実験結果を Fig. 10 (d)-(f) にそれぞれ示す.

SVM と RF は *ham* 分類精度においてほぼ 100% の成績を示し, NN と eLSNB がそれに続いた. SVM, RF, NN は, 実験 1-1 と実験 1-2 と比べ, *ham* 判別成績が向上した. 一方, SVM, RF, LR の *spam* 分類精度は他のモデルよりも低い水準を示し, また, SVM, RF および NN は実験 1-1 と実験 1-2 の結果と比べ *spam* 分類精度が低下した.

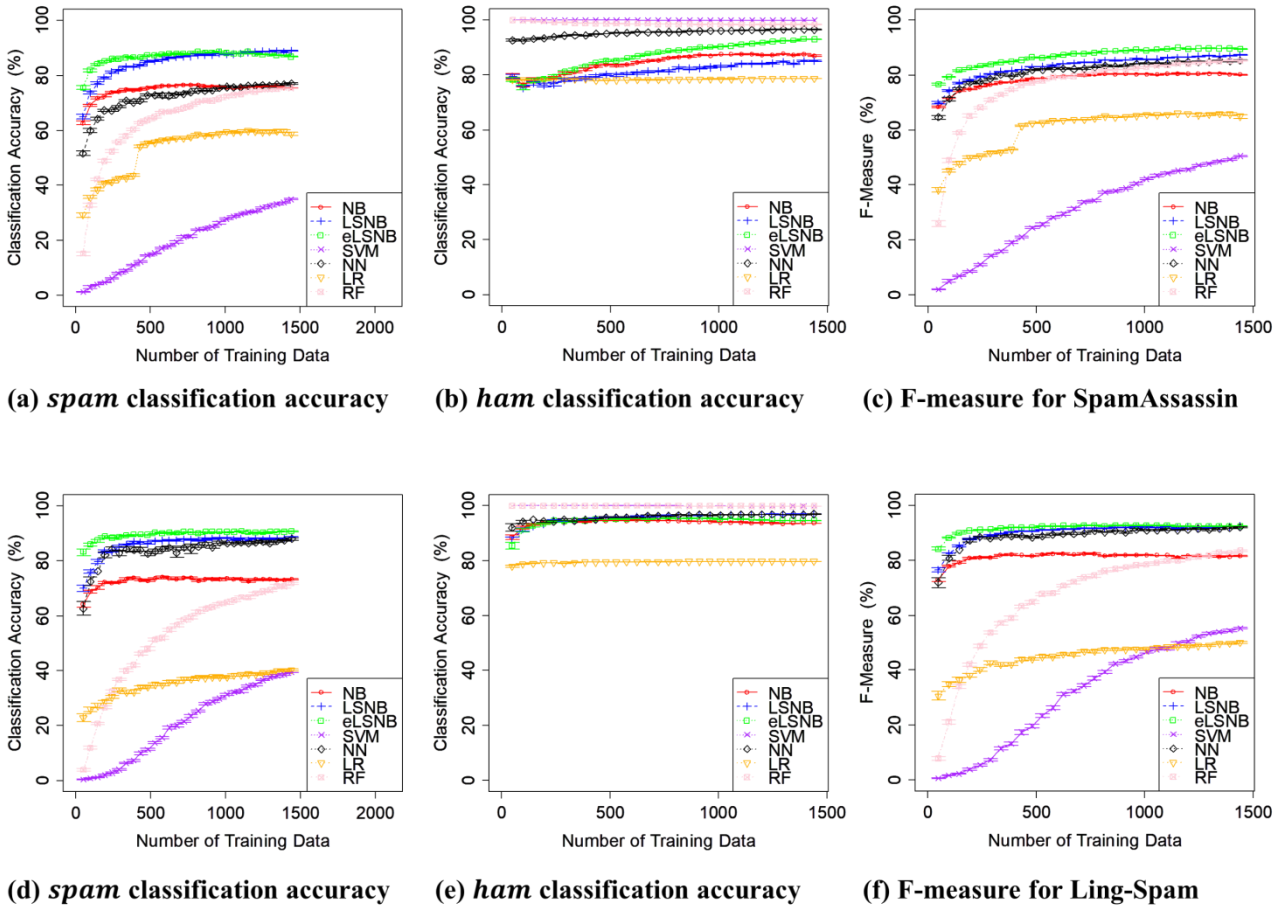
SVM と RF はより大きなトレードオフを見せた. これらのモデルは教師データが少量の段階においても, 100% に近い *ham* 判別成績を見せた. しかしながら, SVM と RF の *spam* 判別成績は非常に低い水準となった. この 2 モデルの *spam* 判別成績は教師データの増加に伴ってわずかに向上したものの, 他の分類モデルの成績には及ばなかった. また, NN は実験 1-1 と実験 1-2 ではトレードオフを見せなかったものの, この実験では *spam* 判別成績が減少し, *ham* 判別成績が向上した.



Figures 8. The results of Exp. 1-1. The error bars indicate the standard errors.



Figures 9. The results of Exp. 1-2. The error bars indicate the standard errors.



Figures 10. The results of Exp. 1-3. The error bars indicate the standard errors.

NB はトレードオフを見せず，高い *ham* 判別成績を示したものの，*spam* 判別成績は他のモデルと比べ低い水準に留まった．一方，提案手法である LSNB と eLSNB はベースとなった NB の性能を向上し，より高い *spam* 判別成績を示すことに成功した．さらに，RF, LR, SVM のトレードオフは，実験 1-2 の結果と比べ拡大した．また，NN は実験 1-1，実験 1-2 では観測されなかったトレードオフを見せた．一方, eLSNB, LSNB, NB は他のモデルのようなトレードオフを見せず，*spam* 分類において高い成績を示した．実験 1 において，NB，LSNB および eLSNB は教師データが不均衡な状況においても，トレードオフを見せなかった．また，提案手法である eLSNB と LSNB は，その基のモデルとなった NB よりも優れた性能を見せた．このことから eLSNB と LSNB は少量かつ不均衡データから既存の分類モデルよりも安定した学習を行い，少量データからの概念学習という，人間のような学習能力を一定の条件下で見せた．実験 1 の結果において，LSNB と eLSNB はトレードオフを見せず， NB よりも優れた性能を見せ，最も高い F-measure を見せた．

2.4.3 データ量を少数かつ固定値とした教師データを用いた実験２

実験 2 では，片方のクラスからのデータ量を少量かつ固定とし，もう一方のクラスからのデータ量のみを増加した際の判別性能の比較を行った．*spam* 教師データ数を固定値とし *ham* 教師データ数を増加させる実験と，*ham* 教師データ数を固定値とし *spam* 教師データ数を増加させる実験の２種を行った．実験 2-1 では *spam* 教師データ数は 100 あるいは 25 のいずれかを取り，*ham* データは 20 刻みで 240 個まで増加させた．実験 2-2 では，実験 2-1 と逆に，*ham* 教師データ数を固定値とし，*spam* 教師データ数を増加させた．*ham* 教師データ数の取る値と *spam* 教師データ数の増加量は，実験 2-1 と同様である．例えば *spam* 教師データ量を 50 に固定し，*ham* 教師データ量を 20 から 40 刻みで

240 まで増加させた場合，教師データ量の最大値は 290 となり，*spam* と *ham* 間のデータ量の差は 190 となる．実験 2 は少量かつスパースな情報のみを用いた際の学習性能の比較が目的である．

2.4.3.1 実験 2-1 の結果と考察

この実験では *spam* 教師データ数を固定値とし，*ham* 教師データ数のみを増加させた．実験 2-1 の *spam* 教師データ数を 100 に固定し，SpamAssasin コーパスを用いた際の *spam* 分類精度，*ham* 分類精度および F-measure を Fig. 11 (a)-(c) に，Ling-Spam コーパスを用いた際の実験結果を Fig. 11 (d)-(f) にそれぞれ示す．また，*spam* 教師データ数を 25 に固定し，SpamAssasin コーパスを用いた際の *spam* 分類精度，*ham* 分類精度および F-measure を Fig. 12 (a)-(c) に，Ling-Spam コーパスを用いた際の実験結果を Fig. 12 (d)-(f) にそれぞれ示す．

spam 教師データ数を 100 に固定した実験結果において，全てのモデルが *ham* 分類精度を，実験を通して上昇させた．一方，NN, LR, RF, SVM は教師データの総数が増加するにつれて，*spam* 分類精度が低下した．このため，これらの分類モデルは不均衡な教師データに対する鋭敏性を見せ，*ham* 教師データ数よりも *spam* 教師データ数の方が少ない状況下において，性能が低下した．教師データのバランスが取れていた場合，分類モデルは特徴に対してより適切な重み付けが可能である．しかしながら，この実験における特徴の分布は，非常に不均衡に設定されていた．このため，既存の機械学習モデルである NN, LR, RF および SVM は特徴に対し適切な重み付けができず，*ham* 教師データ数の増加に伴った *spam* 分類精度の低下を見せた．一方，LSNB と eLSNB は *spam* 分類精度，*ham* 分類精度のいずれも低下せず，最も高い F-measure を示した．

spam 教師データ数を 25 に固定した実験結果において，RF, SVM, NN は，より高い *ham* 分類精度を見せた．しかしながら，これらの分類モデルは *spam* 分類において，成績が低迷した．特に，SVM は非常に高い水準の *ham* 分類精度を見せた一方で，*spam* 分類精度は全モデル中最も低かった．また，NN, RF および LR は SVM と似たトレードオフを見せた．提案モデルのひとつである LSNB

もまた，実験 1 では観測されなかった不均衡データに対する鋭敏性を見せた．LSNB は高い水準の *ham* 分類精度を示したものの，教師データの数が増加するにつれて，*spam* 分類精度が低下した．一方で，NB と eLSNB にはそのような傾向は見受けられなかった．このことから，LSNB は対称性バイアスと相互排他性バイアスの調整に失敗したか，あるいはバイアスの効き具合が強くなりすぎた可能性があり，実験 1 では優れた成績を見せたものの，不均衡データに対する一種の鋭敏性が観測された．eLSNB は LSNB のようなトレードオフを見せず，全モデルの中でも良好な *spam* 分類精度と *ham* 分類精度を見せた．また，F-measure の結果において eLSNB は最良の成績を示した．

2.4.3.2 実験 2-2 の結果と考察

この実験では *ham* 教師データ数を固定値とし，*spam* 教師データ数のみを増加させた．*ham* 教師データ数は 100 あるいは 25 のいずれかを取り，*spam* データは 20 刻みで 240 個まで増加させた．このため，実験 2-1 とは対照的に，*spam* 教師データ量は実験を通じて増加するのに対し，*ham* 教師データ量は固定である．*ham* 教師データ量が $ham = 100$ の SpamAssasin コーパスを用いた際の *spam* 分類精度，*ham* 分類精度および F-measure を Fig. 13 (a)-(c) に，Ling-Spam コーパスを用いた際の実験結果を Fig. 13 (d)-(f) にそれぞれ示す．また，*ham* 教師データ数を 25 に固定し，SpamAssasin コーパスを用いた際の *spam* 分類精度，*ham* 分類精度および F-measure を Fig. 14 (a)-(c) に，Ling-Spam コーパスを用いた際の実験結果を Fig. 14 (d)-(f) にそれぞれ示す．

ham 教師データ数を 100 に固定した実験結果において，全てのモデルが *spam* 分類精度を，教師データ数の増加に伴い上昇させた．一方で，NN, SVM, LR, RF は実験を通じて *ham* 分類精度が低下した．このことから，これらの分類モデルは実験 2-1 のような不均衡データに対する鋭敏性を見せた．

また，*ham* 教師データ数を 25 に固定した実験結果においては，eLSNB と NB は *ham* 分類において優れた成績を示した．NN, SVM, LR, RF は実験 1 および実験 2-1 では安定して高い *ham* 分類精度を見せていたものの，この実験では

成績が低迷した．また，その一方でこれらの分類モデルは最も高い水準の *spam* 分類精度を示した．

実験 2-1 と実験 2-2 におけるデータの割合は対称的である．このため，実験 2-1 における各モデルの *spam* 分類精度と，実験 2-2 における各モデルの *ham* 分類精度は似た水準となった．このような傾向はデータ分布や特徴分布が *spam* 教師データと *ham* 教師データとで対称的でない限り，観測されないと考えられる．例えば，仮に *spam* クラスに属するデータが *ham* クラスに属するデータよりも分類が簡単であった場合，実験 2-1 での *spam* 分類精度と実験 2-2 での *ham* 分類精度は非対称となるはずである．

多くの分類モデルが教師データの増加，並びにデータ分布が不均衡になるに伴って成績が低下し，特徴分布に対する鋭敏性を見せた．実験 1 と実験 2 で観測されたように，SVM, LR, RF, NN は教師データの割合に大きな影響を受けた．NB からは強いトレードオフは観測されなかったものの，その成績は他のモデルと比べ低迷した．また，提案手法のひとつである LSNB は，実験 2 において *ham* 分類精度が向上した一方で，*spam* 分類精度が低下するトレードオフを見せ，認知バイアスの調整が成績の向上に繋がらなかった．一方で eLSNB にはそのような鋭敏性は観測されず，単語密度情報が特徴分布に対する鋭敏性を回避するのに貢献したと考えられる．実験 1，実験 2 の結果から，全ての条件において eLSNB は最も高い水準の F-measure を示した．このことから，eLSNB は認知バイアスと単語密度情報を用い，少量かつ不均衡な教師データからの，優れた学習を実現した．

2.5 10 分割クロスバリデーションによる実験結果と考察

この実験では 10 分割クロスバリデーションによる各機械学習モデルの性能比較を行った．この実験の目的は，特徴ベクトルの揺らぎを抑えることであり，各モデルの性能が上記の２つの実験と比べ，どのように変化するかを検証した．他

の２つの実験同様，この実験においても教師データを少量とした．つまり，教師データは $spam, ham$ それぞれデータセットの 10% にあたる数とし，残りはテストデータとした．また，結果の比較に荷重の調整を行った SVM (weight-initialized)，およびオーバーサンプリング (over sampling)，アンダーサンプリング (under sampling) を調整した LR と RF を比較対象に加えた．Ling-Spam コーパスを用いた 10 分割クロスバリデーションにおける分類精度を Table 3 に示す．各実験結果は 10 試行の平均である．

学習にかかる計算コストは，LSNB は NB の 1.08 倍の CPU 時間，eLSNB は NB の 1.31 倍の CPU 時間であった．

NB と LSNB は高い $spam$ 判別成績を見せ，ham 分類精度は約 0.73 となった．LSNB は NB と比べ $spam, ham$ 両方の分類精度が向上し，より高い F-measure を見せた．また，eLSNB と NN は似た水準の成績となり，約 0.8 の $spam$ 判別成績，および約 0.9 の ham 分類精度となった．この実験で eLSNB は最も高い F-measure のスコアを見せ，NN がそれに続いた．SVM は荷重の初期値を変更することで，$spam$ 分類精度が約 0.6 向上し，ham 分類精度は 0.15 低下した．SVM の F-measure のスコアは荷重の初期値の設定を施したものとそうでないものとで，それぞれ 0.793 と 0.257 となり，大幅な差が見られた．一方で LR と RF はサンプリングを実施した後も，大きな性能の変化は見受けられなかった．

既存の機械学習は分類に失敗し，提案手法は正しく判別できたデータの一例を Fig. 15 に示す．このメールは ham にラベル付けされたメールであるが，多くの機械学習モデルは誤って $spam$ メールであると分類したのに対して，提案手法は正しく ham と判別した．このメールのように，料金の請求や，契約費に関する内容の ham メールは，既存の機械学習手法の多くが，誤って $spam$ メールと判別した．

一方で，提案手法が分類に失敗し，既存の機械学習では正しく判別できたデータの例を Fig. 16 に示す．このメールは $spam$ にラベル付けされたメールであるが，提案手法は誤って ham メールであると分類した．このメールのように，

$spam$ らしい単語が文中に存在しない場合，提案手法は $spam$ とは関係ない特徴に過剰に重み付けしてしまうことが，低頻度ではあるものの観測された．

Table 3. 10 分割クロスバリデーションにおける分類精度 (Ling-Spam コーパス).

