

説得性に基づく情報推薦手法の提案

- 「ぷらっと Plat @ 自由が丘」における統合された行動ログの活用-

A Proposal of Persuasibility-based Recommender

- On Relationships among Users based on Integrated Behavior Logs at "Puratto-Plat @ Jiyugaoka" -

小柴 等[†] 相原 健郎[†] 森 純一郎[‡] 武田 英明^{†‡}
 Hitoshi KOSHIBA Kenro AIHARA Junichiro MORI Hideaki TAKEDA
 小田 朋宏[§] 星 孝哲[¶] 松原 伸人[§]
 Tomohiro ODA Takanori HOSHI Nobuto MATSUBARA

1 はじめに

ICT(Information and Communication Technology)の発展・普及に伴って、情報爆発という問題が顕在化してきた[1]。この情報爆発と呼ばれる問題への対策の一つとして、ユーザの望む、ユーザに適した情報を提供するための情報推薦技術が研究・開発されている。しかしながら、情報爆発の勢いはとどまることを知らず、すでに推薦結果であるユーザに適した情報までが氾濫しつつある。

そこで、従来の推薦手法に加えて、推薦結果である情報がユーザに受け入れられる可能性(情報の受容度)まで考慮した情報推薦手法を提案したので報告する。

2 背景と目的

情報推薦手法として一般的なものに協調フィルタリングが存在する。協調フィルタリングは、“行動や属性が似たものは、同じようなものを好む”といった発想に基づく手法である。協調フィルタリングの具体例とは、例えば以下のようなものである。“年代や趣味、嗜好が非常に近いことがわかっているユーザAとBを想定する。つぎに、このユーザAがリングとミカンを購入しようとした際、ユーザAと類似したユーザであるBが過去に、リングとミカンを購入したときに同時に購入したものがなにかを検索する。その際、ユーザBがリング、ミカンに加えてリキュールも同時に購入していたとすると、ユーザAにもリキュールの購入を推薦する。”上述したとおり、この協調フィルタリングは情報推薦手法として一般的であり、オンラインストアなどでも多く採用されているといわれている。

ところで、近年ICTの発展・普及に伴って、ブログなど人の手によって編集、加工されたコンテンツの他、様々な場所やモノ、ヒトに取り付けられた各種センサの情報など、様々な情報がネット上にあふれてきた。これにより情報検索などの分野では情報爆発といった問題が懸念されており[1]、近い将来、類似するユーザおよび、それらユーザとの差分情報が大量に押し寄せてくることが示唆されている。このことは協調フィルタリングだけにとどまらず、情報推薦全般についていえる問題であり、今後予想される情報爆発時代、すなわち人の手で編集されたコンテンツをはじめセンサ情報など様々な情報が氾濫する時代においては、雪崩や津波のように押し寄せてくる大量の推薦情報の中から、さらに情報を絞り込んでゆく手法が必要になると予測される。

ところで、現実世界(Webを離れた実社会)において人間はWeb以上の多種多様な大量の情報に囲まれているにもかかわらず、それらの環境にある程度適応し、それなりに状況に即した意思決定を行うことができている。したがって、情報爆発時の情報推薦手法を考える上でも、実社会上で我々が情報の取捨選択に使用している手法を考慮することは有用であると考えられる。

実社会において情報が推薦され、判断を行う場面としては例えば“説得”に関連する場面が挙げられる。説得とは“おもに言語的手段によって、他者の態度や行動をある特定の方向へ変化させようとする”ことであり[2]、これに関して、社会心理学の態度変容研究分野で研究が進められている。この研究の成果として、例えば、情報が得られても、受け手に情報そのものを評価する能力がない、情報が曖昧、処理に必要な時間がない、重要度が低く判断にコストをかけたくない、といった場合には、情報そのものではなく、情報に付随する周辺的な手がかりによって意思決定を行うこと[3]や、同種、同質の情報であっても、情報の送り手の持つ特徴やメッセージの送信手段、提示順序、受け手の持つ特徴、などの要因によって、その情報の受け入れられやすさ(受容度)が異なることなどが報告されている[4, 5, 6, 7]。

これら、説得に関わる要素を現在のネット上にある情報をもとに提供することは容易ではないが、様々な情報が氾濫する情報爆発の状態を考慮した場合には、それら大量のデータをもとに、実世界において我々が意思決定を行う際に使用している多様な情報を、ネット上でも同様に提供することが可能になると思われる。

これらの背景を鑑みて、情報推薦を行う際にも情報そのものだけでなく、情報源と情報の受け手の関係性を考慮することが重要であり、また、実世界の様々な情報がネット上に反映される情報爆発時代にあつては、計算機上でそれらの関係性を抽出することが可能であると考え、ユーザ間の説得性を算出するモデル、および、当該モデルによって情報推薦を行う仕組みを開発した。

3 説得性に基づく情報推薦手法の提案

3.1 説得性の定義

一般に説得とは“おもに言語的手段によって、他者の態度や行動をある特定の方向へ変化させようとする”を指す[2]。ここでは上記の定義をふまえた上で、説得性を以下の通り定義する。

定義 人物Aが人物Bに対して説得を行おうという意図を有したとき、同じ意図を有する他者と同じ情報を同じように提供したとして、Aの説得が他者と比較してどの程

[†] 国立情報学研究所, National Institute of Informatics (NII)

[‡] 東京大学, The University of Tokyo

[§] 株式会社 SRA 先端技術研究所, SRA Key Technology Laboratory, Inc.

[¶] 株式会社 SRA, Software Research Associates, Inc.

度 B に影響を及ぼす可能性があるか、相対的に表したものの

したがって、受け手にとって説得性が高い人物の情報ほど受容される可能性が高くなる。また、説得性は基本的に2者間の関係であるので、ある人物について説得性が高い人物が、他の人物についても同じように説得性が高いとは必ずしも言えない。

3.2 説得性の構成要素

本節では説得性を構成する要素について考察する。

説得に関わる要因としては大きく、1. 送り手の属性、2. メッセージ内容、伝達手段、提示順序、3. 受け手の属性、といった要因が指摘されている [5, 7]。

したがって、説得性のモデルを構築するにあたっては、上記のすべての要因を考慮することが望ましい。しかしながら、現実的にはこれらの要因を一度にすべて考慮したモデルを構築することは困難である。そこで、本提案では説得性の構成要素として「1. 送り手の属性」を対象を絞ってモデルを構築することとした。

これは「送り手の属性」がその他の要因に比べて取り扱いきれやすいと考えられることによる。

例えば、メッセージについての操作をなんらかのシステムを用いて自動的に行うためには、システムがメッセージの内容を適切に理解することが不可欠である。しかし、現状ではそのようにメッセージの内容を適切に理解するシステムを構築することが困難であり、さらに、その次の段階である提示順序や提示手段を自動的に操作、演出することも困難である。

受け手の属性に関しては、例えばこれまでに「受け手の心理的状態(気分)」「受け手の知能」といったものが説得の受容に影響を与えることが指摘されている。しかし、受け手の気分については外部からそれを計測することが困難であり、現状すぐに使用可能な推定技術も見あたらない。受け手の知能については、これを測定し、活用することはその計測に関わるコストに加えて、倫理的な面での問題も考えられる。

このように、メッセージや受け手に関わる説得要因は、これを活用するにあたっての技術的なハードルが高い。これに対して「送り手の属性」に関しては後述の通り、外部から観測可能な要因が多く、説得性モデル構築の初期段階として、この送り手の属性に絞ることは適当であると考えられる。

3.3 送り手の属性

説得に関わる送り手の属性としては、大きく信憑性、魅力性の2要素と、それらを構成する下位要素がそれぞれ指摘されている [2, 5, 6, 7]。下位要素を含む各要素の定義/意味はそれぞれ以下の通りである。

信憑性 (Credibility) 信用性と専門性からなる概念 [4]

専門性 (Expertise) 情報源が真実を知りうる立場であり、正確な主張が可能な専門家と認知される程度 [2]
コミュニケーション内容、テーマに関連 [5]

信用 (信頼) 性 (Trustworthiness) 偏らずに誠実な主張をするだろうという情報源の意図への信頼の程度 [2] 送り手のパーソナリティと密接に関連 [5]

魅力性 (Interpersonal Attraction) 人が他者に対して抱く好意や嫌悪のこと [2]

好意 (Likability) 好ましいと思う感情の程度

類似性 (Similarity) 対象との属性の類似の程度

相補性 (Complementarity) 自分にないものや欠けているものを持っている程度

外見的魅力 (Source Attraction) 顔などを初めとする外見的特徴の望ましさの程度

接触回数 実際に目にしたり、ふれたりした程度

これらの属性のうち、外見的魅力に関しては文化的背景や、個人の好みといった問題、倫理的問題があるため今回は採用しない。また、接触回数についても把握が困難であったり、取得、表示することで個人情報やプライバシーを著しく侵害する可能性があるため今回は採用しない。また、類似性と相補性は関連性が強いので、1つの要素として取り扱う。

