

人間関係ネットワークに基づく 情報推薦システムとその実装

堀 田 創^{†1} 萩 原 将 文^{†1}

本論文では、人間関係ネットワークに基づく情報推薦システムの提案とその実装について述べる。提案システムは、(1) ユーザプロフィール構築、(2) アイテム選択、(3) プリファレンス情報の更新の3つのアルゴリズムから構成されている。(1)ではプリファレンス情報と人間関係ネットワークに基づいてユーザプロフィールが構築され、(2)でユーザプロフィールに基づいたアイテム選択が行われる。プリファレンス情報はユーザへのアイテム提示およびクリックを基にした情報であり、(3)においてシステムへのアクセスログが参照され更新される。実装においてはアクセスログの処理および2段階のアイテム選択がバックエンドプロセスとして行われることで、オンラインでの計算量の軽減を図っている。また実装されたシステムは広告配信システムとして実際のWebサービス上で運用された。評価実験では、ランダムな広告配信と比較し1.9倍の推薦効果が得られた。

Design and Implementation of Social-network-based Recommender System

HAJIME HOTTA^{†1} and MASAFUMI HAGIWARA^{†1}

This paper describes the design and implementation of a recommender system using social networks. The proposed system consists of the following three algorithms; (1) Construction of user-profile; (2) Filtering items by profile data, (3) Updating preference data. In (1), user-profiles are constructed based on preference data and social networks. In (2), one item is selected by using constructed user-profile. Preference data are generated from behavior logs, updated in (3). The proposed system has been implemented as an advertisement delivery system. To reduce the calculation cost in one session, processes of log analysis and item selection are previously performed as backend processes. According to experimental results, the efficacy of the proposed system was 1.9 times higher than that of randomly delivering.

1. はじめに

近年のWeb技術の発展にともない、ユーザは様々なWebサービスを利用することが可能となってきた。中でもユーザ同士のコミュニケーションを目的としたサービスは数多く運営されており、近年では友人同士のリンクに基づいて日記等のサービスを利用することのできるSNS (Social Networking Service) が注目を浴びている¹⁾。SNSにはMySpace^{*1}やFacebook^{*2}等、世界的に有名なサービスがいくつも存在しており、多くのユーザがこれらを利用している²⁾。したがってコミュニケーションサービスの中では重要な形態の1つとなっている。

ここで、ユーザ同士の関連性を表す情報の分析は、社会学的見地からも注目されてきた³⁾。特に、ユーザおよび関係性をグラフ構造と見なした人間関係ネットワーク (Social Network) を工学的に取り扱う研究は数多く報告されている⁴⁾⁻⁸⁾。内田ら⁴⁾は複雑ネットワーク分析の見地からSNS内の人間関係ネットワークを分析している。このようなSNSの分析はそのほかにも多く行われているが^{5),6)}、これらは具体的な応用を前提に行われた分析とはなっていない。また大向ら⁷⁾は人間関係ネットワークに基づいたタスクスケジューラを提案しているが、これは特定用途での応用事例となっている。また松尾ら⁸⁾はWebマイニングの一環として、Web上から特定コミュニティ内の人間関係を自動的に抽出する方法を提案している。この手法は人間関係ネットワークが明示的に定義されていない場合でも獲得できることを示唆している。このように人間関係ネットワークの工学的応用はより重要な課題となると考えられる。

一方、近年ではアクセス可能な情報の量は膨大となっているため、効率的な情報獲得を目的とした情報推薦手法に関する研究も数多く行われている^{9),10)}。情報推薦手法はユーザの興味に合致した配信の役割を担うほか^{11),12)}、効果的な広告配信の手法としても注目されている¹³⁾。特に産業界では広告配信による収益確保は一般的となっており、重要なトピックとなっている¹⁴⁾。これら情報推薦の研究では、協調フィルタリングが多く利用されている^{15),16)}。協調フィルタリングは他のユーザのデータを参照することで、高精度にユーザプロフィールを構築するアルゴリズムである。しかしながら、計算時間やデータ量の観点から

^{†1} 慶應義塾大学理工学部

Faculty of Science and Technology, Keio University

*1 <http://myspace.com>

*2 <http://facebook.com>

直接的な応用は難しく、たとえば Amazon.com ではユーザではなくアイテムの特徴を用いたアルゴリズムにより運用している¹⁶⁾。また、協調フィルタリングのようなシステムではユーザ自身のログを必要とするため、新規ユーザに対する推薦には向かない。

そこで本論文では、人間関係ネットワークを利用することで、ユーザの興味に合致した情報推薦を行うシステムを提案し、その実装について述べる。提案システムはすでに運用されている SNS に組み込むことのできる汎用的システムとして実装されており、特に新規ユーザに対しても興味に応じた推薦を行えることが大きな特徴である。

本論文の構成を以下に示す。2 章では提案システムの根幹をなす情報推薦アルゴリズムについて述べる。3 章ではその実装方法について示す。提案システムの一部処理をバックエンドプロセスとして行うことで応答時間の高速化が可能となる。4 章では提案システムの評価について述べ、5 章でまとめる。

2. 人間関係ネットワークに基づく情報推薦システム

提案システムは人間関係ネットワークに基づいたユーザプロフィール構築を行い、それを用いることで推薦すべき情報を選択しユーザに提示する。新規ユーザのような場合でも、関連するユーザに関する情報を参照することによってユーザプロフィールを構築できる点が大きな特徴である。

2.1 概要

提案システムは、SNS 等の人間関係ネットワークを利用した Web サービスに組み込んで運用することが可能な設計となっている。その際、提案システムが必要とする情報はユーザに関する基本的な属性情報と人間関係ネットワークの情報に限定している。商品等の購買履歴やメールの送信履歴等も情報推薦において重要な役割を示す場合もあるが、それらに対応したシステムは、運用が過度に複雑になってしまうため、今回は扱わないこととした。

