

# 不均質な情報源間での情報共有支援

濱崎 雅弘

博士 (情報学)

総合研究大学院大学  
複合科学研究科  
情報学専攻

平成 16 年度  
(2004)

2005 年 3 月

本論文は総合研究大学院大学複合科学研究科情報学専攻に  
博士(情報学)授与の要件として提出した博士論文である。

審査委員：

武田 英明 (主査)

相澤 彰子

角 康之 京都大学

中小路 久美代 東京大学

山田 誠二

(主査以外はアルファベット順)

INFORMATION SHARING AMONG HETEROGENEOUS  
INFORMATION RESOURCES

Masahiro Hamasaki

DOCTOR OF  
PHILOSOPHY

Department of Informatics,  
School of Multidisciplinary Sciences,  
The Graduate University for Advanced Studies (SOKENDAI)

March, 2005

A dissertation submitted to  
the Department of Informatics,  
School of Multidisciplinary Sciences,  
The Graduate University for Advanced Studies (SOKENDAI)  
in partial fulfillment of the requirements for  
the degree of Doctor of Philosophy

Advisory Committee:

Hideaki Takeda (Chair)

Akiko Aizawa

Kumiyo Nakakoji Tokyo University

Yasuyuki Sumi Kyoto University

Seiji Yamada

(Alphabet order of last name except chair)

## 内容梗概

本論文は多種多様な情報源が存在する World Wide Web（以下，Web）における情報収集を支援する技術およびシステムを提案するものである。

近年，ネットワーク技術の発達と普及に伴い，オンライン上に膨大な情報が蓄積されるようになった．それらの情報の多くはネットワークに接続さえすれば誰でも容易にアクセス可能であり，私たちは日々溢れんばかりの情報に囲まれている状態にある．そのような傾向は Web において特に顕著に見られ，その結果，多くの情報の中から必要なものを取り出すのが困難である「情報過多」と呼ばれる状況になっている．この問題を解決するために様々な情報検索や情報推薦技術が提案されているが，Web 上での情報発信量は増す一方で情報過多問題を解決するには至っていない．

そこで本研究では，Web 上の情報源の多くが Web 上の情報を収集し同時に発信もしている「情報収集発信源」である点に着目し，情報収集支援の対象となる利用者が一つの情報源となり，そのような情報源間での情報共有を支援することによって情報収集の支援を行うというアプローチを取る．これは情報収集の支援に積極的に他の情報収集者の知識を利用するアプローチであり，品質にばらつきがあるために内容分類だけでなく価値判断が重要な Web 情報には適していると考えられる．

本論文は6章からなるが，大きく分けると2部からなる．前半では，どのようにして互いに異なる概念体系を持つ不均質な情報源間での情報共有を実現すればよいかということについて述べる．後半では，そのような不均質な情報源が複数存在し，ネットワーク状に情報共有を行う場合において，どのようにして適切な共有相手を発見するか，また，そのようなネットワーク構造を用いた情報共有はどのような特性を持つのかについて述べる．

1章では，本論文の目的と問題の分析，および本研究のアプローチを述べる．この章において，本論文での情報収集支援とは情報を収集する主体的な存在の間での情報共有を支援することによって為されることを示す．その上で本研究で取り組むべき課題として情報共有の問題点を示し，それに関係する既存の研究を概観して位置づけることにより，本論文が扱う領域について明確化を行う．章の最後

に本論文の構成を述べる．

2章では，異なる概念体系を持つ不均質な情報源間での情報共有の方法について議論する．これは情報共有を行う情報源をそれぞれどのように表現するかという問題と密接に関係する．本研究では，情報源をそれぞれ概念階層で表現し，共有対象となる情報はその概念階層の中に格納されているものとした．この手法の有効性を被験者を用いた実験により検証する．

3章では，4章で概念階層を用いた情報共有の有効性が示されたのを踏まえ，類似する概念階層の発見手法について議論する．概念階層内の情報を利用した類似概念階層の発見手法はいくつか提案されているが，本研究では概念階層内の情報の内容的類似性と，概念階層の構造自体の特徴を用いた類似概念階層発見システム *WebHical* を提案する．この手法の有効性を実データを用いた実験により検証する．

4章では，情報共有の相手をどのようにして見つけるかという点について議論する．より良い情報共有を実現するためには共有相手の選択が重要である．しかし，一般に膨大な数の共有相手となる情報源の中から適切なものを見つけるのは困難である．そこで本研究では，友人に友人の友人を紹介してもらおうという私たちが日常生活において行っている新しい知り合いの獲得手法を模倣した，近傍仲介法を提案する．近傍仲介法はボトムアップな共有相手発見手法であり，トップダウンなマッチングシステムとは異なった特徴を持つ．この手法の有効性をシミュレーション実験により検証する．

5章では，情報源である人がもともと知り合いである相手をシステムに登録し情報共有を行った場合にどのような振る舞いが見られるか，また，そのような情報源のネットワーク（情報源が人に限定されるのであればパーソナルネットワーク）が，情報共有にどのように応用することができるか，これら点について実際にシステムを運用し，その利用ログの分析から検証する．

6章では，本論文の結論と展望を述べる．

## Abstract

In recent years, a large amount of information becomes available on computer networks such as World Wide Web. In World Wide Web, a lot of information is provided by a lot of information resources. The number of information resources is overwhelming, but the variety of information resources is also astonishing. Information resources varies in a wide range, i.e., from public database services to personal web pages.

As a result, we bother to find important information from such various and massive information resources. In this research, we model an information resource on the web "Information collector/publisher" as not only gathering information on the web not just as an information publishing entity but an entity working both information publishing and information gathering. Such information entities are connected to each other in various levels, e.g. connected explicitly by links and human network of authors. Information gathering should be understood as an activity in such a community of information entities.

We aim in this research to propose models and methods to enhance information sharing among people. Human is the most rich and flexible information entity that can offer and gather information simultaneously. In order to achieve the goal, we set two topics. One is to identify relationship between information entities. In this topic, we first investigate relationship among people by analyzing WWW bookmarks as topic sharing, and showed that hierarchical structure is important to share topics among people. We also discuss how to identify relationship between large hierarchical information such as Yahoo! Internet Directory. The other is to handle network composed with information entities and their relations. We propose a method to re-configure network to improve network structure. We also build a practical system in which human network is gathered and used in services for participants of conferences.

In Chapter 1, we discuss about information sharing and show two important points of information sharing especially among heterogeneous information entities

and show four researches about these two points in later chapters.

In Chapter 2, we discuss and propose a method to help information sharing among heterogeneous information entities. We propose a system called *kMedia* that can assist users to form knowledge for community by showing shared topics networks (STN) among them. *kMedia* uses WWW bookmarks as information entities. We conducted an experiment to know how *kMedia* can support users. One result is that folder recommendation is more effective than page recommendation. The other is that recommendation is more effective for people belonging to the same real communities than those to different communities. According to these result, we propose a new measurement called "category resemblance" that is recommendation measurement based on resemblance of folder structures. This measurement shows higher than all other system generated parameters and human evaluation to detect human relationship.

In Chapter 3, we propose a method called *WebHical*. This method is for aligning information from one information entities with hierarchical structure to another. It is based on *kMedia* method and *Hical* method. We adopted the  $\kappa$  statistic method and the SMART algorithm to measure the similarity among hierarchical structures. We construct a system to evaluate the performance of our method. The results of this experiment reveal that the proposed method can be used to help sharing information among information entities with different hierarchical structure.

In Chapter 4, we propose an algorithm called *Neighborhood Matchmaker Method* to optimize networks of information entities. Interpersonal network is one of a network of information entities and it is useful in various utilization of information like information gathering. However it is usually formed locally and often independently. In order to adapt various needs for information utilization, it is necessary to extend and optimize it. Using the neighborhood matchmaker method, we can increase a new friend who is expected to share interests via all own neighborhoods on the interpersonal network. Iteration of matchmaking is



used to optimize interpersonal networks. We simulate the Neighborhood Matchmaker Method with the practical data and the random data and compare the results by our method with those by the central server model. The neighborhood matchmaker method can reach almost the same results obtained by the sever model with each type of data.

In Chapter 5, we discuss importance and utilization of interpersonal network in a community system through the result of management and analysis of the scheduling support system for academic conferences. The important feature of the system is generation and utilization of interpersonal network to support information exchanging and information discovery among participants. We applied this system to the academic conference called JSAI2003. We obtained 276 users and their interpersonal networks. We found not only that a lot of participants enjoyed to form interpersonal networks but also that the formed network was useful for them in information browsing and recommendation.

Finally, we conclude this paper in Chapter 6.

# 目次

第 1 章	序言	8
1.1	研究の背景と目的	9
1.2	現状の分析	9
1.3	本研究のアプローチ	10
1.4	本研究の課題と関連研究	11
1.5	本論文の構成	18
第 2 章	共通話題ネットワークを用いた情報共有	21
2.1	はじめに	22
2.2	ネットワーク上の個人としての Web ブックマーク	22
2.3	共通話題ネットワーク	23
2.4	kMedia システム	24
2.5	実験概要	27
2.6	ページとフォルダに関する共通性の分析	31
2.7	人との距離に関する分析	34
2.8	議論	37
2.9	まとめ	41
第 3 章	階層的知識と内容的類似性を用いた異なる概念体系の統合	42
3.1	はじめに	43
3.2	階層的情報源	43
3.3	階層的情報源の統合手法	45
3.4	システム概要	47
3.5	実験	49
3.6	Hical-NB	52
3.7	議論	54
3.8	まとめ	55

第4章	仲介による共有ネットワークの拡張	56
4.1	はじめに	57
4.2	情報交換相手との関係発見	57
4.3	近傍仲介法	58
4.4	実験概要	59
4.5	実験	63
4.6	考察	72
4.7	議論	74
4.8	まとめ	77
第5章	パーソナルネットワークを用いたコミュニティ支援システム	78
5.1	はじめに	79
5.2	スケジューリング支援システム	79
5.3	運用結果の分析	92
5.4	考察	106
5.5	スケジューリング支援システム for JSAI2004	109
5.6	議論	109
5.7	まとめ	111
第6章	結言	113
6.1	結論	113
6.2	課題と今後の展望	115
研究業績		125
付録 A	スケジューリング支援システム for JSAI2004	130

# 目次

1.1	本論文の構成	19
2.1	共通話題による人のつながり	23
2.2	<i>kMedia</i> のユーザインタフェース	25
2.3	共通話題ウィンドウ	25
2.4	共通話題の発見	26
2.5	被験者のグループ分け	28
2.6	評価フォーム 1	30
2.7	評価フォーム 2	30
2.8	フォルダ推薦の有効性	31
2.9	ページ関連度と推薦ページ評価	32
2.10	ページ・フォルダ推薦数	33
2.11	ページ関連度とフォルダ関連度	34
2.12	分類方法の類似	35
2.13	ページ推薦におけるカテゴリ近似度の有効性	37
3.1	階層的情報源のモデル	44
3.2	複数の階層的情報源における問題点	45
3.3	システム概要	48
3.4	内容的類似性による分類の正答率	50
3.5	Hical では適切な分類が困難な例	53
4.1	ノードの振る舞い	59
4.2	シミュレーションの流れ	61
4.3	パスの張り替え	61
4.4	各状態のネットワーク図	62
4.5	被覆率の変化 (仮想データ)	64
4.6	到達率の変化 (仮想データ)	65

4.7	ノード数とパス数と平均被覆率	66
4.8	ノード数とパス数と平均到達率	67
4.9	平均収束ターン数	68
4.10	被覆率の変化(実データ)	69
4.11	収束解と最適解の次数分布	71
5.1	システム構成	80
5.2	Session リソースの HTML ページ例	82
5.3	Paper リソースの HTML ページ例	82
5.4	Person リソースの HTML ページ例	83
5.5	スケジュール表の HTML ページ例	83
5.6	マイページ(人のページ)	84
5.7	マイスケジュール(スケジュール表のページ)	85
5.8	聴講予定の発表論文のページ	86
5.9	論文ページの見え方	87
5.10	人ページの見え方	88
5.11	推薦サービス画面	90
5.12	発表ごとに設置された掲示板	91
5.13	パーソナルメッセージ	92
5.14	1日あたりのログインユーザ数とユーザ数推移	94
5.15	リンク数の推移	94
5.16	登録方法によるリンク状況の違い	96
5.17	被 Check リンク数と被 Know リンク数	98
5.18	Know リンクネットワーク	99
5.19	共著関係ネットワーク	99
5.20	共著関係と Know リンクのネットワーク(一部)	100
5.21	被 Know リンク数の両対数グラフ	101
5.22	アクセス遷移図	102
5.23	推薦結果	103
5.24	推薦手法と被リンク数の関係	104
5.25	利用者とその知り合いが持つ Check リンクの関係	105
6.1	利用者数の推移	134
6.2	一日あたりの PV 数とアクセス者数	135
6.3	Know リンクの累積分布	137

---

6.4 被 Know リンクの累積分布 . . . . . 137

# 表目次

2.1	被験者のブックマークに登録されたフォルダおよび URL の数 . . . . .	29
2.2	推薦者の評価との相関 (1) . . . . .	36
2.3	推薦者の評価との相関 (2) . . . . .	36
2.4	推薦者の評価との相関 (3) . . . . .	36
3.1	インスタンス分割表 . . . . .	46
3.2	実験に用いたデータ . . . . .	50
3.3	内容的類似性の導入で新たに発見された類似階層ペア . . . . .	51
3.4	発見された類似階層ペア数 . . . . .	51
4.1	パラメータ設定 (仮想データ) . . . . .	63
4.2	パラメータ設定 (実データ) . . . . .	69
4.3	孤立したノードの数 . . . . .	70
4.4	初期状態のパス数と収束状態のパス数の相関係数 . . . . .	72
5.1	初期リソース数 . . . . .	93
5.2	初期リンク数 . . . . .	93
5.3	終了時リソース数 . . . . .	93
5.4	終了時リンク数 . . . . .	93
5.5	告知活動 . . . . .	94
5.6	Check リンクと論文リソース . . . . .	97
5.7	Know リンクとリソース . . . . .	97
6.1	基本データの比較 . . . . .	133
6.2	ユーザ数とイベント . . . . .	134
6.3	追加された Check リンクの比較 . . . . .	136
6.4	追加された Know リンクの比較 . . . . .	136
6.5	Know リンクを追加した人の比較 . . . . .	136

---

6.6	各ページのページビュー数 . . . . .	137
6.7	人のページへの移動経路 . . . . .	138
6.8	論文のページへの移動経路 . . . . .	138
6.9	推薦結果 . . . . .	138



# 第1章

## 序言

本章では、本研究の背景と現状の分析について述べ、その上で本論文の目的と研究の基本方針について述べる。その上で本研究で取り組むべき課題として情報共有の問題点を示し、それに関係する既存の研究を概観して位置づけることにより、本論文が扱う領域について明確化を行う。そして最後に本論文の構成を述べる。

## 1.1 研究の背景と目的

World Wide Web (以下, WWW または Web) は急速に社会に浸透しており, すでに私たちの日常生活や業務における情報収集において欠かせない存在となっている. 代表的な Web 検索サービスである Google は 2005 年 1 月の時点で 80 億もの Web ページを収集していると報告しており, WWW 上の正確な情報量はもはや誰も測ることはできないが, WWW が世界最大規模のオンラインデータベースであることは疑う余地はない.

この巨大なオンラインデータベースは, 多くの多様な情報提供源によって支えられている. 有名企業の多くは自社 Web サイトを公開し, Web サイトを通じて業務内容や製品情報を発信している. また, 新聞社や出版社も紙メディアと平行して, ほぼ同レベルのコンテンツを Web サイトを通して提供している.

WWW における情報提供源は企業やマスメディアだけではない. 個人も重要な情報提供源である. 著名人はもちろんのこと, 一般人でも良質のコンテンツを提供する人は注目を集める. また, 一般的に見て価値があるかどうか判断しかねる様なコンテンツ (例えば個人的な日記や, 極めてニッチな技術情報など) を配信する情報提供源も, ある限られた範囲の人々 (提供者の友人や, 提供者と同じ境遇にある人たち等) にとっては, 極めて価値のある情報提供源であったりする. 特に最近では Weblog と呼ばれる個人向けの Content Management System が流行しつつあり, 個人による情報発信はよりいっそう増加するものと予想される.

このように WWW には多様な情報提供源が存在し, それらが様々な思惑で, 様々な内容・性質の情報を発信している. この多種多様な情報源の存在こそが, 未だかつて無いほどの膨大かつ多様な情報を持ち続けるという WWW の特徴を実現している. しかし同時に, このような特性は大量の価値ある情報とそれよりも遙かに膨大な屑情報が混じり合った状態を作り出しており, 情報が多すぎて欲しい情報が見つからない「情報過多」と呼ばれる問題も生み出している.

本研究では, このような多種多様な情報源によって作られた膨大かつ不均質な情報があふれかえった環境における情報収集を支援する技術の提案および検証を行う.

## 1.2 現状の分析

WWW はその性質上, 雑多な情報を膨大に集める特性を持つ [63]. このような問題に対する対策の代表的なものとして情報検索があげられる.

情報検索は現時点においてもっとも効果を発揮している Web 情報収集支援技術と言え, 実際に多くの Web 検索サービスが提供されている. 各 Web 検索サービスはそれぞれ独

自技術を用いているものの、基本的には、WWW から Web ページを機械的に収集してデータベースに蓄積し、利用者からのクエリーに対してデータベース内のアイテムをレイティング（評価）して高いものを提示する、という構成を持つ。

Web 情報に対する情報検索技術は様々な研究がなされており、ここ数年の Web 検索サービスの高性能化は目を見張る物がある。しかしながら、「何か調べ物があるから WWW を使う」ではなく「WWW から何か面白い情報を見つける」という使い方をする場合や、収集可能な膨大な情報の中から不要な物を取り除くといった場合などを考えると、情報検索技術だけでは情報収集支援としては不十分であることがわかる。

そのような膨大な情報のなから必要な物だけを取り出す技術として、情報フィルタリング [11] や協調フィルタリング [27] が挙げられる。

情報フィルタリングは利用者のプロフィールと類似する情報だけを抽出するシステムである。膨大な情報を処理するのに向いているが、そもそもプロフィールとの類似しか判別できないので、屑情報が多い Web コンテンツには不向きである。

協調フィルタリングは複数利用者のプロフィール（コンテンツに対する評価情報）を収集し、利用者に類似した他の利用者（Neighbors）を発見し、その人のプロフィールを利用して情報の推薦を行う。協調フィルタリングは人の評価情報を用いているため、推薦される情報の質はある程度高いものと考えられる。しかし、類似した他の利用者を捜すには利用者間での評価情報の重複が必要であるが、Web コンテンツは膨大であるため重複する割合が少なく、類似する他の利用者が発見困難であるという問題がある。また、協調フィルタリングでは各利用者が積極的に新しい情報を評価していくこと、さらに正しく評価することが必要である（信頼性の問題）。

### 1.3 本研究のアプローチ

本研究では、不均質な情報が膨大に存在する WWW 環境における情報収集支援として、利用者自身が情報源となり、他の情報源との情報共有によって効率よく情報を収集する環境の構築を目指す。これは協調フィルタリングにおける基本的な考え方であるが、本研究では特に個人（個々の情報源）間での情報共有に注目する。

一般に、協調フィルタリングでは個人が作成した評価情報をシステムが一手に収集し、その中から各利用者ごとに類似した利用者達を見つけ出し、それらの評価情報を元にアイテムを推薦する。個人が作る評価情報はあまり多くないため、類似した評価情報を見つけるのは難しい。それゆえにたくさんの利用者から評価情報を集めるという戦略を取る。しかし、Web の様に情報が膨大な場合はどれだけ集めても十分とは言い難いし、また、多くの利用者を集めた場合には評価情報の信頼性や、ただ乗り（新しいアイテムを評価せず、推薦されたアイテムしか評価しない）といった問題が現れてくる。

そこで、本研究では個人がそれぞれめいめいの相手と情報共有を行い、その共有関係の連なりによって間接的に巨大な情報共有関係を作成し、それによる協調フィルタリングの実現を図る。この方法の利点は、個人が選択した相手と情報共有を行うため、評価情報の信頼性やただ乗りの問題がなくなる。しかし逆に、共有相手を自分の手が届く範囲に限定することにより、相手は必ずしも自分にとって最適な共有相手とは限らないため、効率よく情報共有を行う必要があるし、また、より良い共有相手を探す行為が必要である。

本研究の目的とアプローチを整理し直すと、まず目的は不均質な情報が膨大に存在する WWW 環境における情報収集支援を行うことである。次に、そのアプローチとして個々の情報源間（個人間）の情報共有を用いる。このような個人ネットワークを用いた情報共有システムは吉田 [73] や竹内 [65] らも提案しているが、本研究では特にそのような状況における、個々の情報源間の情報共有をいかにして行うかと、個々の情報源によって構成される情報共有ネットワークをいかにして改善するかという基本的な問いに対して検討し、これらの問題を解決する技術の提案および検証を行う。

## 1.4 本研究の課題と関連研究

### 1.4.1 本研究の課題

本章では、まず本研究が取り組むべき課題についての説明を行い、次に関連研究を挙げながら本研究の取り組みを明確にする。

本研究では、情報収集支援のために、情報を収集し発信する個人がそれぞれ互いに情報共有するというアプローチをとる。そこで、ここでまず本研究が意図するような情報共有が持つ本質的な課題について議論し、次に本研究が取り組むべき課題の説明とそれに対する解決案を示す。

情報共有とは他者が収集した情報を互いに取得可能にすることである。そして情報共有による情報収集支援とは、利用者が欲しいと思うような情報を情報共有相手が先に収集してくれることで、情報共有によりわざわざ自ら膨大な情報の山の中からその情報を見つけ出す手間が省けることが期待される。

よって、本研究が意図する情報共有においては、情報共有相手が収集してくる情報の少なくともいくつかは自分にとって興味のある情報でなくてはならないと言える。自分にとって関心のない情報をただ見せられても仕方がなく、ただ皆が集めた情報を集積するだけでは情報過多の繰り返しになる。つまり、いかにして他人の知識および他人が収集した情報を再利用するかが大きな問題になる。

一番単純な情報共有は、皆が集めた情報を一カ所に集めて誰でもアクセス可能にすることである。Windows の共有フォルダや UNIX の Samba などによって行われているファ

イル共有がそれに該当する．共有相手がアクセス可能な場所に共有対象となる情報（ファイル共有では，ファイルに該当するもの）をアップロードすることで共有が行われる．利用者は共有場所へアクセスし，そこから自分に必要な情報を取得する．

このような情報共有によって情報収集支援が行われるためには，情報共有を行うメンバー間でどのような情報を共有するのか意思疎通できてなくてはならない．もし，お互いの意思疎通ができておらず，皆勝手に思い思いの情報を共有スペースに置いた場合，最悪の場合，自分が置いたもの以外は興味のない情報しか共有スペースに置かれていない可能性もある．つまり，共有メンバーは互いに興味を共有しているような情報を共有対象として扱うようにしなくてはならない．

よって，共通する興味を全く持っていない個人同士では効果的な情報共有は難しい．これはつまり，より良い情報共有を実現するためには共有相手の評価および選択が必要であることを示唆している．また，個人は複数の興味を持っていることが考えられるので，個人間で共通している興味の同定が必要になる．例えば，様々な興味を持っている個人が複数集まって無造作に情報共有を行った場合，共有スペースに置かれた情報オブジェクトのうち，どれが自分の興味に合致しているものかを探すのにまた手間がかかってしまい，共有相手や共有している情報オブジェクトが増加してくると，また情報過多に近いような問題が生じてしまう．

ここで問題を整理する．情報共有を行う際には，以下の二つの課題があると考えられる．そしてこれらは同時に，本研究にて取り組むべき課題であると言える．

- 共有相手との間で共通している興味の同定
- 共有相手の選択

以下の節ではそれぞれの問題に関する先行研究について述べ，さらに本研究でのアプローチについて説明する．

#### 1.4.2 共通している興味の同定

共通している興味の同定は，すなわち個人が持つ知識をどのように表現し，また，その個人が持つ知識の関係性をどのように表現するかという問題と言える．このような問題については，情報共有システムやコミュニティ支援システムなどで議論されている．

人と人の関係は，コミュニティを形成する上で非常に重要な要素の一つである．では，コミュニティを形成するためには，人のどのような関係を見つけだせば良いのか，という事が問題になる．

発見すべき人の関係は，抽象過ぎず，かといって特化しすぎてもいけない，しかるべきレベルで記述されなければならないと考えられる．例えば，ネットニュース [39] や Web

ページ [7, 67], 音楽 [56], といったものを扱った多くの推薦システムでは, 情報のオブジェクト自身を共有しているかどうかという事を利用している. しかし, このようなオブジェクトを通じた関係性は, オブジェクトに関する情報を共有するかどうかを決めるのには有効であるが, きわめて個別的なので人と人の関係を理解するのに適しているとはいえない. ReferralWeb[36] ではリンクや著書における引用関係から, 人の関係を発見しようとしている. これは比較的有効な手法と思われるが, 誰にでも適用できる訳ではないという問題がある.

一方, Beehive[28] や Babble[19], Visual Who[17] といったコミュニティ支援システムは, より簡単な関係性を用いている. 例えば, Beehive では単純に e-mail でのコミュニケーションを観察してアクティブなメンバーかどうかを判断している. Babble では2次元空間にユーザの関係を表示する. Visual Who も同様に2次元空間を利用するが, より動的である. このような可視化は直感的で理解し易い反面, 関係性が抽象的になりすぎるという傾向がある. そこで本研究では, 各ユーザが持つ話題に着目し, 話題の共通性によってユーザ間の関係を表現することを試みる. 話題は直感的で理解が容易であると同時に, どのような関係であるかを明示化することが可能であると思われる.

しかし, ユーザの持つ話題の同定は一般に容易ではない. これまでユーザの関心を発見する方法として主に二つの方法が用いられていた. 一つは強化学習やベイジアンネットワークのような機械学習を用いた方法である. 例えば, WebWatcher[33] や Letizia[41] などは Web ブラウジングを対象としたシステムである. この手法の長所であり短所である点は持続した関心をくみ取れる点である. 問題となるのは, ユーザの関心が増加する点を見つけることができないので, 関心を個別に取り出すことが難しい点である. また, 学習したデータはユーザにとってわかりやすいものではないため, ユーザがどういった関心を持っているかを明示化するのも困難である. もう一つの方法は, 階層分類のような分類技術である. 例えば, Scatter/Gather[15] や Webmate[12] などに用いられている. これはユーザの関心をわかりやすく表示する. しかし各分類の意味は明示的ではなく, 理解に努力が必要である.

二つの手法には, 以上のようにそれぞれ問題点が存在している. さらに共通する問題は, 人の持つ話題を同定する方法としてテキスト解析を用いることの限界である. 例えば, ユーザが動物のページを収集していた場合, ページには象や猿などの語は多く現れるだろうが, ユーザの関心を示す動物という語はなかなかあらわれない. このように, 話題を示すような重要な語は出現しない場合がある. というのも「象は動物である」といった一般知識となっている事柄は, 多くはページには現れないからである.

以上のようなことから, 本研究ではユーザの関心を計算により導き出す方法ではなく, ユーザ自身の知識を利用する方法を採ることにした. 具体的には, ユーザは自身の関心を階層構造によって表現する. そしてユーザは興味のある情報を, それが含まれる関心に概

念階層に格納する．このようなイメージに一番近いのは Web ブックマークである．Web ブックマークのフォルダ名とフォルダの中にあるページはユーザが持つ内在的な知識・関心を外在化しようと努力した結果である．例えば，フォルダの名前はユーザの関心を示すものであるし，フォルダ中には，ユーザがその話題に属すると考えたページが入っている．Web ブックマークはツリー構造であり，単一の階層構造に限定されているが，本研究が対象とするような WWW 上の情報に限定した世界でのユーザの知識を示すものとしては十分であると考えられる．以上をふまえ，情報源の一つの例として Web ブックマークを用いた情報共有の実験を行った．これについては，2 章にて述べる．

興味を表現する方法を階層構造に限定した場合，共通している興味を持っているかどうかは，異なる階層構造（を持つ情報源）間で類似している概念階層を発見する事でわかる．もし仮に，概念階層の特徴量を表す事ができ，その特徴量の距離計算が可能であれば，類似する概念階層の発見は容易に行える．しかし，そのような特徴量を求める事は一般に難しい．

例えば概念階層内の情報を特徴ベクトル化し，それらの和や積などを概念階層自体の特徴ベクトルとする方法が考えられるが，どのような演算をすれば適切な概念階層の特徴ベクトルを作れるかというのは非常に難しい問題である．これは，概念階層内の情報の特徴を示す要素が，必ずしも概念階層の特徴を示す要素とは限らないという点に起因すると考えられる．

そこで，本研究では概念階層内の情報間での類似性を求め，それを元に概念階層の類似性を求めるアプローチを取る．本論文の 2 章で用いた Web ブックマーク共有システム *kMedia* では [78]，Web ブックマークのフォルダ（＝概念階層）内の Web ページ（＝情報）間の類似性をキーワードベクトルから求め，一定数以上の類似性のある Web ページを複数持つフォルダ間に類似性があると判断する．同じく Web ブックマーク共有システム *Siteseer* もフォルダ間の類似性を発見するが，これは同一の URL がいくつあるかで判別している．市瀬らの Hical システム [30] は *Siteseer* と同じく同一の URL を手がかりとするが，さらに加えて概念階層が持つ階層構造自体も利用する事で，より複雑な階層構造においても高い精度で類似する概念階層の発見を実現している．しかし，これらは両方同一の情報のある程度持っている事が必要であり，量が膨大である Web コンテンツを対象とした個人ベースの情報共有システムでは，この制約は大きな負荷となる．そこで本研究では新しい概念階層の類似性を発見するアルゴリズムを提案する．これについては，3 章にて述べる．

### 1.4.3 共有相手の選択

情報共有は、その形態の違いにより2種類に分けられる。一つはメールのような一対一型、もう一つはメーリングリストや掲示板のような一対多型である。我々がメールを使うのは大抵の場合、すでに友達であるかもしくはホームページか何かで見知った相手と情報共有をするときである。メールを用いた情報共有では、相手の選択や理解は比較的容易である。対してメーリングリストや掲示板では、情報共有の場を介して多くの人同士で情報共有が行なわれる。この場合、我々は情報共有の新しい相手を比較的容易に獲得することができる。

前者の問題点は、新しい情報共有相手の獲得が困難であることである。このタイプでは、既に知っている相手としか情報共有が行なえない。一般に、人が普通に知りうる人の数には限界があるため、情報共有にとって良い環境を築くのは難しい。対して後者は、新しい情報共有相手の選択が困難であるという問題点がある。多くの参加者の中から自分にとって良い情報共有のパートナーを見つけ出すのは、とても知的かつ時間のかかるタスクであり、人にとって大きな負担となる。

コミュニティ形成や情報共有を支援するシステムは幾つか提案されている。KautzらのREFERRAL WEB[36]はWWWにおける人の関係の重要性に注目したシステムである。REFERRAL WEBは文献の引用関係などを利用して人の関係を見つけ出す。角らは美術館や研究会などで、同じ関心を持った人同士の出会いを人の関係を可視化することで支援するシステムを提案している[60]。また、亀井らは参加者の関係を可視化することでコミュニティの形成を支援するシステムを提案している[34]。

これらのシステムは、支援対象となるグループの存在が明示的もしくは暗示的に仮定されている。グループのサイズが大きければ良い相手と情報共有ができる可能性が高くなるが、大きなグループであるほどより良い相手を見つけ出すことは難しくなる。また、そのような大きなグループを作り、運営することも困難である。特殊なグループを対象とすることでそのような問題を回避している例がある。REFERRAL WEBは研究者のグループが持つ文献データベースという質の高い情報を利用することで、効果的な支援を実現している。角らは研究会の参加者や美術館の訪問者にシステムを適用することで良い結果を得ている。しかしこれらのシステムは誰もが使えるというわけではない。

グループの形を少し異なる方法で作りに出しているものもある。人の情報共有のネットワークを介した情報伝播を利用した情報共有システムである。これは、直接的に情報を共有・交換するのは自分が選択した相手とだけだが、共有・交換対象となった情報が共有したユーザらを介してさらに他の相手に伝わるようになっている。竹内らの情報伝播システム[65]では他人から受け取ったメッセージで興味深いものがあつた場合、システムに対



して関心があることを伝えるだけで、自分と共有関係にある全メンバーに対して自動的にメッセージが転送される。吉田らの WAVE[73] では自分の情報を選択した相手にだけ公開し、その情報に興味を持ったユーザはそれを取り込み、自分が持つ情報として公開する。この行為の繰り返しによって情報が伝播していく。これらのシステムはいわば口コミのネットワークを用いて情報共有をしようというもので、自分を中心とした、境界の緩やかなグループを各ユーザがそれぞれ持っているようなイメージとなる。このモデルでは明示的にグループを作る必要は無い。しかし竹内の報告 [65][64] では、このような口コミによって行なわれる情報共有では、直接情報交換を行なう相手の選択が重要であると指摘されている。つまり、このような緩やかなグループ形態をとった場合でも、相手の選択は重要な課題である。

より良い相手と情報共有をしたいならば、我々は多くの相手と出会い、そして選択しなければならぬ。ゆえに情報共有を支援するシステムには、新しい相手と出会い選び出さなければならぬという潜在的な問題を解決する方法を持つ必要がある。

我々は実生活において、しばしば友達の紹介を介して新しい友達を得る。紹介された友達は、出会う前には当然知らないため相手を選択することはできない。しかし、両者を知っている共通の友達がすでに選択を行った上で紹介を行ってくれているため、ある程度の選択を経た相手と出会うことができる。我々は、このような友達が新しい相手とのマッチメイカーになるというアイデアを用いた、新しいネットワーク最適化手法を提案する。これについては、4章にて詳しく述べる。

本研究で提案する共有相手発見手法は、各情報源が共有相手となる情報源を確保していることが前提となる。提案手法は、その既存の関係性を元に他の情報源と共有関係を結ぶ。このような各情報源がすでに持っている共有相手というのは、情報源が人であるとすると、パーソナルネットワークに該当する。

Pluらの Web of People[52] では、人を1つのコンテンツとして扱い、それらをリンクで繋ぐことにより人の Web を作ろうとしている。Kautz らの ReferralWeb[36] は論文を解析し、共著関係を用いて人のネットワークを生成する。利用者は任意の2人の名前を入力すると、その2人を繋ぐ人のネットワークを見ることができる。松尾らの人間関係発見システム [44] は、入力された2人の名前が共起している Web ページを検索エンジンで探しだし、文中のキーワードから2人の関係を求め、人のネットワークを生成する。緒方らの PeCo-Mediator[50] は電子メールからパーソナルネットワークを発見し、質問を投げべき相手の発見を支援する。このような両者が関わる情報から人と人の関係を導き出すアプローチと異なるものとして原田らのシステム [26] が挙げられる。これは検索結果から人名を取り出し、検索クエリーに対して特に関連の深い人物を提示する。

パーソナルネットワークを用いた情報共有システムとして、沼尾らの WAVE[73] や竹内らの情報伝播システム [65] が挙げられる。これらは人のネットワークによる多段的フィ

ルタリングの効果や、緩やかな情報共有を行うことを目的としている。竹内の報告 [64] では、この様な口コミによって行なわれる情報共有では、直接情報交換を行なう相手の選択が重要であると指摘されている。また、Paolo らは協調フィルタリングにおける信頼性の問題や初期における Neighbor の獲得という点に対してパーソナルネットワークは有効であると述べている [42]。

このように、パーソナルネットワークと情報共有システムの親和性は高いと思われるが、その関係やどのようにパーソナルネットワークが情報共有システムに貢献できるかという点についてはまだ十分に議論されていない。そこで本研究では、実環境でのシステム運営を通してパーソナルネットワークを用いた情報共有システムについて分析を行った。これについては、5章にて詳しく述べる。

#### 1.4.4 ネットワークに関する研究

本研究では、各個人がその時点で共有可能な相手と共有を行い、さらにその共有関係を利用して新しい共有相手を発見していく手法を提案する。これはつまり各個人が情報共有ネットワークを構成し、それをボトムアップに再構築していくという形式になる。この場合、形成されるネットワークの構造が重要な意味を持つと考えられるが、そのようなネットワーク構造の分析に着目した研究が社会学および物理学によってなされている。

社会学では、ネットワークが持つ構造にどのような意味が隠されているかを分析する、ネットワーク分析 [71] と呼ばれる研究分野がある。これは例えば、ネットワークの各ノードがどのようなエッジを持っているかという特徴から、どのノードがこのネットワークにおいて中心的であるかを見つけ出したり [21]、どのノードとノードが同じような関係にあるかを見つけ出したりする。基本的に現在あるネットワーク構造が指し示すものを見つけ出そうとしていると言える。

対して物理学では、Watts [70] や Barabashi [9] らによって自然界にも多く存在するネットワークが、どのような法則によって生成されたのかを、様々なネットワーク生成モデルを提案し、実際に存在するネットワークに当てはめて分析する事で調べようとしている研究が多くなされている。

本研究では、ネットワーク構造を利用して情報共有をより効率よく行えるようにしたいと考える。よって、現時点でのネットワークの構造の良し悪しを評価するという点で社会学のアプローチが、そのネットワークをより良い方向へ変化させるという点で物理学のアプローチが、それぞれ参考になると思われる。

### 1.4.5 最近の Web の動向

最近の Web における注目すべき動きである Weblog と Social Networking Service(SNS) について述べる。

Weblog とは個人が書く日記サイトの様なもので、それに特化した Content Management System(CMS) が用意されているのが特徴である。Weblog 用の CMS として有名なものには米 Six Apart 社の Movable Type[6] がある。ホスティングサービスを使えば誰でも綺麗な Web サイトが作れるという事で、多くの利用者を集めている。Weblog の注目すべき点は CMS を介してコンテンツが作成されているため、HTML 作成と同時に RSS[25] 等のメタデータ配信も同時に行われていることである。

SNS とは知り合いリストをオンライン上に登録するサービスで、知り合いリストを介して友達の友達を見られたり、知り合いだけに限定した日記の公開などが可能である。米国では Friendster[22] や Orkut[24]、日本国内では Gree[74] や Mixi[18] などが有名である。

Weblog も SNS も Web 上での人々の情報発信を支援するサービスであるが、ある特定の場所での発言によってしか発言者が特徴づけられない掲示板などと異なり、発言者と発信情報の関連性をより強く残す性質がある。これは人々がよりネットワーク上の個人として振る舞いやすくするシステムであると言え、本研究と密接な関わりがあると考えられる。

## 1.5 本論文の構成

本論文の構成を図 1.1 に示す。まず、1 章にて本研究の目的と取り組む課題について述べ、さらに関連研究との比較により本研究の位置づけを明確にする。次に続く 2, 3, 4, 5 章では、それぞれ要素技術の提案と検証を行い、最後に 6 章にて本論文をまとめる。

2 章と 3 章では、異なる情報源間における情報共有を支援する技術およびそれをを用いたシステムについて述べる。

本研究における異なる情報源とは、互いに共有可能である情報に対して異なる観点でもって収集・分類を行っている情報源(の集まり)を指す。例えば Web ブックマークは、持ち主が Web ページを閲覧し、気になるものを自分なりの観点で分類したものである。また、Yahoo や Dmoz のようなインターネットディレクトリも、管理組織が Web ページを収集し、独自の観点で分類整理したものである。

このような情報源間では、収集対象となる情報を他の情報源がすでに持っている可能性があるという点で共有を行う価値があるといえるが、異なる観点で収集・分類しているため、互いに持っている情報をそのまま共有するだけでは不必要なものも多く、共有してからさらに情報の選別という作業が必要になる。このような異なる概念体系を持って情報を