したがって、本提案における説得性モデルは信憑性2要素、魅力性2要素の計4要素によって算出するものとし、心理学分野の態度に関する Fishbein らのモデル [8] 等を参考に、説得性をごく単純な線形の多属性補償型モデルとして以下のように定義する。

$$\text{説得性} = \text{信憑性} + \text{魅力性} \quad (1)$$

$$\text{信憑性} = \text{専門性} + \text{信用性} \quad (2)$$

$$\text{魅力性} = \text{類似性} + \text{好意} \quad (3)$$

ただし、ヒューリスティクスを活用するような場面では、必ずしも補償型のモデルが最適であるといえない可能性がある。そこで運用に際しては、説得性そのものを示す指標に加えて、説得性を構築する信憑性、魅力性の下位要素(専門性、信用性、類似性、好意)の各値についても、その高低を把握できる形で提供する。その上で、運用結果を参考に今後モデルを改良してゆくことを考える。

なお、モデルを構成する4要素は必ずしも独立した要素ではなく、互いに強く関連しあう部分もあるため、厳密には別個のものとして取り扱うことができないが、ここでは説明の都合上、独立のものとして扱っている。

ところで、上記のモデルに基づいて考えた場合、推薦手法として一般的な手法である協調フィルタリングは、魅力性の中の類似性と相補性に基づいて、対象者にとって魅力的な候補を推薦する手法であるといえる。このことから、説得性モデルは協調フィルタリングの手法を内包する、より上位の推薦機構と位置づけることができる。

4 評価実験

提案した説得性モデルの妥当性、有効性、実用性などを検証するために、後述の通り実験を実施した。

評価実験は、我々の提案する説得性(情報の受容度)に基づく情報推薦手法のみを評価することを目的として設計されたが、本評価実験は提案手法を包含する、さらに大きなプロジェクトの実証実験の一環として、当該実証実験と一体で実施された。

説得性を算出するための情報の取得や被験者の用いた環境など、当該実証実験と提案手法の評価実験は密接に関連しており、切り離して説明することが困難な部分が多いため、以下ではまず実証実験の内容について説明を行う。

なお、以下で「実証実験」とは提案手法の評価実験を含むプロジェクトの実験全体、「評価実験」とは我々の提案する説得性(情報の受容度)に基づく情報推薦手法の実験、を指すものとする。

4.1 実証実験概要

実証実験について、説得性の算出方法を説明するために必要な範囲で簡単にまとめる。

実証実験では、交通系の非接触 IC カードを用いて実世界

(物理世界)とネット世界の行動ログをつなぎ、来街者の行動からお店の盛り上がりなど街の姿を可視化する取り組みを行った。

具体的な内容は以下の通りである。

まず、実証実験の対象地域である自由が丘(東京都目黒区)の商店街において、実験へのご協力を頂いた52の店舗に交通系非接触ICカードであるPASMOの読み取り装置(PASMOリーダ)を設置した。協力52店舗の業種はレストラン、喫茶店といった飲食店をはじめ、書店や美容室、フィットネスクラブなど、多業種で構成される。これらの協力店舗へ設置したPASMOリーダには、焦電センサとマイクも組み込まれており、PASMOリーダ前の大まかな人の通過量および、周囲の音圧(騒音レベル)を取得することができる。

ユーザは交通系非接触ICカードであるPASMOを携帯し、これらの店に訪店した際にPASMOリーダへPASMOをかざし(PASMOタッチ)、行動の記録(ライフログ)を残す。この際、PASMOタッチに連動して登録ユーザのみがログイン可能な専用ブログサイトにおいて、自動的にその店へ訪れた旨を記載したブログ記事が生成される。

ユーザは帰宅後にこの専用ブログサイトを訪れ、上記、自動生成された自身の記事を編集・公開したり、他者の投稿した記事や各店舗の記事を閲覧する。今回はユーザが手動で生成するブログ記事も協力52店舗のどれかに関するもののみしか投稿できない仕様としたため、基本的にはすべての記事が52店舗のどこかのお店に関連する記事となる。

これらPASMOタッチをはじめとする実世界での行動情報や、ブログサイトでの記事編集・閲覧情報などネット世界での行動情報を用いて街の盛り上がりを算出し、街頭のサイネージやブログサイト上で提供した。

街頭のサイネージは自由が丘駅前のショッピングセンターの一角、一カ所に設置され、サイネージ横の操作端末を操作することで、街の盛り上がり情報を様々な観点(年代、性別、天気など)で閲覧できたほか、気になる店舗の情報や各店舗の発行する割引クーポンなどを印刷して取得できた。被験者の参加登録もこのサイネージを利用して実施された。

また、ブログサイト上では1.自身のライフログ(PASMOタッチのログなど)を要約して表示する機能、2.自動生成された自身の記事に対して、上記の行動情報からその記事が生成されたタイミングでのユーザの内面的な情報(気分、目的、能動・受動)を推定し、提示する機能(推定結果が正しくない場合に、ユーザがそれを修正して申請する機能を含む)、3.他者の記事について、その他者から自己への説得性を算出して表示する機能、4.専用ブログサイト内に開設された協力各店舗のページを閲覧した場合に、説得性の高い順に関連記事を表示する機能、などを提供した。

これらの実証実験は基本的にオープンな形で実施され、PASMOを所有している、登録時に満20歳以上であるといった一定の条件を満たせば、誰もが参加できる形で実施された。

実証実験の実施期間は2009年01月17日~02月08日*。参加協力店舗52店。延べ参加者数は683名であった。(評価実験参加者の年齢、性別、職業構成などは後述する。)

以上が、実証実験の概要である。

表1 説得性の要素と行動ログの対応

要素	対応する行動ログ
専門性	任意の店についての記事数、訪問回数
信用性	対象記事の被参照回数、評価
好意	対象者から自身の記事の参照回数、参照後行動回数
類似・相補性	プロフィール(基本属性)、行動の類似性

4.1.1 行動ログ詳細

以下では実証実験において取得した行動ログについて簡単にまとめる。

実証実験では被験者として、1.独自のアプリケーションを搭載した携帯電話を携帯することで詳細な行動ログを収集し、対価として謝礼を支払う被験者、2.自由が丘駅の乗降データの自動取得とアンケートへの回答など一定の条件付きで謝金を支払う被験者、3.オープン参加の被験者、など複数の区分が存在する。

説得性の算出と表示は被験者の別なく提供することを想定しているため、以下ではすべての被験者に共通して取得可能な行動ログのみを記載する。

実証実験では、実世界の行動ログとして

1. 実証実験協力店舗に置かれたPASMOリーダへのPASMOタッチ情報
2. 実証実験協力店舗に置かれたPASMOリーダ前の人通過量と、周囲の音圧データ
3. 自由が丘駅前のショッピングセンター1カ所におかれたサイネージ端末へのPASMOタッチ情報
4. 自由が丘駅前のショッピングセンター1カ所におかれたサイネージ端末での情報閲覧・印刷履歴

ネット上での行動ログとして

1. 専用ブログサイトでの記事投稿・編集履歴
2. 専用ブログサイトでの他者記事閲覧履歴
3. 専用ブログサイトでの他者記事評価履歴

といったものが利用可能であった。

さらにこれらを組み合わせることで、

1. 特定の店に関する記事閲覧した後に、実際に訪店した履歴
2. 特定の店に訪店した後に、そのお店に関する記事閲覧した履歴

といったものも取得、利用が可能であった。

なお、これらの設定より、実証実験および評価実験で取り扱う行動ログデータは、概ねライフログデータ、すなわち店舗やサイネージ端末でのPASMOタッチの数に比例することになる。

4.2 実証実験・評価実験における説得性の算出方法

以上の事柄を考慮して、説得性の各下位要素と行動ログとを表1に示したとおり大まかに対応付け、評価実験における各下位要素の具体的な算出式を構築した。

それぞれの策定理由は以下の通りである。

まず、今回の実証実験ではブログ記事はすべて特定の「店」に対応づけられる。

これを念頭に、「専門性」という要素を考えると、その店に関する専門家とは、店員を除けば、一般にその店の常連客す

* 01月25日までは実証実験サービスのためのデータ収集期間のため、実サービス提供期間は01月26日から。また専用ブログサイトについては02月17日まで運用した。

