

広告の特徴に基づくクチコミの分類に関する研究

田中成典^{†1} 北野光一^{†2} 寺口敏生^{†2}
今城彰子^{†2} 柳田尚明^{†2}

広告活動では、認知や関心といった購買意思決定過程モデルの各段階に応じて、有効な広告媒体を選択することが重要である。近年、クチコミの増加にともない、クチコミが消費者の購買意思決定に多大な影響を与えていることから、クチコミを広告手段に活用する企業が増加している。しかし、クチコミを活用した広告では、内容の網羅性や詳細性といったクチコミごとの違いを考慮していないため、消費者に対して適切なクチコミ情報を提供できていない課題がある。そこで、本研究では、購買意思決定過程モデルの各段階に対して有効とされる一般的な広告手法を分析し、その特徴を用いてクチコミを分類する手法を提案する。

Research on Classification of Word-of-mouth Based on Advertising Characteristic

SHIGENORI TANAKA,^{†1} KOICHI KITANO,^{†2}
TOSHIO TERAGUCHI,^{†2} AKIKO IMAGI^{†2}
and NAOAKI YANAGIDA^{†2}

In advertising activities, it is important to select an effective advertising media according to different states of the purchase-decision-process model such as cognition or interest. In recent years, since word-of-mouth communications have a great influence on consumer purchase decisions with increase in the word-of-mouth communications, a growing number of companies are utilizing word-of-mouth communications in advertising way. However, there is a problem that appropriate word-of-mouth messages cannot be provided to consumers because an advertisement using word-of-mouth does not take the difference between word-of-mouth messages such as all-inclusiveness or elaborateness of the content into consideration. Thus in this research, we analyze general advertisements that are supposed to be effective for each state of the purchase-decision-process model, and propose a method for classifying word-of-mouth communications using the characteristics obtained from the analysis.

1. はじめに

情報技術の発展とインターネットの普及にともない、SNS (Social Network Service)、ブログやクチコミサイトといった CGM (Consumer Generated Media) を介して、消費者が商品やサービスの感想を発信する機会が増加している。消費者が発信する情報は、消費者目線での商品の評価が含まれているため、企業によるマーケティング分析の対象となってきた。その上、近年、CGM 上のクチコミが消費者の購買意思決定に強い影響^{1),2)} を及ぼしていることが明らかになったことから、企業が用いる新しい広告媒体として、CGM 上のクチコミに関心が寄せられている。企業がクチコミを実際の広告媒体に活用した事例としては、KDDI 社が提供する「クチコミカタログ³⁾」や、リクルート社が提供する「この町のこんなところが好き!⁴⁾」等があげられる。これらの事例では、対象商品に関するクチコミを収集し、一覧を商品ページに掲載することで、クチコミを広告媒体として活用している。しかし、これらの事例では、クチコミの内容を分析しておらず、既存の広告媒体のように効果の高い広告が行えているとはいえない。

クチコミの内容の分析に関する研究では、体験情報に基づき分析する手法^{5),6)}、評価表現に基づき分析する手法^{7),8)} や肯定表現や否定表現に基づき分析する手法^{9),10)} 等が提案されている。

• 体験情報に基づき分析する手法

池田ら⁵⁾ は、「～してみる」や「～したことがある」といった体験を記述する場合に表れる表現を手がかりとして、クチコミ中から体験情報をルールベースで抽出する手法を提案している。また、石野ら⁶⁾ は、ブログデータベースから旅行ブログのエントリを検出し、その中から観光情報である土産物情報と観光名所情報を抽出する手法を提案している。

• 評価表現に基づき分析する手法

那須川ら⁷⁾ は、「文章中に評価表現が存在すると、その周辺に評価表現の連続する文脈が形成されやすく、それらの極性は明示されない限り一致する傾向にある」という仮

^{†1} 関西大学総合情報学部

Faculty of Informatics, Kansai University

^{†2} 関西大学大学院総合情報学研究科

Graduate School of Informatics, Kansai University

定に基づき、文書内からブートストラップ的に評価表現である単語を収集する手法を提案している。また、高野ら⁸⁾は、評価表現が理由や条件といった評価の要因になる単語と共起することが多いことに注目し、評価表現とそれに対応する評価の要因となる単語を抽出することにより、こちらもブートストラップ的に評価表現である単語を収集する手法を提案している。

● 肯定表現や否定表現に基づき分析する手法

Pang ら⁹⁾は、映画のレビューサイトの記事を対象とし、記事とそれらに付加されたレーティング情報に肯定と否定のラベルを付与したものを学習データとして、ナイーブベイズ、最大エントロピー法と SVM (Support Vector Machine) の 3 種類の分類器を用いて、記事内容を肯定的か否定的かを分類する手法を提案している。また、安村ら¹⁰⁾は、評価表現である単語には、商品全般の評価を表す単語と商品のある属性の評価を表す単語の 2 種類に分別できることに注目し、商品全般の評価を表す単語によってクチコミを肯定と否定の 2 値に分類し、もし分類できない場合は、商品のある属性の評価を表す単語によって 2 値に分類する手法を提案している。

以上のように、クチコミの内容を分析する既存研究では、体験情報、評価表現や肯定・否定といった観点から分析することを目的とした研究が行われてきた。しかし、クチコミを広告媒体に利用する場合においては、クチコミが消費者の購買意欲を喚起するかどうかを評価することが重要¹¹⁾であり、クチコミの広告媒体への利用を実現するには次の 2 つの課題がある。

● クチコミが持つ情報の質の違い

クチコミが持つ情報の質はクチコミごとに大きく異なる。クチコミには商品についての評価が自由に記述されるため、「使いやすくて満足です」といったような情報量の少ないものから「CPU の性能が...、ディスプレイの解像度が...」といったような情報量の多いものまで様々なクチコミが存在すると考えられる。このため、クチコミが持つ情報の質の違いを考慮せずにクチコミを広告手段として利用すると、消費者が雑多な情報をすべて確認する必要があるため、高い広告効果を得ることができない。

● 消費者が持つ知識の量の違い

消費者が持つ知識の量によって提供すべき情報量は大きく異なる。たとえば、商品について何も知らない消費者に対して、性能やサイズといった詳細な情報を提供したとしても、意味が理解できず、購買意欲の喚起に結び付かないと考えられる。一方、同様の消費者に対して、商品のイメージやセールスポイントに関する情報を提供すると、商品

についての知識がなくても、大まかなイメージが記憶に残りやすくなるため、購買意欲の喚起につながる可能性が高くなる。したがって、消費者が持つ知識の量によって提供するクチコミを分類しなければ高い広告効果を得ることができない。