Model	Spam Classification Acc	Ham Classification Acc	F-measure
NB	0.926	0.733	0.844
LSNB	0.979	0.738	0.874
eLSNB	0.854	0.911	0.879
NN	0.809	0.955	0.873
SVM	0.148	1.0	0.257
SVM (weight-initialized)	0.753	0.854	0.793
LR	0.321	0.990	0.482
LR (resampled)	0.320	0.990	0.481
RF	0.377	0.999	0.547
RF (resampled)	0.387	0.999	0.558

2.6 提案手法の有用性に関する考察

本節では，どのような問題設定において，提案手法が特に有用に働くのかを考察する．提案手法は上記の３種類の実験において優れた成績を示した．このことから，このことから，提案手法は特徴分布から対称性と相互排他性が観測される場合や，そのいずれか片方が観測される場合において，特に優れた性能を示すと予想される．Fig. 17 (a)-(b)に，SpamAssassin コーパスにおける $p \to q$ と $q \to p$ の関係を示す．図中の各点は単語 w_j の分布を表し，横軸はクラス c_i における単語 w_j の出現確率，縦軸は単語 w_j が文章中に含まれる時，その文章のクラスが c_i である確率を表す．Fig. 17 (a) は $P(W|Spam)$ と $P(Spam|W)$ の関係，Fig. 17 (b) は $P(W|Ham)$ と $P(Ham|W)$ を示す．もしも単語の分布とクラス間で対称性が満たされれば，グラフは比例関係となる．Fig. 17 に示すように，$Spam$ およ

び Ham クラスのメールから得られた特徴分布には，対称性が確認される．また，Fig. 18 (a)-(b) に SpamAssassin コーパスにおける $p \to q$ と $\bar{p} \to \bar{q}$ の関係を示す．この図では $P(W|Spam)$ と Ham クラスにおける W の非出現確率 $P(\overline{W}|Ham)$，および $P(W|Ham)$ と $Spam$ クラスにおける W の非出現確率 $P(\overline{W}|Spam)$ の関係をそれぞれ示す．このグラフにおいても，もしも単語とクラス間で相互排他性が満たされれば，グラフは比例関係となる．Fig. 18 において，SpamAssassin コーパスからは強い相互排他性は観測されなかった．提案手法は教師データに含まれる対称性を，LS モデルを用いて調整し，特徴間の強い差別化を行ったと考えられる．また，このことから，提案手法は対称性と相互排他性の両方ではなく，その片方の対称性のみからでも，優れた学習を行うことが示唆された．

また，LSNB の派生系である eLSNB は単語密度情報を用いて対偶の関係性を強め，$p \to q$ と $\bar{q} \to \bar{p}$，および $q \to p$ と $\bar{p} \to \bar{q}$ の関係性 [Ohmura et al. 2012] を導入することで，対称性バイアスと相互排他性バイアスに補正を加えたと考えられる．

2.7 第 2 章のまとめ

本章では LSNB および eLSNB を提案し，スパム分類タスクにおける認知バイアスの有用性を，実験を通じて示した．実験において，教師あり学習モデルに異なる割合の $spam$ 教師データと ham 教師データを与え，その性能比較を行った．

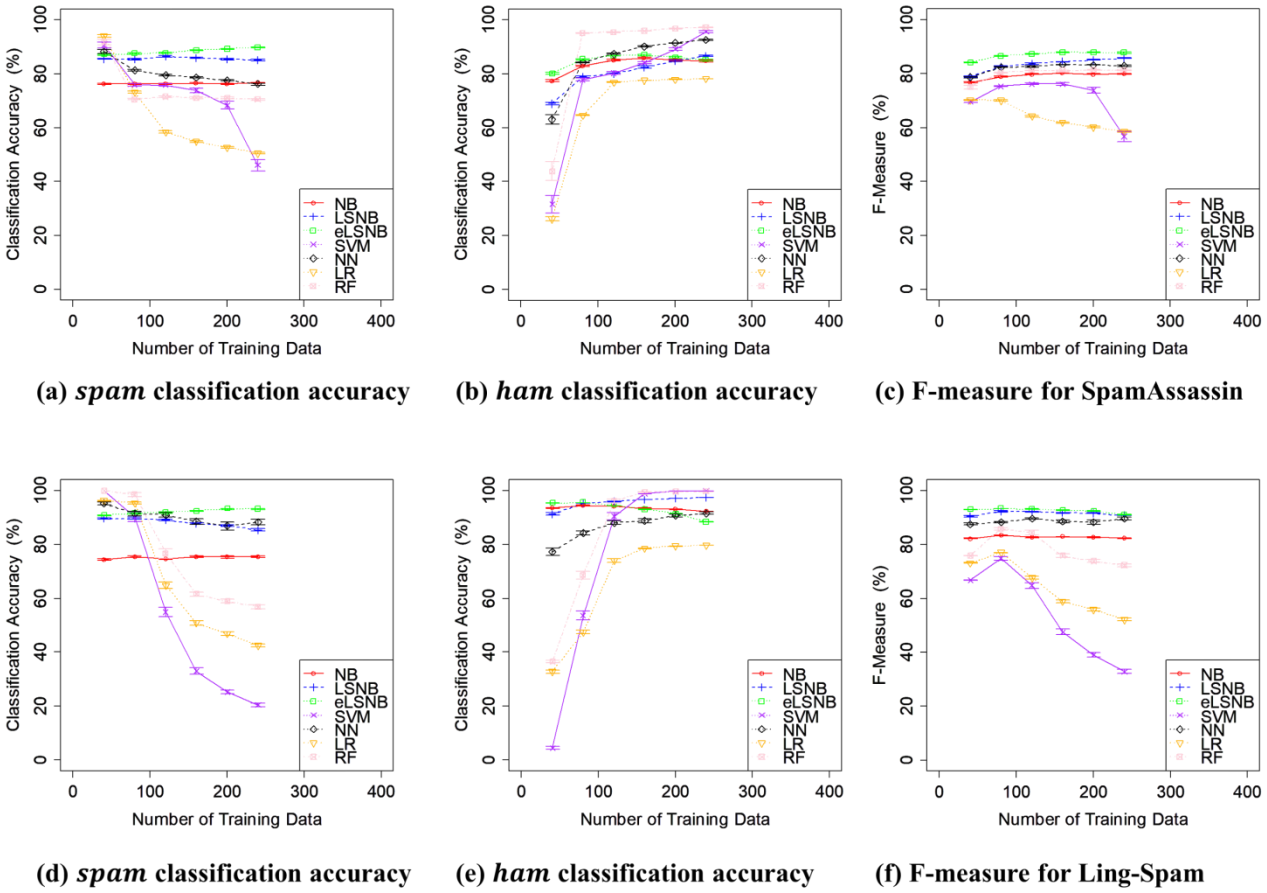
実験 1 において NN, SVM, LR, RF は，特徴の分布に対する一種の鋭敏性を見せ，成績に大きな変動が生じた．NB はトレードオフを見せることはなかったものの，その成績は既存の機械学習モデルの中では低い水準だった．一方で LSNB と eLSNB は NB の性能を大幅に向上させ，全モデル中最良の水準となる F-measure のスコアを示した．

実験 2 においては，少量かつスパースな数の教師データを用いた際の各学習モデルの性能比較を行った．その結果，NB と eLSNB を除く全モデルが実験を通じてトレードオフを見せた．また，$spam$ 教師データから得られた特徴分布と

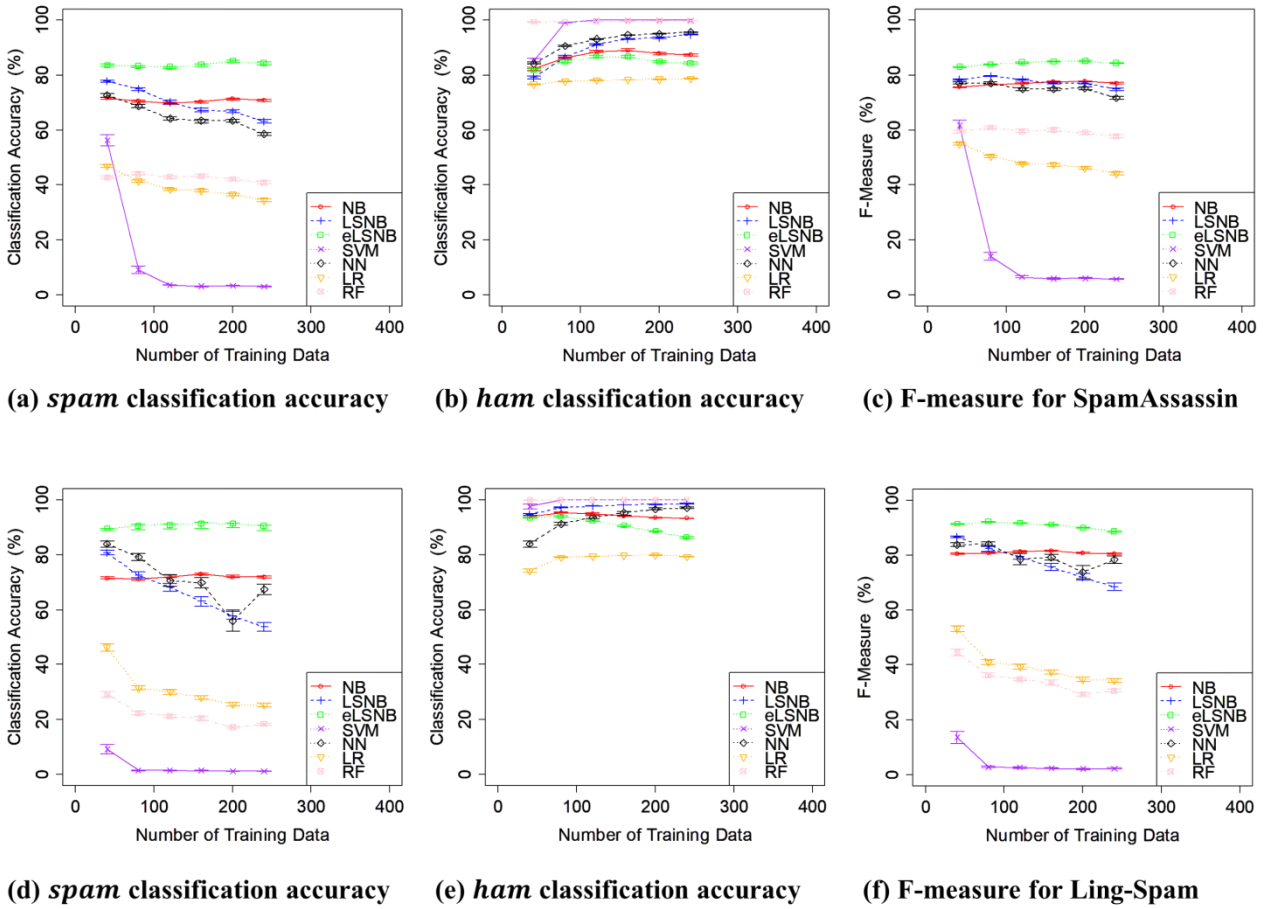
ham 教師データから得られた特徴分布との間に，非対称性は存在しなかったと考えられる．この実験で，SVM, LR, RF, NN は不均衡データに強い影響を受けた．一方，NB はトレードオフを見せることはなかったものの，その成績は低い水準にあった．NN は F-measure の結果においては高い水準であったものの，トレードオフを見せた．LSNB もトレードオフを見せた．一方で eLSNB からはトレードオフが見受けられず，実験を通じて高い水準の F-measure のスコアを示した．

また，実験 3 の特徴ベクトル内のゆらぎを抑えた実験では，SVM は荷重の初期値の設定を施すことで，F-measure が大幅に向上した．LR, RF はサンプリングを調整したものの，大きな精度の変化は見受けられなかった．NB, LSNB, eLSNB, NN は最も高い水準の F-measure を示し，中でも eLSNB が最も優れた精度を見せた．

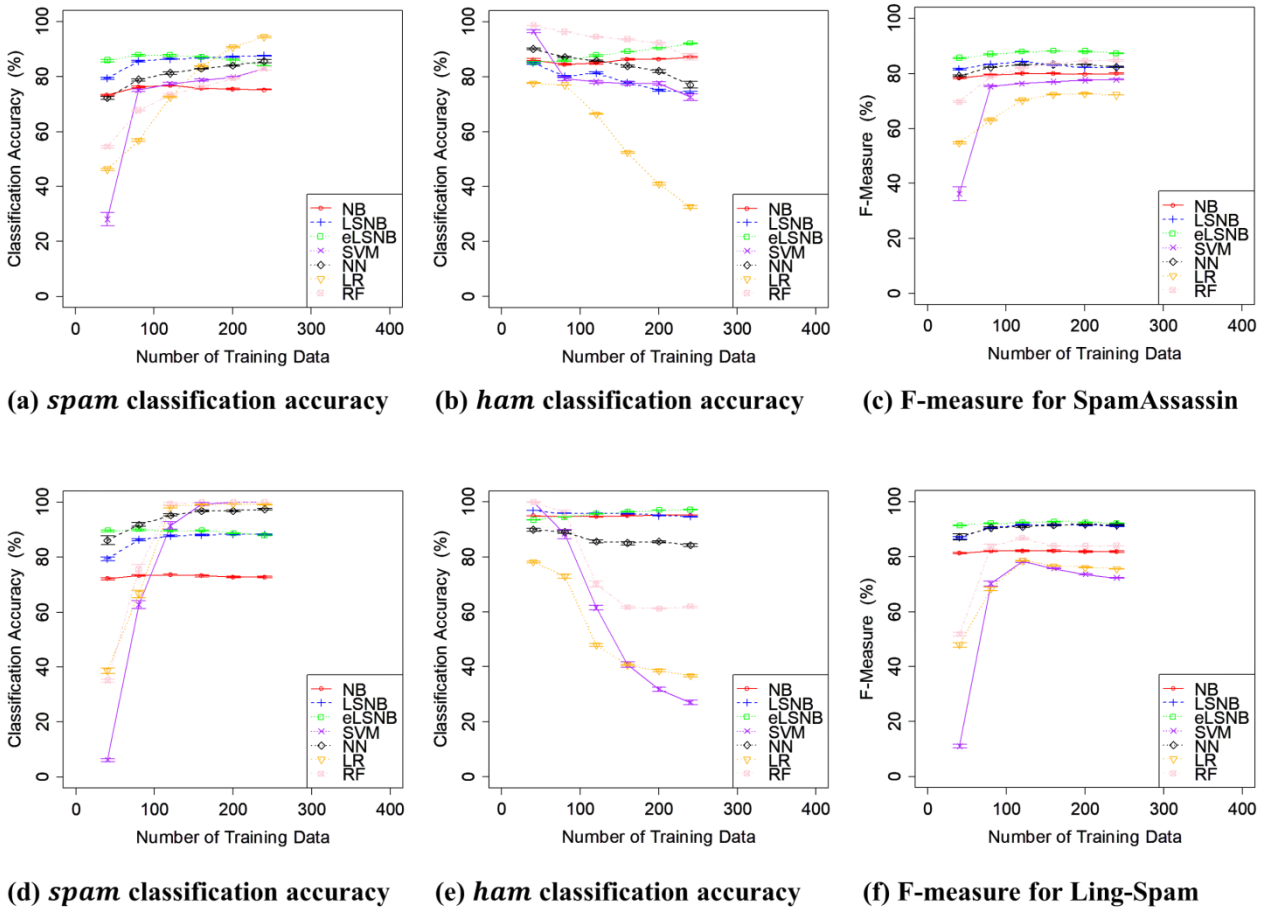
提案モデルは，少量かつ偏りのあるデータからのスパムメール分類タスクにおいて，既存の機械学習モデルと比べ，より優れた性能を見せた．特に，提案手法のひとつである LSNB は特徴分布に対称性あるいは相互排他性のいずれかが観測された時に有用に働くこと示し，またその派生系である eLSNB は認知バイアスと，それに加えて単語密度情報を利用し，対称性バイアスと相互排他性バイアスに更なる補正を加えた．提案手法は少量かつスパースなデータから十分な学習を行うという，人間に近い学習を一定の条件下で達成した．