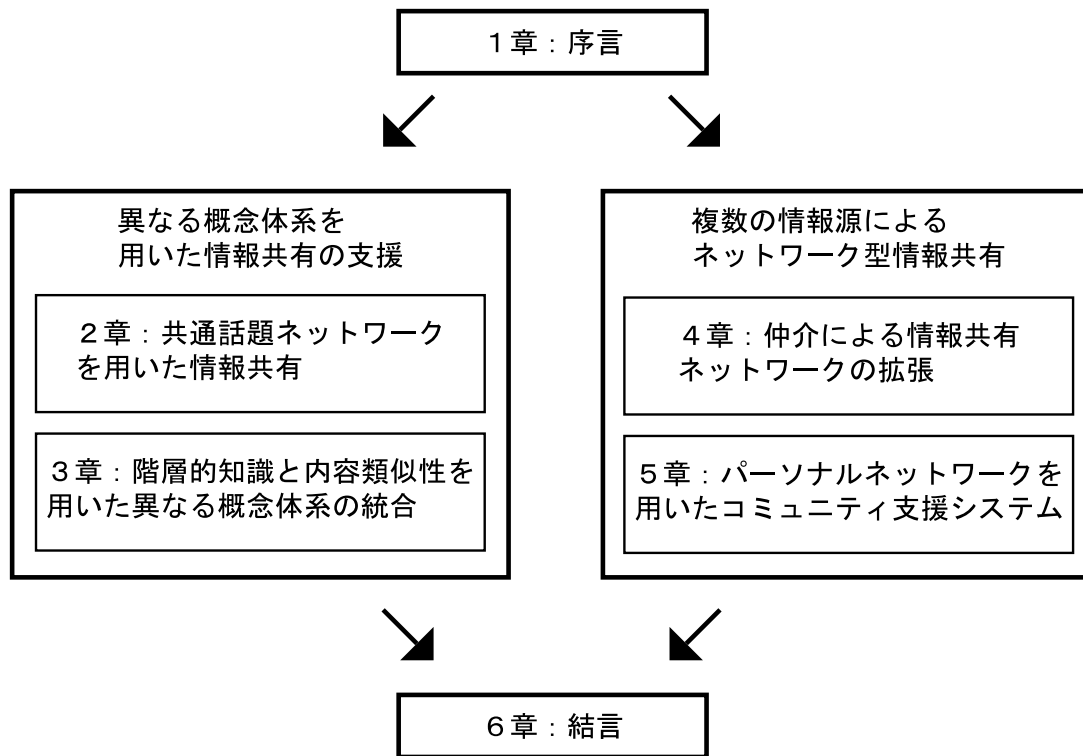


図 1.1 本論文の構成

整理している情報源間での情報共有をどのように支援するのが本研究の課題となる。

2章では、個人間の情報共有の手段として、個人が独自に作成した階層構造を用いることが有効であることを、Web ブックマーク共有システム kMedia を用いた被験者実験により検証する。kMedia は個人が持つ Web ブックマークの共有を支援するシステムであり、Web ブックマークの階層を話題としてとらえ、話題間の類似性を用いて情報の共有を行う。

3章では、より高性能な階層構造の統合技術について検討する。WebHical は木構造によって情報を管理している情報源間での情報共有を支援するシステムであり、階層が持つ構造情報と、階層内に分類されている情報の内容とを使うことで、階層間の類似性を求める。

4章と5章では、人のネットワークを用いた情報共有に関する研究について述べる。

本研究では、自分と同様の立場にある人を利用することで WWW における情報収集の支援を実現しようと考えている。これはオンラインコミュニティや協調フィルタリング等で実現されているアプローチであるが、一つ大きく異なるのは、共有相手を主体的に利用者が選択するという点である。利用者自身が選択した相手との情報共有を行い、その個人ごとの情報共有関係の連なりが大きなネットワークとなり、結果的に多くの人との情報共有関係が結ばれる。

この場合、個人が持つ人のネットワーク（パーソナルネットワーク）を元に情報共有が行われることになるが、そもそも個人が持つパーソナルネットワークは、個人の知りうる範囲で適切であるかもしれないが、全体から見ればより適切なネットワークが存在する。そこでパーソナルネットワークをより改善する手法が必要になる。また、このようなパーソナルネットワークによる情報共有がどのように行われ、どのような特徴を持つかを調べる必要がある。

4章では、情報共有ネットワークを拡張する手法について述べる。提案する近傍仲介法は、すでに個々の情報源間の関係に基づいて情報共有ネットワークが形成された状態において、そのネットワークをさらに発展させるための手法である。シミュレーション実験により提案手法によるネットワークの再構築が機能するかどうかを検証する。

5章では、人（情報源）のネットワークを用いた情報共有システム上で、利用者がどのように振る舞うのか、どのような人のネットワークが獲得されるのかを、実運用したシステムの運用結果から分析する。

本研究は多種多様な情報源が存在する WWW における情報収集支援を目的としている。本研究ではこの課題へのアプローチとして、利用者が発信者にもなりうる WWW の特性を活かし、情報収集者自身が情報提供源の一つとなって信頼できる他の情報源と共に情報共有ネットワークを築き、その「情報源の網」によって情報の収集を容易にするという方法を考えている。2, 3, 4, 5章では、この構想を実現するために必要と思われる技術の提案・検証を行う。

## 第 2 章

# 共通話題ネットワークを用いた情報共有

本章では、異なる観点で情報を収集および管理している情報源（ネットワーク上の個人）間での情報共有をどのようにして行うべきかという点について、本研究が提案する階層構造を用いた情報共有の有効性について検証を行う。提案の有効性を検証する当たり、ネットワーク上の個人として、実際に人が Web ページを収集および管理した結果である Web ブックマークを用い、また情報共有の有効性を測るための試験システムとして Web ブックマーク共有システム *kMedia* を用いた。本章では、まず Web ブックマーク共有システム *kMedia* とそれを用いた実験概要について説明をし、次に実験結果の分析について述べる。分析では、階層構造を用いた情報共有の有効性について述べる。

## 2.1 はじめに

本研究が取り組むべき課題として、情報共有相手と共有している興味の同定があると1にて述べた。そして、本研究では個人の関心を階層構造によって表現し、概念階層単位での共有によって情報共有を行う方法を提案した。本章では、この手法の有効性について、Webブックマーク共有システム *kMedia* を用いた被験者実験により検証する。

本章では、ネットワーク上の個人として Web ブックマークを用いる。Web ブックマークは人が自身の趣味、嗜好に基づいて収集した Web ページを階層構造によって分類整理したもので、本研究が想定する情報源に相当するものであると考える。そして実験には Web ブックマーク共有システム *kMedia* を用いる。*kMedia* では Web ブックマークのフォルダ間の関連性を発見し、情報共有（実際には情報推薦のような形式で情報共有を実現させている）を行う。*kMedia* ではフォルダを「話題」と呼び、関連性のあるフォルダを「共通話題」と呼んでいる。この共通話題の発見は、本研究の課題の一つである「共有している興味の同定」に相当する。よって、本章ではこの共通話題を用いた情報推薦の有効性について検証する事で、本研究が提案する階層構造を用いた情報共有の有効性を示す。

## 2.2 ネットワーク上の個人としての Web ブックマーク

Web ブックマークは持ち主が自身の趣味、嗜好に基づいて収集した Web ページを階層構造によって分類整理したもので、本研究が想定する情報源に相当するものであると考える。Web ブックマークは、持ち主の趣味や嗜好が単に反映されただけでなく、そこに含まれる情報（Web ページ）自体が、すでにその持ち主によってフィルタリングがかけられたものであるため、単なる情報の集まりとして捉えても、利用価値の高いものであると言える。

実際に、このような特徴を持つ Web ブックマークを利用して効果的な情報共有および情報推薦を行おうとする研究がいくつか行われている。PowerBookmarks[40] は、ユーザのコメントが記入されているブックマークデータをデータベースで管理し、情報の推薦を行う。WebTagger[38] は、タグ付けを行ったブックマークを用いて、情報共有を効率よく行おうとしている。BookmarkAgent[46] は、ユーザのブックマーク情報を持つエージェントが、ユーザが現在見ている Web ページに対して情報推薦を行う。Siteseer[54] は、ブックマークのフォルダ構造を利用し、フォルダ内の登録された Web ページの比較によって推薦を行う。また、blink.com<sup>\*1</sup>においては、ブックマーク共有サービスが行わ

---

<sup>\*1</sup> <http://www.blink.com/>

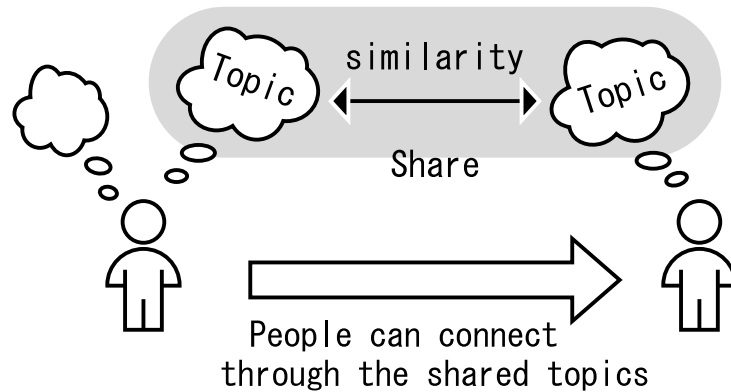


図 2.1 共通話題による人のつながり

れている。

### 2.3 共通話題ネットワーク

本研究で、我々は人のネットワークの実現モデルとして共通話題ネットワークを提案する。ここではユーザの持つ一定のまとまりの興味・関心を話題と呼ぶことにする。<sup>\*2</sup>共通話題ネットワークとは、それぞれのユーザが持つ話題の中で、互いに共通性のあるもの共有したネットワークである。図 2.1 が示すように、ユーザは共通話題を介して他のユーザとつながることができる。

情報交換の点から見た共通話題ネットワークの利点として、以下のようなものが挙げられる。

- ページやファイルのような個々の情報よりも抽象度の高い話題という情報を共有することで、情報の交換を容易にすることができる。
- 人が作り出した話題という情報の関連性を利用することで、既存のテキスト解析では発見できない関連性を利用できる。
- 自分で作った話題に対して新しい情報が入ってくるという形で情報交換が行われるので、受信した情報の理解が容易である。

ここで重要なことは、話題の共通性は単純に話題の名前の比較によって求めることは出来ないという事である。ある情報が与えられた時、その情報がどういう話題に分類されるかは基本的に、人それぞれの判断にゆだねられている。そのため、話題の共通性はその話題に含まれる情報までを調べて初めて判断できる。次章では、この共通話題ネットワーク

<sup>\*2</sup> 話題といった場合、「最近の話題」のような時間軸を考慮したものもあるが、ここでは扱わない。



をブックマークを利用して実装した *kMedia* システムについて説明を行う。

## 2.4 kMedia システム

### 2.4.1 システム概要

本研究では共通話題ネットワークをブックマークデータを用いて構築するためのシステム、*kMedia* を作成した。*kMedia* はクライアント・サーバ型のシステム構成をとっていて、クライアント部は Java2、サーバ部は CGI+Perl5.0 で作成されている。

利用方法は、クライアントを起動させて、設定画面からユーザ名と自分のブックマークデータのある場所を設定して処理を実行させる。クライアントはブックマークデータの解析処理を行い、サーバにその結果を送信する。そして、サーバでの処理が終わり、計算結果が返送されると、クライアントのユーザインタフェースに共通話題ネットワークとウェブページの推薦情報がそれぞれ表示される。

なお、ブックマークデータは Netscape Navigator<sup>\*3</sup>のものを用いているが、Internet Explorer<sup>\*4</sup>のブックマークでも Netscape 形式にブックマークデータを変換する事で利用可能である。

### 2.4.2 インタフェース

*kMedia* のユーザインタフェース (図 2.2) は、共通話題ウィンドウとブックマークウィンドウの 2 つのウィンドウと、ステータスバーで構成されている。

共通話題ウィンドウ (図 2.3) には、共通話題ネットワークが表示される。共通話題ネットワークとは、それぞれのユーザが興味を持つ話題のうち、互いに共通性のあるものを無向グラフで表示したものであり、ユーザを示す黄色のノード (図 2.3 における A, B, C というラベルが付けられたノード) と話題を示す青色のノード (図 2.3 における A, B, C 以外のノード) によって、それぞれのユーザの話題の共通性が示されている。

図 2.3 のグラフは話題を仲介に人がつながっているので、人と人のつながりを表すグラフとしても理解することができる。例えば、図 2.3 では A さんを中心に見ると、UNIX 関係の話題 (UNIX と ftp など) は C さんの話題に繋がっている。このことから、A さんは UNIX で困ったことがあれば C さんに聞けば良いのかな、といった判断を行うことができる。このように共通話題を用いると、人と人の関係をそのつながりの量や内容から理解することができる。

---

\*3 <http://www.netscape.com/>

\*4 <http://www.microsoft.com/>

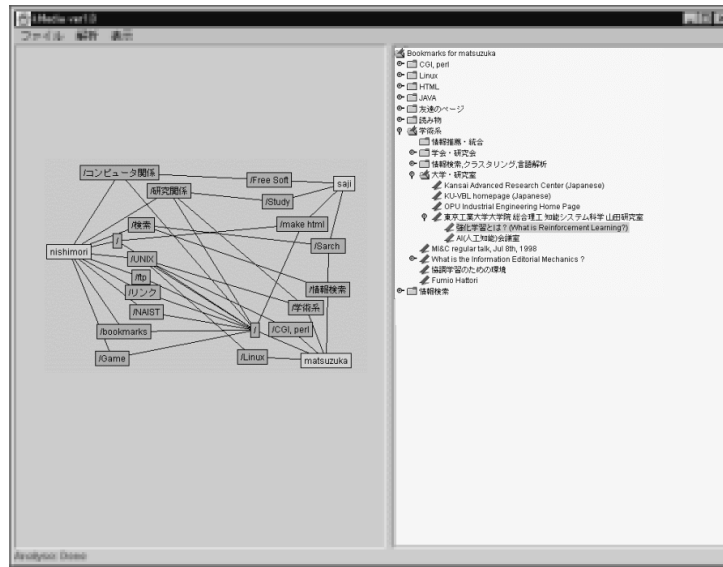
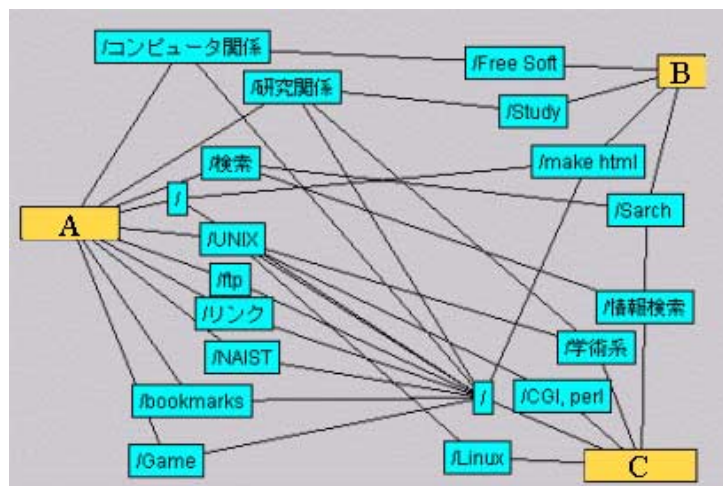
図 2.2 *kMedia* のユーザインタフェース

図 2.3 共通話題ウィンドウ

図 2.2 の右側のウィンドウがブックマークウィンドウである．ここには自分のブックマークデータに，他のユーザのデータで類似したものが推薦情報として付加されて表示される．ユーザは Web ページアイコンをダブルクリックすることで，Web ページをブラウジングすることが可能である．

### 2.4.3 共通話題ネットワークの発見

*kMedia* では，共通ネットワークを発見するために，ブックマークからユーザが興味を

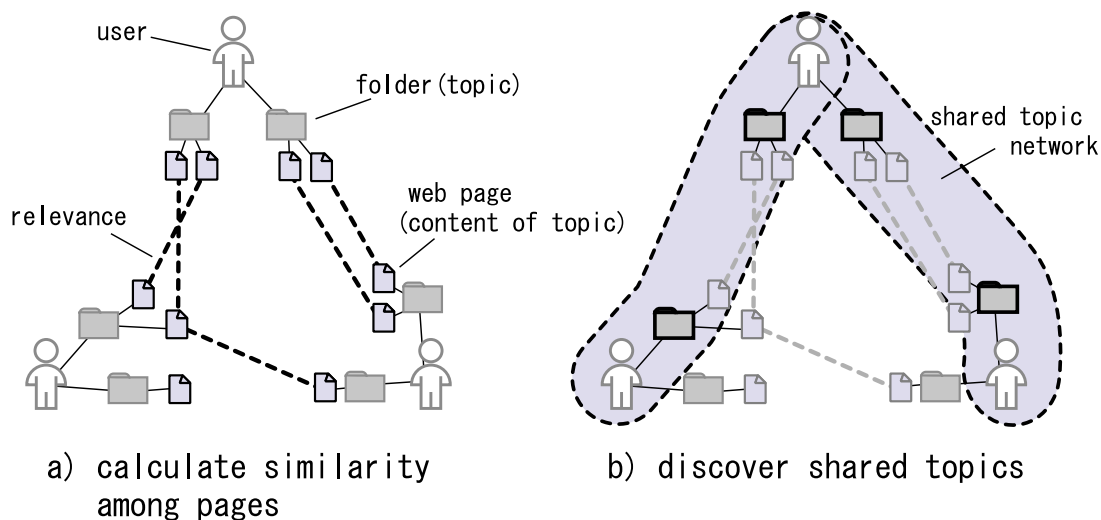


図 2.4 共通話題の発見

持っている話題を調べ、次にユーザ同士の話題の共通性を調べる。

まず、ユーザがどんな話題に興味を持っているかを知るために、*kMedia* はブックマークのフォルダを利用している。フォルダにはユーザが任意の名前を付けることができるが、一般的にフォルダ名には、その中にある Web ページの話題は何であるかが記述されている。そこで、*kMedia* ではブックマークのフォルダにつけられた名前は、その中にある Web ページの話題を表しているとして、また、フォルダ中にある Web ページは、その話題に関する情報であるとしている。

ユーザ同士の話題の共通性は、その話題に含まれる情報がどれだけ類似しているかによって決まる。そこで *kMedia* では、話題の共通性はブックマークフォルダに含まれる情報、すなわちフォルダ中に含まれる Web ページの類似性から求める。つまり、Web ページの類似性が図 2.4-a のように発見された場合、話題の共通性は図 2.4-b のように示される。

なお、*kMedia* では、Web ページの類似性を、ページ中の重要な単語（キーワード）の一致によって決める。Web ページ中の出現頻度の多い単語の中から、上位幾つかをそのページのキーワードとし、互いのページで一致したキーワード数が多いほど、そのページ間には類似性があるとしている。

ここで求められた Web ページの類似性を元に、*kMedia* はブックマークデータの推薦を行っている。すなわち、各ユーザが持っているブックマークに登録された Web ページで、設定した閾値以上のキーワードが一致する Web ページを推薦情報として提示する。

## Web ページ間の類似

*kMedia* は指定されたブックマークファイルを解析し、ブックマークに登録された Web ページを読み込み、その Web ページの文章から単語を切り出す。単語の切り出しには、茶筌 [76] などによる形態素解析ではなく、字種に注目した単語切り出し [35] を用いた。

字種に注目した単語切り出しとは、解析する文章を先頭から一字ずつ読み取り、連続する漢字、カタカナ、全角英数文字を単語として切り出す方法である。この単語切り出し方法は形態素解析に比べて厳密さには欠けるが、特にカタカナの新語、略語に対して柔軟な切り出しができるため、WWW の情報のように、雑多でやわらかい文章の解析に適している。

切り出された単語のうち、頻度順に、頻度が同じ場合は文字長の長い順に最大 10 単語をそのページのキーワードとしている。ただし、「ホームページ」や「カウント」など、あまりその文書の特徴付けられないような頻出単語はキーワードとして除外した。なお、単語と見なす最短文字長は調整可能であるが、今回は 3 文字とした。

このようにして選ばれた各 Web ページのキーワードの一致から、Web ページ間の類似を調べる。今回は、キーワードが 2 語以上一致し、一致したキーワードが両 Web ページで合計 10 回以上現れた場合に、Web ページ間に類似性があったと判断し、Web ページの推薦を行う。なお、このキーワードの出現回数のことを、以後、ページ関連度と呼ぶ。

## 話題の共通性

話題の共通性は、話題を示すブックマークフォルダ中の Web ページの類似によって決める。フォルダ間で互いに類似性がある Web ページの数をフォルダ関連度と呼ぶ。以下の実験では、フォルダ関連度が 3 以上、つまりフォルダ中の Web ページが 3 つ以上、互いに類似性がある場合に、そのフォルダ（話題）間に共通性があるとし、フォルダの推薦を行う。

なお、以下の実験では話題の共通性を調べる際、フォルダの階層は 2 階層以上は考慮していない。つまり、2 階層以上深いフォルダ内の Web ページは、全て親フォルダ内の Web ページと見なした。*kMedia* では、2 階層以上を扱うことも可能であるが、表示や分析の煩雑さを考慮し、このようにした。

## 2.5 実験概要

このような共通話題ネットワーク発見手法の有効性について、実験を行って検証した。まず本章では実験の概要について述べる。

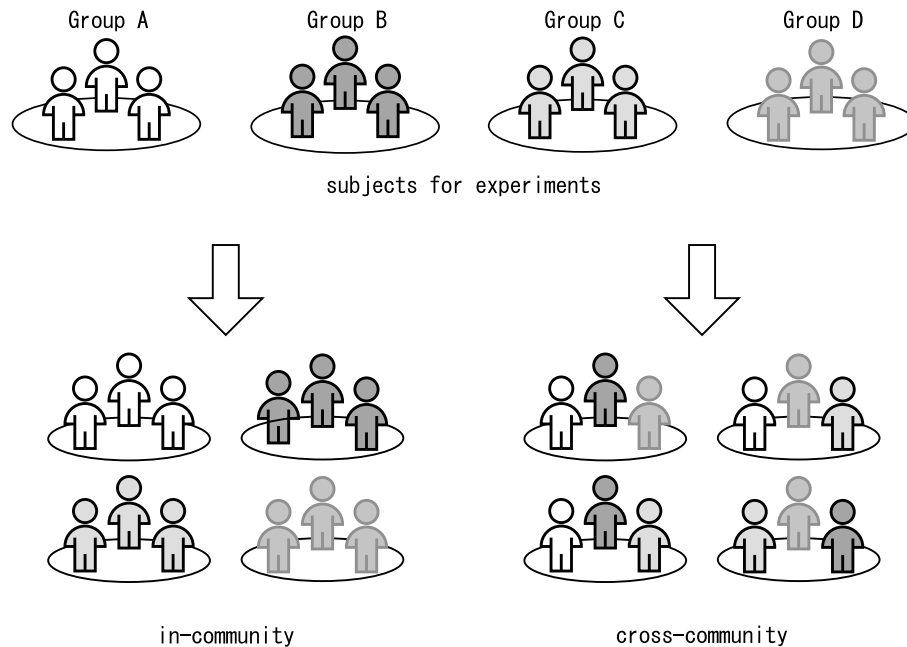


図 2.5 被験者のグループ分け

実験の目的は、人と人を行き交う情報として話題を用いることの有効性を検証すること、人と人との関係を測るための指標を見つけることの2点である。

被験者として、4つの異なるコミュニティから3人ずつを集めた。ここで言うコミュニティとは、実生活上一つのグループに属して活動していることを指す（今回の例では大学・大学院の研究室）。同じコミュニティに所属する人同士（in-community）と、異なるコミュニティに所属する人同士（cross-community）とで、ユーザの評価がどの程度変わるかを調べるために、互いに異なるコミュニティに属する者だけのグループを、同じ12人の被験者を用いて4グループ作った（図2.5）。

表2.1は被験者から集めたブックマークに登録されたフォルダおよびURLの数を示している。共有されているURLの数とは、複数の被験者がブックマークに登録しているURLを指している。グループA～Dの共有されているURLの数は、グループ内で複数登録しているURLの数である。そのため、グループA～Dの共有されているURL数の総和は全体における共有されているURL数と異なっている。

この実験では、*kMedia*を通常の利用法とは異なる方法を用いた。まず、被験者から集めたブックマークを実験者が*kMedia*に与える。そして、その結果をHTML形式に加工し、フォームを用いて評価してもらった。各被験者には、まず同じコミュニティに所属する人同士で*kMedia*を使った場合の推薦結果に対する評価、次に異なるコミュニティに所属する人同士での結果に対する評価をしてもらう。なお、被験者には推薦相手の名前は伏せており、推薦された情報が同じコミュニティに所属している人からのものか、そうでな

表 2.1 被験者のブックマークに登録されたフォルダおよび URL の数

	平均登録フォルダ数	平均登録 URL 数	共有されている URL の数
グループ A	18.7	109.7	21
グループ B	39.7	163.0	20
グループ C	32.7	161.3	3
グループ D	32.3	123.0	18
全体	30.6	139.3	98

いかはわからないようにした。

評価は、以下に示すような推薦されたページに対する評価と推薦されたフォルダに対する評価、そしてページやフォルダを推薦した推薦者に対する評価の合計3種類5項目、それぞれ5段階（5:大変そう思う, 4:そう思う, 3:どちらとも言えない, 2:そう思わない, 1:全く思わない）で評価を行ってもらった。

- あるページに対してのページ推薦が妥当であるか
- 推薦されたフォルダは似ているか（フォルダ評価1）
- 推薦されたフォルダは役立つか（フォルダ評価2）
- 相手と連絡を取りたいか（推薦者評価1）
- 相手と会いたいか（推薦者評価2）

まず被験者には推薦されたページに対して、自分が持つ被推薦ページと推薦ページの組み合わせが適切であるかどうか、それぞれ5段階で評価を行ってもらった。次に、推薦されたフォルダに対して、「似ているか」「役立つか」という2つの観点から、5段階で評価を行ってもらった。なお、評価フォームの下には相手のブックマークが表示されており、被験者はそのフォルダ内の Web ページを見ながら、推薦されたフォルダを評価する。

図 2.7 は、自分の持つフォルダに対して推薦されたフォルダを評価するフォームである。項目は「似ているか」「役立つか」の二つであり、それぞれ5段階評価を行う。ページ推薦と異なり、フォルダ推薦では推薦相手が匿名で表示される。

被験者には、推薦されたフォルダに対して、「似ているか」「役立つか」という2つの観点から、5段階で評価を行ってもらった。なお、このフォームの下には相手のブックマークが表示されており、被験者はそのブックマークファイルを見ながら、推薦されたフォルダを評価する。

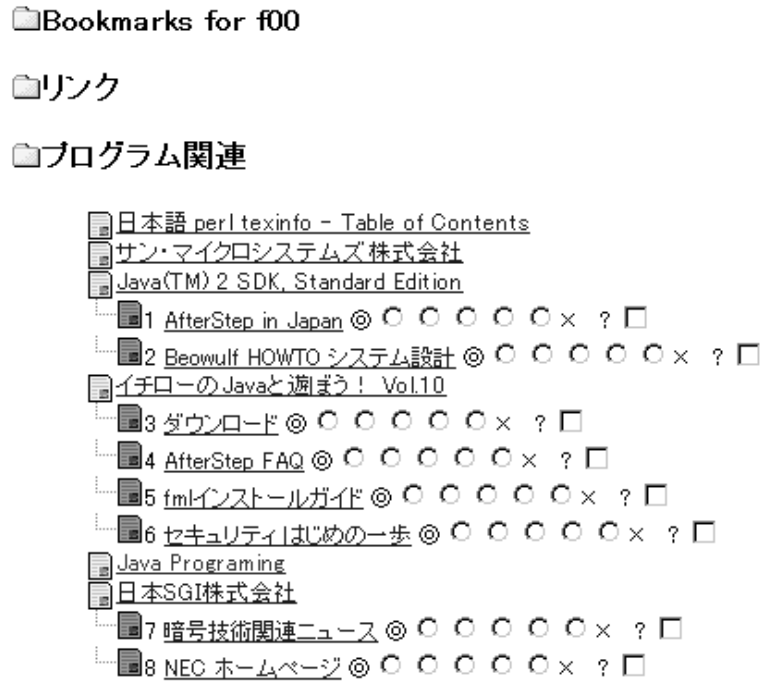


図 2.6 評価フォーム 1

 Human Network

各フォルダについて、以下の観点から評価して下さい。

- 似ているか: 該当するあなたのフォルダーに似ているかどうか
- 役立つか: 該当するあなたのフォルダー内のページを増やしていくのに役立つかどうか

相手のフォルダ内のページリストは、フォルダ名をクリックすると見る事ができます。

あなた	似ているか		役立つか		Aさん
	◎	×	◎	×	
Hobby関係	○ ○ ○ ○ ○ ○		○ ○ ○ ○ ○ ○		奈良
PC関連	○ ○ ○ ○ ○ ○		○ ○ ○ ○ ○ ○		日経関係
奈良先	○ ○ ○ ○ ○ ○		○ ○ ○ ○ ○ ○		研究関連
UNIX & Linux & etc	○ ○ ○ ○ ○ ○		○ ○ ○ ○ ○ ○		UNIX
Hobby関係	○ ○ ○ ○ ○ ○		○ ○ ○ ○ ○ ○		useful
奈良先	○ ○ ○ ○ ○ ○		○ ○ ○ ○ ○ ○		naist
奈良先	○ ○ ○ ○ ○ ○		○ ○ ○ ○ ○ ○		UNIX
奈良先	○ ○ ○ ○ ○ ○		○ ○ ○ ○ ○ ○		new

図 2.7 評価フォーム 2

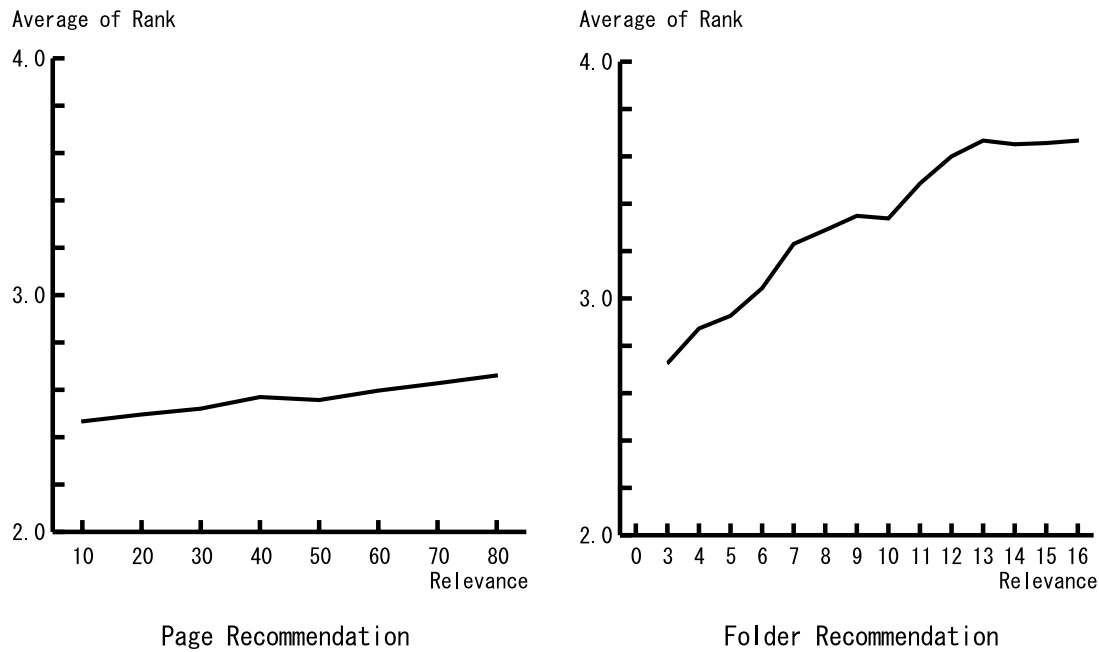


図 2.8 フォルダ推薦の有効性

## 2.6 ページとフォルダに関する共通性の分析

### 2.6.1 フォルダ推薦の有効性

図 2.8 は、ページ推薦とフォルダ推薦の評価結果を示したものである。なお、フォルダ推薦については 2 種類の評価をしてもらったが、両方とも同じ傾向であったため、「似ているか」の評価結果を示す。

それぞれ縦軸が平均評価値、横軸が推薦を行うための関連度の閾値である。今回の実験で用いた閾値（ページ関連度 10、フォルダ関連度 3）で t 検定（有意水準 5%）をしたところ、有意にフォルダ推薦の方が評価が高かった。また、フォルダ推薦に対する評価値はどのような閾値においてもフォルダ推薦の方が高く、また、グラフの傾きは顕著にフォルダ推薦の方が大きい。これはフォルダ関連度の方がユーザの評価により関連している事を示している。このことから、フォルダ（話題）が概念的に Web ページよりも抽象的であるため、ユーザにとって受け入れ易かったのではないかと考えられる。なお、単純にフォルダの方が情報量が大きかったために評価が高くなったとも考えられるが、単にたくさんの情報があるだけ（例えば未分類の情報がたくさんあるだけ、など）では利用者は推薦情報として高く評価しないため、適切な量の情報がまとめられているからこそ評価されたとすれば、やはり人の作ったフォルダという単位が重要であったと言える。



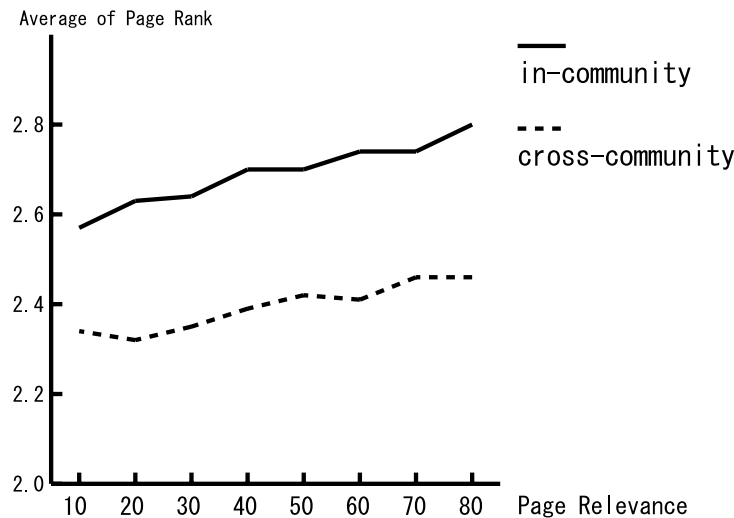


図 2.9 ページ関連度と推薦ページ評価

## 2.6.2 コミュニティの効果

コミュニティの有無で情報推薦の効果がどのように変わるか検証する。同じコミュニティに所属する人同士で行ったページ推薦と、異なるコミュニティに所属する人同士で行ったページ推薦の評価を比較した。図 2.9 は、ページ推薦を行うためのページ関連度の閾値を変化させた場合の、推薦されたページに対する評価の平均値を示したものである。

グラフから、コミュニティの有無に関係がなく、ページ関連度によって設定された閾値が高くなれば、評価も高くなる傾向にあることがわかる。ただし、同じコミュニティ間の方が評価は常に高い。今回用いた閾値（ページ関連度 10）で  $t$  検定（有意水準 5%）を行ったところ、有意であった。情報推薦は、一般に似たような関心をもつ人同士で行った方が有効であるといわれている。*kMedia* によるページ推薦の効果も、同じコミュニティに所属する人同士で行った方が高いという結果が得られた。また、フォルダ推薦においてもほぼ同様の傾向が現れた。

大小の差はあるが、結果としてページ推薦、フォルダ推薦ともにコミュニティの有無による影響が現れた。実生活においても、親しい友人から推薦された情報は他の人からのものよりも有効である可能性が高い。この結果はそれと同様であるといえる。しかし今回の実験においては推薦者が誰であるかは伏せられているので、その所属で判定しているわけではない。ここで行った関係の提示の中に、“コミュニティらしさ”が含まれていたと考えられる。すなわち、人が相手を同じコミュニティだと判断する要素、つまりは人のつながりを示す要素が含まれている筈である。

そこで、人のつながりを示す要素を見つけ出すために、推薦の評価結果以外にどのよう

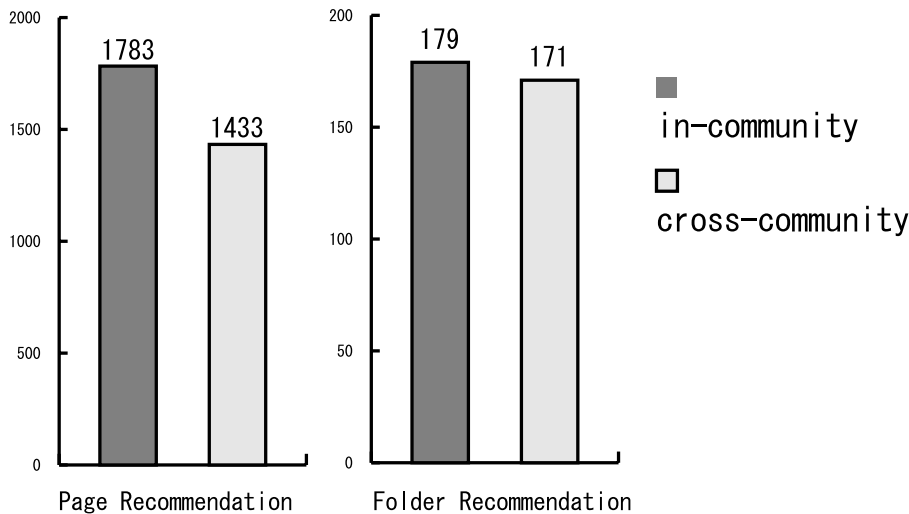


図 2.10 ページ・フォルダ推薦数

な特徴がコミュニティの有無で現れているかを調べる。

#### ページ推薦数とフォルダ推薦数

図 2.10 は、同じコミュニティ間でのページ推薦と異なるコミュニティ間でのページ推薦数、およびフォルダ推薦数を比較したものである。コミュニティの有無で推薦ページ数の差は 300 ほどあるのに対し、推薦フォルダ数の差はわずか 8 であった。

フォルダ推薦はフォルダ中のページの推薦数によって決まる。よって、ページ推薦数が大きいのにフォルダ推薦数が小さいと言うことは、推薦されたフォルダが多くの推薦ページを持っていることを示している。つまり、この結果から、同一コミュニティはそうでない場合と比べて、推薦フォルダがより多くの推薦ページを持つ傾向にあることがわかる。

#### ページ関連度とフォルダ関連度

図 2.11 は、ページ関連度による閾値を変化させた場合、推薦されるページのフォルダ関連度がどのように変化するかを示したものである。推薦ページのフォルダ関連度とは、類似性のあるページ組の、それぞれが登録されているフォルダ間のフォルダ関連度を示す。

ページ関連度による閾値を高くすると、同じコミュニティに所属する人同士で推薦されるページは、フォルダ関連度の高いものが多くなっている。対して、異なるコミュニティでは、そのような傾向は現れない。

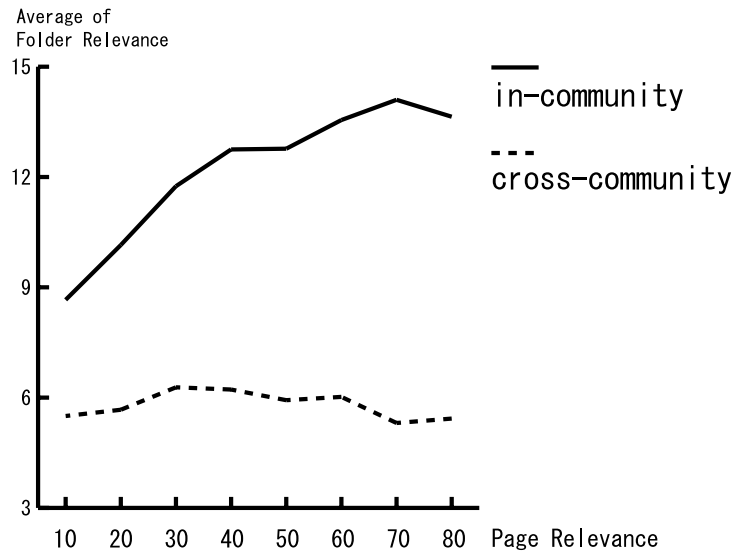


図 2.11 ページ関連度とフォルダ関連度

## 2.7 人々の距離に関する分析

### 2.7.1 カテゴリズ近似度

推薦数および関連度についての分析から，実社会において同じコミュニティに属している（つまり人のつながりがある）ことは，そうでない場合と比較して以下のような二つの傾向があるといえる．

