

EC サイトにおけるユーザの購入可能性の 推定手法に関する研究

外山 諒[†] 中村 健二[‡] 田中 成典^{‡†} 藤本 雄紀^{‡‡}

[†] 関西大学大学院 〒569-1095 大阪府高槻市霊仙寺町 2-1-1

[‡] 大阪経済大学 〒533-8533 大阪市東淀川区大隅 2-2-8

^{‡†} 関西大学 〒569-1095 大阪府高槻市霊仙寺町 2-1-1

^{‡‡} 関西大学大学院 〒569-1095 大阪府高槻市霊仙寺町 2-1-1

E-mail: [†] k943166@kansai-u.ac.jp, [‡] k-nakamu@osaka-ue.ac.jp, ^{‡†} tanaka@res.kutc.kansai-u.ac.jp, ^{‡‡} k398881@kansai-u.ac.jp

あらまし 電子商取引サイトが増加し、ユーザの定着率を高めることや購入の動機付けを与えるマーケティングの重要性が高まっている。しかし、マーケティングには蓄積された情報を利用することが多く、新たにアクセスしたユーザに対する分析が難しい。特に、ユーザの購入可能性を推定する分析には、ユーザの関心や興味、ニーズといった情報が必要であるため、既知のユーザ以外の分析は難しいという課題があった。そこで本研究では、リアルタイムバースト検出手法を拡張して、蓄積された情報を用いずにユーザを分析するための機能を追加する。これを用いて、新たにアクセスしたユーザに対しても適用できる購入可能性の推定手法を開発し、評価実験によりその有用性を明らかにする。

キーワード EC サイト, リアルタイムバースト, データマイニング, 購買

1. はじめに

スマートフォンをはじめとした通信機器の普及によって、インターネットを介した商取引が一般化し、電子商取引サイト（以下、EC サイト）の数も増加している。それにより、各 EC サイト間での競争が激化しており、売り上げ向上のためのマーケティングが重要となっている。このような状況の中で、アクセスログを解析し、関心や興味、ニーズから購入する可能性のあるユーザを推定して、それに応じたコンテンツやサイト構成に更新するなどの効率的なマーケティング手法が望まれている。そのため、各ユーザの購入可能性を推定する研究[1]-[5]が多くなされている。これらの研究ではアクセスログや購入履歴、会員登録情報などを用いている。特にアクセスログに着目した研究では、過去のアクセス頻度に基づいて購入可能性を推定する手法[4][5]などが一般的に用いられる。また、EC サイトではログの蓄積がない状態では分析ができないというコールドスタート問題がある。この問題に対応するための研究[6]が行われているが、ユーザに着目した分析手法は確立されていない。

そこで、本研究では、アクセスログを大量のデータが次々と流れてくるデータストリームとして解析し、その変化に着目することで購入可能性を推定する手法を開発する。変化を検出する手法として、本研究では、

既存研究[7]にならい、リアルタイムバースト検出手法[8][9]を活用する。これによりログ蓄積のないユーザに対しても購入可能性の推定が可能な手法を提案する。

2. 研究の概要

2.1. アクセスログにおけるバーストの概要

本研究では、アクセスログを時系列データの分析に用いられるデータストリームとして解析する。データストリームは、時間情報を持った個別のデータが連続して到着するが、到着間隔は不均一である。そのため、データが全く到着しない疎な時間と、反対にデータが間断なく到着する密な時間が発生する。この密な時間の発生（以下、バースト）を検知する研究[7]-[10]が行われている。

アクセスログをデータストリームとする場合、各ページへのアクセスをデータの到着とみなすことで、アクセスの集中をバーストとして検出できる。特に、本研究対象の EC サイトにおける任意ユーザのアクセスに伴うバースト発生は、短時間に多くの商品に関する情報を収集する行動が要因である可能性が高く、商品に対する興味や関心が高まっていることを示すものと判断できる。このことから、バーストを検出することで、関心や興味の高まった購入する可能性のあるユーザを推定可能であると考えられる。そこで、本研究で

