

# 潜在イベント系列コーパスの自動構築を目的とした負の相関ルールマイニングのオンライン高速化

研究代表者 岩 沼 宏 治 山梨大学 大学院医学工学総合研究部 教授  
共同研究者 山 本 泰 生 山梨大学 大学院医学工学総合研究部 助教

## 1 はじめに

時系列上の現象やイベント間の関係を考える際には、裏に隠れている潜在因子を考慮することが大変重要である。本研究では、その潜在因子も明示的に記述するイベント系列コーパスの自動構築を最終的な目標として、潜在因子の抽出法として負の相関ルールマイニングに着目し、その効果的な計算法について研究を行った。まず、負の相関ルールマイニングの完全かつ効率的なオフライン型アルゴリズムの基礎を確立し、次にオンライン化を試みた。主な結果として、オフライン型アルゴリズムは、頻出アイテム集合の接尾辞木上のトップダウン探索法を開発することができた。オフライン型アルゴリズムとしては、頻出アイテム集合をトランザクション・ストリームから抽出する資源指向型のオンライン近似計算法を提案している。両アルゴリズムの理論的性質の解明、及び実データに対する実証的な性能評価の両面から研究を実施した。これらを通して、良好な性能を得ることができたので報告する。なお、前者のオフライン型アルゴリズムは人工知能学会論文誌上で発表し、後者のオンライン型近似アルゴリズムは、2014 ACM SIGMOD 国際会議に採録され、発表を行っている。

## 2 研究の背景と目的

### 2-1 研究の背景

現実の世界を考えると、潜在因子までも考えた相関ルールの自動抽出は非常に重要である。しかし、これまでのデータマイニングの研究では、潜在的相関ルールの研究は一部を除いて殆ど研究されてこなかった。潜在因子の予測と抽出は、統計的学習での潜在パラメータの推定問題などとは異なる種類の問題であり、その効果的な計算はかなり難しい。関連するデータマイニングの研究としては、負の相関ルールマイニングの研究があげられる。この負の相関ルールのマイニング技術を発展させ、潜在因子の相関ルールのマイニング法を開発することは大変重要なことと考えられる。

負の相関ルールとは、 $X$  と  $Y$  をアイテム集合とするとき、 $\neg X \Rightarrow Y$  (左否定形) または  $Y \Rightarrow \neg X$  (右否定形) のような形のルールのことであり (\*注1)、 $\neg X$  は負のアイテム集合と呼ばれる。ルール  $\neg X \Rightarrow Y$  は「 $X$  が出現しない場合に  $Y$  がよく出現する」ことを意味している。このアイテム集合  $X$  は「 $Y$  の出現に大きな影響を及ぼす隠れ因子」と考えることができる。その効果的な抽出法は非常に重要であるが、出現しないアイテム集合を検出する必要があり、効率的な計算法の開発は容易ではない。

正の相関ルール  $X \Rightarrow Y$  のマイニングは、Agrawal らによる相関ルール抽出法 Apriori [AS94] の提案以後、多くの研究が行われてきた。FP-growth 法 [HPY00] による分割統治法や、頻出飽和アイテム集合などによる抽出データの圧縮などの第2世代の技術が開発され、マイニングの対象となるデータも、グラフや数値データなど多様な形式が考察されている。多系列データの集合を取り扱う手法としては、頻出系列を抽出するものとして Prefix-Span 法、頻出飽和系列を抽出するものとしては CoSpan 法や BIDE 法などが開発されている。一方で、負の相関ルールは 1990 年代末から研究 [AY 98, BMS97, SON98] が開始されている。一般に、負の相関ルールは非常に多数存在し、正ルールと比較しても、その数は格段に多い。そのため、これまでは有用な負ルールを絞り込むための評価尺度について、多くの研究 [AY98, AZ04, BMS97, SON98, YBYZ02] が成されている。残念ながら、それらを応用した形での、負の相関ルールの効率的で高速な抽出計算法はあまり研究されていないが、先行研究 [WZZ04, CYZC06, ZY07] では、Apriori 流 [AS 94] のボトムアップ型の抽出計算法を提案している。これらの手法は、負の相関ルールの膨大な台集合を明示的に生成してルール抽出を行っており、現在のデータマイニングの技術水準からみると非常に効率が悪い。これに対して [WZC08] では、頻出アイテム集合の組合せから負の相関ルールを抽出し、負ルールの台集合の明示的な生成を避けており、高

