DEIM Forum 2023 1a-3-5

ルールマイニングにおける部分正解集合による パラメータバリデーション

高本　綺架†　　細野　湧城†　　廣中　詩織†　　梅村　恭司†

† 豊橋技術科学大学 情報・知能工学系
〒 441–8580 愛知県豊橋市天伯町雲雀ヶ丘 1-1
E-mail: †{takamoto.ayaka.nx,hosono.yuki.ql,hironaka.shiori.ru,umemura}@tut.jp

あらまし　巨大なデータベースから価値あるルールを抽出するルールマイニングは，広く研究されている分野の一つである．ルールマイニングの手法には，頻度の低いアイテムに対応するためのパラメータが存在することがある．このパラメータは抽出の性能に大きな影響を与えるため，適切なパラメータを決定することは重要である．一般的には，このパラメータを決定するために検証データを用いる．これは，正解のルールをあらかじめ用意しておき，各パラメータを設定したときの抽出性能から適切なパラメータを決定する方法である．しかし，新たなルールを抽出したい場合は，パラメータの決定の際に完全な正解ルールを揃えることはできない．そこで我々は，正解の一部だけを用意することは可能な場合もあると考えた．本研究では，ルールマイニングにおいて検証データの数がパラメータの推定に及ぼす影響を調査する．実験では，正解集合が明らかなタスクとして，新聞記事データを対象として県名と市郡名のルールを抽出するタスクを用いた．正解ルールを減らした状態でパラメータ推定を行い，正解ルールを全て用いた場合のパラメータと比較する．実験の結果，正解ルールの一部だけしか含まれていない検証データでも，パラメータの推定が可能であったことを報告する．

キーワード　バリデーション，ルールマイニング，尤度比の直接推定

1　はじめに

巨大なデータベースに存在する複数のアイテムから，関係の強いアイテムの組み合わせを発見するルールマイニングは，データマイニングで主要な技術であり，広く研究されている分野である [1]．POS システムなどのレシートデータベースから発見されるルールの例として ⟨Diaper, Beer⟩,「おむつを買った人はビールも買う傾向がある」というものがある．これは，あるトランザクション（レシート）がビールというアイテムを含んでいるときに，別のアイテムであるおむつも含まれるという関係である．こういったアイテム同士のルールの強さは，レシートを選ぶ試行において，アイテム X とアイテム Y の出現についての条件付き確率 $P(Y|X)$ として表現される．上記の例であれば，P(Beer|Diaper) と表現できる．

この条件付き確率の値が高いほど，2 つのアイテムが同じトランザクションに含まれる可能性が高いことを示している．もし観測できるデータが無限にあり，正しくないルールのペアが出現しないのであれば，P(Beer|Diaper) の最尤推定値をルールの強さとできる．しかし，実際のデータは有限であるため，アイテムが低頻度でしか出現しない可能性は十分にある．低頻度でしか観測できない事象に対して最尤推定を行なった場合，推定値と真値との差が大きくなる可能性が高くなるため，低頻度でしか出現しないルールが問題になる．したがって，低頻度でしか出現しないアイテムを含むルールでは，データ中に偶然現れる可能性を考慮し，ルールの強さを低めに見積もる必要がある．

低頻度のアイテムに対応するために，いくつかのルールマイニング手法ではパラメータを使用する．ルールマイニングでよく使用される Apriori [2] では，設定されたパラメータよりも関係があると考えられる 2 つのアイテムが共に出現するトランザクションの数が少ない場合は，そのルールを無視する．また，Kawakami らの手法 [6] では，前提となるアイテムの出現数にパラメータとして設定した数値を加算することで，低頻度のルールの強さを軽減している．一般的には，このようなパラメータの決定は，正解のルールを集めた検証データを用いて行う．パラメータを変更しながら実際にルール抽出を行い，その結果が検証データに含まれている正解ルールとどの程度一致しているかを判定してパラメータを推定する．

