



## Abstract

Information keeps emerging on the Internet, which urges users to separate the ones they need from the tremendous amount of others, within the extent of their limited information-handling capabilities. Users have developed several ways to reduce the information they must process, such as 1) receive information only from authorities or limited popular sites, and 2) receive information only from several media of their interests, such as SNS or similar communication tools. These actions predefine the range of information the users want to receive, without necessitating them to process information outside of the defined range.

An ideal of information circulation on the Internet is a state in which anyone can transmit to and receive from the world the information as they wish, without necessitating any authorities or economic power. The actions many users are now taking to avoid the information flood raise some issues with respect to the realization of the ideal. The former type of actions may make it difficult to receive information based on different senses, and have a risk of submitting to one single authority. The latter type of actions may cause "Echo Chamber Effect" where predictable information only can be received.

This research has devised a model for information search which utilizes the closeness of senses among users. The research verified, by operation experiments, that the two evasive actions from information flooding: 1) dependencies on set metrics such as authorities or popularity, and 2) predetermining the range of information to receive and never process others, occur less after the model is applied.

[ キーワード ] 情報過多、パーソナライズ検索、ブログ、SNS、共感、信頼

# ユーザ間の主観的センスの共感を用いたブログ検索システム

Blog search system utilizes the closeness of senses among users.

須子 善彦 / Yoshihiko Suko( 慶應義塾大学 政策・メディア研究科 後期博士課程 )

村井 純 / Jun Murai( 慶應義塾大学 環境情報学部 教授 )

1. 目的
  2. 背景
  3. 問題提起
  4. 解決する課題と手法
  5. 設計と実装
  6. 評価
  7. 結論
  8. 今後の展望
- 謝辞

## 1. 目的

本研究では、個々のユーザによって異なる主観的な価値判断を考慮した情報検索・流通システムの原型について考察した。このために特別なRSSリーダ機能を備えたSNS型ブログ検索システムを開発して実証実験を行った。本研究の出発点は、インターネットにおける情報検索・情報取得の際に、情報過多を原因として起きるユーザの各種回避行動がもたらす問題の軽減である。今日、ユーザが情報過多現象に対してとる各種回避行動は、いくつかの点で理想的な情報流通を阻む課題をかかえている。ここでいう各種回避行動は、(1)権威やPage Viewの高い人気のある一部のサイトの情報しか受け取らない、あるいは、Google Page Rank[1][2] などのような画一的な評価軸に従って取得する情報を絞る行動(以下、「画一化された権威的価値基準の再支配」問題と表す)、(2)興味のあるいくつかのメーリングリストに参加して、その情報だけを受け取る。あるいは、SNS(ソーシャルネットワーキングサービス)等を使って知人のブログしか読まないというように、あらかじめ取得する情報の範囲を設定して、その範囲外の情報は処理をしないといった行動(以下、「情報取得範囲の限定」問題と表す)に典型的にあらわれている。本研究では、これらの回避行動がもたらす問題を軽減させるような情報検索・流通システムを開発して実証実験することにより、インターネット上の「理想的」な情報流通の実現の手がかりとしたい。

## 2. 背景

---

今日インターネット上には日々大量の情報が生成されている。しかしインターネットユーザである人間が一定時間内に処理できる情報の量には限界がある。また、その能力は短期間には増加しない。インターネット上でやりとりされる情報量は現在も日々増加しているため、その差はますます広がってゆくであろう[3]。ユーザは大量の情報の中から、自身の限られた情報処理能力の限界範囲内で必要とする情報を取捨選択する必要に迫られている。あらかじめ情報をふるいにかける(以下、この行為を「情報フィルタリング」と表す)情報処理能力の限界の範囲内に情報量を少なくした上で、本格的に情報の内容を吟味する。情報フィルタリングの手段には以下のような方法がある。

- 1)権威や人気のある一部のサイトのみを閲覧する。検索エンジンのアルゴリズム等の画一的な評価軸に従って取得する情報を絞る
- 2)興味のあるいくつかのMLに参加し、その情報だけを受け取る。SNS等を使って知人のブログだけを読む

## 3. 問題提起

---

### 3.1 インターネットがもたらした可能性と理想の情報流通

インターネットの情報過多状態に対して、今日ユーザが行う上記のような情報フィルタリングの行動は、いくつかの点で「理想的」な情報流通の実現を阻む問題を内包している。インターネットにおける情報流通の理想形のひとつは、誰もが既存の権威や経済力などを必要とせずに世界中に情報を発信し、世界中から必要とする情報を得ることである。インターネットの登場によって情報の流通コストが劇的に低下した。これによって従来は困難であったニッチ情報の流通、すなわち、一部の人たちだけが必要とする情報を容易に発信し、また、受け取ることが可能となった。従来は情報発信のコストが高かったため、そのコスト負担に見合うだけの公共性や市場性のある情報、すなわち、マスに受け入れられる情報だけが、テレビや新聞等のマスメディアによって流通していた。しかしインターネットの登場によって、情報発信のコストの低下し、個人による情報発信が可能となった。その結果として多様な情報の流通が実現つつある。たとえば音楽やファッション、映画など、趣味・嗜好は個人によって異なるが、それらに関する情報流通においても多様な価値観の情報が流通している。従来はマスに評価される情報のみであったものが、インターネットは少数派の趣味・嗜好情報の流通を容易にし、より多様な価値観の情報流通を可能にした。また、少数の人間だけがかかる病気の情報など、マスの人間を対象としない情報でも、当事者にとっては非常に重要な情報であるといった、個人個人のニーズにあった情報取得ができる可能性を高めた。これに対して前述した情報過多に対するユーザの回避行動は、このような理想の情報流通の実現を阻む可能性を持つことになる。