- ページ推薦数に対してフォルダ推薦数は少ない
- ページ関連度とフォルダ関連度の相関が高い

前者は，一つの共通話題に対して多くの類似ページがあることを示し，後者は，よく似たページはよく似た話題に分類されていることを示している．この二つの傾向から，人のつながりの尺度として，分類方法の類似性が使えるのではないかと考えられる．例えば図 2.12 において，点線は関連性があるページ組を示しているが，左右両方とも同じ 4 つである．だが，その関連性のあるページの分類の仕方がそれぞれ異なっている．この場合，同じような分類をしている左の方が，右の方と比較して人のつながりが強いと判断できる．

ここで分類方法が類似性を示すものとして，カテゴリズ近似度という指標を新たに作る．これは，共有している情報の量でユーザ間を評価するのではなく，一致する率で評価するものである．ユーザ  $a, b$  間のカテゴリズ近似度  $CR(a, b)$  は，以下の式で求めることができる．

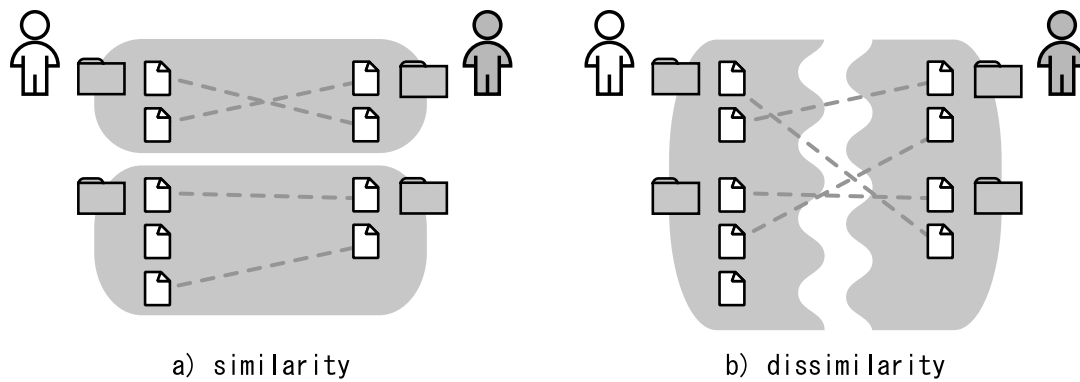


図 2.12 分類方法の類似

$$CR(a, b) = \frac{Nf_{ab} \times Rf_{ab}}{Np_{ab}}$$

$Nf_{ab}$  : 推薦フォルダ数

$Rf_{ab}$  : 推薦フォルダの平均フォルダ関連度

$Np_{ab}$  : 推薦ページ数

## 2.7.2 推薦者評価とカテゴリズ近似度

今回の実験では、推薦されたページやフォルダと同様に、それらのページやフォルダを推薦した推薦者自身の評価も行った。この、推薦者への評価という値は、人のつながりを示す値であると言える。そこで、それらの値とカテゴリズ近似度の相関を調べることで、提案するパラメータが人のつながりを測る指標となりうるのかを調べた。

まず、比較対象としてシステムが計算した四つの基本的なパラメータと被験者による推薦者への評価との相関係数を表 2.2 に示す。最も相関係数が高いパラメータは推薦フォルダ数であった。また、推薦者に対する評価は「連絡したいか」「会いたいか」という2つの項目を用意したが、全てのパラメータにおいて前者の方が相関係数は高いという結果になった。

次に示す表 2.3 は、推薦されたページやフォルダに対する被験者の評価と、その推薦者に対する評価との相関を示したものである。被験者による評価係数は「会いたい」との相関においてシステムが求めたパラメータよりも高い値を出しているが「連絡したい」との相関においては逆に低い結果となった。

「連絡したい」と「会いたい」という二つの評価項目を比較すると、直感的には後者の

表 2.2 推薦者の評価との相関 (1)

	連絡したい	会いたい
推薦ページ数	0.42	0.30
平均ページ関連度	-0.13	-0.19
推薦フォルダ数	0.45	0.35
平均フォルダ関連度	0.38	0.30

表 2.3 推薦者の評価との相関 (2)

	連絡したい	会いたい
平均ページ評価	0.29	0.40
平均フォルダ評価 1	0.28	0.32
平均フォルダ評価 2	0.09	0.20

表 2.4 推薦者の評価との相関 (3)

	連絡したい	会いたい
カテゴリズ近似度	<b>0.49</b>	<b>0.55</b>

方が人のつながりを強くあらわしていると思われる。ならば、人のつながりを測る指標は

(「連絡したい」との相関) < (「会いたい」との相関)

となるべきである。しかし、システムが計算した値では逆の傾向が現れてしまう。このことから、システムが計算した値はなにかしら不十分であると言える。では人が評価した値はどうであるかという、「連絡したい」との相関係数と「会いたい」との相関係数の関係は正しい。しかし、不十分であると判断したシステムが計算した値よりも、「連絡したい」との相関係数において低い結果があらわれた。このことから、人が評価した値も、まだ不十分であると言える。

最後に、今回新たに提案したカテゴリズ近似度と人の評価との相関を表 2.4 に示す。

カテゴリズ近似度の相関係数は両項目とも最高値であり、かつ大小関係も妥当なもの(「会いたい」との相関の方が高い)であった。また、このカテゴリズ近似度は相関の有意性検定において、有意水準 5% で有意であった。

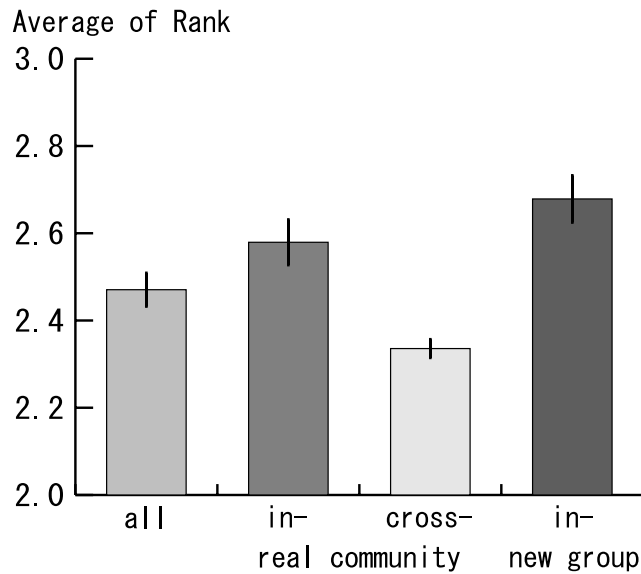


図 2.13 ページ推薦におけるカテゴライズ近似度の有効性

### 2.7.3 ページ推薦とカテゴライズ近似度

さらにカテゴライズ近似度がページ推薦に与える影響を調べた。今回の実験では一人当たり四人の推薦者からの推薦情報を評価してもらったが、その推薦者四人のうちカテゴライズ近似度が高かった者上位二人を、被推薦者にとって同じコミュニティに属する人だと判断して新たに仮想的にコミュニティを作成した。

図 2.13 は、全体 (all)、実際のコミュニティがある場合 (in real community) とない場合 (cross real community)、新しく作った仮想コミュニティがある場合 (in new group) の合計四種類のグループにおけるページ推薦の評価値の平均値を示している。実際のコミュニティ同様、カテゴライズ近似度で作った新しいコミュニティにおいても、その有無によってページ推薦の評価値に差が現れた。この差は  $t$  検定 (有意水準 5%) で有意であった。また、コミュニティを全く考慮しない場合 (図 2.13-all) との差も、同じく  $t$  検定 (有意水準 5%) で有意であった。

## 2.8 議論

### 2.8.1 オンライン上の個人として用いる情報源について

今回の実験ではオンライン上の個人として Web Bookmark を用いた。これはあくまで一つの例題でしかなく、その他にも色々なものが考えられる。次章で取り上げるインター

ネットディレクトリもそうであるし、Samba や WebDAV など情報共有を行った際のディレクトリもその一つである。

Web ブックマークの特徴としては、まず収集対象となる情報が Web ページに限定されることが挙げられる。人の興味は様々であるが、Web ブックマークに登録される Web ページは、その人の興味のなかでも Web 上でコンテンツがあるものに限られる。また、収集対象である Web ページは一般に誰からでもアクセス可能であるため、他の人と同じものを収集する可能性がある。このように収集対象となる情報そのものや、その分野について、他の人との間で重複が十分起こりうる情報源であると言える。この特徴は、本章で提案したような、類似した話題を抽出することで情報共有を支援する手法とは親和性が高いと考えられる。このような特徴は、インターネットディレクトリにも言える。インターネットディレクトリは巨大で幅広い分野にまたがって情報収集しているため重複部分が生じる可能性はより高くなる。しかし、Web ブックマークと比較して構造が詳細化されている可能性があるため、より精度の高い類似する話題（Web ブックマークならフォルダ、インターネットディレクトリならディレクトリ）を抽出する技術が必要になる。これについてはその巨大な構造を利用することが有効であると考えられる。この点については次章にて述べる。

Samba など個人が作成したディレクトリをそのまま共有した場合、共有対象となる情報は、共有しているディレクトリ内に含まれるファイルとなる。なお、Samba 等でファイル共有を行う際にはディレクトリ構造自体を共有していることが多いが、ここではディレクトリはそれぞれ異なるものを用意していると想定する。いわば、お互いに UNIX におけるホームディレクトリや Windows におけるマイドキュメントディレクトリをそのまま共有するようなイメージとなる。このような情報共有の場合、各自が持っている情報はそれぞれ持ち主のオリジナルであることが多い。そのような場合には情報の類似度計算手法の精度が重要になる。テキストに限らず様々なコンテンツにおける類似度の計算方法は色々存在するが、本研究が想定している概念階層の中にあるコンテンツを対象としている状況では、コンテンツ自身の特徴に加えて概念階層の特徴を利用するのが有効であると考えられる。例えば、ある概念階層内の文書には現れるがその概念階層以外では見かけない単語や、概念階層内の文書にはあまり現れないが上位の概念階層の文書には現れる単語などに注目する、といった方法が考えられる。

Web ブックマークとホームディレクトリやマイドキュメントとの両方の特徴を持ったものとして Weblog が挙げられる。Weblog コンテンツは作者によってカテゴリ分けがなされており階層構造を持つ。Weblog コンテンツは基本的にその Weblog の持ち主が書いたオリジナルであり、各 Weblog 間で共通するコンテンツを持っていることはない。また、コンテンツ作者が異なるために同じ話題の Weblog コンテンツであっても、用いる単語などに違いがあって、単純にはコンテンツ間の類似度を求められない可能性が高い。

Weblog コンテンツは一般的に Web にて公開されているので、それを読む側の Web ブックマークには共通するコンテンツが入ることになる。よって、読む側の Web ブックマークが持つ概念階層を踏み台にして Weblog 間の概念階層（カテゴリ）の類似性を求めるという方法も考えられる。

提案する情報共有モデルではコンテンツの類似度をもとに概念階層の類似性を求めている。階層構造があり、その中にあるコンテンツの類似度が計算可能であれば、どのようなものでも対応ができる。そのため Web ブックマークと Weblog と個人のホームディレクトリとの間での情報共有も可能である。しかし、Web ブックマークは読む側が作った概念階層、Weblog は書く側が作った概念階層、ホームディレクトリは持ち主がファイル管理するために作った概念階層、と、それぞれ異なる目的によって作られた概念階層であるため、コンテンツの類似性からそのまま概念階層の類似性が求められるかどうかは今後調べていく必要がある。

## 2.8.2 情報共有によってもたらされる変化について

情報共有が行われることにより新たに情報を得ていくことで、個人は変化していくと考えられる。ここで言う変化とは、今まで知らなかったことを知ったという量的な変化だけでなく、自身の関心や知識構造の変化も含んでいる。これは個人の今後の情報共有の方向性にも影響を与える変化である。

仮に情報共有を継続して行った場合には、統合化と詳細化が行われると考えられる。共通する話題を介して情報共有を行った場合、初期段階においては共通する話題と言っても微妙な差違が存在する。しかし、情報共有を継続して相手から情報を得ることで、その差違は小さくなっていく。これは差違の部分が消えるのではなく、差違を包含するような話題の拡大化が起こることによって生じる（統合化）。このようにして差違を吸収して拡大化した話題には多くの情報が含まれるようになる。情報共有により複数人で情報を集めるので、その収集量は膨大になる。そうすると、その情報を整理する必要が生じ、詳細化した話題が生成されることが考えられる。もし、統合化と詳細化の過程において生じる話題の差違が大きすぎる場合は、異なる話題として情報共有は行われなくなる。逆に差違を生じなかった個人はどんどん集まってきて、一つのコミュニティのようなものを形成すると考えられる。

これに近い現象は Social Bookmarking と呼ばれるサービスにてすでに起こっている [2]。これらのサービスでは、コンテンツに対して自由に複数のラベル（タグ）を付加することができる。ラベルは自由記述であるため、同じ意味のラベルであっても様々な書き方の差違が生じる。しかし、他の人がどのようなラベルを付けているかがわかるため、また、ラベルが共通化されるほど自分が登録したコンテンツと似たコンテンツへのアク



セスが容易になるため、利用者は積極的にその差違を無くそうとする（統合化）。この現象は Folksonomy と呼ばれ、共通のラベル（メタデータ）を付与することが重要である Semantic Web においても注目されている。

今回の実験では、被験者が実際に継続的に利用している Web ブックマークを情報源として用いたが、これはある一時点における情報共有の有効性を測った実験であり、継続的なものではない。本研究が想定しているような情報共有は継続的に行われるものであり、その継続的なインタラクションにより持っている情報や階層構造が変化した場合に、個人および全体においてどのような情報共有が行われるかを観察する必要がある。

### 2.8.3 個人の多様性と発見される共通話題について

今回の実験では、被験者として同年代の 12 人を集めた。被験者の中には互いに面識の無い者もいるが、ある程度似通った集団であった。全く異なるタイプの人たちの間で提案手法を用いた情報共有が有効であるかどうか検討する必要がある。

提案手法では類似した話題を発見して、それを介して情報共有を行う。また、その話題の作り方によって情報共有におけるパートナーとしての適切さを測る指標を提案した。もし情報共有を行う人たちの間で全く関心が異なっている場合は、類似する話題が発見できないことになるので提案手法による情報共有は行えない。だが仮に、タイプの異なる人たちの間で情報共有を行った場合は、話題こそ同じであるが持っている情報そのものが大きく異なる場合もある。例えば、お互いにプログラミング言語の Java に興味はあるが、一方は現役プログラマーで最先端の技術に関する情報を収集しているが、もう一方は初心者で入門的な情報ばかり集めているといった状況が考えられる。この場合、提案手法ではお互いに類似している話題を持っていると判定されるが、実際に共有される情報はお互いに意図するものとは少し異なっている可能性が高い。

このような個人間の比較だけでは発見が困難な話題のズレに対しては、第三の共有相手に注目することが有効であると考えられる。つまり、他にはどのような相手と情報共有を行っているかを調べることでコミュニティの様なものを見つけ出し、それを元にその個人間で共有している話題の類似性を測る。先の例で言えば、現役プログラマーはやはり他の現役プログラマーと情報共有している可能性が高いと考えられるので、一見すると似たような話題を持っている相手でも、自分にとって適切であるとすでにわかっている共有相手との関係（現時点で共有関係にあるかどうか、持っている話題は似ているかどうか、など）を見ることで、判別できる可能性がある。このような見方は、個人と個人との 1 対 1 関係が重なり合っているネットワークを分析対象にすると言い換えることもできる。そして、そのようなネットワークの分析については社会縛の分野において様々な取り組みがなされており、それらの技術を用いることが本研究で提案する情報共有手法の改善にも有効

であると考えられる。

## 2.9 まとめ

本章では、個人間での情報共有を容易にするためには、個人の関係はどのように説明されるべきかという点について議論し、提案する階層構造を用いる事の有効性について、Web ブックマークを元に共通話題を提示するシステム *kMedia* を用いた被験者実験により検証した。

実験結果の分析から、情報共有の単位として概念階層 (*kMedia* では話題と呼ぶ) を用いる事が有効であることがわかった。さらに、階層構造の類似性が情報共有の相手としての適切さを示す指標となりうる事がわかった。これはつまり、階層構造に持ち主の関心が色濃く表されていることを示唆しており、このことから個人知識を階層構造で表現する事が有効であることがわかる。

## 第3章

# 階層的知識と内容的類似性を用いた異なる概念体系の統合

本章では、2章における階層構造を用いた情報共有の有効性を背景に、より高性能な階層構造の統合技術について検討する。提案する *WebHical* は木構造によって情報を管理している情報源の共有を支援するシステムである。*kMedia* による概念（話題）間の類似性を発見する手法は比較的簡単なもので、大きく深い階層構造を扱うには不適切であった。そこで *kMedia* に変わる手法として *WebHical* を提案し、システム構築を行った。

本章では、まず提案する *WebHical* のアルゴリズムについて述べ、次に情報源として実際に利用されているインターネットディレクトリ（Yahoo!と Dmoz）を用いて実験を行い、提案手法の有効性を検証する。

### 3.1 はじめに

本研究では、異なる情報源が異なる概念体系のもとに情報管理をしている状況において、概念体系の概念間の類似性をもとに情報共有を行うことが有用であると考えている。これを、2章では個人（情報源）が持つ WWW ブックマーク（概念体系で管理された情報）を対象として、*kMedia* システムを用いた実験により検証した。

個人では巨大で複雑な概念体系をオンライン上に作るのは難しいが、例えば Yahoo や Dmoz といったインターネットディレクトリは巨大な概念体系をオンライン上に生成している。特に Dmoz は有志によって作られたものであり、これはつまりコミュニティによって巨大な概念体系を作成することが可能であることを示している。本研究が想定する情報源には、個人だけでなくこのようなインターネットディレクトリやコミュニティも含まれるため、巨大な概念体系にも対応した概念（話題）間の類似性発見技術が必要になる。

これに相当する技術として、市瀬らは概念階層が持つインスタンスに基づいて他の知識との相違を調整する規則を学習する手法を提案し、その実装である *Hical* を試作した [30]。*Hical* では、2つの概念階層がインスタンスをそれぞれどのように分類しているかという情報を元に、概念階層間の結合を行っている。この手法の問題として、2つの異なる概念階層間で、十分な数の同じインスタンスを分類していなくてはならないという点が挙げられる。

そこで本研究では、概念階層間で共有されていないインスタンスを、機械的に求めたインスタンス間の類似性に基づいて自動的に共有させることで、同じインスタンスを持っていない概念階層間でも結合が行えるように *Hical* を改良した。本章では、改良したシステム *WebHical* の説明とそれを用いた評価実験について解説をする。

### 3.2 階層的情報源

本研究で仮定する階層的情報源についてモデル化を行う。ここで対象とするのは、概念階層に基づいて分類が行われ、管理されている情報源である。例えば、インターネットディレクトリ、図書の分類目録、オントロジーなどがそのようなものとして挙げられる。これらの情報源における概念階層は、もっとも一般的な概念を最上位として、順により詳細な分類を示す概念からなる階層構造を成している。個々の情報はこの概念階層の中のいずれかに割り当てられて管理される。なお、分野や目的によって最も下位の概念にしか個々の情報が割り当てられていない場合と、任意の概念に割り当て可能な場合があるが、本研究ではより、任意の概念に割り当て可能であるとする。

本章では、分類に使われる概念階層は木構造であると仮定する。先に述べたクラスライ

ブライリーやインターネットディレクトリ，目録分類など様々な分野において，木構造の概念階層は多く用いられている．本研究では，このような概念階層を利用することを目的としているため，その基本構造である木構造を対象とする．なお，複数の上位ノードを持つものは単一の上位ノードを持つものを複数作る形で展開を行うことで木構造として扱う．木構造で示された概念階層を単純化し，グラフを用いて表すと図 3.1 の様に表せる．この図では，概念階層が木構造で表され，インスタンスが木のノードに割り当てられる．黒点がある概念を表し，白点がインスタンスを表す．本章では，以降，情報の実体をインスタンスと呼ぶ．同様に，概念階層によってインスタンスを管理している情報源のことを階層的情報源と呼ぶ．

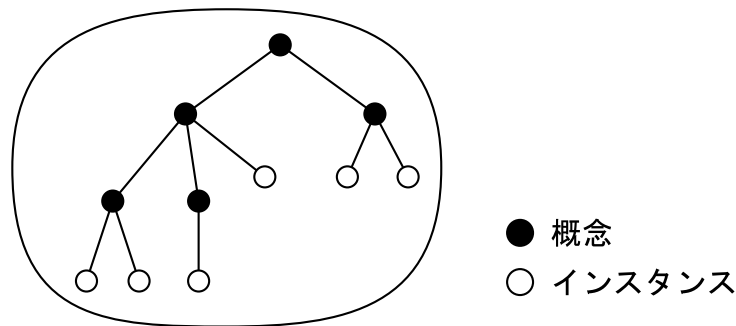


図 3.1 階層的情報源のモデル

階層的情報源は，その中に含まれるインスタンスの種類，概念階層の構築者などによって，異なった階層構造を有する．従って，このような情報源が複数存在する時に，他者の知識を利用するのは困難が伴う．例えば，図 3.2 では，2つの情報源が存在している．概念階層  $H_1$  にはインスタンスとして  $I_1, I_2, I_3$  の3が存在しているが，概念階層  $H_2$  には  $I_1, I_3$  の2つのみが存在しており， $I_2$  は含まれていない．このとき，そのまま  $I_2$  を  $H_2$  に持ってきただけでは， $H_2$  上のどこに位置すべき情報なのかが  $H_2$  にはわからない．従って，そのまま  $I_2$  を利用することはできない．この問題を， $Hical$  は  $j$  情報源の間で共有しているインスタンス  $I_1, I_3$  がどの様に概念階層に分類されているかということから，概念階層間の対応を学習することで解決している．この手法の問題点として，共有するインスタンスが存在しない場合には適用不可能だという点がある．そこで本章では，インスタンスの内容を見ることでインスタンス間の類似関係を見つけ，インスタンスの共有を仮想的に作り出すことで，実際には共有しているインスタンスの無い情報源の間でも概念階層の対応を学習を可能にした *WebHical* を提案する．

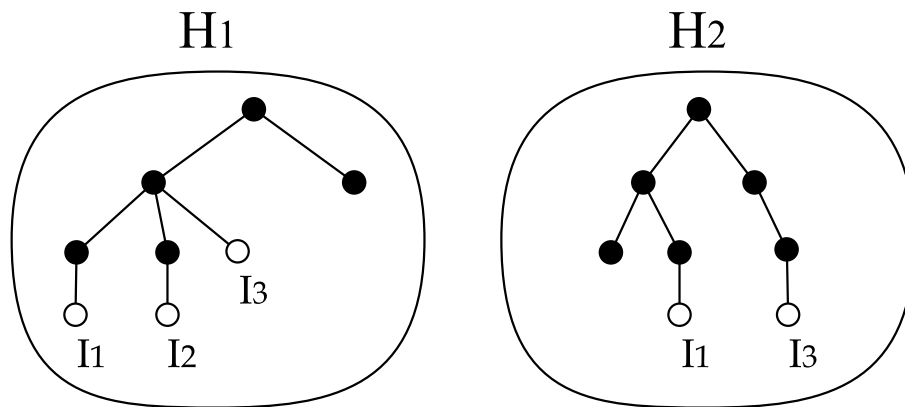


図 3.2 複数の階層的情報源における問題点

### 3.3 階層的情報源の統合手法

*WebHical* は階層的情報源の統合を行うシステムである。ここで、統合とは、ある情報源が持つ知識を異なる情報源に取り込むことを指す。そのためには両者の情報源の相違を吸収する必要がある。提案システムでは、その相違を吸収するために階層的知識と内容的類似性という二種類の情報を利用する。階層的知識とは階層的情報源においてあるインスタンスがどこに入っているかという情報を指し、内容的類似性とは階層的情報源が持つインスタンス間の似ている度合いを指すと、ここでは定義する。この章では、この二種類の情報それぞれを利用した場合の、階層的情報源の統合手法について説明する。

#### 3.3.1 統計量 (階層的知識の利用)

情報源における概念階層は、インスタンスをある概念に従って分類したものである。ここで扱う概念階層は木構造をしているため、ある概念ノードより下に属するインスタンスはその概念ノードに属していると判断できる。そのため、ある概念ノードを選択したときに、任意のインスタンスがその概念に適合するか否かを容易に判定することができる。すると、2つの情報源における任意の2つの概念ノードに対して、共有インスタンスの分類を元に概念基準の類似性の判定を行うことができるようになる。ここで、概念基準とは、ある概念に属するか否かの判断の基準のことを意味する。

*Hical* では、この概念基準の判定に 統計量 [20] を用いる。統計量では、2つの概念に対して表 3.1 のような分割表を作成する。表 3.1 はある概念に含まれるインスタンスの数と含まれないインスタンスの数を一覧にしたものである。 $N_1, N_2$  は、それぞれ概念階層  $H_1, H_2$  中にある概念を表し、 $m_{**}$  はそれぞれの分類に含まれるインスタンスの数を

示している。

表 3.1 インスタンス分割表

		概念 $N_2$	
		含まれる	含まれない
概念 $N_1$	含まれる	$m_{11}$	$m_{12}$
	含まれない	$m_{21}$	$m_{22}$

統計量では、まず、概念基準の一致率  $P$  と偶然に一致率  $P'$  を次の式により計算する。

$$P = \frac{m_{11} + m_{22}}{m_{11} + m_{12} + m_{21} + m_{22}}$$

$$P' = \frac{(m_{11} + m_{12})(m_{11} + m_{21}) + (m_{21} + m_{22})(m_{12} + m_{22})}{(m_{11} + m_{12} + m_{21} + m_{22})^2}$$

その時、統計量は次式で表される。

$$= \frac{P - P'}{1 - P'}$$

この統計量より、概念基準が一致しているかどうかを求める。概念基準が一致したということは、その概念間で知識の交換が可能であることを指す。このようにして階層的知識を用いた異なる階層的情報源の統合は行われる。

### 3.3.2 キーワードマッチング (内容的類似性の利用)

*Hical* の手法は、二つの概念階層が一致する URL を十分な数だけ持っていないとしない。そのため概念的に類似していてもカバーしているインスタンスの領域が全く別である場合には適用できない。そこで *WebHical* では、インスタンスの内容的類似性に基づいて概念階層間でインスタンスの追加を行い、仮想的に一致する URL を作ることで、この問題を解決する手法を用いる。

内容的類似性に基づくインスタンスの追加手法について説明する。まず概念階層  $H_1$  が持つインスタンス  $I_1$  と類似したインスタンス  $I_{1s}$  を、概念階層  $H_2$  が持っているインスタンスの中から抽出する。なお、 $I_{1s}$  にはすでに  $H_1$  が持っているものと同じインスタンスは含まれ得ない。これを  $H_1, H_2$  が持つインスタンス全てに対して行う。類似したインスタンスの抽出は、インスタンスの内容を解析し、その中身を比較することで行う。提案システムでは階層的情報源としてインターネットディレクトリ、インスタンスとして Web

ページを用いることを想定している．Web ページのテキストからキーワード抽出を行い，キーワードベクトル間の類似度を測ることで Web ページ間の類似判定を行う．実験で用いた類似度については次節で説明する．

次に，ある階層  $C_1$  が持つインスタンス  $\{I_{C_1}\}$  と類似したインスタンス  $\{I_{C_1S}\}$  を集める．この中からさらに選ばれたインスタンスが，階層  $C_1$  に追加される．これを概念階層  $H_1, H_2$  が持つ階層全てに対して行う．具体的な選択方法については次節で説明する．

以上の手続きにより，元々共有しているインスタンスが無かった概念階層間にも共有インスタンスができ，統計量による概念階層の統合が可能になる．

### 3.4 システム概要

本研究では，情報源の概念階層としてインターネットディレクトリの分類体系を対象とする，階層的知識と内容的類似性を用いた *WebHical* システムを試作した．本システムの概要を図 3.3 に示す．

システムの処理の流れを説明する．まずインターネットディレクトリの概念体系を取得する．Open Directory の場合は，データが 1 つにまとめられておりダウンロード可能なので，それをそのまま使う．Lycos や Yahoo! の場合は，クローラで Web サイトを巡回して Web ページを取得し，そこから概念体系を取り出す．これら手に入れた概念体系は RDF 形式で保存される．(1)

次に，RDF に記述された URL を元に，クローラにより Web ページを取得する．今回，クローラには Larbin<sup>\*1</sup>を用いた．取得した Web ページから HTML タグを抜き取り，茶筌 [43] により分かち書きをして名詞と未知語だけをキーワードとして抜き出し，リポジトリへ格納する．(2)

格納された Web ページ間の内容的類似性の計算には GETA[1] を用いる．GETA は汎用的な検索エンジンシステムであり，文書をキーとして類似する文書を検索する連想検索などが高速に行える．類似度には SMART[57] の重み値を用いた．(3)

類似度により内容的類似性を計算できるようにした後で，3.3.2 節で説明したように概念階層 (インターネットディレクトリ) にインスタンス (URL) の追加を行う．まず，ある概念階層  $H_1$  の中の概念  $C_1$  が含む  $URLI_1$  のキーワードベクトルをキーとして，概念階層  $H_2$  の中から  $URLI_{C_1S_1}$  を取り出す． $URLI_{C_1S_1}$  を類似度で降順にソートし，その上位 10 個を追加候補  $URLI_{C_1S_2}$  とする．なお， $URLI_{C_1S_2}$  には  $H_1$  が最初から持っている URL は入らないようにする． $URLI_{C_1S_2}$  の中からさらに閾値以上の類似度を持つ  $URLI_{C_1S_3}$  を集め，これを追加 URL として概念  $C_1$  に追加する．(4)

---

\*1 <http://larbin.sourceforge.net/>



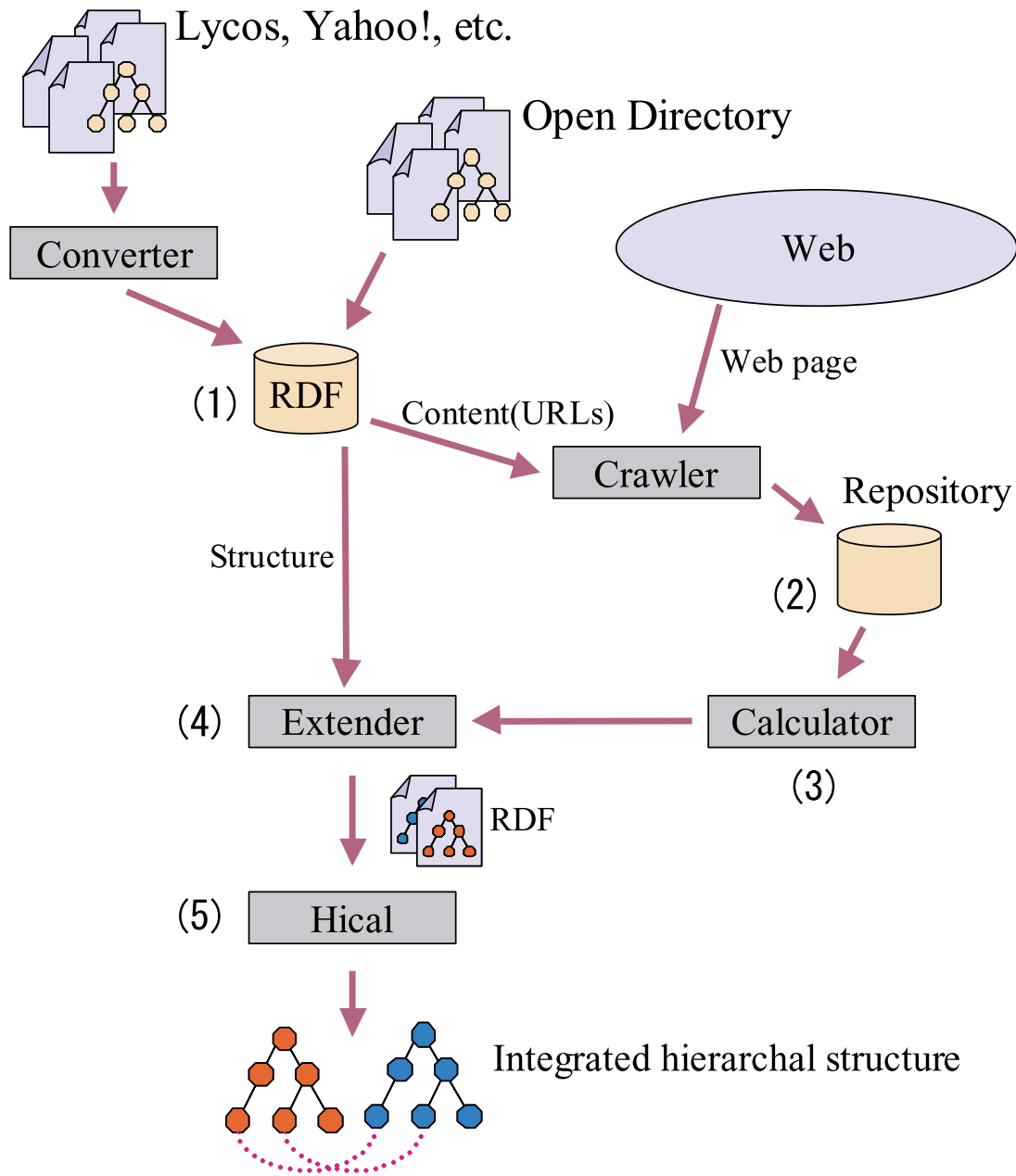


図 3.3 システム概要

最後に、追加処理を終えた概念階層  $H_1$  および  $H_2$  を *Hical* により統合する。(5)

## 3.5 実験

### 3.5.1 実験設定

実験を行う対象として、インターネットディレクトリの OpenDirectory Japan<sup>\*2</sup>と Lycos Japan<sup>\*3</sup>の分類体系を情報源の概念階層として用い、そこに含まれる外部リンク (URL) をインスタンスとして用いた。実験には、OpenDirectory から提供されている RDF データ (2002 年 9 月に生成) と、2000 年の 8 月から 9 月にかけて収集した LYCOS Japan の HTML ページを RDF に変換したものを用いた。ただし、インターネットディレクトリの全データは大き過ぎるので、その中の一部 (OpenDirecotry の文学ディレクトリと Lycos の Literature ディレクトリ) を用いた。

### 3.5.2 インターネットディレクトリの解析

*Hical* では、異なる階層構造の中のある階層間において共有している URL が多くある場合に、その概念は類似関係にあると判断した。今回新たに提案した手法は、異なる階層構造の中のある概念間において類似している URL が多くある場合に、その概念は類似関係にあると判断するものである。ここで問題になるのは、そもそも OpenDirectory や Lycos といった人手で作られた分類体系は、今回用いる内容的類似性に沿って作られた物であるのかどうかという点である。そこで予備実験として、同じ概念に入っている URL 間の内容的類似性を、今回用いた手法で見つけ出すことができるのかどうかを検証する。

予備実験は、次の手順で行った。まず階層構造を一つ用意し、それが持つ URL を GETA に検索対象として登録する。次に、3.3.2 節で解説した手法によって、階層構造に URL を追加する。この時、URL が元々入っていた階層の中に追加された場合に内容的類似性に基づく URL 追加が成功したとみなす。なお、キーと同じ URL は類似検索の結果から省くようにする。

図 3.4 は、OpenDirectory の文学カテゴリと Lycos の文学カテゴリとをデータとした、それぞれの場合の内容的類似性に基づく分類の正答率を示している。正答率とは、追加した URL が元々入っていた階層に追加された割合を指す。

閾値を 0.8 まで挙げた場合の正答率は OpenDirectory で 0.55、Lycos では 0.17 となっている (図 3.4)。これはつまり前者は半分、後者に至っては 8 割の追加 URL がここでいう類似に当てはまらないデータということであり、補完データとして適切なものであると

---

<sup>\*2</sup> <http://www.dmoz.org/>

<sup>\*3</sup> <http://www.lycos.co.jp/>

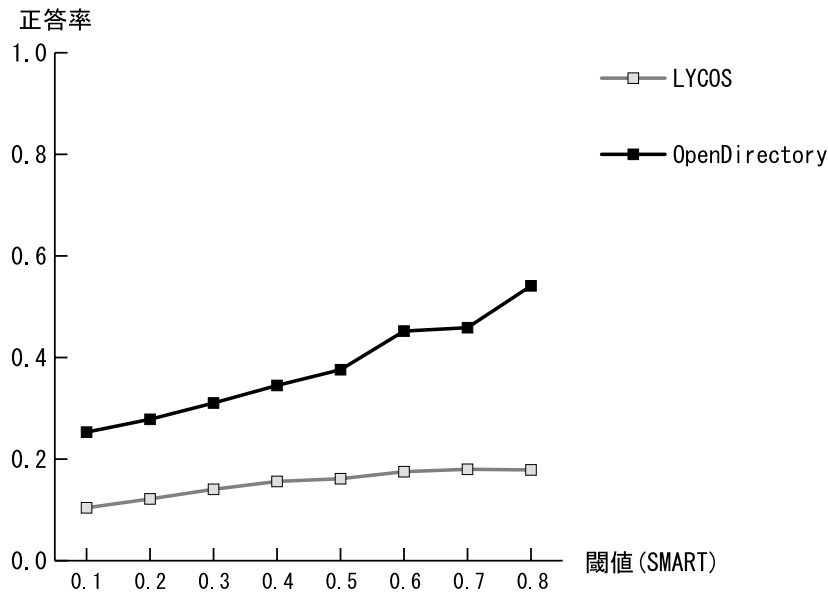


図 3.4 内容的類似性による分類の正答率

はいえない。だが、OpenDirectory、Lycos 共に類似判定の閾値を高く設定することで追加 URL の精度が向上していることから、内容的類似性と階層的分類にはある程度のある関係があることを示していると言える。

本研究での提案手法は、言語処理技術のみを用いたものではなく、人手によって作られた階層的知識と言語処理技術を併用したものである。次章では階層的知識と内容的類似性を併用した場合の実験結果について解説する。

### 3.5.3 インターネットディレクトリの統合

インターネットディレクトリの結合を Hical で行った。インターネットディレクトリには、事前に前章の手法でページを追加している。対象としたデータを表 3.2 に示す。内容的類似性を導入することで、246 個のリンク (URL) が新たに共有されている。

表 3.2 実験に用いたデータ

	OD	Lycos
階層数	43	186
URL 数	454	1119
共有 URL 数	92	
追加共有 URL 数	246	

*Hical* を用いて発見した類似階層ペアの内訳を表 3.4 に、発見された類似階層ペアのうち、内容的類似性により類似ページを追加することで初めて発見されたものを表 3.3 に示す。

表 3.4 は内容的類似性を導入した結果、新たに発見された類似階層ペアを示している。新たに発見された類似階層ペアの中で「OpenDirectory/文学/作家/小説家/横溝正史」と「Lycos/Literature/mystery/yokomizo」に注目してみる。これらディレクトリを詳しく調べてみると、元々は同じ URL を一つも持っていない関係であることがわかった。このペアは内容的類似性に基づいて URL を追加したことで初めて発見できたものであり、これは典型的な階層的知識と内容的類似性を併用することの効果といえる。

表 3.3 内容的類似性の導入で新たに発見された類似階層ペア

OpenDirectory/文学	Lycos/Literature
作家/小説家/横溝正史/ 俳句・和歌・川柳/ 小説/ファンタジー/	mystery/yokomizo/ haiku/ juvenile/

表 3.4 は *WebHical* と *Hical* それぞれの手法で発見された類似階層ペア数を示している。表を見ると、*WebHical* によって見つけた類似階層ペア数は 14 であるのに対し、*Hical* で見つけた類似階層ペア数は 37 と 2 倍以上の差があることがわかる。内容的類似性の導入により本来見つけ出せなかった類似階層ペアを見つけ出せたが、同時にいくつかの *Hical* で発見できた類似階層ペアが見つからなくなっている。これは、人手によって作られた階層的知識と言語処理により見つけ出せる内容的類似性の差が影響しているものと考えられる。