実際のアプリケーションでは，過去のデータを用いてパラメータを推定することが多い．例えば新聞記事データを対象にしたものであれば，前年度のデータから検証データを作成し，パラメータを決定することが考えられる．しかし，適切なパラメータは対象のデータごとに異なる場合があるため，同じ種類のデータであっても年度が異なるデータを用いてパラメータの決定を行うと，検証データの正確性に関わらず適切でないパラメータを算出する可能性がある．例えば，新製品に関するルールは前年度までのデータには登場しないため，過去のデータを用いてパラメータを調整するのは問題がある．もし，異なる年度の全正解ルールを使用して計算したパラメータと，ルールを推定したいデータにおける部分正解集合を用いて計算したパラメータを比べたとき，後者が優れているならば，対象となる

データの部分正解集合を用いてパラメータを推定した方がより適切なパラメータを設定できるのではないかと考えた．

本研究では，年度ごとに出現する正解集合が求まるという観点から，新聞記事データにおける県名と市郡名のルール抽出タスクを対象とし，以下の 2 つの実験を行なった．実験 1 では，データの年度が異なることによる最適なパラメータの違いを調査する．実験 2 では，正解ルール集合から正解ルールをランダムに削減することにより，パラメータがどう変化するかを調査する．実験 1 により，推定されるパラメータの値が年度によって異なることが確認された．実験 2 では，パラメータ推定に用いる正解ルールは 30%程度の欠損があったとしても，推定されるパラメータに大きな違いがないことがわかった．以上の結果から，異なるデータ集合を用いてバリデーションを行うよりも，判定対象となるデータから正解集合を部分的に作成してバリデーションを行なったほうが，より適切なパラメータを設定できる可能性があることが示唆された．

2 問題設定

2.1 新聞記事における地名ルールの抽出

本研究では，都道府県名 x と市郡名 y 間の相関ルールを $\langle x, y\rangle$ と定義し，新聞記事データからこのルールを抽出することを考える．都道府県名と市郡名間に存在するルールは，数あるルールの中でもアイテム間のルールが明確であり，抽出されたルールが適切かを判定しやすい．そのため，本研究でのルール抽出タスクに選んだ．本研究では，x が都道府県名，y が市郡名となるルール $\langle x, y\rangle$ を抽出することを目的とする．

新聞記事の集合をデータ集合 D とし，その中に存在する都道府県名や市郡名の集合をアイテム集合 I とする．このアイテム集合 I は，例えば次のような集合になる．

$$I = \{\text{“愛知県”}, \text{“山口県”}, \text{“東京都”}, \text{“豊橋市”}, \dots\}$$

2.2 正解集合の作成方法

正解ルールの集合である正解集合 R を用意する．都道府県名と市郡名における正解集合 R は，都道府県名とそれに属する市町村などの市郡名の包含関係と考えることができる．したがって，ある県名のアイテム集合と，その県に属する市郡名からなる正解集合 R には，$\langle$“愛知県”, “豊橋市”$\rangle$, $\langle$“愛知県”, “豊田市”$\rangle$ などが含まれる．

正解集合 R に，日本に存在するあらゆる市とその市が属する県名のルールが含まれるように，2000 年の郵便番号データから正解集合を作成した．郵便番号データから，ある都道府県に含まれる市郡という 1 対多関係を抽出した．地名には「県」や「市」などのような，都道府県や市郡を表す語句がついているが，新聞記事に含まれる地名には「愛知」のように県がついていないことがある．そこで,「県」や「市」などを排除したものも併せて正解とした．大阪府と大阪市との関係では，$\langle$“大阪”, “大阪”$\rangle$ や $\langle$“大阪”, “大阪府”$\rangle$ なども正解とした．その結果，1201 件の 1 対多関係のデータが得られた．これをもとに正解集合 R を作成した結果，含まれる正解ルール数は 6094 になった．