提案システムでは、推薦したアイテムおよびそれに対するユーザの反応をログとして保存する。ユーザの反応は、推薦されたアイテムに付加されるハイパーリンクのクリックにより判断される。ここでアイテムの推薦に関するログを viewlog、それに対するユーザクリックに関するログを clicklog と呼ぶこととする。これらは推薦アルゴリズム内で参照される。

提案システムの概要を図 1 に示す。提案システムは Interface 層、Logic 層、Database 層からなる。Database 層は、提案システムが用いる情報群、すなわち人間関係ネットワーク、アイテム、ユーザ情報の 3 つの情報を表す。

提案システムは Web サービスと連携されて運用される。Web サービスがユーザからのア

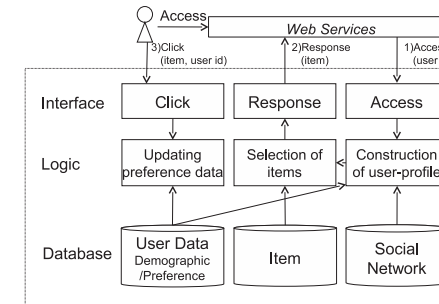


図 1 提案システムの概要

Fig. 1 An overview of the proposed system.

クセスを受けると、それにともない Web サービスが提案システムにアクセスする。その際はユーザを特定するための情報をシステムに送信する。すると提案システムは推薦するアイテムの情報を Web サービスへ出力し、その内容は Web ページ内に組み込まれてユーザに提示される。本論文の実験では、インターネット用のテキスト広告がアイテムとして扱われ、それらは携帯電話用の SNS サイト内に組み込まれた。ユーザがそのアイテムをクリックした場合、ユーザ情報およびアイテムの情報が提案システムに送られる。

人間関係ネットワーク (Social Network) は、ユーザ間の関連を示すネットワーク型のデータである。ここでユーザはノード、ユーザ間のつながりはエッジで表現された無向グラフとして表される。本論文では提案システムにアクセスしてきたユーザを対象ユーザと呼び、人間関係ネットワーク上で対象ユーザと直接にエッジで結ばれるユーザを関連ユーザと呼ぶこととする。

これらのデータは、提案システムを組み込む先のサービスと同期される。SNS のようなシステムであればデータベース内にこのようなデータが存在するため、それらを参照することで提案システムの利用が可能となる。また近年では Facebook^{*1}や Twitter^{*2}等いくつかの SNS では、関連ユーザの情報を取得するための API が公開されている。

アイテムはテキスト内容、URL および特徴ベクトルの 3 つの要素からなり、これらはすべて事前に提案システムへ与えられる。アイテムの特徴ベクトルは 2.2 節で述べるように定

*1 <http://facebook.com>

*2 <http://twitter.com>

義される．

ユーザ情報は Demographic Data と Preference Data の 2 つからなる．Demographic Data は性別・年代・地域等のユーザに関連する基本情報を示しており，あらかじめ提案システムに与えられる．Preference Data はユーザプロフィールを構築する際の根拠となる情報であり，システム内で学習が行われる．Preference Data の構築については 2.3 節で詳説する．

Interface 層はユーザとの直接的な接点を示しており，それぞれ以下のような役割を持つ．

- Access モジュール

アイテム推薦に関するリクエストを受信するモジュールである．リクエストの際には，E-mail アドレスや携帯電話における個体識別 ID，またはサービス内のユーザ ID 等，ユーザを特定するための情報が付加される．

- Response モジュール

アイテムを選択しユーザへ送信するモジュールである．アイテムは概要を示すテキストデータおよびハイパーリンクのための URL の 2 つの情報からなる．これらはユーザのクリック時に，該当アイテムに関する詳細ページへ転送される．リクエスト元のユーザ，そのアイテムと推薦された時刻のタイムスタンプは viewlog として保存され，Preference Data の更新に用いられる．

- Click モジュール

Response モジュールにより送信されたアイテムがユーザによりクリックされると，Click モジュールにより詳細ページへの転送が行われる．クリックしたユーザと推薦されたアイテム，推薦された時刻のタイムスタンプが clicklog として保存され，Preference Data の更新に用いられる．

Logic 層では，先に述べた Database 層のデータと Interface 層のモジュールおよび，以下の 3 つのモジュールによって，ユーザの興味に合致した情報推薦を可能としている．

- ユーザプロフィールの構築

対象ユーザと関連ユーザの Preference Data を用いることで，ユーザプロフィールの構築が行われる．ユーザプロフィールはユーザの興味を示す情報であり，後で示すようにベクトルで表現される．提案システムの大きな特徴である人間関係ネットワークに基づくアルゴリズムは，ここで用いられる．

- アイテム選択

アイテムの選択は確率的な選択により行われる．各アイテムの選択確率は，ユーザプロ

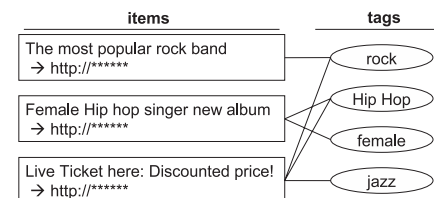


図 2 タグ形式によるアイテム定義の例

Fig. 2 Examples of the item definition with tag relations.