そこで、本研究では、広告媒体へのクチコミの利用を目的とし、クチコミが持つ情報の質の違いや消費者が持つ知識の量の違いを考慮した広告種別の特徴によるクチコミの分類手法を提案する。

2. 研究の概要

2.1 提案手法

本研究では、クチコミが持つ情報の質を定量化する指標と、消費者が持つ知識の量に基づきクチコミを分類する手法を提案することで、クチコミを広告媒体として利用するための 2 つの課題を解決する。

(1) クチコミが持つ情報の質の違いに対する提案手法

1 つ目の課題に対しては、情報の広さを定量化するための指標として網羅度を、情報の深さを定量化するための指標として詳細度を定義する。情報の広さとは、情報がどれほど話題に富んでいるかを表す。また、情報の深さとは、ある話題についてどれだけ詳しく記述されているかを表す。本研究では、クチコミが持つ情報の質を情報の広さと深さにより表現できると定義し、これらの指標に基づきクチコミを評価することで、クチコミが持つ情報の質の違いを解決する。

(2) 消費者が持つ知識の量の違いに対する提案手法

2 つ目の課題に対しては、購買意思決定過程モデルに定義された消費者の購買意欲を喚起する広告種別の特徴によりクチコミを分類する。購買意思決定過程モデル¹²⁾とは、図 1 に

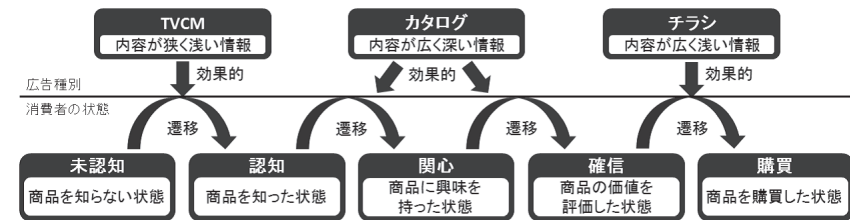


図 1 購買意思決定過程モデル
Fig.1 Purchase-decision-process model.

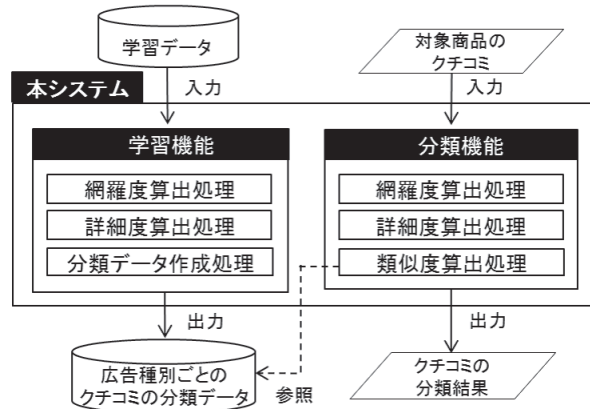


図 2 処理の流れ
Fig. 2 Flow of process.

示すように、消費者が商品を認知してから購買に至るまでの過程を複数の段階に分類し、消費者の段階が次の段階に状態遷移するための有効な広告種別を対応付けたモデルである。企業がこのモデルを用いることで、たとえば、「未認知の段階の消費者に向けて TVCM のような内容の狭く浅い情報を提供することで消費者は商品を認知しやすくなる」や「関心の段階の消費者に向けてカタログのような内容が広く深い情報を提供することで消費者は商品を購入するかどうか判断しやすくなる」というように、消費者がどのような段階のときどのような情報を提供することが購買意欲の喚起につながるのかを具体的に把握できる。

また、ある商品についての消費者が持つ知識の量は、消費者が商品を認知してから購買に至るまでの間に増加していく。そのため、消費者が購買に至るまでにたどる過程を表している購買意思決定過程モデルは、消費者が持つ知識の量が増加していく過程を表していると思えることができ、各段階に有効な広告種別は、その段階の消費者が持つ知識の量に適した情報量を持っていると考えられる。そこで、購買意思決定過程モデルに定義された消費者の購買意欲を喚起する広告種別の特徴によりクチコミを分類することで、消費者が持つ知識の量の違いを解決する。

2.2 処理の流れ

本提案手法のシステムは、図 2 に示すように学習機能と分類機能で構成される。学習機能では、分類機能の類似度算出処理で参照するクチコミの分類データを作成することを目的

に、広告種別に分類したクチコミ群である学習データを入力として網羅度と詳細度を算出し、広告種別ごとのクチコミの分類データを作成する。また、分類機能では、クチコミを広告種別の特徴に基づき分類することを目的に、対象商品のクチコミを入力として、網羅度と詳細度を算出し、クチコミの分類データとの類似度を算出することでクチコミの分類結果を出力する。なお、システムの処理のうち、学習機能と分類機能における網羅度算出処理と詳細度算出処理は、提案手法 (1) を実現するための処理であり、学習機能における分類データ作成処理と分類機能における類似度算出処理は、提案手法 (2) を実現するための処理である。

3. クチコミが持つ情報の質の評価方法

3.1 網羅度の算出方法

情報の広さの度合いを表す網羅度を算出することにより、1 つのクチコミに記述されている内容がどれほど話題に富んでいるかが把握できる。一般的に、消費者は商品のスペックに関する事柄をクチコミの話題として記述する傾向がある。そこで、本研究では、クチコミ中に記述されたスペックを話題と見なし、クチコミに記述されているスペックを網羅度の算出に利用する。網羅度 C の算出式を式 (1) に示す。ここで、 $C(x, y)$ は商品 y に関するクチコミ x の網羅度を、 $S(x)$ はクチコミ x に記述されているスペックの数を、 $allS(y)$ は商品 y におけるスペックの総数を表す。

$$C(x, y) = \frac{S(x)}{allS(y)} \quad (1)$$

網羅度は、その値が高ければ高いほど、クチコミに記述されている話題の量がより豊富であることを示す。ただし、スペックの総数は商品のジャンルごとに異なるため、事前に対象とするジャンルからスペックの総数を取得しておく必要がある。

3.2 詳細度の算出方法

情報の深さの度合いを表す詳細度を算出することにより、1 つのクチコミに記述されている各話題の詳しさが把握できる。ある話題について詳しいかどうかは、その話題の関連語の含有率により決定される。たとえば、CPU が話題にあげられている場合、「種類は…」よりも「種類は…、クロック周波数は…、キャッシュメモリは…」というように、CPU の関連語の記述の多い方がより詳細な内容であると考えられる。そこで、クチコミに記述されているスペックの関連語を詳細度の算出に利用する。詳細度 D の算出式を式 (2) に示す。ここで、 $D(x)$ はクチコミ x の詳細度を、 $SR(x, i)$ はクチコミ x に記述されている i 番目のス

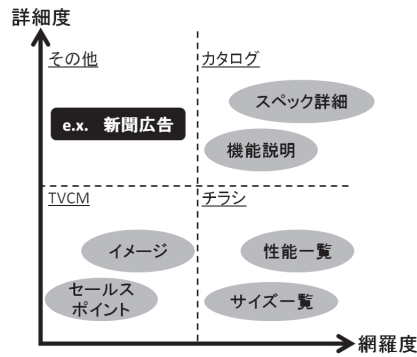


図3 広告種別と情報の質の関係

Fig. 3 Relationship between advertising and quality of information.