なわちその店舗へ足繁く通うユーザである。したがって、この観点から専門性の尺度を構成することとした。なお、特定の店の常連客であって、かつ同業他社に関して広くアクセスしているような人物は、その業種一般に関する権威としてさらに重みを付けるといったことも考えられるが、今回は反映しなかった。

「信用性」の要素については、その記事内容や行動が他者から評価されている人物は信用されていると考えられる。したがって、これらの観点から尺度を構成することとした。

「好意」の要素については、測定が困難な部分があるため、今回は「好意の返報(互惠)性」に着目して、これを好意一般ととらえることとした。「好意の返報性」とは、簡単にいえば自分を好きになってくれる人は、自分も好きになるという性質である[9]。今回の状況に照らし合わせて考えると、自分の記事に反応してくれる人、さらに、その記事を参考に実際に行動(記事のお店へ訪店)してくれた人など、自分の行動になんらかの賛同をしてくれる人(フォロワー)に対しては、好意を抱きやすいと考えられる。したがって、これらの観点から尺度を構成することとした。

「類似・相補性」の要素については、従来多用される、性別、年代、趣味嗜好といった属性の類似性に加えて、どのようなお店に実際に訪れているかという実世界での行動情報を加味することにした。一般に実世界での行動はネット上での行動に加えてコストがかかるため、実世界での行動はネット上の行動よりも、個人の趣味嗜好をより強く、正確に反映できると考える。したがって、これらの観点から尺度を構成することとした。なお、ブログの閲覧履歴等の類似性に関しては今回は反映しなかった。

これらの観点に基づいて作成した各下位要素および説得性の算出式については、付録として添付したとおりであるが、基本的に、各要素は各行動ログをそれぞれ正規化した上で単純に加算する形になっており、説得性はそれらの平均値となっている。

また、以上より本評価実験において使用する説得性は、被説得側の任意のユーザA、説得側のユーザB(ユーザA以外の任意のユーザ)、任意の店舗、という3つの変数を持つ関数となる。

4.3 実証実験・評価実験における説得性の提示方法

本論文において提案する説得性のモデルは、1. 取得した情報を評価する際、情報が曖昧であったり処理に時間がかげられないような状況下では特に、情報そのものだけでなく情報に付随する様々な周辺の手がかりに基づいて、つまりヒューリスティクスによって意思決定を行うことが多い、2. 周辺の手がかりとしては例えば情報の送り手の属性などが存在する、などの社会心理学的な知見をベースに構築されたものである。

したがって、算出された説得性はユーザに不可視の部分で用いられるのではなく、ユーザにとって目に見える形で提供される必要がある。具体的には、ただ単に「あなたへのオススメ記事ランキング」といったものが出てくるのではなく、「なぜ、それがオススメなのか。その記事や記事の投稿者は記事の内容について、また、あなたにとってどのような特徴を持っているのか。」といったことが明示されていなければならない。

そこで、説得性の各要素を「オススメ度グラフ」として可視化し、提示した。また、提示に際しては、一般の被験者にも理解しやすいことや、「説得」という用語を用いた場合、“

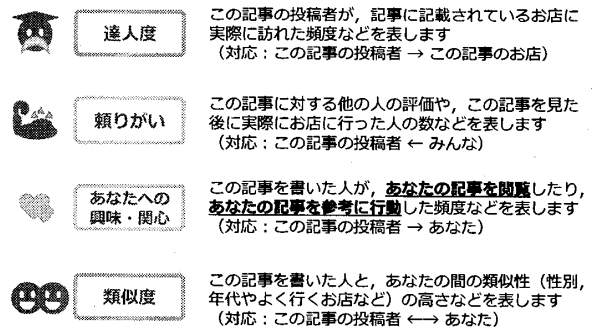


図1 説得性の下位要素に関するユーザへの説明

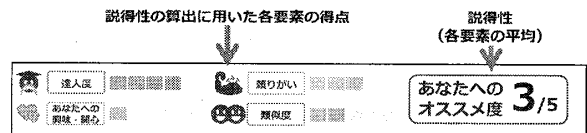


図2 オススメ度グラフの表示例

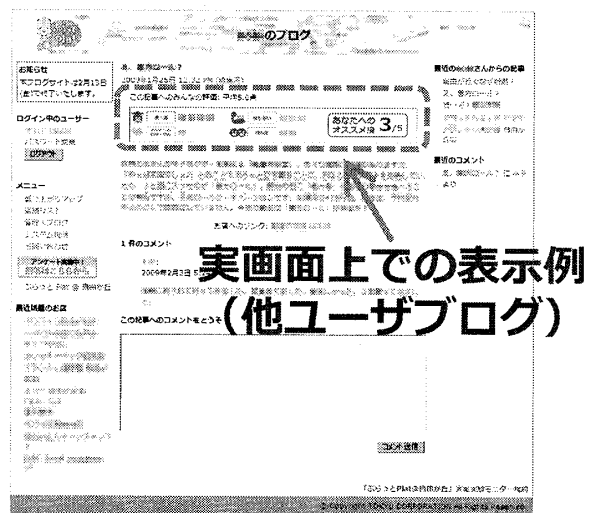


図3 ブログ記事への表示例

自身の態度を恣意的に操作される”と誤解される恐れがあること等を考慮して、「説得性」は「あなたへのオススメ度」、「専門性」は「達人度」、「信用性」は「頼りがい」、「好意」は「あなたへの興味・関心」「類似・相補性」は「類似度」、として説明し、それぞれ1から5の5段階評価の数値やグラフで提示を行った。

この説得性の下位要素に関するユーザへの説明やオススメ度グラフの表示例を図1から、図3に示す。

上述した通り説得性は、あらゆる箇所でもユーザにとって目に見える形で提供される必要があるが、画面デザインの制約上、実証実験サービスである専用ブログの各店舗ページにおけるユーザへの推薦画面では、説得性の高い順に10件の記事タイトルを表示するのみで、説得性及び説得性の各要素を提示することができなかった。

4.4 仮説

評価実験を実施するにあたって想定した仮説は以下の通りである。

仮説1 説得性に基づいて抽出し、オススメ度グラフとともに記事を表示した場合、協調フィルタリング(類似性)

に基づく従来手法で抽出した記事を提示した場合と比較して、ユーザが「自分にとって適切である」、「参考になる」、「このお店へ行ってみよう」と感じる割合が高い(受容度が高い)。

仮説2 一方で、説得性に基づいて記事を抽出した場合でも、オススメ度グラフを表示しないケースでは、協調フィルタリング(類似性)に基づき抽出した記事を提示したケースと比較して、ユーザの主観に有意な違いは見られない。

また、これらの仮説に加えて、説得性モデルの妥当性や行動ログデータ量の観点からの有効性・実用性の検証も行う。

4.5 実験デザイン

4.3節に述べたとおり、実証実験期間中、算出した説得性に基づく記事の抽出および、他者記事への「オススメ度グラフ」(説得性および、説得性の各要素の得点を可視化したもの)の表示を行った。提案する説得性に基づく情報推薦手法の有効性について、本来ならばこれら実証実験期間中に被験者が自由に使用した結果を用いて評価することが望ましい。しかしながら今回の実証実験では、オープン参加の被験者も参加する環境であること、被験者には本提案システム以外にも複数のサービスの体験や、アンケートへの回答を依頼することから、被験者に対してあまり制約をかけられないことなど、様々な要因から統制が困難であり、実証実験中の自由試行結果を基に評価を行うことができなかった。そのため、実証実験期間はコンテンツおよび行動ログデータの収集期間と位置づけ、実施実験期間の最後に行われる事後アンケートにあわせて、説得性のモデルを評価するための実験(評価実験)を別に配置した。

評価実験のデザインとしては被験者内、被験者間混合デザインを採用した。群(被験者群)の配置および要因、水準は以下の通りである。

まず、群は2群を用意した。

群Aは仮説1のうち、「説得性の高い記事をオススメ度グラフと共に提示した場合の影響」を調査するための群である。今回はあわせて、説得性の各要素の高低が、記事評価に及ぼす影響も調査することとした。群Aに適用する水準としては、説得性の要因を信憑性と魅力性に集約し、この2要因についての高低を組み合わせて4水準とした。本来は、今回モデルに採用した下位要素4種類すべてについて、高低を組み合わせ、8水準を設けることが望ましかったが、被験者にかかる負担から上記の4水準にまとめた。

ここで信憑性が高い・低いとは、信憑性を構成する専門性と信用性がともに高い・低い値を示していることを意味する。同様に魅力性が高い・低いとは、魅力性を構成する類似・相補性と好意がともに高い・低い値を示していることを意味する。

群Bは仮説1のうち、「説得性の高い記事をオススメ度グラフなしに提示した場合の影響」の調査と、仮説2を検証するための群である。群Bに適用する水準としては、説得性、類似性それぞれの高低から4水準とした。