ファイルとアイテムの特徴ベクトルの内積により決定される．

- Preference Data の更新

アイテムがユーザに推薦された際の viewlog, clicklog を参照することで，後で示すような Preference Data の更新処理が行われる．

2.2 アイテムの定義

アイテムは提案システムの運用者によって与えられる情報である．本論文の実装ではインターネット広告配信システムとして提案システムを用いており，この場合は各広告がアイテムに該当する．

提案システムでは，運用者の特徴ベクトル定義を簡易化するためにタグ形式を採用している．図 2 はタグ形式のデータ例を示している．各アイテムに対してタグと呼ばれるキーワードを複数指定することにより特徴を表現する．タグ形式は Social Bookmark サービスの Delicious^{*1}や動画共有サービスの youtube^{*2}等，多くのサービスで利用されているフォーマットの 1 つである¹⁷⁾．タグ形式では，アイテムの内容を直接的に表す単語だけでなく，「女性向け」等，対象となる性別や年齢等も同様のフォーマットで表現可能である．

i 番目のアイテムの特徴ベクトル I^i の各要素はタグに対応し，以下のように決定される．

$$I^i = (I_1^i, I_2^i, \dots, I_D^i) \quad (1)$$

ここで D は特徴ベクトルの次元数を示し，タグの種類の総数と一致する．各要素はタグの種類に対応しており，以下の条件を満たすように正規化される．

*1 <http://delicious.com>

*2 <http://www.youtube.com>

$$\sum_{t=1}^D I_t^i = 1 \quad (2)$$

本論文で実装したシステムでは簡単のため、 I^i の t 番目のタグに関する要素 I_t^i を以下のように決定している。

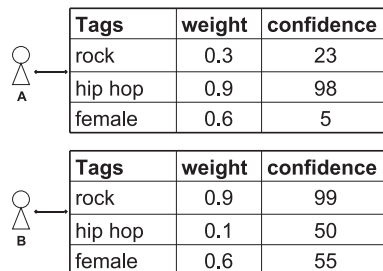
$$I_t^i = \begin{cases} 1/N_i & (i \text{ related}(t, i)) \\ 0 & (\text{otherwise}) \end{cases} \quad (3)$$

ここで $\text{related}(t, i)$ は t 番目のタグが i 番目のアイテムに対して関連づけられていることを表し、 N_i はアイテム i に関連するタグの個数を表している。

2.3 Preference Data の更新

Preference Data はユーザプロファイル構築の際の根拠となる情報であり、ユーザごとに定義される情報である。Preference Data の例を図 3 に示す。Preference Data は weight および confidence の 2 つの要素からなり、それらがタグごとに設定される。これらの情報は全ユーザに設定され、後に示す処理により更新される。weight はタグに対するユーザ興味の強さを表す要素であり、confidence は weight に対する信頼性を示す要素である。ここで weight は、アイテムの提示全体に対する反応の割合によって定めることができ、提案システムではこの決定にアイテムの推薦に対するクリックの割合 (CTR: Click Through Rate) を用いる。また confidence はアイテムの提示回数によって決定される。

具体的には、ユーザ u の t 番目のタグに対する weight $w_{u,t}$ 、confidence $c_{u,t}$ はそれぞれ以下のように計算される。



Tags	weight	confidence
rock	0.3	23
hip hop	0.9	98
female	0.6	5

Tags	weight	confidence
rock	0.9	99
hip hop	0.1	50
female	0.6	55

図 3 Preference Data の例

Fig. 3 Examples of Preference Data.

$$w_{u,t} = \left(\frac{N_{u,t}^{click}}{N_{u,t}^{view}} \right)^\alpha \quad (4)$$

$$c_{u,t} = (N_{u,t}^{view})^\beta \quad (5)$$

ここで α, β は定数であり、 $N_{u,t}^{click}$ はユーザ u が t 番目のタグに関連するアイテムをクリックした回数、 $N_{u,t}^{view}$ はアイテムの提示回数を示している。式 (4) は CTR を α 乗したものである。本論文で実装されたシステムでは、予備実験より $\alpha = 2, \beta = 0.5$ としている。

上記の仕組みの場合、weight は提示回数に応じて収束に向かう。しかしながら実際はユーザの興味は時間に応じて変化する場合も少なくない。そこで $N_{u,t}^{click}, N_{u,t}^{view}$ を以下のような時間減衰を加味した拡張を行うことで、ユーザの時間的な興味の変化に対応させる。

$$N_{u,t}^{click}(T_{now}) = \sum_{\ell \in C_{u,t}} \exp \left(-\frac{T_{now} - T_\ell}{T_{int}} \right) \quad (6)$$

$$N_{u,t}^{view}(T_{now}) = \sum_{\ell \in V_{u,t}} \exp \left(-\frac{T_{now} - T_\ell}{T_{int}} \right) \quad (7)$$

ここで $C_{u,t}, V_{u,t}$ はそれぞれ t 番目のタグに対応したアイテムに関する clicklog, viewlog の集合を示し、 T_{int} は時間減衰の間隔を示す定数を示す。 T_{int} の設定により、ログの時間的な影響を調整することが可能となる。また T_{now} は Preference Data が更新される段階でのタイムスタンプ、 T_ℓ はログ ℓ のタイムスタンプを表している。

2.4 人間関係ネットワークに基づくユーザプロファイルの構築方法

提案システムの大きな特徴の 1 つは、ユーザプロファイルを構築する際に人間関係ネットワークに基づく処理を行っている点である。ここではネットワークの分析によって関連ユーザとの類似度を計算し、それをユーザプロファイルの各要素の計算に用いている。

ユーザプロファイルとは、対象ユーザの興味をベクトルで表現した情報を示す。これはアイテム選択で用いられる。ユーザプロファイルは Preference Data と人間関係ネットワークを参照することで構築される。

人間関係ネットワークとは図 4 に示すような、関係するユーザ間にリンクが張られたネットワークである。代表的な SNS である MySpace や Facebook では友人同士が承認を経てリンクを構築する仕組みとなっている。

人間関係ネットワーク分析を含む社会ネットワークの研究において、ノード間の類似性を

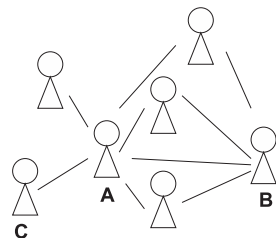


図 4 人間関係ネットワークの例

Fig. 4 An example of social network structure.