### 3.2 「画一化された権威的価値基準の再支配」が引き起こす問題

多くのユーザが利用している検索エンジンは、趣味・嗜好のような個々人によって異なった価値判断をする情報の検索に対応していない。たとえば、多くのユーザが利用しているGoogle検索[4]は、PageRank[1][2]というアルゴリズムに基づいて検索対象のWebページを重み付けし、ランキング形式で表示する。PageRankはWebページ間のリンク構造をグラフ論的アプローチで処理するもので、簡単な概要を述べると、非リンク数の多いページはそれだけ重要なページとして扱われ、重要なページからリンクされたページも重要なページとみなされる、といった具合に、ページの重要性を評価し、検索結果のランキングに反映している。したがって同一の検索キーワードを入力して検索を行えば、例え趣味・嗜好が異なる二人のユーザが検索を行っても、基本的にはランキング結果が同一となる。

また、多くのPage Viewを集める人気の高いサイトとして、Yahoo! [5]等のポータルサイトを利用するユーザも多い。これら情報ポータルは、その性質上、多くのユーザに支持されるコンテンツを重要視して扱う。したがって、万人受けする情報が優先的に扱われ、先に述べたような個々人によって価値判断が異なる情報、ニッチな情報の入手には向いていない。つまり、GoogleのPageRankのような検索エンジンのアルゴリズムや、たくさんのPage Viewを集めるポータルサイトという存在は、前インターネット時代におけるマスメディアのように、マスを対象に画一化された価値基準による情報フィルタリング手段であると考えられる。

これらの手段をもちいて情報過多を回避する行動は、主に効率性の面において有用であるが、一方でニッチな情報や個々人によって価値が異なる情報の流通、検索手段としては有用性が限定される。情報流通が単一の権威に依存しすぎることは、Single Point of Failureのリスクを生む。またたとえば、「Google八分」といった言葉が生まれたような状況、すなわち、Googleに否定されると検索結果に存在できなくなり、事実上インターネットから存在が消されるに等しい状態は、情報流通が単一の権威によって支配されることで生じる権力や失われる自由を示唆する。実際、中国における検閲や反サイレントロジサイトの削除において、Googleへの批判と共にGoogleの支配的権力を危惧する指摘が行われている。

### 3.3 「情報取得範囲の限定」が引き起こす問題

二点目の回避行動については、「情報取得範囲の限定」問題、すなわち、あらかじめ取得する情報の範囲を設定し、その範囲外の情報は処理をしないことも情報流通の理想形の実現を阻むいくつかの問題を引き起こす可能性がある。たとえば、自分が取得したい情報に関する、ディスカッション等が行われているメーリングリストに参加することがある。このこと自身も情報フィルタリングの一つの手段であり、有用な手段の一つである。しかしたとえば、自分が取得したい情報に関して、議論しているインターネット上の場所は、そのメーリングリストに限らない。しかしながら、一旦メーリングリストに参加すると、どうしてもそこで流れている情報だけを処理するようになり、他の場所で議論されている情報が目に触れなくなってしまふ。また、メーリングリストにはブログなどの他のメディアと比較して強いメンバーシップがある。メンバーでないとメールを受け取れなかったり、メールを投稿できないといった設定が、情報の伝達範囲を限定してしまう。また、メンバー内に、発言数の多い人や、管理人といった、いわゆる声の大きい人、議論の進行において影響力の大きい人がでてくると、その人の意見と異なる意見の人はメンバーから外れ、別のメーリングリストで議論を開始するといったことも起きる。この結果、同一メーリングリスト内には、似たような意見を持つ人のみが残し、多様な価値観がやりとりされにくくなる場合も多々あるであろう。このような例からも示唆されるように、自分の参加しているメーリングリストで流通している情報が完全無二

の存在だという保証はなく、また多様な価値観を扱えているとは限らない。

最近登場したSNSは一種の会員制サイトで、サイト上でユーザがお互いの知人関係を公表することで、交友関係が広がる、ないしは信頼感のあるコミュニケーションが行える点などから人気を集めている[6]。SNSには、サイト内でブログを書いたり、サイト外のブログを読み込む機能があり、この機能を用いると、知人に自分のブログを読みやすくさせることができる。ユーザは、SNSサイトに登録し知人を登録すれば、自分の知人の書いたブログを読みやすくなる。これは簡易的なRSSリーダーの機能だということができる。この機能を用いてSNSに知人登録されている人のブログしか読まないといった人も多い。このことも、先に述べたメーリングリストと同様に、多様な情報を受け取りにくくする可能性がある。

メーリングリストや、SNSの問題点は、あらかじめ自分が必要とする情報がどの範囲から入ってくるか、といった「取得する情報の範囲」をあらかじめ設定してしまうことである。このことは、範囲の設定を間違えることによって生じる情報伝達上のリスクの発生や、さらに、自分が必要とする想定内の情報しか受け取らない一種の「タコツボ化」現象[7]を引き起こしてしまう可能性をもつ。またグラノヴェッター (Mark S. Granovetter) [8] が弱い紐帯理論において述べたように、あらかじめ設定した範囲の情報はよく知っている知人と同様の問題を持っている可能性がある。しかし逆に設定した範囲外の情報のほうが、ときに未知で有意義な情報であるということもおおいにあるであろう。

### 3.4 解決する課題

以上のような観点から本研究では、ユーザが情報過多を回避するためにとる二つの行動が持つ問題、「画一化された権威的価値基準の再支配」問題、「情報取得範囲の限定」問題の発生を軽減させるような情報検索・流通システムを開発して実証実験することにより、インターネット上の「理想的」な情報流通の実現の手がかりとしたい。

## 4. 解決する課題と手法

---

前述したユーザの回避行動が引き起こす二つの問題を軽減するため、情報フィルタリングにおいて画一化された権威的価値基準に頼らず、かつあらかじめ取得する情報の範囲を事前決定せずに、より広範囲かつ範囲の境界が動的である情報フィルタリング・情報検索を実現するためのシステムを開発して実証実験を行った。具体的には、知人関係をベースにしたユーザ間のセンスの共感度を用いる情報検索モデルを考案し、ブログ検索システムとして実証環境を構築した。本稿でいう「センスの共感度」とは、同じ情報に対して同じ評価を行う度合いを示す。これは情報へのユーザ個人によって異なる価値判断を考慮した情報フィルタリングを行うためのものである。