アルゴリズムが抽出したルールが正解であるかを判定する際にも，同じく制約を設ける．x が都道府県名，y が市郡名だと考えるため，ルール $\langle x, y\rangle$ が正解かどうかを判定する際，x の頻度が y よりも大きいルールは判定対象にしない．例えば，x を愛知県，y を豊橋市とすると，愛知県が出現した時に豊橋市が出現する可能性は，豊橋市が出た場合に愛知県が出現する可能性より低いと考える．本研究のルール抽出では，ルールにおけるアイテムの順序は重視しないため，この場合であれば豊橋市が出現した時に，愛知県が出現するというルールのみを取得することになる．

3 使用するルールマイニング手法

本研究では，次の 3 つのルールマイニング手法を調査対象とする：Apriori [2]，L1 正則化を用いる尤度比の直接推定手法 [4]，L2 正則化を用いる尤度比の直接推定手法 [6]．ルールマイニングでは，これらの手法でルールの強さを推定した後，ルールの強い順に並べて新しいルールを発見する．以下の節でこれらの手法を順番に説明する．

3.1 尤度比の最尤推定

調査対象とする 3 つの手法は尤度比の最尤推定と関わりが深い．そのため，まず最尤推定によりルールの強さを推定する方法を説明する．

アイテム間のルールの強さは，一般に条件付き確率で推定される．アイテム x と y の間に存在するルール $\langle x, y\rangle$ の強さ $P(X|Y)$ を最尤推定するとき，まず $P(X, Y)$ と $P(Y)$ をそれぞれ求め，その比 $P(X, Y)/P(Y)$ を計算する．$P(X, Y)$ を x と y が同時に出現する確率，$P(Y)$ を y が出現する確率とすると，$\hat{P}(X, Y) = c_{xy}/N$，$\hat{P}(Y) = c_y/N$ と推定できる．ここで，N は対象となるデータ D 内のトランザクションの総数であり，c_y は y が出現したトランザクション数，c_{xy} は x と y が同時に出現したトランザクション数とする．つまり，最尤推定による尤度比は次の式で計算される．

$$\hat{P}_{\mathrm{MLE}}(\langle x, y\rangle) = \frac{c_{xy}}{c_y} \tag{1}$$

3.2 Apriori

Apriori [2] は Agrawal らによって提案された手法であり，一定値以下の低頻度な事象を無視してルールの強さを推定する手法である．推定値は次の式で計算される．

$$\hat{P}_{\mathrm{Apriori}}(\langle x, y\rangle) = \begin{cases} \frac{c_{xy}}{c_y} & \text{if } c_{xy} > \theta \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

ここで，θ は低頻度とみなすアイテムペアの出現回数を決定するパラメータである．θ を大きくすることで，低頻度でしか出現しないルールの強さを 0 に下げる（低めに見積もる）ことになる．ルールの出現頻度が θ より大きいときは，最尤推定によりルールの強さを推定する．

表 1　新聞記事データの詳細情報

記事の取得年	トランザクション数	候補となるペアの種類数
1991	52232	247639
1992	56587	222501
1993	52031	209844
1994	65922	252297
1995	76563	226205
1996	58537	162112
1997	71966	161089

3.3　L1 正則化付き直接推定法

本手法は，尤度比の直接推定に L1 ノルムを正則化項として用いた場合の手法である [4]．推定値は次の式で計算される．

$$\hat{P}_{\mathrm{L1}}(\langle x, y\rangle) = \begin{cases} \frac{c_{xy}-\lambda}{c_y} & \text{if } c_{xy} > \lambda \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

ここで，λ はパラメータである．これは，Apriori と同様に一定の低頻度事象に対して推定を行わずに 0 とすることで，低頻度事象に対応している手法である．

3.4　L2 正則化付き直接推定法

本手法は，尤度比の直接推定に L2 ノルムを正則化項として用いた場合の手法である [6]．推定値は次の式で計算される．

$$\hat{P}_{L2}(\langle x, y\rangle) = \frac{c_{xy}}{c_y + \lambda} \tag{4}$$

L1 正則化項を用いた手法と同様，λ がパラメータである．これは L1 正則化項を用いた推定方法とは逆に，推定式の分母にある一定の値を加算することで，低頻度の事象の確率をより低く見積もっている．

