

## 科学研究費助成事業（科学研究費補助金）研究成果報告書

平成25年 3月31日現在

機関番号：12102

研究種目：基盤研究（C）

研究期間：2010～2012

課題番号：22500117

研究課題名（和文） インターネットビジネス支援技術の開発

研究課題名（英文） Research on Internet Business Support

研究代表者

吉田 健一（YOSHIDA KENICHI）

筑波大学・ビジネスサイエンス系・教授

研究者番号：40344858

研究成果の概要（和文）：

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重要性が高まっている、高速なネットワークに流れる大量のデータから実時間で価値のあるデータを見つけ出しマーケティング等の経済活動に有用な情報を発見するための技術を、それまでの基盤研究で開発した技術をベースに研究した。

研究成果の概要（英文）：

In this study, we have investigated data mining methods for internet marketing. The characteristics of our data mining methods are: 1) They are based on our previous studies on one-to-one marketing techniques using internet data. 2) They also use various techniques we had developed to analyze huge internet data in efficient way.

交付決定額

（金額単位：円）

|        | 直接経費      | 間接経費    | 合 計       |
|--------|-----------|---------|-----------|
| 2010年度 | 1,200,000 | 360,000 | 1,560,000 |
| 2011年度 | 1,100,000 | 330,000 | 1,430,000 |
| 2012年度 | 1,000,000 | 300,000 | 1,300,000 |
| 年度     |           |         |           |
| 年度     |           |         |           |
| 総 計    | 3,300,000 | 990,000 | 4,290,000 |

研究分野：知能情報学

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

## 1. 研究開始当初の背景

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実時間で価値のあるデータを見つけ出し、マーケティング等の経済活動に有用な情報を発見するための技術に対するニーズが高まっている。

提案者は、このような背景の中、平成 15

～17年度 基盤研究(C)「ユビキタス・ネットワーク社会におけるワンツーワン・マーケティング基盤の研究」にて電子商取引活生化のための TV 視聴率の実時間解析技術や、行き過ぎたマーケティング対策としての迷惑メール防止技術を提案し、更に、この2つの技術の高速化・一般化を計った平成 19～20年度 基盤研究(C)「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」にて TV コマーシャルの効果測定技術や P2P サイトの検知技術等を開発してきた。

中でも迷惑メール防止技術は発表時点で世界最高性能の迷惑メール防止技術であり、キャリアで実運用されている。

## 2. 研究の目的

上記研究業績は、いずれも高速なネットワーク(1Gbps, 100 万パケット/秒)に流れる大量のデータ(数百 Tera byte/日)から実時間で価値のあるデータを発見するデータマイニング技術の中核の技術としており、特に迷惑メールの防止や P2P サイトの検知などインターネット・セキュリティの確保については実用上も意味のある成果をあげてきた。

一方電子商取引の活生化は幅広いテーマを包含しており、まだまだ検討すべき事項が残っている。特にネットワークがビデオデータを実時間で転送できる高帯域を獲得した事によりビデオを使ったビジネスが急速に立ち上がっている点は新しい研究課題を作っており、早急な研究の立ち上げを必要とする。

以上を考慮し、本基盤研究では、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、重要性が高まっている、高速なネットワークに流れる大量のデータから実時間で価値のあるデータを見つけ出し、マーケティング等の経済活動に有用な情報を発見するための技術を、それまでの基盤研究で開発した技術をベースに研究した。

## 3. 研究の方法

本提案は、上記背景を考慮し、具体的には当初計画どおり次の2点について研究を行い、実システムにも利用可能な技術を開発した。

- (1) 不正コピーされたビデオの発見技術
- (2) ビデオビジネスを対象としたマーケティング技術

更に(2)については予定以上に一般的な研究成果を得ることができた。具体的には(3) 旅行業におけるインターネットを用いたマーケティング技術として成果を取りまとめた。

## 4. 研究成果

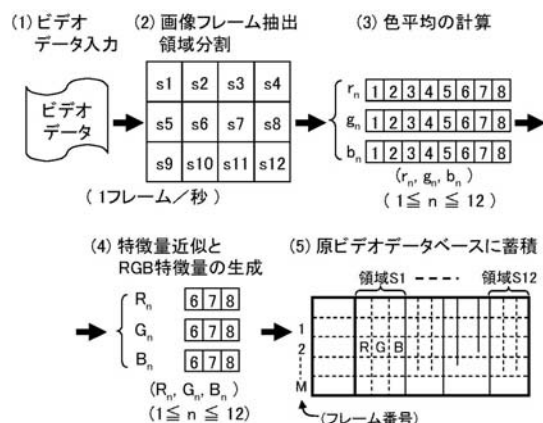
- (1) 不正コピーされたビデオの発見技術

近年、ブロードバンドネットワークの普及により様々なビデオサービスが実現している。このようなサービスは新しいビジネスチャンスを起こすものであるが、それに伴って違法に複製されたビデオコンテンツを検出する技術の重要性が増している。このような