は、バーストを分析する手法の一つであるリアルタイムバースト検出手法[7]を用いて、購入可能性を推定する。これによりリアルタイムにバーストの発生が検出可能となり、商品推薦といった即時性の必要なマーケティングにも対応できると考えられる。

2.2. リアルタイムバースト検出手法をアクセスログの解析に用いる際の課題と対応方策

既存のバースト解析手法[7]をアクセスログの解析に適用する場合、次に示す二つの課題が発生する。

一点目は、ユーザごとのアクセス頻度が異なるため、バーストを検知する期間（以下、ウィンドウサイズ）が同一の場合、バーストの発生を適切に検出できない点である。アクセスログに対して、図1右の様にウィンドウサイズを大きく設定すると、全てのアクセスを内包し、バーストの発生時期を適切に検出できない可能性がある。一方、図1左の様に、ウィンドウサイズを小さく設定すると、バーストとなるアクセス数に不足する可能性があり、バーストの検知が困難となる。そのため、ユーザごとのアクセス傾向に応じたウィンドウサイズを適切に設定する必要がある。この課題には、ユーザのアクセス頻度に応じて、ウィンドウサイズを動的に変更する機能を追加することで対応する。これにより、ユーザごとに適切なウィンドウサイズを設定することで、バースト検知数を向上させる。

二点目は、ユーザが一つのページを長い時間かけて閲覧している場合、イベントが発生しない点である。既存のバースト解析手法をアクセスログの解析に適用する際、ユーザがドメイン内の別のページへアクセスするページ遷移を一つのイベントとして解析する。この際、ユーザが一つのページの閲覧を長く続ける行為に対してはイベントが発生しないため、バーストが発生しない状況となる。しかし、ユーザが一つのページに長く滞在することは、該当ページのコンテンツに関心があると判断できるため、購入可能性の高い行動とみなすことができる。そのため、この行動を考慮したバースト解析手法を考案する必要がある。この課題には、ユーザのページ遷移間の閲覧行動を推定し、イベントを発生させる機能を追加することで対応する。これにより、一つのページに長く滞在するユーザにイベントを補完し、バーストの発生数を増加させる。

2.3. 提案手法の概要

本研究では既存のバースト解析手法[9]に対し、ウィンドウサイズを動的に変更する機能と、閲覧行動を推定し、イベントを発生させる機能を追加することで、アクセスログから購入可能性を推定する手法を開発する。