速化に成功している。しかし、やはりボトムアップ型の計算法を採用しており、負ルールの相互関係の検査が難しいために、やはり計算効率はまだ低いと考えることができる。これまでところ、完全かつ効果的な負の相関ルールのマイニングアルゴリズムは明らかになっていないと考えられる。

## 2-2 研究の目標

本研究では、負の相関ルールの高速かつ効果的な抽出計算法の開発を目的としている。具体的には以下の課題について研究を行った。

- (A) トランザクション・データベースからの負の相関ルールの高速マイニングのための基盤開発
- (B) トランザクション・データストリームからの負ルールのオンライン抽出計算法の開発

トランザクションとはアイテムの集合のことであり、トランザクション・データベースとはトランザクションの多重集合のことである。頻出アイテム集合とは、データベース中のトランザクションに閾値以上の割合で出現するアイテム集合のことを言う。トランザクション・データベースに出現するアイテムの種類数を  $N$  とするとき、頻出アイテム集合の候補は単純には  $2^N - 1$  個存在することになり、 $N$  が大きくなるとその取扱いは非常に困難になることが知られている。正の相関ルール  $X \Rightarrow Y$ 、負の相関ルール  $\neg X \Rightarrow Y$ 、 $Y \Rightarrow \neg X$  の  $X$  と  $Y$  は、通常、頻出アイテム集合に限定して考えるのが自然と認識されている。データベースから正ルールを抽出する場合には、頻出アイテム集合だけを組み合わせるルールを生成すればよいのは明らかであるが、負ルールを抽出する場合には、否定集合を  $\neg X$  を考慮する必要がある。先行研究 [WZZ04, CYZC06, ZY07] では、負ルールの台集合として非頻出アイテム集合を考え、ルール生成アルゴリズムの中で陽に非頻出アイテム集合を抽出して、負ルールを生成している。しかし非頻出集合は膨大な数が存在するので、これらの負ルール生成手法は非常に効率が悪いことは明らかである。これに対して [WZC08] では、頻出アイテム集合の組合せだけから負の相関ルールを抽出し、効率の改善を行っている。しかし [WZC08] を含めた全ての先行研究は、Apriori アルゴリズムに準拠したボトムアップ方式を取っている。これは正ルール抽出で初期に用いられた方式であり、幾つかの本質的な欠点知られており、現在ではそれらを克服したトップダウン型のルール生成方式の方が効率的であるとされている。

そこで本研究では、課題(A)に対して「頻出アイテム集合だけを用いて負の相関ルールを効果的に生成するためのオフライン・トップダウン型のアルゴリズムの開発」を目標として、研究を実施した。Apriori 流のボトムアップ型生成では、負ルール相互の関係の検査が非常に難しい。このため、無駄なルール生成処理を除去することが困難であり、高速化の大きな障害になっていた。本研究では、頻出アイテム集合を頂点とする接尾辞木を用いたトップダウン型の負ルールの生成法を新たに提案する。課題(A)に対する提案手法とその価結果の詳細は、次の3章において説明する。

トランザクション・ストリームとは、トランザクションが次々に到着するデータストリームのことであるが、そこから負ルールをオンラインで抽出することができれば、大変有用である。課題(A)に対して開発した手法は「頻出集合だけを使って負ルールを抽出」するものであるが、これはトランザクション・ストリームのオンライン型処理に本質的に適した性質である。ただオンライン処理を実現するためには、負ルールの抽出に先行して、その土台となる頻出アイテム集合をオンラインで抽出する必要がある。予備的考察の結果、トランザクション・ストリームから頻出アイテム集合をオンライン抽出する計算問題は、「その時間計算量の下限がストリーム上のトランザクションの最大長の指数時間となる」ことを新しく証明することができた。これは、当該問題が本質的に非常に難しい問題であることを示唆しており、トランザクション・ストリームから負ルールをオンライン抽出するにあたって、まず頻出アイテム集合をオンライン抽出する実際的な計算法を開発することが最大の課題となることが、本研究により新たに判明した。このため本研究では、課題(B)の研究を遂行するために、まずトランザクション・ストリームデータから頻出アイテム集合をオンライン抽出するために効果的な技術の開発に取り組んだ。その詳細は以下の4章において説明を行う。