表 3.4 発見された類似階層ペア数

	類似階層ペア数
<i>WebHical</i>	14
<i>Hical</i>	37
<i>WebHical</i> <i>Hical</i>	9

3.5.2 節の予備実験で得られた結果から、今回用いた手法で導かれた内容的類似性による分類はあまり良い精度が出ないことがわかっている。これはつまり、内容的類似性に基づいて階層的情報源にインスタンスを追加するということは、多くの場合において、整然とインスタンスが配置された階層的情報源に対してでたらめな配置でインスタンスを追加

するに等しい．これは結果として階層的情報源の作者が暗黙に持っていた分類ルールをあやふやにしてしまう．あやふやな分類の概念階層間に対応付けを行った場合，そもそもがあやふやであるため類似する階層ペアの数は少なくなる．これが *WebHical* によって発見できた類似階層ペア数が *Hical* よりも少ない原因と考えられる．

*WebHical* は，内容類似性を用いることで *Hical* では発見できなかった類似階層ペアを発見することが目的である．*WebHical* は内容類似性の計算精度によって発見能力が大きく変わる可能性があるが，誤った類似階層ペアを発見してしまうことは階層的知識の利用により防がれるため，*WebHical* と *Hical* の併用により類似階層ペアの発見能力を高めることは可能であるといえる．これは本研究で新たに提案した *WebHical* の有効性を示すものである．

## 3.6 Hical-NB

### 3.6.1 Hical の問題点

*Hical* では，インスタンスが分類しているカテゴリの類似性だけを利用しており，インスタンスが持っている属性は，全く利用せずに概念体系を統合する．一方，Enhanced Naive Bayes [3] では，インスタンスが持っている属性と元カタログの分類の両方の情報を使って統合をする．*Hical* では Enhanced Naive Bayes で利用している分類情報のみを利用して，Enhanced Naive Bayes を上回る性能を出していると言えるが，Enhanced Naive Bayes と同様にインスタンスの属性も利用することができれば，さらに性能を向上させることが可能であると考えられる．

*Hical* が持つ問題点について簡単に説明する．例えば，図 3.5 の様な階層的情報源があったとする．階層的情報源 A の”Automobile/Car/Toyota/”が，階層的情報源 B のどの概念階層に対応するかを *Hical* によって求めた場合，”Auto/Car/”までは判別が可能であるが，それ以下は判別できない．これは，そもそも”Automobile/Car/Toyota/”に対応する概念階層を階層的情報源 B が持たないためである．

概念階層の対応を考えるという点では，ここまで判別できただけで十分であるが，インスタンスの共有を考えるとこのままでは問題がある．”Automobile/Car/Toyota/”のインスタンスとして”Prius”というハイブリットカーと，”DYNA”というディーゼル車があったとすると，これらインスタンスはそれぞれ階層的情報源 B の”Vehicle/Car/Hybrid/”と”Vehicle/Car/Diesel/”に分類されるべきである．しかし *Hical* では”Automobile/Car/Toyota/”に対応するカテゴリとして”Auto/Car/”までしか発見できないため，”Prius”も”DYAN”も共に”Vehicle/Car/”に分類されてしまう．この問題は，*Hical* の拡張版である *WebHical* にも該当することである．

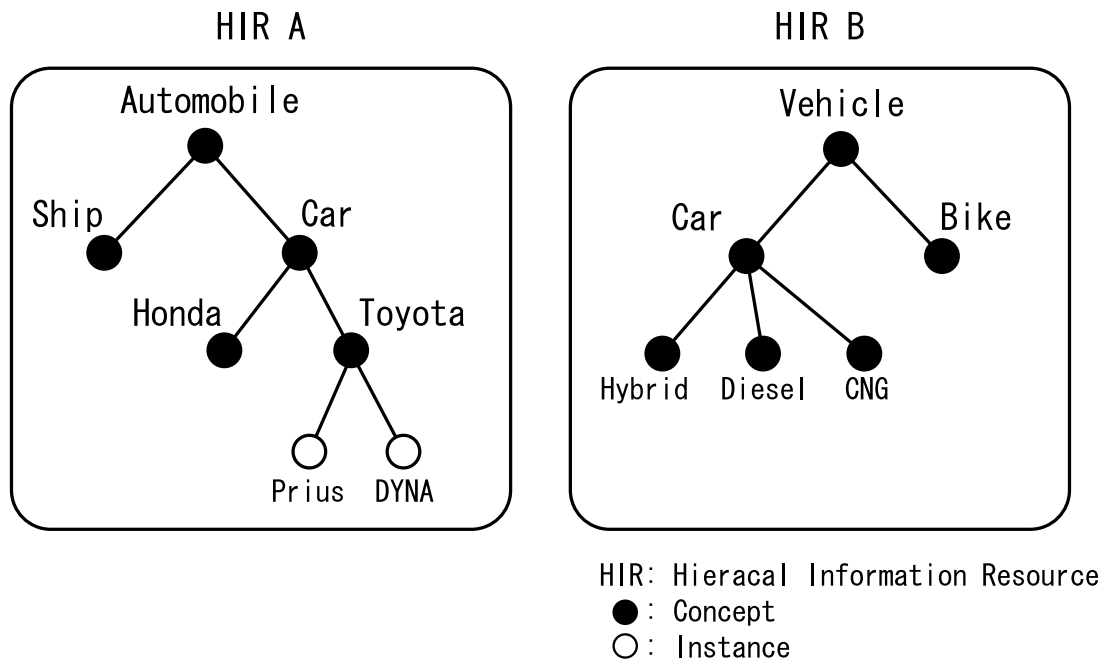


図 3.5 Hical では適切な分類が困難な例

### 3.6.2 Hical-NB

*Hical-NB* では、まずカテゴリ間の類似度を *Hical* により計算し、次にインスタンスがどのカテゴリに分類できるかを Naive Bayes により求める。これにより、正確に対応するカテゴリが無くとも分類することができる。

前節の図 3.5 において、階層的情報源 A の”Prius”というインスタンスを階層的情報源 B のどのカテゴリに分類するかという問題があった場合を考える。まず *Hical* によって”Automobile/Car/Toyota/”に類似するカテゴリとして”Vehicle/Car/”が発見される。次に”Vehicle/Car/Hybrid/”、”Vehicle/Car/Diesel/”と”Vehicle/Car/CNG/”の4つのカテゴリを分類先候補とした分類器を用いて、”Prius”がどのカテゴリに分類されるかを定める。分類器が正しく動けば、”Vehicle/Car/Hybrid/”に分類される。同様にして”DYNA”というインスタンスは、”Vehicle/Car/Diesel/”に分類される。

*Hical-NB* の分類器にはどのような手法を用いることも可能であるが（例えば SVM など）、今回は分類器学習手法として Naive Bayes を、分類器学習プログラムには McCallum 博士によって開発された Rainbow [45] を用いた。*Hical-NB* を [29] にて *Hical* と Naive Bayes それぞれとの分類精度の比較を行った結果、*Hical* および Naive Bayes よりも高い精度で分類できることが確認された。

## 3.7 議論

### 3.7.1 類似階層ペアと情報共有

*Hical* およびそれを元にした *WebHical*, *Hical-NB* が見つける概念階層間の関係性は、インスタンスの類似性を元にしたものである。このようにして発見された関係性を用いた情報共有が有効であることは、2章の被験者実験によりわかった。本節ではさらに、情報共有のタイプという点からインスタンスの類似性を元にした情報共有の有効性について考察する。

本研究が想定する情報共有の目的は、他の情報源が収集した情報を共有することにより自身にとって価値のある情報を容易に取得することである。このような観点で情報共有を捉えると、情報共有には発散指向と収束指向があると考えられる。発散指向の情報共有とは、意外な情報を発見や、一見すると自分の興味とは異なるが実は関わりのあるような情報の発見を支援するものである。発散的思考 [58] を促す情報の発見を支援するような情報共有とも言える。対して収束指向の情報共有とは、自分の興味と一致する情報を集めていくものである。興味ごとに分類して情報を収集するようにすれば、利用者の興味の体系に合わせて情報を整理していくことを支援しているとも言える。

本研究が目的としている情報共有は、個人が持つ概念体系（概念階層）に沿った情報収集を支援するものであり、収束指向の情報共有であると言える。そして、情報源が持つ情報の類似を用いた *Hical* のような手法による情報共有支援は発散指向というよりはむしろ収束指向であり、本研究の目的に合致していると言える。

なお、統計量による概念階層間の関係性の計算では、インスタンスが全く同じである概念階層間で最も高い値が出る。これは同一概念の発見という観点では正しいが、新しい情報を全く得られないという点で情報共有としては適切ではない。ただし、インスタンスが Web ページのように膨大であるとすると、概念階層内のインスタンスが完全に一致する可能性は低いと思われる。また、継続的な情報共有環境では、仮に現時点において概念階層内のインスタンスが完全に一致したといっても、将来的には概念階層に合致する異なるインスタンスをそれぞれの情報源が収集する可能性があると考えられる。これらのことから、本研究が想定する状況では先ほど述べた様な問題点は重要ではないと考えられる。

### 3.7.2 概念階層間の関係について

今回は、いわば同じ話題について情報を交換し合うような情報共有モデルを想定した。そのため、類似性のある階層を見つけ出すことが重要であるとした。しかし、例えばある話題に対して、対極に位置するような話題や補完関係にあるような話題、もしくはわずか

な共通点があるがまったく別の話題も、情報共有という点では価値のある話題といえる。そのような話題間での情報共有を行おうと考えた場合は、今回のような類似性とは異なった指標が必要になると考えられる。

類似性とは異なる関係の一つとして、上位下位関係がある。*Hical-NB*は、十分に類似した概念階層が無い場合に概念階層内の情報ごとに適切な概念階層を探そうとするもので、上位下位関係にある階層しかなかった場合における有効な解決策の一つである。ただし、2章の実験結果では階層構造単位での情報共有の方がより効果的であると示しており、仮に相手の持っている階層構造が自分には無いものであるならば、自らの概念階層に対応する新たな階層を作る方が良いと言える。これは、構造化された情報をもとにその構造を再構築することであると捉えられるが、そのようなことを行う技術として Formal Concept Analysis (FCA) [53] と呼ばれるものがある。FCAは情報の統合やクラスタリングなどにおいて統計的な手法と補完関係にある技術として注目されており、本研究で提案する情報共有手法の改善にも有効であると考えられる。

### 3.8 まとめ

本章では、階層的知識と内容的類似性を使った階層的情報源の結合手法を提案し、それを実装した *WebHical* システムを試作した。さらに階層的情報源としてインターネットディレクトリを用い、提案システムの実験を行った。結果、内容的類似性の導入が階層的知識のみを用いた場合には欠落してしまう部分を補完する効果が見られた一方で、内容的類似性による不正確な情報がかえって正しい発見を阻害してしまうという場面が見られた。

3.5.2 節にて内容的類似性を用いて人手により同一のカテゴリに分類されたページ間の関係を抽出できるか検証した結果、今回用いた手法では高い精度の抽出は行えないという結果となった。人手により作られた階層的知識を機械的処理により自動生成した内容的類似性によって模倣することは不可能であったが、後者により前者を補完するという点に関しては、3.5.3 節の実験結果より、その効果が確認できた。人手で作られた知識をより広範に活用するために、内容的類似性に基づく階層的知識の自動補完は有効な手段であったと言える。



## 第4章

# 仲介による共有ネットワークの拡張

本章では、情報共有の相手をどのようにして見つけるかという点について議論する。より良い情報共有を実現するためには共有相手の選択が重要である。しかし、一般に膨大な数の共有相手となる情報源の中から適切なものを見つけるのは困難である。そこで本研究では、友人に友人の友人を紹介してもらうという私たちが日常生活において行っている新しい知り合いの獲得手法を模倣した、近傍仲介法を提案する。近傍仲介法はボトムアップな共有相手発見手法であり、トップダウンなマッチングシステムとは異なった特徴を持つ。

まず提案する近傍仲介法について述べ、次にシミュレーション実験の概要を述べる。シミュレーション実験では、ランダムデータと2章で用いた実データを利用して提案手法の有効性を検証する。最後に提案手法の特徴について考察を行う。

## 4.1 はじめに

人の間で行なわれる情報交換は、情報過多への効果的かつ現実的な解決策の一つである。本研究が提案する情報共有による情報収集支援は、このような人が日常において行っている情報収集技術の一部をシステム化したものであるとも言える。

人は、情報の収集やフィルタリング、組織化などを行なってくれる知的エージェントであるといえる。我々は自分のパーソナルネットワークで繋がった人たちと情報交換することで、価値ある情報を得ている。より良い情報交換の環境を得るためには、より良いパーソナルネットワークを作る必要がある。しかし個人個人が持つパーソナルネットワークは、個人の知りうる範囲で適切であるかもしれないが、全体から見ればより適切なネットワークが存在しうる。そこで本研究では、ネットワークを自律分散的に最適化する手法（近傍仲介法）を提案し、実験により評価する。これはすでに築かれているネットワークを元に逐次的に拡張していくことで、最適化する手法である。

## 4.2 情報交換相手との関係発見

まず情報交換は、その形態の違いにより2種類に分けられる。一つはメールのような一対一型、もう一つはメーリングリストや掲示板のような一対多型である。我々がメールを使うのは大抵の場合、すでに友達であるかもしくはホームページか何かで見知った相手と情報交換をするときである。メールを用いた情報交換では、相手の選択や理解は比較的容易である。対してメーリングリストや掲示板では、情報共有の場を介して多くの人同士で情報交換が行なわれる。この場合、我々は情報交換の新しい相手を比較的容易に獲得することができる。

前者の問題点は、新しい情報交換相手の獲得が困難であることである。このタイプでは、既に知っている相手としか情報交換が行なえない。一般に、人が普通に知りうる人の数には限界があるため、情報交換にとって良い環境を築くのは難しい。対して後者は、新しい情報交換相手の選択が困難であるという問題点がある。多くの参加者の中から自分にとって良い情報交換のパートナーを見つけ出すのは、とても知的かつ時間のかかるタスクであり、人にとって大きな負担となる。

より良い相手と情報交換をしたいならば、我々は多くの相手と出会い、そして選択しなければならない。ゆえに情報交換を支援するシステムには、新しい相手と出会い選び出さなければならないという潜在的な問題を解決する方法を持つ必要がある。

我々は実生活において、しばしば友達の紹介を介して新しい友達を得る。紹介された友達は、出会う前には当然知らないため相手を選択することはできない。しかし、両者を

知っている共通の友達がすでに選択を行った上で紹介を行ってくれているため、ある程度の選択を経た相手と出会うことができる。我々は、このような友達が新しい相手とのマッチメイカーになるというアイデアを用いた、新しいネットワーク最適化手法を提案する。この手法の説明を次節にて行う。

### 4.3 近傍仲介法

本節では、我々が提案する自律的構築が可能なネットワーク最適化手法の紹介を行なう。この手法は実世界における手法を模倣したものである。我々はこの手法を近傍仲介法 (Neighborhood Matchmaker Method) と呼んでいる。

近傍仲介法を説明する前に、近傍仲介法の適用対象となるネットワークのモデルを定義しておく。まず人をノードとし、人と人との間の情報交換がなされている関係をパスとする。次に、全てのノードは、ノードとノードの関係の計算が可能であると仮定する。計算された関係の強弱を示す値を関係値と呼ぶ。

情報交換にとってより良い環境を作るには、パーソナルネットワークの最適化が必要である。このモデルにおける最適なネットワークとは各ノードにおいて関係値が高いノードと繋がっているネットワークである。全てのノードの情報を知っている場合、ノード間の関係を総当りで計算することで最適なネットワークの発見ができる。以下ではこのような一括計算による最適化手法をサーバ型計算法と呼び、それによって生成された最適なネットワークを最適解と呼ぶ。ここではこのような全ての情報を集約するサーバがなく、個々のノードにおける計算のみで解を求める場合を考える。個々のノードは自分が繋がっているノードしか知り合えないので、最適なネットワークを形成するにはそれ以外の新しいノードを知る方法が必要である。近傍仲介法では、隣接ノードが互いに推薦しあうことで、自律分散的にネットワークの最適化を行なう。

近傍仲介法における、各ノードの振る舞いの例を図 4.1 に示す。まず最初に各ノードが自身と隣接しているノード間の関係を計算する (図 4.1-1)。良い関係にあるノードのペアを発見した場合、それぞれのノードにお互いを推薦する (図 4.1-2)。推薦を受け取ったノードは新しくパスをはるかどうかが決める。推薦を受け入れると推薦されたノードが隣接ノードとして追加される (図 4.1-3)。これらの動作の繰り返しにより、ネットワークは改善されていく。この手法は、前節で述べた友人関係の推薦を模倣したものである。この特徴は自分の友達が選択したものが推薦される点である。このため、推薦されたノードは比較的適切なものであることが期待される。

次節以降において、この手法のシミュレーション実験の解説を行なう。

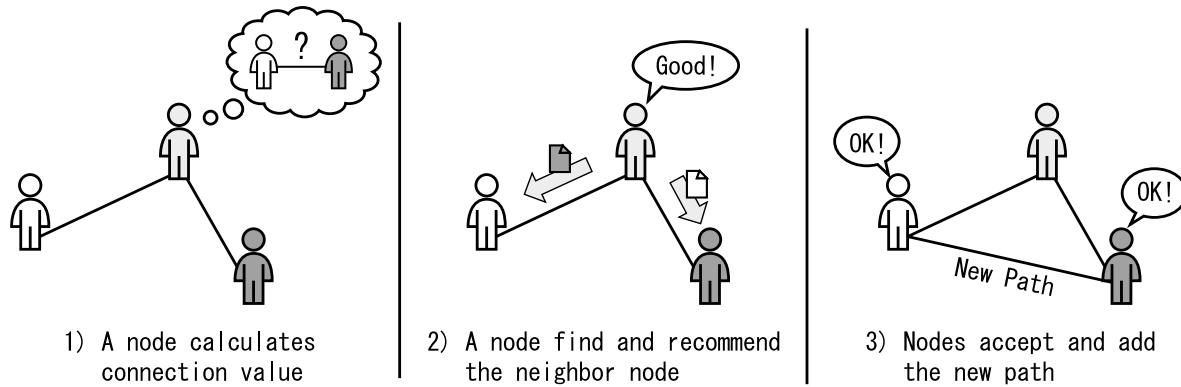


図 4.1 ノードの振る舞い

## 4.4 実験概要

### 4.4.1 シミュレーションの概要

本実験の目的は、提案する手法がどの程度パーソナルネットワークの最適化に有効であるかを調べることにある。そこで実データと仮想データを用いて、近傍仲介法の最適化をシミュレーションにより評価する。

情報共有における最適なネットワークというものを定義するのは難しい。ここでは近傍仲介法というボトムアップな手法がどの程度の能力を持っているかを測ることが目的なので、トップダウンに計算可能な最適解を仮想的に定義し、それに対する最適化能力を計測する。

まず、情報共有ネットワークのパスには関係値（重み）があると考える。これは、2章において提案したカテゴライズ近似度のような情報共有相手としての適切さを示す指標である。このような指標が全てのノード間において計算可能であるとするならば、ネットワークを構成する全てのパスの関係値が計算可能であり、パス数を固定した場合には、パスの関係値の総和が最大であるネットワークというものを定義することができる。このネットワークはネットワークを構成する全ノードのデータを一カ所に集めてトップダウンに計算することで求める事が可能であり、逆に、近傍仲介法のような局所最適化手法では求めるのが困難である。今回のシミュレーションでは、このようなネットワークを最適なネットワークであると定義し、局所最適化手法である近傍仲介法によってどれだけ最適解に近いネットワークを求めることができるかを調べる。

近傍仲介法では、各ノードがどのようなルールに従ってエッジの張り替えをするかが重要になる。今回のシミュレーションでは、最適解が構成するパスの関係値の総和が最大

であるネットワークと定義したので、より関係値の大きいパスを張ることが目的となる。そこで、各ノードは、より関係値の大きいパスを張り、同時に関係値の小さいパスを切るという単純なパス張り替え戦略を取るように決める。つまり、推薦されたノードとの関係値が、隣接ノードとの関係値の中で最低のものよりも高い場合、最低の関係値を持つ隣接ノードとのパスを切り、新たに推薦されたノードとのパスを張る。なお、パスを張ると同時に他のパスを切るのは、ネットワークを構成するパス数を一定に保つためである。

図4.2はこのシミュレーションのフローチャートである。まず最初に各ノードにデータを与える。今回は、10次元の特徴ベクトルとWWWブックマークの2種類のデータを用いた。次に、ノード間にランダムにパスを張る。パス数はパラメータとして設定する(1)。

ネットワークの再構築は以下の手順で行なう。まず、全てのノードは自分と隣接する(パスが張られている)ノード間の関係値を計算する。計算を行なったノードは、各隣接ノードに対して最も高い関係値のパスを張ることができるノードをそれぞれ一つ選び、そのノードに推薦する(2)。次にパスの張替えを行なうノードをランダムに選択する(3)。選択されるノードは1ターンに1つだけで、そのノードはかならず新しいパスを生成可能でなくてはならない。選ばれたノードは隣接ノードから推薦されたノードの中から最も関係値の高いものを選び、新しくパスを張る。そして既に張っているパスの中から最も関係値の低いものを選び、そのパスを切る(4)(5)。1ターンに1つのパスが新しく生成され、1つのパスが削除されるので、全体としてはパス数は常に一定に保たれる。

各ターンに選択されるノードは新しいパスを生成可能でなくてはならない。そして、新しいパスの関係値は、既に自分が持っている削除候補となるパスよりも高くないでなくてはならない。ネットワークを構成する全てのノードが新しいパスを生成できなくなれば、そのネットワークは収束したとみなし(6)、シミュレーションは終了する(7)。

図4.3はパスの張替えの例を示している。図中でノードAはすでにノードB、C、D、Eとのパスを持っている。ノードAが新たに生成可能なパスは、ノードBまたはFがマッチメイキングしてくれたノードEまたはGとのパスである。ノードAにとってノードEおよびGは未知のノードであり、それらと自身とのパスの関係値は計算不可能である。だが両ノードを知るマッチメイカーであるノードBまたはFがそれぞれ関係値を計算しており、ノードAには伝えられている。この値を元にノードAは新しいパスの選択を行なう。

まず、ノードAはすでにあるパスのなかで最も関係値の低いパスを選ぶ。この例では、ノードBとのパスは0.2で最も低い。次にマッチメイカーが推薦するノードの中から最も関係値の高いパスを持つノードを選ぶ。この例では、ノードGとのパスとなる。ノードBとの関係値よりもノードGとの関係値の方が高いので、このパスの張り替えは成立する。具体的にはノードBとのパスは削除され、新たにノードGとのパスが生成される(図4.3-右)。なお、パスの追加および削除はパスの一端のノードの判断によって行なわ

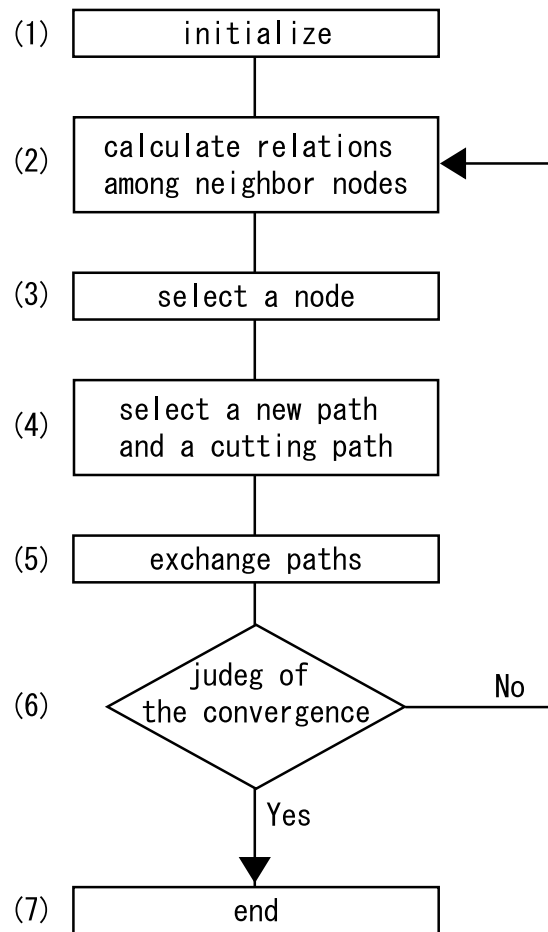


図 4.2 シミュレーションの流れ

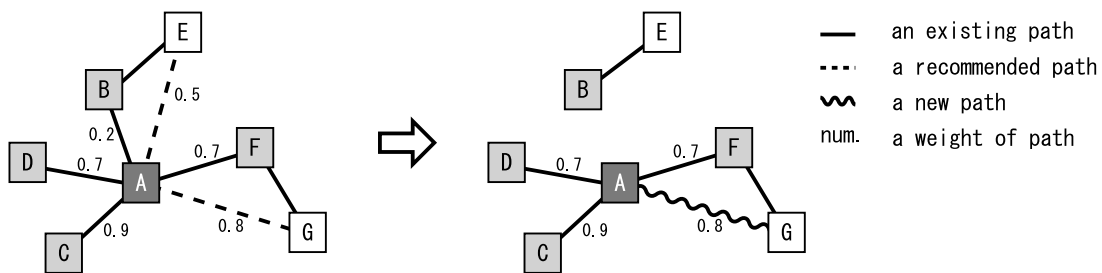


図 4.3 パスの張り替え

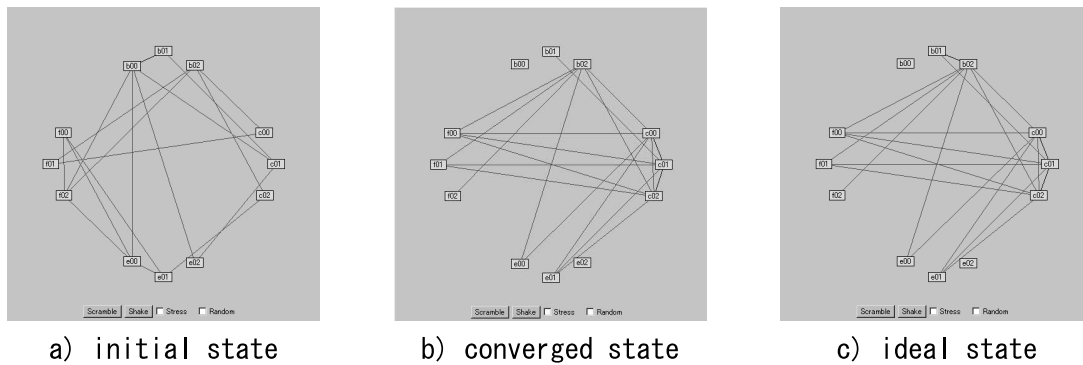


図 4.4 各状態のネットワーク図

れる．そのため，場合によってはどこからもパスの無い孤立ノードが生まれる可能性もある．本シミュレーションでは，いったん孤立したノードが新たにパスを張る方法はない．

我々はシミュレーション結果を可視化するために，Java アプレットを実装した．図 4.4 がシミュレーションの結果を Java アプレットによって可視化したものである．それぞれ，初期状態 (a)，収束状態 (b)，最適解 (c) のネットワークを示している．図 4.4-b(converged) のネットワークが図 4.4-a(initial) よりもずいぶん図 4.4-c(ideal) に似ていることがわかる．

#### 4.4.2 評価値

全てのノード間の関係値が一箇所で計算可能であるならば，関係値において最適なネットワークというものを作ることができる．そこで，サーバ型計算法で計算した最適なネットワークと提案手法で最適化したネットワークと比較することで，我々の提案する手法の有効性を評価する．評価は二つの評価値を用いて行う．一つは被覆率という指標で測るもので，これは最適解に対してどの程度似たようなネットワークが提案手法によって作られているかを示す割合である．もう一つは到達率という指標で，これはノード間の関係値の総和における割合である．ネットワークがどれだけ関係値の高いパスから構成されているかという指標となる．被覆率は形式的に二つのネットワークの類似を示すのに対し，到達率は実質的に類似していることを示している．被覆率と到達率は直接関係していないので，被覆率が悪くても到達率が良いことや，その逆もありうる．被覆率，到達率はそれぞれ以下の数式により求められる．なお，最適解の被覆率，到達率はともに 1.0 である．

$$\text{被覆率} = \frac{|\{P_{current} \cap P_{best}\}|}{N}$$

$p$	: パス
$N$	: パス数
$\{P\}$	: パスの集合
$\{P_{best}\}$	: 最適解のパスの集合
$\{P_{current}\}$	: その時点でのパスの集合
$f(p)$	: パスの関係値

表 4.1 パラメータ設定 (仮想データ)

ノード	10次元特徴ベクトル
ノード間の関係値	特徴ベクトルの内積
ノード数	10, 20, 30, 40, 50, 60, 70, 80, 90, 100
パス数	$\times 1, \times 2, \times 3, \times 4, \times 5$

$$\text{到達率} = \frac{\sum_{l=1}^N f(p_l | p_l \in \{P_{current}\})}{\sum_{m=1}^N f(p_m | p_m \in \{P_{best}\})}$$

## 4.5 実験

### 4.5.1 仮想データを用いた場合

本シミュレーションでは、ノード間の関係値を計算するためにデータが必要である。仮想データの場合、各ノードは乱数を用いて生成された10次元の特徴ベクトルを持つ。ノード間の関係値はノードがそれぞれ持つ特徴ベクトルの内積によって求められる。特徴ベクトルの要素は0.0~1.0の乱数である。ただし各要素は0.5の確率で0.0になるようにした。これはノード間の関係にばらつきを作るためである。

実験のために二つのパラメータを設定する必要がある。一つはノード数であり、もう一つはパス数である。本実験では、ノード数を10~100とし、パス数をノード数の1~5倍とした。実験の諸設定を表4.1に示す。シミュレーションは各設定ごとに10試行し、その平均値を結果とした。

図4.5の二つのグラフはともに横軸がターン数で縦軸が被覆率のグラフである。図4.5-上はパス数をノード数の3倍に固定したときの結果であり、図4.5-下はノード数を60に固定したときの結果である。



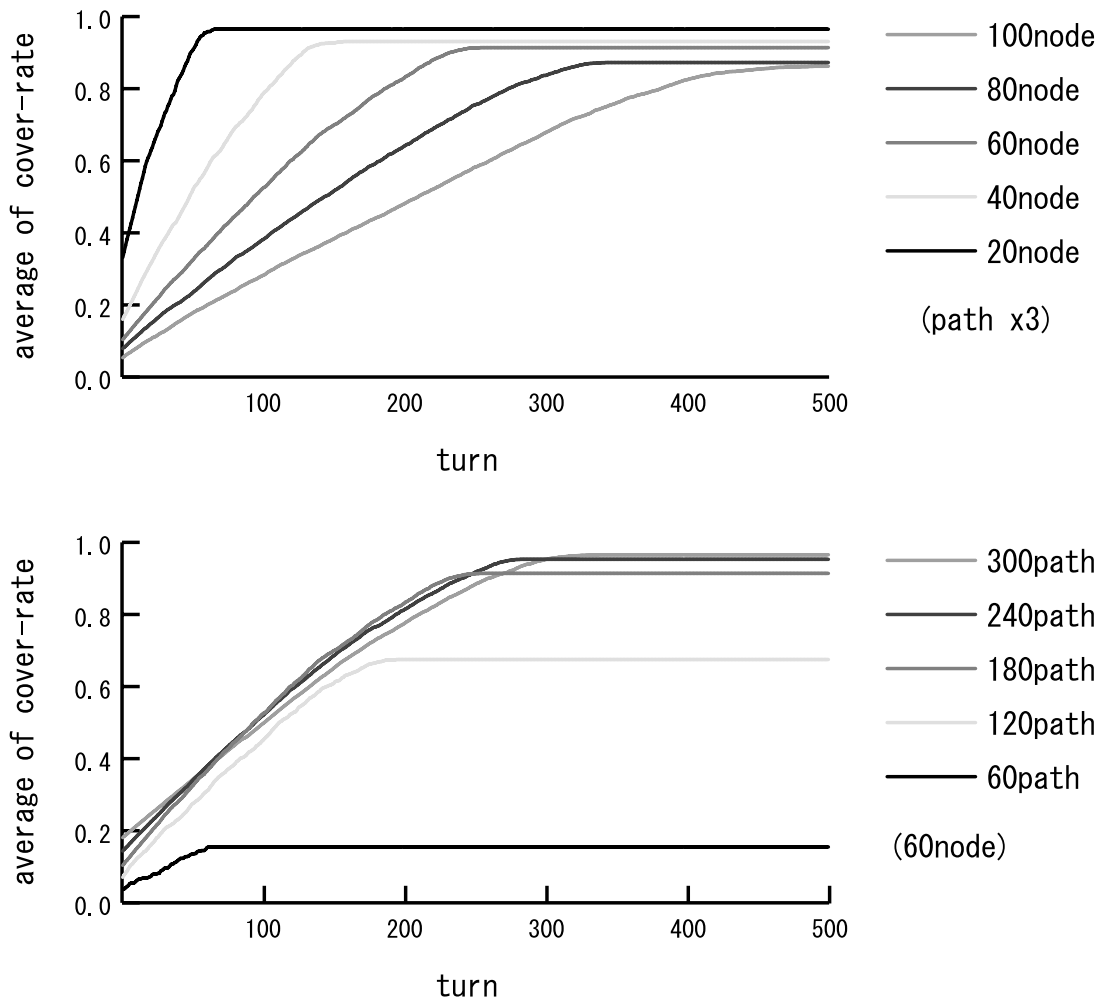


図 4.5 被覆率の変化 (仮想データ)

図 4.6 は横軸にターン数，縦軸に到達率をとったグラフである．先ほどの被覆率に関するグラフと同様，それぞれパス数とノード数を固定した場合の結果を示している．グラフを見ると，パス数やノード数が到達率に与える影響は，被覆率の場合よりも比較的小さいことがわかる．

図 4.7 および図 4.8 のグラフはは評価値とノード数およびパス数の関係を示している．図 4.7-a のグラフは初期状態，図 4.7-b のグラフは収束状態におけるノード数と被覆率の関係をそれぞれ示している．可能なパス数はノード数の 2 乗に比例するので，確率的に生成される初期状態における被覆率はノード数に反比例する (図 4.7-a)．図 4.7-b を見ると，この初期状態における差はパス数が 2 倍程度までは影響を受けるが，それ以上では問題がないことがわかる．同様に図 4.8 の二つのグラフはそれぞれの状態のノード数と到達率の関係を示している．到達率も初期状態においてノード数と反比例の関係にあるが (図 4.8-a)，その影響が現れるのはパス数が 1 倍程度のときだけであることが図 4.8-b からわ

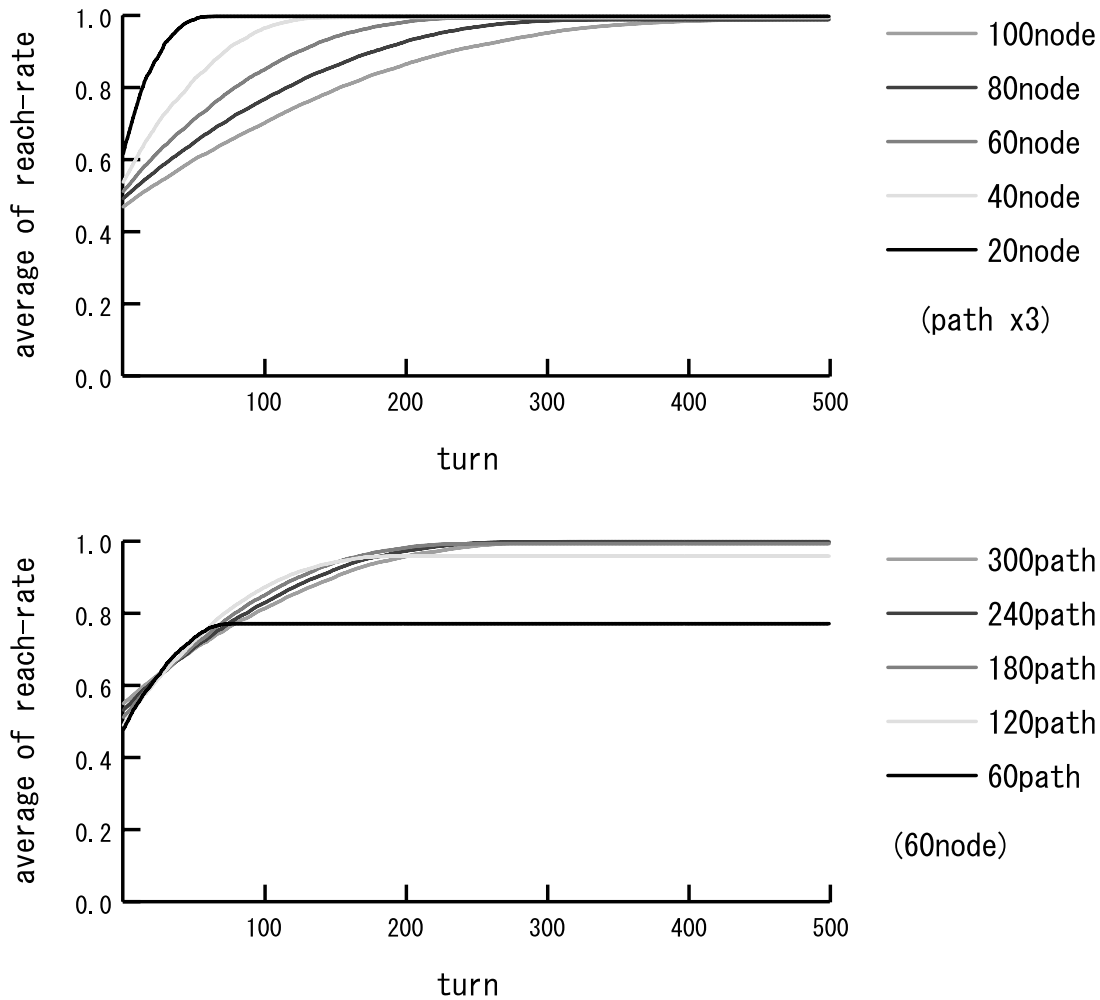


図 4.6 到達率の変化 (仮想データ)

かる．以上の様にパス数がノード数の 2~3 倍以上ある場合，ほぼ最適解に近い解に至ることがわかった．

図 4.9 のグラフは，パスおよびノード数と収束ターン数の関係を示している．グラフの横軸はノード数，縦軸は収束ターン数である．なお，本実験では全てのケースにおいて収束が確認された．グラフから，ノード数の増加にしたがって収束ターン数は線形に増加していることがわかる．

以上が仮想データにおける近傍仲介法の最適化の性能についての実験結果である．被覆率および到達率という二つの指標で近傍仲介法の最適化過程を評価したが，両指標において十分な結果が得られている．