ここでいう類似性は、説得性の算出に用いる下位要素の類似性そのものを意味するが、前述したとおり、提案手法における類似性の項目は一般的な推薦手法であり、「属性の似たユーザは似たようなものを好む」という主張に基づいて推薦を行う協調フィルタリングに相当するものであることから、提案手法を評価する上での従来手法として適当であると考へ、これを使用した。

これらの群の設定目的や特徴について、表2にまとめた。

表2 群の設定目的と特徴

	設定目的	オススメ度 グラフ	水準数
群A	仮説1の検証データ(一部)取得(説得性を構成する要因の高低が及ぼす影響も調査)	あり	4
群B	仮説1の検証データ(一部)取得と、仮説2の検証データ取得	なし	4

図4 アンケート画面

なお、A,B両群の各水準における高低の具体的な基準・抽出手法については後述する。また、各群への被験者の割り付け手続きなどは実験プロトコルの項において詳述する。

4.6 評価指標

評価には、アンケートによる被験者評価の値を用いる。

このアンケートの取得方法であるが、各水準に基づいて抽出された記事を提示、閲覧させ、続いて、それらの記事に対するアンケートを行うという形で実施する。手続きの詳細は実験プロトコルの項で述べる。

評価用アンケートの項目および、回答項目については

- この記事の内容に興味は持てたか
- この記事を読んで訪店意欲に変化はあったか
- この記事は参考になりそうか
- この記事の投稿者の他の記事も読んでみたいと思うか
- この記事は推薦記事として適切であったか

という5項目を設定し、各項目について従来手法との比較によって提案手法の評価を行うこととした。

図4に、実際に被験者に提示した画面の一例を示す[†]。

4.7 実験プロトコル

以下では実験のプロトコルについて述べる。

実験は大きく3つに切り分けられる。詳細は後述し、以下では概要を説明する。

まず第1段階として初期態度を取得する。初期態度とは、実験開始前の被験者の態度を意味する。本実験では被験者に“店”に関連する記事の閲覧と評価を行わせるため、ここでは被験者の店に対する興味関心を問い、その結果をもとに、実

[†] 被験者は個人のPCを用いて回答を行うため、それぞれの環境によって、フォントの書体や大きさ、一画面内に表示される領域は異なる。この画面は横幅が800pxに収まるようにデザインされている。

験で用いる対象店舗の抽出を行う。

つぎに第2段階として本実験を実施する。ここでは各群の各水準に応じて抽出された記事の閲覧とアンケートへの回答を水準の数(群A,B共に4水準)だけ繰り返して実施する。

最後に第3段階として簡単な総合アンケートを実施する。

4.7.1 第1段階：エントリーから初期態度の取得まで

実験は実験者の管理するサーバ上のサイトへ、被験者がPCを用いネットを介してアクセスすることで行う。

被験者は専用ブログへログインした状態であれば自動的に、そうでない場合は専用ブログと同様のログインID等を入力することで、評価実験用システムにエントリーする。

この際、被験者はシステムによって自動的に各群へと振り分けられる。振り分けは群A、群Bそれぞれへの割り付け比が1:2となるように工夫されたブロックランダム表を用いて、ある程度ランダムに振り分けられる。これは、群Bに関する検出力を向上させるための措置である[†]。

つぎにシステムは第2段階の本実験において記事を抽出するために必要な、店舗の候補を抽出する。このとき、店舗に対する思い入れや予備知識に大きな偏りが生じないように、店舗の候補は、被験者が未訪問の店の中からランダムに抽出する。仮に被験者が未訪問の店が規定の数に満たない場合は、訪問回数の少ない店から順に規定の店舗数までを抜き出す仕様である。その後、被験者にはランダムに選択された6店舗の名称を提示する。

被験者はシステムから提示された6店舗について、各店舗への興味の度合いに応じて1から6まで順位を付ける。これによって被験者の初期態度を取得する。

被験者から初期態度を回収すると、システムは被験者から「1番目に興味がある」「6番目に興味がある」と回答を得た店舗を実験の候補から除外する。これは、説得のおよび方に関するモデルである精緻化見込みモデル[3]などにおいて、「対象者が強い興味関心を持つ場合は情報を論理的に精査して意思決定に及ぶ」といった指摘がなされており、もともと強い興味関心をもつ店舗の記事を用いて実験を行うことは、ヒューリスティクスの活用を想定する説得性のモデルを検証する上で適切ではなさそうなこと、被験者が論理的に思考しない場合でも、他店に比べて正負を問わず強い思い入れがある場合、評価に強いバイアスがかかると想定されること、などを考慮し、これらの要因を排除するために行った措置である。

このようにして初期態度を取得した6店舗の内、残った4店舗を本実験の候補とする。

4.7.2 第2段階：本実験

群A,Bともに4水準を有するため、本実験は両群とも4回行われる。各回は記事の提示・閲覧および閲覧度の自己申請フェーズと、アンケートフェーズという2段階で構成される。各回の提示・閲覧フェーズにおける水準、表示店舗(興味の度合い)の組み合わせについてはランダム化を行った。また提示・閲覧フェーズの滞留時間はマイクロ秒単位で記録されており、事後に何秒間提示していたかを表示することが可能である。

被験者が記事の閲覧に使う画面を図5に示す。なお、図5は群Aの被験者に示されるタイプのものであり、群Bの被験者の場合は全水準について画面中央のオススメ度グラフが

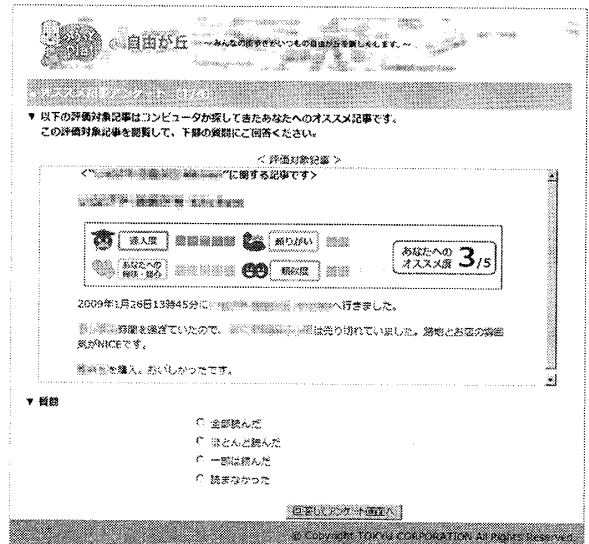


図5 記事の提示・閲覧画面

表3 各水準の記事抽出基準・手法

群	水準	抽出基準
群 A ₁	憑:高 魅:高	4要素の総計が最大
群 A ₂	憑:高 魅:低	信憑性と魅力性の差が最大
群 A ₃	憑:低 魅:高	魅力性と信憑性の差が最大
群 A ₄	憑:低 魅:低	4要素の総計が最小
群 B ₁	類似性:高	類似性とその他3要素の平均の差が最大
群 B ₂	類似性:低	類似性が最低
群 B ₃	説得性:高	4要素の総計が最大
群 B ₄	説得性:低	4要素の総計が最小

※憑:信憑性, 魅:魅力性

表示されない。アンケート回答(評価入力)画面は図4に示したとおりである。

なお、これらの記事は計算結果に基づいて自動的に抽出されるため、被験者毎に提示される記事内容は異なる。また、本評価実験においては提案手法、従来手法ともに記事内容(情報の量や質など)は一切考慮していない。

群、および水準によって提示される記事の抽出基準・手法については、表3に示したとおりである。なお、同値のものが存在する場合には表3の条件に加えて、信憑性、魅力性を構成する各要素間の分散が最小のものを選択するといった条件もそれぞれに付与してある。

4.7.3 第3段階：第2段階終了後から評価実験終了まで

本実験終了後は、追加のいくつかのアンケートに回答をさせた後、評価実験がすべて終了した旨の画面を表示して、被験者をログアウトさせる。これらの内容については、本稿の内容に直接関連しないので割愛する。

以上の実験プロトコルについて概略を図6に示した。

4.8 実施条件

本評価実験は実証実験の事後アンケートの一環として実施する。

実証実験の被験者には、1. 独自のアプリケーションを搭載した携帯電話を携帯することで詳細な行動ログを収集し、対価として謝礼を支払う被験者、2. 自由が丘駅の乗降データの自動取得とアンケートへの回答など一定の条件付きで謝金を支払う被験者、3. オープン参加の被験者、など複数の区分が

[†] 群Bは主に“(各手法がもたらす効果について、実用上大きな)差がない”という仮説を検証する事を念頭に用意された。そのため、検出力を高めに設定したものである。