具体的には、個々のユーザがある情報(たとえばブログ記事)に対して行う価値判断(自分にとって役に立つ情報であったか否か。以下、「評価」と表す場合もある)をシステムが記録・ユーザ間で共有し、同じ情報に対して同じような価値判断を行うユーザの発見を容易化する。システムに記録された価値判断の履歴をもとに、ユーザ間の「センスへの共感の度合いの大きさ」、すなわち、同じ情報に同じ価値判断を行うであろうという確率を予測し、確率の大きいユーザ間で情報を優先的に流通させることで、パーソナライズ化された情報流通を実現し、情報過多による問題を解決するシステムである。これらは、たとえば「あの人の映画評はいつも役に立つから、別の友達の映画評よりも重要だ」といった、われわれの日常において行われている情報流通モデルに基づいている。

われわれは日常、全ての他人からの情報を平等に扱うのではなく、知人、それも過去の関係の経験が

ら、「あるテーマの情報に関してはこの知人」という具合に、いわば「師匠」のような存在(先の定義によれば「センスへの共感の度合いの大きい」人)を見つけ、その人からの情報を優先的に処理している。言い換えると、その人に情報の価値判断を委譲しているともいえる。(図1)なお、映画、音楽、グルメなどなど、情報のトピック毎、すなわちセマンティクスのクラスター毎に別の人が「師匠」になることもあり、「師匠」の数は複数人いるだろう。また、ある人の「師匠」も同様に自身の「師匠」を持つ。それらの「師匠」間の関係がネットワークとして数珠繋ぎになれば、「師匠」関係ネットワークによるセンスの適材適所化ともいえるべき情報流通ネットワークが形成されることになる。実際に「師匠」関係ネットワークは、今までもわれわれの日常において存在しているが、インターネットが登場しブロードバンドが普及するまでの長い間それを可視化し、その効力を増大する手段が存在しなかった。本研究では、「師匠」関係ネットワークモデルをシステムとして実装し可視化し、情報過多時代における情報検索のためのプラットフォームとして応用化したものだといえることができる。

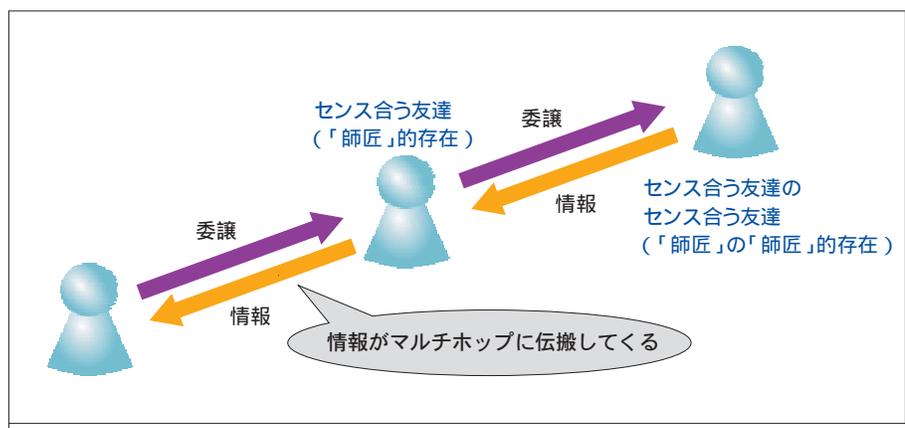


図1 センスの共感に基づく情報流通モデル

先に述べた情報検索・情報流通モデルを、SNSを用いたWebベースのRSSリーダ兼ブログ検索システムとして実装した。本システムの名称を「STN(Social Trust Networking)」と名付けた。本システムはユーザ間の知人関係を管理するいわゆるSNSの機能を持ち、その知人関係の情報を活用したRSSリーダ機能をユーザに提供する。ユーザは、SNS機能によって知人やお気に入りのブログ作者を登録し、WebベースのRSSリーダと同様に、自分が読みたいブログ等のRSSフィードをSTN上に登録する。ユーザは、知人関係ネットワーク経由で自分の知人や知人の知人、他のユーザ等が書いたブログや普段使っているRSSフィード等を読む。これらブログやRSSフィードの記事は、知人ネットワーク上を自動的にマルチホップに伝播してきたもので、ユーザが最初にログインした画面で自動的に表示される。ユーザはこれらの記事を読み、記事ひとつひとつに対して、その情報に対する価値判断を個々に行う。つまり、その記事の情報が自分にとって価値があったかどうかを評価付けする。ユーザが行った価値判断をユーザ毎に統計処理し、ユーザ間の主観的センス共感度の近さをアルゴリズムによって算出する。算出された主観的センス共感度の近さに従って、先ほどのユーザ間で伝播させたブログ記事(ブログエントリーともいう)等の優先度を変える。つまり、主観的センス共感度のユーザが書いた、あるいは読んで高く評価している記事が、ユーザに優先的に伝播されるようになる。つまり、各ユーザが主観的に記事を評価していくことで、マルチホップの情報取得のネットワークが最適化され、個々のユーザによって異なった情報への価値判断基準にしたがって、重要だと思ふ情報を優先的に目にするようになる。言い換えれば、情報のほうから磁石に

吸い寄せられてくるかのように個々のユーザに近づいていくような情報検索・情報流通を実現するブログ検索・RSSリーダーだということができるであろう。(図2)