## 3 負の相関ルールの高速マイニングのための基盤開発

本研究では、頻出集合だけを用いたトップダウン型の負ルールの生成法を新たに提案する。また、この準備として、生成すべき“妥当な負ルール”の定義について初めに考察を行った。先行研究 [WZZ04, CYZC06, WZC08] において考えられている正と負ルールの妥当性には、 $X \Rightarrow Y$  と  $X \Rightarrow \neg Y$  が同時に抽出される矛盾が生じる可能性があることが、考察の結果明らかとなった。本研究では、このような矛盾を回避するための新しい制約条件を導入している。我々の知る限りにおいて、生成する正と負のルールの間の無矛盾性を明示的に考慮した研究はこれまでに無く、本研究が初めてと考えられる。

以上の“負ルールの妥当性”の条件のもとに、全ての妥当な負ルールを高速に生成することを考える。生成するルールの集合には、全てのルールが陽に含まれる必要は無いこと、即ち、ある種の圧縮を行ってルールの集合を生成することが可能である。頻出アイテム集合に関する圧縮形式としては、頻出極大集合と頻出飽和集合がよく知られている。飽和集合の枠組みを用いると、アイテム集合の出現頻度の情報まで保存できるが、相関ルール、特に負ルールの飽和性の性質は全く調べられておらず、適用が困難である。そこで本研究ではまず、負ルールの極小性に注目してルール集合の圧縮を図った。例えば、 $X$  と  $Y$  を頻出アイテム集合とすると、右否定形  $Y \Rightarrow \neg X$  の妥当な負ルールには、包含関係の意味で極小な負ルールが必ず存在することが知られている。これは右否定形の負ルールのマイニング問題を、極小で妥当な負ルールの抽出問題に置き換えられることを意味し、アルゴリズムの高速化に極めて有効である。但し、その実現のためには負ルールの間の包含関係を効果的に検査する必要がある。この包含性の検査は、我々が考慮するトップダウン型探索の枠組みの中で効果的に実現できる。本研究では、頻出アイテム集合を節点に持つ接尾辞木の利用を考える。接尾辞木上で左優先深さ優先探索を行うと、訪問頂点  $N$  を訪問した時点では、 $N$  が表現するアイテム集合の全ての部分集合の訪問はそれ以前に完了しているという有用な性質 [KS12] があり、極小性の検査を非常に効率的に実現できる。この極小性検査を行うトップダウン型探索システムを試験的に実装し、性能評価実験を行ったところ、100 倍から 1000 倍に及ぶ高速化の効果を確認することができた。これらの中間的結果は 2013 年人工知能学会全国大会において発表 [発表資料 3] している。

左否定形  $\neg X \Rightarrow Y$  の負ルールは、右否定形と異なり、通常のルール評価尺度は逆単調性を満たさないことが多い。そのため極小なルールも存在せず、右否定形の負ルールで利用した手法は適用できない。そこで本研究では、新たに接尾辞木上の分枝限定法の導入を試みた。逆単調でない評価尺度に対して、逆単調性を満たす上界関数となる尺度が存在すれば、それを用いて分枝限定法を効果的に実行できることが既に知られている [MS00]。本研究では、左否定形の負ルールの確信度などの評価尺度に対して、逆単調性を満たす上界関数を新たに考案し、それらを利用した分枝限定型の探索アルゴリズムを新しく提案した。注意すべき点は、左否定形と右否定形の負ルールを別々に生成すること、即ち、接尾辞木全体を 2 回探索して別々に生成することは無駄が多い点である。本研究で提案する探索アルゴリズムでは、負ルールの妥当性の検査手順を工夫し、重複なしに 1 回の探索で左右の妥当な負ルールを同時に生成することができる。これらのアルゴリズムを実装し、性能評価実験を行った結果、良好な性能を確認することができた。以上の成果を取りまとめたものは人工知能学会論文誌に原著論文として採録され、発表 [発表資料 1] している。

#### 4 トランザクション・データストリームからの負ルールのオンライン抽出計算法の開発

2 章で述べた負ルールのオフライン型抽出アルゴリズムは「頻出集合だけを使って負ルールを抽出」するが、これはトランザクションが次々に流れるストリームデータ上でのオンライン型処理に本質的に適した性質である。負ルールを非頻出アイテム集合から生成する場合は、非頻出アイテム集合の数が多すぎて、オンライン型の処理は殆ど不可能と考えられる。但し、実際に負ルールをオンライン処理で抽出するためには、負ルールの抽出に先行して、その土台となる頻出アイテム集合をオンラインで抽出する必要がある。