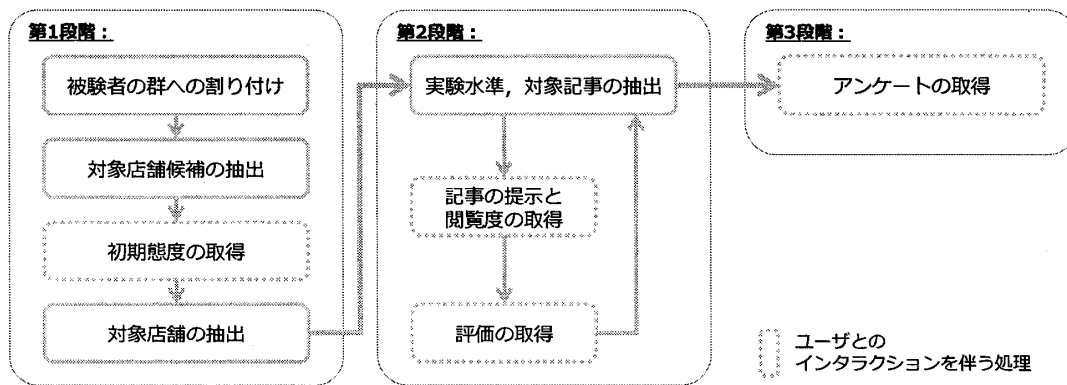


図6 実験プロトコルの概略

存在するが、これらすべての被験者が評価実験の対象である。なお、謝金の支払いをとまなう被験者には、本評価実験への参加も謝金支払い条件の一部であることを説明した。

これらの事後アンケート及び評価実験のシステムはそれぞれ特定のサーバ上に構築されており、被験者はPC及びネットを介してこれらのサーバへアクセスして実験を行った。

上述の通り、本評価実験は実証実験の事後アンケートの一環として実施され、被験者には実証実験終了後の2009年02月06日～13日までの1週間の間で、任意の時間に1回実施するよう求めた。

4.9 結果

実験から得られた結果を以下に記述する。

4.9.1 評価対象被験者の絞り込み

実証実験の全参加者683名のうち、評価実験に参加した被験者の総数は184名であった。実験の結果をとりまとめるにあたってはこれらの被験者全体をそのままとりまとめるのではなく、ライフログ(実世界でのPASMOタッチ)の数を基準として3つの層に分割し、実験結果をとりまとめる。

具体的な被験者層の分割理由、手法は以下の通りである。

まず、ライフログ数を基準に採用した理由であるが、そもそも説得性は実世界とネット上、両世界の行動ログが十分に蓄積された世界で、かつ、類似の情報が氾濫した状況を想定して開発されたものである。したがって、本手法が有効に作用するためには行動ログと情報の蓄積が不可欠である。しかしながら、実証実験ではデータの蓄積に用いる期間が2週間程度に限定されたため、全被験者に対して十分にデータを蓄積することができなかった。そこで、想定する状況に近い被験者層のみを対象とすることで、我々の想定する状況下での手法の有効性を精度良く検証した。ここで、本実証実験ではPASMOタッチによってブログ記事が生成されるという特性上、基本的に行動ログデータ量はライフログに比例する。また、実世界における行動は、ネット上での行動に比べて基本的にコストが高いため、行動ログの中でも実世界での行動ログの方が相対的に高い価値を有すると考えられる。以上より、ライフログ数を基準として採用した。

つぎに、分割手法であるが、ライフログ数によって被験者をソートし、四分位点を基準に分割を行った。その上で75パーセンタイル点(第3四分位点)以上の被験者のみを仮説検証の分析対象に採用することとした。これは、上述したとおり説得性を算出するにあたって十分な行動ログが収集されていることを想定していることによる。

また、25パーセンタイル点(第1四分位点)以上75パーセ

表4 被験者属性(全体:性別)

性別	人数	割合
男性	29	16.5%
女性	147	83.5%
合計	176	100.0%

ンタイトル点未満の被験者層についても、仮説検証の分析対象と同様の分析を行い、これと比較することで、行動ログデータ量が十分に集まらず理想的でない状況下でのモデルの有効性、実用性を検証することとした。

そもそもライフログが無い、ほとんど無い場合には、算出された説得性が意味を持たないため、ライフログ数が25パーセンタイル点未満の者は完全に排除した。

4.9.2 欠損値の除去

アンケートサイトでは、アンケートの回答に不備があった場合には再入力を促したため、アンケートの回答に関する欠損は生じなかった。

一方、被験者に表示した記事について、今回使用した各手法ではコンテンツの中身を考慮していないため、コンテンツの投稿者が後からコンテンツを削除するなどした場合に、中身を持たない、空のコンテンツが表示されることがあった。

そこで、4回の本実験のうち、1回でも空のコンテンツが表示された被験者については、これを欠損値としてリストワイズで被験者の全データを削除することとした。

結果、184名の評価実験被験者のうち、8名の被験者のデータが欠損値として排除され、解析対象の被験者は176名となった。

4.9.3 被験者の属性

欠損値を除いた176名の被験者の属性を表4から表9に示す。各群及びライフログ数で被験者を分割した際にも、各グループ内の比率は表4から表9に示した比率におおむね当たっていた。

被験者全体の特徴としては、実証実験地域である“自由が丘”という地域の特性を反映して、年齢層では20代から30代、性別では女性の割合が特に多くなっている。また、実証実験がブログなどネットを介したサービスを提供するものであるためか、被験者の多くがもともと日常的(週1回以上)にブログもしくはSNSといったサービスを利用している事がわかる。

4.9.4 提案手法(グラフあり)と既存手法の比較結果

評価対象であるライフログ数が75パーセンタイル点以上の被験者のアンケート結果等について、表10および表11に

表5 被験者属性(全体:年代)

年代	人数	割合
20代	39	22.2%
30代	94	53.4%
40代	34	19.3%
50代以上	9	5.1%
合計	176	100.0%

表6 被験者属性(全体:婚姻)

既婚	人数	割合
既婚	95	54.0%
未婚	81	46.0%
合計	176	100.0%

表7 被験者属性(全体:職業)

職業	人数	割合
会社員・公務員・団体職員	84	47.7%
自営業・会社経営	4	2.3%
派遣社員・契約社員	20	11.4%
パート・アルバイト	11	6.3%
自由業・フリーランス	6	3.4%
主婦・家事従事	35	19.9%
学生	14	8.0%
無職	0	0.0%
その他	2	1.1%
合計	176	100.0%

表8 被験者属性(全体:謝礼)

謝礼	人数	割合
あり	146	83.0%
なし	30	17.0%
合計	176	100.0%

表9 被験者属性(全体:ブログ・SNSの使用頻度)

ブログ, SNS 使用頻度	人数	割合
ほぼ毎日(週6~7日)	93	52.8%
週に4~5日くらい	23	13.1%
週に2~3日くらい	21	11.9%
週に1日くらい	18	10.2%
月に2日くらい	7	4.0%
月に1日くらい	5	2.8%
月に1日未満	3	1.7%
ブログやSNSは閲覧しない	6	3.4%
合計	176	100.0%

示した。

これらの表において、群 A₁ は「信憑性:高, 魅力性:高」の水準、群 B₁ は「類似性:高」の水準の結果のみを抜粋したものである。有意確率は上記2水準の代表値の差を Wilcoxon の順位和検定によって算出したものである[§]。

4.9.5 提案手法(グラフなし)と既存手法の比較結果

評価対象であるライフログ数が75パーセンタイル点以上の被験者のアンケート結果等について、表12および表13を示した。

これらの表において、群 B₃ は「説得性:高」の水準、群

[§] 表10の File Size および閲覧時間についてはウエルチの方法による t 検定。

表10 コンテンツサイズ及び閲覧時間、閲覧度の比較(提案手法(グラフあり) vs. 既存手法)

	File Size (byte)		閲覧時間 (sec)		閲覧度 [*]	
	群 A ₁	群 B ₁	群 A ₁	群 B ₁	群 A ₁	群 B ₁
平均値	565.8	500.1	25.2	18.0	2.9	3.2
標準偏差	243.8	280.8	26.4	11.4	1.2	1.2
最頻値	—	—	—	—	4	4
有意確率	0.403		0.275		0.309	
標本数	19	27	19	27	19	27

※4段階評価 1:読まなかった-4:全部読んだ

表12 コンテンツサイズ及び閲覧時間、閲覧度の比較(提案手法(グラフなし) vs. 既存手法)

