

選好商品のクラスタリングに基づく嗜好の変化の検出

松村冬子^{†1,†2} 廣安知之^{†3} 三木光範^{†4}
佐々木康成^{†4} 大向一輝^{†5,†6} 武田英明^{†5,†6,†7}

E-Commerce (EC) サイトでは、商品推薦を用いることで、より個々のユーザの嗜好に合った商品を提示し、売上げの向上を図っている。しかし、商品を探検中のユーザの嗜好や興味は、そのサイトへのアクセス開始時から提示された商品やその提示方法などの要因によって変化する場合がある。本論文ではユーザの嗜好がどのような要因によって変化するかを特定するため、その基礎的検討としてユーザの選好情報をクラスタリングすることで嗜好を把握し、そのクラスタリング結果の時系列変化を嗜好の変化として検出する手法について検討した。実験では、嗜好の変化を表現できるクラスタの特徴量を特定した。この特徴量に基づいて、提案手法によりエージェントによる嗜好の変化を検出可能であることを確認した。ただし、被験者実験においては、多くの被験者に対して探索のコンセプトの変化を検出することが困難であった。

Detection of the Preference Shift Timing Based on Clustered Products Preferred by a User

FUYUKO MATSUMURA,^{†1,†2} TOMOYUKI HIROYASU,^{†3}
MITSUNORI MIKI,^{†4} YASUNARI SASAKI,^{†4}
IKKI OHMUKAI^{†5,†6} and HIDEAKI TAKEDA^{†5,†6,†7}

E-commerce sites improve their sales by presenting products that match users' preferences using recommendation methods. A user's preference may change during the shopping. In this paper, this change is called a "preference shift". However, conventional methods suppose that user's preferences are static. Here, a new method is proposed, which responds to the preference shift and this method leads users to remain at the site longer than before. This paper discusses the detection method for finding the preference shift timing based on the transition of the clustering result of the products preferred by a user. It was confirmed that the proposed method can detect the preference shift of the agents using a feature of clusters which can describe the preference shift. However, the proposed method could not detect the concept drift of actual users in the subjective experiment.

1. はじめに

Web 上で商品の販売や購買を行う E-Commerce (EC) サイトの売上規模は年々増加している。また、多くの EC サイトでは、Amazon^{*1} のように商品推薦を用いることで、ユーザの過去の購買履歴やサイト上での行動履歴から嗜好のプロファイルを獲得し、それに基づいてユーザが購入する可能性が高いと予測される商品を提示している。本論文では、ユーザがあるときに EC サイトにアクセスし、商品購入を決定するまで、もしくはその EC サイトを離れるまでを「1 回のアクセス」と表現する。実際には 1 回のアクセスにおける商品探索中のユーザの嗜好は、過去にシステムを利用した時点やアクセス開始時から変化する可能性がある。しかし既存の推薦手法では、過去の様々な状況の嗜好のプロファイルを基に推薦を行うため、これらの嗜好の変化を考慮していない。本研究では、ユーザの動的な嗜好に対応した商品推薦を行うため、1 回のアクセスという短期間におけるユーザの嗜好の変化を検出することを目的とする。

EC サイトにおいては、サイトデザインなどの変化によってユーザが購入した商品が変化したという報告がある^{7)~9)}。そのため、提案手法によって嗜好が変化した時点を検出することで、その時点でユーザに提示されていた商品や提示方法などの情報から、そのユーザがどのような要因によって嗜好が変化しやすいか分析できる可能性がある。このようにして得られた嗜好の変化の分析結果を用いると、たとえばユーザの嗜好の変化を促すような商品提示を行うことで、サイトでの滞在時間を延ばし商品の販売機会の増加が促進できると考えら

†1 同志社大学大学院工学研究科

Graduate School of Engineering, Doshisha University

†2 日本学術振興会

Japan Society for the Promotion of Science

†3 同志社大学生命医科学部

Department of Life and Medical Sciences, Doshisha University

†4 同志社大学理工学部

Department of Science and Engineering, Doshisha University

†5 国立情報学研究所

National Institute of Informatics

†6 総合研究大学院大学

The Graduate University for Advanced Studies

†7 東京大学

The University of Tokyo

*1 <http://amazon.com/>

れる。

本論文では、ユーザの選好商品をクラスタリングすることで嗜好を把握し、そのクラスタリング結果の時系列変化を嗜好の変化として検出する手法について提案する。具体的には実験においては blog のデザインのテンプレートを選択するサイトを構築し、ユーザの嗜好の変化のモデルを設定したエージェントおよび被験者に対して、提案手法による嗜好の変化の検出性能を検討した。

2. EC サイトにおける嗜好の変化

人間の嗜好は様々な要因によって変化するが、これらの嗜好の変化は機械学習などの分野において古くから研究されており^{15),18)}、下記のように“段階性”¹⁶⁾という観点から2つに大別できる。また、EC サイトにおける嗜好の変化については、“期間”という新たな観点によっても分類可能と考えられる。

- 段階性
 - － 緩やかに嗜好が変化する場合
 - － 急激に嗜好が変化する場合
- 期間
 - － 1回のアクセス中に嗜好が変化する場合
 - － 複数回のアクセスを通して嗜好が変化する場合

本研究で扱う嗜好の変化は、上記の分類における嗜好の段階性としては双方を考慮しつつ、1回のアクセス中という短期間でユーザが理想とする商品の傾向が変化する場合として定義する。たとえば、あるユーザがドレスを探している場合を考える。図1に示したように、そのユーザは最初は黒いドレスばかりを探していたが、閲覧しているうちに画面下部に小さく提示されていた鮮やかな色のドレスも魅力的だと気づき、ピンクやオレンジなどのドレスを探すようになる状態を本研究で対象とする嗜好の変化とする。

嗜好の変化を考慮した商品推薦としては、Koren⁶⁾により開発された、生存期間中のデータの変化をすべて考慮する協調フィルタリングが存在する。この研究では Netflix^{*1}における商品のレーティングデータに対して、嗜好の変化の段階性が急激な場合と緩やかな場合、さらに様々な嗜好の変化の期間に合わせた適用方法やその有効性が示されている。しかし、最も変化が急激な場合であっても1日単位のログを考慮することにとどまっていることが



図1 EC サイトにおける嗜好の変化の例

Fig. 1 An example of preference shift on e-commerce sites.