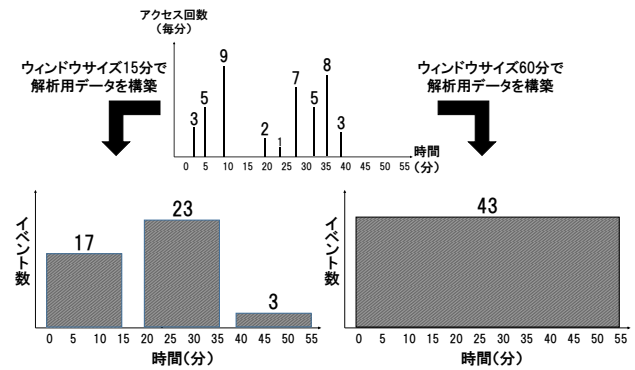


図1 ウィンドウサイズによる違い

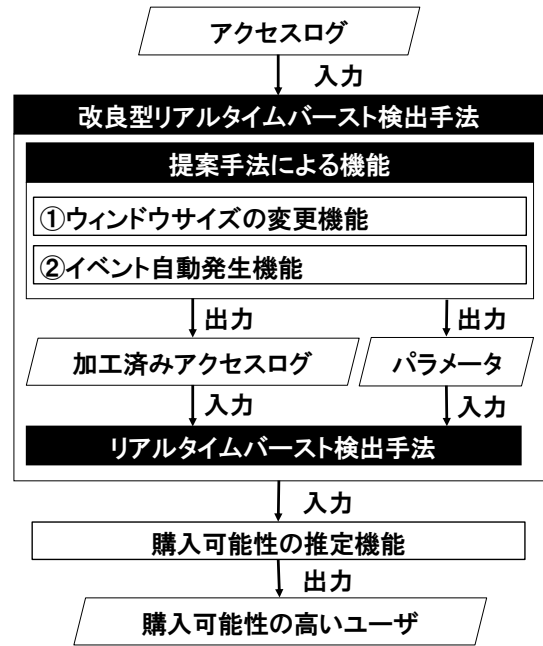


図2 提案手法の処理の流れ

提案手法の処理フローを図2に示す。提案手法では既存のバースト解析手法にウィンドウサイズ変更機能、イベント自動発生機能、購入可能性の推定機能を追加する。入力データは既存のバースト解析手法と同様にアクセス（イベント）のデータを用いる。なお、リアルタイムバースト検出手法の詳細については、文献[9]を参照されたい。

2.4. ウィンドウサイズ変更機能

ウィンドウサイズ変更機能は入力されたアクセス情報に基づき、ユーザごとのアクセス総数からウィンドウサイズを変更する。既存手法で用いられているウィンドウサイズの最小値 W_{min} および最大値 W_{max} を各ユーザのアクセス頻度ごとに決定することで、意図しないバーストの発生を抑制できると考えられる。本機能では、 W_{min} および W_{max} は同値とするが、ウィンドウサイズは複数を用意し、ユーザごとに適切な値

を判定して適用する．その結果，ユーザごとのバースト検出精度の向上が見込める．ウィンドウサイズの自動判定の指標には，ユーザのアクセス総数，アクセス時間帯別のアクセス総数の二つの指標のいずれかを用いる．過去にアクセスがあるユーザの場合は，アクセス総数に基づき，ウィンドウサイズを決定する．この際，当該の時間帯のアクセス総数が一定（ AC_u ）以上になった場合，ウィンドウサイズの設定を W_u に更新する．これにより，ユーザごとに適切なウィンドウサイズを常時設定可能であると共に，急激なアクセスの増加によるバースト検出の誤りを抑制できる．

一方，初めてアクセスしたユーザの場合は，規定値（ W_{def} ）を採用する．これにより，ユーザのアクセス総数や時間帯毎のアクセス総数に基づき，ウィンドウサイズを動的に設定することができ，ユーザのアクセス頻度に基づき適切なバースト検知が可能となる．

2.5. イベント自動発生機能

イベント自動発生機能は入力されたアクセス情報に基づき，アクセスのイベントが発生しないアクセス以外の行動を想定し，追加でイベントを発生させる．本機能では同一ユーザの直前のアクセス日時を取得し，現在のアクセス日時までの経過時間を算出する．まず，アクセスの発生を検出する．さらに，同一ユーザの直前のアクセス日時を取得し，現在のアクセスまでの経過時間を算出する．経過時間があらかじめ設定した時間の範囲内である場合，単一ページを長時間閲覧した後の遷移によるアクセスと判断し，新たなイベントを発生させ，アクセスログに格納する．設定した時間の範囲外である場合は閲覧を終了していると判断し，処理を行わない．これにより，ページ遷移の回数が少ないが，長時間にわたり閲覧を続けている購入可能性の高いユーザのイベント発生回数が増加でき，興味や関心に基づくバーストを発生させ易くできる．また，イベントの新規追加を行うために，イベント発生間隔を示すパラメータ $Eterm$ とイベントの追加発生数を算出するためのパラメータ $Etime$ を追加する． $Eterm$ はイベントが自動発生する判定秒数で，ウィンドウサイズとは異なる値を設定する．直前のページ遷移からこの値の時間までに再度ページ遷移が行われた場合，その間はページ上の情報を閲覧していたとし，自動的にイベントを追加する．また， $Etime$ はイベントの自動発生間隔秒数を表し， $Eterm$ と $DT_i - DT_{i-1}$ の比較により，追加のイベント発生が必要と判断された場合，この値で指定された間隔でイベントを発生させる．各ユーザのアクセス回数を i ，そのアクセスの時刻を DT_i ， $DT_i - DT_{i-1}$ 間のイベント数 E_f は式(1)で表される．

$$E_f = \begin{cases} \frac{(DT_i - DT_{i-1})}{Etime} & (if\ DT_i - DT_{i-1} > Eterm) \\ 1 & (if\ DT_i - DT_{i-1} \leq Eterm) \end{cases} \quad (1)$$