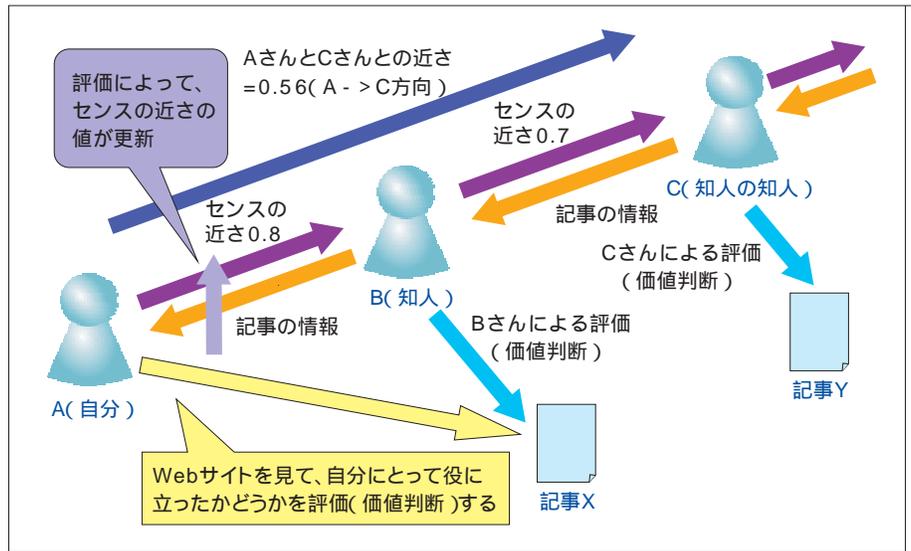


図2 システムの概要

多くのソーシャルウェアは、ユーザが生成する知をシステム全体のひとつの集合知として扱うが、STNは、ユーザが生成する知を全体で共有するのではなくセンスの近いユーザ毎に局所的に利用する点の特徴になっている。次節以降はシステムの詳細について述べたい。

## 5. 設計と実装

### 5.1 機能一覧

STNシステムでは以下の機能を設計して実装した。

#### 1) ユーザ情報の管理、知人関係の管理およびメッセージング機能

ユーザが本システムを利用するために必要なユーザアカウントを作成する機能。アカウント作成後、ユーザが自分のプロフィールや顔写真に関する情報を作成、表示、更新する機能。これらのプロフィールは、一般的なソーシャルネットワーキングサービスと同様に、他のユーザから参照される。ユーザの知人関係の構築、表示、管理。その他、ユーザ間のメッセージ交換といった一連のソーシャルネットワーキングサービス機能を備える。

#### 2) RSS登録機能

ユーザ自身が書いているブログや、通常のRSSリーダーのように読みたいブログのRSSフィードを登録する「お気に入りブログ登録」機能。ユーザが登録したURI等をリレーショナルデータベース(以下、DBと表す)に保存する。また、ユーザがお気に入りの他ユーザを登録する「お気に入りブログ作者登録機能」がある。お気に入りブログ作者として登録すると、システム的にはセンス共感度ネットワークのリンク関係として扱われる。つまり、知人関係と同様に扱われる。リンク関係を結ぶと後述するランキングアルゴリズム的に優先され、その結果お気に入りのユーザの書いたブログの記事やお気に入りのユーザがRSSフィ

ード登録したブログの記事が優先的に取得できる。

### 3)登録されたRSSフィードを取得し、本システムのDBに保存するクローラ

2)の機能でシステムに登録されたRSSフィードURIを対し、定期的にRSSのチャンネル、ブログ記事の情報をHTTPで取得し、その内容をDBに保存するRSSクローラ。本システムはランキングの計算や検索機能などを行う際、および、ユーザに表示するRSSを取得する際に、DBに保存されたRSSデータを対象に情報処理をする。

### 4)ユーザ間のセンス共感度の大きさを手動設定・管理する機能

ユーザが自分の知人や、2)の「お気に入りブログ作者登録機能」で登録した対象ユーザに対して、自分とのセンス共感度の大きさを手動で設定してDBに保存する機能。本システムは、ユーザとのセンス共感度の大きさに応じてランキングが自動計算され、機能5)によってユーザが受け取るブログ記事の表示順番が変わる。本機能は、自動計算されたセンス共感度の値をユーザが手動設定し上書きすることを可能にする。

### 5)センス共感度の大きさを処理して、ユーザに最適なブログ記事を表示する機能

機能7)のアルゴリズムで計算されたランキング結果と機能4)に基づき、ユーザに最適なブログ記事を順番に表示してゆく機能。ユーザは、本システムにログインしたら最初に、この機能によって最初の順番に表示される(すなわちその時点のランキングが最も上位な)ブログ記事の画面に飛ぶ。(図3参照)。なお、ランキング順の一覧表示も可能である。(図4参照)



図 3 最初に表示されるブログ記事の例



図 4 ブログ記事のランキング画面

## 6) ブログ記事をユーザに評価させる機能

5)の機能と連携し、ユーザに5)の機能で提示されたブログ記事に対し、価値判断、すなわち、ユーザが自分にとって有意義な情報であったか否かを、はい( )、いいえ(×)、どちらでもない( )で評価させる機能と、その情報をリレーショナルデータベースに保存する機能。ユーザはブログ記事の本文を読んだ後、次のブログ記事に進む際、かならず読んだ記事の評価をしないといけない。ブログ記事の評価することで、自動的にランキング順番の次点のブログ記事の本文が表示されるようになる。また、ブログ記事の評価することで、その記事は既読状態としてシステムに記憶されるため、ユーザは、未読記事のみを読む、既読記事を含めて読む、といった二つのモードでブログ記事を読むことができる。本システムは、このように、ブログ記事の評価しながら、次々とブログ記事をランキング順番にしたがって読んでゆく、という利用スタイルのブログ検索・RSSリーダーシステムになる。



図 5 ブログ記事の評価

## 7) ユーザからの記事評価とセンス共感度の手動設定値をもとに、5) のブログ記事を選出するランキングを計算するアルゴリズム

機能5) の評価の累計を元に、ユーザ間のセンス共感度を自動的に算出・更新するアルゴリズム。また、前述のセンス共感度の自動算出の結果と、機能4) の設定値、さらにブログ記事の最終更新時刻に基づいて、ユーザ毎の最適なブログ記事のランキングを算出するアルゴリズム。詳細は5.2で述べる。