表す考え方に連結性 (cohesion) と構造同値 (structural equivalence) が多く用いられる。連結性¹⁸⁾とは多くユーザ間でコミュニケーションをとっている等の関連の強さを示す概念である。また構造同値性¹⁹⁾とは、ノードどうしを入れ替えてもネットワークの構造が変化せず、ネットワーク的に同じ役割を果たしているような関係を示す。図 4 の例では、ユーザ A から見てユーザ B、ユーザ C はともに直接連結されている。しかし他のノードとの関連性を考えると、共通して連結されるノードはユーザ B の方がより多い。このような場合に、ユーザ C と比較してユーザ B の方がユーザ A との構造同値性が高いとされる。一般に連結性よりも構造同値性のほうが類似度指標として有効であることが報告されている¹⁹⁾。

提案システムにおけるユーザ A とユーザ B の間の類似度 $\text{rel}(A, B)$ を、以下のようなシンブソン係数によって定義する。

$$\text{rel}(A, B) = \frac{n(R^A \cap R^B)}{\min(n(R^A), n(R^B))} \quad (8)$$

ここで R^A , R^B はそれぞれ、ユーザ A に関連するユーザ、ユーザ B に関連するユーザの集合を表しており、 $n(\cdot)$ は要素数を表す。また、 $n(R^A \cap R^B)$ は A, B の双方に関連するユーザの数を表している。

シンブソン係数は構造同値性を表現している指標の 1 つであり、類似度の評価においても多く用いられている²⁰⁾。ネットワーク内でのノード間の構造同値性に注目した類似度指標としては他にも提案されている²¹⁾が、計算コストの面で大規模なネットワークの場合には実用的でない。

このように求められた類似度に基づき、ユーザ u のユーザプロフィール P^u が計算される。これはアイテムと同じ空間内で表現されるベクトルとして表現され、各要素はタグに対する興味の強度を表す。

ユーザプロフィールの計算は、対象となるユーザの weight と、その関連するユーザの weight に関する加重平均によって決定される。その加重は式 (8) で求められる関連するユーザの類似度、および confidence に基づいたものである。提示回数が非常に少ない状態での weight の値は、1 回のクリックだけで大きく左右されることもあり概して不安定であるため、提示回数に基づいた confidence を加重の根拠に用いることで安定化を図っている。

ユーザ u に関連するユーザの集合を R^u とすると、ユーザ u の t 番目のタグに対するユーザプロフィール P^u の要素 P_t^u は以下のように計算される。

$$P_t^u = \frac{\gamma w_{u,t} c_{u,t} + \sum_{v \in R^u} \text{rel}(u, v) w_{v,t} c_{v,t}}{\gamma c_{u,t} + \sum_{v \in R^u} \text{rel}(u, v) c_{v,t}} \quad (9)$$

ここで γ は定数、 $w_{u,t}$, $c_{u,t}$ はそれぞれ式 (4), (5) の定義式によるものである。関連するユーザの weight は、対象ユーザのユーザプロフィールに必要なログが不足している場合に特に有効になる。対象ユーザのログが十分である場合は関連するユーザよりも対象ユーザ自身の weight が優先的に用いられるべきである。そのため γ は関連ユーザ数に対して大きな値であることが望ましい。なお、実験で用いたシステムでは $\gamma = 10.0$ と設定されている。本式により各要素を計算することで、ユーザプロフィールを構築している。

2.5 アイテムの選択

提案システムでは、ユーザ情報等に基づき確率的にアイテムが選択される。ユーザプロフィール構築からアイテム選択までの流れを図 5 に示す。

まずユーザのリクエストが来ると、ユーザの Preference Data および関連ユーザの Preference Data を参照し、ユーザプロフィールを構築する。その後、以下の 2 通りの方法でフィルタリングが行われる。

1) Demographic Filtering

ここでは Demographic Data に基づいて、選択されるアイテムの候補が絞り込まれる。たとえば男性ユーザに対して化粧品の推薦を行わない、等、的外れな推薦を防ぐことを目的としている。提案システムでは、Demographic Data とタグの組合せのうち禁止されるべき組合せをルールとしてあらかじめ定義しておくことでこの処理を行う。

2) Preference Filtering

構築されたユーザプロフィールと対応するアイテムの特徴ベクトルに基づき、確率的

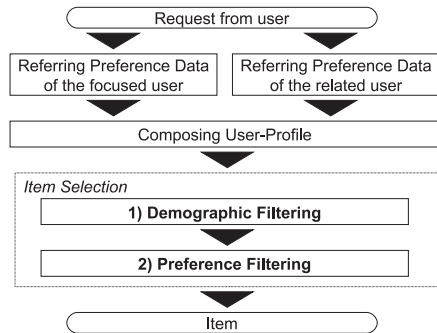


図 5 ユーザプロフィール構築からアイテム選択までの流れ
Fig. 5 A flow of the item selection.

にアイテムを選択する。

Preference Filtering では、アイテムの選択確率は該当するアイテムとユーザプロフィールの類似性によって判断される。提案システムでは類似度指標に特徴ベクトルと構築されたユーザプロフィールの内積を用いており、

これに比例した確率でアイテムが選択される。すなわち、ユーザ u に対し総アイテム数 M 個から j 番目のアイテムが選択される確率 $Prob_u(j)$ は以下の式で決定される。

$$Prob_u(j) = \frac{P^u \cdot I^j}{\sum_{i \in Item} P^u \cdot I^i} \quad (10)$$

ここで $Item$ はアイテム全体の集合、 P^u はユーザ u のユーザプロフィール、 I^j は j 番目のアイテムの特徴ベクトルである。

Demographic Filtering は推薦精度に直接的に寄与するフィルタリングではないが、推薦効果以外の理由で Demographic フィルタリングを行う場合に必要となる。たとえばウェブサイトのイメージを保つため、システム利用者自身が女性用化粧品の推薦を男性に行いたくない場合等があげられる。

3. 提案システムの実装

本研究では 2 章で述べた提案システムを、コミュニティサービスを対象とした広告配信システムとして実装した。広告配信システムはテキストおよびハイパーリンクによって構成される広告を対象にしており、実験においては携帯電話サービス用のアフィリエイトプログラ

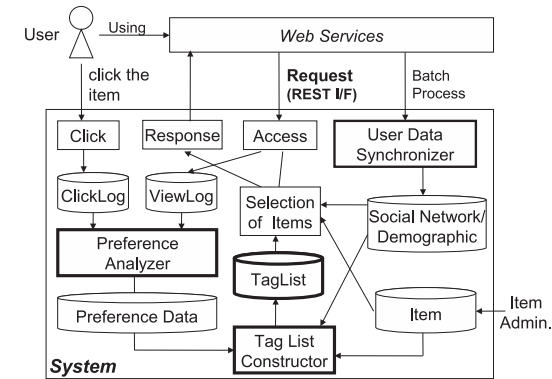


図 6 実装された提案システムの全体的な構成
Fig. 6 An overview of the architecture of the implemented system.