トランザクションに含まれるアイテムの数を  $n$  とするとき、頻出アイテム集合の候補は単純には  $2^n - 1$  個存在し、 $n$  が大きくなるとそれら全体のリアルタイム処理は非常に困難になる。我々が予備的な考察を行った結果、トランザクション・ストリームからの頻出アイテム集合を抽出する問題の時間計算量の下限は、 $L$  をトランザクション・ストリーム中に出現する各トランザクションの最大長（即ち、各トランザクションが含むアイテムの最大数）とすると、 $2^L$  という指数オーダーとならざるを得ないことが明らかになった。この指数時間計算量の壁は、本質的に対応が非常に難しい。この問題の実際的な解決法の開発が、頻出アイテム集合、ひいては負ルールのオンライン抽出計算のための最大の課題となることが明らかになった。

本研究ではこの問題に対処するために、まず抽出した頻出アイテム集合を順次、非可逆型圧縮する準オンライン型の抽出アルゴリズムを開発し、計算時間と抽出データの指数オーダ的な増加へ対応を試みた。評価実験を行った結果、一定の効果は確認できたが、同時に限界も明らかになった[発表資料 4]。そこで本研究では、課題(B)を解決するための中心的な部分課題として、出力された頻出アイテム集合の圧縮ではなく、トランザクション・ストリームの個々のトランザクションの効果的な近似計算法の開発を目標として、以下の2つの課題について研究を進めた。

- (B-1) トランザクション・ストリームデータのオンライン型の縮退近似計算法
- (B-2) 計算資源指向型の頻出アイテム集合のオンライン近似抽出計算法

#### 4-1 トランザクション・ストリーム縮退手法の開発

頻出アイテム集合は必ず頻出アイテムだけを要素として構成される。そのため、個々のトランザクションから事前に非頻出アイテムを削除し、そこから候補アイテム集合を抽出すれば、その数を大幅に減らすことができる。その実現のためには、トランザクション・ストリーム中の頻出アイテムをオンラインで高速に識別抽出する必要がある。現在までのところ、使用可能なメモリ容量をある定数に固定した下で、頻出アイテムをトランザクションから識別・抽出するオンライン計算法は提案されていない。そこで本研究では、まず上記の(B-1)の課題について研究を行った。トランザクション(即ちアイテム集合)ではなく、アイテムから成るストリームデータから頻出アイテムをオンライン抽出する近似計算法、既に数多くの既存研究がある。その中で、Metwally 他 の Space-Saving 法 [MAA05] はメモリの使用効率及び計算スループットなどの観点から、現時点で最善のアルゴリズムと評価されている [CH10]。我々は、この Space-Saving を拡張し、アイテムではなくトランザクション、即ちアイテム集合が次々に到着するストリームデータ上で、頻出アイテムをオンラインで抽出する近似計算法を開発している [発表資料 5]。

#### 4-2 計算資源指向型のトランザクション・ストリームからの頻出アイテム集合のオンライン抽出

上で述べた手法により、ストリーム上の個々のトランザクションの即時の縮退処理が容易になるが、縮小しても、まだトランザクションが大きすぎる状況は多々生じる。トランザクションが大きい場合、それに含まれる部分アイテム集合は組合せ爆発により膨大な数となってしまふ。このような場合、まずメモリが足りなくなり、マイニングシステムは異常停止せざるを得なくなる。これに対処するために、上記の(B-2)の研究課題、即ち、使用可能メモリ量を事前に定数値  $K$  に固定し、メモリの使用状況とトランザクションの状態を常に観察しながら、状況に応じて候補アイテム集合を積極的に棄却・スキップする適応的なオンライン型近似計算法を開発した。