2.6. 購入可能性の判定機能

購入可能性の判定機能では，ウィンドウサイズ変更機能により決定されたパラメータおよび，イベント自動発生機能により生成されたアクセスログを対象に，リアルタイムバースト検出手法にてバースト解析した結果から購入可能性を出力する．本研究ではバーストの発生はユーザによる興味や関心に関する行動を示すため，バーストが発生したユーザを購入可能性のあるユーザとして抽出する．

3. 提案手法の評価実験

3.1. 実験目的

本実験の目的は，提案手法の有用性を確認するため，次の2点に関して評価することである．1つ目は，アクセスログの解析のため，機能追加を行った提案手法と既存手法を比較し，提案手法の有効性を評価する．2つ目は，提案手法および一般的にアクセスログ解析に用いられる手法で実際のアクセスログを解析し，提案手法の有効性を評価する．

3.2. 実験内容

本実験では，既存手法と提案手法にて抽出されたユーザが，購入可能性の高いユーザであるかどうかを評価するため，あらかじめ蓄積されたアクセスログを用いて実験を行う．既存手法には，一定期間でのアクセス総数が多いユーザを購入可能性の高いユーザとして判定する方策を採用する．本実験では，商品を買う物かごに追加する行動を購入可能性が高いと判断し，正解データとする．

各手法の有効性は，各手法にて推定した結果と正解データとを比較し，推定結果を購入可能性として利用できるかを適合率，再現率，F値を算出することで評価する．ここで，提案手法はウィンドウサイズを決定する基準により，「ユーザごとのアクセス総数に基づく手法A」と「時間帯毎のアクセス総数に基づく手法B」の2種類の決定基準による抽出結果を用いる．手法Aでは，対象ユーザのアクセス総数を集計し，その値に基づきユーザのウィンドウサイズを設定する．また，手法Bでは対象ユーザのアクセスを時間帯（0時～6時，6時～9時，9時～12時，12時～13時，13時～18時，18時～24時）ごとに集計し，その時間帯毎のアクセス総数に基づき，ウィンドウサイズを設定する．

表 1 実験結果

手法	アクセス 総数	ウィンドウ サイズ	対象 ユーザ数	購入 ユーザ数	時間帯	判定結果	正解数	再現率	適合率	F値
既存 手法	1回以上		803,126	15,456		803,126	15,456	<u>1.0000</u>	0.0192	0.0378
	2回以上		358,350	14,959		358,350	14,959	0.9678	0.0417	0.0800
	3回以上		222,235	14,489		222,235	14,489	0.9374	0.0652	0.1219
	5回以上		123,091	13,188		123,091	13,188	0.8533	0.1071	0.1904
	10回以上		55,795	10,082		55,795	10,082	0.6523	<u>0.1807</u>	<u>0.2830</u>
提案 手法 A	1回以上	60分, 30分, 10分, 5分	803,126	15,456		10,455	8,066	<u>0.5219</u>	0.7715	0.6226
	2回以上		358,350	14,959		10,455	8,066	<u>0.5219</u>	0.7715	0.6226
	3回以上		222,235	14,489		10,425	8,064	0.5217	0.7735	0.6232
	5回以上		123,091	13,188		10,248	8,050	0.5208	0.7855	<u>0.6264</u>
	10回以上		55,795	10,082		8,530	6,858	0.4437	<u>0.8040</u>	0.5718
提案 手法 B	30回未満	30分	1,159	832	0時～6時	999	257	0.3089	<u>0.8317</u>	<u>0.4505</u>
			704	457	6時～9時	701	96	0.2101	0.6076	0.3122
			453	304	9時～12時	451	75	0.2467	0.6098	0.3513
			382	243	12時～13時	381	18	0.0741	0.5806	0.1314
			560	411	13時～18時	559	137	<u>0.3333</u>	0.6587	0.4426
			200	143	18時～24時	199	36	0.2517	0.5806	0.3512
	30回～99回	10分	330	258	0時～6時	330	143	0.5543	<u>0.7259</u>	<u>0.6286</u>
			241	70	6時～9時	121	24	0.3429	0.4800	0.4000
			374	199	9時～12時	374	117	0.5879	0.5021	0.5417
			281	82	12時～13時	161	33	0.4024	0.4342	0.4177
			604	307	13時～18時	523	191	<u>0.6221</u>	0.5306	0.5727
			372	215	18時～24時	372	121	0.5628	0.5426	0.5525