	File Size (byte)		閲覧時間 (sec)		閲覧度 [*]	
	群 B ₃	群 B ₁	群 B ₃	群 B ₁	群 B ₃	群 B ₁
平均値	538.7	500.1	19.4	18.0	3.2	3.2
標準偏差	360.6	280.8	13.6	11.4	1.2	1.2
最頻値	—	—	—	—	4	4
有意確率	0.681		0.674		1.000	
標本数	27	27	27	27	27	27

※4段階評価 1:読まなかった-4:全部読んだ

表14 コンテンツサイズ及び閲覧時間、閲覧度の比較(提案手法(グラフあり) vs. 既存手法, ライフログ数少)

	File Size (byte)		閲覧時間 (sec)		閲覧度 [*]	
	群 A ₁	群 B ₁	群 A ₁	群 B ₁	群 A ₁	群 B ₁
平均値	639.0	636.2	26.4	32.6	3.1	3.1
標準偏差	368.4	378.1	20.0	113.8	1.0	1.3
最頻値	—	—	—	—	4	4
有意確率	0.975		0.678		0.446	
標本数	25	63	25	63	25	63

※4段階評価 1:読まなかった-4:全部読んだ

B₁ は「類似性:高」の水準の結果のみを抜粋したものである。有意確率は上記2水準の代表値の差を Wilcoxon の符号順位和検定によって算出したものである[¶]。

4.9.6 ライフログ数が少ない場合の比較結果

評価対象としたライフログ数が75パーセンタイル点以上の被験者と、ライフログ数が25パーセンタイル点以上75パーセンタイル点未満と、ライフログ数(行動ログデータ量)が少ない被験者のアンケート結果等について、表14および表15に示した。

これらの表において、群 A₁ は「信憑性:高, 魅力性:高」の水準、群 B₁ は「類似性:高」の水準の結果のみを抜粋したものである。有意確率は上記2水準の代表値の差を Wilcoxon の順位和検定によって算出したものである[¶]。

4.9.7 提案手法の構成要素に関する交互作用

評価対象としたライフログ数が75パーセンタイル点以上の被験者のうち、群 A の被験者の各水準のデータを用いて、説得性の構成要素である信憑性、魅力性についての高低の組み合わせがアンケート結果に及ぼす影響を、対応のある二元配置分散分析によって処理した。結果を表16に示す。

[¶] 表12の File Size および閲覧時間については対応のある t 検定。

[¶] 表14の File Size および閲覧時間についてはウエルチの方法による t 検定。

表 11 本実験のアンケート集計結果 (提案手法 (グラフあり) vs. 既存手法)

	Q1		Q2		Q3		Q4		Q5	
	群 A ₁	群 B ₁	群 A ₁	群 B ₁	群 A ₁	群 B ₁	群 A ₁	群 B ₁	群 A ₁	群 B ₁
平均値	3.9	3.2	3.9	3.3	3.9	3.1	3.8	2.9	3.9	3.4
標準偏差	1.1	1.4	0.7	1.0	0.7	1.4	1.0	1.4	1.1	1.4
最頻値	4	4	4	3	4	4	4	4	4	4
高評価占有率*	78.9%	48.1%	73.7%	33.3%	84.2%	51.9%	78.9%	44.4%	78.9%	59.3%
有意確率		0.052 ⁺		0.018 [*]		0.045 [*]		0.018 [*]		0.210
標本数	19	27	19	27	19	27	19	27	19	27

※ 5 を高評価, 3 を「どちらでもない」とする 5 段階評価で 4 もしくは 5 の評価の占める割合

Q1~Q5 は 5 を高評価, 3 を「どちらでもない」とする 5 段階評価

Q1:この記事へ興味を持てたか

Q4:この記事の投稿者の他の記事を読みたいと思うか

Q2:記事閲覧前に比べ店に行きたくなくなったか

Q5:この記事は推薦記事として適切であったと思うか

Q3:この記事は参考になると思うか

+ : p < 0.10, * : p < 0.05, ** : p < 0.01

表 13 本実験のアンケート集計結果 (提案手法 (グラフなし) vs. 既存手法)

	Q1		Q2		Q3		Q4		Q5	
	群 B ₃	群 B ₁	群 B ₃	群 B ₁	群 B ₃	群 B ₁	群 B ₃	群 B ₁	群 B ₃	群 B ₁
平均値	3.5	3.2	3.7	3.3	3.6	3.1	3.3	2.9	3.5	3.4
標準偏差	1.5	1.4	0.9	1.0	1.3	1.4	1.3	1.4	1.4	1.4
最頻値	5	4	3	3	4	4	4	4	4	4
高評価占有率*	59.3%	48.1%	55.6%	33.3%	70.4%	51.9%	55.6%	44.4%	66.7%	59.3%
有意確率		0.390		0.080 ⁺		0.094 ⁺		0.237		0.644
標本数	27	27	27	27	27	27	27	27	27	27

※ 5 を高評価, 3 を「どちらでもない」とする 5 段階評価で 4 もしくは 5 の評価の占める割合

Q1~Q5 は 5 を高評価, 3 を「どちらでもない」とする 5 段階評価

Q1:この記事へ興味を持てたか

Q4:この記事の投稿者の他の記事を読みたいと思うか

Q2:記事閲覧前に比べ店に行きたくなくなったか

Q5:この記事は推薦記事として適切であったと思うか

Q3:この記事は参考になると思うか

+ : p < 0.10

表 15 本実験のアンケート集計結果 (提案手法 (グラフあり) vs. 既存手法, ライフログ数少)

	Q1		Q2		Q3		Q4		Q5	
	群 A ₁	群 B ₁	群 A ₁	群 B ₁	群 A ₁	群 B ₁	群 A ₁	群 B ₁	群 A ₁	群 B ₁
平均値	3.6	3.5	3.5	3.6	3.6	3.7	3.4	3.2	3.2	3.5
標準偏差	0.9	1.3	0.8	1.0	1.0	1.1	1.1	1.3	1.1	1.3
最頻値	4	4	3	3	4	4	4	4	2	4
高評価占有率*	56.0%	58.7%	48.0%	52.4%	60.0%	69.8%	48.0%	49.2%	40.0%	55.6%
有意確率		0.817		0.422		0.623		0.645		0.287
標本数	25	63	25	63	25	63	25	63	25	63

※ 5 を高評価, 3 を「どちらでもない」とする 5 段階評価で 4 もしくは 5 の評価の占める割合

Q1~Q5 は 5 を高評価, 3 を「どちらでもない」とする 5 段階評価

Q1:この記事へ興味を持てたか

Q4:この記事の投稿者の他の記事を読みたいと思うか

Q2:記事閲覧前に比べ店に行きたくなくなったか

Q5:この記事は推薦記事として適切であったと思うか

Q3:この記事は参考になると思うか

表 16 交互作用分析結果

	平均値 (標準偏差)						F 値 (有意確率)						標本数
	信憑性: 高			信憑性: 低			信憑性		魅力性		交互作用		
	魅力性: 高	魅力性: 低		魅力性: 高	魅力性: 低								
Q1	3.9 (1.1)	3.3 (1.2)		3.7 (1.2)	3.7 (1.4)		0.085 (0.774)	1.000 (0.331)	6.642 (0.019)*			19	
Q2	3.9 (0.7)	3.5 (0.8)		3.9 (0.8)	3.7 (1.1)		0.063 (0.804)	3.071 (0.097) ⁺	0.487 (0.494)			19	
Q3	3.9 (0.7)	3.4 (1.1)		3.9 (1.0)	3.8 (1.2)		0.607 (0.446)	1.018 (0.326)	2.497 (0.132)			19	
Q4	3.8 (1.0)	3.3 (1.3)		3.6 (1.3)	3.7 (1.3)		0.096 (0.760)	0.762 (0.394)	4.434 (0.050) ⁺			19	
Q5	3.9 (1.1)	3.1 (1.3)		3.5 (1.3)	3.6 (1.3)		0.077 (0.784)	1.270 (0.275)	9.852 (0.006)**			19	

Q1~Q5 は 5 を高評価, 3 を「どちらでもない」とする 5 段階評価

Q3:この記事は参考になると思うか

Q1:この記事へ興味を持てたか

Q4:この記事の投稿者の他の記事を読みたいと思うか

Q2:記事閲覧前に比べ店に行きたくなくなったか

Q5:この記事は推薦記事として適切であったと思うか

+ : p < 0.10, * : p < 0.05, ** : p < 0.01

5 考察

5.1 仮説の検証

4.4節に挙げた仮説について、評価実験の結果を基に考察する。これについては当初設定したとおり、ライフログ数が75パーセントポイント以上の被験者のデータに限って分析する。