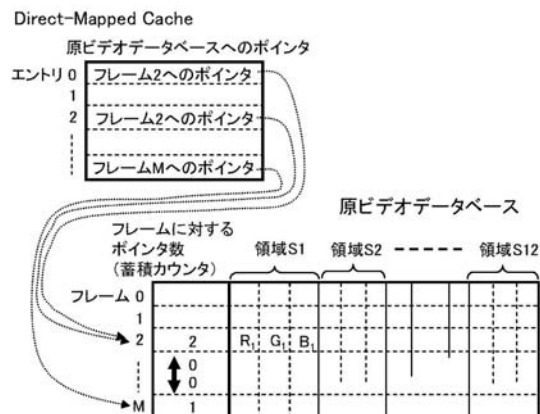
背景において技術的な課題は、WWW 上に存在する膨大な量のビデオコンテンツから違法に複製されたビデオを検出することである。

本研究では、大規模なビデオデータベースから所望のビデオ断片を実用的な精度と速度で検索可能な手法を提案した。

図に提案手法で用いた RGB 色空間に基づいた画像特徴量の抽出方法を示す。はじめに、入力した著作権保護の対象ビデオから1秒に1回画像フレームを抽出して、その画像フレームを4x3の領域に分割する。次に領域ごとに各8ビットのRGB色データのそれぞれの平均値を計算し、計算した各平均値データから上位3ビットのデータだけを取り出し36次元の特徴ベクトルとする。この36次元特徴ベクトルを、検索に用いるため、取り出したフレームの順番で原ビデオデータベースに蓄積する。

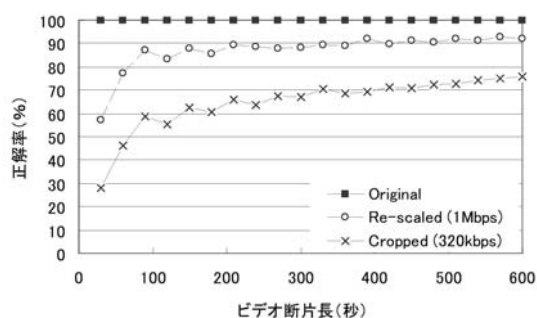


実用システムを構築するには、この原ビデオデータベースを高速に検索する仕組みが必要であるが、この部分は迷惑メール防止技術として開発した高速検索技術(下図)を利用した。



この結果、発表論文[2]では実際のビデオコンテンツを用いた実験の結果、次の図に示すように、改変されていないビデオの場合、

平均検索時間 0.01 秒, 正解率 100%で該当するビデオの検索ができること, 改変されたビデオの場合もサイズの変換程度であれば著しい性能低下は無いことを確認した。



## (2) ビデオビジネスを対象としたマーケティング技術

本基盤研究では, ビデオビジネスを対象としたマーケティング技術として, DNS クエリデータを用いた映画に対する社会からの関心度の測定手法を提案する事ができた. 従来の手法は, 主にインターネット上の仮想コミュニティへの投稿者という「書き手側」の情報を分析することで関心度の測定を行うが, そこには「書き手側」の意図的な情報操作による影響を受けやすいという課題があった. そこで本研究では, 意図的な情報操作の難しい DNS クエリデータを用いた「読み手側」からの関心度測定手法を提案した.

提案手法では, DNS キャッシュの影響を排除するため DNS クエリデータの異なり数を用いる. また, 提案手法を検証するために日本映画の公開初週の観客動員数予測を行い, 公開前2ヶ月間における予測精度の推移についても考察し, 提案手法が実務へ適用できる精度を持つ事を示した.