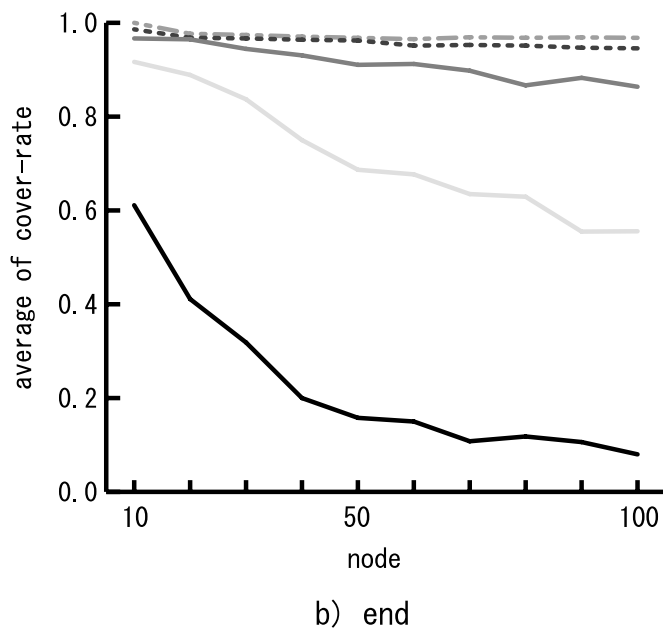
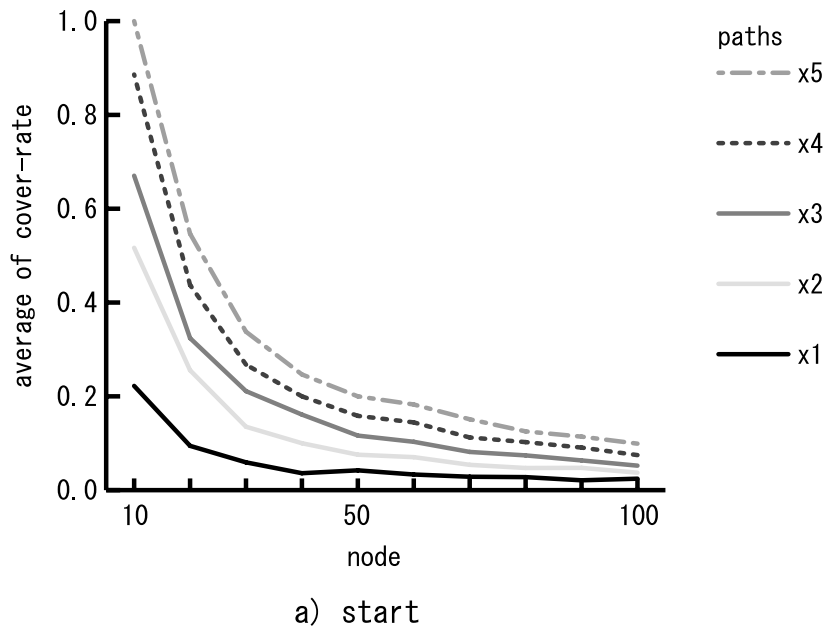
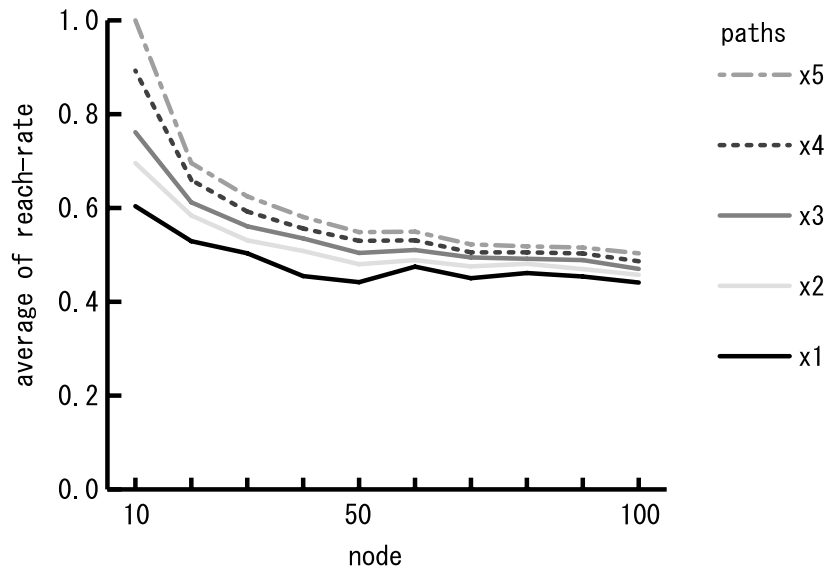
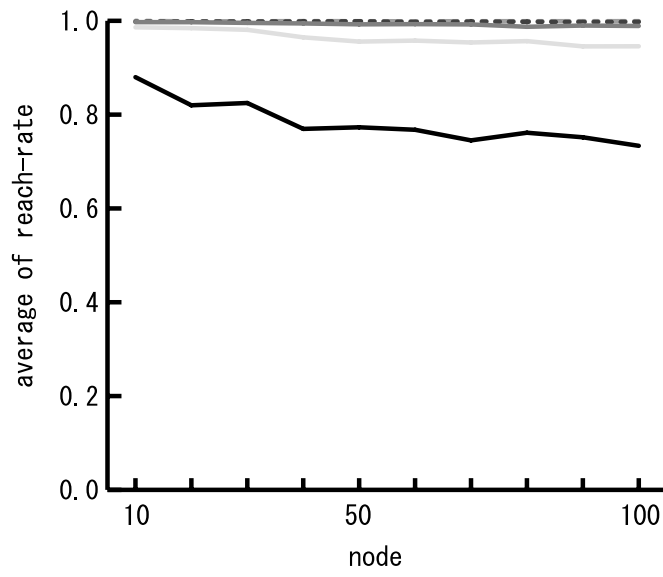


図 4.7 ノード数とパス数と平均被覆率



a) start



b) end

図 4.8 ノード数とパス数と平均到達率

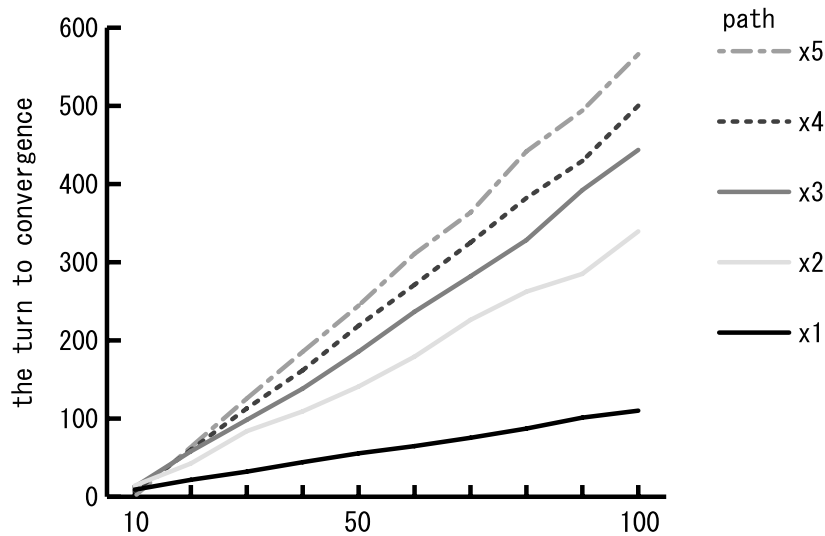


図 4.9 平均収束ターン数

#### 4.5.2 実データを用いた場合

前節では乱数で作った仮想データを用いてシミュレーション実験を行なった．次の実験では，実際に人が用いているデータを使って実験を行なう．実験データとしてここでは WWW ブックマークを用いる．WWW ブックマークはユーザが興味を持った Web ページを集めたデータであり，利用者の関心が表れているデータであるといえる．本実験では各ノードが WWW ブックマークをデータとして一つずつ持つ．ノード間の関係値はカテゴライズ近似度（2.7.1 節）を用いる．これは WWW ブックマーク中にある Web ページの類似と分類の仕方の類似を元にした，WWW ブックマークを持つ人同士の関係を示す値である．

我々は 12 人の被験者から WWW ブックマークを収集し（2 章で用いた物と同じ），システムによってそれぞれの WWW ブックマークの関係値（カテゴライズ近似度）を計算した．収集した WWW ブックマーク数が 12 であるため，ノード数は 12 に固定し，パス数を 12, 16, 20, 24 とした．シミュレーション実験の諸設定を表 4.2 に示す．シミュレーションは各設定ごとに 10 試行し，その平均値を結果とした．

図 4.10 は横軸にターン数，縦軸に被覆率および到達率をとったグラフである．仮想データを用いた場合と同様の傾向が現れている．

表 4.2 パラメータ設定 (実データ)

ノード	WWW ブックマーク
ノード間の関係値	カテゴライズ近似度
ノード数	12
パス数	12, 16, 20, 24

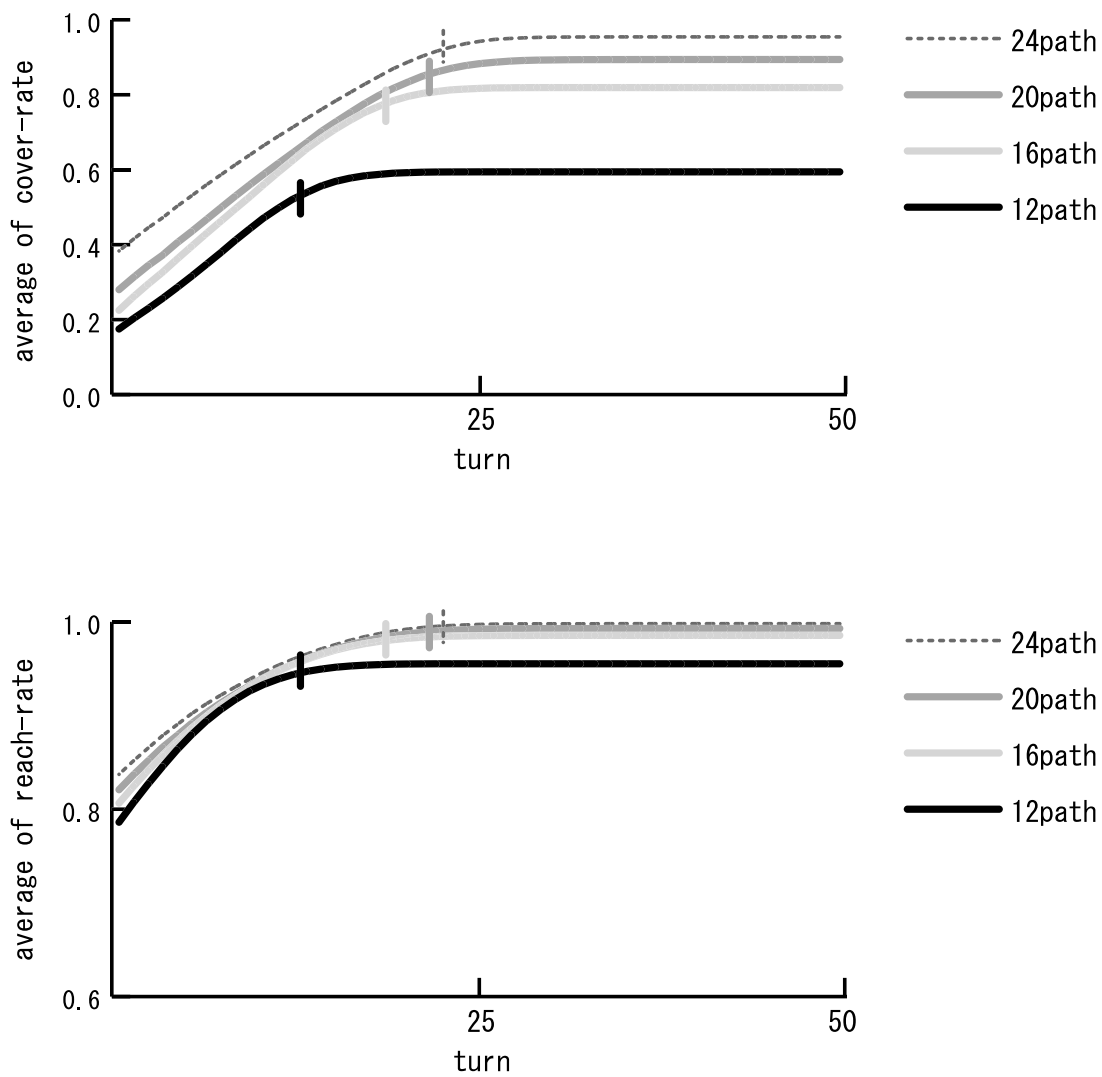


図 4.10 被覆率の変化 (実データ)

表 4.3 孤立したノードの数

パス数	100	200	300	400	500
収束解	161	107	29	8	5
最適解	116	16	8	4	4

### 4.5.3 ネットワーク構造の分析

これまで、被覆率と到達率という二つの指標を用いて近傍仲介法の分析を行ってきた。ここで、そのような評価値とは異なった観点からの分析を行う。具体的には、近傍仲介法がどのようなネットワークを生成しているのかを分析する。

図 4.11 はノード数が 100、パス数が 100～500 の時の、シミュレーション 10 試行分の収束解（近傍仲介法が生成したネットワーク）と最適解それぞれのネットワークにおける度数分布を示したものである。

パス数が多い 400 や 500 の場合にはほぼ同じ度数分布を描くが、パス数が少ない場合には異なった分布を描いている。特にパス数が 200 の時に大きく異なっている。ノード数が 100 でパス数が 200 の時には被覆率は低い (0.56) にもかかわらず到達率は比較的高い (0.95) ネットワークが作られている (図 4.7-b および図 4.8-b)。これは形式的には異なるが、実質的には類似したネットワークであることを示している。このネットワークは近傍仲介法特有の最適化の結果といえる。

100 個のノードに対して 200 本のパスなので、平均して 1 ノードあたり 4 つのパスを持つ。最適解では最頻値が 4 となっているが、収束解では最頻値は 4 ではなく、ヒストグラムもより平坦である。つまり必要以上に多くのパスを持っているノードが多数存在し、それによって必要なパス数を持っていないノードができていく。これは、局所的に優れたノードにパスが集まり、ネットワークが収束した（局所解に陥った）結果といえる。

もう一つの特徴として、次数が 0 であるノード（パスを一つも持たない、ネットワークから孤立したノード）が近傍仲介法において多いことが挙げられる。表 4.3 は、10 試行のシミュレーション結果から次数が 0 であるノードの数をパス数ごとに示したものである。

特にパス数が少ないときに次数が 0 であるノードが多い。近傍仲介法ではネットワークから孤立してしまったノードは新しいパスを得るチャンスを失う。最適解に似たネットワークを作るという点では、パス数が少ない場合に孤立したノードを多く作ってしまうことが近傍仲介法にとって問題であることがわかる。

表 4.4 は、各ノードの初期状態におけるパス数と収束状態でのパス数の相関係数を示している。

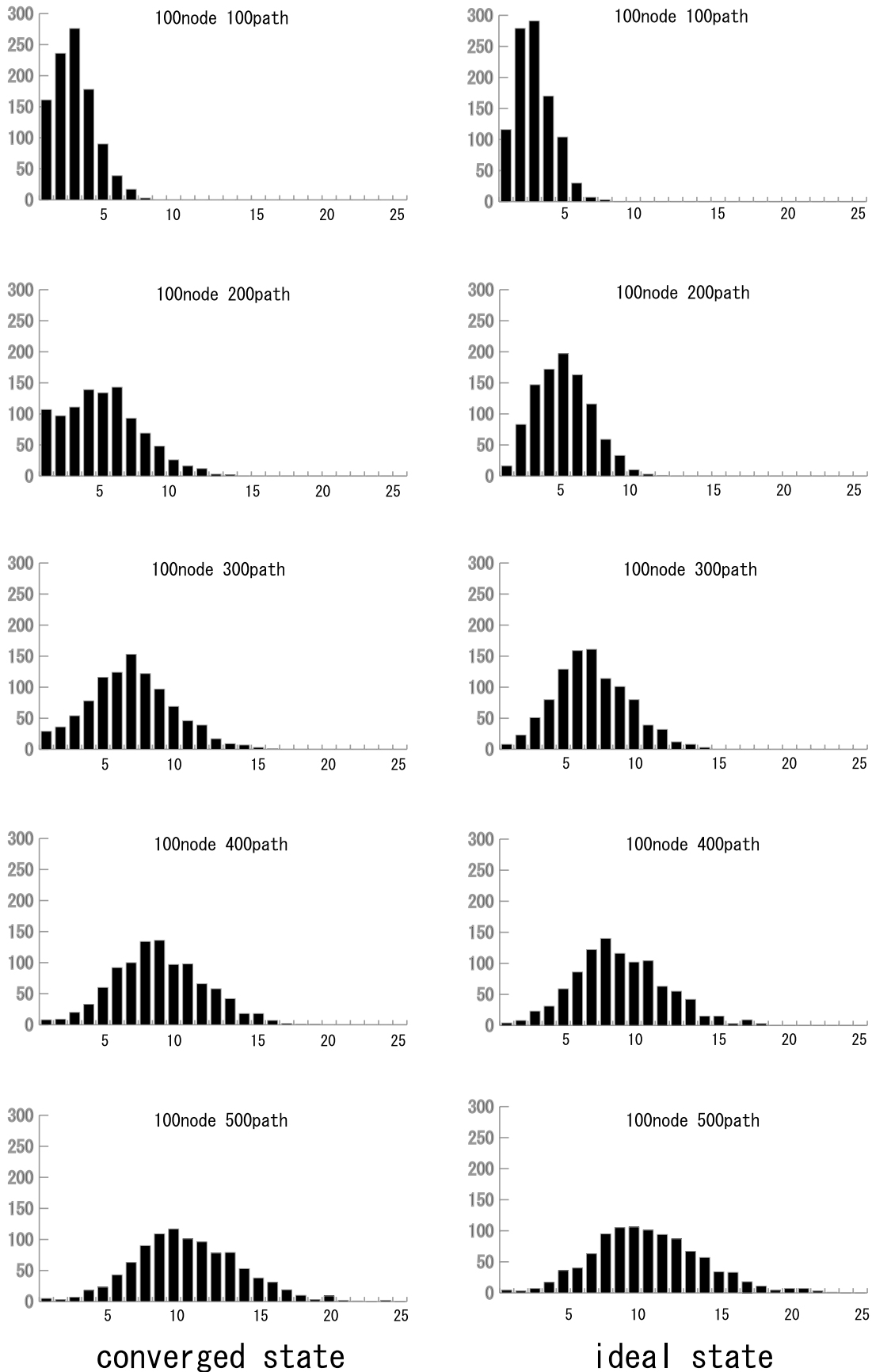


図 4.11 収束解と最適解の次数分布



表 4.4 初期状態のパス数と収束状態のパス数の相関係数

パス数	100	200	300	400	500
相関係数	0.36	0.26	0.06	0.01	0.01

パス数が少ないときは相関係数が高い。初期状態と収束状態の相関係数が高いということは、初期に多くのパスを持っていたノードは収束時にも多くのパスを持ち、逆に初期にパスを多く持っていないノードは収束時にも多くのパスを持っていないということである。

このような傾向がパス数が少ないときに起きているということから、十分な数のパスがネットワーク中に無いときには、あるノードが適切な（関係値の高い）隣接ノードを発見できるかどうかは初期状態でのネットワーク内の位置に依存しているということがわかる。逆に、十分なパス数があれば仲介により初期における偏りは無くなることがわかる。

## 4.6 考察

### 4.6.1 評価値

図 4.7 および図 4.8 のグラフを見ると、ノード数が増えるに従い評価値は減少傾向にあり、逆にパス数が増えると評価値は増加傾向にある。このことから、たくさんのノードによって構成されたネットワークを最適化するためにはたくさんのパス数が必要になるということがわかる。被覆率の場合、パス数がノード数の1倍である場合、かなり悪い結果となったが、到達率の場合は0.8程度であり、ほぼ十分な結果が得られているといえる。特にパス数が2倍以上の場合はほとんど最適解と同じような値となっている。到達率はネットワークを構成するノード間の関係値の高さを示す値である。これらのグラフから、パス数がノード数の2~3倍以上である場合、提案手法は十分に効果を発揮することがわかった。

### 4.6.2 収束性

図 4.5 および図 4.6 を見るとが正しくネットワークが収束していることがわかる。仮想データではなく、実際に人が作ったデータ（Webブックマーク）を用いた場合でも収束していることが図からわかる。仮想データにおける収束ターン数とパス数およびノード数については図 4.10 において説明しているが、どのようなノード数、パス数であっても十分なターン数があれば問題なく収束している。近傍仲介法による最適化は収束を保証していないが、これらの結果から現実にはほぼ確実に収束するといえる。

### 4.6.3 計算量

図 4.9 はノード数およびパス数と収束に必要なターン数の関係を示したグラフである。ノード数の増加にしたがって収束ターン数は線形に増加していることがわかる。線形増加であるため、ノード数が  $n$  倍になった場合は収束に必要なターン数も  $n$  倍になる。本シミュレーションではパスの張替えは 1 ターンで 1 回行なわれる。よって新しいパスが追加されるノードは 1 ターンにつき 2 つとなる。パスが追加されたノードは、推薦のために新たな隣接ノード間の関係値計算を行なわねばならない。ここで 1 つのノードが持つ平均パス数を  $r$  とすると、1 ターンで平均  $2r$  回の関係値計算が行なわれることになる。

ノード数が  $N$  の時に収束ターン数が  $kN$  で求められるとすると、近傍仲介法による最適化に必要な関係値計算回数は  $2r(kN)$  となる。対してサーバ型計算法では全てのノード間の関係値を計算するので計算回数は  $N^2$  となる。 $r$  および  $k$  は定数であるため、近傍仲介法による最適化に必要な計算量は  $O(N)$  となりサーバ型計算法の  $O(N^2)$  を下回る。

### 4.6.4 Anytime Algorithm

近傍仲介法による最適化は常にその時点で計算できた最高の解を出力し続ける Anytime Algorithm である。これは動的にノードの状態が変化するようなネットワーク、例えば人のネットワークなど、において重要な特徴であると考えられる。動的にノードの状態が変化する環境では、ある瞬間において最適であったネットワーク構造が、次の瞬間においても最適とは限らない。ゆえに時間をかけて完全な最適解を求める方法は有効であるとはいえない。その点、Anytime Algorithm である近傍仲介法は動的に変化するネットワークに対しても柔軟に追従することができると考えられる。また、必ずしも収束するまで計算する必要がないという点も計算量軽減に大きく役立つといえる。

### 4.6.5 自律分散モデル

近傍仲介法では個々のノードに対して色々なパス張り替えの戦略を組み合わせることが可能である。人のネットワークに対して適用することを想定した場合、ノードに該当する人は、それぞれ異なる目標を持つはずである。各ノードで異なる評価尺度を持つネットワークの最適化はサーバ型計算法では実現困難であると考えられる。

## 4.7 議論

### 4.7.1 Nearest-Neighboring Model との比較

本研究が提案する近傍仲介法と似た方法でネットワークを生成する, Nearest-Neighboring Model というネットワーク生成モデルが Alexei によって提案されている [68]. 本節では近傍仲介法と Nearest-Neighboring Model との類似点と相違点について議論する.

Nearest-Neighboring Model とは, Alexei によって 2003 年に提案されたスケールフリーネットワーク生成モデルの一つである. Nearest-Neighboring Model におけるネットワークを生成はこのように行われる: 確率  $u$  で新規ノードがネットワークに追加され, 追加された新規ノードはランダムに選択した既存ノードとエッジを結ぶ. さらに確率  $1 - u$  でネットワーク上で最短距離が 2 であるノードのペアのうちランダムで選択された 1 組の間に新たにエッジを追加する.

このモデルは友達に友達を紹介するという現実における社会ネットワークの拡張モデルを模したネットワーク生成モデルであり, 形成されたネットワークは平均結合相関において正相関のスケールフリーになることが確認されている [77]. なお, この平均結合相関は近接するノードとの次数の関係を示す指標で, 社会的ネットワークでは正相関が出ることが報告されている [47].

Nearest-Neighboring Model と近傍仲介法は, 共に友達の友達を紹介するという方法でネットワークを生成または再構築しており, ネットワーク生成手法としては非常に似たものと考えられる. この Nearest-Neighboring Model と今回のシミュレーション実験で実装した近傍仲介法とを, エッジの追加ルールとノードおよびエッジの特徴という点で見ると以下のような違いがある.

- エッジの追加ルール
  - Nearest-Neighboring Model ではパスの関係値がなく, 仲介によって得られたパス候補の中からランダムで追加する
  - 近傍仲介法では追加するパスの一方のノードはランダムで選ぶが, もう一方のノードは仲介によって得られたパス候補の中からパスの関係値により一意に決定される
- ノードとエッジの特徴
  - Nearest-Neighboring Model ではノードは全て同質であり, エッジに重みはない

– 近傍仲介法では個々のノードに違いがあり、エッジに重みがある

Nearest-Neighboring Model はスケールフリーネットワークを生成するモデルである。しかし、それと類似している近傍仲介法が生成したネットワークは、図 4.11 が示すとおり、スケールフリー性は見られない。これは近傍仲介法がパスの関係値の大小によってパスの追加判断を行っていることと、今回のシミュレーションではパスを追加した分だけ削除するというルールがあることに起因すると考えられる。

Nearest-Neighboring Model では、仲介により新たなパスが張られるが、仲介によってどのノードとノードの間にパスが張られるかは単純に確率的に決められる。これはつまりパスを多く持つノードほど仲介によりパスを増やすチャンスが増えるということである。対して近傍仲介法では、パスが多いほど仲介を受けるチャンスは増えるが、パスを張るかどうかはノードとノードの関係値によって判断される。この関係値がノードが持つパス数と正相関がある様な場合は Nearest-Neighboring Model 同様に、パスを多く持つノードほどパスを多く集め、スケールフリーネットワークを生成すると思われる。しかし、例えば今回のシミュレーション実験の様にノードが持つ特徴ベクトルによって固定的に決まってしまう場合は、持っているパスの多さは仲介により新しいパスが張られるかどうかと直接的には無関係であるため、スケールフリー性は生じない。

今回の近傍仲介法の実装ではパスの追加に加えて削除も行う様にし、また、各ノードが持つパスの関係値を高くするようなパスの選択を行う様にした。結果、パスが十分にある状態では最適なネットワークに近いものを作成した。近傍仲介法でもパスを多く持つノードが新しいパスを持つチャンスを多く得ることになるが、パスが実際に追加されるかどうかは関係値に依存し、今回のシミュレーション実験ではパスの関係値は各ノードの特徴ベクトルの内積によって決まるため、スケールフリーネットワークにはならなかった。今回のシミュレーション実験ではネットワークをランダムに生成したので、最適解となるネットワークの次数分布は正規分布に近い形となる。よって、近傍仲介法によって生成されるネットワークも正規分布に近い形となる。ただし、パス数が少ないときはこの限りではない。

パスの関係値の性質によって生成されるネットワークが変わる例として、[37] が挙げられる。まずノードにカテゴリのようなものがあり、同じカテゴリのノードとは関係値の高いパスを持てるがそうでないノードとは関係値の低いパスしか持てないといった性質の関係値が設定されているとする。この場合において、近傍仲介法と似た手法によりネットワークの再構築を行ったところ、同じカテゴリに属するノードは近くに集まり、さらに集まったノードのクラスタ内部ではスモールワールドネットワークが形成されたと報告されている。

以上の考察から、近傍仲介法はスケールフリーネットワークを生成する可能性を持つ

が、最終的にどのようなネットワークを形成するかは、仲介を十分に行えるだけのパス数の有無とパスの関係値の特徴によって決められると言える。

#### 4.7.2 シミュレーション実験での設定について

本章では、ボトムアップな方法による情報共有ネットワークの改善方法について検討を行った。提案手法の有効性を検討するにあたっていくつかの仮定をおいてシミュレーションを行ったが、現実の情報共有ではより複雑な環境になると思われる。そこで、今回のシミュレーション実験では取り扱わなかったが情報共有ネットワークを構築する上で重要と思われる点を以下に述べる。

##### ネットワーク最適解の定義

今回のシミュレーションでは、ネットワークを構成するパス数を固定し、パスの重み（関係値）の和を最大化したネットワークを最適解とした。しかし実際の情報共有においては最適解はもっと複雑なものになると考えられる。まず、今回は最適解を計算可能にするためにパス数を固定にしたが、実際にはパス数は全体として一定ではなく、増減を繰り返すと思われる。また、その際にはネットワークを構成するノードにあたる個人が持てるパス数（次数）には適切な上限があると考えられる。

誰にとっての最適解かという点でも議論が必要である。今回はネットワークを構成するパスの重みを最大化するとしたが、これはネットワーク全体にとっての最適解の一例と言える。この場合、最適解にはネットワークから切断された孤立ノードが含まれる場合もありうる。このネットワークが情報共有ネットワークであるとする、情報共有を行っている個人にとって情報共有ネットワークから切断されることは最も悪い状態である。

このような点を考慮すると、最適なネットワークは構造として定常的ではなく、何らかの条件を満たし続けながら変化し続ける動的なもので、ネットワークの利用モデルと合わせて定義する必要がある。

##### 初期ネットワークの構造

今回のシミュレーションでは、初期ネットワークはランダムに生成した。しかし実際にはすでに出来上がったなんらかのネットワークを用いることになるため、初期ネットワークはランダムとは異なる何らかの意味ある構造を持っていると思われる。そのような偏りがある場合における、ネットワーク再構築手法についても検討する必要がある。

### 各ノードの個性

今回のシミュレーションでは、ネットワークを構成するノードは全て同じルールでパスの張り替えを行うようにし、また、任意のノード間のパスの重みはシミュレーション開始から終了時まで変化しないようにした。しかし、情報共有ネットワークを想定すると、各ノードである個人は様々な戦略を取ることが考えられ、また、情報共有を行うことにより相手との情報共有の相性の良さ（シミュレーションではパスの重みとして表現される）も変化していくと考えられる。このようなネットワークを構成するノードが多様で動的に変化する場合における、ネットワーク再構築手法についても検討する必要がある。

## 4.8 まとめ

本章において、我々は新しい情報交換のネットワークを拡張する手法として近傍仲介法（近傍仲介法）を提案し、シミュレーションによるサーバ型計算法との比較実験により有効性を検証した。近傍仲介法は各ノードが協調的かつ自律的にマッチングを行なうだけで最適化を行なう。しかしながら、それでも十分なネットワークを構築することが可能であることが実験からわかった。

本最適化手法は自律分散的であり中央サーバを必要としていない。これには二つの利点がある。一つは簡単にこの手法を利用できるという点である。ユーザはインターネット上にコミュニティをいつでもどこからでも動的に生成していくことができる。もう一つは、この手法はどのようなサイズのグループに対しても適用可能という点である。なぜなら全ての計算やデータの記録は各ノードが自分の周囲に対して行なうだけで、全てのデータを集め全ての計算をするような中央サーバを必要としないためである。参加するノード数が多いほど、良いペアが生まれる可能性は高くなるので、これは大きな利点であるといえる。

## 第5章

# パーソナルネットワークを用いたコミュニティ支援システム

本章では、異なる情報源が個々が持つ関係性を元に情報共有を行った場合にどのようなのかを、実際にシステムを稼働させて観察および分析することを試みる。

稼働システムとして、新たに開発した学会会議用のスケジューリング支援システムを用いる。このシステムはパーソナルネットワークを利用した情報共有支援という点に特色があり、本研究の観点から言うと情報源は人、情報源間の関係性は個人の知り合い関係、そして共有対象となる情報は学会会議におけるスケジューリングデータとなる。

スケジューリング支援システムは2003年度と2004年度人工知能学会全国大会にて実運用したが、本章では、2003年度人工知能学会全国大会における運用結果の分析を中心に、コミュニティ支援システムにおけるパーソナルネットワークの重要性および有用性について論じる。

## 5.1 はじめに

学会会議内での議論活発化のために重要なこととして参加者が自身の興味に合った発表の存在に気づくことが挙げられる。だが、ある程度の規模の学会会議になると数百の発表があり、その中から自身が興味のある発表を見つけ出すのは困難である。

参加者間での交流を促進するためには、どのような人が参加しているのか、どのような人が自身と似たような関心を持っているのかを知ることが重要であると考えられる。しかし参加者数は一般に発表論文数以上に多く、そのような人の発見は興味のある発表の発見以上に難しい問題となる。

本研究では、この2つの問題を解決するためのシステムとしてパーソナルネットワークを用いたコミュニティシステムを提案・開発し、問題解決に向けて実際に運用した。パーソナルネットワークを用いたコミュニティシステムとは、参加者間のインタラクションを促進させるために知り合い関係という新しい情報源を取り込んだ情報共有システムを指す。

パーソナルネットワークとは社会学の用語で、人と人との関係によって作られる、人のネットワークを指す。人と人との関係といっても様々あるが、今回は特に知り合い関係に注目した。コミュニティも同様に社会学の用語で、様々な解釈があるが、ここでは、参加者間のインタラクションにより情報共有が促進される情報共有システムのことをコミュニティシステムと呼ぶ。

コミュニティシステムにパーソナルネットワークを導入することにより、参加者はシステム内でいわば“コンテンツ化”され“ネットワーク化”される。コンテンツ化によって参加者を知るための情報が詳細化され、ネットワーク化によってその情報の流通が促進される。その結果、参加者間のインタラクションが促進されると考える。

## 5.2 スケジューリング支援システム

### 5.2.1 システムの特徴

提案システムの特徴は、多くの人に使ってもらうためのシンプルな操作性と、参加者間の交流促進と情報発見容易化のための共有モデルである。

提案システムはシステムは実際に会議聴講を側面から支援するものであるため、いかに簡単に使えるかは多くの利用者を獲得できるかどうかの鍵になる。そこで我々はほとんど全ての操作をボタンのクリックのみで行えるようにした。システムの利用方法については5.2.3節にて説明する。



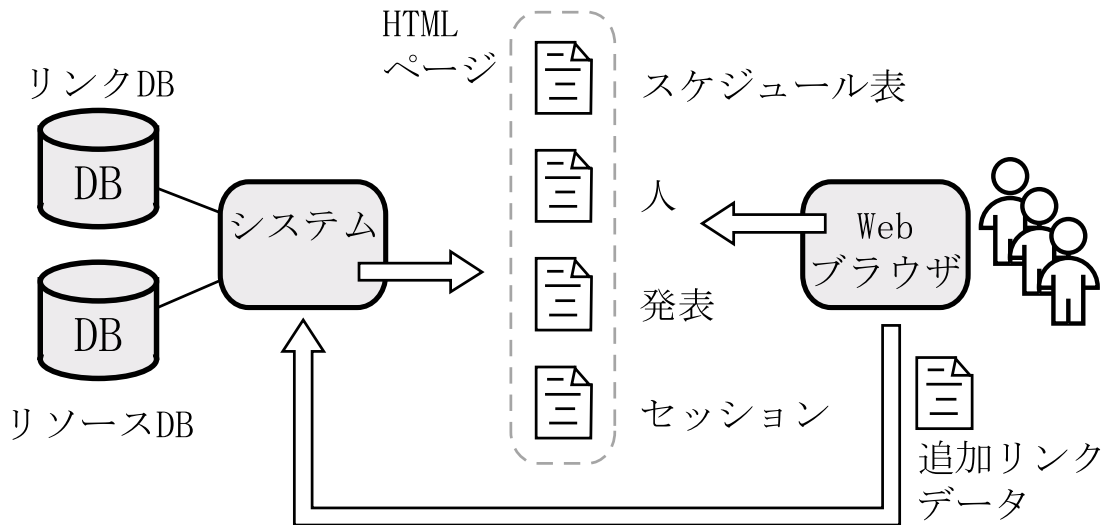


図 5.1 システム構成

提案システムの最も特徴的な点は、知り合いネットワークの構築とそれに基づく柔軟な情報共有機構の提供である。

- インタラクティブな知り合いネットワークの構築：簡便なリンク追加に加え、リンクが自分宛に追加されたことが即座にわかる。また、自分と友人を共有してる人を発見したり、自分と聴講予定の論文を共有している人を発見し、知り合いネットワークに追加することも可能である。
- 知り合いネットワークを利用した多様な情報共有：他の人の聴講スケジュールや知り合いリストを見ることができる。同様に、自分の聴講スケジュールを見せることもできる。また、自分と聴講論文を共有している人たちが誰であるかもわかる。
- 知り合いネットワークを利用した柔軟な共有のコントロール：アクセスコントロールを知り合いネットワークを利用して実現する。例えば、相手から友人と指定されたときのみ、その人のスケジュールを知ることができる。

### 5.2.2 システム概要

システムは図 5.1 のような構成をとる。システムは MySQL データベースと Perl で記述されたプログラムにより構成される。データは全て MySQL データベースで管理され、利用者からのアクセスは CGI プログラムが受け付ける。利用者は Web ブラウザを用いてシステムにアクセスする。

データベースは大きく分けて 2 種類ある。1 つはリソース用データベース、もう 1 つはリンク用データベースである。リンクとはリソースとリソースの関連性を示す情報であ

り、常に（主格リソース、リンクの種類、目的格リソース）の 3 つ組の形式を取る。これは RDF (Resource Description Framework) [14] の書式を参考にしている。

リソースおよびリンクの種類について以下に示す。リソースは 3 種類あり、それぞれ独自の属性値を持つ。

- Session : セッション。属性値としてセッション名や部屋番号、開催日時を持つ。
- Paper : 発表論文。属性値としてタイトルや原稿ファイルの URL を持つ。
- Person : 人。最初に登録されるのは著者と座長。聴講者も後から登録される。

リンクは 5 種類ある。Contains と Chairs および Authors は事前に登録されており追加不可能であるのに対し、Check と Know はシステム稼働前には登録されておらず、利用者が追加することによってのみ登録することができる。

- Contain : セッション - 発表論文関係。セッションからそのセッション内の発表論文に対して作られる。
- Author : 著者 - 著作関係。著者 (人) から著作 (発表論文) に対して作られる。
- Chair : 座長 - セッション関係。座長 (人) からセッションに対して作られる。
- Check : 参加者 - 発表論文間関係。聴講者が気になる発表論文に対して作る関係。自分用の聴講スケジュールを作成する際に用いられる。
- Know : 参加者 - 参加者間関係。参加者が知っている参加者に対して作る関係。

これらリソースデータベースとリンクデータベースに格納されたデータを元に、システムは著者と発表論文とセッション、さらにスケジュール表の 4 種類の HTML ページを動的に生成する。以下に各ページの画面イメージを示す。HTML ページ間にはリンク情報に基づいてハイパーリンクが張られている。

利用者は生成された HTML ページを自由に閲覧することができ、さらに新しいリンクを追加することもできる。この利用者によって追加されたリンクを元に個人用のスケジュール表は変化していく。

### 5.2.3 システム利用の流れ

システムを利用するためには、利用者登録をする必要がある。今回は著者または共著者、座長は事前に登録してその情報を通知した。そうでない場合は、フォームに氏名、所属、メールアドレスを入力して利用者登録を行う。利用者登録が済むとシステムからの ID とパスワード通知メールが送信される。