## 8) ユーザの利用状況を把握するため管理者用分析ツール

主にユーザの利用状況と、本システムのランキングアルゴリズム(機能7) のチューニングのために必要な情報を収集し、分析するシステム管理者用の分析ツール。ユーザ毎のログイン数、どのページを閲覧したか、どのような評価を行ったか、といった情報に加え、ネットワークの変化を時系列に取得し、可視化する機能などを持つ。

### 5.2 アルゴリズム

本システムのアルゴリズム機能には二つの処理がある。一つ目は、5.1で記述した機能6) の評価の累計をもとに、ユーザ間のセンス共感度を自動的に算出・更新する処理であり、二つ目はセンス共感度の自動算出の結果と、機能4) の設定値、さらにブログ記事の最終更新時刻に基づいて、ユーザ毎の最適なブログ記事のランキングを算出する処理である。一つ目のユーザ間のセンス共感度を自動的に算出・更新する処理では、ユーザが機能6) によってブログ記事の評価を行う度に毎回実行される。機能4) に当該するリンク関係のセンス共感度が手動設定されている場合は計算されない。センス共感度の指定方法が手動から自動に戻されたりする度には、即座に再実行される。

二つ目のブログ記事のランキングの算出処理では、前述の機能によって自動計算されたユーザ間のセンス共感度の値と機能4) で設定された手動設定のセンス共感度の値を用いてユーザ間の距離を計算する。具体的には、ユーザをノード、ユーザ間の知人関係やお気に入り関係をエッジとした有向グラフを形成し、センス共感度をエッジの重みづけ(ノード間の距離)として実現する。そして自ユーザを始点ノードとし、このネットワークグラフの他のノードへの最短経路をDijkstra法によって計算する。また最短経路におけるノード間の距離を合計して、直接のリンク関係のない2クリーク以上のユーザ間のセンス共感度を算出

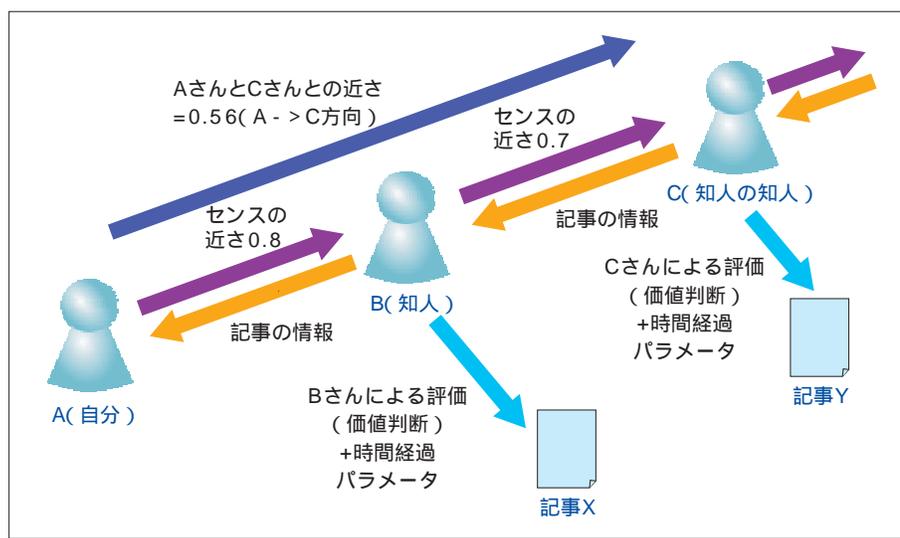


図6 ランキングアルゴリズムの概要図

する。

それぞれのユーザが書いたブログ記事やお気に入りRSS機能に登録しているブログ記事も、ノードとしてそのユーザのノードとの間にエッジをはる。ブログ記事に対して付けられた価値判断や、ブログ記事の最終更新時刻からの経過時間などをパラメータとして(投稿後時間経過の少ない新しい記事ほどランキングが高くなる)エッジの重みづけ(ノード間の距離)を設定する。このことで、自ユーザを始点ノードとして、他のすべてのノードとの最短パスにおける距離を計算し、距離の短い順にランキングを生成する。

## 6. 評価

---

### 6.1 評価項目と評価方法

本システムの評価として、個々のユーザによって異なる価値判断を考慮した情報フィルタリング・情報検索を実証し、情報過多におけるユーザの二つの回避行動が引き起こす問題を、どのように軽減したかを検証した。具体的には、以下の二点を検証した。

- 1) 「画一化された権威的価値基準の再支配」問題を回避しながら多様な価値判断に基づく情報フィルタリングを行う機能
- 2) 「情報取得範囲の限定」問題を回避しながら情報取得範囲を拡大した程度

1)については、個々のユーザの異なった価値判断に従って、異なる情報検索・情報取得が実現できているかどうかを検証した。2)については、ユーザが情報過多を防ぐ手段として、あらかじめ情報収集範囲を設定することなく、従来と比較して広い情報収集範囲から情報取得を実現できていることを示した。また、その前提として、情報収集範囲を広くしても、従来と比較して取得した情報への満足度が低下しないことも必要な条件であり、この点も検証した。

評価実験は、101名のテストユーザによって本システムを1ヶ月間利用し、ユーザの行動を分析した。なお、101名のテストユーザは筆者や共同開発者の知人を中心に集めたユーザで、普段ユーザ自身がブログを書き、かつ、他者のブログを日常的に読んでいるユーザである。一般的なインターネットユーザと比較して、先進的でヘビーユーザである点を特記したい。1ヶ月間の評価実験期間中は、ユーザに対する一斉送信メールなどで、ユーザに本システムの積極的な利用を呼びかけた。また、この間においても機能拡張は行われたが、極力実験結果に影響を与えないバグフィックスや、ユーザからの要望のうち、評価実験のための利用に関して利用率を妨げるような原因に関する要望を中心に扱った。

## 6.2 評価結果

1ヶ月間の実験での利用率は表のようになった。

期間	31日間 (2005/7/27-2005/8/26)
テストユーザ数	101人
1日あたりのPV	2,400
1 ユーザあたりのPV	約24
1日あたりのログインユーザ数	40ユーザ