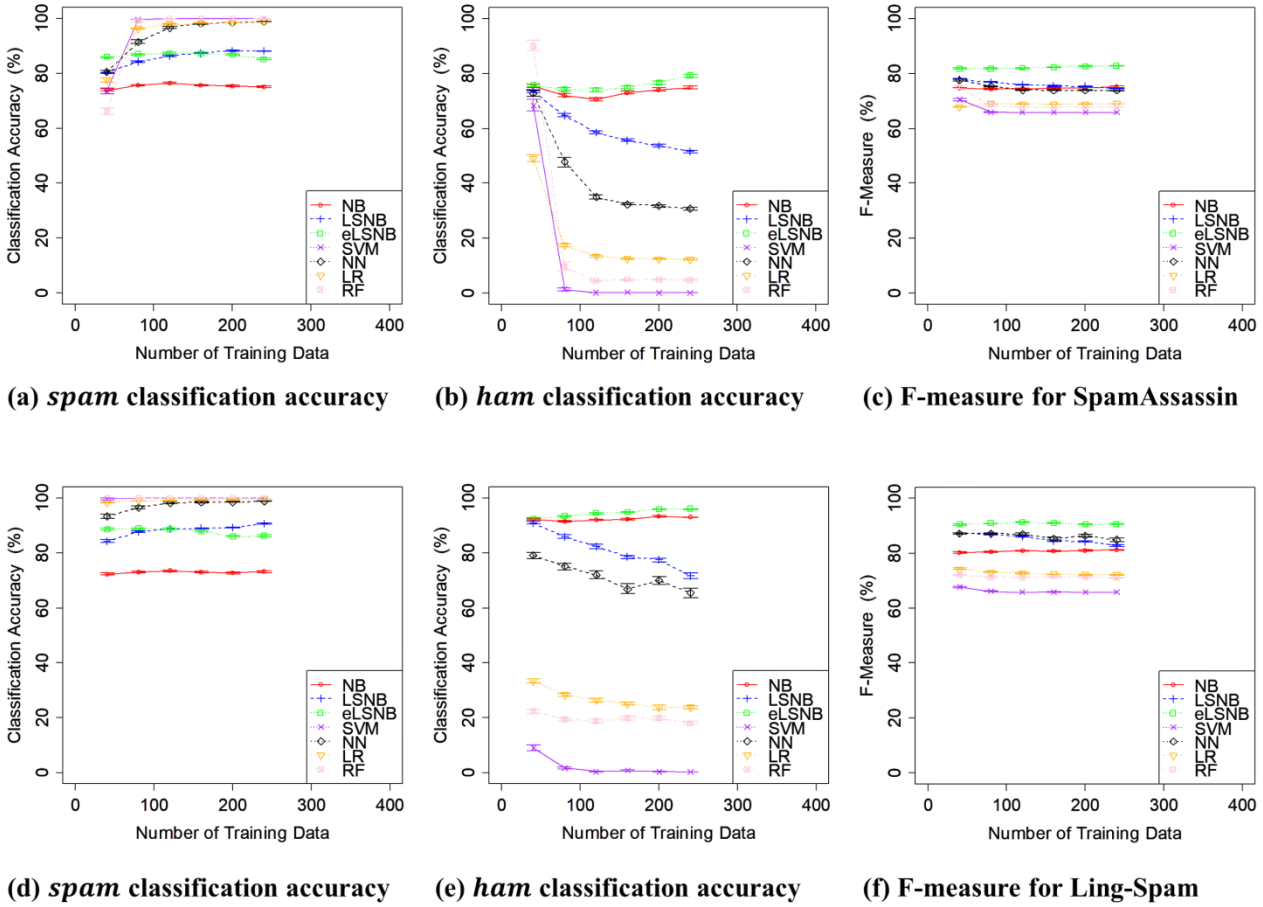
Figures 11. The results of Exp. 2-1. Error bars indicate the standard error.



Figures 12. The results of Exp. 2-2. Error bars indicate the standard error.



Figures 13. The results of Exp. 2-3. Error bars indicate the standard error.



Figures 14. The results of Exp. 2-4. Error bars indicate the standard error.

Subject: avail for review : phonology , semantics , dong , interpreting

the books listed below are in the linguist office and now available for review .

if you are interested in reviewing a book (or leading a discussion of the book) ;

please contact our book review editor , andrew carnie , at : carnie @ linguistlist . org please include in your request message a brief statement about your research interests , background , affiliation and other information that might be valuable to help us select a suitable reviewer .

please do not simply provide a url for an electronic cv or web page .

these will be ignored .

phonology pier marco bertinetto , livio gaeta , georgijetchev & david michaels (eds) , certamen phonologicum iii .

papers from the third cortona phonology meeting , april 1996 .

torino , rosenberg & sellier 1997 , pp . 291 , price lit . 63 . 000 (approximately us $ 37) .

isbn 88-7011 - 717 - 0 .

semantics pier marco bertinetto , il dominio tempo-aspettuale . demarcazioni , intersezioni , contrasti . torino , rosenberg & sellier 1997 , pp 252 , price lit 48 . 000 isbn 887011726x (approximately us $ 25 . 50) contents - introduzione i .

demarcazioni - aspect vs . actionality - statives , progressives , habituals - the progressive as a ' partialization ' operator ii .

intersezioni - neutralizations and interactions in temporal-aspectual categories - metafore tempo-aspettuali - l ' interazione tra azionalita e aspetto nella perifrasi ' continua ' iii .

contrasti - le strutture tempo-aspettuali dell ' italiano e dell ' inglese - le perifrasi abituali in italiano e in inglese - l ' espressione della ' progressivita / continuita ' : un confronto tripolare (editor 's note : the reviewer of this book must be fluent in both italian and english) .

dong language long yaohong and zheng guoqiao (translated by d . n . geary) (1998) the dong langauge in guizhou province china . sil / u texas austin . interpreting / translation harris , brian (compiler) (1997) translation and interpreting schools .

language international world directory .

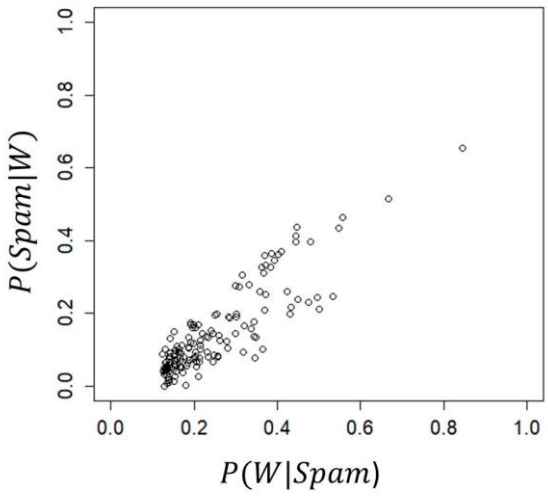
john benjamins : amsterdam .

Figure 15. 既存の機械学習は分類に失敗し，提案手法は正しく判別できた例

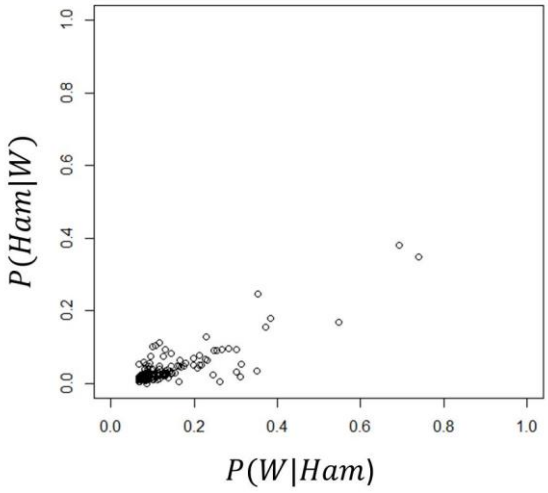
Subject: scitech international , inc . - your resource for scientific , engineering and technical software

hello , welcome to scitech international , your source for more than 2 , 000 of the most popular and some of the most obscure tools for scientific , engineering , education , and technical computing . if you wish to be removed from this list , please let us know immediately . we would be happy to do so . whether you ' re interested in astronomy or zoology , we can almost guarantee that you ' ll find at least one tool that will satisfy your needs . please check out our web site : http : / / www . scitechint . com you can browse our web site and do full text keyword searches of all the information in our product database . some products can be purchased online . we have over 100 different downloadable demos on our site to help you to make the right choice for your applications . if you can't find a product - or if you find the choice overwhelming - be sure to drop us an e-mail to info @ scitechint . com , or pick up the phone (7 : 00 am to 6 : 00 pm u . s . central time) and call our technical sales department for help at 1 . 773 . 486 . 9191 (international) or 1 . 888 . 462 . 6232 (domestic) . we ' re not trying to win any awards for the most fun or exciting web site . what we are trying to do is to build the ultimate database information about the best scientific and engineering software products available , and keep it up-to - date . if you are a scientist , engineer , or educator , be sure to bookmark the scitech internet catalog today . one final word about scitech . we do n't write software or manufacture hardware - service is our product . our goal , plain and simple , is to get you the right product , quickly , at the best price . best regards , christian staudinger senior product manager scitech international , inc . phone : 773-486 9191 / ext . 252 web : http : / / www . scitechint . com

Figure 16. 提案手法が分類に失敗し，既存の機械学習では正しく判別できた例

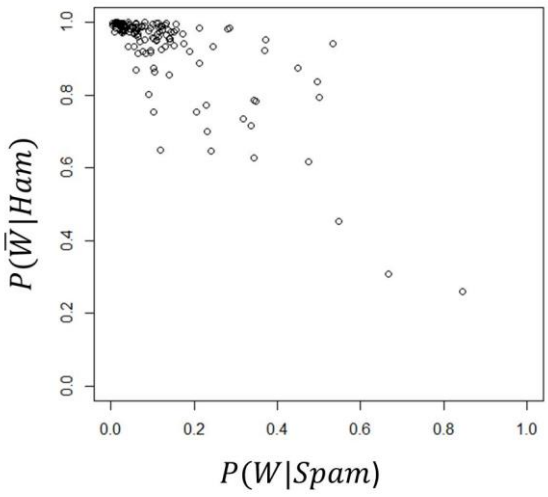


(a) Relation between $P(W|Spam)$ and $P(Spam|W)$

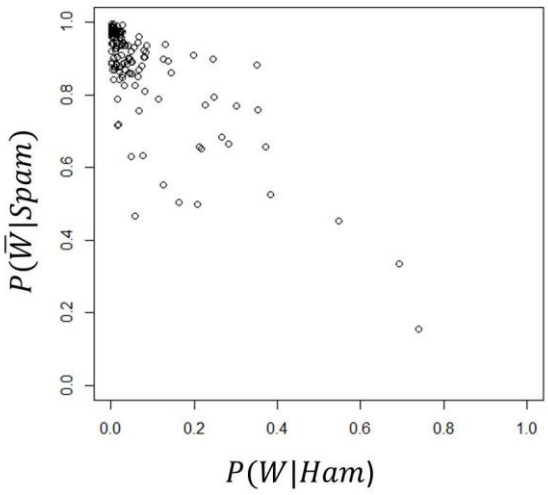


(b) Relation between $P(W|Ham)$ and $P(Ham|W)$

Figures 17. Symmetric relationships represented in an email corpus.



(a) Relation between $P(W|Spam)$ and $P(\bar{W}|Ham)$



(b) Relation between $P(W|Ham)$ and $P(\bar{W}|Spam)$

Figures 18. Mutually exclusive relationships represented in an email corpus.

第 3 章　医療データ分類タスクにおけるニューラルネットワークへの認知バイアスの適用

本章では，対称性バイアスと相互排他性バイアスを導入した NN の提案を行い，認知バイアスの観点からヘッブの法則 [Hebb 1949] を再現する．NN はその発明以来，様々なタスクに適用され，多くの派生モデルが開発された [LeCun et al. 2015]．その一例に，Dropout [Hinton et al. 2012, Srivastava et al. 2014] や Batch Normalization [Ioffe & Szegedy 2015] があり，これ等の手法は手書き文字や文章の分類タスクにおいて優れた性能を見せてきた．しかしながら，これ等の NN から派生した手法は，動物の脳の働きや，ヘッブの法則に，どの程度近しいものなのかについて，議論の余地がある．例えば，Dropout は生物学のアイデアに触発されており [Goodfellow et al. 2014]，生物は遺伝子の取捨選択を行うことで，生存環境に対して過剰となりすぎないように適応進化するとしている．しかしながら，こうしたシステムが人間あるいは生物の脳内に存在するかについては議論の余地がある．また，NN の派生手法により性能向上と，生物の脳へと近似させる試みには隔たりがあると言え，NN の手法の多くはむしろ，動物の学習とはまったく異なる構造を持つのではないかという指摘がある [Lake et al. 2015a, Lake et al. 2015b]．そこで，本章では動物の脳およびヘッブの法則の再現に主眼を置きつつ，同時により優れた学習能力を持つ NN の提案を行い，Dropout [Hinton et al. 2012, Srivastava et al. 2014, Hinton et al. 2015] および Batch Normalization [Ioffe & Szegedy 2015] を適用した NN との性能比較を行った．本章では，NN の評価に度々用いられる，医療データの分類タスクにおける実験結果，並びにその考察を通じて，提案手法の妥当性を示す．

3.1　はじめに

NN は機械学習モデルの中でも特に注目される手法であり，その原型は形式ニューロン [McCulloch & Pitts 1943] やパーセプトロン [Rosenblatt 1958] に遡る．

パーセプトロンの発明以来，その学習能力の向上のため多数の手法が考案された [Hopfield 1982, Ackley et al. 1985, LeCun et al. 1998, Werbos 1975, Hinton et al. 2006, Hinton & Salakhutdinov 2006, LeCun et al. 2015]．これ等のパーセプトロンから派生した手法は，病気の分類など様々なタスクにおいて優れた成果を挙げてきた [Marcano-Cedeno et al. 2011, Han et al. 2017]．しかしながら，上記の手法の多くは，学習時に多量かつバランスの良い教師データを必要とし [Mitchell 1997]，データが少量しか存在しない場合，性能が低下する [Japkowicz & Stephen 2002]．NN およびその派生モデルを扱うためには，金銭的，時間的コストがかかるのが現状である．こうした状況の解決のため，データが少量しかない状況でも高い性能を得られる NN の手法として，Dropout [Hinton et al. 2012, Srivastava et al. 2014, Hinton et al. 2015] が考案された．Dropout は，ネットワーク内のノードを一定の確率で無視する手法である．Dropout は，NN の overfitting および underfitting を防ぐことが可能であるとして，これまで一定の成果を挙げてきた [Goodfellow et al. 2016]．また，この手法は教師データの数が比較的少量の場合においても， NN の性能を向上させることができた．しかしながら，Dropout は教師データの数が極少量である場合，その有効性が低くなることが指摘されている [Srivastava et al. 2014, Goodfellow et al. 2016]．また，Dropout はバランスの良い教師データが十分に存在する場合，必ずしも性能向上に貢献する訳ではなく [Swietojanski et al. 2014]，むしろ性能が低下してしまう場合がある [Gal & Ghahramani 2016]．こうしたことから，Dropout は比較的少量のデータからの学習には貢献するが，データが十分である場合や，極少量である場合，その挙動はむしろ不安定であると言える．

本章では，NN に認知バイアスを導入することで，教師データが極少量である場合や不均衡である場合においても，高い性能を持つフレームワークを提案する．第 2 章で述べたように，認知バイアスを実装した機械学習モデルは，少量かつ不均衡な教師データからより優れた学習を行うと考えられる．NN に LS を導入した提案手法は，Dropout と同様ノードを”無視”することが可能であり，更に学習状況に応じて，それらを”復活”をさせることができる．この手法は，ヘッブの法則

[Hebb 1949] を認知バイアスの観点から，より最適に NN で再現する試みである．ヘッブの法則において，あるニューロンがそれに隣接するニューロンを繰り返し発火させた場合，前者のニューロンの軸索の小頭が生成・発達する．また，これ等のニューロンの間で一定の期間発火が発生しなかった場合，軸索の繋がりは弱まるとされている．このため，ヘッブの法則は

(i) “もしもニューロン x がニューロン y を発火させたら，ニューロン x のニューロン y への繋がりが生成・発達する”

(ii) “もしもニューロン x がニューロン y を発火させなかったら，ニューロン x のニューロン y への繋がりは弱まる”

という２つの事象として解釈できる．これ等の事象のうち (i) は対称性バイアスとして表現でき，(ii) は相互排他性バイアスとして表現できる．原因 p を“ニューロン x が強い信号を送った”，結果 q を“ニューロン y が発火した”と見なすとする．この時，対称性バイアスは“もしもニューロン x が強い信号を送ったならば (p)，ニューロン y が発火する (q)” という事象から “もしもニューロン y が発火したならば (q)，ニューロン x が強い信号を送った (p)” を対称的に導くことができる．また，相互排他性バイアスは “もしもニューロン x が強い信号を送らなかったならば，(p)，ニューロン y は発火しない (q)” を相互排他的に導く．ヘッブの法則は，NN のフレームワーク内で既に再現されているものの，上記の推論形式はよりヘッブの法則の概念に近しいものと考える．提案手法は対称性バイアスと相互排他性バイアスを用いてヘッブの法則の再現を行い，この仕組みがより優れた概念学習に貢献すると仮定する．本章では，医療データの分類タスクを題材に，提案手法と NN, SVM, RF, Dropout, Batch Normalization の性能比較を行った．これ等の機械学習手法を比較対象に選んだ理由は，まず NN は提案手法のベースであり，両者の特性を比較するために用いた．SVM と RF は機械学習の代表的な手法であり，これも比較対象とした．また，提案手法は Dropout と強い関連性を持ち，この手法は教師データが少量の時，ネットワークにノイズを与えることで overfitting および underfitting を防ぎ，分類精度の低下を防ぐことができる．Batch Normalization は NN の判別性能を向上させる最も強力な手法のひと

