

平成21年4月1日現在

研究種目：基盤研究（C）  
研究期間：2007～2008  
課題番号：19500115  
研究課題名（和文） リアルタイム・インターネット・マーケティング技術の研究  
研究課題名（英文） Research on Real-Time Internet Marketing  
研究代表者  
吉田 健一（YOSHIDA KENICHI）  
筑波大学・大学院ビジネス科学研究科・教授  
研究者番号：40344858

## 研究成果の概要：

本研究では、高速・大容量のデータに適用可能なストリームマイニング技術を研究し、単に大量のデータをオンライン処理可能な技術だけでなく、刻々と変化するデータの重要度に自動的に追従する機能を開発し、ネットワーク計測とTV視聴率の実時間解析を中心とした広告効果解析へ応用した。

## 交付額

(金額単位：円)

	直接経費	間接経費	合計
2007年度	1,700,000	510,000	2,210,000
2008年度	1,500,000	450,000	1,950,000
年度			
年度			
年度			
総計	3,200,000	960,000	4,160,000

研究分野：総合領域

科研費の分科・細目：情報学、知能情報学

キーワード：インターネット、マーケティング、データマイニング

## 1. 研究開始当初の背景

近年の高速インターネット環境の急速な整備、および、電子商取引やネットオークションにみられるような社会インフラとしてのインターネットの重要性の増加に伴い、高速なネットワークに流れる大量のデータから高速に実時間で価値のあるデータを見つけ出す技術に対するニーズが高まっていた。

しかしながら現在のインターネットは10Gbps以上の高速回線により毎秒100万パケット以上、1日あたりに換算すると数十～数

百テラバイト単位のデータが流れており、このような高速・大容量のデータから実時間で価値のあるデータを見つけ出す技術は確立されていなかった

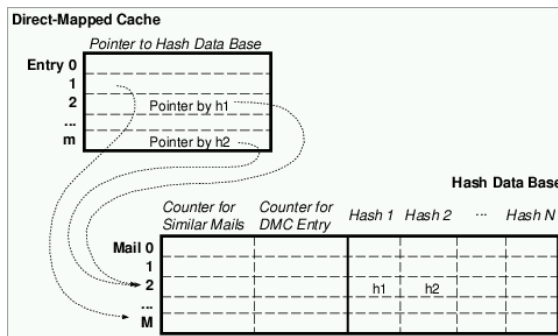
## 2. 研究の目的

本研究は、上記のような高速・大容量のデータに適用可能なストリームマイニング技術の研究を目的とした。研究においては、

- ① 単に大量のデータをオンライン処理可能な技術だけでなく、刻々と変化するデータの重要度(コンセプトドリフト)に自動的に追従する機能の検討を1つ目の目的とした。
- ② さらにデータ重要度の変化に自動追従する事により、TV 視聴率の実時間解析を中心とした広告効果解析と購買データ分析のためのデータマイニング技術への応用を検討し、インターネットを用いた商業活動の活性化に貢献する事を2つ目の目的とした。

### 3. 研究の方法

本提案に関連して解決すべき第1の技術課題は高速なストリームマイニング技術であり、初年度に研究した。図に開発した手法で用いているデータ構造を示す。このデータ構造は Hash Data Base と呼ぶメールに特化した記憶構造と、 Direct-Mapped Cache と呼ぶ高速検索の仕組みから構成されている。この中で Hash Data Base の部分はメール情報に特化しており、研究では、この部分を汎用的な Frequent Itemset 解析に利用できる形に一般化することを検討した。高速検索については Direct-Mapped Cache を基礎として改良を行った。この仕組みは高速性だけでなく、新しい迷惑メールに自動的に追従すると言ったコンセプトドリフト対応の為の基本的な機能も持っている。

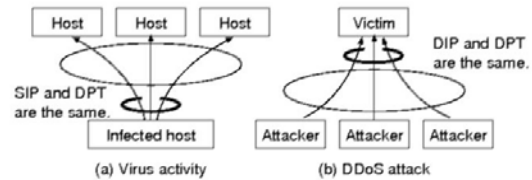


2年目は初年度の研究をベースとした TV 視聴率の実時間解析と広告効果の解析研究に着手した。これについては、研究代表者のそれまでの研究でインターネット上のニュースの出現回数を観測することで、実時間で俳優の人気指数とも言えるデータが取れることを明らかにしており、その基礎検討の知見をベースに一般的な視聴率調査と購買行動の予測などインターネットを使った商業活動活性化の研究に発展させた。