4　実　　験

本節では，複数年度の新聞記事データを使用して，年度ごとの最適なパラメータの変化と，正解が欠損しているときの最適なパラメータの変化を調査する．実験 1 では，欠損のない正解集合を使用しパラメータ推定を行うことで，年度ごとのパラメータの差を明らかにする．実験 2 では，部分正解集合を使用してパラメータ推定を行い，完全な正解集合を用いて求めたパラメータと比較する．

4.1　使用するデータと前処理

本研究では，文献 [8], [9] と同様に，新聞記事データ集合として，1991 年から 1997 年の毎日新聞の記事コーパスを用いる．この新聞記事コーパスに含まれる記事を 1 つのトランザクションとし，記事中に出現する地名をアイテムとする．また，記事に出現する地名以外の語句は，本研究では不要であるため前処理の段階で削除している．前処理を行なった新聞記事から作成したデータの詳細情報を表 1 に示す．

4.2　パラメータの決定方法

調査対象のルールマイニング手法に存在するパラメータ λ（L1，L2 正則化項を用いた手法）及び θ（Apriori）の推定では，最も性能が良いときの λ を λ_{Opt} とし，適合率–再現率曲線 [5] を用いて以下のように推定を行う．まず，それぞれの手法を用いて，地名の各組に対しルールの強さを推定し，推定したルールの強さが強い順に並べ，ランク付けを行う．そして，上位 k 位までにランキングされているルールの正誤判定を行い，適合率@k 及び 再現率@k を求める．正誤判定には検証データを用いる．各ランクの適合率及び再現率は次の式で計算する．

$$\text{適合率@}k = \frac{\text{上位からランク } k \text{ までの正解ルール数}}{\text{上位からランク } n \text{ までに含まれる正解ルール数}}$$

$$\text{再現率@}k = \frac{\text{上位からランク } k \text{ までの正解ルール数}}{\text{正解集合に含まれる正解ルール数}}$$

縦軸を適合率，横軸を再現率として，k を変化させて適合率–再現率曲線を描き，この曲線の下部面積 (AUC of PRC) が最大値をとるときの λ を λ_{Opt} とする．AUC of PRC は，適合率と再現率が共に高く推定できているものほど良い性能であると評価するものであり，正解のペアを正しく判定できていることに主眼を置く評価指標である．

本研究では最適なパラメータ決定に，ハイパーパラメータ最適化フレームワークである Optuna [3] の TPESampler を用いた．実験に使用した Optuna に関する設定は，探索回数を 100 回，適応度は AUC of PRC，パラメータの探索範囲は 10^{-5} から 10000 とし，対数変換を行い探索している．Apriori については，取りうるパラメータが整数値であるため，探索範囲は 0 から 100 とした．

4.3　実験 1：年度ごとに算出されるパラメータの確認

本実験は，年度ごとに算出されるパラメータにどれくらいの差があるかを確認することを目的とする．具体的には，正解集合に手を加えずに各年の記事データにおける 3 つの手法のパラメータ λ_{Opt} を予測する．なお，この実験で使用する正解集合は削減しないため，本実験は 1 回のみ行う．

4.4　実験 2：正解集合を部分的に削減した場合のパラメータの変化

本実験では，パラメータを推定する際に正解集合を削減した場合の，削減率がパラメータの推定結果に与える影響を調査する．部分正解集合は次のように作成する．2.2 節で定義した 1201 件の正解集合から全体の 10%に相当する 120 件をランダムに選出し，正解集合から削除する．作成したデータからさらに 120 件をランダムに選出して正解ルールを削減していき，正解集合が残り 10%になった時点で削除をやめる．この処理を行うことで，部分正解集合として削減率が $10\%, 20\%, \ldots, 90\%$ の 9 件のデータが作成できる．