つであり，提案手法と Batch Normalization には，学習時に各層の入力に特殊な処理を行うという，共通のプロセスが存在する．提案手法と既存の機械学習手法との性能比較に医療データを用いた理由は，このタスクが機械学習の分野において長年研究されてきたことと，他の多くのタスクと同様，高い判別性能を得るため，多量の教師データを必要とすることである [Weiss & Provost 2003]．しかしながら，乳がんの診察データなどの医療データは，プライバシー保護などの理由から多量のデータを手に入れることが難しい [Hrovat et al. 2014]．このため，少量の医療データから高い判別性能を得られる NN のフレームワークの実現は，非常に重要であると考えられる．本章では LS モデルを NN に適用し，少量かつ不均衡な教師データからも安定して学習を行う手法を提案する．

3.2 Dropout Neural Networks

Dropout は, NN の正則化を行う手法のひとつであり，少ない計算量で優れた成績を示す手法として注目されている [Hinton et al. 2012, Srivastava et al. 2014]. Dropout は，ニューラルネットワークの学習時に，中間層または入力層のノードを一時的に，一定の確率で無視する．

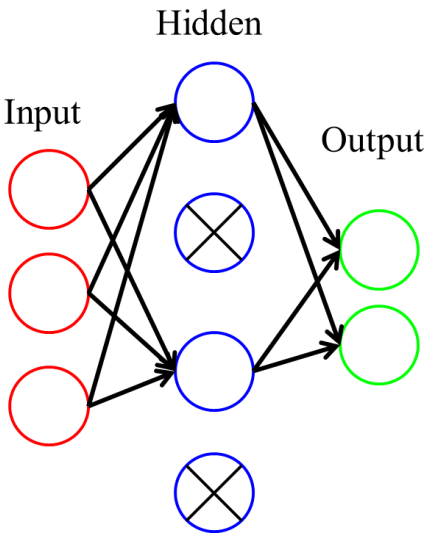


Figure 19. Dropout を適用した ニューラルネットワーク．

Fig. 19 に Dropout を適用した NN の図を示す. Dropout したノードの持つ値, およびそのノードからの結合荷重は 0 となり, フォワード学習にもバックプロパゲーションにも参加しない. Dropout するニューロンの選択は, 学習毎に行われる. つまり, Dropout を用いたニューラルネットワークは, 毎回異なる形状のネットワークを使い学習を行う [Hinton et al. 2012, Srivastava et al. 2014]. このことから Dropout は, 複数のモデルを組み合わせることで, 汎化誤差を抑える手法であるバギング (Bagging) [Breiman 1996] と似た手法であると考えられる. バギングは, 学習の際に複数のモデルを個別に学習し, テストの際には各モデルの多数決によって判別を行う. Dropout は, バギングを NN のフレームワークに導入した手法と言え, 学習時にはネットワーク内のノードをランダムに脱落させ, 毎回異なるネットワーク構造を用いて学習を行う [Goodfellow et al. 2014]. この仕組みにより, データに対して NN が過適合することを防ぐ [Dahl et al. 2013]. また, Dropout は教師データの数が限られている時に, 特に有用に働くとされている [Hinton et al. 2012, Dahl et al. 2013, Srivastava et al. 2014].

3.3 Batch Normalization

バッチ正則化 (Batch Normalization, BN) は NN の学習を促進する手法であり, 各層の入力を式 (49) のように正則化する [Ioffe & Szegedy 2015].

$$\widehat{x^{(k)}} = \frac{x^{(k)} - \mathrm{E}\left(x^{(k)}\right)}{\sqrt{\mathrm{Var}[x^{(k)}]}} \tag{49}$$

ここで, $\widehat{x^{(k)}}$ は各層への入力であり, $\mathrm{E}\left(x^{(k)}\right)$ と $\sqrt{\mathrm{Var}[x^{(k)}]}$ はそれぞれミニバッチにおけるノードの平均値と標準偏差である [Gitman & Ginsburg 2017]. ミニバッチ学習とは, 教師データからサンプリングされた一部のデータを用い, 結合荷重の更新を行う手法である [Goodfellow et al. 2014]. BN の目的のひとつは internal covariate shift [Ioffe & Szegedy 2015] と呼ばれる現象を避けることにあり,

この現象は，ネットワークの活性の分布に，パラメータ更新が影響を与えることを指す．例えば，クラス y がある時，分類器は $P(y|x)$ を学習する．もしも結合荷重やバイアス値の更新によって $P(x)$ が頻繁に変更される場合，$P(y|x)$ の推定は難化し，ネットワークの学習が遅くなる [Arpit et al. 2016]．この現象を防ぐため，BN は各層の入力の平均と標準偏差をもとに正則化し，ネットワークの重みの初期値に依存せず，また Dropout の必要性を下げることができるとしている [Ioffe & Szegedy 2015].

3.4 Loosely Symmetric Neural Networks

本研究では篠原らの LS モデルを NN に適用し，Loosely Symmetric Neural Networks (LSNN) を開発した．提案モデルである LSNN モデルでは，各層のノードの出力値を LS モデルを使い調整を行う．LSNN は LSNB と強い関連性を持ち, LSNB が尤度の計算に LS を利用するように, LSNN はノードの取る値を LS を使って調整する．提案手法は，この動作をフォワード学習とバックプロパゲーションの両方，あるいはそのいずれかにおいて実行できる．この仕組みは，動物のニューロンから観測される特性としての対称性 [Reigl et al. 2004]，相互排他性 [Sommer & Wurtz 2000] に着目している．人間の認知において観測されるバイアスが，ニューロン単位でも観測されることに着目している．ニューロンにおける対称性とは，”ニューロン y が発火し，それに隣接するニューロン x も発火する” という傾向性のことである．また，ニューロンにおける相互排他性とは ” ニューロン x が発火しなかったならば，それに隣接するニューロン x もまた発火しない” という傾向性のことである．これ等の関係は “もしもニューロン x が発火したならば，それに隣接するニューロン y も発火する” という事象から，対称性バイアスと相互排他性バイアスを用いて導くことができる．本研究で開発した LSNN は，ニューロン間で観測されるこれらの生理的な因果関係を LS モデルを基に再現する．LSNN は対称性バイアスと相互排他性バイアスに基づき，式 (50)-(54) のようにノードの出力値の調整を行う．

$$a = y_i^{k-1} \tag{50}$$

$$b = 1 - y_i^{k-1} \tag{51}$$

$$c = 1 - x_j^k \tag{52}$$

$$d = x_j^k \tag{53}$$

$$LS(y_i^{k-1}) = \frac{a + \frac{bd}{b+d}}{a + b + \frac{ac}{a+c} + \frac{bd}{b+d}} \tag{54}$$

ここで, a は $k-1$ 層目のノード y_i^{k-1} が取る値であり, あるノードの活性具合を表す指標である. b はそのノードの非活性化具合を表す指標である. また, c, d は k 層目のノード x_j^k の非活性化具合, および活性化具合である. この時, LSNN はノード y_i^{k-1} とノード x_j^k の間で因果推論を行う. もしも, ノード y_i^{k-1} がノード x_j^k に信号を送り, x_j^k が活性化した場合, LSNN は y_i^{k-1} は活性化に「貢献した」と推定し, $LS(y_i^{k-1})$ は y_i^{k-1} よりも大きい値を出力する. 一方, もしもノード y_i^{k-1} がノード x_j^k に信号を送り, x_j^k が活性化しなかった場合, LSNN は y_i^{k-1} を「弱い信号を出力するニューロン」だと推定し, $LS(y_i^{k-1})$ は y_i^{k-1} よりも小さい値を出力する. また, 結合荷重の更新式は式 (55) のように変更される.

$$\Delta_{LS} w_{ij}^{k-1,k} = -\epsilon \delta_j^k y_j^k \left(1 - y_j^k\right) LS(y_i^{k-1}) \tag{55}$$

ノード y_i^{k-1} とノード x_j^k 間の結合荷重の変化量 $\Delta_{LS} w_{ij}^{k-1,k}$ は LS モデルに基づき変更される. この更新を行う手順として

(i) フォワード学習を行い k 層目のノードの値を求める.

(ii) LS モデルを基に $k-1$ 層目のノードの持つ値の調整を行う.

(iii) バックプロパゲーションの際に (ii) で修正したノードを用いる.

という流れを取る．この更新式の特徴のひとつは，ノード y_i^{k-1} が 0 を取る場合でも，x_j^k の状態によっては結合荷重が更新されることである．1 章で示した式 (28) の一般的なフィードフォワード NN のバックプロパゲーションでは，y_i^{k-1} の取る値が 0 の場合，その結合荷重の変化量 $\Delta_{LS} w_{ij}^{k-1,k}$ も 0 を取る．一方で LSNN は，仮に y_i^{k-1} が 0 を取る場合でも，次の層のノード y_j^k が発火した場合，y_i^{k-1} が発火に「貢献した」と考え，結合荷重の更新を行う．Dropout と LSNN の大きな違いは，Dropout がランダムにノードとその結合荷重を 0 に置き換えるのに対し，LSNN は層と層の状態から，対称性バイアスと相互排他性バイアスを基にノードの調整を行うことである．LSNN は Dropout 的な挙動も取れば，逆に Dropout したノードの復活も行い，より学習状況に応じた調整を行う．

3.5 少量の教師データによる医療データの分類実験

3.5.1 実験設定

NN, SVM, RF, Dropout を用いた NN (Drop-NN), Batch Normalization を用いた NN (NN-BN) と提案モデル LSNN の計 6 モデルを用い，乳がんの良性 (Benign) ・悪性 (Malignant) の分類の結果を比較した．実験データは医療データ分類タスクにたびたび用いられる Wisconsin Breast Cancer Database (*https://archive.ics.uci.edu/ml/datasets/breast+cancer+wisconsin+(original)*) を用いた．Wisconsin Breast Cancer Database は 458 サンプルの Benign (良性) データと 241 サンプルの Malignant (悪性) データの計 699 サンプルのデータを含み，それぞれデータベース全体の 65.5%, 34.5% を占める．また，各データの特徴数は 10 個であり，各データの ID ナンバーである Sample code number 以外の各特徴は，1~10 の整数を取る．今回の実験では未知データが含まれる 17 サンプルおよび特徴のひとつである Sample code number を事前に取り除いた．このため，実験に用いたデータの母集団は 682 サンプルであり，特徴数は 9 個である．Wisconsin

Breast Cancer Database の詳細情報を Table 4 に示す.

比較対象に用いた各アルゴリズムのパラメータ設定を以下に示す.

- NN, Drop-NN, NN-BN, LSNN の中間の層数は 1 層とし, 活性化関数はシグモイド関数とした. この理由として, 今回用いたデータセットの特徴は 9 個であり, また不均衡データを扱うため, 過剰な overfitting を防ぐため層数は 3 層とし,シグモイド関数は, 二値分類に頻繁に用いられるため利用した.
- また, 中間層のノード数は 30 個とした.
- Drop-NN は中間層のノードを 50% の確率で無視するよう設定した.
- NN-BN のミニバッチ数は, 各実験で使用した教師データ数の都合から, 実験 4 では 32, 実験 5 では 3, 実験 6,7 では 16 とした. これ等の数値はミニバッチ数を 3, 6, 10, 16, 32, 64 に設定した中で, 最も良い成績となったものを選んだ. また, 各 NN は 100 エポック学習を行った.
- SVM にはガウシアンカーネルを用い, コストパラメータは $C = 0.1, gamma = 0.1$ とした. RF はツリーの数を 5 とし, この値は 3, 5, 10, 30 に設定した中で, 最も良い成績となったものを選んだ. 各アルゴリズムのパラメータは複数回の試行のもと, 最良となるものを選んだ.
- 教師データの数は, 実験 4 において Benign=150, Malignant=150 とし, 実験 5 では Benign=6, Malignant=6, 実験 6 では Benign=150, Malignant=6, 実験 7 では Benign=6, Malignant=150 とし, 4 種類の実験を行い, 教師データのクラス毎の割合が分類に与える影響を調査した. つまり, 実験 4 では比較的多量かつバランスの良い教師データ, 実験 5 では少量かつバランスの良い教師データ, 実験 6,7 では不均衡な教師データを用いた.
- テストデータの数は, 各実験とも Benign=50, Malignant=50 の計 100 データである.
- 分類精度の指標には, 再現率 (recall) [Alpaydin 2014] を用いた. 各条件における試行回数は 50 回とし, その平均を成績とした. また, Malignant クラスを陽性, Benign クラスを陰性として扱った.

Table 4. Wisconsin Breast Cancer dataset の詳細

特徴	平均	標準偏差	平均 (benign)	標準偏差 (benign)	平均 (malignant)	標準偏差 (malignant)
Clump thickness	4.44	2.82	2.96	1.67	7.19	2.44
Uniformity of cell size	3.15	3.07	1.30	0.86	6.58	2.72
Uniformity of cell shape	3.22	2.99	1.41	0.96	6.56	2.57
Marginal adhesion	2.83	2.86	1.37	0.92	5.59	3.20
Single epithelial cell size	3.23	2.22	2.11	0.88	5.33	2.44
Bare nuclei	3.22	2.15	2.41	1.22	4.71	2.66
Bland chromatin	3.45	2.45	2.08	1.06	5.97	2.28
Normal nucleoli	2.87	3.05	1.26	0.95	5.86	3.35
Mitoses	1.60	1.73	1.07	0.51	2.60	2.56

3.5.2 実験結果と考察

本章の実験では，教師データの割合を４パターン設定し，各モデルの挙動の変化を比較した．実験 4,5,6,7 の４種の結果を, Table 5-8 に示す.

実験 4 では，実験４種のうち，最も多量かつバランスの良い教師データを用いた．この実験では, F-measure の結果において, LSNN, Drop-NN, NN-BN, SVM, NN, RF の順番となった．この時, NN-BN と SVM のF-masureの値は同じであった.

Table 5. The results of experiment 1.

	Malignant	Benign	F-measure
NN	0.928	0.952	0.939
SVM	0.994	0.929	0.963
RF	0.767	0.882	0.814
Drop-NN	0.979	0.953	0.966
NN-BN	0.955	0.972	0.963
LSNN	0.995	0.941	0.969

Table 6. The results of experiment 2.

	Malignant	Benign	F-measure
NN	0.823	0.912	0.861
SVM	0.797	0.970	0.872
RF	0.556	0.793	0.631
Drop-NN	0.820	0.947	0.876
NN-BN	0.808	0.960	0.868
LSNN	0.977	0.946	0.962

Table 7. The results of experiment 3.

	Malignant	Benign	F-measure
NN	0.363	0.993	0.530
SVM	0.527	0.994	0.688
RF	0.170	0.962	0.281
Drop-NN	0.820	0.949	0.877
NN-BN	0.312	0.998	0.459
LSNN	0.997	0.901	0.951

Table 8. The results of experiment 4.

	Malignant	Benign	F-measure
NN	0.998	0.827	0.919
SVM	1.0	0.473	0.791
RF	0.949	0.316	0.721
Drop-NN	0.996	0.678	0.859
NN-BN	1.0	0.511	0.813
LSNN	0.996	0.886	0.944

LSNN は Malignant 分類の結果においてほぼ 100 % に近い成績となり, Benign 分類におけるエラー率は約 5 % に留まり, 全モデル中最も高い F-measure の値を示した.

Drop-NN は最も高い Benign 分類の成績を示し, また高水準の Malignant 分類の成績を示した. Drop-NN の F-measure の結果は全モデル中 2 番目となった.

NN-BN は Drop-NN 同様, 高水準の Benign 分類, Malignant 分類の成績を示し, F-measure の結果は全モデル中 3 番目となった.

NN は非常に高い Benign 分類の成績を示し, そのエラー率は 5% 未満となったものの, Malignant 分類の成績と F-measure は他の NN モデルと比べ低水準となった.

また, SVM は LSNN と似た水準の成績を示し, Benign 分類と Malignant 分類の両方において高い成績を見せ, NN-BN と同じく 3 番目に高い F-measure の値を示した.

RF の Benign 分類精度は 88%, Malignant 分類精度は 77%, 81% の F-measure のスコアとなったが, これ等の値は他のモデルと比べて低水準となった.

LSNN は最良の Malignant 分類精度を示し, また Benign 分類精度は実験 4 と同様, 約 5% のエラー率に留まった.