基本的な利用法は以下の通りである。Web ブラウザでログインページへアクセスし、ID とパスワードを入力してシステムにログインする。ログイン後、利用者はまずマイ

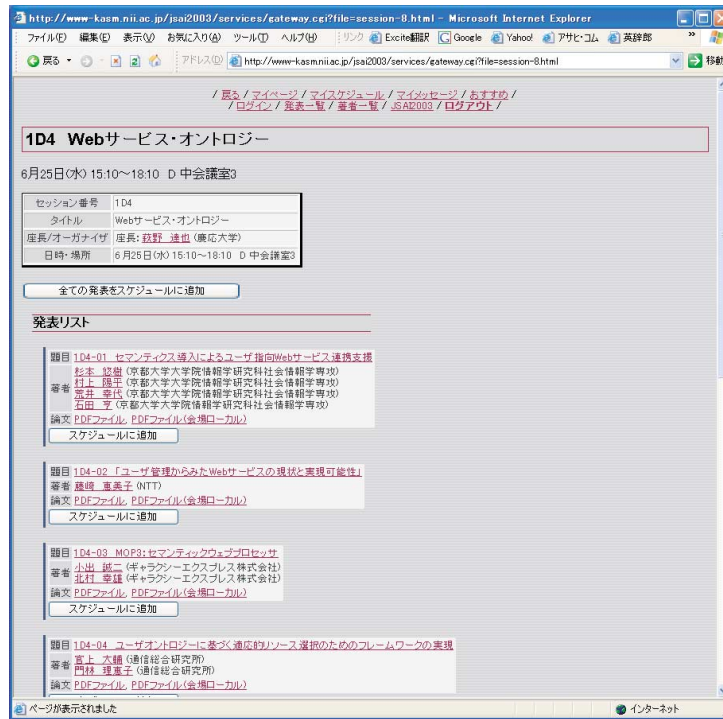


図 5.2 Session リソースの HTML ページ例

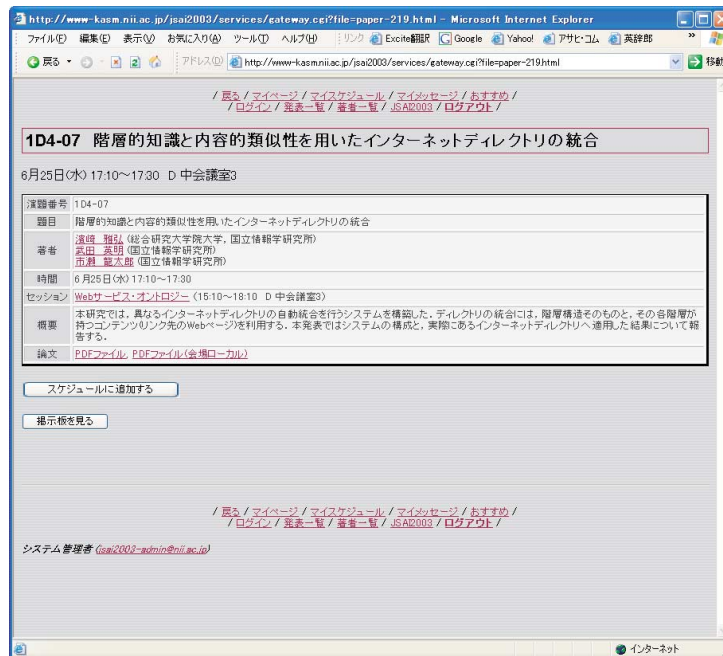


図 5.3 Paper リソースの HTML ページ例

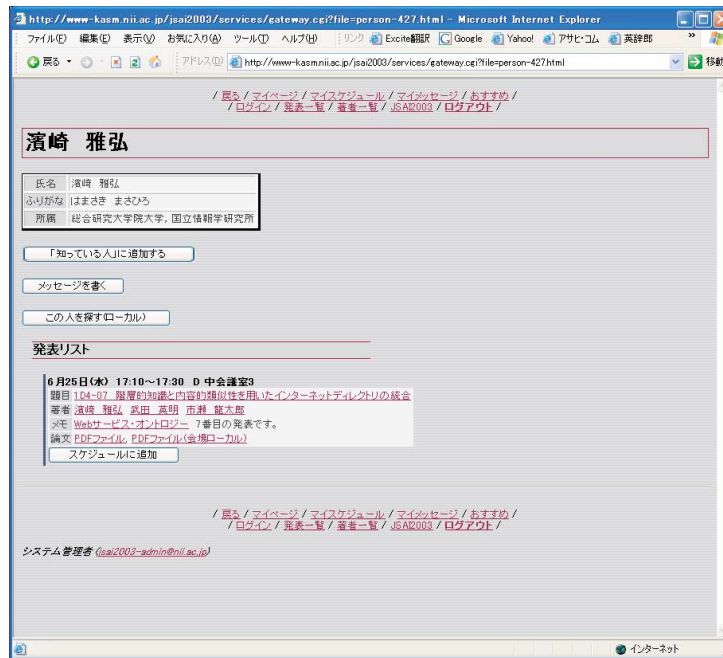


図 5.4 Person リソースの HTML ページ例

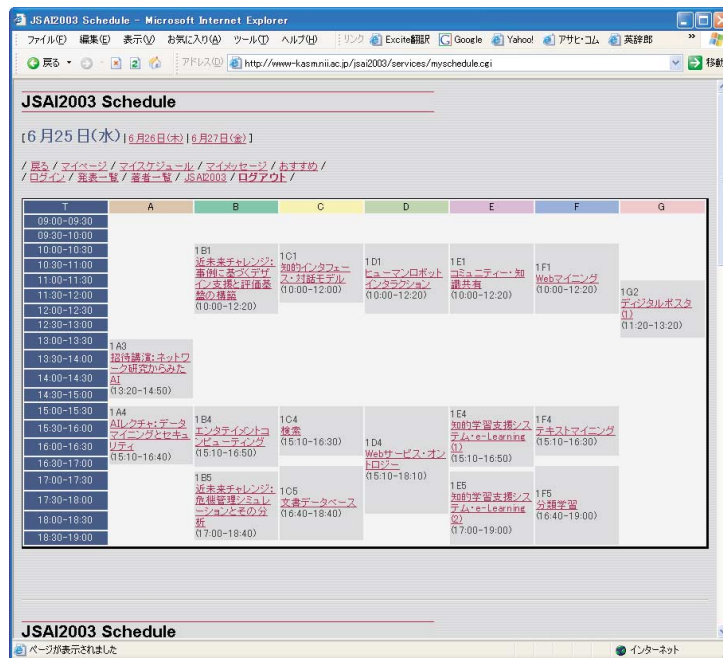


図 5.5 スケジュール表の HTML ページ例

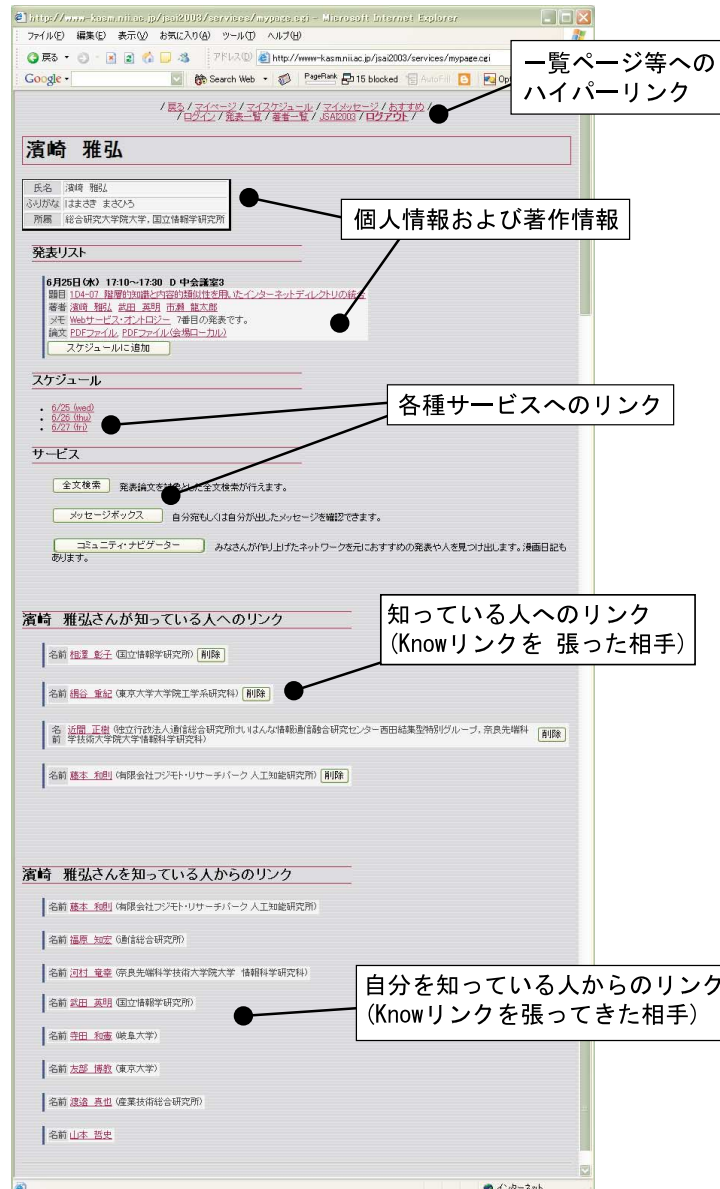


図 5.6 マイページ (人のページ)

ページへ移動する。マイページとは本システムにおける自分用のポータルとなるページである。自分の著作論文や自分が追加したリンクの一覧、自分用のスケジュール表へのリンクの他、全文検索やメッセージャー等の他のサービスへのリンクがある（図 5.6）

本システムを用いて興味のある発表を探す場合、一覧ページを使う方法と、発表のタイムテーブルから興味あるセッションを選び、そのセッションの発表一覧から見つける方法がある。

前者の場合は各 HTML ページ上部にあるハイパーリンクを利用する。このリンクから発表論文へ移動することができる。発表論文は演題番号順に一覧表示されている。

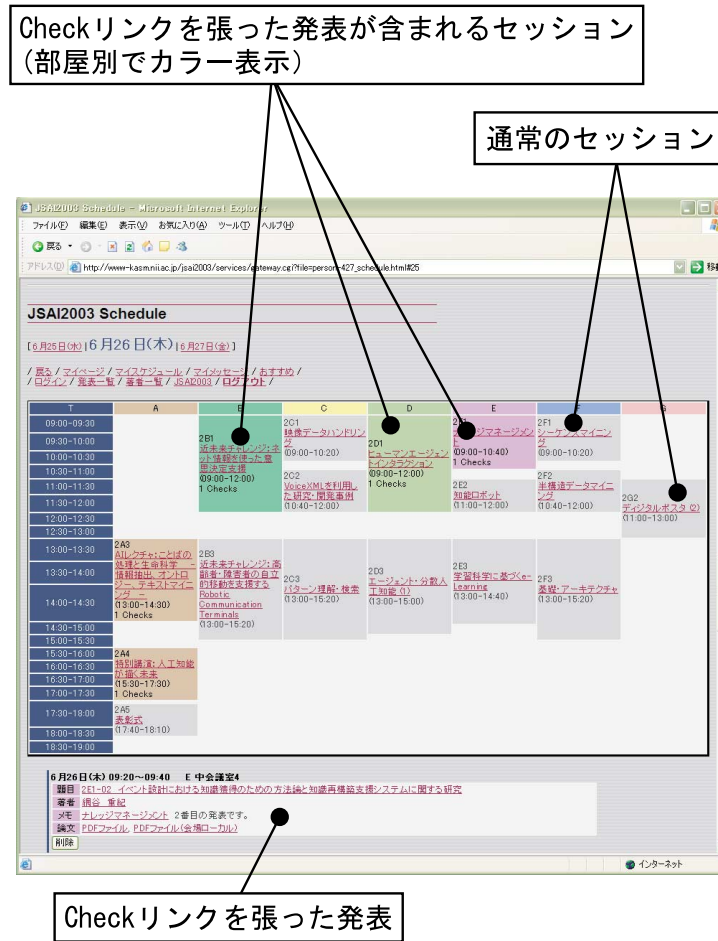


図 5.7 マイスケジュール (スケジュール表のページ)

後者の場合は、まず自分用のスケジュール表のページへ移動する。スケジュール表のページでは部屋と時間で分類されたセッション一覧を見ることが出来る。セッション名をクリックするとセッションのページへ移動する。そこにはそのセッションの発表一覧が表示される。発表タイトルをクリックすると発表論文のページへ移動する。各発表にはリンク追加ボタンが付いており、クリックすると Check リンクが生成される。Check リンクは利用者がその発表を自分の個人用スケジュール表に追加したい場合に生成する。本論文では、この Check リンクを張ることを、聴講予定に追加すると呼ぶ。

Check リンクが追加されると、その利用者の個人用スケジュールページに聴講予定の発表論文のページへのハイパーリンクが追加される。同時に、その発表が行われるセッションがタイムテーブル上で強調表示される (図 5.7)

また、聴講予定の発表論文のページへ再度訪れてみると、その発表論文を聴講予定であるに他の参加者を見ることが出来る。これは発表をきっかけとした人の出会い支援機能である (図 5.8)



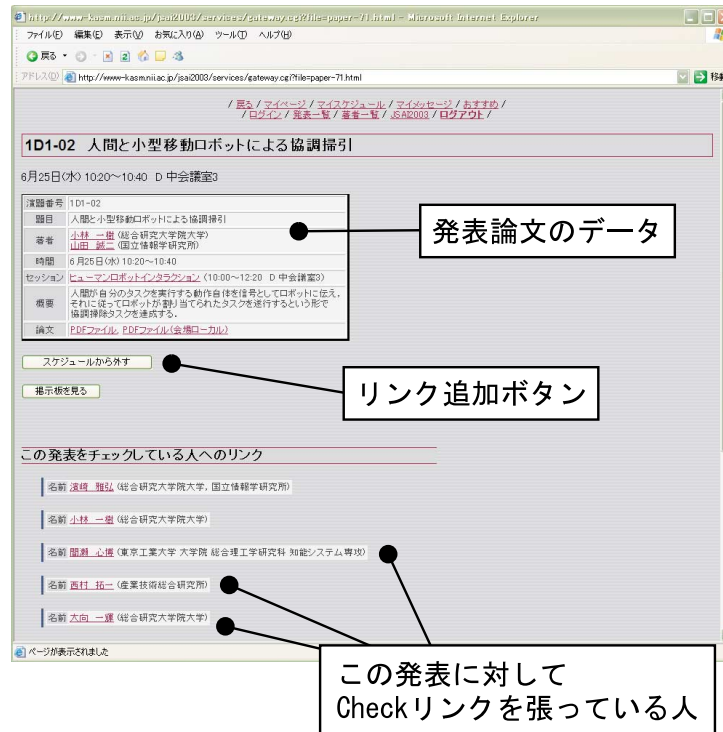


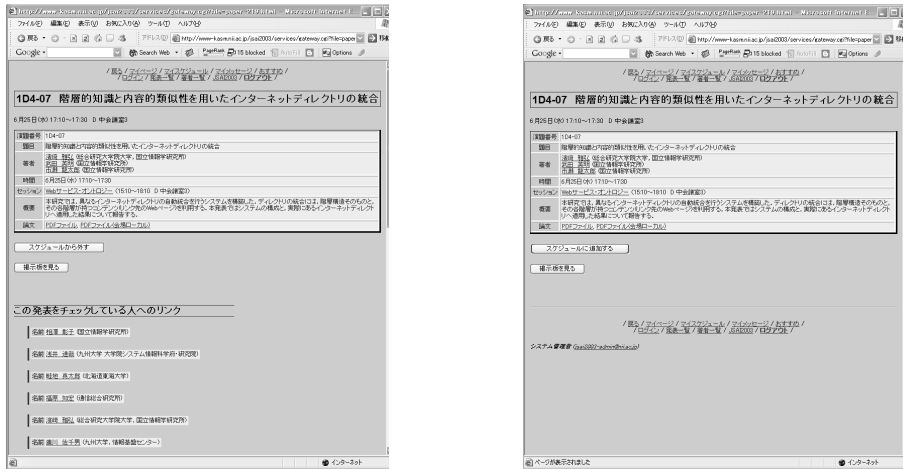
図 5.8 聴講予定の発表論文のページ

著者を発見する場合は、同様に一覧ページから探すが、発表論文データにある著者へのリンクを利用する。著者のページには発表ページ同様、リンク追加のボタンがある。ボタンをクリックすれば Know リンクが生成される。Know リンクは、自分が知り合いだと思う相手に対して生成される。Know リンクを追加した場合、自分のマイページに Know リンク先としてリストに追加される。同時に相手のマイページにも Know リンク元としてリストに追加されるので、Know リンクを張られた側は誰が自分に向けて Know リンクを張っているかを容易に気づくことができる。

利用者がリンクを追加していくことで、利用者自身のページが充実していく。利用者のページはすなわち人リソースのページであり、Know リンクでつながった相手と共有されているコンテンツである。利用者は、他の利用者が作ったコンテンツおよびリンクを通じて、新たな論文や著者のページを発見することができる。このように、本システムでは知り合いネットワークの構築と情報共有は同時に行われる。

#### 5.2.4 情報共有のコントロール

追加されたリンク情報は個人情報であり、そのまま公開するには適さない。利用者による情報共有のコントロールが必要である。このコントロールに手間がかかる場合、利用者



(1) 自分がCheckリンクを張っている論文

(2) 自分がCheckリンクを張っていない論文

図 5.9 論文ページの見え方

は情報共有に億劫になり、結果、必要以上に共有が制限されてしまう可能性がある。本システムのような情報共有システムには、以下のような点に注意する必要がある。

- 不必要に公開を制限させない
- 利用者に負荷をかけない
- 利用者が自由にコントロールできる

そこで本システムでは、グループへの所属に基づくアクセスコントロールを行った [51]。本システムにおけるグループには 2 種類ある。1 つは友人グループで、これはある人の知り合いグループである。自分が友人として登録されている場合には、その人の個人情報や発表情報に加え、知り合いネットワークも見ることができる。そうでない場合は D や E は表示されない。もう 1 つは論文の聴講予定者グループである。自分がその論文を聴講予定としていたときのみ、他の聴講予定者を知ることができる。

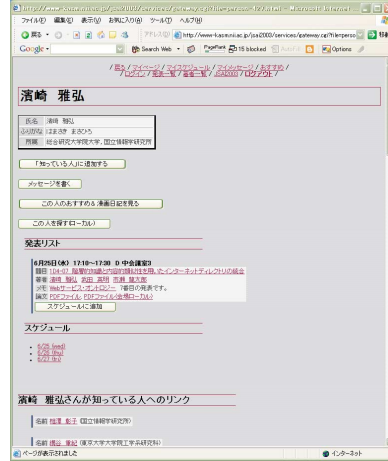
図 5.9 は論文のページの見え方の違いを示している。自分と論文のページとの関係には 2 種類ある。すなわち Check リンクを張った場合 (1) とそうでない場合 (2) である。アクセスした利用者とアクセスした先の論文の関係がどちらであるかによって、表示されるが異なる。(1) の場合、論文の基本情報 (タイトル、著者、概要等) に加え、その論文に対して Check リンクを張っている人のリストや、その論文用の掲示板へジャンプするボタンが表示される。(2) の場合は基本情報と Check リンク追加のボタンが表示される。

図 5.10 は人のページの見え方の違いを示している。自分と人のページとの関係には 3 種類ある。1 つは自分自身のページ (1)、もう 2 つは相手が自分に対して Know リンクを張っている人のページ (2) とそうでない人のページ (3) である。この 3 種類の関係の内、





(1) 自分のページ



(2) 知人のページ



(3) 他人のページ

図 5.10 人ページの見え方

アクセスした利用者がどれに該当するかで見える内容が異なる。(1)の場合、人の基本情報(名前, 所属)に加え, Know リンクを張っている / 張られている人のリストや, 個人用時間割表やパーソナルメッセージャー, 全文検索等のジャンプするリンク / ボタンが表示される。(2)の場合は基本情報に加え, 個人用時間割表と Know リンクを張っている / 張られている人のリストが表示される。(3)の場合は基本情報と Know リンク追加のボタンが表示される。

### 5.2.5 推薦機能

リンク追加による個人用ページの生成に加えて, 利用者が追加したリンクを利用した情報推薦サービスを行った。この推薦サービスでは, 発表論文だけでなく人の推薦も行われ

る．利用者はブラウジングだけでなく，推薦によっても興味ある発表論文や人の発見を行うことができる．

本サービスでは，人と論文の2種類のリソースを推薦した．さらに推薦には2種類の方法を用いた．一つは GroupLens[39] 等の協調フィルタリング的手法である．これは同じような関心を持っている人のデータを参考に推薦を行う．論文推薦の場合，利用者が Check リンクの張った論文に対して，同様に Check リンクを張っている人たちがどの論文に Check リンクを張っているか，という情報を利用する．もう一つはパーソナルネットワークを利用する手法であり，知り合いが持つデータを参考に推薦を行う．具体的には Know リンクを張っているまたは張られている相手が持っているリンクを用いる．本論文では，前者の Check リンクを用いた推薦を関心類似型推薦，後者の Know リンクを用いた推薦をパーソナルネットワーク型推薦と呼ぶ．

このように，2種類のリソースを2種類の手法で，合計4種類の情報推薦を行った．

推薦するリソースの見つけ方について説明する．人  $h_0$  が論文  $p$  に対して Check リンクを張っている場合は  $Check(h_0, p) = 1$ ，そうでない場合は0を取る  $Check$  関数，人  $h_0$  が人  $h_1$  に対して Know リンクを張っている場合は  $Know(h_0, h_1) = 1$ ，そうでない場合は0を取る  $Know$  関数を定義する．さらに  $Know(h_0, h_1) = 1$  または  $Know(h_1, h_0) = 1$  の時にのみ  $Relate(h_0, h_1) = 1$  となる  $Relate$  関数を定義する．

$V_{hc}(h_0, h_x)$  と  $V_{pc}(h_0, p_x)$  は人  $h_0$  に対する関心類似型推薦における，人  $h_x$  および論文  $p_x$  の推薦度をそれぞれ計算する．

$$V_{hc}(h_0, h_x) = \sum_{\{p|Check(h_0, p_x)=1\}} Check(h_x, p_x)$$

$$V_{pc}(h_0, p_x) = \sum_{\{h|V_{hc}(h_x, h_0)>0\}} Check(h_x, p_x)$$

$V_{hk}(h_0, h_x)$  と  $V_{pk}(h_0, p_x)$  は人  $h_0$  に対するパーソナルネットワーク型推薦における，人  $h_x$  および論文  $p_x$  の推薦度をそれぞれ計算する．

$$V_{hk}(h_0, h_x) = \sum_{\{h_k|Relate(h_0, h_k)=1\}} Relate(h_k, h_x)$$

$$V_{pk}(h_0, p_x) = \sum_{\{h_k|Relate(h_0, h_k)=1\}} Check(h_k, p_x)$$

今回の推薦サービスでは，それぞれ4種類の推薦ごとに推薦度  $V$  が高い上位5つを取り出し，そのうち推薦度  $V$  が閾値（今回は2）以上のものを推薦情報として利用者に提示した．

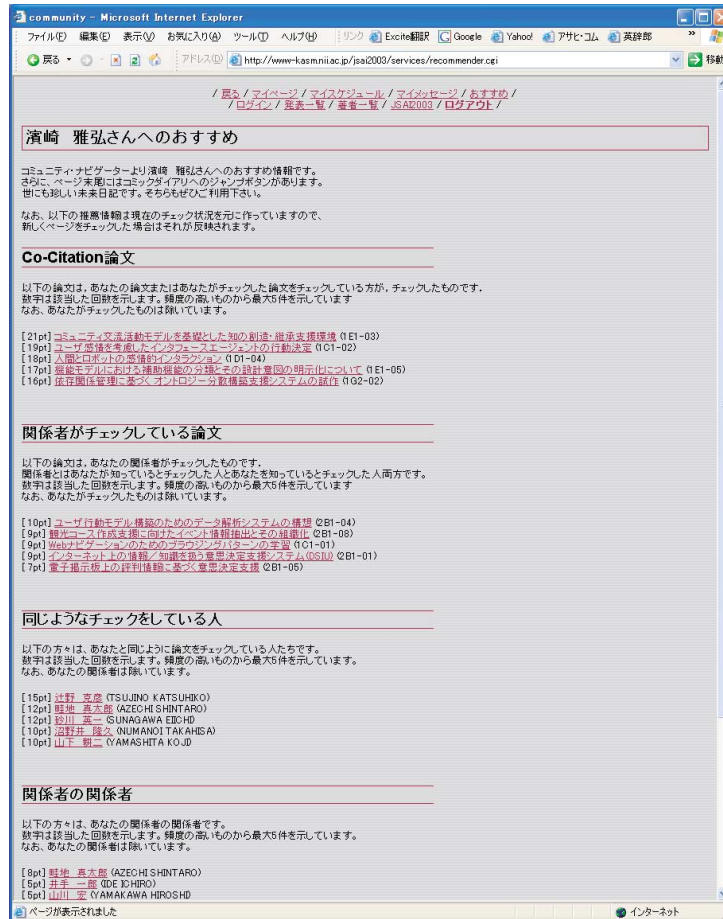


図 5.11 推薦サービス画面

図 5.11 は推薦サービスの画面イメージを示している。推薦には 4 種類あるが、関心類似型推薦による人推薦の結果を「同じようなチェックをしている人」、同じく論文推薦の結果を「Co-Citation 論文」とし、パーソナルネットワーク型推薦による人推薦の結果を「関係者の関係者」、同じく論文推薦の結果を「関係者がチェックしている論文」として表示している。複数の推薦手法で同一アイテムを推薦しているため、各種法については簡単な解説も書いてある。

### 5.2.6 その他のサービス

本システムでは、利用者間でのコミュニケーション手段として、掲示板とパーソナルメッセージャーを提供した。

図 5.12 は掲示板の画面イメージである。掲示板は各発表ごとに設置されており、聴講予定者だけが閲覧および書き込み可能である。掲示板には書き込まれたコメント内容だけでなく、聴講予定者に向けた情報がいくつか表示される。一番上に聴講予定者のリストが



図 5.12 発表ごとに設置された掲示板

表示されているが、このリストは何人の人（この発表の聴講予定者に限る）から知られているかによって降順にソートされている。よって、一番上に位置する人はこの聴講予定者の中で一番多くの人から認知されている人であることを示している。さらに、聴講予定者たちを対象としたおすすめ論文および人のリストが表示される。これは 5.2.5 節で示した方法により聴講予定者全員に対して推薦アイテムを求め、その結果を足し合わせたものに相当する。

図 5.13 はパーソナルメッセンジャーの画面イメージである。パーソナルメッセンジャーは、個人宛にメッセージを送る機能であり、誰でも利用することができる。パーソナルメッセンジャーは他人からのメッセージの受信箱として機能するだけでなく、自身の発言

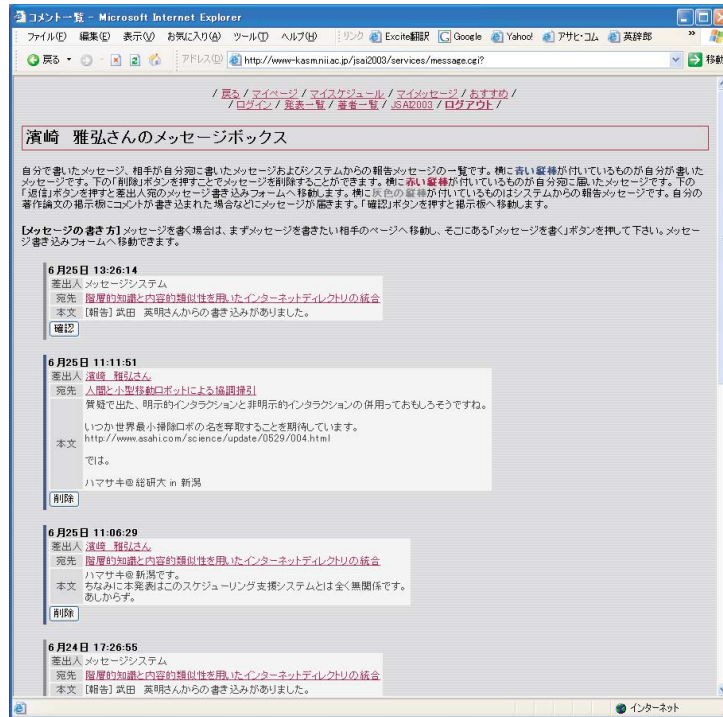


図 5.13 パーソナルメッセンジャー

の管理システムとしても機能する．自分が送信したメッセージや掲示板での発言は全てパーソナルメッセンジャーにリストアップされており，それらの確認および削除が行える．

さらに漫画日記 [62] や namazu<sup>\*1</sup>を用いた論文検索サービスを行った．前者は本システムが持つリンク情報や，推薦情報を漫画という手法で表現したものである．後者はキーワード検索に加え，マイクロクラスタリング [4] を用いたクラスタリングサービスも行った．

会期中には西村らによる CoBIT[49] を用いた位置情報サービスが行われた．同時に無線 ID タグを用いた参加者の位置検索サービスも提供された．その検索サービスへのインタフェースも用意されており，人の現在位置を知りたい場合はスケジューリング支援システム内の人のページにアクセスして「この人を探す」ボタンをクリックすると，CoBIT が提供する位置検索サービスへ移動し，その人の現在位置が表示される．

### 5.3 運用結果の分析

本システムを 2003 年 4 月 30 日から，人工知能学会全国大会開催の 6 月 23～27 日を挟んで 6 月 30 日までの 2ヶ月間運用した．本章では，この間の利用ログの分析結果を述

<sup>\*1</sup> <http://www.namazu.org/>

表 5.1 初期リソース数

Session	49
Paper	259
Person	510

表 5.2 初期リンク数

Contain リンク	259
Chair リンク	40
Author リンク	770

表 5.3 終了時リソース数

Person	558
--------	-----

表 5.4 終了時リンク数

Check リンク	1840
Know リンク	840

べる。

### 5.3.1 基本データ

2003 年人工知能学会全国大会では、49 のセッションがあり、259 件の発表が行われた（特別講演等含む）。本システムは各セッションおよび発表論文ごとにページを作成した。著者および共著者、座長等は合計 510 人であり、こちらも同様に 1 人 1 ページずつ作成した。最終的に 818 ページが初期データを用いて作成された。

5.1 と 5.2 はシステム稼働開始時にデータベースに登録されたデータの数である。5.3 と 5.4 は学術会議終了後の 6 月 30 日の時点でデータベースに登録されたデータの数である。なお、Person を除く初期登録されたリソースおよびリンクの数は基本的に不変であるため省略している。

### 5.3.2 利用者数の推移

図 5.14 は、4 月 30 日の開始から 6 月 30 日までの利用者数<sup>\*2</sup>の変化および毎日のログインユーザ数を示している。表 5.5 は、利用者増のために行った告知活動と、その活動 2 日後の利用者数を示している。

本サービスは 2003 年人工知能学会全国大会のオンラインプログラム公開と同時に開始した。その際には、登録した著者および共著者のうちメールアドレスがわかっている 257 人に、メールにてサービス開始の通知を行った (1)。ある程度のリンク追加が行われたのを見計らって被リンク通知を行った (4)。被リンク通知とは、自分または自分が著者・共著者である論文に対するリンクがあることを通知するサービスである。その後も告知活動

<sup>\*2</sup> ここでいう利用者とは、システムに登録されており、かつ、1 回以上システムにログインしたことがある人を指す。

表 5.5 告知活動

活動	日付	利用者数
1. システム始動	4/30	29(+29)
2. AI学会 ML 通知	5/04	37(+8)
3. AI学会 ML 通知二回目	5/07	73(+36)
4. 被リンク通知	5/26	176(+84)
5. 推薦サービス開始	6/04	217(+35)
6. PDF 公開&論文検索・掲示板開始	6/17	238(+16)
7. 漫画日記開始	6/23	263(+17)

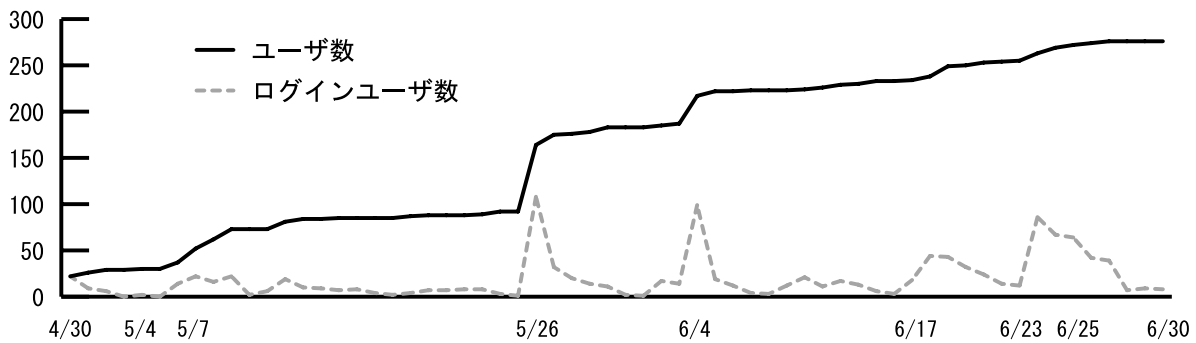


図 5.14 1日あたりのログインユーザ数とユーザ数推移

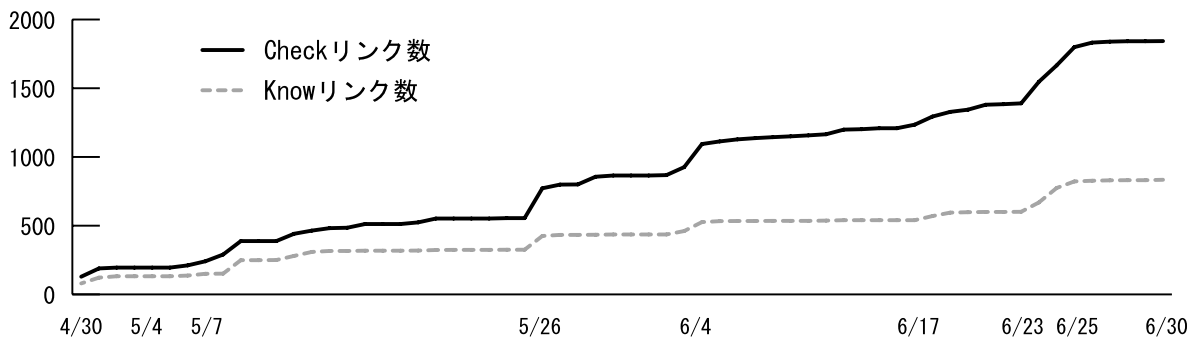


図 5.15 リンク数の推移

を行い、6月30日の時点で1度でもシステムにログインしたことのある利用者の数は276人となった。

最終的に全登録ユーザのシステムへのログイン率は49%であった。これは2000年度に同じ学術会議において角らが行ったオンラインサービス [59] のログイン率(33%)よりも高い。システムに自動登録されている人の中には共著者の様に会議には参加しない人も含まれていることや、会議のオンラインプログラムは本システムとは別に同等の物が用意されているにも関わらず、半数近くの人がログインしたという結果は、本システムが学術会議の支援システムとして参加者に受け入れられたことを示しているといえる。

### 5.3.3 登録方法による利用状況の違い

本システムでは登録方法により登録されている人を二種類に分けることができる。一つは座長、著者、共著者などの初期データとして自動登録されている人たちである。もう一つはそれ以外の人たちで、彼らは自発的にシステムにアクセスし、登録手続きを済ませることでシステムに登録される。

自動登録された人々には、システムに自動登録された旨がメールにより伝えられる<sup>\*3</sup>。だが実際にログインするかどうかは任意であるため、システムに登録されている人の中にはシステムに一度もログインしたことがない者が含まれる。以後、システムに登録されている人を登録者、システムにログインしたことがある人を利用者と呼ぶ。

全登録者558人中のうち、利用者は276人であった。内訳は自動登録された利用者が228人、自発的に登録した利用者が48人である。

5.16は自動登録された利用者とは自発的に登録した利用者の、リンク状況の違いを示している。リンク追加は本システムにおいて利用者が行うことができる能動的動作であり、リンク追加回数の多さは本システム内におけるアクティビティの高さを示しているといえる。自発的に登録を行った利用者の方がCheckリンク、Knowリンク共に追加する割合が高い。これはシステム利用の動機が高いためと考えられる。この傾向は、2000年度に角らが行った参加者支援サービスで提供されたオンラインサービスでも同様であった [59]。

Knowリンク所持の割合、つまりKnowリンクを追加したかもしくは自分宛にKnowリンクを追加されている割合、は自動登録された利用者の方が高かった。これは、自動登録された利用者は論文の著者または座長であるためセッションや論文のページからハイパーリンクが張られている可能性が高いことを勘案すると、妥当な結果である。しかし、むしろプログラム上に出現しない自発的登録ユーザ<sup>\*4</sup>の約3分の1が他者からKnowリ

<sup>\*3</sup> ただし半分はメールアドレスを確認できなかったため通知はしていない。

<sup>\*4</sup> メール通知の不備等の理由から、自発的登録ユーザがプログラム上の人物と同一人物である可能性はある。しかし、システムでは両者が同一であるかといった処理は行っていない。



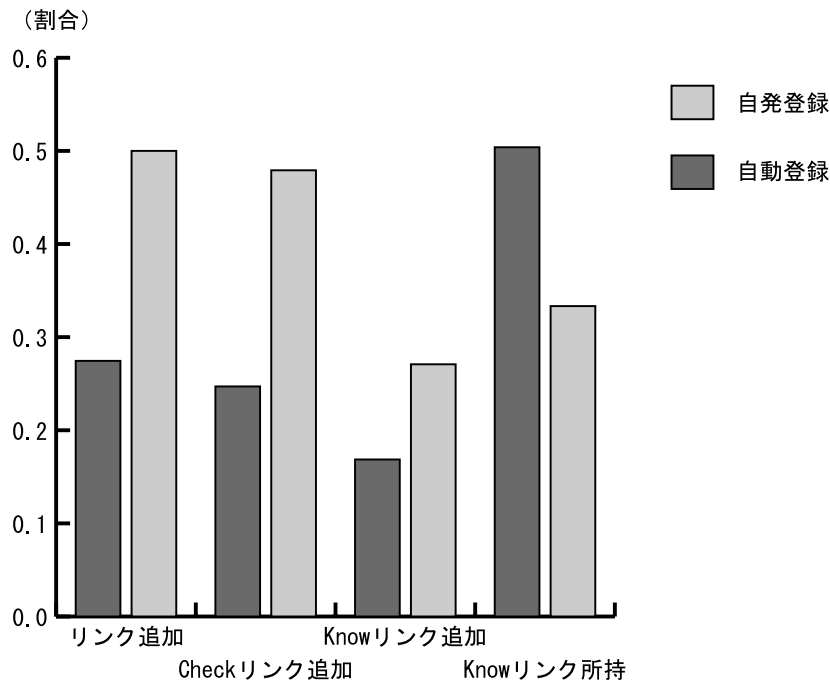


図 5.16 登録方法によるリンク状況の違い

リンクを張られてたという結果は興味深い。

これは、これらのユーザが積極的に Check リンクおよび Know リンクを作った結果、他者から認知された割合を示していると考えられる。この点でシステムはコミュニティ形成に一定の効果があったといえる。

活発な（多くのリンクを追加する）利用者は、他の情報とのつながりが増えていくことによって情報源として重要な存在になっていく。本システムでは自発的に新規登録した利用者は活発な利用者である傾向がみられたので、例えば新規参加者に注目が集まるような仕組みを設けるなどすれば、情報流通の点で効果的であると考えられる。

#### 5.3.4 生成されたリンク構造の分析

本節では、利用者により追加されたリンクによって、どのようなネットワークが作られたのかを分析する。

人は Check リンクと Know リンクの始点、および Know リンクの終点となることができる。同様に、論文は Check リンクの終点になることができる。5.6 は Check リンクを追加したことのある利用者数と、少なくとも 1 人以上から Check リンクを張られている論文数を示している。同じく 5.7 は Know リンクを追加したことがある人の数と、少なくとも 1 つ以上 Know リンク又は被 Know リンクを持っている人の数を示している。

Check リンクも Know リンクも共に、リンク追加を行った人は登録されている人に対

表 5.6 Check リンクと論文リソース

	持っている 人/論文	全体に占める 割合
Check リンク	149 人	0.27
被 Check リンク	245 件	0.95

表 5.7 Know リンクとリソース

	持っている 人	全体に占める 割合
Know リンク	99 人	0.18
被 Know リンク	260 人	0.47

して3分の1にも満たない。しかし論文では95%が、人では49%が、それぞれ Check リンクまたは Know リンクのネットワークに接続されている。利用者がどこにリンクを張るかは自由であるが、結果的にこのように幅広いネットワークを作ることができた。これは利用者によって追加されたリンクで作られたネットワークによって新しい情報探索のルートを提示することを目的としている本システムにとって、システムの有用性を示す結果であるといえる。

5.17 は論文が持つ被 Check リンク数の度数分布と、人が持つ被 Know リンク数の度数分布を並べたものである。被 Check リンク数は4~12個に集中していることがわかる。対して被 Know リンク数は、最小値である被リンク数1のノードが一番多く、極端に多く持つノードがわずかにあるという、典型的なスケールフリー構造を持っている[8]。知り合いネットワークは著名人に集中する一方、発表論文は一定のばらつきがあることがうかがえる。このような異なる性質を持つリンクを両方持つことは幅広い探索を実現するのに有効であると思われる。Check リンクも Know リンクも利用者によって自由に追加可能であるが、両者の情報の性質は異なっていることがわかる。

なお、Know リンクによって作られた人のネットワークは直径が9、クラスタリング係数が0.39であり、5つのクラスタから形成されていた。クラスタのサイズはそれぞれ262, 4, 3, 2, 2であった。対して共著関係で作られた人のネットワークは73のクラスタから形成されており、直径が7であった\*5。