ペックの関連語のうち、クチコミ x に記述されている関連語の数を、 $allSR(x, i)$ はクチコミ x に記述されている i 番目のスペックの関連語の総数をそれぞれ表す。また、 n はクチコミ x に出現するスペックの数である。

$$D(x) = \sum_{i=1}^n \frac{SR(x, i)}{allSR(x, i)} \quad (2)$$

詳細度は、その値が高ければ高いほど、1つのクチコミに記述された各話題がより詳しいことを示す。ただし、スペックの関連語の総数はスペックごとに異なるため、事前にスペックの関連語の総数を取得しておく必要がある。

3.3 広告種別と情報の質の関係

購買意思決定過程モデルに定義された消費者の購買意欲を喚起する広告種別を情報の質でそれぞれ評価すると図3のようになると考えられる。ここで、「その他」にあてはまる広告種別には、たとえば、新聞広告のような内容が狭く深い情報が該当する。本提案手法では、図3の関係に基づきクチコミを分類することで、消費者が持つ知識の量に適した情報量を持つクチコミを抽出する。この関係に基づいた場合の広告種別ごとの特徴は次のようになると考えられる。

(1) TVCMの特徴

TVCMは、未認知の状態から認知の状態に遷移させやすい広告種別である。未認知の状態の消費者に対しては、商品のイメージやセールスポイントといった内容が狭く浅い情報

を提供することで消費者の記憶に残りやすくなる。このことから、TVCMは、網羅度が低く、詳細度も低いという特徴を持つ広告種別であるといえる。

(2) カタログの特徴

カタログは、認知の状態から関心または確信の状態に遷移させやすい広告種別である。認知の状態の消費者に対しては、スペックの詳細な情報や機能の説明といった内容が広く深い情報を提供することで自身にとって必要な商品かどうかの判断がしやすくなる。このことから、カタログは、網羅度が高く、詳細度も高いという特徴を持つ広告種別であるといえる。

(3) チラシの特徴

チラシは、確信の状態から購買の状態に遷移させやすい広告種別である。確信の状態の消費者に対しては、性能一覧やサイズ一覧といった内容が広く浅い情報を提供することで購買への後押しとなる。このことから、チラシは、網羅度が高く、詳細度が低いという特徴を持つ広告種別であるといえる。

4. 消費者が持つ知識の量に基づくクチコミの分類方法

4.1 分類データの作成方法

入力されたクチコミを分類するときに参照する広告種別ごとのクチコミの分類データを作成する。広告種別ごとのクチコミの分類データは、網羅度と詳細度を素性とした広告種別ごとの特徴ベクトルを登録したものである。本提案手法で用いる特徴ベクトル V_{char} を式(3)に示す。ここで、 X_{cov} は網羅度、 X_{dit} は詳細度を表す。

$$V_{char} = \{X_{cov}, X_{dit}\} \quad (3)$$

まず、学習データとして入力したクチコミの網羅度と詳細度を算出し、特徴ベクトルを作成する。そして、広告種別ごとの特徴ベクトル群から重心ベクトルを算出する。この重心ベクトルは、広告種別のクラスタの重心を表すベクトルであるため、広告種別の特徴ベクトルと見なすことができる。最後に、広告種別の特徴ベクトルを登録することでクチコミの分類データを構築する。

4.2 類似度の算出方法

ベクトル空間モデル (VSM: Vector Space Model)³⁾ を用いて広告種別ごとの類似度を算出することでクチコミを分類する。まず、入力データであるクチコミの網羅度と詳細度を算出し、特徴ベクトルを作成する。そして、そのクチコミの特徴ベクトルと、広告種別ごとのクチコミの分類データに登録された広告種別の特徴ベクトルとの類似度を算出する。ベクトル間の類似度の算出には、コサイン距離とユークリッド距離を用いる。

コサイン距離は、ベクトルの向きの近さを表す指標であり、ベクトルの向きが類似するほど距離が近くなる。コサイン距離 $CosDict$ の算出式を式 (4) に示す。ここで、 V_a と V_b は類似度を算出するために入力するベクトルであり、 $V_a * V_b$ はベクトルの内積を表す。

$$CosDict(V_a, V_b) = \frac{V_a * V_b}{\sqrt{V_a^2} \times \sqrt{V_b^2}} \quad (4)$$

また、ユークリッド距離は、ベクトルのノルムの近さを表す指標であり、ベクトルのノルムが類似するほど距離が近くなる。ユークリッド距離 $NormDict$ の算出式を式 (5) に示す。ここで、 X_{cov} と X_{dit} は、ベクトル V_a の網羅度と詳細度、 Y_{cov} と Y_{dit} は、ベクトル V_b の網羅度と詳細度をそれぞれ表す。

$$NormDict(V_a, V_b) = \sqrt{(X_{cov} - Y_{cov})^2 + (X_{dit} - Y_{dit})^2} \quad (5)$$

コサイン距離とユークリッド距離は、ベクトル間の類似度が高いほど距離が近くなることから、これらの積を類似度と定義する。類似度 Sim の算出式を式 (6) に示す。

$$Sim(V_a, V_b) = CosDict(V_a, V_b) \times NormDict(V_a, V_b) \quad (6)$$

この類似度を広告種別ごとに算出し、その中で最も距離に近い広告種別にクチコミを分類する。

5. 実験と考察

5.1 実験目的

実験を通じて、クチコミが持つ情報の質の評価方法と、消費者が持つ知識の量に基づくクチコミの分類方法の有用性を確認する。そこで、網羅度と詳細度の妥当性に関する実験とクチコミの分類精度に関する 2 つの実験を行う。前者の実験では、網羅度と詳細度が、購買意思決定過程モデルに定義された消費者の購買意欲を喚起する広告種別にクチコミを分類するときの有用な指標であるかどうかを検証する。後者の実験では、網羅度と詳細度を用いたときのクチコミの分類精度を検証する。

5.1.1 網羅度と詳細度の妥当性に関する実験概要

本実験では、2.1 節 (1) で提案した 2 つの指標における提案指標の妥当性を評価することを目的に、広告種別の特徴に基づいたクチコミの分類結果と情報の質に基づいたクチコミの分類結果が一致するかどうかを確かめる。また、3.3 節の広告種別と情報の質の関係が成立するかどうかを検証する。前者を検証することで、情報の質に基づきクチコミを分類した場合であっても、広告種別の特徴に基づいてクチコミを分類したときと同程度の分類ができる

表 1 テストデータの件数
Table 1 Number of test data.