本研究で開発した近似算法の基本は、Space-Saving 法と Manku 他 の Lossy Counting 法 [MM02] を融合させたもの(以下、LC-SS 法と呼ぶ)である。LC-SS 法では、ストリーム上の各トランザクションからその部分集合、即ち頻出アイテム集合の候補を順次抜出して、メモリ上のエントリ表にその出現頻度と共に記憶する。ストリームの処理が進んでくると、記録しておくべきアイテム集合の数が増え、エントリ表が拡大していく。ある時点で、予め容量を  $K$  に固定したメモリ空間にエントリ表が納まらない状況となる。LC-SS 法では、このような状況においては Space-saving 法と同様に、エントリ表の中で出現頻度が最小のエントリ、即ち最小の出現回数を持つ候補アイテム集合を削除して、強制的に空きスペースを作り、そこに新しい候補集合を登録する。この際に削除したエントリの出現頻度を、その時点での出現頻度の最大可能誤差として記録しておく。以下、エントリ表に新しく記録されるアイテム集合の出現頻度は、その時点の最大可能誤差に 1 を加えたものとして、エントリ表に記録される。これにより、各時点での頻出アイテム集合は、その時点での最大可能誤差が閾値(最小サポート値)を超えない限り、常にエントリ表中に登録されていることが保証(完全性)できる。またメモリ使用量を定数値に固定しているが、エントリ表中のアイテム集合の出現頻度値の誤差も、最大可能誤差以下であることが保証できる。

LC-SS 法の欠点は、ストリーム上の各トランザクションの部分アイテム集合を全て一度は調べる点にある。このため各トランザクションの処理に要する処理時間が指数関数的に増加する現象が頻出する。その解決法として、本研究ではスキップ動作を導入した Skip LC-SS 法を開発した。Skip LC-SS 法はオンライン処理す

べきトランザクションの大きさと現時点でのエントリ表の内部状況を比較し、適宜、トランザクション中の部分アイテム集合の処理をスキップして省略中止する。これにより  $L$  をトランザクションの大きさ（即ち、内部のアイテムの数）とするとき、個々のトランザクション処理に要する時間を  $O(K \cdot L)$  という線形時間に抑え込むことが可能になる。

以上、Skip LC-SS 法の完全性を保証する条件、および線形の時間計算量と定数値となる空間計算量、更に頻出アイテム集合として出力した集合の出現頻度の誤差を理論的に保証すると同時に、多くの実データを用いて実証実験を行い、提案手法の有効性を確認することができた。既存の関連手法である Stream Mining [JA05] や Adaptive Frequency Counting [LHKC07] 等と比較しても良好な結果を得ている。この研究成果は、データ科学・工学の分野で世界のトップ会議である 2014 ACM SIGMOD に採録され発表を行っている [発表資料 2]。

本研究では、研究計画申請時には、トランザクション・ストリームから負ルールをオンライン抽出するアルゴリズムとシステムを最終的に開発する予定を立てていたが、残念ながら当初の研究予定期間内には完了せず、今後の課題として残った。現在、研究を継続して精力的に進めている途中であり、結果がまとまり次第、適宜発表していく予定である。

## 【参考文献】

- [AS 94] R. Agrawal and R. Srikant: Fast Algorithm for Mining Association Rules. *Proc. 20th Intl. Conf. on Very Large Data Bases*, pp.487–499 (1994)
- [AY 98] C. C. Aggarwal and P. S. Yu: A New Framework for Itemset Generation. *Proc. 17th ACM SIGACT-SIGMOD-AIGART Symp. on Principles of Database Systems*, pp.18–24 (1998)
- [AZ04] M.-L. Antonie and O. R. Zaïane: Mining Positive and Negative Association Rules: An Approach for Confined Rules. *Proc. 8th Euro. Conf. on Principles and Practice of Knowledge Discovery in Databases*, pp.27–38 (2004)
- [BMS97] S. Brin, R. Motwani and C. Silverstein: Beyond Market Baskets: Generalizing Association Rules to Correlations, *Proc. ACM SIGMOD Conf.*, pp. 265-276, May 1997.
- [CH10] G. Cormode and M. Hadjieleftheriou: Methods for finding frequent items in data streams, *Int. J. on Very Large Data Bases*, Vol.9, No.1, pp.3–20 (2010).
- [CYZC06] C. Cornelis, P. Yan, X. Zhang and G. Chen: Mining Positive and Negative Association Rules from Large Databases. *Proc. IEEE Int. Conf. on Cybernetics and Intelligent Systems (CIS 2006)*, pp.613–618 (2006)
- [HPY00] J. Han, J. Pei and Y. Yin: Frequent Patterns without Candidate Generation. *Proc. 2000 ACM-SIGMOD Intl. Conf. on Management of Data*, pp.1–12 (2000)
- [JA05] R. Jin and G. Agrawal: An algorithm for in-core frequent itemset mining on streaming data, *Proc. of Intl. Conf. on Data Mining (ICDM)*, 210–217 (2005).
- [KS12] Y. Kameya and T. Sato: RP-growth: Top-k Mining of Relevant Patterns with Minimum Support Raising. *Proc. of 2012 SIAM Intl. Conf. on Data Mining (SDM 2012)*, pp.816-827 (2012)
- [LHKC07] B. Lin, W.-S. Ho, B. Kao and C.-K. Chui: Adaptive frequency counting over bursty data streams, *Proc. of the 2007 IEEE Symp. on Computational Intelligence. and Data Mining (CIDM)*, 516–523 (2007).
- [MAA05] A. Metwally, D. Agrawal and A. E. Abbadi: Efficient computation of frequent and top-k elements in data streams, *Proc. of the 10th Int. Conf. on Database Theory (ICDT)*, pp.398–412 (2005).
- [MM02] G. S. Manku and R. Motwani: Approximate frequent counts over data streams, *Proc. of the 28th Int. Conf. on Very Large Data Bases (VLDB2002)*, pp.346–357 (2002).
- [MS00] S. Morishita and J. Sese: Traversing itemset lattices with statistical metric pruning. *Proc. 19th ACM SIGACT-SIGMOD-AIGART Symp. on Principles of Database*, pp.50-60 (2000)