### 4. 研究成果

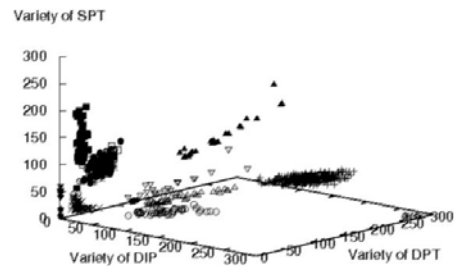
- ① 初年度の高速なストリームマイニング技術は、2年度の研究のベースにしただけでなく、ネットワーク運用管理など別種の応用にも利用し、査読付き論文1本および査読付き国際会議2本の成果を得た。

下図にネットワーク運用管理に開発手法を適用するためのアイデアを示す。



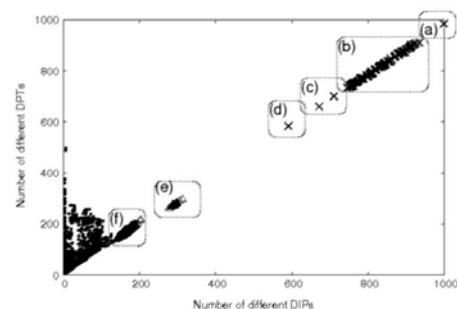
毎年インターネットを經由して感染する新種のウィルスが報告されているが、それらはすべて、感染したホストから新しいホストを探すための IP パケットを送出している。また分散型 DoS 攻撃といった問題も、複数のホストと通信しているホストとして問題のあるホスト計算機を検出可能である

下図に実際のネットワークに流れるパケットの情報を開発手法で解析した時の結果の例を示す。



開発したストリームマイニング機構により、頻出するパケットの異なり数情報を3次元表示するだけで、P2P ソフトや、インターネットウィルスが、その他のパケットの情報とは異なる位置に3次元表示される結果が見てとれる。

次図は更に詳細に提案手法の表示結果を分析したものである。



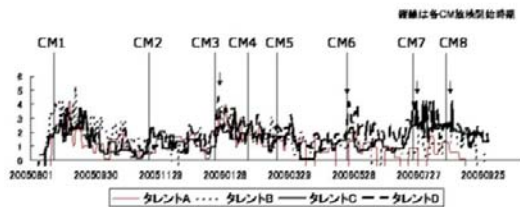
異なり数の情報を分析するだけで、

- 異なり数が 600 以上の (a, b, c, d) は DDoS 攻撃を受けているか、Port Scan を実施している問題のあるサイトであり、
- 異なり数が 200 前後の (e, f) は、P2P ネットワークのハブホストである

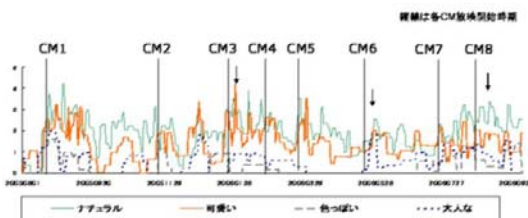
こと等が判別できる。

- ② 2 年度目に関しても計画どおりの研究成果を得、2 年間の査読付き論文総数 4 本、査読付き国際会議 6 本の成果を得た。

下図は開発手法により TV コマーシャルがどの程度人々の関心をひいたか、その効果の測定を試みた結果の例である。



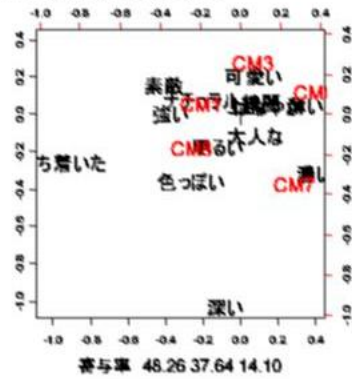
シリーズ化されたコマーシャルのうち 1 番めの CM1、最後の CM7, CM8 等が人々の興味を引く事に成功したのに比べ、CM2, CM4, CM5 は、あまり効果がなかった事などが見てとれる。また、