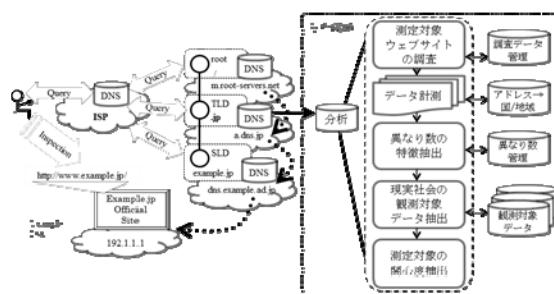
近年, インターネット上のコミュニティ形成を支援するブログや SNS などのサービスが, 日常的に利用されるようになってきた. このようなインターネット上のコミュニティ (以下, 仮想コミュニティ) での現象と現実社会での現象との関係についての研究は多く行われている. これら仮想コミュニティにおける現象と現実社会での現象の関連性に関する研究は, 主として仮想コミュニティにおける「書き手側」の視点から記述内容や記述頻度, 時間などの特徴を抽出し, 現実社会での現象との関係を分析する事が主流であった.

ここでの疑問は, 「仮想コミュニティにおける書き手は, 作為的な書き手ではないと言えるのだろうか?」ということである. たとえば, インターネット掲示板への書き手が, CM に関する好意的な書き込みを掲示板に連続投稿し続けることで関心度を見かけ上長

引かせ, CM による効果が継続しているように見せかける操作を行うことは, 技術的には可能である. 特に, 分析する内容 (商材や人物など), 分析する掲示板やブログなど分析対象などを絞り込めほど, 「書き手側」による作為的な影響は高くなると考えられる.

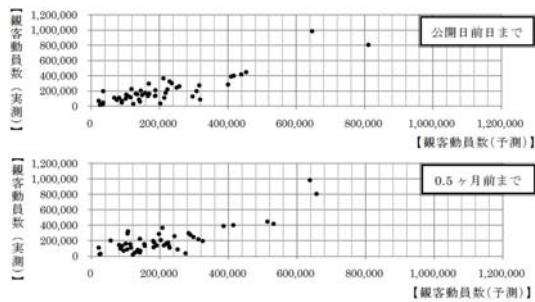
本研究では, このようなコンテンツへの「書き手側」を中心とした関心度分析法の持つ課題を解決する手法を提案した. 具体的にはインターネットにおける現象と現実社会での現象の関係において, 作為的な意思が入りにくい「読み手側」を中心とした関心度分析法を提案した. 提案する手法では, 「読み手側」の関心を表わす情報として DNS へのクエリデータを用いる.

一般的に DNS へのクエリは, ウェブの閲覧やメールの送受信などで発生する. つまり, コンテンツの読み手がアクセスする場合に DNS へのクエリは常に発生する. また DNS へのクエリは, 特定のブログや掲示板などにアクセスする場合にのみ発生するものではなく, すべてのウェブのアクセスで共通に発生するものである. このことから, DNS へのクエリから抽出したデータには, コンテンツへの「書き手側」の意思は入りにくいと言える.



図に一般的な DNS クエリの仕組みと, 本研究における関心度測定のためのシステム構成を示す. この事例では, まずウェブの閲覧者によるブラウザの操作により, ISP の DNS にウェブの URL である `www.example.jp` と IP アドレスの変換が要求される. これに対して ISP の DNS は, 初めに `m.root-servers.net` から `.jp` のデータを持つ `a.dns.jp` の IP アドレスを取得し, 次に `a.dns.jp` から `dns.example.ad.jp` の IP アドレスを取得する. 最後に `dns.example.ad.jp` から `www.example.jp` の IP アドレスを取得し, 最終的に `www.example.jp` のウェブサイトの IP アドレスをブラウザが動作する PC に通知する.

発表論文[3]では, このような仕組みにより次の図のような映画の興行収入予測が可能である事を報告した. また, この結果は実際の映画配給会社に確認いただき, 実務にも応用可能である事を確認いただいた.



### (3) 旅行業におけるインターネットを用いたマーケティング技術

宿泊検討者がインターネットより宿泊予約を行う際、宿泊利用者が投稿した宿泊レビューが重要視されている。宿泊レビューは、主に宿泊利用者のプロフィール、数項目の数値評価および感想コメントの3要素からなる。多くの宿泊検討者は、容易に情報を取得できる数値評価を重要視するが、数値評価のみでは当該宿泊施設の情報を十分に把握できるとは言い難い。そこで本論文では、テキストマイニングを用い感想コメントを分析することで、宿泊検討者にとって有意義な情報を抽出する手法を提案する。さらに、提案手法を用いて感想コメントを分析した結果、ビジネスエリアとレジャーエリアでは同じ数値評価の項目でも全く異なる指標で採点されていること、エリア間で評価の差異があること、数値評価の項目にない事項が重要視されていることなどを導くことができた。