ら、本研究において目的としている1回のアクセス中での嗜好の変化の検出やその利用には該当しない。また、本研究と同様に飽きのこない情報推薦を目的として、稲村ら²⁰⁾はユーザが飽きない弁当の献立作成を支援する研究を行っているが、この試みも最小の時間単位が1日であることから、本研究で対象とする短期間での嗜好の変化を扱っていない。

3. 本研究が対象とするシステムと嗜好を表現するモデル

3.1 対象とするシステム

本研究では、ユーザがECサイトにアクセスすると商品候補が提示され、ユーザが何らかの方法で自身の選好情報をシステムに提示すると、次の商品候補が提示されるようなシステムを想定している。たとえば、Amazonでは、商品の詳細情報にアクセスしたり商品を購入したりすると、その商品の関連商品が提示される。このような手続きの中で、Amazonのシステムではユーザの選好情報を獲得することが可能である。本研究では、商品を探している瞬間に理想とする商品の持つ傾向を嗜好として定義する。このため、選好商品の情報に基づいて特徴空間において嗜好に合う領域を対話的に特定することで、ユーザの嗜好を把握する。

3.2 一般的な嗜好のモデル

商品推薦に用いられる情報推薦手法は協調フィルタリング (collaborative filtering)^{5),14)}とコンテンツに基づいたフィルタリング (contents-based filtering)、そしてそれらを組み合わせたハイブリッドなアプローチの3つに大別できると報告されている¹⁾。コンテンツに基づいたフィルタリングでは、行動履歴から獲得した嗜好情報をユーザプロファイルとしてモデル化する。まず、対象となる商品がその複数の特徴量を要素とするベクトルで表現される場合、その空間で嗜好をモデル化するアプローチの1つとして、ユーザが興味を持ってク

*1 <http://www.netflix.com/>

リックするなどした嗜好商品の情報から特徴空間において嗜好に合う領域を特定する方法がある。この方法ではユーザの嗜好商品から対象となる特徴空間において嗜好に合う領域を特定したり、未知の商品と嗜好商品の類似度を余弦で表現したりすることでユーザの嗜好を表現している。

3.3 特徴空間におけるクラスタリングによる嗜好の把握

本研究では、先に述べたように、ユーザの行動から対話的に嗜好の情報を獲得するシステムを対象としている。一般に、ユーザは興味を持った商品をクリックし、その詳細を確認するため、ここではそれらの嗜好商品の集合からユーザの嗜好のモデルを特定するものとした。また、商品は特徴空間におけるベクトルで表すものとする。たとえば、対象商品が衣服である場合、生地の色や袖の長さなどが商品の特徴となり、その商品のベクトルに各特徴の値を設定する。

このように嗜好を嗜好商品の集合で表現する場合、それらの特徴量の選定や特徴空間における距離の扱いは嗜好の把握の精度に大きく影響する。特徴ベクトルの集合から嗜好を把握し、それに基づいて最適化を行う手法の1つとして対話型遺伝的アルゴリズム (interactive Genetic Algorithms: iGAs)¹⁷⁾があるが、一般にこれらの手法では最適化対象を表現する特徴量は適切に嗜好を表現できるように、専門家の意見や購買理由のアンケートの分析結果などから手動で決定される。また、各特徴量における心理的な距離の情報を被験者実験で求めて近傍を設計することは、色などパラメトリックな表現が可能な特徴量においても重要となるが、特に名義尺度のようなノンパラメトリックな特徴量では不可欠である。しかし、商品の種類や数が増えた場合にはこれらの特徴量の決定コストが増加してしまう。これより、著者らはECサイトが保持している商品どうしの関連を表現した隣接行列などから、主成分分析などを用いて自動的に特徴空間を抽出する手法などについても検討している³⁾。

ただし、ユーザの嗜好はつねに1つの傾向を示しているとは限らない。そのため、嗜好商品の集合に対して特徴空間上でクラスタリングを行うことで、ユーザが同時に持つ複数の嗜好を把握する。図1に示したドレスの探索の例において、ドレスが色と価格の2次元の特徴ベクトルで表現できる場合、閲覧したドレスの履歴は図2のように特徴空間にプロットすることができる。つまり、嗜好商品の時系列データは、ユーザが商品をクリックして閲覧するごとに、閲覧された商品がインクリメンタルにプロットされていくことになる。ただし、新しい傾向を持つデータの数や過去のデータに対して極端に少ない場合、その新たな傾向の影響が淘汰されてしまうため、最新 n 個のデータを嗜好の抽出の対象とするスライディングウィンドウ方式をとる。図2の例では、最新6個のデータを対象としている。ユーザの

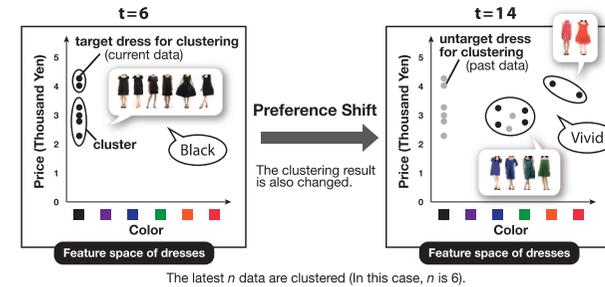


図2 特徴空間上でのクラスタリングと嗜好の変化の例

Fig. 2 An example of preference shift and clustering of preferred products on feature space of products.

嗜好が単一でない場合には、図2右側の $t = 14$ における特徴空間に示しているように多数のクラスタが存在する。すでに著者らは、Tシャツを対象商品とするiGAを用いたECサイトの実験により、嗜好商品へのクラスタリングによりユーザの複数の嗜好を把握可能であることを確認している^{4),19)}。このように、嗜好のモデルを嗜好商品のクラスタリングにより対話的に特定することが可能ではあるが、特定した嗜好のモデルは様々な要因により変化することがある。次章ではこれらの嗜好の変化を特定する方法について提案する。

4. クラスタリング結果の変化による嗜好の変化の検出

4.1 提案手法の概要

本研究では2章に示したように、商品推薦におけるユーザの嗜好の変化を1回のアクセス中での理想とする商品の傾向の変化と定義し、この嗜好の変化をユーザの嗜好商品のクラスタリング結果の変化を基に検出する手法を提案する。本研究では、この嗜好の変化をユーザが商品をクリックするごとに行われるクラスタリング結果の変化として表現する。そのため、例で示した閲覧したドレスのクラスタリング結果も、図2のように変化する。

4.2 提案手法の処理手順

提案手法では、嗜好情報の時系列データにデータが追加されるごとにクラスタリングを適用し、その変化を検出する。以下に提案手法の具体的な処理を示す。なお、時間は離散時間とし、嗜好商品が1つ追加されるごとに時刻 t が1だけ進むものとする。

Step 1 時刻 t において嗜好商品が追加されたら、特徴空間にプロットする。

Step 2 ウィンドウに含まれる最新 n 個の商品を特徴空間においてクラスタリングを行い、

クラスタ $A_0(t), A_1(t), \dots, A_i(t), \dots, A_m(t)$ を得る .