5.1.1 仮説1に関する結果の考察

仮説1は提案手法を適用した条件である群 A_1 と、既存手法を適用した条件である群 B_1 の2つを比較することで検証できる。この比較は被験者間で行われる。

結果は表10および表11の通りである。

これらの結果からQ1からQ5の各項目で、信憑性、魅力性が共に高い、すなわち説得性が高いユーザの記事を抽出し、さらにオススメ度グラフを組み込んで表示する提案手法(群 A_1)と、類似性のみが高いユーザの記事を抽出し、そのまま表示する従来手法(群 B_1)では、すべての項目で提案手法に基づいて記事を抽出、提示した群Aの被験者の方が、好意的な回答を寄せる割合が多かった。また、複数の項目についてその割合の差は統計的に有意、有意な傾向を有するものであることが確認された。

なお記事の持つ情報量が単純に記事のサイズに比例すると考えた場合、それらの偏りが結果に影響を及ぼすことが考えられるが、表10に示したとおり、記事サイズの間に統計的な有意差は認められなかった。また、記事のサイズがほぼ同じ場合、記事の閲覧時間や被験者が記事を閲覧したと感じた度合い**と、記事に対する興味関心が比例することなどが考えられるが、これらについても統計的に有意な差は認められなかった。

したがって、表11に示された割合の差は、コンテンツのサイズや興味関心を暗に示唆する、閲覧時間、閲覧したと感じた度合いの差に起因するものではなく、実験条件に起因するものである可能性が高い。

これにより、我々の仮説が支持された。

5.1.2 仮説2に関する結果の考察

仮説2は説得性が高いユーザの記事を抽出するが、オススメ度グラフは組み込まずにそのまま表示する条件である群 B_3 と類似性のみが高いユーザの記事を抽出し、そのまま表示する条件である群 B_1 を比較することで検証できる。この比較は被験者内で行われる。

結果は表12および表13の通りである。

これらの結果から、Q1からQ5の各項目で説得性の高い記事を提示した場合の方が、類似性の高い記事を提示したときよりも、選好される割合が多い傾向が見える。また、複数の項目についてその割合の差は統計的に有意な傾向を有するものであることが確認された。

一方で、記事サイズや閲覧時間、閲覧したと感じた度合いには有意な差が確認されなかった。従って、5.1.1節と同様に、表13に示された割合の差は、実験条件に起因するものである可能性が高い。

これにより、我々の仮説は支持されなかった。ただし、これは説得性の明示をともなわない場合にも提案手法が有効に作用することを示すものであり、提案手法の有効性についてはこれを支持するものである。

5.2 行動ログデータ量が及ぼす影響

前節までは分析の対象として行動ログデータが十分に蓄積された被験者のデータのみを分析対象として採用し考察を行ったが、データが十分ではない場合、すなわちデータ収集システムの立ち上がり期間など不安定な状態において、提案手法がどの程度の性能を確保できるのかという点も、実用性を考える上で重要な課題である。

そこで、4.4節に述べたとおり、分析の対象から外したライフログ数が25パーセントポイント以上75パーセントポイント未満の被験者層についても、仮説検証の分析対象と同様の分析を行い、実用性を考察する。

結果は表14および表15の通りである。これらの結果からは、行動ログデータ量が少ない場合、Q1からQ5について提案手法よりも従来手法の方が選好される割合が高いことが確認される。ただし、その差は数%から10%程度で、統計的な有意差も確認できなかった。

従来手法の方がやや選好された理由としては、以下を考察することができる。

説得性を算出するには前述の通り、ある程度のライフログ(行動ログデータ)の蓄積が必要である。このライフログが十分ではない場合はデータのS/N比が低い状態となり、そのデータをもとに算出した結果も、当然S/N比が低いものとなる。

一方、提案手法の構成要素であり従来手法としても使用する類似性については、他の説得性算出基準とは違い、「ユーザの(静的)属性」も使用している。「ユーザの(静的)属性」とは具体的には、年代や性別、興味のあるジャンルなどである。これらのデータは実験開始前に被験者に入力されたものであり、ライフログとは無関係に類似性を算出することができる。したがって、この部分に限定すればライフログの影響を受けないため、S/N比が相対的に高く、ライフログがない状態でも比較的妥当な推薦が行えたものと考えられる。

ただし、従来手法と提案手法の差は統計的に有意なものではなく、かつ、提案手法は従来手法である協調フィルタリングの考え方を内包するものであるため、データ量が少ない場合でも、従来手法程度の能力を確保できていると考えられる。また、データ量が少ない場合には類似性のみを使用するといった使い方も可能である。

以上より、提案手法は行動ログデータ量が少ない状態であっても、従来手法と同程度の性能(実用性)を保持しているといえる。

5.3 モデルの妥当性

最後に、今回モデルに採用した信憑性、魅力性という大きく2つの要素について、その交互作用からモデルの構成要素としての妥当性を考察する。

これは同一の被験者に、信憑性、魅力性の高低4つの組み合わせ、それぞれに基づいて抽出した記事を閲覧し、評価してもらった群Aの結果について分析することで検証できる。

結果は表16に示したとおりである。この結果からは、Q2「記事閲覧前に比べ店に行きたくなったか」という問いに対しては、魅力性の単純主効果が認められ、信憑性の単純主効果や交互作用は認められない。一方で、Q1、Q4、Q5といった項目では単純主効果は認められない一方で、交互作用が認められる。また、Q1からQ5について説得性が高い(信憑性、魅力性がともに高い)ものが最も高く評価を得ている傾向も確認できる。

以上の結果より、複数の項目で単純主効果が認められず交

** 4.7.2節における「提示・閲覧および閲覧度の自己申告フェーズ」で被験者が入力した記事の閲覧度

相互作用が認められていることから、今回、説得性の構成要素として信憑性と魅力性を採用したことは、ある程度妥当なものであったと考えられる。

モデル全体としても、上記の複数の考察で提案手法の有効性が支持されており、妥当なものであったと考えられる。

5.4 関連研究

説得性を構成する要素である信憑性と魅力性について、既存の推薦研究で対応するものとしては、信憑性に関して「信頼度」(trust)、魅力性に関して「親しさ」(familiarity、もしくは social relation) を考慮したものが挙げられる。

「信頼度」については推薦の受け入れやすさではなく、cold start (使い始めの時にデータが乏しく効果が出ない) といった推薦の問題や推薦の精度向上のために用いられている。例えば、Massaらはユーザ間の評価を元にユーザ間の信頼度を伝搬させ、ユーザ間類似度に信頼度を加味した推薦手法を提案している [10]。同様に Golbeckらはユーザ間の評価を元に計算された信頼度を用いた映画推薦システムを提案している [11]。ユーザ間の明示的な評価に対して、O'Donovanらはユーザのプロファイルに基づく過去の推薦精度を元に信頼度を非明示的に計算し協調フィルタリングに利用する推薦手法を提案している [12]。この他、信頼度は推薦システムにおける悪意あるユーザを防ぐために利用されることもあるが、いずれも信頼度は推薦結果の説得性(受容性)を考慮したのではなく、推薦精度の向上を目的として利用されている。

「親しさ」については情報推薦の分野でも自身に近い、親しい相手からの推薦は受け入れられやすいことが示されている [13, 14]。またプロファイルの類似性(profile similarity)が高い相手からの推薦についても受け入れられやすいことが示されている [15]。以上のように、魅力性は親しさや社会的つながり(social tie)という形で研究がなされているが、相手に対する魅力や好意を直接に用いた推薦手法ではない。

その他、オンライン情報を用いた推薦研究は多く存在するが、実世界のユーザ行動も考慮した研究はあまり行われていない。

以上より、実世界の様々な情報がネット上に反映される情報爆発時代を想定し、実世界とネット上の行動ログを活用して推薦を行う提案手法は先進的な取り組みであると言える。

6 おわりに

本論文では、行動ログデータを基にユーザ間の説得性と呼ばれる指数を算出するモデルと、この説得性を用いてユーザからの受容のされやすさを考慮した情報推薦手法を提案し、評価実験を行った結果について述べた。

評価実験の結果は、おおむね我々の仮説を好意的に支持するものであり、本論文において提案した“説得性を考慮した情報推薦手法”が、有効に作用することが確認できた。

ただし、今回は様々な制約条件から、十分なデータを得ることができず、説得性の推定精度や適用条件が適切であったとは言い難い面もある。今後は、さらに大規模かつ長期の実証実験を行い、モデルの妥当性、有効性を検証すると共に、モデルの精緻化を行ってゆく予定である。

謝辞

本研究は経済産業省 情報大航海プロジェクトにおける「地域活性化を支える e 空間サービス ーぷらっと Plat ー」(受託企業: 株式会社エス・ピー・シー) 実証事業の一環として、東京急行電鉄株式会社、日本電気株式会社、株式会社東急エー