具体的な抽出結果を以下に示す。

下表には「ビジネスエリア数値評価平均と特徴表現分析」結果を示す。数値評価では部屋と接客は同じ4.1と評価されている。しかし、口コミ分析の結果の否定表現は、部屋は約32%、接客は約5%と大きな開きがある(発表論文[1])。

| 数値評価 | 部屋  | 風呂  | 料理  | 接客  | 清潔感 |
|------|-----|-----|-----|-----|-----|
| 平均点  | 4.1 | 3.9 | 4.1 | 4.1 | 4.2 |

| 感想評価 | 部屋    | 風呂  | 料理  | 接客  | 清潔感 | 立地    |
|------|-------|-----|-----|-----|-----|-------|
| 肯定表現 | 813   | 439 | 694 | 916 | 350 | 1,404 |
| 否定表現 | 406   | 91  | 78  | 47  | 34  | 40    |
| 中立・他 | 48    | 32  | 22  | 23  | 1   | 11    |
| 総計   | 1,267 | 562 | 794 | 986 | 385 | 1,455 |

下記には「レジャーエリア数値評価平均と特徴表現分析」結果を示す。

| 数値評価 | 部屋  | 風呂  | 料理  | 接客  | 清潔感 |
|------|-----|-----|-----|-----|-----|
| 平均点  | 4.1 | 4.3 | 4.4 | 4.3 | 4.2 |

| 感想評価 | 部屋  | 風呂    | 料理    | 接客    | 清潔感 | 立地  |
|------|-----|-------|-------|-------|-----|-----|
| 肯定表現 | 526 | 1,118 | 1,900 | 1,309 | 274 | 471 |
| 否定表現 | 164 | 233   | 184   | 74    | 29  | 19  |
| 中立・他 | 35  | 65    | 67    | 33    | 2   | 5   |
| 総計   | 725 | 1,416 | 2,151 | 1,416 | 305 | 495 |

ビジネスエリアとは異なり、料理や風呂に対する口コミが多いことから、顧客が気にする項目がビジネスエリアとは異なることが見えてくる。さらには、ビジネスエリア同様、数値評価と口コミの否定表現の割合が異なることが分かる。

このことから、数値評価だけでは見えてこない真の評価が、口コミの分析を実施したことで見えてくることを意味する。

今後は、この技術を更に研鑽し、インターネットを用いた宿泊を含む旅行アイテムのマーケティングに活用したいと考えている。

### 5. 主な発表論文等

〔雑誌論文〕(計5件)

[1] 辻井康一, 津田和彦, “テキストマイニングを用いた宿泊レビューからの注目情報抽出方法” 情報処理学会デジタルプラクティス, 査読有, 3(4), 289-296 (2012-10-15, [https://ipsj.ixsq.nii.ac.jp/ej/index.php?active\\_action=repository\\_view\\_main\\_item\\_detail&item\\_id=89387&item\\_no=1&page\\_id=13&block\\_id=8](https://ipsj.ixsq.nii.ac.jp/ej/index.php?active_action=repository_view_main_item_detail&item_id=89387&item_no=1&page_id=13&block_id=8))

[2] 村林, 吉田, “Tiny LSH に基づいた大規模ビデオ検索”, 電子情報通信学会和文論文誌, 査読有, J94-D(8), pp. 1461-1472 (2011.8, <http://ci.nii.ac.jp/naid/110008686495>)