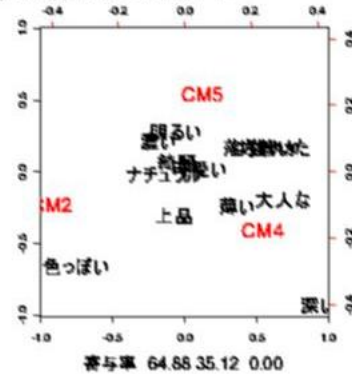
は、各 CM がどのようなイメージ効果を視聴者にもたらしたかの分析結果を示している。

次図は TV コマーシャルの効果測定やイメージ効果の測定結果を使い、実務において良く用いられるコレスポンド分析を行った結果である。各 CM が視聴者にどのようなプロダクトイメージを形成できたかを分析できており、開発手法がマーケティング等の実務にそのまま適用可能である事を示している。

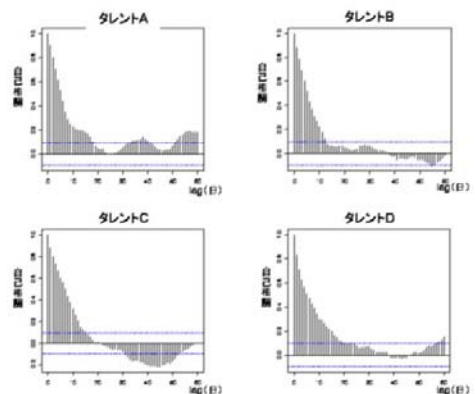
(a) 高關心度CMとプロダクトイメージ



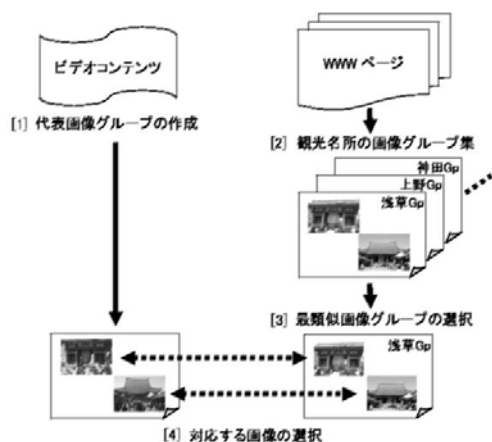
(b) 低關心度CMとプロダクトイメージ



提案手法の実務への適用可能性は上記に止まらない。例えば下図は TV コマーシャルの効果測定結果の自己相関を分析し、各コマーシャルがオンエア開始後約 17 日間で、広告効果をなくしている事を明らかにしている。視聴者の受けの良い TV コマーシャルであっても 17 日たてば次のコマーシャルに変更しないと効果を失う事をデータから客観的に示せる事の実務的意味は大きい。

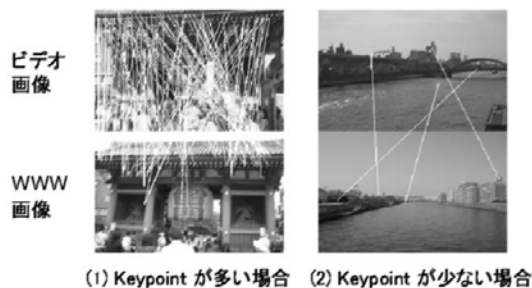


本研究では、さらに近年ネットワークを経由したやりとりが増えている各種画像についても、マイニング技術の一環として、類似画像の検索技術を検討した。



開発した技術は複数の画像グループで構成した情報を用いて検索性能をあげるグループ法というアイデア（上図参照）で、従来検索が難しかった類似画像の検索精度向上を実現した。

さらに開発手法は、sift 特徴と wavelet 特徴を組み合わせる事で、sift 特徴が苦手とする keypoint の少ない画像（下図参照）の検索精度を向上させている。