本システムでは、人と人をつなぐリンクには共著関係と Know リンクとの二種類があ

\*5 ノード数は462、リンク数は885。

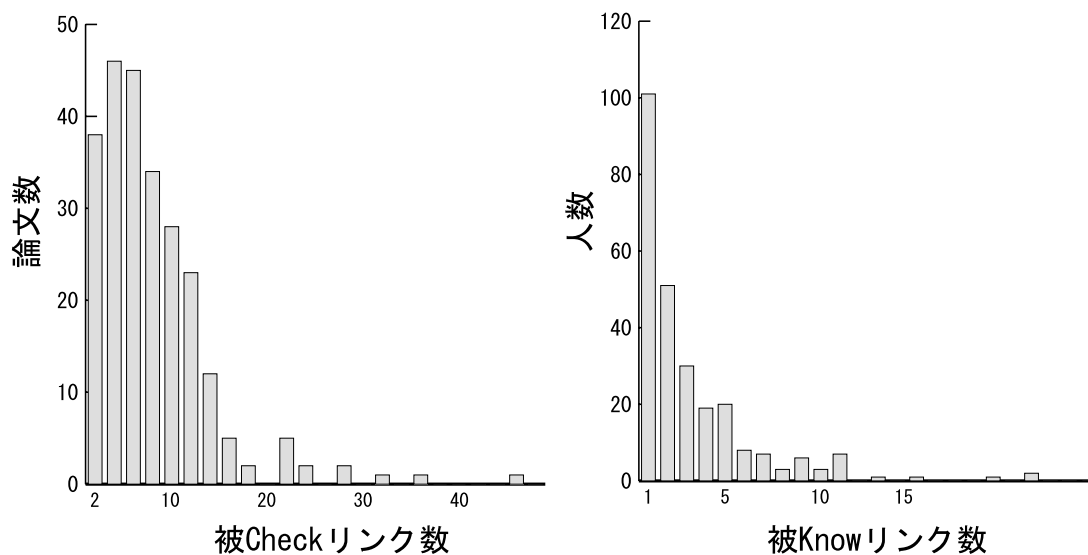


図 5.17 被 Check リンク数と被 Know リンク数

る．図 5.18 と図 5.19 はそれぞれ Know リンクのネットワーク（図 5.18）と共著関係のネットワーク（図 5.19）をネットワーク分析ツール Pajek[10] で描画したものである．共著関係ネットワークと Know リンクネットワークとで共有しているリンクは 135 個であった．すなわち Know リンクの約半数は共著関係からは見つけられないものであったことがわかる．また，Know リンクネットワークは 5 つのクラスタ（各クラスタに含まれるノード数はそれぞれ 262, 4, 3, 2, 2）から形成されているのに対し，共著関係で作られた人のネットワークは 73 のクラスタから形成されていた\*6．

Know リンクと共著関係を合わせたネットワークは，クラスタ数は 19 個になった\*7．図 5.20 は，共著関係と Know リンクによって構成された人のネットワークの一部を示している．細い実線が共著関係，破線が Know リンク，太い実線が共著関係があり，かつ Know リンクが張られている関係である．ノードは共著ネットワークによって作られたクラスタごとに色分けされている．

共著ネットワークでは 8 つのクラスタに別れてしまっているが，Know リンクによって 1 つのクラスタに結合できることがわかる．共著関係は人のネットワークにおける重要なリンクであるが，それだけでは断片的なネットワークを構成してしまい，Know リンクの追加によって初めて大きなパーソナルネットワークを構築することができた．

\*6 ノード数は 462，リンク数は 885．

\*7 共著ネットワークには Know リンクネットワークにはないノードがあるため，Know リンクネットワーク単体よりもクラスタ数が多くなっている．

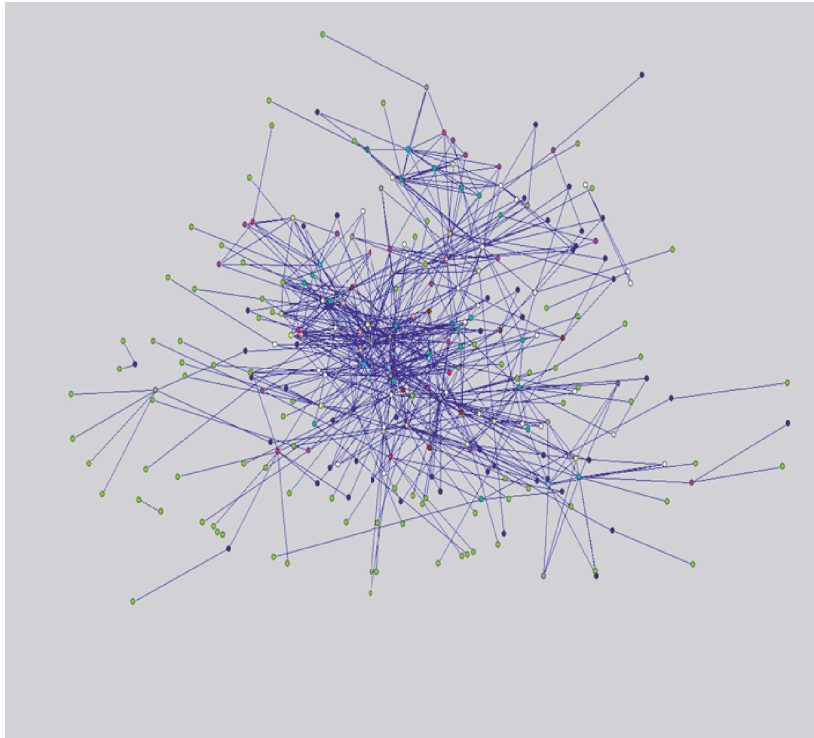


図 5.18 Know リンクネットワーク

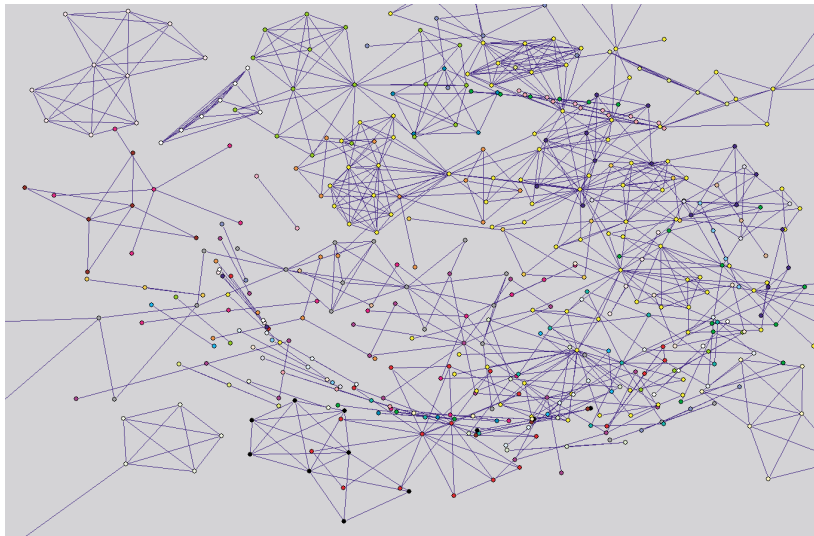


図 5.19 共著関係ネットワーク

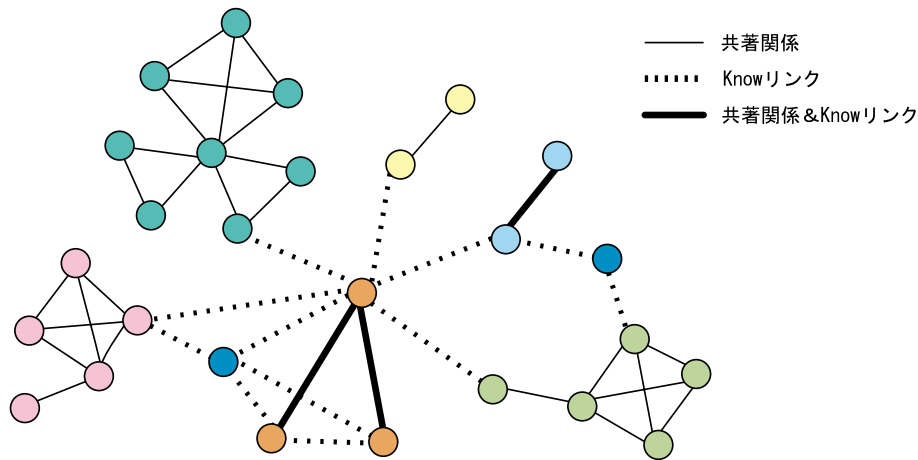


図 5.20 共著関係と Know リンクのネットワーク (一部)

### 5.3.5 Know リンクネットワークのネットワーク構造

Know リンクが作るネットワークの構造について分析を行う。

Watts はネットワークの特徴を示す指標として、特性パス長  $L$  (characteristic path length) とクラスタリング係数  $C$  (clustering coefficient) を提案している [69]。特性パス長はネットワークを構成する全ノード間の最短距離平均を、クラスタリング係数は後者は隣接するノード間が直接パスで結ばれている割合の平均を指す。同じノード数およびリンク数を持つランダムネットワークと比較して、 $L$  が同程度、 $C$  が高い場合はスモールワールドと呼ばれる。これは、ノードがクラスタ状に集まっているにもかかわらず、ノード間のパスが短いという特徴を持つ構造であり、自然界や社会における様々なネットワークにこの傾向が見られる。

Know リンクによって作られた人のネットワークは  $L = 3.69$ ,  $C = 0.47$  であった。ノード数 280, リンク数 669<sup>\*8</sup> のランダムネットワークでは  $L = 3.79$ ,  $C = 0.01$  となり、Know リンクネットワークがスモールワールドであることがわかる。

Watts らは様々なネットワークに対してこの値を求めており、俳優ネットワークは  $L = 3.65$ ,  $C = 0.79$ , アメリカの電力網は  $L = 18.7$ ,  $C = 0.28$ , 線虫の神経細胞ネットワークは  $L = 2.65$ ,  $C = 0.28$  であると報告している [70]。

Amaral らはスモールワールドネットワークを scale-free, broad-scale, single-scale の三種類に分類している [5]。scale-free はノード数とリンク数の関係はべき法則に従う。broad-scale は初期にはべき法則に従うが、ある程度以上は指数的に急速に減衰する。

\*8 Know リンク数は 840 だが、うち 342 は双方向リンクなので  $804 - 342 / 2 = 669$ 。

single-scale はリンク数が指数的に減衰する．このような違いが生まれるのは，ノードがリンク作成可能な期間とノードがリンクを作るコストの存在によるとされている．

5.21 は横軸に被 Know リンク数，縦軸に人数をとった両対数グラフである．グラフ中の直線は被リンク数 1~7 までの近似曲線で，式は  $y = 14.176x^{-1.3542}$  となる．グラフを見ると初期にはべき法則的に振る舞っていることが，その後，急激に減衰していることがわかる．これは broad-scale に見られる傾向である．Amaral らの調べによると，知り合いネットワークは single-scale であった．本システムの Know リンクネットワークは，知り合いネットワークであると同時にオンライン上にて作られるネットワークであるために，その性質に違いが生じたのだと考えられる．

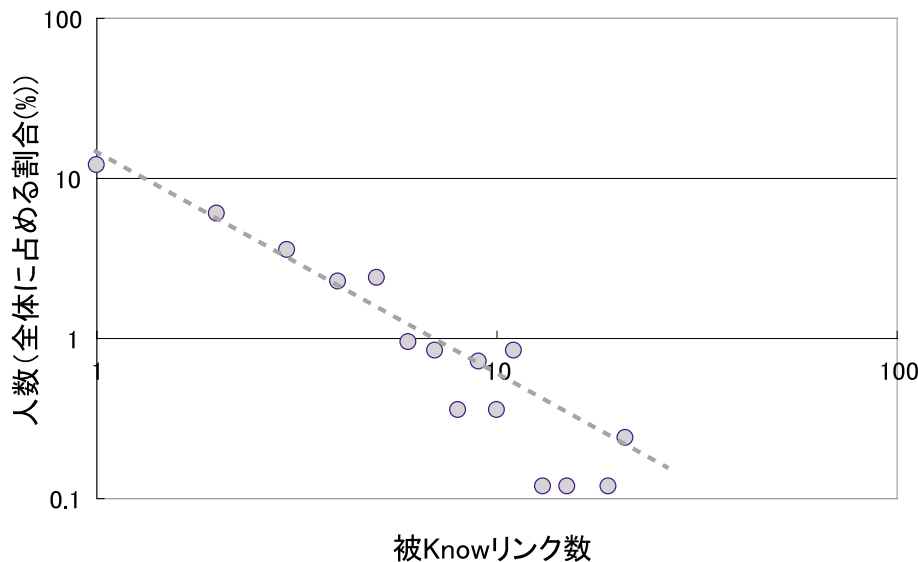


図 5.21 被 Know リンク数の両対数グラフ

### 5.3.6 利用者によるリンク構造の利用の分析

本節では，利用者のページの移動パターンから，人のネットワークがどの様に活用されたかを示す．

HTTP サーバのアクセスログから利用者の移動パターンを解析した<sup>\*9</sup>．図 5.22 は各種ページ間の利用者による移動回数のうち値が十分大きいもののみを示している．

図から，スケジュールからセッション，そして論文のページへと移動するのが大きな流れとなっていることがわかる．それぞれ 2~3000 件程度の移動履歴が残っていた．スケ

<sup>\*9</sup> この手法では Web ブラウザで「戻る」ボタンを押して移動した場合は追跡できない．また，ブックマークやメールから URL を直接指定した場合にも追跡できない．このため図 5.22 の各ノードの入出力の和は 0 にならない

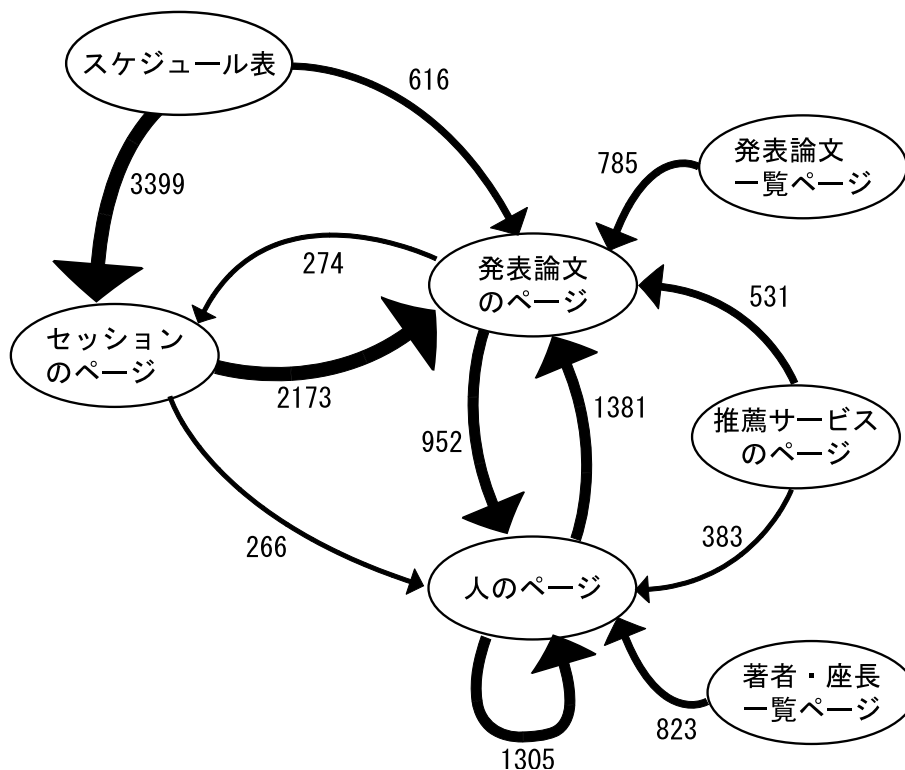


図 5.22 アクセス遷移図

ジュールの中からセッションを選び、セッションの中から論文を選ぶという流れは典型的なオンラインプログラムの閲覧の仕方であるといえる。論文のページから人のページには1000件近くの移動履歴があり、逆も同様であった。さらに人のページから人のページへの移動も1300件程度あることがわかった。

セッションから論文への一般的な移動パターンの半分に相当する回数のアクセスが、人のページを介して行われた。また、人から人への移動もそれに相当する回数行われた。前者は論文（のページ）へのリンクを持った人（のページ）というコンテンツが利用者に多く利用されたことを示し、後者は人と人のネットワークが利用者に多く利用されたことを示している。これらの結果は、本システムの目的であった人のコンテンツ化およびネットワーク化による利用者の情報収集支援がなされたことを示している。

### 5.3.7 パーソナルネットワークを用いた推薦の分析

システムの推薦サービスによって推薦されたリソースが、利用者に受け入れられたかどうかを調べた。推薦は5.2.5で説明した手法で計算した推薦度の高いリソースを列挙することで行われる。利用者は推薦リソースの名前をクリックすることで推薦リソースのページへ移動することができる。そこでリンク追加ボタンをクリックすれば、Check/Know

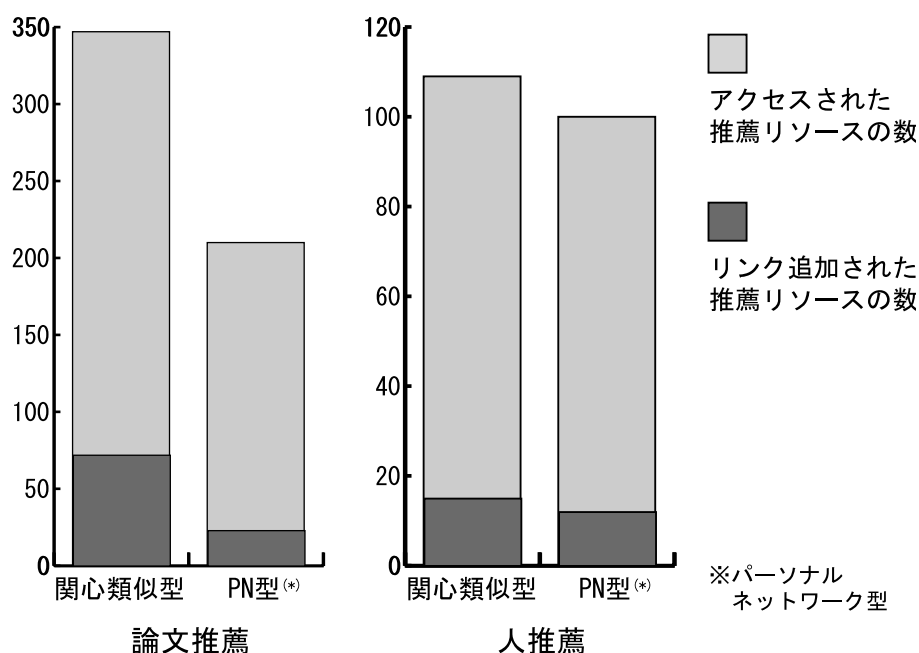


図 5.23 推薦結果

リンクが生成される。推薦リソースのページにアクセスしたということは、そのタイトルを見てある程度の関心を持ったことであり、推薦結果の弱い受理と言える。さらにその推薦リソースに対してリンクを生成した場合は、推薦されたリソースに対して強い関心があったといえ、これは推薦結果の強い受理と言える。

5.23 は、推薦を受けた利用者が、実際に論文または人のページにアクセスした数と、それに対してリンク追加した数を示している。推薦されたリソースにアクセスした人は 135 人おり、うち 37% の 51 人が推薦リソースに対してリンクを張ったことがある。なお、推薦サービスは全利用者に対して提供しているが、推薦サービスにアクセスするかどうかは任意であるため、利用者全員が推薦を受けたわけではない。

論文推薦においては、パーソナルネットワーク型推薦 (PN 型) は関心類似型推薦と比べて、アクセスされた数では 60% 程度、リンク追加された数では 30% であった。しかし人推薦においてはほぼ同程度の結果となった<sup>\*10</sup>。

次に、関心類似型推薦とパーソナルネットワーク型推薦との推薦結果について比較分析をした。5.24 は両手法で推薦された論文うち、利用者がアクセスした論文および Check リンクを追加した論文が持つ、被 Check リンク数を示している。

グラフを見ると、受理された推薦コンテンツが持つ被リンク数は関心類似型推薦の方

\*10 ただし推薦時の表示が、まず関心類似型推薦で次にパーソナルネットワーク型推薦であったことと、両手法が同じコンテンツを推薦する可能性があるため、関心類似型に有利な設定であった。



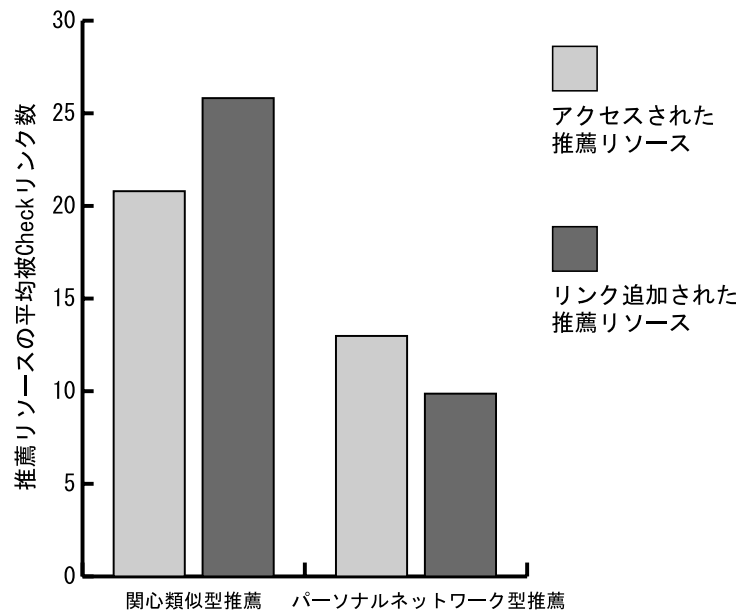


図 5.24 推薦手法と被リンク数の関係

が多い。関心類似型推薦では全体で人気のあるリソースが推薦されがちであるのに対し、パーソナルネットワーク型推薦では自分の周りで人気のあるリソースが推薦される。それらは必ずしも全体においても人気のあるリソースであるとは限らないため、このような結果が現れたのだと考えられる。

また、関心類似型手法による推薦では、受理されたリソースはより被リンク数の多いものであったのに対し、パーソナルネットワーク型推薦では逆の傾向になった。これも関心類似型手法が、より人気のある物を推薦しようとするのに対し、パーソナルネットワーク型推薦ではそれとは異なった尺度で推薦しようとしていることの現れであると考えられる。

ここで、パーソナルネットワーク型推薦の推薦能力について今回の運用結果に基づいて推定する。

利用者  $U$  が Know リンクを張っている先の人たちを  $F$  とする。  $U$  が Check リンクを張っている論文を  $U_c$ 、  $F$  が Check リンクを張っている論文を  $F_c$  とする。このとき  $U_c \cap F_c$  が利用者  $U$  とその知り合い  $F$  が共に Check リンクを張っている論文になる (5.25)。同様に  $U$  の Know リンク先を  $U_k$ 、  $F$  の Check リンク先を  $F_k$  とする。

今回の結果では、各個人が Check リンクを張った論文に対して、自分が Know リンクを張っている人のうち少なくとも 1 人が Check リンクを張っている割合は  $\frac{U_c \cap F_c}{U_c} = 0.67$  であり、自分が Know リンクを張った人に対して Know リンクを張っている割合は  $\frac{U_k \cap F_k}{U_k} = 0.76$  であった。

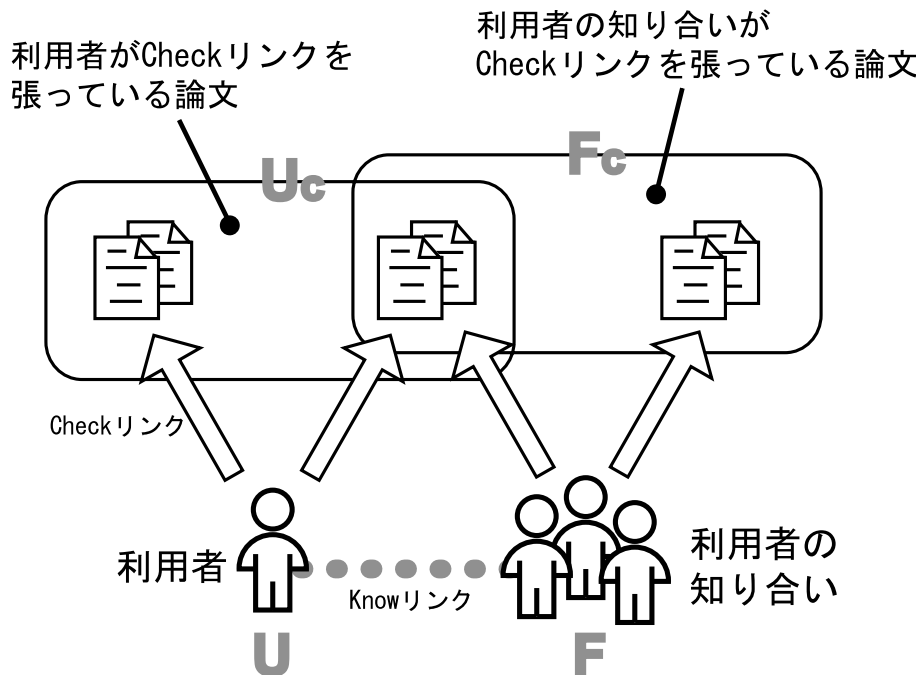


図 5.25 利用者とその知り合いが持つ Check リンクの関係

また、各個人が持っている Check リンクの平均値は  $U_c = 12.4$ 、そして各個人が Know リンクを張っている先の人たちが Check リンクを張っている論文の数の平均値は  $F_c = 52.8^{*11}$ であった。

$F$  が Check リンクを張っている論文  $B$  が、利用者  $U$  に Check リンクを張られている割合は、 $\frac{U_c \cap F_c}{F_c} = \frac{U_c \cap F_c}{U_c} \times \frac{U_c}{F_c} = \frac{12.4 \times 0.67}{52.8} = 0.16$  となる。ここから、パーソナルネットワーク型推薦によって推薦された論文  $B$  が利用者  $U$  にとって Check リンクを張る対象となる（すなわち利用者  $U$  にとって関心のある）論文であるかどうかの確率はだいたい 16% であることが予想できる。

同様に人を推薦する場合も考える。自分が Know リンクを張った人に対して Know リンクを張っている割合は  $\frac{U_k \cap F_k}{U_k} = 0.76$  であった。各個人が持っている Know リンクの平均値は  $U_k = 8.5$ 、各個人が Know リンクを張っている先の人たちが Know リンクを張っている人の数の平均値は  $F_k = 42.5$  であった。従って知り合いの知り合いが自分の知り合いであった確率は  $\frac{U_k \cap F_k}{F_k} = \frac{U_k \cap F_k}{U_k} \times \frac{U_k}{F_k} = \frac{8.5 \times 0.76}{42.5} = 0.15$  となり、だいたい 15% であることが予想できる。

以上の考察から、パーソナルネットワークを用いた推薦は、協調フィルタリング的な関心類似型推薦とは異なる性質を持ち、またその精度は 15~6% 程度と予想されることがわかった。

\*11 重複があるのでこの値は  $U_k \times U_c$  とは等しくない。

## 5.4 考察

### 5.4.1 サービスとしての成果

本システムは276人に利用され、さらに能動的行動であるリンク追加を行った利用者を164人獲得できた。発表件数259件の会議における非強制的サービスとしては優れた結果であったと思われる。

2000年度人工知能学会全国大会にて角らが行ったオンラインサービスでは、利用者は自動登録者のうち利用したのが87人(18.16%)、自発的登録者では116人(85.29%)であった。対して本システムは自動登録者で228人(44.7%)、自発的登録者で48人(100.0%)であった。サービスの内容、提供の仕方および提供期間が異なるので単純には比較できないが、利用者数の点では以前に運営されたシステムを上回る結果となった。

### 5.4.2 コンテンツネットワークにおける人ノードの価値

本システムでは学会のプログラム表示に、セッションや発表論文のページだけでなく、著者のページも導入した。発表時間を調べるためのオンラインプログラムという観点からいえば著者のページは特に必要ではない。しかし発表論文間をセッションと論文間の関係とは異なる関係でつなぐ役割を果たせる。5.3.4より、利用者が作成したリンクネットワークによって、全論文の約95%がカバーされたことがわかった。さらにアクセスログの解析5.3.6から、このリンクネットワークが多く利用されていることがわかった。

また、Checkで追加する論文を知り合いが追加している割合、Knowで追加する相手を知り合いが知っている割合が、共に7割近くあった。これは知り合いから情報を得ることが有用であることを示す結果といえる。また、実際にパーソナルネットワークを用いた推薦を行った結果、一般的な協調フィルタリングと比較すると論文推薦で3分の1程度、人推薦で同程度の効果を上げることができた。

前者の結果は人のネットワークが有効利用されたことを示し、後者の結果は人が価値あるコンテンツであることを示しているといえる。

### 5.4.3 コミュニケーション支援

本システムでは利用者間のコミュニケーションを支援する機能として、掲示板とパーソナルメッセージャーを提供した。掲示板は各発表ごとに用意されており、発表論文に関する議論を各発表論文用の掲示板でしてもらおうことが筆者らの意図であったが、実際には掲示板はほとんど利用されなかった。

CHI2002にて運用された掲示板システム CHI Place[13][23]においても、盛況なのは自由に設定した話題に関して議論する部分がほとんどで、各発表論文を話題とした掲示板はあまり盛況ではなかった。角らが2002年度に行ったデジタルアシスタントプロジェクトにおいても各発表論文ごとに議論するための書き込みスペースを用意されたが、あまり利用されなかった[59]。これらの結果から、このような形式の掲示板は盛り上がりにくいものであることがわかる。これに関しては今後、考察を深める必要がある。

#### 5.4.4 パーソナルネットワークと情報共有支援

本システムでは、利用者が作成したパーソナルネットワークを用いた情報推薦を行った。5.3.7節の結果より、パーソナルネットワークを用いた情報推薦の有効性が認められたが、ここでは推薦性能とは異なる観点からパーソナルネットワークを用いた情報推薦（情報共有支援）の有効性について考察する。

情報推薦では、一般に利用者のプロフィールを作成する必要がある。情報フィルタリングではキーワード登録やカテゴリ登録などによってプロフィール作成が行われる。協調フィルタリングは少し形式が異なるが、利用者が行ったアイテムに対する評価を用いることから、個々のアイテムに対する評価そのものがプロフィールであると言える。

情報推薦の課題の一つは、このユーザプロフィール作成の問題である。自動的に作成するには、利用者の嗜好を機械的に取得する必要があるが、それは一般に困難である。また、利用者に作成してもらうことである程度詳細なプロフィールを得る事も可能であろうが、利用者の負荷という問題がある。キーワード登録や興味のあるカテゴリの登録などは利用者の手によるプロフィール作成の比較的負荷の少ない方法である。これは適切なキーワードを用意しないと利用者の特徴を上手く表現できないが、かといって大量のキーワードを用意しては作成者の負荷が大きくなってしまふという問題がある。

パーソナルネットワークの場合、登録するのは自分の知り合いであるから数にはある程度限りがあり、また、システムが特に登録対象を選別して用意する必要はない。システムと利用者の両方においてパーソナルネットワークによるプロフィール作成は比較的負荷の低いものであると考えられる。知り合いが多い場合には負荷が大きくなる事が考えられるが、その場合はめばしい知り合いだけをまずは登録すればよいと考える。なぜならば、キーワードの場合には登録されなかったキーワードに関する分野がすっぱり抜け落ちる可能性があるが、人の場合は一人で複数の意味合いを持つ存在であるためそのようなことが起こりにくいと考えられる。ただし、このように登録した人数を増減させた場合に推薦結果がどのように変化するか（どの程度性能が落ちるか）は今後調査が必要であると考えられる。

アイテムの評価を登録する協調フィルタリングとは、補完関係にあると思われる。協調

フィルタリングはアイテムに対する評価が類似したユーザ (Neighbor) を発見することが必要である。しかしユーザが十分な数のアイテムを評価していない初期状態においては適切な Neighbor の発見が難しいという問題がある。そのような場合にパーソナルネットワークを利用する (例えば知り合いを仮想的に Neighbor として扱う, 等) という戦略は有効であると考えられる。

#### 5.4.5 学術会議支援システム

学術会議における出会い支援を行うシステムはいくつか提案されており, 実際に学術会議等にて用いられ成果を得ているものがある。

石田らは, ICMAS'96 にて ICMAS Mobile Assistant Project を行った [31]。これはは現実のコミュニティのメンバーが情報ネットワークによってどの様に支援可能であるかを実験したプロジェクトである。国際会議の参加者に携帯端末を持たせ, その上でコミュニケーションを促進するような実験プログラム, 通常の電子メールや掲示板の他に周辺情報案内や個人カスタマイズ型情報共有, 出会い支援システムなど, を稼働させた。

Dey らは携帯端末と無線タグによる位置情報を利用した, 会議参加者支援システムを試作した [16]。システムは取得した位置情報などを用いて, 携帯端末に利用者向けにカスタマイズした発表スケジュール表を表示する。

角らは 2000 年度人工知能学会全国大会にてデジタルアシスタントプロジェクトを行った [59][61]。デジタルアシスタントは, 会議における出会い支援を目的としたカンファレンス支援システムである。利用者は PDA を持って会期中にリアルタイムで情報共有サービスを受けられると同時に, 会場内に設置された情報キオスクやエージェントサロンを通してサービスを受けることができる。

会期中は利用者が同じ場所と時間を共有しており, そこにおいてリアルタイムにサービスを行うことができるモバイルコンピューティングを利用した出会い支援は, 有効な手法であると考えられる。しかしながら会場用のモバイルコンピュータを配布し, それを参加者に利用してもらうコストはまだまだ大きい。我々は学術会議開催前に公開されるオンラインプログラムを用いることで, 会期前の比較的長い期間において利用者の関心を引きつけることのできるシステムの構築を目指した。学術会議開催前に提案システムのような利用負荷の小さい非同期型コミュニティウェアを用いて情報共有を行い, 会期中に角らの提案するような同期型コミュニティウェアを用いて情報交流を行うことで, より効果的に学術会議における参加者支援が行えるのではないかと考える。

オンラインプログラムをコンテンツとした学術会議支援システムとして Jameson らの UM2001 website [55][32] が挙げられる。これは会議における個人用スケジュール作成を支援するシステムである。このシステムでは, 利用者のアクセス履歴や他の利用者のアク

セス履歴を用いて、利用者が関心があると思われる発表を見つけ出し、推薦する。このシステムでは利用者が適切な発表聴講計画表を作られることを目的としている。本システムは発表推薦に加え、人の推薦も行うことにより、参加者間の情報共有を支援しようとしている点で異なっている。

## 5.5 スケジューリング支援システム for JSAI2004

本章で示した解析結果は、スケジューリング支援システムを JSAI2003 にて運営し、そこで得られた運用データを分析したものである。翌年の JSAI2004 (2004 年度人工知能学会全国大会) でも同様にスケジューリング支援システムを運用した。スケジューリング支援システム for JSAI2004 (以下, SSS2004) では、本論文で解説したシステムよりも、他のシステムとの連携を強化したところに特徴がある。SSS2004 では SOAP インタフェースを用意しており、松尾らの人間関係ネットワークシステムや坂本らの Comic FOAF[66] との連携を行った。なお、SSS2004 の運用結果およびその解析結果については、巻末の付録 A にて述べる。

## 5.6 議論

### 5.6.1 近傍仲介法の実装としての人推薦

本システムでは、Know リンクを用いた人の推薦を行った。これは、知り合いの知り合いを推薦するということであり、近傍仲介法の一実装例ともいえる。5.3.7 節の考察から、知り合い (Know リンクを張った相手) が他の知り合いの知り合いであった割合は 7 割を越えていることがわかる。これはつまり近傍仲介法により推薦可能であったノードがそれだけ多いことを示している。5.3.7 節では人推薦の成功率を 15~16% と挙げたが、これは推薦対象となるノード (人) を候補の中から無作為に選んだ場合であり、人手やなんらかのアルゴリズムによる選別が入ることによって、より高くすることは可能であると考えられる。これについては、より多くの利用例を集めて分析する必要がある。

### 5.6.2 Social Networking Services との関係

Social Networking Services (SNS) は互いに友人として承認し合ったリンク (双方向 Know リンク) により人のネットワークを構築していくサービスである。SNS は本システムと似ているが、いくつか異なっている点がある。多くの SNS は参加するにはすでにメンバーとなっている人の招待が必要な仕組みとなっている。つまり、すでにできあがっているメンバーのネットワークと必ず繋がった状態で参加することとなる。そのため、途

中脱退するメンバーがいなければ、巨大なたった一つのクラスタができあがることになる。対して本システムでは、誰でも紹介無しにメンバーとなることができる。SNSではリンクを張るのに両方の承認が必要であるが、本システムはどちらか一方からだけのリンクを張ることができる。

湯田らの報告 [75] によると、国産 SNS の mixi のネットワークは Connecting Nearest-Neighbor モデルに近い成長モデルであり、スケールフリー性を持つ。また、クラスタリング係数は 0.33 であり、Know リンクネットワークの 0.47 と同様に高い。mixi 利用者は mixi を使うためにメンバーになるが、学会聴講スケジュールの支援をベースとした本システムの利用者にとっては Know リンクはあくまで追加機能にすぎない、という利用者の性質の違いを反映していくつかルールが異なっているが、結果的には概観としては似たようなネットワークが構築されることがわかる。ただし、本システムは人工知能学会という一つのコミュニティ内で運営されたシステムのため、コミュニティがすでに持っている構造（組織構造や人のネットワーク）などが深く関わっていると考えられる。これは今後分析進めていく必要がある。また、そのような分析から得られた知見が mixi 内に潜在的にある構造の解明につながれば、SNS の大規模なネットワークをコミュニティ支援や情報フィルタリングなどへ活用する助けになると考えられる。

### 5.6.3 他の観点からのネットワークとの関係

イベント空間情報支援プロジェクトでは、スケジューリング支援システム以外にもいくつかのサービスが提供された。そのサービスの中でも、人間関係ネットワークシステムと CoBIT を用いた位置情報支援システムは、本システムとは異なった観点で参加者のネットワークを抽出することができるシステムである [48]。前者は Web 上での名前の共起関係によりネットワークを、後者は学会参加者の移動履歴からネットワークを、それぞれ作成することができる。

このような異なる手法で集められたパーソナルネットワークは、それぞれを補完し合うことが可能である。スケジューリング支援システムでは、利用者が自発的に任意の相手に対して Know リンク（知り合いリンク）を追加することでパーソナルネットワークを形成する。これは利用者による明示的なパーソナルネットワークの作成であり、できあがったネットワークの信憑性は高いが、積極的に参加しない利用者の周辺のネットワークは抽出不可能である。対して人間関係ネットワークシステムを用いれば大規模なネットワークを抽出することができ、例えば Know リンクの無い部分を補完したり、もしくは Know リンクの候補として利用者に提示することで、Know リンク追加に関わる利用者の負荷を軽減することができる。移動履歴を用いることで会場内で会ったかどうかのネットワークが作成可能であるが、たまたま同じ場所にいただけなのか一緒に会っていたのかを判別する

のは難しい．そこでスケジューリング支援システムや人間関係ネットワークシステムが作成したネットワークを参照し，同じ場所にいた時間が短くとも知り合い関係であれば会って話した可能性が高く，逆に長時間一緒にいても知り合い関係でなければたまたま同じ場所にいただけ，といった推測によるネットワークの補正も可能になる．

このような同一対象をノードとしたネットワークを複数の手法によって作成する利点は，個々のネットワーク作成を支援するだけでなく，そのネットワーク自体の情報量をより豊かにするため，例えば今回のスケジューリング支援システムで行った情報推薦のような，パーソナルネットワークを用いたサービスの強化にもつながる．複数手法によりパーソナルネットワークを作成した場合にどのようにして補完し合うことが可能か，また，それらをまとめて用いてどのような情報サービスができるかといったことについては，今後検討していく必要がある．