	ノート PC	デスクトップ PC	プラズマテレビ
TVCM	600 件	600 件	600 件
カタログ	600 件	600 件	600 件
チラシ	600 件	600 件	600 件
合計	1,800 件	1,800 件	1,800 件

かを明らかにする。そして、後者を検証することで、購買意思決定過程モデルに定義された消費者の購買意欲を喚起する広告種別が網羅度と詳細度で特徴づけることができるかを明らかにする。これら 2 つを実証することで網羅度と詳細度が、購買意思決定過程モデルに定義された消費者の購買意欲を喚起する広告種別にクチコミを分類するときの有用な指標であることを評価する。

5.1.2 クチコミの分類精度に関する実験概要

本実験では、2.1 節 (2) で提案した広告種別の特徴に基づくクチコミの分類における分類精度を評価することを目的に、分類精度に影響があると考えられる類似度の算出方法を比較し、本提案手法で採用した類似度算出手法の有用性を検証する。本実験で比較する類似度は、コサイン距離のみを考慮した場合、ユークリッド距離のみを考慮した場合とコサイン距離とユークリッド距離を考慮した場合の 3 つを対象に行う。

5.1.3 実験対象

2 つの実験で用いるテストデータ (表 1) には、クチコミポータルサイトである価格.com¹⁴⁾ が提供するクチコミ掲示板から、クチコミの投稿数が多いノート PC、デスクトップ PC とプラズマテレビの 3 つのジャンルに関するクチコミを収集し、特徴が類似する広告種別に手作業で分類したクチコミを用いる。

手作業でのクチコミの分類には、広告種別の特徴を整理した項目 (表 2) を使用し、あてはまる項目の数が最も多い広告種別にクチコミを分類する。この項目を用いて、クチコミを手作業で分類した TVCM、カタログとチラシの典型事例をジャンルごとにまとめたものを表 3 に示す。ただし、あてはまる項目の数がいずれの広告種別においても 2 つ未満であった場合は、「該当する広告種別なし」と判定する。なお、表 2 に示した項目は、消費者の印象やイメージに近い観点からクチコミを広告種別に分類するため、消費者が広告種別に対して持つ印象やイメージを尋ねたアンケートである「2009 年全国メディア接触・評価調査¹⁵⁾」、「広告市場の変遷からみた通信と放送の今後¹⁶⁾」、「2009 年メディア定点調査¹⁷⁾」および「メ

表 2 広告種別の特徴
Table 2 Advertising characteristics.

広告種別	項目
TVCM	形容詞が記述されている。 記述されているスペックの種類が少ない。 商品のコンセプトが分かる。
カタログ	他の商品との差異が分かる。 商品のコンセプトが分かる。 1つの事柄について詳細に記述されている。
チラシ	価格に関する情報が記述されている。 記述されているスペックの種類が多い。 1つの事柄について簡潔に記述されている。

ディア別広告評価¹⁸⁾」に掲載されている項目をグループ化して整理したものである。

また、2つの実験において、網羅度の算出に必要なスペックの総数は、価格.comの商品スペックのページから取得し、詳細度の算出に必要な関連語の総数は、ヤフー社が提供する関連検索ワード Web API¹⁹⁾を用いて収集した関連語の総数を使用する。

5.2 網羅度と詳細度の妥当性に関する実験

5.2.1 実験内容

提案指標の妥当性を評価するために、表2の広告種別の特徴に基づいた手作業によるクチコミの分類結果と情報の質に基づいたシステムによるクチコミの分類結果が一致するかどうかを検証する。また、広告種別と情報の質の関係が成立しているかどうかを検証する。

広告種別の特徴に基づいたクチコミの分類では、表2の広告種別の特徴に基づきクチコミを分類する。また、情報の質に基づいたクチコミの分類では、事前にテストデータの網羅度と詳細度を算出し、その網羅度と詳細度に基づいてクラスタリングを行う。クラスタリングには、一般的なクラスタリング手法である k-means 法²⁰⁾を用いる。また、本研究では、クチコミを TVCM、カタログとチラシの3つに分類することを目的としているため、本実験で行うクラスタリングのクラス数も3とする。なお、k-means 法は、そのアルゴリズムの特性として、与える中心値の初期値によってクラスタリング結果が変化する。そこで、テストデータとは異なるデータセットを対象に k-means 法によるクラスタリングを10回試行し、最もうまくデータセットを分類できたときの初期値を採用する。ただし、データセットを対象にクラスタリングを行うときの初期値は、データセットから無作為に選択したデータの網羅度と詳細度を用いる。なお、本実験では、ジャンルごとに作成したテストデータ1,800件すべてを使用し、各テストデータは、事前に網羅度と詳細度を算出したものを用

表 3 TVCM, カタログやチラシの典型事例
Table 3 Typical examples of TVCM, catalog and flyer.