ムの 1 つである Pocket Affiliate^{*1}からデータを参照した。実装においては以下の点に注意しながら設計を行った。

- 独立性

提案システムは既存の SNS のようなコミュニティサービスの内部で用いられることを想定している。そのため提案システムを利用する Web サービス（利用元サービスと呼ぶ）とは独立に設計され、利用元に対して最小限の作業で利用可能となることが望ましい²²⁾。

- 応答時間の短さ

近年の Web サービスは特に、リクエストからレスポンスまでの時間が短いことが望まれている^{23)–24)}。そこで一定の処理をバックエンドで行うことにより処理を高速化する必要がある。

3.1 全体の構成

実装されたシステムの全体の構成を図 6 に示す。

システムは利用元となる Web サービスとは独立に構築され、REST アーキテクチャによって構成される。REST は HTTP 上で動作するソフトウェアアーキテクチャであり広く用いられている方式の 1 つである²⁵⁾。

*1 <http://smaf.jp>

また Web サービスから人間関係ネットワークおよび Demographic Data を抽出する必要があるため、バックエンドプロセスによってデータの同期化が User Data Synchronizer によって定期的に行われる。

提案システムの工程をすべてリアルタイムに実行する場合、多くの計算資源が必要となり、応答時間が長くなってしまふ。そこで実装では Preference Analyzer および Tag List Constructor の 2 モジュールにより必要な計算の一部がバックエンドでバックエンドプロセスとして行われ、オンラインでの処理が軽減される仕組みとなっている。

3.2 Preference Analyzer

提案システムにおいては、アイテム推薦時とクリック時に Preference Data が逐次更新されることが理想的であるが、そのたびに過去のログを参照すると大きな計算コストがかかる。そこで図 6 における Preference Analyzer では、バックエンドプロセスとして Preference Data の更新が行われる。

式 (6) は、時刻 $T_{now} - \Delta T$ 時点で集計された $N_{u,t}^{click}(T_{now} - \Delta T)$ を用いると以下のように変形できる。

$$\begin{aligned} N_{u,t}^{click}(T_{now}) &= \sum_{\ell \in S} \exp\left(-\frac{T_{now} - T_{\ell}}{T_{int}}\right) + \sum_{\ell \in \bar{S}} \exp\left(-\frac{T_{now} - T_{\ell}}{T_{int}}\right) \\ &= \sum_{\ell \in S} \exp\left(-\frac{T_{now} - T_{\ell}}{T_{int}}\right) + \exp\left(-\frac{\Delta T}{T_{int}}\right) N_{u,t}^{click}(T_{now} - \Delta T) \end{aligned} \quad (11)$$

ただし S は、ユーザ u 、タグ t に関する clicklog の集合 $C_{u,t}$ の中で、 $T_{now} - \Delta T$ より後のタイムスタンプを持つものの集合を表しており、 $\bar{S}$ は $T_{now} - \Delta T$ 以前のログの集合を表している。この式は、 $N_{u,t}^{click}(T_{now} - \Delta T)$ の値をサーバ内に保持しておくことにより、 $T_{now} - \Delta T$ 以前の clicklog を参照をしなくても Preference Data を計算することが可能であることを示している。

実装においてはさらに簡略化し、次のような更新式を用いた。

$$\hat{N}_{u,t}^{click}(T_{now}) = N_S + \exp\left(-\frac{\Delta T}{T_{int}}\right) \hat{N}_{u,t}^{click}(T_{now} - \Delta T) \quad (12)$$

ここで N_S は、 S の要素数を示している。この更新式がバッチ処理として、 ΔT ごとに行われる。このような更新式を用いることにより、Preference Data の更新時には過去のログをすべて参照することなく処理を行うことが可能となる。 ΔT は、 T_{int} に対して十分小さい値であれば簡略化による計算結果の誤差は非常に小さいものになる。

なお、本論文での実装においては、 T_{int} を 48 時間、 ΔT を 1 時間として更新を行って

る。 $N_{u,t}^{view}$ に対しても同様の処理が行われる。

3.3 Tag List Constructor

アイテム選択においては、各アイテムが式 (10) で定まる確率に従って選択される。ここで簡単な計算の高速化方法として、アイテムについての選択確率の計算をリクエストごとに行うのではなく、あらかじめ計算確率を計算しておく方法がある。しかしながらこの方式を用いた場合でも、選択処理においてはアイテム全体を検索対象としなければならない。アイテム数が多い場合はこの選択処理が高負荷となってしまう。

そこで実装においては図 7 に示すようなフローで処理が行われる。ユーザからのリクエストを受信すると、まずすべてのタグの中から、そのユーザに推薦される可能性の高いアイテムに関連するタグが選択される。次に選択されたタグに関連するアイテムの中から確率的な選択が行われる。最後に Demographic Filtering を行い、アイテムが配信可能であれば配信を行う。

ユーザ u からリクエストが来た際、タグ t が選択される確率 $Prob_u(t)$ を以下のように定める。

$$Prob_u(t) = \frac{p_t^u v_t}{\sum_{k=1}^D p_k^u v_k} \quad (13)$$

$$v_t = \sum_{i=1}^M I_t^i \quad (14)$$

ここで D はユーザプロフィールの次元数を示しており、タグの種類数と一致する。 p_t^u は

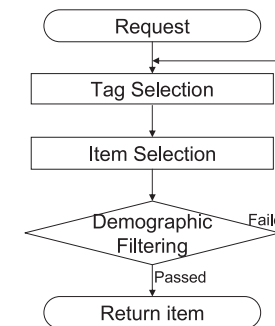


図 7 実装されたシステムにおけるアイテム選択の流れ

Fig. 7 Implemented flow of item selection.