Step 3 時刻 t におけるクラスタ $A_0(t), A_1(t), \dots, A_i(t), \dots, A_m(t)$ と、時刻 $t-1$ におけるクラスタ $A_0(t-1), A_1(t-1), \dots, A_j(t-1), \dots, A_k(t-1)$ のすべての組合せにおいて、Jaccard 係数に基づき次式で 2 つのクラスタ間の相関を算出し^{12),13)}、相関が高い組合せから同一のクラスタであるとする .

$$C_{-A_{ij}}(t) \equiv \frac{|A_i(t-1) \cap A_j(t)|}{|A_i(t-1) \cup A_j(t)|} \quad (1)$$

Step 4 時刻 $t-1$ と時刻 t における同一のクラスタ $A_i(t-1)$ と $A_i(t)$ の差の特徴量を算出し、特徴量もしくはその特徴量の過去のデータから予測されたモデルに対する残差が閾値よりも大きければクラスタリング結果が変化したと判断する .

Step 5 嗜好商品の追加が終わるまで、Step 1 から 4 を繰り返す .

上記の手順のとおり、提案手法では時刻 $t-1$ におけるクラスタ $A_i(t-1)$ と時刻 t におけるクラスタ $A_i(t)$ が同じクラスタであると見なせるときに、 $A_i(t-1)$ と $A_i(t)$ を比較することで変化を検出する . 最も単純な時系列クラスタリングの方法としては、ユーザの嗜好商品が追加されるごとに、その時点までに蓄積されている嗜好商品のデータに対してクラスタリングを行うことが考えられる . しかし、長期間にわたって大量のデータが蓄積されている場合、多くのデータによって形成されているクラスタ群とはまったく異なる性質を持つ新しいデータが少しずつ追加されても、クラスタリング結果はあまり変化せず、嗜好の変化をとらえられない可能性がある . そのため、クラスタリングに適用するデータの選択が必要となる . これに対して、本研究では上記のように新しい順から一定数 n のサンプルデータを用いるスライディングウィンドウ方式をとる .

また、提案手法では時刻 $t-1$ におけるクラスタ $A_i(t-1)$ と時刻 t におけるクラスタ $A_i(t)$ を比較して嗜好の変化を表現できるクラスタの特徴量が必要となる . ここでは、嗜好を表現するクラスタが含むデータの分布から、それらの分散共分散行列の固有値を長径および短径とする超楕円体で近似する . 時刻 $t-1$ と時刻 t における同一のクラスタ $A_i(t-1)$ と $A_i(t)$ を比較するため、

- $A_i(t-1)$ と $A_i(t)$ に含まれるデータの重心間の距離 $d_c(t)$
- $A_i(t-1)$ と $A_i(t)$ を近似した超楕円体が特徴空間に占める領域の差 $d_S(t)$

といった 2 つの超楕円体の物理的特徴量を嗜好の変化を表すクラスタの差の特徴量の候補とする . 特徴空間が 2 次元の場合の 2 つのクラスタの変化の特徴量の例を図 3 に示した . これら以外にも物理的特徴量は存在するが、各クラスタの距離情報および領域情報の代表とし

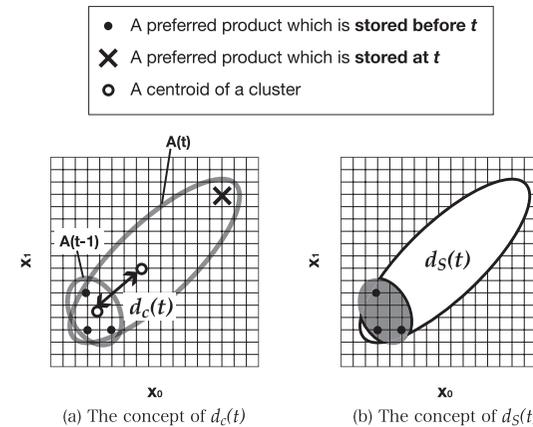


図 3 2 つのクラスタの重心間の距離 $d_c(t)$ および領域の差 $d_S(t)$

Fig. 3 $d_c(t)$ is the distance between centroids of two clusters. $d_S(t)$ is the difference between the regions of two clusters.

てこれらを選択した .

次章以降では、 $d_c(t)$ および $d_S(t)$ についてクラスタの特徴量としての検討を行う . 次に、その特徴量を用いて嗜好が変化した時刻を検出する方法について検討を行い、最後に被験者実験によって実際の人間の嗜好の変化を検出可能か検討を行う .

5. 嗜好の変化を検出可能なクラスタの特徴量の検討

5.1 検討概要

提案手法の中で、どのような特徴量がユーザの嗜好の変化を適切に表現できるのかは分かっていない . そのため本章では、現時点と前の時点との比較でクラスタリング結果の変化を表現するために用いるクラスタの特徴量について検討を行う . 具体的には、エージェントに設定した嗜好の変化に合わせて、それぞれのクラスタの特徴量に変化するかを確認する . また、クラスタの特徴量の変化が人間の主観により観察される嗜好の変化と一致するかを、被験者にエージェントの嗜好履歴を閲覧させて嗜好の変化を観察した時刻を調査することで検討する .

5.1.1 実験システム

本検討は実験システムを構築し、実験を通じて行う . 本実験では、嗜好の変化を表現可能

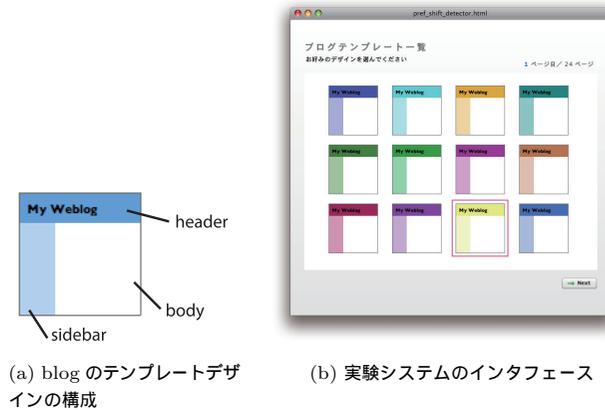


図 4 実験システムのインタフェースと対象商品の表現
Fig. 4 Experimental setup and the expression of a product.