ジェンシーの協力の下に行われた。また、実証実験は自由が丘商店街振興組合の協力の下に行われた。記して感謝する。

参考文献

- [1] 喜連川優, 松岡聡, 松山隆司, 須藤修, 安達淳: 情報爆発時代に向けた新しい IT 基盤技術の研究, 人工知能学会誌, Vol. 22, No. 2, pp. 209-214 (2007).
- [2] 中島義明, 子安増生, 繁榊算男, 箱田裕司, 安藤清志, 坂野雄二, 立花政夫 (編): 心理学辞典, 有斐閣 (1999).
- [3] Petty, R. E. and Cacioppo, J. T.: The Elaboration Likelihood Model of persuasion, *Advances in experimental social psychology*, Vol. 19, pp. 123-205 (1986).
- [4] Hovland, C. I., Janis, I. L. and Kelley, H. H.: *Communication and Persuasion: Psychological studies of opinion change*, Yale University Press, Connecticut (1956). 辻正三, 今井省吾 (訳): コミュニケーションと説得, 誠信書房, 1960.
- [5] 榊博文: 説得と影響—交渉のための社会心理学, プレイン出版 (2002).
- [6] 深田博己 (編): 説得心理学ハンドブッカー説得コミュニケーション研究の最前線, 北大路書房 (2002).
- [7] 今井芳昭: 依頼と説得の心理学—人は他者にどう影響を与えるか, セレクション社会心理学, Vol. 10, サイエンス社 (2006).
- [8] Fishbein, M.(ed.): *Readings in attitude theory and measurement*, John Wiley and Sons, Inc, New York (1967).
- [9] 奥田秀宇: 人をひきつける心—対人魅力の社会心理学, セレクション社会心理学, Vol. 17, サイエンス社 (1997).
- [10] Massa, P. and Avesani, P.: Trust-aware recommender systems, *RecSys '07: Proceedings of the 2007 ACM conference on Recommender systems*, New York, NY, USA, ACM, pp. 17-24 (2007).
- [11] Golbeck, J.: Filmtrust: movie recommendations from semantic web-based social networks, *The 3rd IEEE Consumer Communications and Networking Conference (CCNC 2006)*, Vol. 2, pp. 1314-1315 (2006).
- [12] O'Donovan, J. and Smyth, B.: Trust in recommender systems, *IUI '05: Proceedings of the 10th international conference on Intelligent user interfaces*, New York, NY, USA, ACM, pp. 167-174 (2005).
- [13] Bonhard, P. and Sasse, M. A.: Knowing me, knowing you' - Using profiles and social networking to improve recommender systems -, *BT Technology Journal*, Vol. 24, No. 3, pp. 84-98 (2006).
- [14] Woerndl, W. and Groh, G.: Utilizing Physical and Social Context to Improve Recommender Systems, *WI-IATW '07: Proceedings of the 2007 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology - Workshops*, Washington, DC, USA, IEEE Computer Society, pp. 123-128 (2007).
- [15] Bonhard, P., Harries, C., McCarthy, J. and Sasse, M. A.: Accounting for taste: using profile similarity to improve recommender systems, *CHI '06: Proceedings of the SIGCHI conference on Human Factors in computing systems*, New York, NY, USA, ACM, pp. 1057-1066 (2006).

Appendix

専門性

$$\begin{aligned} & \text{専門性 (評価対象者, 対象店)} \\ & = \text{パーセンタイル} \{ \text{正規化記事数 (評価対象者, 対象店)} \\ & \quad + \text{正規化訪問回数 (被験者, 対象店)} \} \end{aligned}$$

※パーセンタイルは所与の対象店に関する記事を書いた被験者 (評価対象者) を母集合とする

$$\begin{aligned} & \text{正規化記事数 (評価対象者, 対象店)} \\ & = \frac{\text{記事数 (評価対象者, 対象店)}}{\text{Max}_{\text{被験者} \in \text{全被験者}, \text{店} \in \text{全店舗}} \text{記事数 (被験者, 店)}} \end{aligned}$$

$$\begin{aligned} & \text{正規化訪問回数 (評価対象者, 対象店)} \\ & = \frac{\text{訪問日数 (評価対象者, 対象店)}}{\text{Max}_{\text{被験者} \in \text{全被験者}, \text{店} \in \text{全店舗}} \text{訪問日数 (被験者, 店)}} \end{aligned}$$

信用性

$$\begin{aligned} & \text{信用性 (評価対象者, 対象店, 評価者)} \\ & = \text{パーセンタイル} \{ \text{正規化被参照回数 (評価対象者)} \\ & \quad + \text{正規化評価 (評価対象者, 評価者)} \\ & \quad + \text{正規化参照後他者行動回数 (評価対象者, 評価者)} \} \end{aligned}$$

※パーセンタイルは所与の対象店に関する記事を書いた被験者 (評価対象者) を母集合とする

$$\begin{aligned} & \text{正規化被参照回数 (評価対象者)} \\ & = \frac{\text{被参照回数 (評価対象者)}}{\text{Max}_{\text{被験者} \in \text{全被験者}} \text{被参照回数 (被験者)}} \end{aligned}$$

$$\text{被参照回数 (評価対象者)} = \Sigma(\text{アクセス時刻} - \text{基点時刻})^{-1}$$

※基点時刻は説得性更新の時刻の7日前の同時刻, アクセス時刻も説得性更新の時刻の7日前の同時刻まで。

$$\begin{aligned} & \text{正規化評価 (評価対象者, 評価者)} \\ & = \frac{\text{全評価 (評価対象者)} - \text{評価者自身による評価 (評価対象者, 評価者)}}{\text{Max}_{\text{被験者} \in \text{全被験者}} \text{全評価 (評価対象者)}} \end{aligned}$$

$$\begin{aligned} & \text{全評価 (評価対象者)} \\ & = \Sigma_{\text{記事} \in \text{評価対象者が書いた記事}, \text{被験者} \in \text{全被験者}} \text{評価 (記事, 被験者)} \end{aligned}$$

$$\begin{aligned} & \text{対象者自身による評価 (評価対象者, 評価者)} \\ & = \Sigma_{\text{記事} \in \text{評価対象者の記事}} \text{評価 (記事, 評価者)} \end{aligned}$$

$$\begin{aligned} & \text{正規化参照後他者行動回数 (評価対象者)} \\ & = \frac{\text{参照後他者行動回数 (評価対象者)}}{\text{Max}_{\text{被験者} \in \text{全被験者}} \text{参照後他者行動回数 (被験者)}} \end{aligned}$$

好意

$$\begin{aligned} & \text{好意 (評価対象者, 対象店舗, 評価者)} \\ & = \text{パーセンタイル} \{ \text{正規化記事参照回数 (評価対象者, 評価者)} \\ & \quad + \text{正規化参照後対象者行動回数 (評価対象者, 評価者)} \} \end{aligned}$$

※パーセンタイルは所与の対象店に関する記事を書いた被験者 (評価対象者) を母集合とする

$$\begin{aligned} & \text{正規化記事参照回数 (評価対象者, 評価者)} \\ & = \frac{\text{記事参照回数 (評価対象者, 評価者)}}{\text{Max}_{\text{被験者} 1, \text{被験者} 2 \in \text{全被験者}} \text{記事参照回数 (被験者 1, 被験者 2)}} \end{aligned}$$

$$\begin{aligned} & \text{正規化参照後対象者行動回数 (評価対象者, 評価者)} \\ & = \frac{\text{参照後対象者行動回数 (評価対象者, 評価者)}}{\text{Max}_{\text{被験者} 1, \text{被験者} 2 \in \text{全被験者}} \text{参照後対象者行動回数 (被験者 1, 被験者 2)}} \end{aligned}$$

類似性

$$\begin{aligned} & \text{類似性 (評価対象者, 対象店, 評価者)} \\ & = \text{パーセンタイル} \{ \text{プロフィール類似度 (評価対象者, 評価者)} \\ & \quad + \text{行動類似度 (評価対象者, 評価者)} \} \end{aligned}$$

※パーセンタイルは所与の対象店に関する記事を書いた被験者 (評価対象者) を母集合とする

$$\begin{aligned} & \text{プロフィール類似度 (評価対象者, 評価者)} \\ & = \text{プロフィール一致項目数 (評価対象者, 評価者)} \end{aligned}$$

$$\begin{aligned} & \text{行動類似度 (評価対象者, 評価者)} \\ & = | \text{評価対象者がタッチ履歴を残した店舗} \\ & \quad \cap \text{評価者がタッチ履歴を残した店舗} | \end{aligned}$$

説得性

$$\text{説得性 (評価対象者, 対象店, 評価者)} = \frac{\Sigma \text{下位要素}}{\text{下位要素の個数}}$$