表1 評価実験における利用率

次に実験結果について述べたい。6.1における評価項目1)に関しては、個々のユーザの異なった価値判断に従って、情報検索・情報取得が実現できているかを検証した。このことは、ユーザが本システムを利用し、異なった価値判断をシステムに記憶していった結果、ある同一のブログ記事が、異なるユーザにおいて、異なった重要度をもつ情報だと判断されていることを示すことによって行った。ある同一のブログ記事が、異なるユーザの検索結果ランキングにおいて、異なった順位で表示されていることを示した。図7は、あるブログ記事である。この同一のブログ記事は、表2のように異なったユーザにおいて、異なった順位で表示されている。



図7 評価項目1) ブログ記事

ユーザ	順位
H	81
K	268
M	334
S	347
N	818

表2 評価項目1) ブログ記事に対するユーザ別の順位

なお、表2に示した順位は、図7に示したブログ記事の各ユーザ別の順位の上位5位までを抜き出したものである。つまり、このランキングによれば、このブログ記事はユーザHにとって最もセンスが近いということになる。評価項目2)に関しては、ユーザが本システムを通して読み、内容の価値判断を行ったブログ記事について、ユーザから知人関係ネットワーク上のクリーク数(Hop数)において、どれだけ離れたユ

ーザが発信したブログ記事かを調べることによって情報取得範囲が拡大していることを示す。このことによって、ユーザが情報過多を防ぐ手段として、あらかじめ情報収集範囲を設定することなく、より広い情報収集範囲から情報取得を実現できていることを示す。

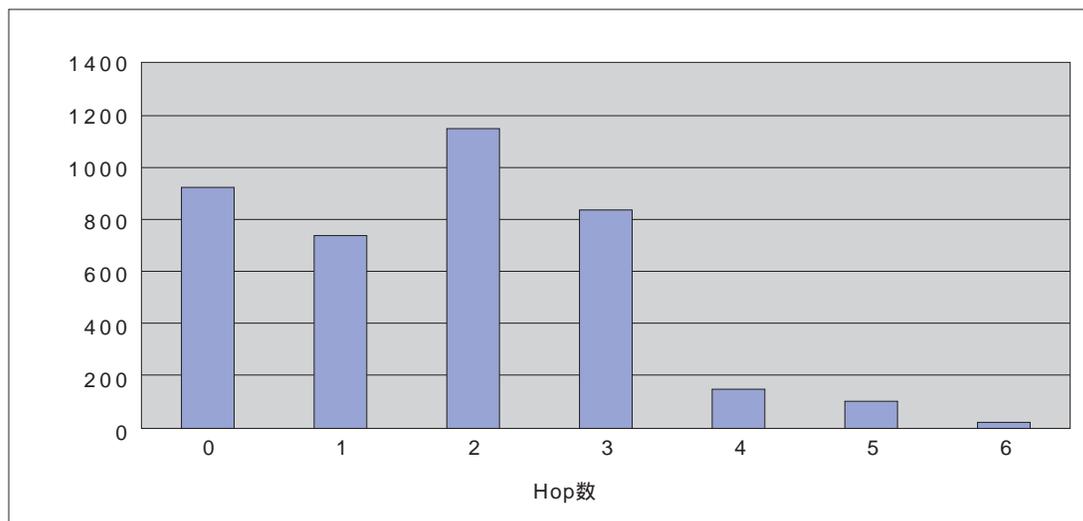


図8 評価項目2) ブログ記事と知人ネットワーク上の距離との関係

図8がその結果である。横軸の「Hop数」が、ユーザが内容の価値判断を行ったブログ記事と、ユーザ自身との知人関係ネットワーク上の距離をクリーク数(Hop数)で表現したものである。縦軸は、各ブログ記事がすべてのユーザに内容の価値判断をされた総数を示している。したがって、図全体は各ブログ記事がどの距離でユーザに価値判断されかを示している。ユーザの視点から言い換えると、自分が内容の価値判断を行ったブログ記事は、自分とどの程度の距離のユーザからやってきたものか、を示している。横軸の「Hop数」が0というものは、ユーザ自身がそのブログを「お気に入りブログ」として登録している、すなわち、通常のRSSリーダーのように自分でそのブログを選択して読んだ場合の数となる。横軸の「Hop数」が1は、知人が書いたブログ記事、ないし、知人がそのブログを「お気に入りブログ」として登録している際に、その記事を読んだ場合の数となる。横軸の「Hop数」が2の場合は、知人の知人となり、3の場合は知人の知人の知人となる。

従来の方法の一つ、通常のRSSリーダーでブログ記事を読む場合は、あらかじめユーザ自身が選択したブログの記事を読むことになる。つまり、横軸の「Hop数」が0の数となる。また、SNS等で知人のブログをのみ読む場合は、横軸の「Hop数」が1の数となる。それらと比較して本システムでは、横軸の「Hop数」が2や3といった距離のブログ記事も数多く読まれ、ユーザに内容の価値判断をされていることが分かる。

このことによって、ユーザは、情報収集範囲をあらかじめ設定することなく、従来と比較して、より広い情報収集範囲から情報取得を実現できていることが示された。さらに、より広い情報収集範囲から情報取得をしたとしても、従来と比較して、取得した情報への満足度が低下しないことを示すために、取得したブログ記事の発信元クリーク数(Hop数)別に、それぞれのブログ記事をユーザがどのように価値判断したかを示した。図9がその結果である。