なクラスタの特徴量について基礎的な検討を行うため、図 4(a) に示した blog のテンプレートデザインを対象商品とする EC サイトのシステムを用いた。テンプレートは、図 4(a) に示したようにヘッダ、サイドバー、ボディの 3 つの部分から構成されている。その色の表現には人間の色彩感覚に類似しているとされる HSB (Hue Saturation Brightness) 表色系を用いた。ただし、自然な配色となるようにヘッダとサイドバーの色は同じとし、サイドバーのみ透明度を 0.4 とした。また、題字が黒である場合のテンプレートデザインとして自然な色調になるように彩度 (Saturation) と明度 (Brightness) を 0.5 以上に設定した。ボディの色は白色で固定した。

HSB 表色系では、色相 (Hue) ($0 \leq H \leq 360$)、彩度 ($0 \leq S \leq 1$)、明度 ($0 \leq B \leq 1$) の 3 つの値で色を表現する。なお、色相の軸は 0 と 360 が等しく、循環している。そのため、テンプレートの特徴空間はそのヘッダの色の色相 H 、彩度 S 、明度 B の 3 次元で表現される。ヘッダとサイドバーの色相の範囲は $0 \leq H \leq 360$ 、彩度の範囲は $0.5 \leq S \leq 1.0$ 、明度の範囲は $0.5 \leq B \leq 1.0$ とすべて同じ実数の範囲である。

次に実験システムでは画面の単位をページとし、図 4(b) に示したように 1 ページごとに 12 個のテンプレートを提示するようにした。なお、1 ページに提示するテンプレートの多様性を保つため、各ページの 12 個のテンプレートには色相の範囲から均等に色相の値を設定した。エージェントもしくは被験者は 24 ページにわたって 288 個のテンプレートを閲覧

し、各ページにおいて最も嗜好に合うテンプレートを 1 つだけクリックして選択するものとした。選択されたテンプレートは図 4(b) に示したように、赤色の枠で囲んで表示した。

5.1.2 エージェントにより生成される嗜好情報の時系列データ

本実験では、ユーザの嗜好の変化のモデルとして下記の各嗜好を持つエージェントに全 24 ページの提示テンプレートから 24 個のテンプレートを選択させて 4 種類の嗜好情報の時系列データを用意した。なお、単位時間は 1 つのテンプレートの追加とするため、 $1 \leq t \leq 24$ となる。

(1) 単一の嗜好が変化する場合

探索の前半は、特徴空間上でエージェントの嗜好に合う範囲を R_{1a} ($60 \leq H \leq 150$, $0.5 \leq S \leq 0.75$, $0.75 \leq B \leq 1.0$) とし、後半から R_{1b} ($270 \leq H \leq 360$, $0.75 \leq S \leq 1.0$, $0.5 \leq B \leq 0.75$) に変化すると定義する。実際には、エージェントは実験システムにより提示されたテンプレートの中から、12 ページにわたって R_{1a} に相当するテンプレートを選択し、残りの 12 ページにわたって R_{1b} に相当するテンプレートを選択する。これより、このエージェントの嗜好が変化した時刻は $t = 13$ となる。このエージェントの嗜好は急激な変化を表現している。

(2) 複数の嗜好のうち一方が変化する場合

探索の最初はエージェントの嗜好に合う範囲が R_{2a} ($90 \leq H \leq 150$, $0.5 \leq S \leq 0.75$, $0.75 \leq B \leq 1.0$) と R_{2b} ($180 \leq H \leq 240$, $0.5 \leq S \leq 0.75$, $0.75 \leq B \leq 1.0$) とし、残りの期間は R_{2a} の嗜好が R_{2c} ($300 \leq H \leq 360$, $0.5 \leq S \leq 0.55$, $0.5 \leq B \leq 0.55$) に変化すると定義する。なお、 R_{2b} は探索中に一貫して嗜好に合う範囲であると定義する。実際には、エージェントは提示されたテンプレートの中から、12 ページにわたって R_{2a} に含まれるテンプレートを 8 個、 R_{2b} に含まれるテンプレートを 4 個選択する。次の 12 ページにわたっては、 R_{2c} に含まれるテンプレートを 8 個、 R_{2b} からテンプレートを 4 個選択する。これより、このエージェントの嗜好が変化した時刻は $t = 14$ となる。このエージェントの嗜好は緩やかな変化を表現している。

(3) 複数の嗜好が別の単一の嗜好に変化する場合

探索の最初はエージェントの嗜好に合う範囲が R_{3a} (R_{2a} に同じ) と R_{3b} (R_{2b} に同じ) であるとし、残りの期間は R_{3a} , R_{3b} の嗜好の双方が R_{3c} (R_{2c} に同じ) に収束すると定義する。実際にはエージェントは、16 ページにわたって、 R_{3a} と R_{3b} それぞれから 8 個のテンプレートをランダムな順序で選択し、17 ページ以降では R_{3c} に含まれる 8 個のテンプレートを選択する。これより、この時系列データにおいて嗜好が変化し

た時刻は $t = 17$ である．このエージェントの嗜好は急激な変化を表現している．

(4) 嗜好が変化しない場合

エージェントは一貫して R_{4a} ($180 \leq H \leq 270, 0.5 \leq S \leq 0.75, 0.75 \leq B \leq 1.0$) に含まれる 24 個のテンプレートを選択し続ける．これより，この時系列データにおいては嗜好は変化しない．

5.1.3 クラスタリング手法

クラスタリング手法としては，ネットワークからコミュニティを抽出する手法として用いられている Newman 法¹¹⁾ を，重み付きネットワークに適用できるように改良した手法を用いる．Newman 法は，階層的クラスタリング手法の 1 つであり，クラスタの分離性と均質性を考慮した *modularity* Q ¹⁰⁾ という指標の最大化を行うことで最も適当なクラスタ分割を求めることができる．そのため Newman 法においては，クラスタ数は自動的に決定される．