広告種別	典型事例 (ノート PC)
TVCM	動作も速いし液晶もキレイ。動画編集も出来そうですね。
カタログ	ギガビット LAN も HDMI 端子も無い G550 はシンプルすぎるかと。個人的には AMD の CPU ですがグラフィック性能が高く 64 Bit OS でギガビット LAN ・ HDMI ・ Bluetooth まで装備した (商品名) が当初より 15k 程安くなっていますし、このクラスでは頭一つ抜けているように思います。
チラシ	(企業名) 直販モデル。型番: (型番号) です。売価 189,800 円 - 65,000 円 (クーポン分) 使用だと 124,800 円ですね。現在最安値 (約 127,500 円) で HDD250 GB ・ 重量 1.9 kg ・ さらにバッテリー持続時間も短縮。
広告種別	典型事例 (デスクトップ PC)
TVCM	仕事柄いろいろなディスプレイ見ますが断然良いですよ! 映り込みもかなり抑えられています。明るさは調整できますしね。
カタログ	とうとう本日届きました。CPU: core i7 860 グラフィック: NVIDIA Quadro FX580 メモリ: 8 GB です。今のところ快調です。静かです。コンパクトで軽いです。カードリダが露出ではなくて、プッシュカバーを開けると現れるので、デザインがすっきりしています。また、DVD ドライブも、カバーに隠れていて、フロントパネルの開閉式になっている点も、すっきりしていて好みます。このまま使えば、これ、けっこうよいですよ。
チラシ	AMD Athlon XP 2400+ Windows XP Professional メモリ 256 MB (128 MB × 2) NVIDIA 2 チップセット NVIDIA 4 (チップセット内蔵) 2 画面対応 40 GB Ultra ATA/100 最大 48 倍速 CD-ROM 3 年間保証サービス (1 年目はオンサイト, 2,3 年目はパーツ保証) (商品名) は C/P が高いと思う。
広告種別	典型事例 (プラズマテレビ)
TVCM	こんばんは。1366 × 768 ピクセルです。HDMI 繋げた場合ブルーレイもゲームも綺麗に写りますよ。音声もいい音ですよ。
カタログ	小さい子供がいらっしやるならプラズマを薦めます。画面強度なら断然プラズマが優位で多少の衝撃でも液晶に比べ強固です。また PX70 はハーフ HD ですが、画質は液晶のフル HD と比べて遜色ないと思いますよ。DVD レコーダー (商品名) も評判 (安定駆動) が良いみたいなので、ピエラリンクも視野にいれるなら、私は (企業名) をオススメします。
チラシ	(企業名).com にて..... (商品名) [22V 型 地上デジタルハイビジョン液晶テレビ サクラピンク BS・CS110 度 CS チューナー非搭載] 特価: ¥23,800 (税込) 配達料金 無料 10%還元 (2,380 ポイント)

価格.com のクチコミ掲示板より引用 (一部改変)

いる。

このように分類した2種類の分類結果から次の2つの考察を行い、提案指標の妥当性を検証する。1つ目は、網羅度と詳細度に基づきグラフ化した分類結果を比較することで、それぞれの分類結果が同じ結果であるかどうかを考察する。2つ目は、手作業における平均値とシステムにおける中心値を分析することで、広告種別と情報の質の関係が成立しているかどうかを考察する。

28 広告の特徴に基づくクチコミの分類に関する研究

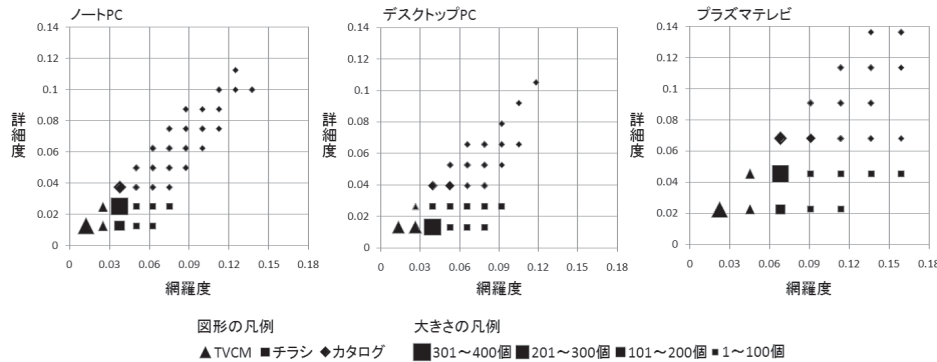


図 4 広告種別の特徴に基づいたクチコミの分類結果

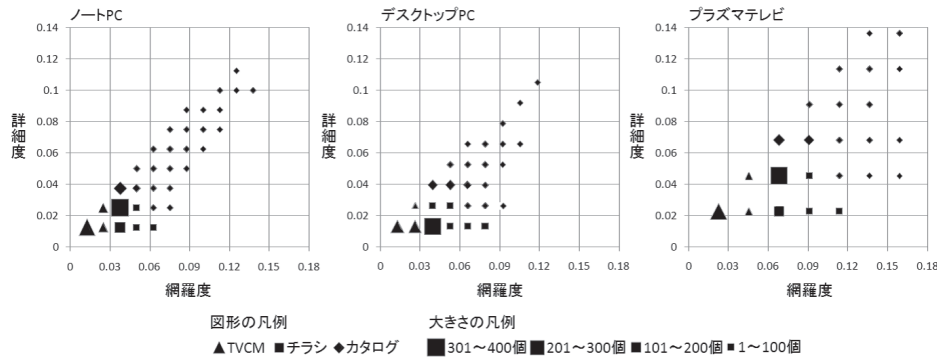


図 5 情報の質に基づいたクチコミの分類結果

5.2.2 結果と考察

広告種別の特徴に基づいたクチコミの分類結果を図 4 に、情報の質に基づいたクチコミの分類結果を図 5 に示す。また、情報の質に基づいたクチコミの分類により得られた各クラスタに属する広告種別の割合を表 4 に示す。

広告種別の特徴に基づいたクチコミの分類結果と情報の質に基づいたクチコミの分類結果を比較するには、情報の質に基づいたクチコミの分類により得られたクラスタと広告種別との対応付けを行う必要がある。そこで、表 4 を確認すると、クラスタ 1 は TVCM、クラ

表 4 各クラスタに属する広告種別の割合

Table 4 Ratio of advertisings according to each cluster.

	ノート PC			デスクトップ PC			プラズマテレビ		
	TVCM	チラシ	カタログ	TVCM	チラシ	カタログ	TVCM	チラシ	カタログ
クラスタ 1 (TVCM)	600 (100)	0 (0)	0 (0)	600 (100)	0 (0)	0 (0)	600 (100)	0 (0)	0 (0)
クラスタ 2 (チラシ)	0 (0)	586 (100)	0 (0)	0 (0)	509 (100)	0 (0)	0 (0)	584 (100)	0 (0)
クラスタ 3 (カタログ)	0 (0)	14 (2)	600 (98)	0 (0)	91 (13)	600 (87)	0 (0)	16 (2)	600 (98)

() 内は各クラスタにおける広告種別の割合 (%)

表 5 広告種別の特徴に基づき分類したクチコミの各広告種別の平均値

Table 5 Average of each advertising of classified word-of-mouth based on advertising characteristic.