実験 5 では, 実験 4 と比べ大幅に少ない教師データを用いたものの, LSNN では大幅な成績低下は見受けられなかった. LSNN は極めて少量の教師データか

ら，より適切に結合荷重の更新を行い，最も高い F-measure の値を示したと言える．

Drop-NN は LSNN と同水準の Benign 分類精度を見せたものの，Malignant 分類精度は実験 4 の結果と比べ 15% 低下した．Drop-NN の F-measure のスコアは実験 4 のスコアに比べ減少した．Drop-NN はネットワークのサイズを小さく調整することにより，少量データからの学習により適した構造へと変化したと考えられる．

NN-BN はDrop-NN と似た水準の結果および傾向を見せ，その Malignant 分類精度は実験 4 と比べて 15% 減少した．Drop-NN と NN- BN は，教師データの制限により，似た傾向の影響を受けた．[Loffe & Szegedy 2015] で示唆されたように，少量データからの正則化に成功したと考えられる．[Gitman & Ginsburg 2006] は Batch Normalization はバッチ数が少量の場合，利用に適していないとするが，実験 5 においてバッチ数 3 の NN-BN は優れた成績を示した．

また，NN は Drop-NN と似た成績を示したが，その Benign 分類精度は後者を下回るものであった．NN は実験 5 において，実験 4 と比べて Benign 分類精度が 0.04 減少し，Malignant 分類精度は 0.1，F-measure のスコアは 0.08 減少した．Drop-NN と NN の性能差は実験 5 において，実験 4 のそれよりも小さくなった．このことから，Dropout の効果は実験 5 において小さくなり，これは教師データ量の減少に起因すると考えられる．

SVM は他の 5 つの機械学習モデルとは異なる傾向を示し，Benign 分類精度が実験 4 のものと比べ向上したものの，Malignant 分類精度は大幅に減少した．RF は Benign 分類精度，Malignant 分類精度，F-measure スコアの全てが実験 4 のものよりも減少した．

RF は Benign 分類精度，Malignant 分類精度，F-measure の全てが実験 4 のものよりも減少した．

SVM と RF の成績の減少は，NN およびその派生モデルにおける成績の減少よりも大きいものであった．

NN，Drop-NN，NN-BN，LSNN は，SVM や RF ほどの成績の低下を見せなかっ

た. NN およびその派生モデルは, 少量データからより安定した学習を行ったと考えられ, SVM や RF よりも優れた成績を示した.

提案手法である LSNN はネットワークのサイズを調整する仕組みを有しており, ノードを無視すること, およびノードを復活させることが可能である. この仕組みが少量データからの優れた学習に貢献し, 特徴分布に適したネットワーク構造を構築できたと考えられる.

実験 6 において, LSNN を除く全機械学習モデルが, 他の 3 つの実験と比較して高い Benign 分類精度を見せた. 一方で LSNN の Malignant 分類精度は全モデル中最良であり, この数値は 4 つの実験中, 最も高い数値となった.

また, Drop-NN は実験 4 の結果と比較しても大きな性能の減少は見受けられず, 実験 5 と同水準の成績となった.

NN-BN は実験 6 において最も高い Benign 分類精度を示したが, Malignant 分類精度は実験 4 と比べて 0.64 減少し, 実験 5 と比べて 0.50 と, 大幅な低下を見せた. NN-BN の F-measure の値もまた大幅な減少を見せ, 全モデル中 4 番目の成績となった.

NN は実験 4 と実験 5 と比べ, 大幅に Benign 分類精度が向上したが, 一方で Malignant 分類精度および F-measure は大きく低下した. NN と NN-BN は不均衡な教師データより得られた特徴分布から強い影響を受け, データのバランスに対する鋭敏性を見せた.

SVM も同様に, 実験 4 と実験 5 と比べ Malignant 分類精度が低下し, 一方で Benign 分類精度は 100 % に近く, 全モデル中最も高かった. しかしながら SVM の F-measure のスコアは, 実験 4 と実験 5 で示した成績と比べて低水準であった.

NN と同様, NN-BN, SVM, RF は実験 4 と実験 5 と比べ, Benign 分類精度が大幅に向上したものの, Malignant 分類精度は低迷した. また, RF の F-measure のスコアは全モデル中最も低いものとなった. 実験 6 の結果において, 機械学習の代表的な手法である NN, SVM, RF は実験 4, 実験 5 の結果と比較して Benign 分類精度を向上させた一方, Malignant 分類精度を低下させた. NN-BN も

また不均衡データに対する強い鋭敏性を見せ，通常の NN を下回る成績となった．また，Drop-NN には大きな性能低下は見受けられなかったものの，やはり実験 4 および実験 5 の結果と比較すると，成績が低下した．一方で提案手法である LSNN には他のモデルのように大きな性能低下は見受けられず，最も高い F-measure のスコアを示した．

実験 6 では，教師データに偏りがある場合での性能比較を実施し，Malignant の教師データのみを極めて少量とした. NN, SVM, RF は非常に高い Benign 分類精度を見せた一方で, Malignant 判別成績は大幅に減少した. NN, SVM, RF はよりデータが多量に存在した Benign データの分類において高い性能を見せ，この結果はオーバーフィッティング [Alpaydin 2014] によるものだと考えられる．

NN-BN もまた，上記の機械学習の代表的な手法と似た傾向を見せ，BN は不均衡データを用いた状況においては有用性が薄らぐと考えられる．NN-BN の性能は各ミニバッチにおける各層のノードの入力値に依存する [Salimans et al. 2016]. このことから，NN-BN は入力データがスパースだったことによる $\mathrm{E}\left(x^{(k)}\right)$ と $\sqrt{\mathrm{Var}[x^{(k)}]}$ の不規則な推移により，学習中に多数の internal covariate shift が発生し，ノードの入力値の正規化に失敗したと考えられる．

Drop-NN と LSNN には Malignant 判別性能と Benign 判別性能との間で大きな差は見受けられず, Drop-NN はわずかに Malignant 判別成績が低下したものの, 両モデルとも高い性能を示した．Dropout は学習毎に一定の確率でノードを脱落させ，毎回異なるネットワークを用いて学習を進める．一方で LSNN はネットワークの学習状況を基に，ノードの脱落と復活を行う．LSNN は, Dropout よりもさらに柔軟なネットワーク構造の構築を行い，より優れた成績を示した．

実験 7 において，ほとんどの機械学習手法が，実験 6 とは逆に高い水準の Malignant 分類精度と，低い水準の Benign 分類精度を示した．実験 6 と実験 7 の間で用いたデータの比率は対称的であり，実験 7 の結果において, LSNN を除く全ての機械学習モデルの成績が実験 6 と対称的となった．

LSNN は Benign 分類精度が実験 4 と比べ 0.06 下がったが，この数値は他の機械学習モデルの成績と比べ，大幅に小さかった．また，LSNN は Malignant 分

類においてほぼ 100 % に近い成績を示した．実験 7 における LSNN の F-measure の値は非常に高い水準にあり，実験 4-6 の結果と比較しても大きな差が無い．

一方で，LSNN と同じく NN の派生系である Drop-NN は Benign 分類精度が大幅に低下し，4 つの実験の中で最も低い数値となった．また，Drop-NN の Benign 分類精度および F-measure は NN の成績よりもむしろ悪かった．

NN-BN と同様，Drop-NN もまた不均衡なデータ分布に対する鋭敏性を実験 7 の結果において見せた．NN-BN は 100 % の Malignant 分類精度を示したものの，Benign 分類精度は比較的低い水準に留まり，NN およびその派生モデルの中では最も低い数値となった．一方で NN は高い Malignant 判別性能，Benign 判別性能を持ち，Drop-NN や NN-BN よりも高い F-measure のスコアを示した．また，4 つの実験を通じて，NN は Drop-NN や NN-BN のような Benign 判別性能の低下が見受けられなかった．

SVM は Malignant 分類において 100% の精度を見せたが，Benign 分類の成績は大幅に低下した．また，実験 4-6 において，SVM の Benign 分類精度は 0.9 以上と非常に高かったものの，実験 7 においては低迷する結果となった．

RF も SVM, NN-BN, Drop-NN と同様の傾向を見せ，Malignant 分類精度が向上し，同時に Benign 分類精度が向上した．実験 6 における結果とは対称的に，NN, NN-BN, SVM, RF は Benign 分類精度を犠牲に，Malignant 分類精度を向上させる結果となった．この傾向は実験 4-6 で安定した成績を示していた，Drop-NN からも観測された．実験 6，実験 7 で観測されたように，NN，SVM，RF，NN-BN，Drop-NN は教師データが不均衡な場合においてはトレードオフが発生した．一方で LSNN は NN の派生モデルの中で唯一トレードオフが見受けられず，偏りのあるデータ分布からより安定した学習を行うことに成功した．

4 種の実験を通じて，LSNN は不均衡データに対する鋭敏性を見せず，また全ての実験において，最も高い F-measure のスコアを示した．

実験 7 では，実験 6 とは対称に Benign の教師データのみを極めて少量とした．Table 7, 8 が示すように，ほぼ全てのモデルが実験 6 と実験 7 との間で対称

的な成績を示した．つまり，多くのモデルにおいて Malignant 分類精度が向上し，一方で Benign 分類精度は低下した．Dropout は NN の不均衡データからの学習を促進する手法であるが，実験 7 において Benign 分類精度の低迷が見受けられた. Table 8 が示すように, Drop-NN の F-measure のスコアは通常の NN よりもむしろ低かった．この結果から，Dropout の不均衡データに対する有効性は，議論の余地があると考える．既存研究において，Dropout は手書き文字画像の分類タスクや，テキスト分類タスクにおいて有用性を示してきた [Dahl et al. 2013, Srivastava et al. 2014]．これ等のタスクにおける特徴ベクトルの次元数は 100 あるいは 1000 を超える．一方で，本章の実験で用いたデータセットから得られる特徴ベクトルの次元数は 9 個である．実験結果から，Dropout の有用性は，特徴ベクトルが非常に高次元の場合に，より強く発揮されると考察する．また, Table 4 に示したように, Malignant データにおける値の変動は Benign データにおけるものよりも大きく，これによりデータ特性を学習するのがより困難となった可能性がある．また，Dropout は少量の教師データからの学習の促進に貢献するとされている [Hinton et al. 2012, Sun et al. 2016] ものの, [Goodfellow et al. 2016] は Dropout は極少量の教師データ下では，その有用性は小さくなるとしている．また, [Zhang et al. 2015] の研究では Dropout は十分な量の教師データが与えられた場合，その有用性が低下すると報告している．Dropout は特定の条件下ではむしろ逆効果となる可能性があり，本章の実験 7 においてその現象が確認された．また，実験 6 と実験 7 の結果において，NN-BN もまた NN を下回る成績を見せた．特に実験 7 の結果において，NN-BN の Benign 分類精度は 0.4 以上減少し，NN および LSNB よりも大幅に低いものとなった．実験 6 の結果と同様，NN-BN は不均衡なデータ分布から強い影響を受け，学習時に多数の internal covariate shift [Loffe & Szegedy 2015] が発生したと考えられる. NN-BN もまた不均衡データに対する鋭敏性を見せ，その性能は通常の NN をむしろ下回る結果となった．一方で LSNN にはそのような鋭敏性は見受けられず, 4 つの実験全てにおいて最良の F-measure のスコアを示した．LSNN は少量かつ不均衡なデータからの学習という観点で，2 章における LSNB と近しい挙動を見せた．これ等の

提案手法は特徴ベクトルから得られた分布を LS を用いて調整し，少量かつ不均衡なデータから，高い成績を示した．

3.6 提案手法の有用性に関する考察

本節では，提案手法である LSNN の働き，および LSNN がどのような問題設定において，特に有用に働くかを考察する．３層以上の層から構成される NN は，線形分離不可能な問題設定に頻繁に用いられる．NN の研究で扱われる，線形分離不可能な問題設定の代表的なものに XOR (exclusive or) 問題がある．XOR は非常にシンプルな問題であるものの NN の学習の題材として，長年研究されてきた [Werbos 1975]．本節では，提案手法の線形分離不可能な問題における有用性を考察するため，XOR 問題における NN, Drop-NN, LSNN の働きを考察した．入力層のノード数は 2 個，中間層のノード数は 5 個，出力層のノード数は 1 個とし，活性化関数はシグモイド関数とした．また，ネットワークの大きさを考慮し，Drop-NN が中間層のノードを無視する確率は 10% とした．試行回数は 500 回とした．各モデルにおける結合荷重の初期値は，試行毎に同一のものを用いた．

Fig. 20 (a)-(c) に，各ノードの結合荷重の値を示す．Fig. 20 (a) は学習完了時の NN の結合荷重の値，Fig. 20 (b) は Drop-NN の結合荷重の値，Fig. 20 (c) は LSNN の結合荷重の値である．図中の棒グラフは各ネットワーク内の結合荷重の位置であり

- w_i1_h1 は入力層１ノード目から中間層１ノード目への結合荷重,
- w_i1_h2 は入力層１ノード目から中間層２ノード目への結合荷重,
- w_i1_h3 は入力層１ノード目から中間層３ノード目への結合荷重,
- w_i1_h4 は入力層１ノード目から中間層４ノード目への結合荷重,
- w_i1_h5 は入力層１ノード目から中間層５ノード目への結合荷重,
- w_i2_h1 は入力層２ノード目から中間層１ノード目への結合荷重,
- w_i2_h2 は入力層２ノード目から中間層２ノード目への結合荷重,
- w_i1_h3 は入力層２ノード目から中間層３ノード目への結合荷重,
- w_i1_h4 は入力層２ノード目から中間層４ノード目への結合荷重,

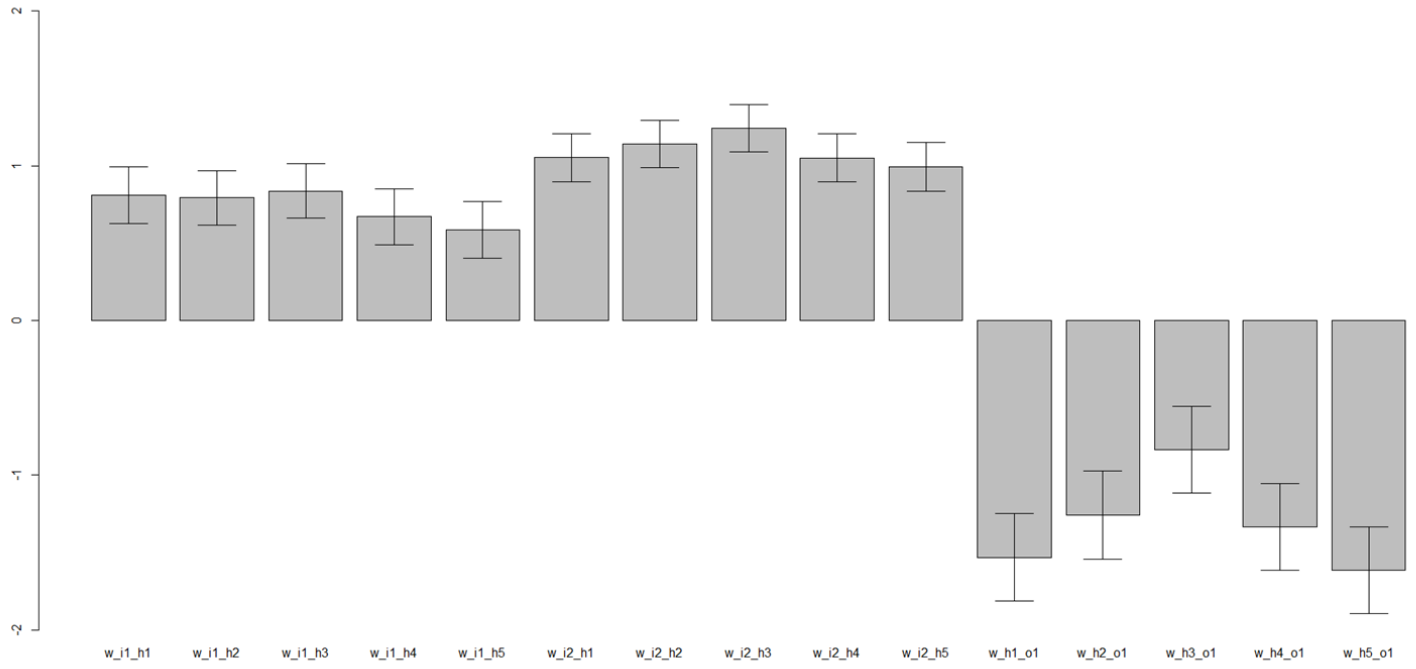
- w_i1_h5 は入力層２ノード目から中間層５ノード目への結合荷重,
- w_h1_o1 は中間層１ノード目から出力層１ノード目への結合荷重,
- w_h2_o1 は中間層２ノード目から出力層１ノード目への結合荷重,
- w_h3_o1 は中間層３ノード目から出力層１ノード目への結合荷重,
- w_h4_o1 は中間層４ノード目から出力層１ノード目への結合荷重,
- w_h5_o1 は中間層５ノード目から出力層１ノード目への結合荷重,