[3] N. Suzuki, K. Tsuda, “Expressing the mutual understanding gap with the human egocentrism”, Journal of Computer Applications in Technology, 査読有, V38-1, pp. 316-323, (2010, [http://books.google.co.jp/books?id=mIKYxNLuM4IC&pg=PA318&lpg=PA318&dq=N.Suzuki,+K.Tsuda,+%E2%80%9CExpressing+the+mutual+understanding+gap+with+the+human+egocentrism&source=bl&ots=JehnNCDRWL&sig=2rmAKgFNt01Vboxv\\_sgqSSmJERo&hl=ja&sa=X&ei=NY1iUd\\_6BcnirAeQ9oH4DQ&ved=0CDsQ6AEwAA#v=onepage&q=N.Suzuki%2C%20K.Tsuda%2C%20%E2%80%9CExpressing%20the%20mutual%20understanding%20gap%20with%20the%20human%20egocentrism&f=false](http://books.google.co.jp/books?id=mIKYxNLuM4IC&pg=PA318&lpg=PA318&dq=N.Suzuki,+K.Tsuda,+%E2%80%9CExpressing+the+mutual+understanding+gap+with+the+human+egocentrism&source=bl&ots=JehnNCDRWL&sig=2rmAKgFNt01Vboxv_sgqSSmJERo&hl=ja&sa=X&ei=NY1iUd_6BcnirAeQ9oH4DQ&ved=0CDsQ6AEwAA#v=onepage&q=N.Suzuki%2C%20K.Tsuda%2C%20%E2%80%9CExpressing%20the%20mutual%20understanding%20gap%20with%20the%20human%20egocentrism&f=false))

[4] 三田村, 吉田, “DNS クエリデータに基づくコンテンツへの関心度分析”, 電子情報通信学会和文誌, 電子情報通信学会, J93-B, 10, 査読有, pp. 1368-1377, (2010.10, <http://iss.ndl.go.jp/books/R100000002-I000000105381-00>)

[5] K. Suzuki, M. Jibiki, K. Yoshida, “Selective Precomputation of Alternate Routes using Link-State Information for IP

Fast Restoration”, IEICE Trans. Communications, IEICE, 査読有, E93-B(5), pp.1085-1094 (2010.5, DOI:IEICE transactions on communications Vol.E93.B No.5 2010-05 p.1085-1094 ISSN:0916-8516 NCID:AA10826261)

〔学会発表〕(計 9 件)

[1] T.Mitamura, K.Yoshida, “Viewers’ side analysis of social interests”, Proc. of ICDM 2012 Data Mining for Service Workshop (2012.12.10, Sheraton Brussels Hotel, Belgie)

[2] M.Okawa, K.Yoshida, “Off-line Writer Verification Using Shape and Pen Pressure Information”, Proc. of International Conference on Frontiers in Handwriting Recognition (2012.9.18, Congress Palace of the hotel-village Porto Giardino, Italiana)

[3] K.Fujiwara, A.Sato, K.Yoshida, “DNS traffic analysis -- Issues of IPv6 and CDN --”, Proc of SAINT2012, pp.129 - 137 (2012.7.16, Swiss Hotel in Izmir, Turkiye)

[4] N.Murabayashi, K.Yoshida, “Copied Video Detection with MPEG-7 Video Signature and Multiple Direct-mapped Cache”, Proc of HIS2011, IEEE SMCS, pp 190-195 (2011.12.8, Universiti Teknikal Malaysia Melaka, Malaysia)

[5] N.Murabayashi, K.Yoshida, “Edited internet video detection using MPEG-7 video signature”, Proc of IA2011, pp.33-37 (2011.10.21, Dalian University of Technology, China)

[6] 辻井康一, 津田和彦, “ユーザーレビューを用いた宿泊施設の情報補完手法”, 2011 情報科学技術フォーラム, E-003, (2011-09-08, 函館未来大学, 函館)

[7] K.Suzuki, M.Jibiki, K.Yoshida, “Comparing Proactive and Reactive Methods for IP Fast Restoration using Localization Algorithm”, Proc. of Signal Processing and Communication Systems, pp.1-10 (2010.12.14, Radisson Resort Gold Coast, Australia)

[8] Y.Shomura, K.Yoshida, A.Sato, S.Matsumoto, K.Itano, “A Traffic Analysis Using Cardinalities and Header

Information”, The First International Conference on Networking and Computing, pp.55-62, (2010.11.18, Hiroshima University, Hiroshima)

[9] S.Matsumoto, A.Sato, Y.Shinjo, H.Nakai, K.Itano, Y.Shomura, K.Yoshida, “A Method for Analyzing Network Traffic Using Cardinality Information in Firewall Logs”, The First Workshop on Company, Campus and Community Networking (C3NET 2010 in SAINT 2010), pp.241-244 (2010.7.19, JW Marriott Hotel Seoul, Korea)

## 6. 研究組織

### (1) 研究代表者

吉田 健一 (YOSHIDA KENICHI)

筑波大学・ビジネスサイエンス系・教授

研究者番号：40344858

### (2) 研究分担者

津田 和彦 (TSUDA KAZUHIKO)

筑波大学・ビジネスサイエンス系・教授

研究者番号：50302378