具体的には，クラスタリング対象のノードが k 個ある場合，最初に各ノード x_i をそれぞれクラスタ A_i として， e_{ij} を A_i と A_j の間にあるエッジの重みの和とする $k \times k$ の対称行列 e を作成する．次に， $a_i = \sum_j e_{ij}$ として計算する．つまり， e_{ii} はクラスタ A_i 内のノード間で結ばれているエッジの重みの和を， a_i はクラスタ A_i 内のノードが持つすべてのエッジの重みの和を表しており， Q は以下に示すように，クラスタ内でのつながりを強め，クラスタ間のつながりを弱くするように設計されている．

$$Q = \sum_i (e_{ii} - a_i^2) \tag{2}$$

Newman 法では，重みなしネットワークのノード間の関係をエッジの有無で表現していたが，本研究では各商品のデータ間の類似度をエッジの重みとするネットワークを対象とする．そのため，特徴空間の各データの座標から，2 つのデータ x_i と x_j の間の距離の逆数を類似度として用いる．また，クラスタリングの対象とするデータはスライディングウィンドウ方式により最新の n 個のデータを用いるものとする．本実験では $n = 9$ とした．

$$\text{Similarity}(x_i, x_j) = \frac{1}{\text{Distance}(x_i, x_j)} \tag{3}$$

5.1.4 実験手順

被験者は 20 歳代の男女 8 名 (男性 6 名，女性 2 名) であった．まず，エージェントが各ページにおいて提示されたテンプレート群と，エージェントによりそのページで選択された商品を各ページについて図 4 (b) のように表示し，紙に印刷した．1 ページから 24 ページ

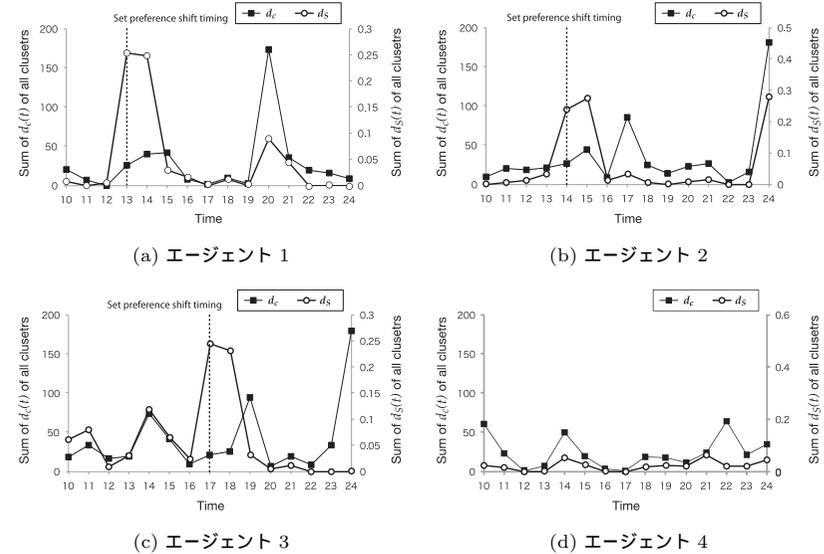


図 5 エージェントの嗜好に対する $d_c(t)$ と $d_S(t)$ の合計値の推移
Fig. 5 Transitions of sum of $d_c(t)$ and sum of $d_S(t)$ of preferred products of agents.

にわたり，印刷されたテンプレート群と選択されたテンプレートを被験者に選択の時系列が分かるように一度に提示した．被験者は提示されたエージェントによるテンプレートの選択過程を見て，「エージェントの嗜好が一定していたか」「エージェントの嗜好はどのページ間で変化したか」を回答した．

5.2 実験結果と考察

各エージェントによるテンプレートの選択についてのクラスタの特徴量の変化を検討する．図 5 に各エージェントによる嗜好情報の時系列データの時刻 t におけるすべてのクラスタの $d_c(t)$ ， $d_S(t)$ の合計値の推移を示した．図 5 の左側の縦軸は $d_c(t)$ の合計値を，右側の縦軸は $d_S(t)$ の合計値を表している．横軸は時刻 t を表している．図中の点線はエージェントに嗜好の変化を設定した時刻を示している．なお，クラスタリングはノード数がウィンドウのサンプル数 $n = 9$ と等しくなった時点から適用を開始している．

まず，単一の嗜好が変化するエージェント 1 についてクラスタの特徴量がどのように変化しているかを検証する．図 5 (a) から分かるように，エージェントの嗜好が変化する $t = 13$

で $d_S(t)$ は急激に増加している．一方、 $d_c(t)$ は $t = 13$ では変化がなく、逆に嗜好の変化が設定されていない $t = 20$ で大きく増加していることが分かる．

次に、複数の嗜好のうち一方が変化するエージェント 2 について検証する．図 5 (b) より、エージェントの嗜好が変化する $t = 14$ で $d_S(t)$ は急に増加するが、 $d_c(t)$ では $t = 14$ においてはそのような変化は見られなかった．また、 $d_c(t)$ と $d_S(t)$ の双方ともに $t = 24$ で大きく増加してしまう現象が見られた．これはスライディングウィンドウ方式を採用しており、最新 n 個のデータのみをクラスタリングの対象としていることから、探索序盤のみに好まれていたクラスタに含まれるデータがウィンドウから除外された際に特徴量が大きく変化したと考えられる．

複数の嗜好が別の単一の嗜好に変化するエージェント 3 については、図 5 (c) より、エージェントの嗜好が変化する $t = 17$ で $d_S(t)$ は急に増加するが、 $d_c(t)$ では $t = 17$ においてそのような変化はなく、 $t = 24$ で大きく増加してしまうことが分かった．

嗜好が変化しないエージェント 4 については、図 5 (d) より分かるように $d_c(t)$ 、 $d_S(t)$ の双方ともに変化が少ないことが分かる．

これより、すべてのデータに対して、 $d_c(t)$ と比較して $d_S(t)$ の方がエージェントに設定した嗜好の変化に沿って増減していることが分かる．また、エージェント間で $d_S(t)$ の推移を比較すると、エージェント 1 および 3 については嗜好の変化を設定した時刻において大きく増加しているのに対して、エージェント 2 では増加はしているものの他の時刻においてより大きく増加していることが分かる．これはエージェント 2 のような緩やかな変化と比較して、エージェント 1 および 3 のような嗜好の急激な変化の方が $d_S(t)$ で表現しやすいことを示唆している．

次に、クラスタの特徴量が被験者の主観により観察された嗜好の変化を表現できているかを検証する．ここでは $d_c(t)$ より $d_S(t)$ の方がエージェントに設定した嗜好の変化を反映していることから、図 6 に $d_S(t)$ と時刻 t において嗜好が変化したと判断した被験者数の推移を示した．左側の縦軸は時刻 t において嗜好が変化したと回答した被験者数を、右側の縦軸は $d_S(t)$ の合計値を表している．横軸は時刻 t を表している．これより、どのエージェントのデータの $d_S(t)$ も嗜好の変化をより多くの被験者が判断した時刻において大きく変化していることが分かった．これより、 $d_S(t)$ は人間の主観から観察される嗜好の変化を表現可能であると考えられる．