提案手法の対象ユーザはバーストが発生する可能性があるアクセス総数が2回以上あるユーザとする。手法Bでは各時間帯でアクセス総数が100回以上のクローラなどの可能性があるユーザは対象外とした。また、ウィンドウサイズ変更基準であるアクセス総数 AC_u とユーザごとのウィンドウサイズ W_u はアクセス総数の分布を目視で確認し、経験的にバースト発生の可能性が見込まれる値を設定した。初めてアクセスしたユーザのウィンドウサイズ W_{def} は30分とした。

3.3. 実験データ

本実験では、テックジャム社 (<https://www.tech-jam.com>) から提供された同社のECサイトのアクセスログを用いる。アクセスログの期間は、2013年9月26日から2015年8月3日とする。また、アクセスログは、アクセスに使われた端末のIPアドレス、アクセス先URL、アクセス元URL、アクセス日時で構成されており、4,174,521件のPV、803,126件のユーザが含まれている。この中で、買い物かごに追加した件数は15,456件となっており、約2%のユーザが購入した状況である。

3.4. 実験結果と考察

本実験の結果を表1に示す。表において、判定結果は各手法にてバーストが発生したユーザ数、正解数は判定結果の中で購入があったユーザ数を指す。ウィンドウサイズは提案手法Aではアクセス総数 AC_u に基づき、 W_u を60分、30分、10分、5分に設定した。提案手法Bでは時間毎のアクセス総数 AC_u が30回未満のユーザの W_u を30分、30回～99回のユーザの W_u を10分に設定した。実験結果を確認すると次に示す内容が明らかとなった。

(1) 既存手法と比較して提案手法Aは誤判定が少ないことが明らかとなった

各手法の適合率を確認すると、提案手法Aは、既存手法の7倍以上になっている。また、再現率は提案手法Aが既存手法の半分程度となっている。

これらのことから網羅性と判定精度にトレードオフの関係があることが明らかとなった。そのため、提案手法は網羅性が低い代わりに誤判定が少ない手法であることが明らかとなった。

(2) 既存手法と比較して提案手法は購入可能性の高いユーザを取得可能であることがわかった。

各手法の F 値の最良の値を比較すると、既存手法が 0.2830 であるのに対して、提案手法 A は 0.6264、提案手法 B は 0.6286 となっている。このことから、二つの機能を追加したリアルタイムバースト解析がより購入可能性の高いユーザを抽出可能であることがわかった。その一方で、提案手法 A では、購入可能性が高いユーザとして検知できたユーザ数は 8,066 人に留まっている。このことから、提案手法では検知することができないユーザが半数以上いることが明らかとなった。そのため、追加機能で設定するウィンドウサイズの最適化が必要であると考えられる。

(3) アクセス総数に応じてウィンドウサイズを動的に変更する手法が有効であることがわかった。

提案手法 A でアクセス総数が 10 回以上のユーザの適合率を確認すると、0.8040 となっており、アクセス総数に応じてウィンドウサイズを動的に変更することで購入可能性のあるユーザを適切に推定できていることがわかった。特に、アクセス総数が多いユーザの中でもバーストするユーザは、購入に至る可能性が高い傾向がみられるため、リピータ対策に活用できると考えられる。