- [SON98] A. Savasere, E. Omiecinski and S. Navathe: Mining for Strong Negative Associations in a Large Database of Customer Transactions. *Proc. Intl. Conf. on Data Engineering*, pp.494-502 (1998)
- [YBYZ02] X. Yuan, B. P. Buckles, Z. Yuan and J. Zhang: Mining Negative Association Rules. *Proc. 7th Intl. Symp. on Computers and Communication*, pp.623-629, (2002)
- [WZC08] H. Wang, X. Zhang and G. Chen: Mining a Complete Set of Both Positive and Negative Association Rules from Large Databases, *Proc. 12th Pacific-Asia Conf. on Advances in Knowledge Discovery and Data Mining (PAKDD 2008)*, LNAI, Vol.5012, pp.777-784 (2008)
- [WZZ04] X. Wu, C. Zhang and S. Zhang.: Efficient Mining of Both Positive and Negative Association Rules. *ACM Transactions on Information Systems*, Vol.22, No.3, pp.381-405, July 2004.
- [ZY07] L. Zhou and S. Yau: Efficient Association Rule Mining among Both Frequent and Infrequent Items. *Computers and Mathematics with Applications*, Vol.54, pp.737-749 (2007)

(+注1) 先行研究[WZZ04,CYZC06, WZC08, ZY07] では、負の相関ルールの形は左否定形 $\neg X \Rightarrow Y$ , 右否定形 $X \Rightarrow \neg Y$ の他に、両否定形 $\neg X \Rightarrow \neg Y$ の3種類を対象としている、本研究では両否定形の負ルールは考察の対象としていない。両否定形のルール $\neg X \Rightarrow \neg Y$ は支持度と確信度による選別機構が有効に働かず、その数は左否定形および右否定形の負ルールと比較しても、通常は格段に多くなる。また、前件 $\neg X$ と後件 $\neg Y$ の間の統計的関係性を見出すことも一般にはかなり難しく、その観点からも両否定形の負ルールの有効性は低いと考えられる。そこで本研究では左右の否定形の負ルールに焦点を絞り、抽出計算の高速化を考察している。

### 〈発 表 資 料〉

題 名	掲載誌・学会名等	発表年月
1. 負の相関ルールを抽出する高速トップダウン型アルゴリズム	人工知能学会論文誌, Vol29, No. 4, pp. 406-415	2014年6月
2. Resource-oriented Approximation for Frequent Itemset Mining from Bursty Data Streams	Proceedings of the 2014 ACM SIGMOD Intel. Conf. on Management of Data (SIGMOD'14), pp.205-216	2014年6月
3. 可変長トランザクションからなるストリーム上のオンライン型頻出アイテムマイニング	人工知能学会 人工知能基本問題研究会 (第92回) 資料集 SIG-FPAI-B303	2014年1月
4. 極小性を用いた負の相関ルールの効率的な抽出法	人工知能学会全国大会 (第27回) 論文集, 2C1-4	2013年6月
5. 頻出アイテム集合の即時圧縮を行う準オンライン型ストリームマイニング	人工知能学会全国大会 (第27回) 論文集, 2N5-0S-21b-3	2013年6月