上記の部分正解集合を用いて，パラメータ推定の実験を次の流れで実施する．

（1）　部分正解集合を作成する．

（2）　各手法における最適なパラメータを**部分正解集合**を用いて推定する．

この手順を L1，L2 正則化項を用いた手法では 5 回実施した．Apriori を用いた場合に求まるパラメータは整数であるため，

確率的な変動を抑えるため 50 回実施した．その後，それぞれの平均と分散を求めた．なお，使用する部分正解集合は，実験を行うたびに新しく作成し直している．

5 実験結果

5.1 実験 1

実験 1 では，使用する正解集合に手を加えずにパラメータの推定を行なった．前述の通り，各年の新聞記事データにおけるパラメータの差に注目する．各手法で算出されたパラメータの最適値を表 2 の 0%の列に示す．算出されたパラメータを見ると，どの手法でも年度によってパラメータに差があることがわかる．Apriori では最小 2.0，最大 3.0 であり，L1 では最小 0.787，最大 1.641 である．L2 では最小 2.025 最大 6.448 である．

5.2 実験 2

実験 2 では，正解集合から 10%ずつ正解を削減した部分正解集合を用いてパラメータの推定を試みた．L1，L2 正則化項を用いた手法ではそれぞれ 5 回，Apriori を用いた推定では 50 回行い，それぞれの平均と不偏分散を求めた．図 1 から 3 に，それぞれの手法で推定されたパラメータの平均及び 95%信頼区間をプロットしたものを示す．さらに，削減率 0%から 90%でそれぞれ求めた最適なパラメータの平均値と不偏分散を付録の表 A·1 から A·6 に示す．なお，各表における 10%などの数値は，元の正解集合である 1201 件からの削減率を表している．

図を見ると，どの手法でも正解ルールの削減率が 30%までのパラメータ λ_{Opt} の値は，データの年度による差に比べて変化が小さいことがわかる．例えば，1995 年をとると，Apriori では削減なしのとき 2.0 であるが，3.0 となる年度の方が多い．L1 では，削減なしのとき最適なパラメータは 0.787 で，一番近い 1996 年でも 0.813 であるが，1995 年の 30%削減した検証データを使ったときは，最適なパラメータは 0.770 である．L2 も同様であるが，1995 年度に限り削減の影響が大きい．それでも，L2 では削減なしのとき最適なパラメータは 2.025 で，一番近い 1996 年でも 3.008 であるが，1995 年の 30%削減した検証データを使ったときは，最適なパラメータは 1.428 である．
また，分散の大きさから，これらの傾向は削減対象となる正解ルールの選び方に依存しないこともわかる．

6 議　　論

本実験で調査したパラメータは，尤度比の直接推定 [7] における正則化係数及び Apriori の低頻度とみなす際の閾値である．これは，正解の出現分布と不正解の出現分布の両方に依存するものである．実際，極端に正解集合の削減率を高くすると影響を受けるが，正解集合を 30%まで削減してもその影響はあまり観測できない．これは，検証データを用意する際に，ヒューリスティックスと人手による誤りの除去によって得られた部分正解集合で正則化係数を求めることが実際的であることを示している．