広告種別	ノート PC		デスクトップ PC		プラズマテレビ	
	網羅度	詳細度	網羅度	詳細度	網羅度	詳細度
TVCM	0.017	0.016	0.020	0.015	0.033	0.027
チラシ	0.041	0.021	0.051	0.019	0.073	0.040
カタログ	0.050	0.044	0.059	0.043	0.090	0.077

スタ 2 はチラシ、クラスタ 3 はカタログのクラスタであることが分かった。これをふまえて、それぞれの広告種別の分布を比較すると、図 4 と図 5 から広告種別の分布がほぼ一致していることが確認できた。しかし、表 4 からチラシのテストデータが最大で 13% の割合でカタログのクラスタに属していることも分かった。これは、図 4 からチラシのテストデータが偏った分布であることにに対し、カタログのテストデータが均一な分布であることより、密度の低い箇所に分布するチラシのテストデータをカタログのテストデータと誤判定したことが原因であると考えられる。

次に、広告種別の特徴に基づき分類したクチコミの各広告種別の平均値を表 5 に、情報の質に基づき分類したクチコミの各クラスタの中心値を表 6 に示す。また、各広告種別の平均値 (表 5) と各クラスタの中心値 (表 6) を図 6 に示す。

図 6 から各広告種別の平均値と各クラスタの中心値がほぼ一致していることが確認できた。このことから情報の質に基づいたクチコミの分類結果において、チラシのテストデータをカタログに誤分類したものの、広告種別の特徴に基づきクチコミを分類した場合と情報の質に基づきクチコミを分類した場合の分類結果がほぼ同じであることを示していると考えられる。

表 6 情報の質に基づき分類したクチコミの各クラスターの中心値

Table 6 Median of each cluster of classified word-of-mouth based on quality of information.

広告種別	ノート PC		デスクトップ PC		プラズマテレビ	
	網羅度	詳細度	網羅度	詳細度	網羅度	詳細度
TVCM	0.018	0.016	0.020	0.015	0.033	0.027
チラシ	0.040	0.021	0.047	0.017	0.072	0.040
カタログ	0.051	0.043	0.059	0.040	0.090	0.076

表 7 比較する類似度算出手法

Table 7 Adapting methods.

	コサイン距離を考慮	ユークリッド距離を考慮
システム A		×
システム B	×	
システム C		

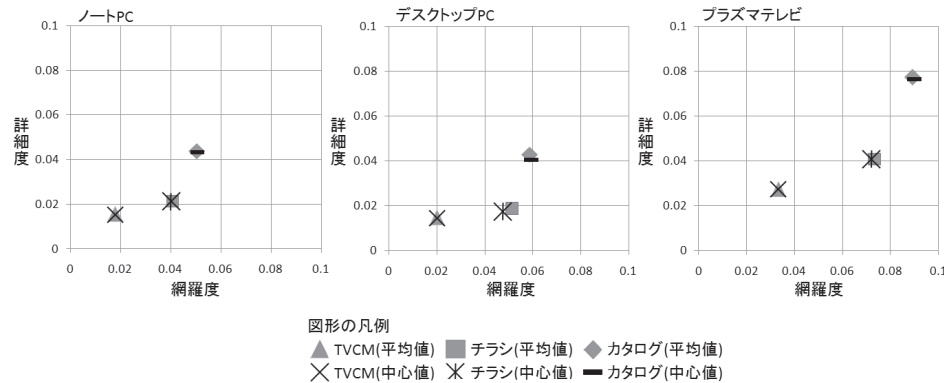


図 6 各広告種別の平均値と各クラスターの中心値

Fig. 6 Average of each advertising and median of each cluster.

また、ジャンルごとの網羅度と詳細度の値に注目すると、ノート PC とデスクトップ PC に関しては、網羅度と詳細度がほぼ同じ値であることが確認できた。一方、プラズマテレビに関しては、ノート PC やデスクトップ PC と比べて網羅度と詳細度の値が高いことが確認できた。これは、ノート PC とデスクトップ PC のようにジャンルが比較的近い場合であれば、それらの商品に対して記述されるクチコミが持つ情報の質にそれほど差異が見られないが、ノート PC とプラズマテレビのようにジャンルが異なる場合では、クチコミが持つ情報の質に違いが表れることを示していると考えられる。さらに、広告種別ごとの網羅度と詳細度の値に注目すると、TVCM の網羅度と詳細度がともに低いこと、チラシの網羅度が高く詳細度が低いこと、カタログの網羅度と詳細度がともに高いことが分かった。これは、3.3 節で示した関係に基づいた場合の広告種別ごとの特徴と一致している。これらのことから、テストデータと広告種別の間には、広告種別と情報の質の関係が成立しており、この関

係は、他のジャンルにおいても成立すると予想される。

以上のことから、広告種別の特徴に基づいたクチコミの分類結果と情報の質に基づいたクチコミの分類結果が一致することが実証された。また、3.3 節の広告種別と情報の質の関係が成立することも実証された。したがって、本提案手法で導入した網羅度と詳細度が、購買意思決定過程モデルに定義された消費者の購買意欲を喚起する広告種別にクチコミを分類するときの有用な指標であることが実証できた。

5.3 クチコミの分類精度に関する実験

5.3.1 実験内容

クチコミと広告種別との類似度算出手法を評価するために、比較する類似度算出手法(表 7)ごとにシステムを構築し、それぞれのクチコミの分類精度について検証する。なお、本実験では、ジャンルごとに作成したテストデータ 1,800 件すべてを使用し、10 分割交差検定を行う。10 分割交差検定とは、テストデータを 10 個のデータ群に分割し、データ群の 1 つを実験データ、残りの 9 つを学習データとして実験を行うという試行を 10 回繰り返し、各試行の精度の平均値を実験の精度とする検定方法である。本実験では、分類精度として適合率、再現率と F 値を用いる。ただし、本研究は、企業がクチコミを広告に活用するときに利用されることを想定しているため、どれだけ正しい分類結果を算出できたかという指標である適合率を重視する。

5.3.2 結果と考察

比較システムの分類精度を図 7 と表 8 に示す。広告種別ごとの分類精度の平均値に注目すると、比較システム A, B と C では、A や B よりも C の適合率、再現率と F 値がすべてにおいて優れていることが確認できる。したがって、クチコミと広告種別との類似度の算出には、コサイン距離とユークリッド距離を併用した類似度が有用であることが分かった。コサイン距離やユークリッド距離のみを考慮した場合に分類精度が低下した原因は、図 4 で示した広告種別ごとのデータの分布から、TVCM とカタログのベクトルの向きが類似していることや、チラシとカタログのベクトルのノルムが類似していることにより、これらを誤