ユーザ u のユーザプロファイルの t 番目の要素を示し、 I_t^i はアイテム i の特徴ベクトル I^i の t 番目の要素を示している。

次に、選択されたタグ t に従ってアイテム j が選択される確率 $Prob_u(j|t)$ を以下の式で計算する。

$$Prob_u(j|t) = \frac{I_t^j}{v_t} \quad (15)$$

以上 2 段階の選択により、ユーザに推薦するアイテムが決定される。このようにタグの選択を先に行ってからアイテムを選択することにより確率的選択の候補数を減らし計算効率を向上させている。この場合のアイテム選択確率 $Prob_u(j)$ は式 (13)～式 (15) を用いると以下のように計算できる。

$$\begin{aligned} Prob_u(j) &= \sum_{t=1}^D Prob_u(j|t) Prob_u(t) = \sum_{t=1}^D \frac{p_t^u v_t}{\sum_{k=1}^D p_k^u v_k} \cdot \frac{I_t^j}{v_t} \\ &= \frac{\sum_{k=1}^D p_k^u I_k^j}{\sum_{i=1}^M \sum_{k=1}^D p_k I_k^i} = \frac{\mathbf{P}^u \cdot \mathbf{I}^j}{\sum_{i=1}^M \mathbf{P}^u \cdot \mathbf{I}^i} \end{aligned} \quad (16)$$

これは式 (10) と一致する。特にアイテムに関連するタグが少数である場合、タグに関連するアイテム数が少なくなり、大幅な計算量の削減が可能となる。また実装においてはユーザごとのタグの選択確率 $q(t)$ はバックエンドプロセスとして計算することが可能であり、その計算結果は Tag List として保存される。

4. 評価

4.1 推薦アルゴリズムの効果

実装された提案システムの推薦効果を評価するため、実際のコミュニティサイトに組み込んで効果測定を行った。

実験条件を表 1 に示す。実験に用いたのは運用されている携帯電話向けの SNS であり、アイテムは実験時に Pocket Affiliate に掲載されていたアイテムを登録することで、アイテムの定義を行った。この実験では、ユーザが携帯電話からサービスにアクセスすると、ページ内にテキストによって提案システムからの広告が表示される。携帯電話向けの SNS を選んだ理由は、携帯電話では一般に広告提示に対するユーザの反応が大きく、またアクセス回数も多いためである²⁶⁾。

図 8 は、実験期間 10 日間による、友人数別の推薦効果を示した図である。推薦効果の指

表 1 実験条件と実験環境の詳細

Table 1 Environment and conditions of the experiment.

ユーザ数	949 人
平均リンク数	3.8
実験日数	10 日
アイテム数	150
アイテムあたり平均タグ数	3.0
タグの種類	32

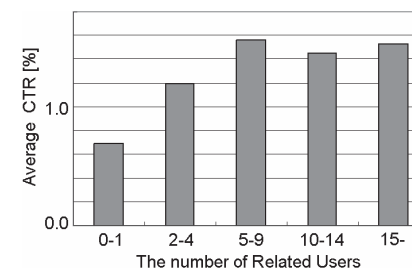


図 8 推薦アルゴリズムによる CTR 向上の効果

Fig. 8 Experimental results of CTR by users.

標として CTR を採用している。本実験での全ユーザの平均 CTR は 0.98% であった。

図 8 によると、友人が 0 人または 1 人のユーザに対する広告効果よりも、友人が 2 人以上の広告効果のほうが高かったといえる。また全ユーザに対してアイテムのランダム配信を行った場合の CTR は 0.52% であった。これにより提案システムの推薦効果は 1.9 倍となっており、ユーザの興味に合致したアイテム推薦が行われていることが確認できる。友人数が多いほど CTR が高くなっているが、本実験では友人数 5 人以上で効果が飽和していた。これは本システムではユーザプロファイルがタグごとでしか行われていないため、細かいユーザの興味まで注目されていなかったことが理由の 1 つであると考えられる。

また友人を考慮せずにユーザプロファイルの構築を行った場合のシステムでは、平均 CTR は 0.67% となった。したがって関連ユーザ情報の利用は有効であったことが確認できる。

図 9 は、実験期間中における日付ごとの CTR の推移を示したものである。Proposed は提案法、Profiling は関連ユーザ情報を利用しない場合のプロファイルシステム、Random はランダムな配信を示している。一般に広告等の推薦システムは提示回数に応じて CTR が下がる傾向にあるが、提案法は CTR1% 近くを維持している。

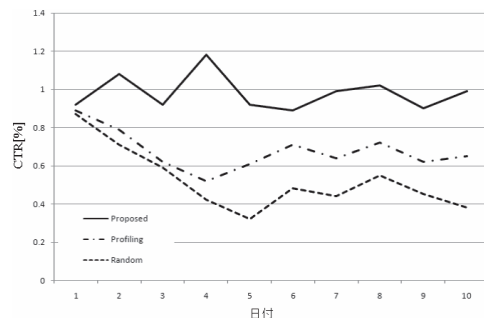


図 9 推薦アルゴリズムによる CTR 向上の効果
Fig. 9 Experimental results of CTR by users.

表 2 計算機環境の詳細
Table 2 Specification of the calculation test.