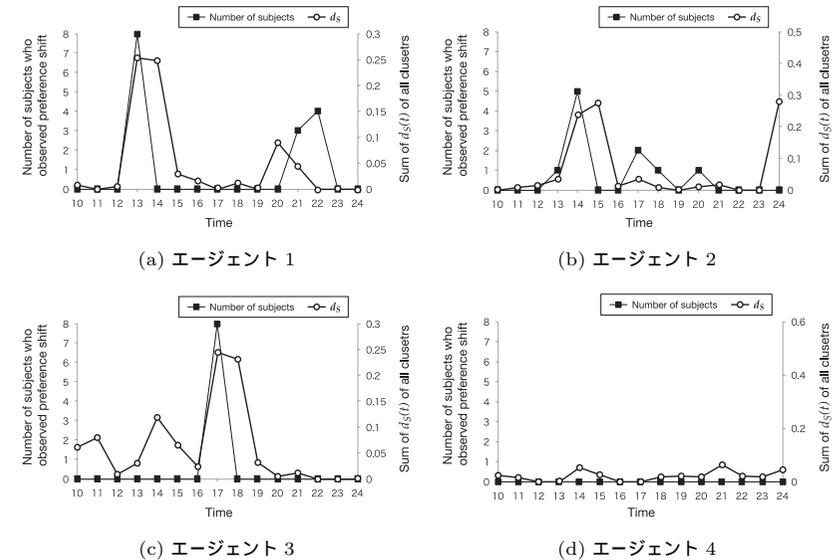


図 6 エージェントの選好に対する $d_S(t)$ の合計値と嗜好の変化を観察した被験者数の推移

Fig. 6 Transitions of sum of $d_S(t)$ of preferred products of agents and the number of subjects who observed preference shift.

6. クラスタの特徴量による嗜好の変化の検出方法の検討

6.1 AR モデルによる嗜好の変化の検出

5 章の実験結果より、クラスタ内のデータが特徴空間に占める領域の差 $d_S(t)$ により嗜好の変化を表現できる可能性が確認されたことから、本章では $d_S(t)$ を用いて嗜好の変化が生じた時刻を検出する手法について検討する．一般に時系列データからの異常検出などを行う場合には、自己回帰モデル (Auto Regressive Model, 以降 AR モデル)²⁾ や自己回帰移動平均モデル (Auto Regressive Moving Average Model, 以降 ARMA モデル) などの線形離散時間システムを用いて時系列データのモデルを予測し、現在時刻の予測値と実測値の残差を求め、その残差が大きい際に異常として検出する．ここでは、時刻 $t - 1$ までの $d_S(t)$ を用いて得た次式の p 次の AR モデルによる時刻 t における残差を算出する．なお、 $y(t)$ は時刻 t におけるシステムの出力を表し、 $a(k)$ とは任意の定数であり AR パラメータと呼ばれる．ここでは Yule-Walker 方程式を用いて、最適な AR パラメータと次数 p を求めた．

32 嗜好商品のクラスタリングに基づく嗜好の変化の検出

表 1 $d_S(t)$ および $d_S(t)$ により予測した AR モデルの残差 $r(t)$ の閾値により嗜好の変化が検出された時刻 (下線は嗜好の変化の設定時刻を件検出した時刻, *は嗜好を検出しなかった時刻を指す)

Table 1 Preference shift timings are detected by thresholds of $d_S(t)$ and its residual error $r(t)$.

The detected timing is underlined when it is equal to set preference shift timing.

* is marked when no preference shift is detected.

threshold	$d_S(t)$		$r(t)$		correct timing
	$mean + 1SD$	$mean + 2SD$	$mean + 1SD$	$mean + 2SD$	
agent 1	<u>13</u> , 14	<u>13</u> , 14	<u>13</u> , 14	<u>13</u>	13
agent 2	<u>14</u> , 15	15	22, 23	22, 23	14
agent 3	<u>17</u> , 18	<u>17</u> , 18	<u>17</u>	<u>17</u>	17
agent 4	14, 21, 24	21	18, 19, 20, 21	*	*

$$y(t) = \sum_{k=1}^p a(k)y(t-k) + \epsilon(t) \quad (4)$$

6.2 クラスタの特徴量と AR モデルによる嗜好の変化の検出の比較

ここでは、嗜好の変化が生じた時刻を特定するために、クラスタの特徴量 $d_S(t)$ の時系列データ、および $d_S(t)$ から AR モデルを予測して得た残差 $r(t)$ の時系列データの中で外れ値が生じた時刻を嗜好の変化した時刻として検出した。ここで外れ値を判定する閾値の候補として、それぞれの時系列データの平均 $+1SD$ (Standard Deviation, 標準偏差), 平均 $+2SD$ を用いた。これらの閾値により、5 章における各エージェントの嗜好情報の時系列データから嗜好が変化すると検出された時刻 t の一覧を表 1 に示した。なお、エージェントに嗜好を設定した時刻を検出できた場合は表中の t に下線を付与した。

表 1 より、 $d_S(t)$ の外れ値では平均 $+1SD$ のときにエージェント 1, 2, 3 についてエージェントに設定された嗜好が変化する時刻を検出できているが、嗜好が変化しないエージェント 4 の場合に多くの誤検出が生じていることが分かる。一方、 $d_S(t)$ の予測モデルの残差である $r(t)$ の外れ値では、平均 $+2SD$ のときにエージェント 1, 3 についてエージェントに設定された嗜好が変化する時刻を検出でき、嗜好の変化が設定されていないエージェント 4 については嗜好の変化をいっさい検出しなかった。ただし、緩やかに嗜好が変化する場合のエージェント 2 の場合においてのみ $r(t)$ では設定された嗜好が変化した時刻を検出することができなかった。これは、エージェント 2 において嗜好の変化が設定された $t = 14$ よりも $t = 22, 23$ において $r(t)$ が増大していたためと考えられる。今後、ARMA モデルや自己回帰和分移動平均モデル (Auto Regressive Integrated Moving Average Model, ARIMA モデル) などその他の線形離散時間システムによって緩やかな変化を検出可能か検討する。

7. 被験者実験による嗜好の変化の検出の検討

本章では実際に被験者に blog のテンプレートデザインを選択させ、その嗜好の変化を検出可能かを検証した。ここでは、被験者に探索途中に理想とするテンプレートのコンセプトを変更するように指示することで、意図的な探索目標のコンセプトの変化を発生させ、5 章や 6 章において検討した手法を用いて検出が行えるかを確認する。