のように表される．また，各グラフのエラーバーは標準誤差である．

この実験において，全モデルが全試行で XOR の出力を再現した.また，Fig. 20 (a)-(c) の結果から, LSNN は NN および Drop-NN と比べ, 結合荷重の取る値の範囲が小さいことがわかった. LSNN の結合荷重は NN および Drop-NN よりも収束しやすいと言え，特に LS モデルを導入した中間層と出力層の間の結合荷重においてその傾向が顕著であった．また，LSNN はノードの値の調整によって，正則化に近い効果を，結合荷重に与えることがわかった．

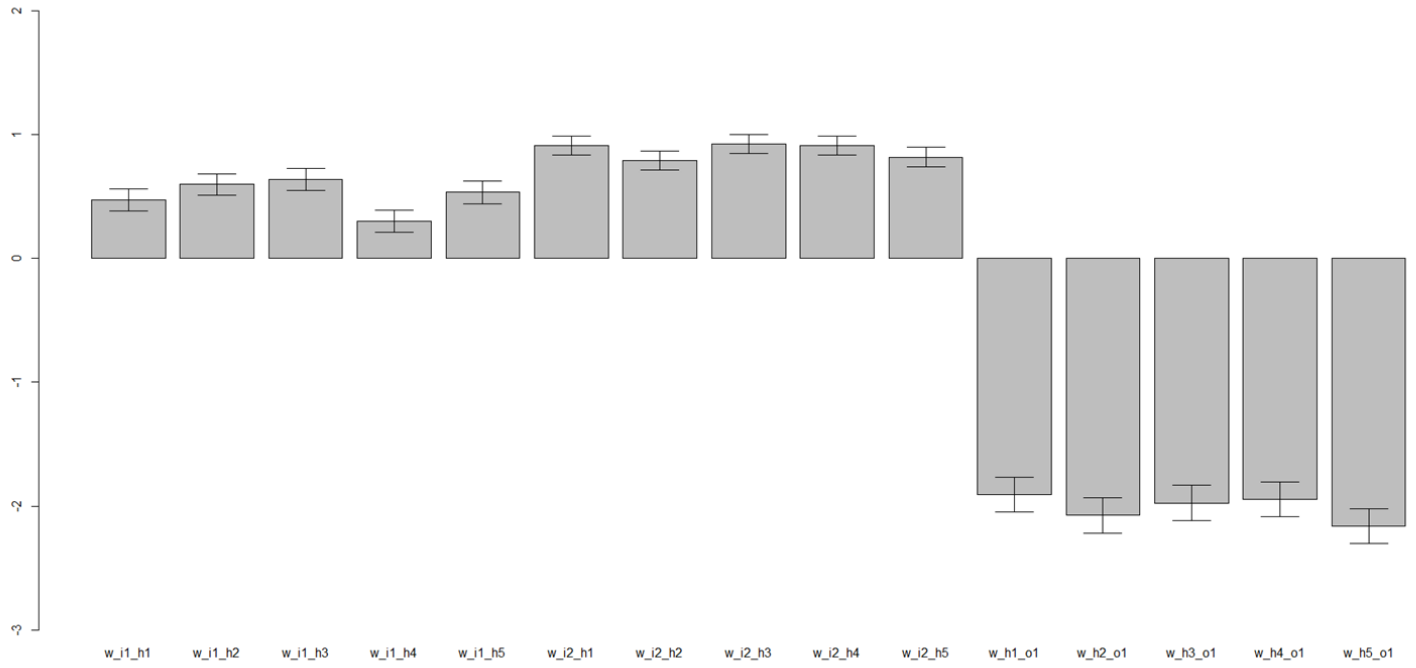
また, Drop-NN は NN と比べ, 入力層と中間層の間の結合荷重の取る値の範囲が小さかったものの，入力層と中間層の間の結合荷重が取る範囲と，中間層と出力層の間の結合荷重の取る範囲に，大きな差があった．このことから，Drop-NN は入力層と中間層の間の結合荷重の収束を助ける働きがあることが予想される．しかしながら，中間層と出力層の結合荷重が取る値が，入力層と中間層の間の結合荷重が取る値と比べ非常に大きい場合，荷重の更新量が結合荷重そのものの取る値を大幅に上回り，学習が失敗してしまうリスクがある．

一方で提案手法である LSNN は入力層と中間層，ならびに中間層と出力層の間で結合荷重が収束しやすく，より効果的に過学習を防ぐことがわかった．提案手法は線形分離不可能な問題設定において，より安定した学習を行うことがわかった．



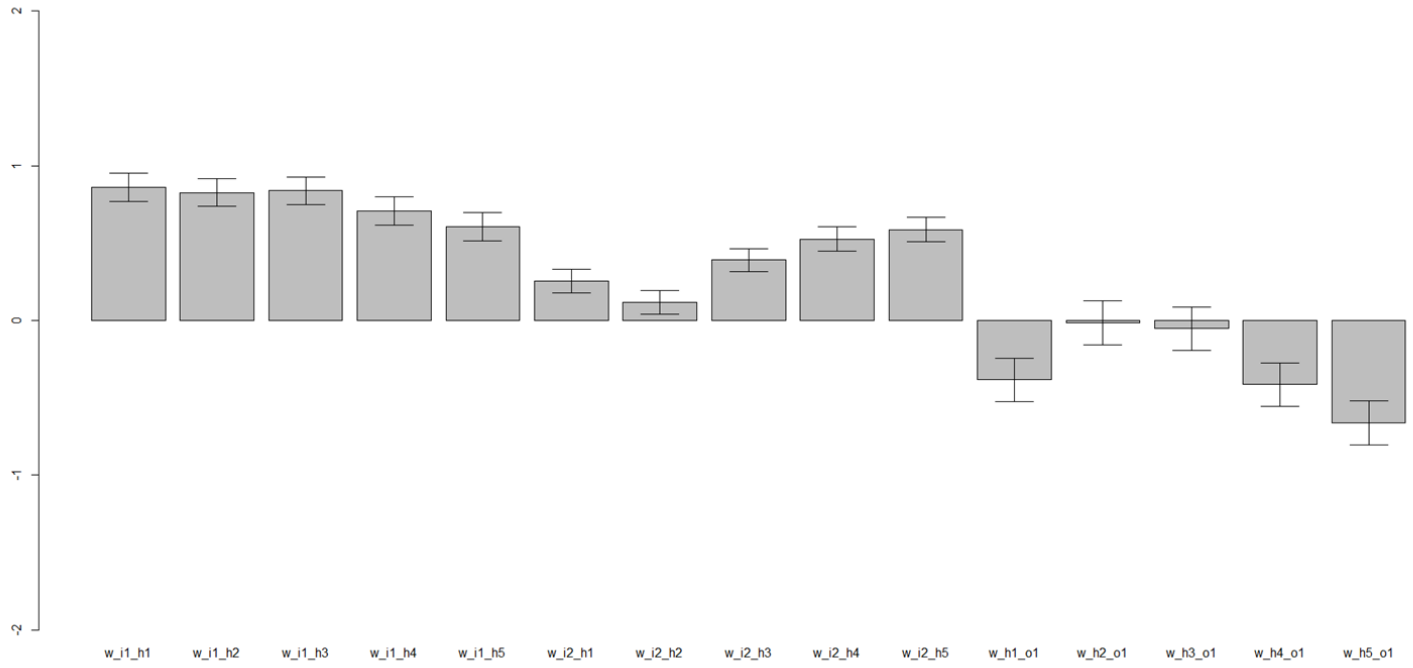
(a) Weights of NN after Training

Figures 20 (a). Weights of NN after training. The error bars indicate the standard errors.



(b) Weights of Drop-NN after Training

Figures 20 (b). Weights of Drop-NN after training. The error bars indicate the standard errors.



(c) Weights of LSNN after Training

Figures 20 (c). Weights of LSNN after training. The error bars indicate the standard errors.

3.7 第 3 章のまとめ

本章では NN に認知バイアスを導入した LSNN を提案し，医療データ分類タスクを題材に，NN, SVM, RF, Drop-NN, NN-BN との比較を行った．教師データの割合を比較的多量かつバランス良く設定した実験 4，少量かつバランスの良い教師データを用いた実験 5，教師データを不均衡とした実験 6,7 を行い，各モデルの挙動の変化を考察した．

提案手法では，ニューロン間で観測される対称性・相互排他性に着目し，ヘッブの法則の認知バイアスの観点からの再現を試みた．LSNN はノードの脱落を行うこともあれば，逆にノードの復活も行い，より学習状況に応じた調整を行うことが可能である．

実験 4 において，全モデルが高い分類精度を見せた．中でも，NN の派生系である LSNN, Drop-NN, NN-BN は最も高い水準の F-measure を示した．

実験 5 において，多くのモデルが教師データ数の減少による影響を受け，実験 4 に比べ分類精度が低下した．一方，提案手法である LSNN からは大きな精度の減少は観測されなかった．

実験 6 の教師データを不均衡とし，比較的多量の Benign 教師データと少量の Malignant 教師データを用いた実験では，LSNN と Drop-NN を除く全てのモデルがトレードオフを見せた．

実験 7 の，実験 6 とは教師データの割合を逆転させた実験では，Drop-NN を含め，全ての既存の機械学習モデルがトレードオフを見せた．一方，LSNN からは大きな精度の減少は見受けられず，不均衡データから安定した学習を行った．

また，LSNN はノードが取る値を調整することで，正則化に近い効果を結合荷重にもたらすことがわかった．提案手法は Dropout と比べ，より効果的に過学習を防ぐことが示唆された．

3 章のまとめとして，人間の認知バイアスを導入し，ニューロンにおける対称性と相互排他性を再現した LSNN が，全ての実験において最も優れた成績を示した．提案手法は機械学習の不均衡データに対する問題点を克服し，人間の優れ

た学習能力の再現を行った.

第 4 章　結論

本研究では機械学習モデルに人間の認知バイアスを実装し，スパムメール分類と病気の良性･悪性分類を題材に実験を行った．提案手法は，認知バイアスの中でも特に注目される対称性バイアスと相互排他性バイアスを導入判別することで，他の機械学習手法と比べ，大幅に高い分類精度を見せた．特に，提案手法は教師データが少量である場合や，不均衡な場合においても，安定した学習を行うことに成功した．このことから，人間の認知バイアスを利用した機械学習モデルは，教師データを確保することが難しいタスクにも，既存の手法と比べ，より効果的に利用できることを明らかにした．

従来の機械学習の代表的な手法は，多量かつバランスの良い教師データを用意することで，高い判別性能を得ることができる．一方で本研究では，教師データが少量かつ不均衡な状況に対して，人間の認知能力に着目した，より柔軟な学習を行うモデルを提案し，第 2 章では，スパムメール分類タスクを題材とし，ナイーブベイズ分類器に認知バイアスを適用し，NN, SVM, NB, LR, RF との比較を行った．教師データ数を少量に限定した実験 1 において，比較対象としたモデルの多くは判別性能が低迷したものの，提案手法は高い精度を見せた．また，特定のクラスに属する教師データの数を固定にした実験 2 において，提案手法からは判別成績の減少は見受けられなかった．特徴ベクトル内のゆらぎを考慮した実験 3 においても，提案手法は比較対象とした機械学習モデルと比べ，高い分類精度を示した．

第 3 章では医療データ分類を題材とし，このタスクにおいて頻繁に用いられる手法である NN に認知バイアスを適用した．提案手法は，ニューロン間で観測される生理的な因果関係を，認知バイアスの観点から再現したものであり，認知バイアスの観点からのヘッブの法則の再現を行う試みである．実験では SVM と RF, および NN とその派生系である Dropout と Batch Normalization との性能比較を行った．教師データのバランスが良く，比較的多量のデータを用いた実験 4 および，バランスは良いもののデータの総量が少量であった実験 5 において，提案手法は最も高い精度を見せた．また，教師データの数を不均衡にした実験 6,

実験 7 において，提案手法は NN の不均衡データに対する鋭敏性を克服した．データ量とそのバランスを変化させた４種の実験全てにおいて，提案手法は最も優れた成績を見せ，この仕組みがより優れた概念学習に貢献することを示した．本研究では，人間の認知バイアスを機械学習に適用することの有効性について議論した．第２章と第３章の実験結果から，認知バイアスを適用した機械学習モデルは，教師データが少量である場合や不均衡である場合においても，安定した学習を行うことが明らかになった．この結果から，認知バイアスを適用した機械学習モデルは，一定の条件下において，より人間に近い学習能力を持つことを示した．

今後の発展として，スパムメール分類や病気の良性･悪性分類の以外にも，他のデータセットへの適用，およびより多くの機械学習手法への認知バイアスの適用･調査を行う．一例として，医療データからの入院日数の推定といった回帰タスクや, Long Short Term Memory (LSTM) による時系列データの分析が考えられる．これ等の手法に認知バイアスならびに Loosely Symmetric モデルを適用することで，「手元の情報から未来に発生すること」を予測することが可能である．例えば対称性バイアスは「患者からある症状が観測されると，入院期間が長くなる」という事象から「入院期間が長引いたのは，ある症状が観測されたからだ」という推論形式を導き，相互排他性バイアスは「患者にある症状が発生しなかったから，入院期間が長引かなかった」という推論形式を導く．これ等の推論は 「手元の情報から状況判断を下す」という認知バイアスの形式に類似しており，その実現が期待できる．

付録 A: ストップワード

本研究では Table 11 に示す単語をストップワードと定義した.

Table 11. 実験に用いたストップワードの一覧

A	about	Above	across	after	again	against
All	almost	alone	along	already	also	although
always	am	among	an	and	another	any
anybody	anyone	anything	anywhere	are	area	areas
aren't	around	as	ask	asked	asking	asks
At	away	B	back	backed	backing	backs
Be	became	because	become	becomes	been	before
began	behind	being	beings	below	best	better
between	big	both	but	by	C	came
Can	cannot	can't	case	cases	certain	certainly
clear	clearly	come	could	couldn't	D	did
didn't	differ	different	differently	do	does	doesn't
doing	done	don't	down	downed	downing	downs
during	E	each	early	either	end	ended
ending	ends	enough	even	evenly	ever	every
everybody	everyone	everything	everywhere	F	face	faces
Fact	facts	far	felt	few	find	finds
First	for	four	from	full	fully	further
furthered	furthering	furthers	G	gave	general	generally
Get	gets	give	given	gives	go	going
good	goods	got	great	greater	greatest	group
grouped	grouping	groups	H	had	hadn't	has
hasn't	have	haven't	having	he	he'd	he'll
Her	here	here's	hers	herself	he's	high

higher	highest	him	himself	his	how	however
how's	I	i'd	If	i'll	i'm	important
In	interest	interested	interesting	interests	into	is
isn't	it	its	it's	itself	i've	J
Just	K	keep	keeps	kind	knew	know
known	knows	L	large	largely	last	later
latest	least	less	let	lets	let's	like
likely	long	longer	longest	M	made	make
making	man	many	may	me	member	members
Men	might	more	most	mostly	mr	mrs
much	must	mustn't	my	myself	N	necessary
need	needed	needing	needs	never	new	newer
newest	next	no	nobody	non	no one	nor
Not	nothing	now	nowhere	number	numbers	O
Of	off	often	old	older	oldest	on
once	one	only	open	opened	opening	opens
Or	order	ordered	ordering	orders	other	others
ought	our	ours	ourselves	out	over	own
P	part	parted	parting	parts	per	perhaps
place	places	point	pointed	pointing	points	possible
present	presented	presenting	presents	problem	problems	put
Puts	Q	quite	R	rather	really	right
room	rooms	S	said	same	saw	say
Says	second	seconds	see	seem	seemed	seeming
seems	sees	several	shall	shan't	she	she'd
she'll	she's	should	shouldn't	show	showed	showing
shows	side	sides	since	small	smaller	smallest
So	some	somebody	someone	something	somewhere	state
states	still	such	sure	T	take	taken
Than	that	that's	the	their	theirs	them

themselves	then	there	therefore	there's	these	they
they'd	they'll	they're	they've	thing	things	think
thinks	this	those	though	thought	thoughts	three
through	thus	to	today	together	too	took
toward	turn	turned	turning	turns	two	U
under	until	up	upon	us	use	used
Uses	V	very	W	want	wanted	wanting
wants	was	wasn't	way	ways	we	we'd
Well	we'll	wells	went	were	we're	weren't
we've	what	what's	when	when's	where	where's
whether	which	while	who	whole	whom	who's
whose	why	why's	will	with	within	without
won't	work	worked	working	works	would	wouldn't
X	Y	year	years	yes	yet	you
you'd	you'll	young	younger	youngest	your	you're
yours	yourself	yourselves	you've	Z		