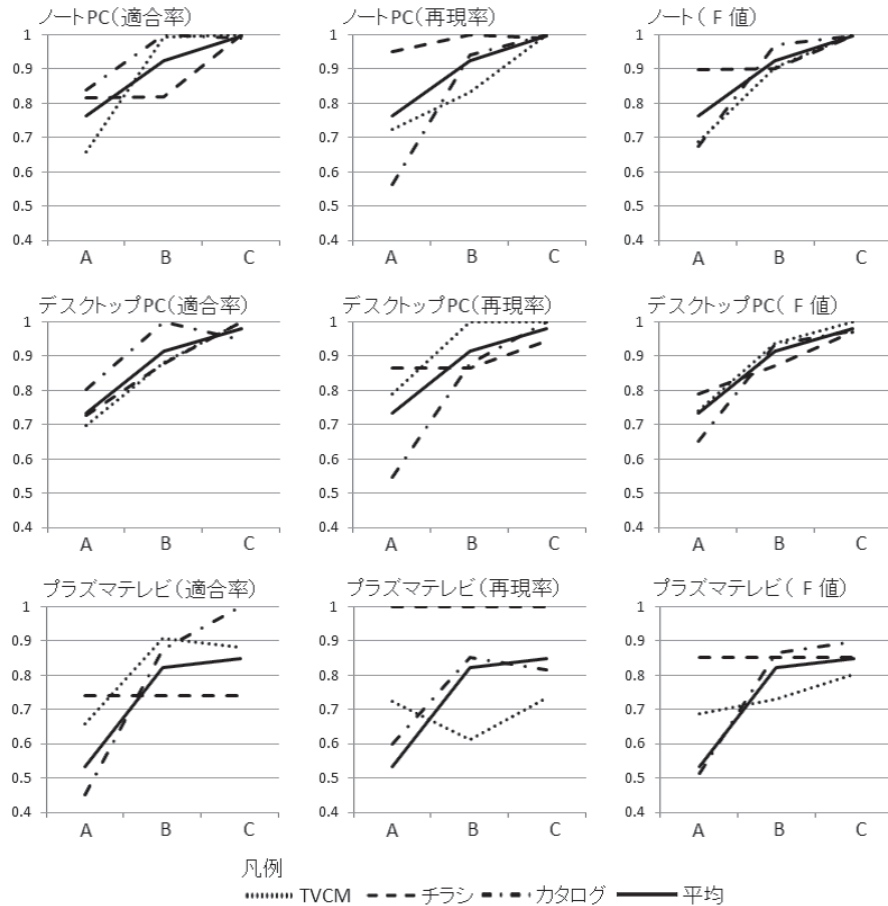


図 7 比較システムの分類精度の可視化

Fig. 7 Visualizing of accuracy of classification among three methods.

分類したためであると考えられる。

また、広告種別ごとの分類精度に注目すると、デスクトップ PC におけるカタログの適合率のように、比較システム B が C の分類精度を上回る場合があることが分かった。これは、システム B と C の類似度算出方法の違いや、システム A の分類精度がシステム B よ

表 8 比較システムの分類精度
Table 8 Accuracy of classification among three methods.

システム A	ノート PC			デスクトップ PC			プラズマテレビ		
広告種別	適合率	再現率	F 値	適合率	再現率	F 値	適合率	再現率	F 値
TVCM	0.66	0.72	0.69	0.70	0.79	0.74	0.66	0.72	0.69
チャラシ	0.82	0.95	0.90	0.73	0.87	0.79	0.74	1.00	0.85
カタログ	0.84	0.56	0.67	0.80	0.55	0.65	0.45	0.59	0.51
平均	0.76	0.76	0.76	0.73	0.73	0.73	0.53	0.53	0.53
システム B	ノート PC			デスクトップ PC			プラズマテレビ		
広告種別	適合率	再現率	F 値	適合率	再現率	F 値	適合率	再現率	F 値
TVCM	0.99	0.83	0.91	0.88	1.00	0.94	0.91	0.62	0.73
チャラシ	0.82	1.00	0.90	0.88	0.87	0.87	0.73	1.00	0.84
カタログ	1.00	0.94	0.97	1.00	0.87	0.93	0.87	0.85	0.86
平均	0.92	0.92	0.92	0.91	0.91	0.91	0.82	0.82	0.82
システム C	ノート PC			デスクトップ PC			プラズマテレビ		
広告種別	適合率	再現率	F 値	適合率	再現率	F 値	適合率	再現率	F 値
TVCM	1.00	1.00	1.00	1.00	1.00	1.00	0.88	0.73	0.80
チャラシ	1.00	0.99	0.99	1.00	0.94	0.97	0.74	1.00	0.85
カタログ	0.99	1.00	0.99	0.94	1.00	0.97	1.00	0.81	0.90
平均	0.99	0.99	0.99	0.98	0.98	0.98	0.85	0.85	0.85

り低いことから、システム C のみが考慮しているコサイン距離が分類精度に影響を与えていると思われる。この問題に対しては、コサイン距離とユークリッド距離に重み付けを行うことで解決できると考えられる。

さらに、ジャンルごとの分類精度に注目すると、いずれの比較システムにおいても、ノート PC やデスクトップ PC よりもプラズマテレビの分類精度が低下していることが分かった。これは、クチコミに含まれる話題の種類が他の 2 つよりも偏っていたことが原因だと考えられる。そこで、プラズマテレビに関するクチコミとして用意したテストデータを分析すると、およそ 50% のクチコミに画面に関する話題が記述されていたことが分かった。これは、テレビという性質上、消費者は、明るさや画質といった画面に関する話題に注目する傾向が強かったためと考えられる。このことから、プラズマテレビの網羅度の値は低く、詳細度の値は高いと予測される。しかし、図 6 においては、ノート PC やデスクトップ PC よりもプラズマテレビの網羅度の値が高く見える。これは、プラズマテレビのスペックの総数がノート PC やデスクトップ PC よりも少ないため、スペックの総数を分母とする網羅度の値が反比例して高くなったと考えられる。これらのことから、網羅度と詳細度は、クチコミに含まれる話題を利用した指標であるため、必然的にクチコミに含まれる話題に強く影響を

受けることが分かった。したがって、プラズマテレビのような話題に偏りがあるジャンルに対しては、網羅度と詳細度だけでなく、そのジャンルの話題に依存した指標を新たに導入する必要があると考えられる。

以上の実験結果から、特徴が類似する広告種別にクチコミを分類するために、本提案手法であるシステム C の類似度算出手法が有用であることが分かった。

6. 今後の課題

本研究では、クチコミを活用することで、クチコミが持つ情報の質の違いや消費者が持つ知識の量の違いを考慮していないという 2 つの課題を解決することを目的に、クチコミが持つ情報の質を定量化する指標に加えて、消費者が持つ知識の量に基づき、クチコミを分類する手法を提案した。また、実験により、網羅度と詳細度が購買意思決定過程モデルに定義された消費者の購買意欲を喚起する広告種別にクチコミを分類するときの有用な指標であることと、本研究で採用した類似度算出手法が有用であることを示した。しかし、網羅度と詳細度はクチコミに記述された話題に依存する指標であるため、対象商品のジャンルによっては分類精度が低下してしまう可能性があることが分かった。また、情報の質を情報の広さや深さととらえてクチコミを解析したが、他にも「情報の信頼性」といった観点でもとらえることができると考えられる。そこで、今後は、情報の信頼性を評価する指標についての検討と、本提案手法で分類したクチコミを消費者に提示することで購買意欲を喚起するかどうかを評価する被験者実験を行う予定である。