7.1 実験概要と手順

本実験では被験者に 5 章におけるエージェントでの実験と同様に、5.1.1 項に示した実験システムにおいて blog のテンプレートを 24 個選択させた。テンプレートの特徴量は 5.1.1 項と同様にテンプレートのヘッダの色相, 明度, 彩度の 3 次元とした。このとき、探索前半の $t = 1$ から $t = 12$ では「読者が男性 (女性) である blog」というコンセプトで、探索後半の $t = 13$ から $t = 24$ においては逆に「読者が女性 (男性) である blog」というコンセプトでテンプレートを選択させ、 $t = 13$ でコンセプトの変化が生じるように被験者に指示した。なお、被験者は 20 歳の男女 10 名 (男性 6 名, 女性 4 名) であり、これらのコンセプトの提示順は被験者間でカウンタバランスをとった。

7.2 実験結果と考察

本実験では被験者に $t = 13$ においてコンセプトを変更してテンプレートを選択するように指示した。クラスタの特徴量 $d_S(t)$ とその予測モデルの残差 $r(t)$ により 6 章の実験で検討した閾値によってコンセプトが変化した時刻の検出を行った。まず、 $d_S(t)$ による検出結果としては、被験者 10 名のうち平均 $+1SD$ により検出できたのは 1 名, 平均 $+2SD$ により検出できた被験者はいなかった。また、 $r(t)$ による検出結果において、平均 $+1SD$ により検出できたのは 4 名, 平均 $+2SD$ により検出できたのは 3 名にとどまった。これより、双方ともに多くの被験者のコンセプトの変化をとらえることができなかったが、 $d_S(t)$ と比較すると $r(t)$ の方がコンセプトの変化を検出できていることが分かった。

図 7 にある被験者による「読者が男性である blog」というコンセプトに基づいた $t = 13$ から $t = 22$ までの選択履歴を示した。図 7 から、同じ「読者が男性である blog」というコンセプトの中でも、この被験者は $t = 18$ からはそれ以前と比較して彩度と明度が低い暗めのテンプレートを選択していることが分かる。図 7 の例からも分かるように、多くの被験者についてコンセプトの変化を検出できなかった原因として、被験者によってはコンセプトに合うテンプレートデザインの特徴が多岐にわたる場合が考えられる。つまり、同じコンセプトでの探索においても、複数の嗜好に対する選好がコンセプトの変化よりも強い嗜好の変

33 嗜好商品のクラスタリングに基づく嗜好の変化の検出

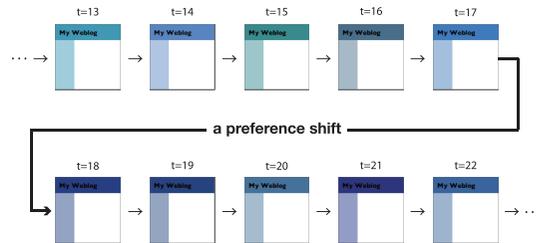


図7 ある被験者による「読者が男性であるblog」のコンセプトに基づいたテンプレートの選好履歴
Fig. 7 A history of template selection by a subject based on a concept of “blog for male.”

化として検出される場合があったためと推測される。今後は単一のコンセプトに基づいて探索する場合について被験者実験により検証する必要がある。

8. 結 論

本研究では、嗜好の変化に対応した新しい推薦手法を実現するため、ユーザの嗜好の変化の検出を目標としている。そこで本論文ではユーザの嗜好商品をクラスタリングすることで嗜好を把握し、嗜好の変化を表現できるクラスタの特徴量、およびクラスタリング結果の時系列変化を嗜好の変化として検出する手法を提案し、blogのテンプレートデザインを対象商品とする実験システムを構築して提案手法の検討を行った。

まず嗜好の変化を表現可能なクラスタの特徴量について、4種類の嗜好の変化を設定したエージェントにより検討を行った結果、クラスタに含まれるデータが特徴空間に占める領域の差である $d_S(t)$ により人間の主観に合った嗜好の変化を検出できることが確認された。また、 $d_S(t)$ による嗜好の変化の検出においては緩やかな変化よりも急激な変化を正確に把握できることが分かった。

さらに嗜好が変化した時刻の検出には $d_S(t)$ の時系列データにより予測されたARモデルの残差 $r(t)$ を用いることで、エージェントに嗜好の変化を設定した時刻を検出可能であることが分かった。

一方、被験者実験において意図的に探索途中に導入したコンセプトの変化は多くの被験者において提案手法により検出することが困難であった。これは、男性向けや女性向けというそれぞれのコンセプトが複数の嗜好によって構成されていることから、1つのコンセプトによる探索中においても被験者の嗜好が変化する場合があり、誤検出などを生じることが原因だと考えられる。

本論文における実験では、blogのテンプレートデザインという特定の特徴空間において、提案手法が嗜好の変化を検出できるかを確認し、また嗜好を表現する適切なクラスタの特徴量などを検討した。他の特徴空間においても、提案手法および検討されたクラスタの特徴量が有効であるかどうかは確認が必要であるが、近傍が適切に設計されている空間であれば、定性的に見て提案手法により嗜好の変化を検出できる可能性が示唆された。

今後はウィンドウのサンプル数の検討、ARモデル以外の線形離散時間システムの利用の検討、被験者実験による単一のコンセプトに基づいた嗜好の変化の検出可能性の検討などを引き続き行う。

参 考 文 献

- 1) Adomavicius, G. and Tuzhilin, E.: Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions, *IEEE Trans. Knowledge and Data Engineering*, Vol.17, No.6, pp.734–749 (2005).
- 2) Box, G.E.P. and Jenkins, G.M.: *Time Series Analysis; Forecasting and Control*, Holden-Day, San Francisco (1970).
- 3) Hiroyasu, T., Tanaka, M., Miki, M. and Yokouchi, H.: Extraction of Design Variables using Collaborative Filtering for interactive Genetic Algorithms, *Proc. 18th IEEE International Conference on Fuzzy Systems*, pp.1579–1584 (2009).
- 4) Ito, F., Hiroyasu, T., Miki, M. and Yokouchi, H.: Discussion of Offspring Generation Method for Interactive Genetic Algorithms with Consideration of Multimodal Preference, *Simulated Evolution and Learning*, Lecture Notes in Computer Science, Vol.5361, pp.349–359, Springer (2008).
- 5) Konstan, J.A., Miller, B.N., Maltz, D., Herlocker, J.L., Gordon, L.R. and Riedl, J.: Grouplens: applying collaborative filtering to usenet news, *Comm. ACM*, Vol.40, No.3, pp.77–87 (1997).
- 6) Koren, Y.: Collaborative Filtering with Temporal Dynamics, *The 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp.447–456 (2009).
- 7) Koufaris, M., Kambil, A. and LaBarbera, P.A.: Consumer Behavior in Web-Based Commerce: An Empirical Study, *International Journal of Electronic Commerce*, Vol.6, No.2, pp.115–138 (2002).
- 8) Mandel, N. and Johnson, E.J.: When Web Pages Influence Choice: Effects of Visual Primes on Experts and Novices, *Journal of Consumer Research*, Vol.29, No.2, pp.235–245 (2002).
- 9) Mandel, N. and Nowlis, S.M.: The Effect of Making a Prediction about the Outcome of a Consumption Experience on the Enjoyment of That Experience, *Journal*

34 選好商品のクラスタリングに基づく嗜好の変化の検出

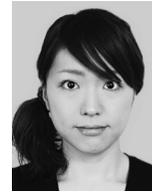
of *Consumer Research*, Vol.35, No.1, pp.9–20 (2008).