## 5. 主な発表論文等

（研究代表者、研究分担者及び連携研究者には下線）

〔雑誌論文〕（計 4 件）

- (1) ビデオ画像へのアノテーションのためのグループに基づいた画像検索手法, 村林昇, 吉田健一, 人工知能学会誌, Vol.24, No.2, pp203-213(2009)
- (2) Analyzing the Number of Varieties in Frequently Found Flows, Y. Shoumura, Y. Wanatabe, K. Yoshida, IEICE Transactions on Communication, Vol.E91-B, No.6, pp1896-1905 (2008)

- (3) インターネット・コミュニティ・データを使ったテレビCMの商品イメージ形成効果測定, 上原 宏, 佐藤 忠彦, 吉田健一, 人工知能学会論文誌, Vol.23, No.3, pp205-216 (2008)
- (4) A Proposal of Wireless Network Routing Protocol for Heterogeneous Mobility, Tomohiko Yagyu, Masahiro Jibiki, Kenichi Yoshida, 電子情報通信学会英文誌, Vol.E90-B, No10, pp2693-2701 (2007)

〔学会発表〕（計 12 件）

- (1) Acquiring Marketing Knowledge from Internet Bulletin Boards, H.Uehara, K.YOSHIDA, Proc of The 2008 Pacific Rim Knowledge Acquisition Workshop, (2008年12月16日、ベトナム)
- (2) Egocentrism Presumption Method with N-Gram for e-Business, N. Suzuki, K. Tsuda, 12th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems, pp.1002--1009, (2008年9月3日、クロアチア)
- (3) Tiny LSH for Content-based Copied Video Detection, K. Yoshida, N. Murabayashi, Proc. of SAINT2008, (2008年8月1日、フィンランド)
- (4) Group-Based Image Retrieval Method for Video Annotation, N. Murabayashi, S. Kurahashi, K. Yoshida, Proc. of SAINT2008, (2008年8月1日、フィンランド)
- (5) Evaluation of a Hierarchical Shaper as a Policy Execution Point, T. Aimoto, T. Yazaki, T. Isobe, Y. Sakata, K. Yoshida, 11th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems, (2007年9月13日、イタリア)
- (6) Visualizing Network Status, K. Yoshida, Y. Shoumura, Y. Watanabe, International Conference on Machine Learning and Cybernetics 2007 (2007年8月21日、香港)

以上査読付き国際会議

- (7) 画像グループの類似度に基づいたインターネット上の画像検索, 村林昇, 吉田健一, 映像情報メディア学会冬季大会, 講演番号 8-6 (2008年12月10日、東京)
- (8) 画像のグループ化処理を適用した類似画像検索の性能改善, 村林 昇, 倉橋

- 節也, 吉田 健一, 情報処理学会全国大会予稿, (2008年3月14日、茨城)
- (9) 画像のグループ化に基づいた画像検索法, 村林昇, 倉橋節也, 吉田健一, 信学技報, vol. 107, no. 290, IE2007-83, pp. 7-12, (2007年10月26日、会津)
- (10) インターネットコミュニティにおけるテレビCMイメージの伝播効果測定, 上原宏, 佐藤忠彦, 吉田健一, 人工知能学会データマイニングと統計数理研究会予稿 (2007年7月26日、旭川)
- (11) Frequent Itemsets Mining for Network Risk Management, K. Yoshida, Y. Shoumura, Y. watanabe, International Workshop on Risk Informatics (2007年6月20日、宮崎)
- (12) 無線マルチホップネットワークにおけるアプリケーション適応型メトリックを用いたリアクティブ型経路制御方式, 柳生智彦・地引昌弘・吉田健一, 電子情報通信学会インターネットアーキテクチャ研究会技術報告, IA2007-6, pp29-34 (2007年5月30日、東京)

## 6. 研究組織

### (1) 研究代表者

吉田 健一 (YOSHIDA KENICHI)

筑波大学・大学院ビジネス科学研究科・教授  
研究者番号：40344858

### (2) 研究分担者

倉橋 節也 (KURAHASHI SETSUYA)

筑波大学・大学院ビジネス科学研究科・准教授  
研究者番号：40431663

### (3) 連携研究者

津田 和彦 (TSUDA KAZUHIKO)

筑波大学・大学院ビジネス科学研究科・教授  
研究者番号：50302378