(4) ユーザのアクセス時間帯毎に動的にウィンドウサイズを変更する手法が有効であることがわかった。

30 回～99 回のユーザでは 0 時～6 時の時間帯は 18 時～24 時の再現率 0.5628 と比べ、再現率は 0.5543 と同程度であるのに対し、適合率 0.5426 と比べ、0.7259 と 1.4 倍程度となっている。しかし、6 時～9 時や 12 時～13 時の時間帯では適合率は 0.4800 や 0.4342 と他と同程度であるのに対し、再現率が 0.3429 や 0.4024 と他の時間帯の 3 割程度に低下している。このことからユーザの特定時間帯のアクセス総数を考慮することが有効であることが明らかとなった。

4. おわりに

本研究では EC サイトにおける購入可能性を推定する手法として、リアルタイムバースト検出手法を用いることを提案した。さらに、アクセスログ解析のための二つの機能を追加した。これらの機能を組み込んだ提案手法は、評価実験の結果から既存手法と比較して購入可能性の高いユーザを推定できることが明らかとなった。しかし、今回の実験ではウィンドウサイズの変更基準となるアクセス総数とウィンドウサイズを経

験的に決定してしまったため、今後は実際のデータからのパラメータ決定と検証が必要であると考えられる。

また、今後は追加した機能が他の分野のデータストリームの分析に対しても有用であるかを検証する予定である。

謝辞 本研究は JSPS 科研費 JP 15K00163 の助成を受けたものです。本研究の遂行にあたり、アクセスログを提供頂いたテックジャム社 (<https://www.tech-jam.com>) に深く感謝します。

参 考 文 献

- [1] 新美潤一郎, 星野崇宏: ユーザ別アクセス・パターン情報の多様性を用いた顧客行動の予測とモデリング, 応用統計学, 応用統計学会, Vol.44, No.3, pp.121-143, 2015.
- [2] 新美潤一郎, 星野崇宏: マルコフ潜在クラスモデルに基づく EC サイトにおける施策実施効果分析に関する一考察, 情報処理学会論文誌, 情報処理学会, Vol.58, No.12, pp.2034-2045, 2017.
- [3] 新美潤一郎, 星野崇宏: 顧客行動の多様性変数を利用した購買行動の予測: Deep Learning を応用した実店舗・Web・モバイルの多面的な分析, 人工知能学会論文誌, 人工知能学会, Vol.32, No.2, pp.1-9, 2017.
- [4] Wendy, M. and Peter, F.: Dynamic Conversion Behavior at E-commerce Sites, Management Science, Vol.50, No.3, pp.326-335, 2004.
- [5] 佐藤翔太, 朝日弓未: 消費者向け電子商取引サイトにおける顧客の来店有無を考慮した購買モデル, オペレーションズ・リサーチ: 経営の科学, 日本オペレーションズ・リサーチ学会, Vol.58, No.2, pp.80-86, 2013.
- [6] 三宅悠介, 松本亮介, 力武健次, 栗林健太郎: 特徴抽出器の学習と購買履歴を必要としない類似画像による関連商品検索システム, 研究報告インターネットと運用技術 (IOT), 情報処理学会, Vol.37, No.4, pp.1-8, 2017.
- [7] 中村健二, 外山諒, 田中成典, 藤本雄紀: EC サイトにおけるユーザの消費者関与の分析に関する研究, 第 33 回ファジィシステムシンポジウム講演論文集, 日本知能情報ファジィ学会, Vol.34, pp. 282-283, 2017.
- [8] 蝦名亮平, 中村健二, 小柳滋: リアルタイムバースト検出手法の提案, 日本データベース学会論文誌, 日本データベース学会, Vol.9, No.2, pp.1-6, 2010.
- [9] 蝦名亮平, 中村健二, 小柳滋: リアルタイムバースト解析手法の提案, 情報処理学会論文誌データベース, 情報処理学会, Vol.5, No.3, pp.86-96, 2012.
- [10] 柳本豪一, 泉愛: バースト解析を用いたサイト流入に貢献したクエリの評価, 平成 27 年電気学会電子・情報・システム部門大会講演論文集, 電気学会, Vol.4, No.4, pp.420-425, 2015.