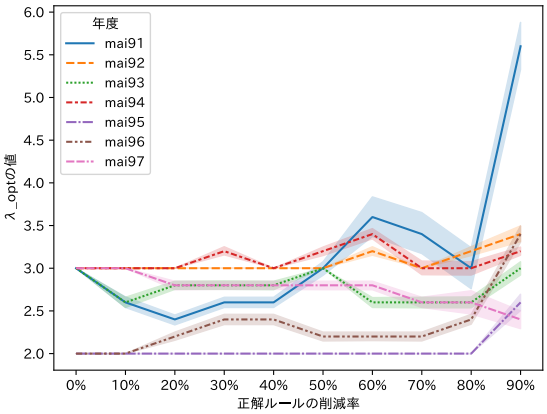

図 1　Apriori のパラメータの推定値

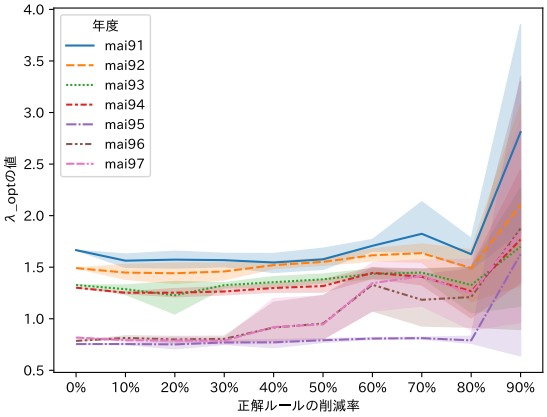

図 2　L1 推定のパラメータの推定値

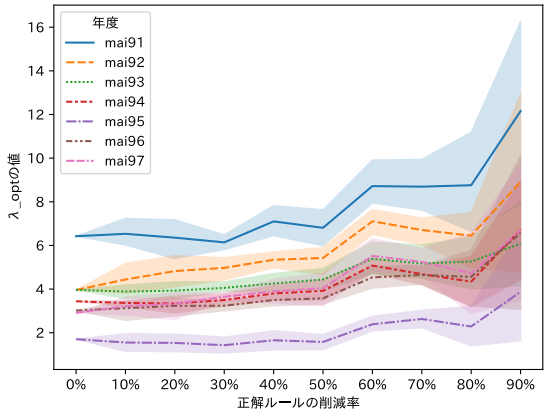

図 3　L2 推定のパラメータの推定値

次に興味深いところは，実際のアプリケーションで考えられる前年度のデータを用いて検証し，本年度の処理を行うことに対する示唆である．実験 1 から，適切な正則化係数の値は，年度によって異なることが示唆されている．このことから，年度が異なる同種の記事という比較的同質なデータであっても，適切な正則化係数の値が異なると言える．本実験の正解は地理的

表 2　各パラメータの年度ごとの差と 30%正解ルールを削減した場合の差

	Apriori		L1		L2	
削減率	0%	30%	0%	30%	0%	30%
1991	3.000	2.573	1.641	1.568	6.448	6.140
1992	3.000	3.060	1.478	1.459	4.884	4.972
1993	3.000	2.653	1.286	1.326	3.662	4.055
1994	3.000	3.067	1.303	1.265	3.453	3.498
1995	2.000	2.000	0.787	0.770	2.025	1.428
1996	2.000	2.487	0.813	0.805	3.008	3.236
1997	3.000	2.913	0.822	0.787	2.902	3.650

な関係であり，データの年度によって正解が変化することはない．したがって，誤って正解と検出される不正解のルールが，年度ごとに大きく異なると考えられる．

最後に興味深いところは，30%という比較的大きな正解ルールの削減率であっても，パラメータの推定結果の差が小さいところである．これは，L1，L2 正則化項を用いた手法及び Apriori に共通の結果となっている．今回は 1 つのタスクでの実験報告であるが，部分正解集合を用いたバリデーションの適用範囲を明確にするには，正解と不正解の分布について，これらの事象を説明できるようなモデルを考え，正則化係数のパラメータの理論的な解析をする必要がある．

7 おわりに

本研究では，新聞記事データから県名と市郡名のルールを抽出するタスクにおいて，尤度比の直接推定に L1，L2 正則化項を用いた手法と，Apriori によってルール抽出を行う際の最適なパラメータについて調査をおこなった．実験 1 では，まず年度ごとにパラメータ推定を行い，各手法におけるパラメータの年度ごとの差を確認した．実験 2 では，正解集合を部分的に削除したデータを用いてパラメータの推定を行なった．その結果，いずれの手法であっても 30%程度の欠損は容認できることがわかった．したがって，ルールを上位から抽出するようなタスクにおいては，同様に正解集合が部分的であっても適当なパラメータを推定できる可能性がある．また，新聞記事データの年度の差がパラメータに及ぼす影響よりも，部分正解集合を用いた場合のパラメータの推定値の差が及ぼす影響の方が小さいことから，前年度のデータを用いてルールの推定を行うよりも，対象となるデータから部分的に正解を抽出し，そのデータを用いてパラメータの推定を行なった方が良いと考えられる．

文　献

[1] Rakesh Agrawal, Tomasz Imieliński, and Arun Swami. Mining association rules between sets of items in large databases. In *Proceedings of the 1993 ACM SIGMOD international conference on Management of data*, pages 207–216, 1993.