参 考 文 献

- 1) 経済産業省：消費者購買行動調査—リーマンショック以降の日本の消費者の実像 (2010).
- 2) 内閣府：平成 20 年度版国民生活白書 (2008).
- 3) KDDI：クチコミカタログ, 入手先(<http://au-is.jp/kuchikomi.html>) (参照 2011-02).
- 4) リクルート：この街のこんなところが好き! (高槻), 入手先(http://suumo.jp/edit/area/osaka/takatsuki/takatsuki_2420/) (参照 2011-02).
- 5) 池田佳代, 田邊勝義, 奥田英範, 奥 雅博：Blog からの体験情報抽出, 情報処理学会論文誌, Vol.49, No.2, pp.838-847, 情報処理学会 (2008).
- 6) 石野亜耶, 難波英嗣, 竹澤寿幸：旅行ブログエントリからの観光情報の自動抽出, 日本知能情報ファジィ学会誌, Vol.22, No.6, pp.667-679, 日本知能情報ファジィ学会 (2010).
- 7) 那須川哲哉, 金山 博：文脈一貫性を利用した極性付評価表現の語彙獲得, 自然言語処理研究会報告, Vol.2004, No.73, pp.109-116, 情報処理学会 (2004).
- 8) 高野敦子, 池奥渉太, 北村泰彦：因果関係に着目した口コミ Web サイトからの評価

表現抽出, 人工知能学会論文誌, Vol.24, No.3, pp.322-332, 人工知能学会 (2009).

- 9) Pang, B., Lee, L. and Vaithyanathan, S.: Thumps up?, Sentiment Classification using Machine Learning Techniques, *Proc. Conference on Empirical Methods in Natural Language Processing*, pp.77-86, Association for Computing Machinery (2002).
- 10) 安村禎明, 坂野大作, 上原邦昭：評判情報のレベルを考慮した評価文書の分類と評価情報の信頼性評価への応用, 自然言語処理, Vol.14, No.3, pp.297-313, 言語処理学会 (2007).
- 11) Schullz, D., Tannenbaum, S. and Lauterborn, R.: *Integrated Marketing Communications*, McGraw-Hill (1993).
- 12) 青木幸弘, 上田隆穂：マーケティングを学ぶ 下 売れ続ける仕組み, pp.186-210, 中央経済社 (2009).
- 13) Salton, G. and McGill, M.: *Introduction to Modern Information Retrieval*, McGraw-Hill (1983).
- 14) カカクコム：価格.com, 入手先(<http://kakaku.com/>) (参照 2011-02).
- 15) 日本新聞協会広告委員会：2009 年全国メディア接触・評価調査 (2010).
- 16) KDDI 総研：広告市場の変遷からみた通信と放送の今後 (2007), 入手先(<http://www.kddi-ri.jp/pdf/KDDI-RA-200704-21-PRT.pdf>) (参照 2011-02).
- 17) 博報堂 DY メディアパートナーズメディア環境研究所：2009 年メディア定点調査 (2009).
- 18) 新聞広告データアーカイブ：メディア別広告評価 (2009), 入手先(<http://www.pressnet.or.jp/adarc/data/data02/07.html>) (参照 2011-02).
- 19) ヤフー：関連検索ワード Web API, 入手先(<http://developer.yahoo.co.jp/webapi/search/assistsearch/v1/webunitsearch.html>) (参照 2011-02).
- 20) Mackay, D.: *Information Theory, Inference and Learning Algorithms*, pp.284-292, Cambridge University Press (2003).

(平成 23 年 3 月 20 日受付)

(平成 23 年 7 月 2 日採録)

(担当編集委員 関 洋平)



田中 成典 (正会員)

1986年関西大学工学部土木工学科卒業，1988年関西大学大学院工学研究科土木工学専攻博士課程前期課程修了．同年（株）東洋情報システム（現在，TIS）に入社．人工知能に関する研究受託開発業務に従事．1994年関西大学総合情報学部専任講師として着任，1997年助教授，2004年教授，2006年から学生センター副所長，現在に至る．2002年8月から1年間，カナダのUBCにて客員助教授．博士（工学）．専門は知識工学と社会基盤情報学．CAD/CG，GIS/GPS，画像処理およびWebソリューションビジネスに関する研究に従事．2000年（株）関西総合情報研究所を起業，設立当初から現在まで取締役会長．2006年（株）フォーラムエイトの顧問に就任，現在に至る．建設省土木研究所CAD製図基準検討委員会委員長，土木学会土木情報システム委員会幹事長，同委員会土木CAD小委員会委員長，ISO/TC184/SC4国内委員等を歴任．現在，国土交通省日本建設情報総合センター社会基盤情報標準化委員会委員，同委員会CAD/データ連携小委員会委員長，土木学会情報利用技術委員会副委員長．主に，ISOに準拠したCAD製図基準とCADデータ交換基盤の開発に従事．



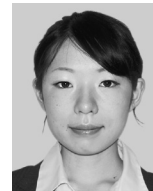
北野 光一 (学生会員)

1983年生．2006年関西大学総合情報学部総合情報学科卒業．2008年関西大学大学院総合情報学研究科知識情報学専攻博士課程前期課程修了．修士（情報学）．現在，関西大学大学院総合情報学研究科総合情報学専攻博士課程後期課程在学中．自然言語処理，データマイニングの研究に従事．



寺口 敏生

1984年生．2007年関西大学総合情報学部総合情報学科卒業．2009年関西大学大学院総合情報学研究科知識情報学専攻博士課程前期課程修了．修士（情報学）．現在，関西大学大学院総合情報学研究科総合情報学専攻博士課程後期課程在学中．WebマイニングおよびGISの研究に従事．



今城 彰子

1987年生．2009年関西大学総合情報学部総合情報学科卒業．2011年関西大学大学院総合情報学研究科知識情報学専攻博士課程前期課程修了．修士（情報学）．Webマイニングの研究に従事．



柳田 尚明

1988年生．2011年関西大学総合情報学部総合情報学科卒業．学士（情報学）．現在，関西大学大学院総合情報学研究科知識情報学専攻博士課程前期課程在学中．Webマイニングの研究に従事．