- 10) Newman, M.E.J. and Girvan, J.: Finding and evaluating community structure in networks, *Physics Review E*, Vol.69, Issue 2, 026113 (2004).
- 11) Newman, M.E.J.: Fast algorithm for detecting community structure in networks, *Physics Review E*, Vol.69, Issue 6, 066133 (2004).
- 12) Palla, G., Derenyi, I., Farkas, I. and Vicsek, T.: Uncovering the overlapping community structure of complex networks in nature and society, *Nature*, Vol.435, No.7043, pp.814–818 (2005).
- 13) Palla, G., Barabasi, A.L. and Vicsek, T.: Quantifying social group evolution, *Nature*, Vol.446, No.7136, pp.664–667 (2007).
- 14) Sarwar, B., Karypis, G., Konstan, J. and Reidl, J.: Item-based collaborative filtering recommendation algorithms, *Proc. 10th international conference on World Wide Web*, pp.285–295 (2001).
- 15) Schlimmer, J. and Granger, R.: Beyond incremental processing: Tracking concept drift, *Proc. 5th National Conference on Artificial Intelligence*, pp.502–507 (1986).
- 16) Stanley, K.O.: Learning concept drift with a committee of decision trees, Tech. Report UT-AI-TR-03-302, Department of Computer Sciences, University of Texas at Austin, USA (2003).
- 17) Takagi, H.: Interactive evolutionary computation: Fusion of the capabilities of ec optimization and human evaluation, *Proc. IEEE*, Vol.89, No.9, pp.1275–1296 (2001).
- 18) Widmer, G. and Kubat, M.: Learning in the presence of concept drift and hidden contexts, *Machine Learning*, Vol.23, No.1, pp.69–101 (1996).
- 19) 伊藤冬子, 廣安知之, 三木光範, 横内久猛: 対話型遺伝的アルゴリズムにおける嗜好の多峰性に対応可能な個体生成方法, *人工知能学会論文誌*, Vol.24, No.1, pp.127–135 (2009).
- 20) 稲村博央, 野間裕子, 荻野晃大, 庄司裕子: 飽きずに継続利用できる情報推薦の実現に向けた試み, *電子情報通信学会第二種研究会技術研究報告 (第 15 回 Web インテリジェンスとインタラクション研究会)*, WI2-2009 (2009).

(平成 21 年 8 月 16 日受付)

(平成 21 年 10 月 3 日再受付)

(平成 21 年 10 月 16 日採録)



松村 冬子 (学生会員)

2005 年同志社大学工学部 3 年次修了, 退学. 2007 年同大学院工学研究科博士前期課程修了. 現在, 同大学院同研究科博士後期課程および日本学術振興会特別研究員 (DC2). Web における知識流通のための感性情報の獲得と利用に興味を持つ. IEEE, 人工知能学会, 日本知能情報フレンジイ学会各学生会員.



廣安 知之 (正会員)

1997 年早稲田大学大学院理工学研究科後期博士課程修了. 早稲田大学理工学部助手, 同志社大学工学部インテリジェント情報工学科准教授を経て, 2008 年から生命医科学部医情報学科教授. 進化的計算, 最適設計, 並列処理, 設計工学, 医療画像工学等の研究に従事. IEEE, 電子情報通信学会, 計測自動制御学会, 日本機械学会, 超並列計算研究会各会員.



三木 光範 (正会員)

1978 年大阪市立大学大学院工学研究科博士課程修了, 工学博士. 大阪市立工業研究所研究員, 金沢工業大学助教授を経て, 1987 年大阪府立大学工学部航空宇宙工学科助教授, 1994 年同志社大学理工学部教授. 進化的計算手法とその並列化, および知的なシステム的设计に関する研究に従事. 著書は『工学問題を解決する適応化・知能化・最適化法』(技法堂出版)等多数. IEEE, 米国航空宇宙学会, 人工知能学会, システム制御情報学会, 日本機械学会, 計算工学会, 日本航空宇宙学会等各会員. 超並列計算研究会代表. 経済産業省産業技術審議会委員等を歴任. 知的オフィス環境コンソーシアム会長.



佐々木康成

2000年筑波大学大学院博士課程心理学研究科退学，2003年博士（学術），同志社大学工学部ポスドクター，同特別研究員（PD）を経て，2005年同知識工学科（現インテリジェント情報工学科）任期付講師，現在，同大学理工学部研究員．場所情報の作業記憶と海馬系に関する生理心理学的研究，霊長類の音コミュニケーションと齧歯類の聴覚に関する行動科学的研究に従事，音声知覚，潜在記憶，注意機能，学習と記憶，言語等の認知科学的研究，さらに情報分野との共同研究も幅広く展開．日本生理心理学会，日本音響学会，日本認知科学会，日本認知心理学会，日本光脳機能イメージング研究会等各会員．



大向 一輝（正会員）

2000年同志社大学工学部卒業．2002年同大学院工学研究科博士前期課程修了．2005年総合研究大学院大学複合科学研究科博士後期課程修了．博士（情報学）．2005年国立情報学研究所助手．2007年同助教．2009年同准教授，現在に至る．セマンティックウェブ，ソーシャルウェブの研究ならびに学術情報サービスの開発に従事．人工知能学会，電子情報通信学会各会員．



武田 英明（正会員）

1991年東京大学大学院工学系研究科博士課程修了．1993年4月奈良先端科学技術大学院大学助手，1995年4月同助教授．2000年4月国立情報学研究所助教授，2003年5月同教授，現在に至る．同研究所学術コンテンツサービス研究開発センター長（兼任），東京大学人工物工学研究センター教授（兼務）．人工知能，特に知識共有，オントロジ，ネットワークコミュニティ等の研究に従事．AAAI，Design Society，電子情報通信学会，人工知能学会等各会員．