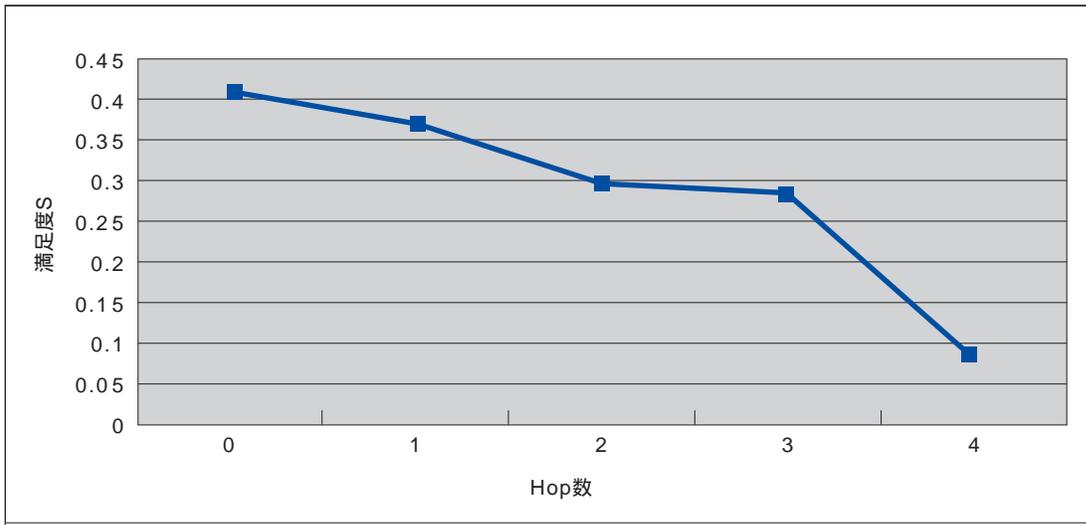


図9 評価項目2) ブログ記事への満足度と知人ネットワーク上の距離との関係

図9の縦軸の満足度Sは、ユーザがブログ記事に対して行った価値判断を、以下の・・・(1)によって数値化した値である。ただし、Egがはい( )評価を行った数、Eb がいいえ( × )評価を行った数、Enがどちらでもない( )評価を行った数である。

$$S = \left( Eg + \frac{En}{4} - \frac{Eb}{2} \right) / \left( Eg + \frac{En}{4} + Eb \right) \dots (1)$$

図9の結果から、Hop数2,3の記事においては、通常のRSSリーダーのように自分でそのブログを選択して読んだ記事(横軸の「Hop数」が0の場合の満足度)や、SNSで知人のブログ記事を読んだときの満足度(横軸の「Hop数」が1の場合の満足度)と比較しても、それほど満足度の低下が起きていないことが示された。上記の結果として、1)「画一化された権威的価値基準の再支配」問題を回避し、多様な価値判断に基づく情報フィルタリングの実現が実現されたこと、2)「情報取得範囲の限定」問題を回避し、情報取得範囲の拡大が実現されたことが示唆された。

### 6.3 評価結果の限界と課題

本評価の限界については以下になるであろう。

#### 1) 評価期間が限定性

今回の評価はユーザ数を出来る限り固定し、かつ高頻度に使ってもらったため、1ヶ月という評価期間ならびに101人という評価規模となり、限定されたものとなった。このことが評価結果の一般性に与える影響は考慮しなければならない。なお、今後はシステムの継続運用を通して継続的に評価を行う。さらに、今回の評価実験によって得られたデータを元に、シミュレーションモデルを構築し、シミュレーションによる評価を行うことを予定している。

#### 2) 評価環境への外部要因の排除

先に述べたように、SNSの知人招待機能によるユーザ数の変化や、実験過程でのバグフィックス、機能

拡張といった外部要因によって評価へ影響が与えている可能性がある。また、開発者からユーザへ利用を呼びかけるメール等を数回送信した点も、本システムが想定する通常の利用形態とは異なるものである。

### 3) 対照実験を行っていない点

情報過多におけるユーザの回避行動が引き起こす問題が低減されたことを検証するには、従来の情報取得の手段と同一条件下において比較することが望ましい。しかしながら、今回の検証内容では同一のほぼ同一のユーザに対してほぼ同一のブログ記事を提示することが困難な上に、従来の情報取得の手段におけるデータを収集することが難しい。したがって、取得したブログ記事の発信元クリーク数(Hop数)が0ないし1の場合を従来の情報取得の手段と見なすことによって評価をおこなった。このことによって、検証結果が限定的なものとなる。なお、本システムにおいて、テストユーザを二つのグループに分け、一つのグループには本システムのアルゴリズムによる検索結果を、もう一つのグループにはランダムにブログ記事を表示させる、といった手法による比較も考えられたが、従来の情報取得の手段がまったくのランダム取得ではないため、その比較にも限定的なものだと考える。

### 4) テストユーザの一般性

先に述べたように、テストユーザは筆者や共同開発者の知人を中心に集めたユーザで、一般的なインターネットユーザと比較して、先進的でヘビーユーザである。したがって、一般的なインターネットユーザが本システムを利用した場合と比較して実験結果が異なる可能性がある。本システムの知見を現実のサービスとして広くユーザに提供する場合を考えると、この点は検討すべき課題である。一方で、サービスとして普及させる際に、最初にコアユーザとして利用するユーザ層は今回のテストユーザのユーザ層であると考えられる。従って、今回の検証も一定の意義がある。

### 5) トピック別のセンス共感度を測定していない点

本項目が今後の最大の課題である。本システムが採用するユーザ間のセンスの共感度を用いる情報検索モデルにおいて、扱う情報のトピックを考慮する必要があるかもしれない。ここでいうトピックとは、流通される情報のセマンティクスのクラスターのことであり、たとえば、映画のトピック、音楽のトピック、グルメのトピックなどがある。ここで問題となるのは、トピック毎によってユーザ間のセンスの共感度は異なる可能性があるということである。たとえば、映画に対する評論に関してセンスの合う2ユーザが、グルメにおいてもセンスが合うかといった点に関して疑問が残る。よってトピック毎に異なるセンスの共感度のネットワークを構築しなければならない可能性がある。トピック毎に異なるネットワークを構築する際は、どのようにトピックを分類するか、といった、セマンティックの扱い方による複雑な問題を引き起こす。たとえば映画というトピックを作ったとする。すると、一部のユーザから邦画と洋画を分けて扱いたいという要望がくるかもしれない。逆のケースも起こりうる。この問題に対して、工学的には人工知能的アプローチとして、セマンティック間の距離を連続値として示し、センスの共感度ネットワーク上のノードとして扱うことで、事実上、トピックの分類を不要にしてしまう方法などが考えられるが、セマンティック間の距離を、計算量の限界やユーザの容易な理解が可能な手段の中で、測定する方法は、今後の研究課題となろう。