CPU	AMD Sempron(tm) Processor 2800+ 1.6 GHz
メモリ	500 MB
Web サーバ	2.0.53
プログラム環境	PHP5.2.3
データベース	MySQL5.0.27

4.2 計算時間の評価

次に提案システムを Web サービスで利用する際の実用性を評価するため、アイテム数・タグ数ごとの計算コストを計測した。

利用した計算機は 1 台であり、Web サーバおよびデータベースサーバの役割を備えた Linux サーバを用いた。開発プラットフォームは標準的なオープンソースソフトウェアを用いた。計算時間評価のための各ハードウェア仕様および用いた各ソフトウェアの詳細を表 2 に示す。

計算時間の測定には、仮想的なアイテムを擬似的に生成することで行われた。以下の 3 つのシステム間の比較を行った。

- System 1：実装した提案システム。タグ数を 50, 1 アイテムあたりの関連タグ数を 3 とした場合。
- System 2：実装した提案システム。タグ数を 200, 1 アイテムあたりの関連タグ数を 10 とした場合。

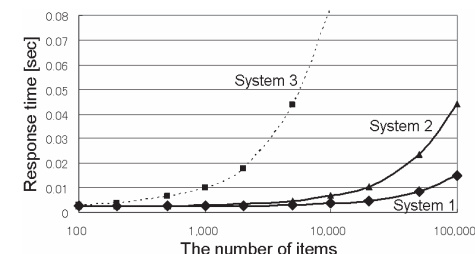


図 10 アイテム数ごとの応答時間
Fig. 10 A graph of response time corresponding to the number of items.

- System 3：提案システムにおいて Tag List を用いずにユーザプロフィールからアイテムの選択を行うシステム。タグ数を 50, 1 アイテムあたりの関連タグ数を 3 とした場合。アイテム数を 100 から 100,000 まで推移させて平均応答時間を測定した。ここでの応答時間は、データベース接続からアイテム選択までの一連のプログラムの流れの処理に要した時間を示しており、プログラムの内部で計測されたものである。

図 10 に、アイテム数と平均応答時間のグラフを示した。図 10 によると Tag List を導入した場合、導入しない場合に比べて計算時間が軽減できている。System 3 で 100,000 アイテムの場合のシステム応答時間は 0.61 秒であった。System 1 と System 2 を比較すると、System 2 の状況で計算コストが大きくなっているが、その場合でも 100,000 アイテムで 0.05 秒以内に収まっており、多くのアイテム数を扱う場合でもユーザビリティの観点で軽快に動作する推薦システムとなっている²⁴⁾。

4.3 考 察

提案システムは、協調フィルタリングのようなアルゴリズムにおける計算資源の問題を解決することが動機の 1 つとなっている。そこでまず、計算コストの観点から提案システムと協調フィルタリングの比較についての考察を行う。

ユーザ数を U 、平均の関連ユーザ数を $\bar{R}$ とする。一般的な協調フィルタリングでは、ユーザ間の関連度を相関等を用いて計算し、それらと既知の情報の加重平均をもってユーザプロフィールリングを行う。この考え方は提案システムと類似しているが、関連度の計算を関連ユーザに限定せず、全ユーザを対象にしている点で異なる。この場合、協調フィルタリングではユーザ間の関連度計算のコストは $O(U^2)$ となる一方、提案システムの関連度計算は

$O(U\bar{R})$ にとどまる．たとえば mixi^{*1}を例とした場合，そのユーザ数は2008年7月13日の段階でユーザ数が $U = 15,000,000$ を突破したと発表されている一方で，松尾らのmixiの分析²⁷⁾によれば平均の友人数は $\bar{R} = 10.48$ と非常に小さい．この場合は計算コストに大きな開き生まれることは明らかである．

また提案システムが協調フィルタリング等従来のアルゴリズムに比べて優れている点として，計算コスト以外にも新規ユーザに対する推薦精度が上げられる．これは対象ユーザに対するログが存在しない場合でも，友人の情報を参照することでユーザプロフィールを構築できるためである．実験期間中の新規ユーザは13人であり，その平均CTRは0.70%であった．これは全ユーザの平均CTRよりは小さいものの，ランダム配信よりも効果が高いことを示唆している．

次にパラメータについての考察を行う．本実験では， γ, T_{int} は予備実験を通して暫定的に設定されたものを用いたが，これらの値は長期運用を通して経験的に定めることが望ましい．またこれらの最適値はユーザによって大きく異なることも考えられる．したがって今後の課題として，これらのパラメータに関するシステムの拡張および最適値の設定があげられる．

5. おわりに

本論文では人間関係ネットワークに基づく情報推薦システムの構築について述べた．ユーザプロフィール構築において，人間関係ネットワークで定義される関連ユーザの情報をを用いて補完する推薦アルゴリズムを採用している．新規ユーザに関するログ情報が存在していない場合でも，関連ユーザの情報をを用いることでユーザプロフィールが構築可能である点が特徴である．

提案システムは実用的なWebサービスとして実装され，広告配信として実際のコミュニティサービスで運用された．評価実験においては推薦効果が確認された．またアイテム数が多い場合でも高速にアイテムが推薦できることが確認されている．

今後の研究課題としてユーザ間の類似度計算があげられる．提案システムでは構造同値性に基づいた一律的な類似度計算を行っているが，実際には嗜好の異なるユーザも多く存在すると考えられる．そのため，ログに基づいて類似度を学習する仕組みが実現されると，より効果的なアイテム推薦が可能になると考えられる．

*1 <http://mixi.jp>.