[2] Rakesh Agrawal and Ramakrishnan Srikant. Fast algorithms for mining association rules. In *Proc. 20th int. conf. very large data bases, VLDB*, volume 1215, pages 487–499. Santiago, Chile, 1994.

[3] Takuya Akiba, Shotaro Sano, Toshihiko Yanase, Takeru Ohta, and Masanori Koyama. Optuna: A next-generation hyperparameter optimization framework. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, pages 2623–2631, 2019.

[4] Toshiki Aoba, Masato Kikuchi, Mitsuo Yoshida, and Kyoji Umemura. Improving association rule mining for infrequent items using direct importance estimation. In *2020 7th International Conference on Advance Informatics: Concepts, Theory and Applications (ICAICTA)*, pages 1–5. IEEE, 2020.

[5] T.C.Bell I.H.Witten, A.Moffat. *ManagingGigabytes*. Morgan Kaufman, second edition, 1999.

[6] Kento Kawakami, Masato Kikuchi, Mitsuo Yoshida, Eiko Yamamoto, and Kyoji Umemura. Finding association rules by direct estimation of likelihood ratios. In *2017 International Conference on Advanced Informatics, Concepts, Theory, and Applications (ICAICTA)*, pages 1–5. IEEE, 2017.

[7] Masashi Sugiyama, Taiji Suzuki, and Takafumi Kanamori. *Density ratio estimation in machine learning*. Cambridge University Press, 2012.

[8] 菊地真人, 山本英子, 吉田光男, 岡部正幸, and 梅村恭司. 条件付き確率の保守的な推定. **電子情報通信学会論文誌** *D*, 100(4):544–555, 2017.

[9] 山本英子 and 梅村恭司. コーパス中の一対多関係を推定する問題における類似尺度. **自然言語処理**, 9(2):45–75, 2002.

付　録

表 A·1　部分正解集合を用いて推定した Apriori パラメータの平均値

削減する割合	0%	10%	20%	30%	40%	50%	60%	70%	80%	90%
1991	3.000	2.600	2.570	2.573	2.635	2.772	2.913	2.977	3.075	3.247
1992	3.000	3.000	3.020	3.060	3.080	3.128	3.133	3.149	3.168	3.238
1993	3.000	2.640	2.640	2.653	2.640	2.660	2.637	2.629	2.630	2.693
1994	3.000	3.000	3.020	3.067	3.095	3.124	3.133	3.149	3.155	3.249
1995	2.000	2.000	2.000	2.000	2.000	2.004	2.007	2.009	2.033	2.142
1996	2.000	2.420	2.500	2.487	2.480	2.484	2.487	2.486	2.498	2.600
1997	3.000	2.980	2.950	2.913	2.885	2.872	2.867	2.846	2.848	3.042

表 A·2　部分正解集合を用いて推定した Apriori パラメータの分散

削減する割合	0%	10%	20%	30%	40%	50%	60%	70%	80%	90%
1991	0	0.245	0.253	0.371	1.661	2.467	3.506	3.704	4.839	6.159
1992	0	0.000	0.039	0.123	0.164	0.385	0.545	0.676	1.071	3.673
1993	0	0.235	0.235	0.222	0.245	0.196	0.255	0.412	0.480	6.939
1994	0	0.000	0.039	0.137	0.151	0.227	0.314	0.431	0.531	16.449
1995	0	0.000	0.000	0.000	0.000	0.020	0.020	0.020	0.367	22.428
1996	0	0.249	0.249	0.253	0.253	0.255	0.255	0.255	0.575	11.840
1997	0	0.020	0.075	0.137	0.204	0.232	0.464	0.655	1.511	43.061