付録 B: SpamAssassin コーパスに含まれる *spam* メールデータの例

```
From ilug-admin@linux.ie  Fri Aug 23 11:08:03 2002
Return-Path: <ilug-admin@linux.ie>
Delivered-To: zzzz@localhost.spamassassin.taint.org
Received: from localhost (localhost [127.0.0.1])
        by phobos.labs.spamassassin.taint.org (Postfix) with ESMTP id D843A4416F
        for <zzzz@localhost>; Fri, 23 Aug 2002 06:06:37 -0400 (EDT)
Received: from phobos [127.0.0.1]
        by localhost with IMAP (fetchmail-5.9.0)
        for zzzz@localhost (single-drop); Fri, 23 Aug 2002 11:06:37 +0100 (IST)
Received: from lugh.tuatha.org (root@lugh.tuatha.org [194.125.145.45]) by
    dogma.slashnull.org (8.11.6/8.11.6) with ESMTP id g7N7fDZ14920 for
    <zzzz-ilug@jmason.org>; Fri, 23 Aug 2002 08:41:13 +0100
Received: from lugh (root@localhost [127.0.0.1]) by lugh.tuatha.org
    (8.9.3/8.9.3) with ESMTP id IAA13857; Fri, 23 Aug 2002 08:38:52 +0100
X-Authentication-Warning: lugh.tuatha.org: Host root@localhost [127.0.0.1]
    claimed to be lugh
Received: from relay.dub-t3-1.nwcgroup.com
    (postfix@relay.dub-t3-1.nwcgroup.com [195.129.80.16]) by lugh.tuatha.org
    (8.9.3/8.9.3) with ESMTP id IAA13820 for <ilug@linux.ie>; Fri,
    23 Aug 2002 08:38:43 +0100
Received: from mail.com (unknown [64.86.155.148]) by
    relay.dub-t3-1.nwcgroup.com (Postfix) with SMTP id 1AE6470047 for
    <ilug@linux.ie>; Fri, 23 Aug 2002 08:38:15 +0100 (IST)
From: "MR.Johnson S. Abu" <coll2001ng@mail.com>
To: <ilug@linux.ie>
MIME-Version: 1.0
Content-Type: text/plain; charset="ISO-8859-1"
Date: Fri, 23 Aug 2002 20:41:55 +0100
Reply-To: "MR.Johnson S. Abu" <smith_j@mailsurf.com>
```

Message-Id: <20020823073815.1AE6470047@relay.dub-t3-1.nwcgroup.com>
Subject: [ILUG] BUSINESS
Sender: ilug-admin@linux.ie
Errors-To: ilug-admin@linux.ie
X-Mailman-Version: 1.1
Precedence: bulk
List-Id: Irish Linux Users' Group <ilug.linux.ie>
X-Beenthere: ilug@linux.ie
Content-Transfer-Encoding: 8bit

CENTRAL BANK OF NIGERIA
FOREIGN REMITTANCE DEPT.
TINUBU SQUARE, LAGOS NIGERIA
EMAIL-smith_j@mailsurf.com
23TH OF August 2002

ATTN:PRESIDENT/CEO

STRICTLY PRIVATE BUSINESS PROPOSAL

I am MR.Johnson S. Abu, the bills and exchange Director at the ForeignRemittance Department of the Central Bank of Nigeria. I am writingyou this letter to ask for your support and cooperation to carrying thisbusiness opportunity in my department. We discovered abandoned the sumof US$37,400,000.00 (Thirty seven million four hundred thousand unitedstates dollars) in an account that belong to one of our foreign customers,an American late Engr. John Creek (Junior) an oil merchant with the federal government of Nigeria who died along with his entire family of a wifeand two children in Kenya Airbus (A310-300) flight KQ430 in November2000.

Since we heard of his death, we have been expecting his next of kin tocome over

and put claims for his money as the heir, because we cannotrelease the fund from his account unless someone applies for claims asthe next of kin to the deceased as indicated in our banking guidelines. Unfortunately, neither their family member nor distant relative hasappeared to claim the said fund. Upon this discovery, I and other officialsin my department have agreed to make business with you release the totalamount into your account as the heir of the fund since no one came forit or discovered either maintained account with our bank, other wisethe fund will be returned to the bank treasury as unclaimed fund.

We have agreed that our ratio of sharing will be as stated thus: 30%for you as foreign partner and 70% for us the officials in my department.

Upon the successful completion of this transfer, my colleague and I willcome to your country and mind our share. It is from our 60% we intendto import computer accessories into my country as way of recycling thefund. To commence this transaction we require you to immediately indicateyour interest by calling me or sending me a fax immediately on the aboveTelefax # and enclose your private contact Telephone #, Fax #, full nameand address and your designated banking co-ordinates to enable us fileletter of claim to the appropriate department for necessary approvalsbefore the transfer can be made.

Note also, this transaction must be kept strictly confidential becauseof its nature.

NB: Please remember to give me your Phone and Fax No

MR.Johnson Smith Abu

--

Irish Linux Users' Group: ilug@linux.ie

http://www.linux.ie/mailman/listinfo/ilug for (un)subscription information.

List maintainer: listmaster@linux.ie

付録 C: Ling-Spam コーパスに含まれる *spam* メールデータの例

Subject: dear website operator

hi , i thought this could help your success . feel free to call me with any questions . sincerely , jennifer powers 904-441 - 8080 env associates you will never receive a message from me again . * * * first time ever offered ! * * * keep your prospect pipeline - tm filled ! disappointed with traditional marketing ? maybe it 's time to consider ' business to business ' direct e - mail . forget the " get rich quick " schemes and $ 395 + software . forget the " 60 million " address cd 's that are filled with duplicates and even invalid , " generated " addresses , hidden in many different files that rarely add up to even a million prospects which are still unqualified . over 90 % are private personal addresses of people who do not want to be invaded and unless you have duplicate filtering software , you would be mailing many of them multiple times , with the same message ! no wonder they call it spam . you should respect their privacy . prospect pipeline - tm gets you started with the contact e-mail addresses from each of 100 , 000 unique commercial web sites (e . g . www . mysite . com) and a free 5 day trial of e - mail pump : software that does what every business needs done - - it keeps a pipeline of prospects coming . maybe it 's time you filled your pipeline ? prospect pipeline is the most reasonable marketing / announcement tool you ' ll ever find at $ 49 . 95 (+ s&h) . you can continue to receive a fresh cd (100 , 000 new commercial addresses) each month thereafter , at a 20 % discount . we can even deliver them to you automatically ! prospect pipeline - tm addresses are contact addresses from commercial web sites (100 % ' . com ') . a commercial domain (. com) is a business by definition and business people love to do business . you know the value of qualified prospects and how much time and money you can waste if they ' re not . this is an extremely reasonable offer ! get down to business today . stop waiting for prospects to find you . prospect pipeline - tm business to business package includes : 100 , 000 highly refined (no duplicates) commercial contact e - mail addresses in

plain text files ready for mailing . a free , fully functional , 5 day trial of ' e - mail pump , ' the latest in direct mail software technology . start prospecting immediately ! e - mail pump includes a built in ' instant ' registration process via the internet . it 's also priced reasonably at $ 49 . 95 , should you decide to register . if you have any further questions or to place an order by phone , please do not hesitate to call us at : 904-441 - 8080 business hours are monday - saturday 9 : 00 am - 9 : 00 pm . to order by fax or postal mail , simply print out the order form below and fax or mail it to our office today . * we accept us checks by fax , telephone and postal mail . money orders in us dollars drawn on us or canadian banks only , are accepted by postal mail . * - order form - env associates - voice telephone : 904-441 - 8080 business hours are monday - saturday 9 : 00 am - 9 : 00 pm . complete this form and follow the fax instructions at the bottom . all orders are sent us postal service 3 day priority mail or global priority mail outside of the us . _ _ _ _ _ yes ! please send me the ' prospect pipeline - tm ' cd-rom of 100 , 000 fresh , new , commercial addresses and free - 5 day trial of e - mail pump for only $ 49 . 95 (us dollars) name _ company name _ address _ address _ city , state , zip _ country _ phone number (s) _ fax number _ e-mail address _ 2nd e-mail address _ * please select the appropriate shipping for your location and make your check payable for the respective total . . . _ _ _ _ i am in the united states , so i will add $ 3 . 00 for us postal priority mail for a total of $ 52 . 95 (us dollars) . _ _ _ _ i am in canada ,

so i will add $ 6 . 95 for global priority mail for a total of $ 56 . 90 (us dollars) . _ _ _ _ i am outside the us and canada , so i will add $ 12 . 00 for global priority mail for a total of $ 61 . 95 (us dollars) . - * * * 24 hour ordering by fax * * * - 1 . print this order form 2 . paste or tape your check here 3 . be sure the above form is complete 4 . fax to 1-904 - 441-6481 (24 hours , 7 days a week) - you need not mail the original check when using check - by - fax . our banking software drafts a special check , with the exact information from your original . orders are shipped at the time funds clear . if you feel uncomfortable with check - by - fax or check - by - phone payment , send this form with your check or money order to : env associates 171 east granada boulevard ormond beach , florida 32176 904-441 - 8080 voice 904-441 - 6481 fax * -

謝辞

本論文は防衛大学校理工学研究科後期課程において，白川智弘 講師および佐藤浩 准教授のご指導のもと，研究成果をまとめたものです.

白川講師と佐藤准教授には筆者の以前の所属である東京電機大学理工学研究科修士課程在籍時から，約５年に渡りご指導いただきました．両先生のもとで研究科後期課程を過ごせたことを大変嬉しく思います．厚く御礼申し上げます.

学位論文審査において，本学理工学部の黒川恭一 教授，三村守 准教授，大阪府立大学工学部の森直樹 准教授には論文作成にあたり多大なるご指導と助言をいただきました．厚く御礼申し上げます.

また，研究室のメンバーである陸上自衛官の室道徳 二等陸尉，本科学生の松尾拓哉 学生，橋本真冶 学生，若林真純 学生，ブイ・ドク・ヴェト 学生には，公私共に多くの助言とサポートをいただきました．厚く御礼申し上げます.

研究室在籍中には，理化学研究所言語情報アクセス技術チームおよび New York University の関根聡 准教授および，WorkFusion の Abby Levenberg 博士，Tianhao Wu 博士, Ilya Vadeiko 博士, Nelson Chen 氏, Aaron Yu 氏, Artsiom Strok 氏, Farnaz Abtahi 博士，Kenichi Moriya 氏，Hiroaki Morita 氏，Marshall Zheng 氏，Motokazu Ishikawa 氏，Kazumasa Sakai 氏，Akira Senoo 氏，Hiroshi Watanabe 氏，Hisatou Watanabe 氏，Aleh Hlushkou 氏，Andrey Protasevich 氏，Facebook の Fadi Botros 氏から，機械学習の技術に関し，大変多くの助言をいただきました．上記の皆様のおかげで，多くの知識を培い，研究のアイデアを考えることができました．この場をお借りして，厚く御礼申し上げます.

早稲田大学理工学部の郡司ペギオ幸夫 教授，京都大学デザイン学ユニットの川上浩司 教授，ほりのうちカイロプラクティックの杉山成久 先生，東京大学の篠原修二 助教には，学会発表や研究会の場で，大変お世話になりました．また，博士後期課程に進学する際に，大変お世話になった東京電機大学理工学部の柴山拓郎 准教授，ホンダ・リサーチ・インスティチュート・ジャパンの船越孝太郎 博士，茨城大学農学部の朝山宗彦 教授に，この場をお借りして厚く御礼申し上げま

す.

最後に，私の研究生活を支えてくれた両親，祖母，愛犬のくるみ，そしていつも心の支えとなってくださった 20th Century Fox および The Walt Disney Company の Chris Collins 氏に，心から感謝します.

参考文献

[Ackley et al. 1985] Ackley, D. H., Hinton, G. E., & Sejnowski, T. J. A learning algorithm for Boltzmann machines, Cognitive science **9**(1), 147-169 (1985).

[Alpaydin 2014] Alpaydin, E. Introduction to machine learning (MIT press, 2014).

[Androutsopoulos et al. 2000] Androutsopoulos, I., Koutsias, J., Chandrinos, K. V., Paliouras, G., & Spyropoulos, C. D. An evaluation of naive Bayesian anti-spam filtering. arXiv preprint cs/0006013 (2000).

[Arpit et al. 2016] Arpit, D., Zhou, Y., Kota, B. U., & Govindaraju, V. Normalization propagation: A parametric technique for removing internal covariate shift in deep networks, ArXiv, 1603.01431 (2016).

[Barrouillet & Gauffroy 2015] Barrouillet, P., & Gauffroy, C. Probability in reasoning: a developmental test on conditionals. Cognition **137**, 22-39 (2015).

[Bayes & Price 1763] Bayes, M. & Price, M. An essay towards solving a problem in the doctrine of chances. By the late Rev. Mr. Bayes, FRS communicated by Mr. Price, in a letter to John Canton, AMFRS. Philos. Trans., 370-418 (1763).

[Birch et al. 2008] Birch, S. A., Vauthier, S. A., & Bloom, P. Three-and four-year-olds spontaneously use others' past performance to guide their learning. Cognition **107**, 1018-1034 (2008).

[Breiman 1996] Breiman, L. Bagging predictors, Machine learning **24**(2), 123-140 (1996).

[Breiman 2001] Breiman, L. Random forests, Machine learning **45**, 5-32 (2001).

[Conway & White 2012] Conway, D., & White, J. Machine learning for hackers. (O'Reilly Media, 2012).

[Cox 1958] Cox, D. The regression analysis of binary sequences (with discussion). Journal of the royal statistical society: series b **20**, 215-242 (1958).

[Dahl et al. 2013] Dahl, G., Sainath, T. & Hinton, G. Improving deep neural networks

for LVCSR using rectified linear units and dropout, Acoustics, Speech and Signal Processing (ICASSP 2013), 8609-8613 (2013).

[Diesendruck & Markson 2001] Diesendruck, G. & Markson, L. Children's avoidance of lexical overlap: a pragmatic account. Developmental psychology **37**, 630-641 (2001).

[Domingos & Pazzani 1997] Domingos, P. & Pazzani, M. On the optimality of the simple Bayesian classifier under zero-one loss. Machine learning **29**, 103-130 (1997).

[Dougherty 2013] Dougherty, G. Pattern Recognition and Classification-An Introduction (Springer, 2013).

[Eberhardt 2015] Eberhardt, J. J. Bayesian spam detection, Scholarly Horizons: University of Minnesota, Morris Undergraduate Journal, **2**(1), (2015).

[Edgington 1995] Edgington, D. On conditionals. Mind, **104**, 235-329 (1995).

[Feldman 2000] Feldman, J. Minimization of boolean complexity in human concept learning. Nature **407**, 630-633 (2000).

[Firoiu et al. 2017] Firoiu, V., Whitney, W. F., & Tenenbaum, J. B. Beating the World's Best at Super Smash Bros. with Deep Reinforcement Learning, arXiv, 1702.06230 (2017).

[Friedman et al. 1997] Friedman, N., Geiger, D., & Goldszmidt, M. Bayesian network classifiers. Machine learning **29**(2-3), 131-163 (1997).

[Gal & Ghahramani 2016] Gal, Y., & Ghahramani, Z. Dropout as a Bayesian approximation: Representing model uncertainty in deep learning, International conference on machine learning, 1050-1059 (2016).

[Gerken et al. 2015] Gerken, L., Dawson, C., Chatila, R. & Tenenbaum, J. Surprise! Infants consider possible bases of generalization for a single input example, Developmental Science, **18**, 80-89 (2015).

[Gitman & Ginsburg 2017] Gitman, I. & Ginsburg, B. Comparison of batch normalization and weight normalization algorithms for the large-scale image classification, arXiv, 1709.08145 (2017).

[Goodfellow et al. 2014] Goodfellow, et al: Generative adversarial nets. In Advances in

neural information processing systems, 2672/2680 (2014).

[Goodfellow et al. 2016] Goodfellow, I., Bengio, Y., Courville, A. & Bengio, Y. Deep learning (MIT press, 2016).