謝辞 本研究は日本学術振興会特別研究員奨励費(19・11037)の助成を受けて行われた．

参考文献

- 1) Oreilly, T.: *What is web 2.0: design patterns of business models for the next generation of software*, Oreilly Media (2007).
- 2) 梅田望夫：ウェブ進化論—本当の大変化はこれから始まる，筑摩書房(2006).
- 3) 安田 雪：社会ネットワーク分析—何が行為を決定するか，新曜社(2005).
- 4) 内田 誠，白山 晋：SNSのネットワーク構造の分析とモデル推定，情報処理学会論文誌，Vol.47, No.9, pp.2840–2849 (2006).
- 5) Albert, R. and Barabasi, A.-L.: Statistical mechanics of complex networks, *Reviews of Modern Physics*, Vol.74, No.1, pp.47–97 (2002).
- 6) Gomez, V., Kaltenbrunner, A. and Lopez, V.: Statistical analysis of the social network and discussion threads in slashdot, *Proc. WWW2008*, pp.645–654 (2008).
- 7) 大向一輝，武田英明：人間関係ネットワークに基づく情報フィルタリングを用いた協調的タスクスケジューラ，電子情報通信学会論文誌 DI，Vol.87, No.11, pp.1020–1029 (2004).
- 8) 松尾 豊，友部博教，橋田浩一，中島秀之，石塚 満：Web上の情報からの人間関係ネットワークの抽出，人工知能学会論文誌，Vol.20, No.1, pp.46–56 (2005).
- 9) Riecken, D.: Introduction: personalized views of personalization, *Comm. ACM*, Vol.43, No.8, pp.26–28 (2000).
- 10) 土方嘉徳：情報推薦・情報フィルタリングのためのユーザプロフィール技術，人工知能学会論文誌，Vol.19, No.3, pp.365–372 (2004).
- 11) 奥 健太，中島伸介，宮崎 純，植村俊亮：状況依存型ユーザ嗜好モデリングに基づく Context-Aware 情報推薦システム，情報処理学会論文誌：データベース，Vol.48, pp.162–176 (2007).
- 12) 村上知子，古岡信和，折原良平，古川康一：CAM法を用いた個人嗜好モデルに基づく商品推薦システム，人工知能学会論文誌，Vol.20, No.5, pp.346–355 (2005).
- 13) Ribeiro-Neto, B., Cristo, M., Golgher, P.B. and de Moura, E.S.: Impedance coupling in content-targeted advertising, *Proc. SIGIR*, pp.496–503 (2005).
- 14) Nazerzadeh, H., Saberi, A. and Vohra, R.: Dynamic cost-per-action mechanisms and applications to online advertising, *Proc. WWW2008*, pp.175–186 (1994).
- 15) Resnick, P., Iacovou, N., Suchak, M. and Peter Bergstrom, J.R.: GroupLens: An open architecture for collaborative filtering of netnews, *Proc. CSCW94*, pp.175–186 (1994).
- 16) Linden, G., Smith, B. and York, J.: Amazon.com recommendations-Item-to-item collaborative filtering, *IEEE Internet Computing*, Vol.7, No.1, pp.76–80 (2003).
- 17) Hotho, A., Jaschke, R., Schmitz, C. and Stumme, G.: *The Semantic Web: Research and Applications*, Lecture Notes in Computer Science, pp.411–426, Springer

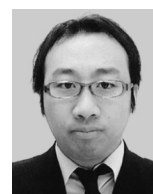
(2006).

- 18) Scott, J.: Social Network Analysis, *Sociology*, Vol.22, No.1, pp.109–127 (1988).
- 19) Maoz, Z., Kuperman, R.D., Terris, L. and Talmud, I.: Structural Equivalence and International Conflict: A Social Networks Analysis, *Journal of Conflict Resolution*, Vol.50, No.5, pp.664–689 (2006).
- 20) Nishimura, T., Matsuo, Y., Hamasaki, M., Fujimura, N., Ishida, K., Hope, T. and Nakamura, Y.: A method of social network extraction via Internet and networked sensing, *Proc. INSS2006* (2006).
- 21) Leicht, E.A., Holme, P. and Newman, M.E.J.: Vertex similarity in networks, *Physical Review*, pp.1–10 (2006).
- 22) Felfernig, A.: Koba4MS: Selling Complex Products and Services Using Knowledge-Based Recommender Technologies, *Proc. IEEE International Conference on E-Commerce Technology*, pp.92–100 (2005).
- 23) Ethiera, J., Hadayaa, P., Talbotb, J. and Cadieuxa, J.: *Interface design and emotions experienced on B2C Web sites: Empirical testing of a research model*, Computers in Human Behavior, Elsevier (2008).
- 24) Palmer, J.W.: Web site usability, design and performance metrics, *Information Systems Research*, Vol.13, No.2, pp.151–167 (2002).
- 25) Fielding, R.T. and Taylor, R.N.: Principled design of the modern Web architecture, *Proc. International Conference on Software Engineering*, pp.407–416 (2000).
- 26) モバイル・コンテンツ・フォーラム：ケータイ白書 2008，インプレス R&D (2007).
- 27) 松尾 豊，安田 雪：SNS における関係形成原理—mixi のデータ分析，人工知能学会論文誌，Vol.22, No.5, pp.531–541 (2007).

(平成 20 年 9 月 18 日受付)

(平成 21 年 1 月 9 日採録)

(担当編集委員 瀬本 明代)



堀田 創 (正会員)

2005 年慶應義塾大学理工学部情報工学科卒業．2006 年同大学大学院修士課程修了．2008 年同大学院博士課程修了．日本学術振興会特別研究員 DC1．現在，同大学訪問研究員．インターネットサービス・ソフトコンピューティング・感性工学の研究に従事．電子情報通信学会，ヒューマンインタフェース学会各会員．



萩原 将文 (正会員)

1982 年慶應義塾大学工学部電気工学科卒業．1987 年同大学大学院博士課程修了．工学博士．同年同大学助手．以来，ニューラルネットワーク，ファジィシステム，進化計算，感性工学の研究に従事．現在，同大学教授．1991～1992 年度スタンフォード大学訪問研究員．1986 年丹羽記念賞，1987 年電子情報通信学会学術奨励賞，1990 年 IEEE Consumer Electronics Society 論文賞，1994 年安藤博記念学術奨励賞，1996 年日本ファジィ学会著述賞，2003 年日本感性工学会技術賞，2004 年同学会論文賞受賞．電子情報通信学会，日本知能情報ファジィ学会，人工知能学会，電気学会，日本神経回路学会，日本感性工学会，日本デザイン学会各会員．IEEE シニアメンバ．