現実的には、インターネット上のサービスやコミュニティの多くは、映画、音楽、グルメのトピック別に作られており、一般的な検索ニーズにおいては、同様の分類で、十分にユーザのニーズを満たすことができるのかもしれない。また、必ずしもトピック別にセンス共感度ネットワークを分割しないと、ユーザに満足な検索結果を与えられないかどうかといった点も、厳密には検証されていない。本モデルのような単一のネットワーク形成によって、どの程度のニーズを満足させることができるのかといった点からの検証も必要である。さらに本システムの知見を現実のサービスとして広くユーザに提供する場合には、計算量といった運用の際のコストを踏まえた単純さ、実現可能性、ならびに、ユーザの容易な理解と利用となど

を踏まえ、バランスのとれた実現が不可欠である。なお、今回は4)で述べたように、テストユーザ層が先進的でヘビーなインターネットユーザであったため、システム上でやりとりされた情報の多くが、インターネット技術に関する話題や、最新のインターネットサービスに関する話題であった。4)で述べたように、ユーザ層の拡大、一般化に従って、本項目の課題の重要度が明らかになってくると考えられる。

## 7. 結論

---

本研究では、個々のユーザによって異なる主観的な価値判断と他者への価値判断の委譲を情報検索に応用するパーソナライズ化された情報検索・流通システムを構築して実証した。そのことを通してインターネットにおける情報検索・情報取得において、情報過多を原因として起きるユーザの各種回避行動の軽減を試みた。具体的には、<sup>1)</sup>権威や高いPage Viewといった人気のある一部のサイトの情報しか受け取らない、あるいは、Google Page Rank、という具合に、画一的な評価軸に従って取得する情報を絞る行動(「画一化された権威的価値基準の再支配」問題)<sup>2)</sup>興味のあるいくつかのMLに参加して、その情報だけを受け取る。あるいは、SNS等を使って知人のブログしか読まない。という具合に、あらかじめ取得する情報の範囲を設定し、その範囲外の情報は処理をしない(「情報取得範囲の限定」問題)といった行動の軽減を試みた。運用実験による検証によって、いくつかの課題と限界が明らかになったが、そのもとで、上記2点のユーザの各種回避行動の軽減が示唆された。

## 8. 今後の展望

---

本研究は、「画一化された権威的価値基準の再支配」問題や「情報取得範囲の限定」問題といった情報過多に対してわれわれが日々とっている回避行動のデメリットを克服し、インターネットの理想の情報流通へ一歩前進することを目指して始まった。今回の検証した点はその目標と比較して非常に限定されたものである。特に「情報取得範囲の限定」問題に関しては、予め情報の取得の範囲を限定しないこと、という要件を満たしたに過ぎない。この要件を技術的に解決したことが、3.3で「情報取得範囲の限定」問題を定義した際に述べた問題の一つである「タコツボ化」、すなわち、多様な価値観に触れる機会が低下し価値観が画一化されてしまうリスクを、低減することにつながるのか、あるいは、かえって増大させるのではないか、本研究を進めるうちにそのような重要な議論点が浮かび上がってきた。主観的や共感と、多様な価値観との関係は如何なるものか、というテーマである。

この点は、ユーザー一人一人の主観的な価値判断や、価値判断の他者への委譲に際して、各ユーザがどのような思考のもとでそれらを行うかによって大きく左右されることから、システム的设计によって一概に決定されるものではないだろう。また本システムの使い方によっても結果は大きく異なるものになるであろう。たとえば、現状のシステムでは、センス共感度ネットワークはランキングの生成の手段に過ぎない。すなわち、ユーザにとっての効用、すなわち、パーソナライズ検索を実現するための手段に過ぎないわけであるが、このセンス共感度ネットワークをユーザに効果的に可視化し提示することによって、センスの共感度が高くないユーザの情報に触れる機会を提示することができるかもしれない。インターネットの発展と共に、たびたび指摘される「タコツボ化」、価値観の画一化の懸念に関して、議論・考察を進めていく上で、本システムが実際のデータ、事例を提供することができれば幸いである。また筆者自身も今後は「タコツボ化」、価値観の画一化の懸念に関して考察、研究を進めていきたい。

## 謝辞

本研究におけるシステムSTNは、独立行政法人情報処理推進機構による未踏ソフトウェア創造事業の支援を受けて開発されたものである。共同開発者の佐藤毅氏、ならびに、関係者に謝意を表す。また、株式会社サルガッソーの鈴木健氏には、アルゴリズム設計の議論や実装に関してご尽力いただいた。産業技術総合研究所の濱崎雅弘氏、アーティキュレート株式会社の鈴木貴晶氏には主に評価に関してご尽力いただいた。この場を借りて謝意を表したい。

---

[ 参考文献 ]

- [ 1 ] Lawrence Page, Sergey Brin, Rajeev Motwani, Terry Winograd, 'The PageRank Citation Ranking: Bringing Order to the Web', 1998.
- [ 2 ] Taher H. Haveliwala, 'Efficient Computation of PageRank', Stanford Technical Report, 1999.
- [ 3 ] ハーバート・A. サイモン『システムの科学』パーソナルメディア、1999年
- [ 4 ] Google. <http://www.google.com/>
- [ 5 ] Yahoo!. <http://www.yahoo.com/>
- [ 6 ] 大向一輝「SNSの現在と展望 -コミュニケーションツールから情報流通の基盤へ-」情報処理、 Vol.47、 No.9、 pp.993-1000、 2006年
- [ 7 ] 丸山眞男『日本の思想』岩波書店、1961年
- [ 8 ] Mark S. Granovetter, 'Getting a Job: A Study in Contacts and Careers', University of Chicago, 1995年