表 A·3　部分正解集合を用いて推定した L1 パラメータの平均値

削減する割合	0%	10%	20%	30%	40%	50%	60%	70%	80%	90%
1991	1.641	1.563	1.573	1.568	1.546	1.576	1.707	1.824	1.626	2.810
1992	1.478	1.448	1.441	1.459	1.520	1.551	1.615	1.636	1.492	2.103
1993	1.286	1.287	1.227	1.326	1.355	1.382	1.436	1.448	1.327	1.699
1994	1.303	1.252	1.253	1.265	1.298	1.317	1.445	1.407	1.267	1.769
1995	0.787	0.755	0.751	0.770	0.771	0.792	0.808	0.812	0.790	1.627
1996	0.813	0.814	0.799	0.805	0.915	0.954	1.330	1.184	1.211	1.876
1997	0.822	0.792	0.789	0.787	0.920	0.946	1.345	1.413	1.247	1.841

表 A·4　部分正解集合を用いて推定した L1 パラメータの分散

削減する割合	0%	10%	20%	30%	40%	50%	60%	70%	80%	90%
1991	0.000	0.007	0.012	0.007	0.013	0.017	0.005	0.118	0.039	1.814
1992	0.000	0.008	0.015	0.011	0.002	0.002	0.006	0.015	0.132	1.260
1993	0.000	0.004	0.041	0.001	0.005	0.003	0.003	0.003	0.089	0.507
1994	0.000	0.000	0.002	0.003	0.003	0.006	0.005	0.009	0.097	0.532
1995	0.000	0.000	0.002	0.001	0.003	0.000	0.000	0.000	0.002	3.655
1996	0.000	0.001	0.001	0.001	0.074	0.087	0.080	0.111	0.146	2.641
1997	0.000	0.000	0.002	0.001	0.097	0.094	0.095	0.104	0.195	1.703

表 A·5　部分正解集合を用いて推定した L2 パラメータの平均

削減する割合	0%	10%	20%	30%	40%	50%	60%	70%	80%	90%
1991	6.448	6.532	6.356	6.140	7.102	6.807	8.719	8.693	8.759	12.156
1992	4.884	4.439	4.824	4.972	5.344	5.426	7.106	6.708	6.446	8.913
1993	3.662	3.885	3.929	4.055	4.254	4.442	5.386	5.173	5.262	6.085
1994	3.453	3.366	3.340	3.498	3.800	3.919	5.078	4.686	4.341	6.738
1995	2.025	1.550	1.534	1.428	1.654	1.577	2.387	2.631	2.285	3.876
1996	3.008	3.124	3.232	3.236	3.498	3.574	4.536	4.646	4.560	6.582
1997	2.902	3.219	3.324	3.650	3.894	4.042	5.520	5.246	4.721	6.705

表 A·6　部分正解集合を用いて推定した L2 パラメータの分散

削減する割合	0%	10%	20%	30%	40%	50%	60%	70%	80%	90%
1991	0.000	0.694	1.356	0.193	0.848	1.129	1.682	2.351	10.101	28.626
1992	0.000	0.788	0.778	0.458	0.234	0.699	0.586	0.662	2.563	30.834
1993	0.000	0.139	0.235	0.092	0.271	0.444	0.844	1.047	2.467	6.654
1994	0.000	0.033	0.256	0.111	0.212	0.333	0.211	0.322	2.154	8.162
1995	0.000	0.262	0.252	0.220	0.339	0.200	0.185	0.279	1.479	11.525
1996	0.000	0.389	0.248	0.078	0.115	0.100	0.430	0.303	2.532	20.436
1997	0.000	0.296	0.687	0.044	0.531	0.836	0.850	0.601	6.688	16.828