[Goodman et al. 2007] Goodman, J., Cormack, G. V. & Heckerman, D. Spam and the ongoing battle for the inbox. Communications of the ACM, **50**, 24-33 (2007).

[Goodman et al. 2008] Goodman, N. D., Tenenbaum, J. B., Feldman, J. & Griffiths, T. L. A rational analysis of rule-based concept learning. Cognitive science **32**, 108-154 (2008).

[Halberda 2003] Halberda, J. The development of a word-learning strategy. Cognition **87**, 23-34 (2003).

[Han et al. 2017] Han, Z., Wei, B., Zheng, Y., Yin, Y., Li, K. & Li, S. Breast cancer multi-classification from histopathological images with structured deep learning model, Scientific Reports, **7**, 4172 (2017).

[Hebb 1949] Hebb, D. The organization of behavior: A neuropsychological theory, (John Wiley and Sons, 1949).

[Hattori & Oaksford 2007] Hattori, M. & Oaksford, M. Adaptive non-interventional heuristics for covariation detection in causal induction: model comparison and rational analysis. Cognitive science **31**, 765-814 (2007).

[He et al. 2008] He, H., Bai, Y., Garcia, E. A., & Li, S. ADASYN: Adaptive synthetic sampling approach for imbalanced learning, 2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence), 1322-1328 (2008).

[Hinton et al. 2006] Hinton, G. E., Osindero, S., & Teh, Y. W. A fast learning algorithm for deep belief nets, Neural computation, **18**(7), 1527-1554 (2006).

[Hinton & Salakhutdinov 2006] Hinton, G. & Salakhutdinov, R. Reducing the dimensionality of data with neural networks, Science, **313**, 504-507 (2006).

[Hinton et al. 2012] Hinton, G., Srivastava, N., Krizhevsky, A., Sutskever, I. & Salakhutdinov, R. Improving neural networks by preventing co-adaptation of feature

detectors, arXiv, 1207.0580 (2012).

[Hinton et al. 2015] Hinton, G., Vinyals, O., & Dean, J. Distilling the knowledge in a neural network, arXiv, 1503.02531 (2015).

[Hopfield 1982] Hopfield, J. Neural networks and physical systems with emergent collective computational abilities, Proc. the national academy of sciences, 2554-2558 (1982).

[Hrovat et al. 2014] Hrovat, G., Stiglic, G., Kokol, P. & Ojstersek, M. Contrasting temporal trend discovery for large healthcare databases, Computer methods and programs in biomedicine, **113**, 251-257 (2014).

[Japkowicz & Stephen 2002] Japkowicz, N. & Stephen, S. The class imbalance problem: A systematic study, Intelligent data analysis, **6**, 429-449 (2002).

[Jenkins & Ward 1965] Jenkins, H. M. & Ward, W. C. Judgment of contingency between responses and outcomes. Psychological monographs: general and applied **79** (1965).

[Kahneman 2002] Kahneman, D. Thinking, fast and slow (Macmillan, 2002).

[Kanaris et al. 2006] Kanaris, I., Kanaris, K., & Stamatatos, E. Spam detection using character n-grams, Hellenic conference on artificial intelligence, 95-104 (2006).

[Katz 1993] Katz, V. J. (1993). A history of mathematics: an introduction. *New York*.

[Katz 1996] Katz, S. M. Distribution of content words and phrases in text and language modelling. Natural language engineering **2**, 15-59 (1996).

[King & Zeng 2001] King, G., & Zeng, L. Logistic regression in rare events data, Political analysis **9**(2), 137-163 (2001).

[Kwan et al. 2002] Kwan, K. Y., Lee, T., & Yang, C. Unsupervised n-best based model adaptation using model-level confidence measures. Seventh International Conference on Spoken Language Processing, 69-72 (2002).

[Lake et al. 2015a] Lake, B., Salakhutdinov, R., Gross, J., & Tenenbaum, J. One shot learning of simple visual concepts. Proc. Cog. Sci. Soc. USA **33**, 1332-1338 (2015).

[Lake et al. 2015b] Lake, B. M., Salakhutdinov, R., & Tenenbaum, J. Human-level

concept learning through probabilistic program induction. Science, **350**, 1332-1338 (2015).

[Lake et al. 2017] Lake, B. M., Ullman, T. D., Tenenbaum, J. B., & Gershman, S. J. Building machines that learn and think like people. Behavioral and Brain Sciences, **40** (2017).

[LeCun et al. 1998] LeCun, Y., Bottou, L., Bengio, Y., and Haffner, P. Gradient-based learning applied to document recognition, Proc. the IEEE, 2278-2324 (1998).

[LeCun et al. 2015] LeCun, Y., Bengio, Y. & Hinton, G. Deep learning, Nature, **521**, 436-444 (2015).

[Lin et al. 2014] Lin, D., Dechter, E., Ellis, K., Tenenbaum, J., & Muggleton, S. Bias reformulation for one-shot function induction. Proceedings of the twenty-first ECAI. Czech Republic **263**, 525-530 (2014).

[Ioffe & Szegedy 2015] Ioffe, S. & Szegedy, C. Batch normalization: Accelerating deep network training by reducing internal covariate shift, arXiv, 1502.03167 (2015).

[Ma et al. 2009] Ma, J., Saul, L. K., Savage, S., & Voelker, G. M. Identifying suspicious URLs: an application of large-scale online learning, In Proceedings of the 26th annual international conference on machine learning, 681-688 (2009).

[Marcano-Cedeno et al. 2011] Marcano-Cedeno, A. Quintanilla-Dominguez, J. & Andina, D. WBCD breast cancer database classification applying artificialmetaplasticity neural network, Expert Systems with Applications, **38**, 9573-9579 (2011).

[Markman & Wachtel 1988] Markman, E. M. & Wachtel, G. F. Children's use of mutual exclusivity to constrain the meanings of words. Cognitive psychology **20**, 121-157 (1988).

[Markman 1990] Markman, E. M. Constraints children place on word meanings. Cognitive science **14**, 57-77 (1990).

[Mason 2003] Mason, J, SpamAssassin Public Corpus, http://spamassassin.apache.org/publiccorpus (2003).

[Matoba et al. 2011] Matoba, R., Nakamura, M., & Tojo, S. Efficiency of the symmetry bias in grammar acquisition, Information and Computation **209**(3), 536-547 (2011).

[McCulloch & Pitts 1943] McCulloch, W. S., & Pitts, W. A logical calculus of the ideas immanent in nervous activity, The bulletin of mathematical biophysics **5**(4), 115-133 (1943).

[McGrayne 2011] McGrayne, S. B. The theory that would not die: how Bayes' rule cracked the enigma code, hunted down Russian submarines, & emerged triumphant from two centuries of controversy, (Yale University Press, 2011).

[Merriman et al. 1989] Merriman, W. E., Bowman, L. L., & MacWhinney, B. The mutual exclusivity bias in children's word learning. Monographs of the society for research in child development **54**, (1989).

[Mitchell 1997] Mitchell, T. M. Machine learning (McGraw Hill, 1997).

[Ng & Jordan 2002] Ng, A. Y. & Jordan, M. I. On discriminative vs. generative classifiers: a comparison of logistic regression and naive Bayes. Advances in neural information processing systems, 841-848 (2002).

[Ohmura et al. 2012] Ohmura, H., Shibayama, T., Takahashi, T., Shibuya, S., Okanoya, K., & Furukawa, K, Melody generation system based on generalization by human causal intuition, SICE Annual Conference (SICE), 2005-2010 (2012).

[Over & Evans 2003] Over, D. E., & Evans, J. S. B. The probability of conditionals: The psychological evidence. Mind & Language **18**, 340-358 (2003).

[Over et al. 2007] Over, D. E., Hadjichristidis, C., Evans, J. S. B., Handley, S. J., & Sloman, S. A. The probability of causal conditionals. Cognitive psychology **54**, 62-97 (2007).

[Pitsillidis et al. 2010] Pitsillidis, A. et al. Botnet Judo: Fighting Spam with Itself, NDSS Symposium, (2010).

[Rao & Reiley 2012] Rao, J. M. & Reiley, D. H.. The economics of spam. Journal of Economic Perspectives, **26**, 87-110 (2012).

[Reigl et al. 2004] Reigl, M., Alon, U. & Chklovskii, D. Search for computational

modules in the C. elegans brain, BMC biology, **2**, (2004).

[Rosenblatt 1958] Rosenblatt, F. The perceptron: a probabilistic model for information storage and organization in the brain. Psychological review, **65**, 386-408 (1958).

[Rumelhart et al. 1986] Rumelhart, D. E., Hinton, G. E., & Williams, R. J. Learning representations by back-propagating errors, Nature **323**(6088), (1986).

[Salakhutdinov et al. 2012] Salakhutdinov, R., Tenenbaum, J., & Torralba, A. One-shot learning with a hierarchical nonparametric Bayesian model. Proceedings of ICML workshop on unsupervised and transfer learning. USA **27**, 195-206 (2012).

[Salimans et al. 2016] Salimans, T., Goodfellow, I., Zaremba, W., Cheung, V., Radford, A., & Chen, X. Improved techniques for training gans, Advances in Neural Information Processing Systems 2016, 2234-2242 (2016).

[Salton 1989] Salton, G. Automatic text processing: the transformation, analysis, and retrieval of information by computer (Addison-Wesley, 1989).

[Sarkar et al. 2005] Sarkar, A., Garthwaite, P. H. & De Roeck, A. A Bayesian mixture model for term re-occurrence and burstiness. Proceedings of the Ninth Conference on Computational Natural Language Learning, 48-55 (2005).

[Schneider 2005] Schneider, K. M. Techniques for improving the performance of naive bayes for text classification, International Conference on Intelligent Text Processing and Computational Linguistics, 682-693 (2005).

[Sermanet et al. 2014] Sermanet, P., Frome, A., & Real, E. Attention for fine-grained categorization, ArXiv, 1412.7054 (2014).

[Sidman et al. 1982] Sidman, M. *et al.* A search for symmetry in the conditional discriminations of rhesus monkeys, baboons, and children. Journal of the experimental analysis of behavior **37**, 23-44 (1982).

[Silver et al. 2016] Silver, D. et al. Mastering the game of Go with deep neural networks and tree search, Nature **529**(7587), (2016).

[Sommer & Wurtz 2000] Sommer, M. & Wurtz, R. Composition and topographic organization of signals sent from the frontal eye field to the superior colliculus,

Neurophysiology, **83**, 1979-2001 (2000).

[Srivastava et al. 2014] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever and R. Salakhutdinov: Dropout: A simple way to prevent neural networks from overfitting, Machine Learning Research, **15**, 1929-1958 (2014).

[Sun et al. 2016] Sun, W., Shao, S., Zhao, R., Yan, R., Zhang, X. & Chen, X. Sparse auto-encoder-based deep neural network approach for induction motor faults classification, Measurement, **89**, (2016).

[Swietojanski et al. 2014] Swietojanski, P., Li, J., & Huang, J. T. Investigation of maxout networks for speech recognition, Acoustics, Speech and Signal Processing (ICASSP) 2014, 7649-7653 (2014).

[Takahashi et al. 2010] Takahashi, T., Nakano, M., & Shinohara, S. Cognitive symmetry: illogical but rational biases. Symmetry: culture and science **21**, 275-294 (2010).

[Takahashi et al. 2011] Takahashi, T., Oyo, K., & Shinohara, S. A loosely symmetric model of cognition. Advances in artificial life Darwin meets von Neumann. Hungary **5778**, 238-245 (2011).

[Tanner & Wong 1987] Tanner, M. A., & Wong, W. H. The calculation of posterior distributions by data augmentation, Journal of the American statistical Association **82**, 528-540 (1987).

[Tenenbaum 1999] Tenenbaum J., Bayesian modeling of human concept learning. Advances in neural information processing system, 59-65 (1999).

[Tversky & Kahneman 1973] Tversky, E. & Kahneman, D. Availability: a heuristics for judging frequency and probability. Cognitive psychology **5**, 207-232 (1973).

[Tversky & Kahneman 1974] Tversky, E. & Kahneman, D. Judgement under uncertainty: heuristics and biases. Science **27**, 1124-1131 (1974).

[Vapnik 1963] Vapnik, V. The nature of statistical learning theory (Springer, 1963).

[Weiss & Provost 2003] Weiss, G. & Provost, F. Learning when training data are costly: The effect of class distribution on tree induction, Journal of Artificial Intelligence

Research, **19**, 315-354 (2003).

[Werbos 1975] Werbos, P. J. Beyond regression: new tools for prediction and analysis in the behavioral sciences. Doctoral thesis, Harvard University (1975).

[Zhang 2004] Zhang, H. The optimality of naive Bayes, Proceedings of the Seventeenth International Florida Artificial Intelligence Research Society Conference, (2004).

[Zhang et al. 2015] Zhang, H., Miao, Y. & Metze, F. Regularizing dnn acoustic models with gaussian stochastic neurons, Acoustics, Speech and Signal Processing (ICASSP 2015), 4964-4968 (2015).

[郡司 2008] 郡司 ペギオ 幸夫, 因果推論とアドホック論理: 世界内に留まり続ける意識の構造, 人工知能学会, (2008).

[篠原 & 中野 2007] 篠原 修二 & 中野 昌宏, 2本腕バンディット問題に対する「緩い対称性モデル」の有効性 : 因果推論における対称性バイアスと相互排他性バイアス, 進化経済学論集 **11**, (2007).

[篠原ら 2007] 篠原修二, 田口亮, 桂田浩一, & 新田恒雄. 因果性に基づく信念形成モデルと N 本腕バンディット問題への適用. 人工知能学会論文誌 **22**(1), 58-68 (2007).

研究業績

学術論文 (査読あり)

[1] H. Taniguchi, T. Shirakawa, T. Takahashi: Implementation of Human Cognitive Bias on Naïve Bayes, EAI Endorsed Trans. Creative Technologies, **3**(7), (May 2016).

[2] H. Taniguchi, H. Sato, T. Shirakawa: A machine learning model with human cognitive bias that is capable of learning from small and biased datasets, Scientific Reports, **8**(7397), (May 2018).

[3] H. Taniguchi, H. Sato, T. Shirakawa: Implementation of Human Cognitive Bias on Neural Network and its Application to Breast Cancer Diagnosis, SICE Journal of Control, Measurement, and System Integration, **12**(2), (March 2019).

国際会議 (査読あり)

[1] H. Taniguchi, K, Oyo, Y. Kohno, T, Takahashi: Causal cognition and spam classifier, Inetrnational Conference of Numerical Analysis and Applied Mathematics 2015, 23-29 Sep 2015, Rodos Palace Hotel, Rhodes, Greece (2015).

[2] H. Taniguchi, H, Sato, T. Shirakawa: Application of human cognitive mechanisms to Naïve Bayes text classifier, Inetrnational Conference of Numerical Analysis and Applied Mathematics 2017, 25-30 Sep 2017, Rodos Palace Hotel, Rhodes, Greece (2017).

国内学会

[1] 谷口英貴, 白川智弘, 高橋達二: 認知特性の付与によるナイーブベイズ分類器の性能向上, 計測自動制御学会 システム・情報部門 学術講演会 2015, 2015 年 11 月 18 日-20 日, 函館アリーナ, 函館 (2015).

[2] 谷口英貴, 佐藤浩, 白川智弘: 人間の認知バイアスを利用したナイーブベイズ分類器の性能向上, 計測自動制御学会 システム・情報部門 学術講演会 2016, 2016 年 12 月 6-8 日, ウカルちゃんアリーナ, 大津 (2016).

[3] 谷口英貴, 佐藤浩, 白川智弘: Improvement of Naïve Bayes Spam Filtering Based on Human Cognitive Biases, 22th International Symposium on Artificial Life and Robotics, 2017 年 1 月 19-21 日, ビーコンプラザ, 別府 (2017).
[4] 谷口英貴, 佐藤浩, 白川智弘: 人間の認知バイアスを用いることによるメール分類器の性能向上, 第 44 回知能システムシンポジウム プログラム, 2017 年 3 月 13-14 日, 東海大学 高輪キャンパス, 港区 (2017).
[5] 谷口英貴, 佐藤浩, 白川智弘: 人間の認知バイアスを利用したニューラルネットワークの性能向上, 計測自動制御学会 システム・情報部門 学術講演会 2017, 2017年11月25-27日, 静岡大学浜松キャンパス, 浜松 (2017). SSI研究奨励賞受賞.
[6] 谷口英貴, 佐藤浩, 白川智弘: 認知バイアスの実装による Generative Adversarial Nets の性能向上, 計測自動制御学会 システム・情報部門 学術講演会 2018, 2018 年 11 月 25-27 日, 富山国際会議場, 大手町 (2018).