#### 5.6.4 社会ネットワークでの調査および分析

今回はパーソナルネットワークを用いたコミュニティ支援システムということで，パーソナルネットワークを用いた情報推薦やシステム内の利用者の挙動についての分析を行った．社会学では社会ネットワーク分析と呼ばれる，調査対象となる人や企業のネットワークの構造をもとに人や企業の間を調べていく分野があり，パーソナルネットワークの分析に関しては様々な分析手法や知識が蓄積されている．これらを用いることで，今回用いたコミュニティ支援システムの利用者分析をより深く行えると考えられる．また，論文推薦のようなネットワークを利用した情報サービスについても，その性能向上に関して社会ネットワーク分析の手法が利用可能であると考えられる．

例えば，社会ネットワーク分析では中心性と呼ばれる，ネットワークにおいて中心的な位置づけにあるノードを見つける手法がいくつか存在するが [21]，これはそのネットワークにおいて注目すべき存在ということで情報推薦にそのまま用いることができる．また，あるノードとあるノードが競合関係にあるのか，それとも補完関係にあるのかといった構造的な関係の分析手法もある [72]．情報推薦を行う際にこのような技術を用いれば，どのような関係にあるから推薦しているのかといった説明に用いることができるし，より個人に特化した情報推薦を行う上でも有用である．

### 5.7 まとめ

本章ではパーソナルネットワークを導入したイベント支援システムを提案・開発し，2003年度人工知能学会全国大会にて運用し，生成されたパーソナルネットワークの分析およびその利用方法の分析を行った．



本システムの運用結果の解析により、パーソナルネットワークの重要性および有用性について検証を行った。この二つのポイントについてそれぞれまとめる。

まずパーソナルネットワークの重要性について述べる。Know リンク追加を行った人(99人)は、システム利用者(276人)の36%、リンク追加を行ったことがある人(164人)の60%を占めた。Check リンクが追加によって個人用聴講スケジュールを作成するのに対し、Know リンクは追加によって知り合いリストを作成する。つまり Know リンクを追加する見返りとして得られるものは可視化されたパーソナルネットワークだけである。しかしこれに興味を示した人が、積極的なシステム利用者(リンク追加を行った人)の半数以上を占めた。これは利用者のパーソナルネットワークへの関心の強さを示していると考えられる。

次にパーソナルネットワークの有用性について述べる。本システムではパーソナルネットワークを二種類の方法で利用した。一つはパーソナルネットワークをハイパーリンクとしてオンラインプログラム中に埋め込むことで、新しい情報アクセスの経路として提供したこと。もう一つはパーソナルネットワークを利用者のプロフィール情報と見立てて情報推薦を行ったこと、である。

前者は明示的なリンクであり一定の利用がある程度見込めるものであったが、後者の方の結果は興味深い物であった。すなわちパーソナルネットワークネットワークによる情報推薦は関心型の推薦とは異なる性質を持っていることが、利用者の行動履歴よりわかった。この結果から、パーソナルネットワークは既存のサービスを行うにしても、既存のものとは異なる新しい情報源として利用可能であることがわかった。

## 第 6 章

# 結言

### 6.1 結論

本研究では、不均質な情報が膨大に存在する WWW 環境における情報収集支援として、利用者自身が情報源となり、他の情報源との情報共有によって効率よく情報を収集する環境の構築に必要な、他の情報源との情報共有手法と、他の情報源との情報共有ネットワークの改善手法および利用方法の提案と検証を行った。

2 章では、個人間の情報共有の手段として、個人が独自に作成した階層構造を用いることが有効であることを、WWW ブックマーク共有システム kMedia を用いて被験者実験により検証した。異なる観点で情報を収集した情報源の間で情報を共有する方法は様々考えられるが、本研究では情報源が概念階層によって情報を管理している場合、その階層間の類似性に基づいて情報を共有することが有効であることを被験者実験を通して検証した。

3 章では、2 章における階層構造を用いた情報共有の有効性を背景に、より高性能な階層構造の統合技術について検討した。WebHical は木構造によって情報を管理している情報源間での情報共有を支援するシステムであり、階層が持つ構造情報と、階層内に分類されている情報の内容とを使うことで、階層間の類似性を求める。WebHical は市瀬らが提案した Hical の拡張版であるが、Hical では発見不可能だった階層間の類似性を検出できるようになった。本研究ではさらに高精度な共有を実現する手法として Hical に分類学習器を用いた Hical+NB を提案した。

4 章では、情報共有ネットワークを拡張する手法について述べた。提案する近傍仲介法は、すでに個々の情報源間の関係に基づいて情報共有ネットワークが形成された状態において、そのネットワークをさらに発展させるための手法である。全体を把握できていない状態で情報共有を行った場合、新しい共有相手の発見が大きな問題となる。そこで本研究では、すでに情報を共有している相手が自分にあった新しい共有相手を探索してくれる、

つまり友達の友達を紹介してもらうことでこの問題を解決する手法を提案した。シミュレーション実験により本手法によるネットワークの再構築が機能するかどうかを検証し、トップダウン的に作られた仮想的な最適ネットワークとの比較によりその有効性が確認できた。

5章では、実際に人（情報源）のネットワークを用いた情報共有システム上で、利用者がどのように振る舞うのか、どのような人のネットワークが獲得されるのかを実験した。分析結果より、人のネットワークはスケールフリーに似た構造を形成すること、人がボトムアップに作ったネットワークの構造を用いた情報推薦は、一般的な協調フィルタリングによる情報推薦とは異なった性質を持つことがわかった。この結果は、個々の情報源がネットワーク状に連なって情報共有をするというモデルは、単なる分散型の協調フィルタリングとは異なった性質を持っている事を示唆する結果と言える。

情報共有システムの様な利用者の積極的参加が不可欠なシステムにおいては、利用者のシステム利用に対するモチベーションの維持が重要な課題となる。そして、利用者のシステム利用の一番シンプルなモチベーションは自分に見返りがあることである。

システム利用による見返りには、Instant Gratification（即時的な見返り）と、Delayed Gratification（長期的な見返り）の二種類があると考えられる。前者はシステムを利用したらすぐにわかるメリットであり、利用者の獲得に不可欠なものである。後者はシステムを継続的に利用していると得られるメリットであり、利用者を継続利用させるのに不可欠なものである。

情報共有システムが提供するサービスは、情報が蓄積されるほどに効果が増大するという点で、どちらかというとも Delayed Gratification の傾向が強いと言える。しかしながら、まず情報を増やすには利用者を獲得する必要があり、組織内システムのように利用者を強制的に集められるような利用携帯でない限り、Instant Gratification を提供する仕組みが不可欠である。このことから、情報共有システムは明確な Instant Gratification を用意する必要があり、さらにそこから自然に Delayed Gratification へとつなげる必要があると言える。

2章で取り上げた *kMedia* では Web ページを対象とした情報推薦を行うことで Instant Gratification を提供している。そして概念階層単位での情報推薦、カテゴライズ近似度や近傍仲介法による共有相手の選別は、この情報推薦をより良くするものであると言える。同時に、概念階層単位での情報共有を他の複数の情報源との間で行うことで、一度きりのアイテムの推薦に留まらず、継続的に関連する情報を得ることができる。これは自分が持つ概念体系に合わせた概念階層によるものなので、自分の観点で情報を整理していくことにつながる。3章にて提案した、より高性能な概念階層の類似性発見手法を用いることで巨大な概念階層にも対応することができるが、これは長期的に自分や周辺の情報源が収集した情報の管理を支援するもので、より強力な Delayed Gratification を提供する。

また、本研究が提案するモデルでは、共有情報を一カ所に集めるのではなく情報源単位で管理し、その情報源間で共有ネットワークを構成し、結果的に全体で情報共有を行う。このような分散型の情報共有は、スタートアップの容易性向上に貢献すると共に、できあがった情報源ネットワーク自体が新しい価値を持つ。5章の実験結果より、情報源のネットワークを新しい情報へのアクセス経路として用いたり、ネットワークの情報をを用いた新しい情報推薦の提供が可能であることがわかった。これらは、利用者がシステムを継続的に使うことで新しいメリットを享受できるようになるということで、一つの Delayed Gratification であると言える。

本研究では、情報共有における基礎的な問題として、興味の間定および共有相手の選択を支援する技術の提案および検証を行ってきたが、本研究で提案する手法は、情報共有において同じく不可欠な利用者のモチベーション維持という問題に対しても有効な解決策を提示していると考えられる。

## 6.2 課題と今後の展望

本研究は大きく二つに分かれる。一つは異なる概念体系を持つ情報源間での情報共有支援であり、もう一つは情報源ネットワークの改善およびネットワーク構造を利用した情報収集支援である。それぞれについて今後の展望を述べる。

まず前者については、今後は二つの概念体系間での情報共有に留まらず、より多くの概念体系間での情報共有を実現するための手法を提案したいと考える。方法としては、三つの概念体系を統合して一つの新しい概念体系を作り出すことが考えられる。これを実現するためには、類似性が発見されなかった階層が、相手の概念階層においてどこかに追加可能な概念なのか、それともまったく追加できない物なのかを判別できる必要がある。このような統合による新しい概念体系の作成が可能になれば、一般に個人が巨大な概念階層を構築するのは困難であるが、本手法を用いることで互いに補完し合うことで巨大な概念階層も構築可能ではないかと考える。

次に後者について述べる。近傍仲介法の実験ではエッジの評価値のみを用いてネットワークの改善を図った。しかし今回想定しているネットワークは情報共有ネットワークであり、実際にはネットワークの上を情報が流通する。よってネットワークはその流通した情報の良し悪しで評価されるのが、より正確であると考えられる。今後はノードとエッジに加え、コンテンツという概念を取り込んだネットワークシミュレーションによって、近傍仲介法と情報共有ネットワークおよび情報流通の関係について検証を行いたい。

また、本研究では要素技術の提案と検証に留まっているが、今後は手法のブラッシュアップと併行して、各手法を用いたシステムの実装および運用により、実環境にて提案手法の有効性を確かめることが重要であると考えられる。Web コンテンツを対象とした情報共

有システムに関する研究は様々なされているが、実サービスとして利用できるものはまだ見かけないのが現状である。これは、本論文の冒頭に述べたように、Web コンテンツが膨大かつ多様であるがために効率の良い情報共有が困難であるからと考えられる。また、個人が収集できる情報には限りがあり、またそれを管理する環境も共有する風潮も無かったのがこれまでであった。しかしながら、最近ではRSSのような Semantic Web 技術により情報流通が向上し、個人が収集できる情報は飛躍的に増加した。また、Weblog や Social Networking Service の普及により、人が持つ情報がオンライン上に集まるようになってきており、「ネットワーク上の個人」が持つ情報はリッチになりつつある。これは情報共有による情報収集支援環境が構築される土台ができつつある兆しと思われ、情報共有支援技術を実環境にて試す絶好の時期であると考えられる。

# 謝 辞

論文を締めくくるに当たり、本論文を作成するに至った5年間の研究生活でお世話になった方々への感謝の気持ちを述べたいと思います。

まず、奈良先端科学技術大学院大学での博士前期課程在学中の2年間、指導教官をしてくださいました木戸出正継教授、総合研究大学院大学での博士後期課程在学中に指導教官をしてくださいました上野晴樹教授、山田誠二教授に心より感謝いたします。木戸出教授には研究内容に関する鋭いご意見を数多くいただきました。ありがとうございました。上野教授は、私の博士後期課程進学での指導教官を快く引き受けて下さりました。心より感謝いたします。山田教授には本論文を作成するにあたり重要なアドバイスを数多くいただきました。心より感謝いたします。

審査委員に加わって下さった国立情報学研究所の相澤彰子教授、東京大学の中小路特任教授、京都大学の角康之助教授には、本論文を作成するにあたり、適切なご指摘および有益なコメントをいただきました。心より感謝いたします。

武田英明教授には奈良先端科学技術大学院大学での博士前期課程からの5年間に渡る長い間、ご指導いただきました。本研究を進めるに当たって参考になった先生との議論やコメントは数知れません。研究に行き詰まる事が幾度と無くありましたが、そのたびに暖かくかつ適切な指導により方向性を示して下さい、5年間の研究を導いて下さりました。未熟な私を博士論文まで指導して下さい事に深く感謝いたします。ありがとうございました。

市瀬龍太郎助手は、本研究で *Hical* の改良版を開発するにあたり、親身になって手助けをして下さいました。特に3章の研究は市瀬助手の協力無しではありえませんでした。ありがとうございました。

本研究の一部は独立行政法人産業技術総合研究所との共同研究として行われました。同研究所の西村拓一研究員と松尾豊研究員には、研究に関する有益なたくさんアドバイスをいただきました。深く感謝いたします。本共同研究では、他にも数多くの方と出会い、様々な場面でお世話になりました。厚く感謝いたします。

東京大学ものづくり経営研究センターの安田雪特任助教授には、5章の研究をまとめる

上で不可欠なネットワーク分析に関する専門的なご意見を多数いただきました。心より感謝いたします。

私が3年間すごした総合研究大学院大学情報学専攻では多くの研究スタッフや学友に出会えることができました。ムリアディ・ヘンドリーさん、事務補佐員の内田昭子さん、平井直子さん、同期の大向一輝君、小林一樹君、山田太造君、後輩の沼晃介君、鈴木聡君、川崎直之君ほか、多くの方々との出会い、研究に関する相談や議論をしたり、日常の様々な場面で共に経験を共有したりすることを通して有意義な学生生活を送ることができました。ありがとうございました。

同様に、博士前期課程2年間を過ごした奈良先端科学技術大学院大学でも多くの研究スタッフや学友に出会えることができました。河野恭之助教授、上野敦志助手（現、大阪市立大学勤務）、久米出助手、浮田宗伯助手、事務補佐員の谷村優香里さん、峰真知子さん、先輩の高橋徹さん（現、ATR勤務）、河村竜幸さん（現、奈良先端科学技術大学院大学勤務）、谷口雄一郎さん（現、プラエセンス勤務）、同期の上岡隆宏君ほか、多くの方々との出会いました。皆さんとの議論や共に過ごした日々が、私が研究者として歩き出すきっかけを作ってくださいました。ありがとうございました。

研究者になるきっかけという点では、同志社大学での卒業研究が全ての始まりでした。私に研究する楽しさを教えて下さりました、三木光範教授と廣安知之助教授に心より感謝いたします。

最後に、学生生活を送るにあたって経済的・精神的に支援をして下さった両親に心より感謝します。

## 参考文献

- [1] 汎用連想計算エンジン(第2版)導入・操作マニュアル, 2001. <http://geta.ex.nii.ac.jp/getaN2001/gdoc/manual.html>.
- [2] del.icio.us, 2004. <http://del.icio.us/>.
- [3] Rakesh Agrawal and Ramakrishnan Srikant. On integrating catalogs. In *Proceedings of The 10th International World Wide Web Conference (WWW-10)*, 5 2001.
- [4] 相澤彰子. 双対的クラスタリングによる情報空間のモデル化. 2002年度人工知能学会全国大会論文集, 2002.
- [5] Luis A. Nunes Amaral, A. Scala, M. Barthelemy, and H. E. Stanley. Classes of small-world networks. *Applied Physical Science*, Vol. 97, No. 21, pp. 11149–11152, 10 2000.
- [6] Six Apart. Movable type, 2003. <http://www.movabletype.org/>.
- [7] Marko Balabanovic and Yoav Shoham. Fab: Content-based, collaborative recommendation. *Communications of the ACM*, Vol. 40, No. 3, pp. 66 – 72, 1997.
- [8] Albert-Laszlo Barabasi. *Linked: The New Science of Networks*. Perseus Publishing, 5 2002.
- [9] Albert-Laszlo Barabasi. 新ネットワーク思考. NHK 出版, 12 2002.
- [10] Vladimir Batagelj and Andrej Mrvar. Pajek - program for large network analysis. *Connections*, Vol. 21, pp. 47–57, 1998.
- [11] Nicholas J. Bekin and W. Bruce Croft. Information filtering and information retrieval: Two sides of the same coin? *Communication of the ACM*, Vol. 35, No. 12, pp. 29–38, 12 1992.
- [12] L. Chen and K. Sycara. Webmate: A personal agent for browsing and searching. In *Proceedings of the 2nd International Conference on Autonomous Agents and Multi Agent Systems, AGENTS'98*, pp. 132–139, 1998.
- [13] CHI2001. Chi place, 2001. <http://www.chiplace.org/index.jsp>.



- [14] World Wide Web Consortium. Resource description framework (rdf) model and syntax specification, 1999. <http://www.w3c.org/TR/REC-rdf-syntax/>.
- [15] Douglass R. Cutting, David R. Karger, Jan O. Pederson, and John W. Tukey. Scatter/gather: A cluster-based approach to browsing large document collections. In *15th Annual International ACM/SIGIR Conference*, pp. 318–329, 1992.
- [16] Anind K. Dey, Masayasu Futakawa, D. Salber, and G. D. Abowd. The conference assistant: Combining context-awareness with wearable computing. In *Proceedings of the 3rd International Symposium on Wearable Computers*, pp. 21–28, 10 1999.
- [17] Judith S. Donath. Visual who: Animating the affinities and activities of an electronic community. In *ACM Multimedia 95*, 1995.
- [18] eMercury. Mixi, 2004. <http://mixi.jp/>.
- [19] Thomas Erickson, David N. Smith, Wendy A. Kellogg, Mark Laff, John T. Richards, and Erin Bradner. Socially translucent systems: Social proxies, persistent conversation, and the design of “babble”. In *CHI*, pp. 72–79, 1999.
- [20] J.L. Fleiss. *Statistical Methods for Rates and Proportions*. John Wiley & Sons, 1973.
- [21] Linton C. Freeman. Centrality in social networks conceptual clarification. *Social Networks*, Vol. 1, pp. 215–239, 1978.
- [22] Friendster. Friendster, 2004. <http://www.friendster.com/>.
- [23] Andreas Girgensohn and Alison Lee. Making web sites be places for social interaction. In *the 2002 ACM Conference on Computer Supported Cooperative Works*, pp. 136–145, 2002. Available from <http://www.webcollab.com/alee/papers/cscw02.html>.
- [24] Google. Orkut, 2004. <http://www.orkut.com/>.
- [25] RDF Site Summary 1.0 Specification Working Group. Rdf site summary (rss) 1.0, 2001. <http://web.resource.org/rss/1.0/spec>.
- [26] 原田昌紀, 佐藤進也, 風間一洋. Web 上のキーパーソンの発見と関係の可視化. 情報処理学会研究会報告 DBS-130/FI-71, 2003.
- [27] Jonathan L. Herlocker, Joseph A. Konstan, Al Borchers, and John Riedl. An algorithmic framework for performing collaborative filtering. In *Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval*, pp. 230–237, 1999.
- [28] B. A. Huberman and M. Kaminsky. Beehive: A system for cooperative filtering and sharing of information, 1996. <ftp://parcftp.xerox.com/pub/dynamics/beehive.html>.

- [29] 市瀬龍太郎, 濱崎雅弘, 武田英明. 階層的分類データを統合するための規則学習機構. 人工知能学会論文誌, Vol. 19, No. 6, pp. 521–529, 2004.
- [30] 市瀬龍太郎, 武田英明, 本位田真一. 階層的知識間の調整規則の学習. 人工知能学会論文誌, Vol. 17, No. 3, pp. 256–278, 2002.
- [31] 石田亨, 西村俊和, 八槇博史, 後藤忠広, 西部喜康, 和氣弘明, 森原一郎, 服部文夫, 西田豊明, 武田英明, 沢田篤史, 前田晴美. モバイルコンピューティングによる国際会議支援. 情報処理学会論文誌, Vol. 39, No. 10, pp. 2855–2865, 1998.
- [32] Anthony Jameson and Eric Schwarzkopf. Pros and cons of controllability: An empirical study. In *Proceedings of Adaptive Hypermedia and Adaptive Web-Based Systems*, 2002. Available from <http://dfki.de/~jameson/abs/JamesonS02.html>.
- [33] T. Joachims, D. Freitag, and T. Mitchell. Webwatcher: A tour guide for the world wide web. In *Proceedings of the Fifteenth International Joint Conference on Artificial Intelligence (IJCAI-97)*, 1997.
- [34] Koji Kamei, Eva Jettmar, Kunihiko Fujita, Sen Yoshida, and Kazuhiro Kuwabara. Community organizer: Supporting the formation of network communities through spatial representation. In *the 2001 Symposium on Applications and the Internet (SAINT'01)*, 2001.
- [35] 片岡充照, 今中武, 水谷研治, 若見昇. テキスト情報を対象としたキーワード抽出と関連情報収集システム. 日本ファジイ学会誌, Vol. 9, No. 5, pp. 710–717, 1997.
- [36] Henry Kautz, Bart Selman, and Mehul Shah. Referral web: Combining social networks and collaborative filtering. *Communications of the ACM*, Vol. 40, No. 3, pp. 63–65, 1997.
- [37] 小島一浩. 自己組織化型 p2p コンテンツ共有・検索エンジン-tellagate-の開発. 情報処理学会研究報告, 第 2004-ICS-136 巻, pp. 123–130, 2004.
- [38] Richard M. Keller, Shawn Wolfe, James R. Chen, Joshua L. Rabinowitz, and Nathalie Mathe. A bookmarking service for organizing and sharing urls. In *Proceedings of The 6th International World Wide Web Conference (WWW-6)*, 1997.
- [39] Joseph A. Konstan, Bradley N. Miller, David Maltz, Jonathan L. Herlocker, Lee R. Gorden, and John Riedl. Grouplens: Applying collaborative filtering to usenet news. *Communications of the ACM*, Vol. 40, No. 3, pp. 76–87, 1997.
- [40] Wen-Syan Li, Quoc Vu, Divyakant Agrawal, Yoshinori Hara, and Hajime Takano. Powerbookmarks: A system for personalizable web information organization, sharing, and management. In *Proceedings of The 8th International World Wide Web Conference (WWW-8)*, 1999.

- [41] Henry Lieberman. Letizia: An agent that assist web browsing. In *Proceedings of IJCAI-95*, pp. 924–929, 1995.
- [42] Paolo Massa and Paolo Avesani. Trust-aware collaborative filtering for recommender systems. In *Proceedings of International Conference on Cooperative Information Systems (CoopIS)*, 2004.
- [43] 松本裕次, 北内啓, 山下達雄, 平野善隆. 日本語形態素解析システム『茶筌』version2.2.0 使用説明書, 1999. <http://isw3.aist-nara.ac.jp/IS/TechReport2/report/99012.ps>.
- [44] 松尾豊, 友部博教, 橋田浩一, 石塚満. Web から人間関係ネットワークの抽出と情報支援. 第 17 回人工知能学会全国大会講演論文集, 2003.
- [45] Andrew Kachites McCallum. Bow: A toolkit for statistical language modeling, text retrieval, classification and clustering. <http://www.cs.cmu.edu/mccallum/bow>, 1996.
- [46] Mikihiro Mori and Seiji Yamada. Bookmark-agent: Information sharing of urls. In *Poster Proceedings of The 8th International World Wide Web Conference (WWW-8)*, 1999.
- [47] M. E. J. Newman. Assortative mixing in networks. *Physical Review Letters*, Vol. 89, No. 20, p. 208701, 11 2002.
- [48] 西村拓一, 濱崎雅弘, 松尾豊, 大向一輝, 友部博教, 武田英明. 2003 年度人工知能学会全国大会支援統合システム. 人工知能学会学会誌, Vol. 18, No. 1, 2004.
- [49] 西村拓一, 伊藤日出男, 山本吉伸, 中島秀之. 無電源小型通信端末を用いた位置に基づく情報支援システム. 情報処理学会 知的都市基盤研究グループ研究報告, pp. 1–6, 2002.
- [50] 緒方広明, 古郡延子, 金群, 矢野米雄. 分散型人脈活用支援システム peco-mediator-ii の構築. 電子情報通信学会論文誌, 第 J80-D-I 巻, pp. 551–560, 7 1997.
- [51] 大向一輝, 武田英明, 三木光範. 多様かつ曖昧な個人タスクのための管理システムの提案と実装. エージェント合同シンポジウム (JAWS2002) 講演論文集, pp. 502–509, 2002.
- [52] Michel Plu, Pascal Bellec, Layda Agosto, and Walter Van De Velde. The web of people: A dual view on the www. In *Proceedings of The 12th International World Wide Web Conference*, 2003.
- [53] Uta Priss. Introduction to and overview of formal concept analysis. *Annual Review of Information Science and Technology*, Vol. 40, , to appear.
- [54] James Rucker and Marcos J. Polanco. Siteseer: Personalized navigation for the web. *Communications of the ACM*, Vol. 40, No. 3, pp. 73–75, 1997.

- [55] Eric Schwarzkopf. An adaptive web site for the um2001 conference. In *Proceedings of the UM2001 Workshop on Machine Learning for User Modeling*, pp. 77–86, 7 2001.
- [56] Upendra Shardanand and Patti Maes. Social information filtering: Algorithms for automating “word of mouth”. In *CHI*, pp. 210–217, 1995.
- [57] A. Singhal, C. Buckley, and M Mitra. Pivoted document length normalization. In *Proceedings of SIGIR*, pp. 21–29, 1996.
- [58] 杉山公造, 永田晃也, 下嶋篤. ナレッジサイエンス. 紀伊国屋書店, 12 2002.
- [59] 角康之. Jsai2000 デジタルアシスタントプロジェクトの報告. 人工知能学会誌, Vol. 15, No. 6, pp. 1012–1026, 11 2000.
- [60] Yasuyuki Sumi and Kenji Mase. Collecting, visualizing, and exchanging personal interests and experiences in communities. In *2001 International Conference on Web Intelligence (WI-01)*, 2001.
- [61] Yasuyuki Sumi and Kenji Mase. Conference assistant system for supporting knowledge sharing in academic communities. *Interacting with Computers*, Vol. 14, No. 6, pp. 713–737, 12 2002.
- [62] 角康之, 坂本竜基, 中尾恵子, 間瀬健二. コミックダイアリ: 経験や興味を伝え合うための漫画日記. インタラクシオン 2002 論文集, pp. 101–108, 3 2002.
- [63] 武田英明. インターネットにおける情報の知的利用 -情報収集から情報 統合へ-. 専門講習会講演論文集「インターネットの最新技術と 家庭への浸透」. 電子情報通信学会 関西支部, 1998.
- [64] 竹内亨, 鎌原淳三, 佐伯勇, 寺岡伸悟, 原田隆司, 下條真司, 宮原秀夫. 携帯端末を用いた情報伝播モデルによる実験に基づいた情報伝達力の評価. 第 13 回データ工学ワークショップ, 2002.
- [65] 竹内亨, 鎌原淳三, 下條真司, 宮原秀夫. ユーザの関連性に基づいた情報伝播モデル. 第 12 回データ工学ワークショップ, 2001.
- [66] 田中郁, 坂本竜基, 小暮潔, 國藤進. 複数の学会支援サービス群からのパーソナルネットワークの抽出及び視覚化. 第 5 回セマンティックウェブとオントロジー研究会, 第 SIG-SWO-A402-02 巻, 2004.
- [67] Loren Terveen, Will Hill, Brian Amento, David McDonald, and Josh Creter. Phoaks: A system for sharing recommendations. *Communications of the ACM*, Vol. 40, No. 3, pp. 59–62, 1997.
- [68] Alexei Vazquez. Growing network with local rules: Preferential attachment, clustering hierarchy, and degree correlations. *Physical Review E*, Vol. 67, p. 056104, 2003.

- [69] Duncan J. Watts. *Small Worlds: The Dynamics of Networks Between Order and Randomness*. Princeton University Press, 1999.
- [70] Duncan J. Watts and Steven H. Strogatz. Collective dynamics of 'small-world' networks. *nature*, pp. 440–442, 6 1998.
- [71] 安田雪. ネットワーク分析 何が行為を決定するか. 新曜社, 1997.
- [72] 安田雪. 実践ネットワーク分析 関係を解く理論と技法. 新曜社, 2001.
- [73] 吉田匡志, 伊藤雄介, 沼尾正行. 口コミによる分散型情報収集システム. 第 10 回マルチ・エージェントと協調計算ワークショップ (MACC2001), pp. 61–68, 11 2001.
- [74] 田中良和. Gree, 2004. <http://gree.jp/>.
- [75] 湯田聡夫, 小野直亮, 藤原義久. ソーシャル・ネットワーキング・サービスのリンク特性とクラスター構造. ネットワーク生態学 2005 シンポジウム, pp. 1–8, 2005.
- [76] 松本裕次, 北内啓, 山下達雄, 平野善隆. 日本語形態素解析システム『茶筌』version2.2.0 使用説明書. NAIST Technical Report NAIST-IS-TR99012, 奈良先端科学技術大学院大学情報科学研究科, 1999.
- [77] 林幸雄, 宮崎敏幸. Sf ネットワークモデルの特徴比較. 情報処理学会研究報告, 第 2004-ICS-136 巻, pp. 9–16, 2004.
- [78] 濱崎雅弘, 武田英明, 松塚健, 谷口雄一郎, 河野恭之, 木戸出正継. Bookmark からの共通話題ネットワークの発見手法の提案とその評価. 人工知能学会論文誌, Vol. 17, No. 3, pp. 276–284, 2002.

# 研究業績

## 主著論文

### 学術論文（査読付き）

1. 濱崎雅弘, 武田英明, 松塚健, 谷口雄一郎, 河野恭之, 木戸出正継. Bookmark からの共通話題ネットワークの発見手法の提案とその評価. 人工知能学会論文誌, Vol. 17, No. 3, pp. 276–284, 2002.
2. 濱崎雅弘, 武田英明, 大向一輝, 市瀬龍太郎. パーソナルネットワークを利用したコミュニティシステムの提案と分析. 人工知能学会論文誌, Vol. 19, No. 5, pp. 389–398, 2004.
3. 濱崎雅弘, 武田英明, 大向一輝, 市瀬龍太郎. 学術会議における共有型スケジューリング支援システムの開発と運用. 日本データベース学会 Letters (DBSJ Letters), Vol. 2, No. 4, pp. 7–10, 11 2004.

## 国際会議

### 口頭発表

1. Masahiro Hamasaki and Hideaki Takeda. Experimental results for a method to discover human relationship based on www bookmarks. In *Fifth International Conference on Knowledge-Based Intelligent Information Engineering Systems & Allied Thchnologies (KES-2001)*, pp. 1291–1295, 2001.
2. Masahiro Hamasaki and Hideaki Takeda. Find better friends? - re-configuration of personal networks by the neighborhood matchmaker method -. In *INTERNATIONAL WORKSHOP ON SEMANTIC WEB FOUNDATIONS AND APPLICATION TECHNOLOGIES (SWFAT)*, 3 2003.
3. Masahiro Hamasaki and Hideaki Takeda. Neighborhood matchmaker method: A decentralized optimization algorithm for personal human network. In *Pro-*

*ceedings of Seventh International Conference on Knowledge-Based Intelligent Information & Engineering Systems (KES2003)*, pp. 929–935, 9 2003.

4. Masahiro Hamasaki and Hideaki Takeda. Proposal of decentralized information sharing system using local matchmaking. In *Proceedings of 2003 IEEE Pacific Rim Conference on Communications, Computers and Signal Processing (PACRIM'03)*, pp. 631–634, 8 2003.

#### ポスター発表

1. Masahiro Hamasaki, Hideaki Takeda, Ikki Omukai, and Ryutaro Ichise. Scheduling support system for academic conferences based on interpersonal networks. In *Demonstration and Poster Proceedings of Hypertext2004*, pp. 50–51, 8 2004.

#### 全国大会・研究会

1. 濱崎雅弘, 市瀬龍太郎, 武田英明. 階層的知識と内容的類似性を用いたインターネットディレクトリの統合. 第17回人工知能学会全国大会論文集, 6 2003.
2. 濱崎雅弘, 武田英明. ネットワークの自律的構築のための近傍仲介法の提案とシミュレーションによる検証. エージェント合同シンポジウム (JAWS2002), 11 2002.
3. 濱崎雅弘, 武田英明. 仲介者を用いた人のネットワーク拡張手法の提案. 第16回人工知能学会全国大会論文集, 5 2002.
4. 濱崎雅弘, 武田英明, 河野恭之, 木戸出正継. kmedia: ブックマークからの共通話題ネットワークの発見手法とその評価. 電子情報通信学会技術研究報告, 第100巻, pp. 65–72, 2001.
5. 濱崎雅弘, 武田英明, 河野恭之, 木戸出正継. ブックマークを用いた人の繋がり発見手法. 第15回人工知能学会全国大会論文集, 5 2001.
6. 濱崎雅弘, 武田英明, 河野恭之, 木戸出正継. 知識の階層構造を用いた協調的情報推薦ネットワークの提案. 第10回マルチエージェントと協調計算ワークショップ, pp. 69–70, 2001.
7. 濱崎雅弘, 武田英明, 大向一輝, 市瀬龍太郎. 学術会議における共有型スケジューリング支援システムの開発と運用. データベースとWeb情報システムに関するシンポジウム (DBWeb2003), pp. 147–154, 11 2003.
8. 濱崎雅弘, 武田英明, 大向一輝, 市瀬龍太郎. 流通情報に基づいた自律分散最適化を行うネットワーク型情報共有システムの提案. 合同マルチエージェントシンポジウ

△ (JAWS2003), 10 2003.

9. 濱崎雅弘, 武田英明, 大向一輝, 市瀬龍太郎. 2003 年度スケジューリング支援システムの開発と運用. 第 18 回人工知能学会全国大会論文集, 6 2004.
10. 濱崎雅弘, 武田英明, 大向一輝, 市瀬龍太郎. コミュニティシステムのためのパーソナルネットワークの利用とその分析. 第 18 回人工知能学会全国大会論文集, 6 2004.

## 主著以外の論文

### 学術論文 ( 査読付き )

1. 市瀬龍太郎, 濱崎雅弘, 武田英明. 階層的分類データを統合するための規則学習機構. 人工知能学会論文誌, Vol. 19, No. 6, pp. 521–529, 2004.

## 国際会議

### 口頭発表

1. Ryutaro Ichise, Masahiro Hamasaki, and Hideaki Takeda. A hybrid algorithm for alignment of concept hierarchies. In *In Proceedings of the 14th International Conference on Knowledge Engineering and Knowledge Management*. Springer, 2004.
2. Ryutaro Ichise, Masahiro Hamasaki, and Hideaki Takeda. Discovering relationships among catalogs. In *Proceedings of the 7th International Conference on Discovery Science*, Vol. 3245 of *LNAI*, pp. 371–379. Springer, 2004.
3. Ikki Ohmukai, Hideaki Takeda, Kosuke Numa, Masahiro Hamasaki, and Shin Adachi. Personal knowledge publishing suite with weblog. In *Workshop on Weblogging Ecosystem, Held in conjunction with the Thirteenth International World Wide Web Conference (WWW2004)*, 2004.
4. Ikki Ohmukai, Hideaki Takeda, Kosuke Numa, Masahiro Hamasaki, and Shin Adachi. Metadata-driven personal knowledge publishing. In *Proceedings of the Third International Semantic Web Conference (ISWC2004)*, 2004.
5. Yutaka Matsuo, Masahiro Hamasaki, Junichiro Mori, Hideaki Takeda, Koiti Hasida. Ontological Consideration on Human Relationship Vocabulary for FOAF. In *1st Workshop on Friend of a Friend, Social Networking and the Semantic Web*, 2004.



### ポスター発表

1. Ryutaro Ichise, Masahiro Hamasaki, and Hideaki Takeda. A multi-strategy approach for catalog integration. In In Chengqi Zhang, Hans W. Guesgen, and Wai K. Yeap, editors, *Proceedings of 8th Pacific Rim International Conference on Artificial Intelligence*, 2004.
2. Kosuke Numa, Ikki Ohmukai, Masahiro Hamasaki, and Hideaki Takeda. Ego-centric search based on rss. In *Proceedings of the Third International Semantic Web Conference (ISWC2004)*, 2004.

### 全国大会・研究会

1. 大向一輝, 濱崎雅弘, 武田英明, 三木光範. Social scheduler - p2p モデルを用いた協調的リソースプランナの提案. 第16回人工知能学会全国大会講演論文集, 2002.
2. 市瀬龍太郎, 濱崎雅弘, 武田英明. 階層的分類体系の統合手法の比較. 人工知能学会研究会資料, 第SIG-FPAI-A04巻, pp. 79-84, 2004.
3. 沼晃介, 大向一輝, 濱崎雅弘, 武田英明. Weblogにおけるエゴセントリック検索の提案と実装. 人工知能学会セマンティックウェブとオントロジー研究会, SIG-SWO-A401-06, 2004.
4. 沼晃介, 大向一輝, 濱崎雅弘, 武田英明. Weblogにおける文書作成支援のためのエゴセントリック検索. 第18回人工知能学会全国大会論文集, 2C2-06, 2004.
5. 市瀬龍太郎, 濱崎雅弘, 武田英明. Web文書を階層的に分類するための複数分類器の利用. 第18回人工知能学会全国大会講演論文集, 2004.
6. 武田英明, 松尾豊, 濱崎雅弘, 西村拓一. イベント空間支援の可能性. 第18回人工知能学会全国大会講演論文集, 2004.
7. 安随晋太郎, 濱崎雅弘, 大向一輝, 武田英明, 山口高平. 領域オントロジーと個人嗜好情報に基づくモバイル型レコメンデーションシステムの構成. 電気情報通信学会技術報告「知能ソフトウェア工学」, Vol. 104, No. 588, pp. 43-48, 2005.

### その他

#### 解説記事

1. 西村拓一, 濱崎雅弘, 松尾豊, 大向一輝, 友部博教, 武田英明. 2003年度人工知能学会全国大会支援統合システム. 人工知能学会学会誌, Vol. 18, No. 1, 2004.

## 表彰

1. 第 18 回人工知能学会全国大会 優秀賞

# 付録 A スケジューリング支援システム for JSAI2004

## 昨年度システムとの違い

本プロジェクトでは、昨年度も同様のコンセプトで同じシステムを提供した。昨年度と今年のものとの違いを説明する。

### アクセスコントロール

JSAI2003 は個人が作成したスケジュールと知り合いリストは、その持ち主が知り合いと認めた人（Know リンクを張った相手）しか見られなかったが、JSAI2004 では誰でも見られるようになった。

これは昨年度のシステムの分析結果より、被 Know リンクは多いが自分では Know リンクを張らない人（有名な先生に多い）の存在が、人のネットワークによるアクセス経路に行き止まりを作ってしまうことがわかったためである。また、昨今 Social Networking Service (SNS) の普及により、人のネットワークをオンラインにのせて公開することに対して抵抗が少なくなりつつあるということも理由の一つとして挙げられる。

### 掲示板機能

JSAI2003 では各論文ごとにスケジュールリング支援システム独自の掲示板を設置していた。JSAI2004 では、発表 Weblog と呼んでいる Weblog ツール (Movable Type) を利用した日記&コメント掲示板システムに、掲示板機能を全て任せた。

なお、掲示板ではなく個人宛にメッセージを送信する機能は、JSAI2003 も JSAI2004 も同様に実装した。

## 知り合い推薦

JSAI2003 では Check リンクや Know リンクを用いて人の推薦を行った。しかし、推薦に対して行えるアクションが Know リンクの追加のみであったため（正確にはメッセージ送信もあったが、ほとんど使われなかったのでここでは考えないでおく）、Check リンクを元にして興味の近い人を推薦したとしても、それが Know リンクを張る対象にはなりにくいという問題があった。

そこで JSAI2004 では、人間関係ネットワークシステムが抽出した人間関係を元に、人間関係ネットワークシステムのデータでは利用者と関係が深いと判断されているのにまだ Know リンクを張っていない相手を推薦するサービスを行った。

## 人との関係表示機能

出会い支援の一環として、自分と任意の相手との関係を表示する機能を新たに追加した。人間関係ネットワークシステムから得た知り合いネットワークを元に、自分と任意の相手との間をつなぐ人を表示する。また、自分と相手とで共通する Check リンク一覧も表示する。

## SOAP インタフェース

JSAI2003 では、利用者が異なるシステム間をハイパーリンクを介して移動する際に、GET リクエストや POST リクエストを用いてシステム間のデータ交換を行っていた。JSAI2004 では、SOAP インタフェースを各サービスが用意することによって、各サービスが自由にデータ交換を行えるようになった。人の推薦や、人と人との関係表示機能は、いずれも SOAP インタフェースを介して人間関係ネットワークシステムから受け取ったデータを元に行っている。

## 他のシステムとの連携

本章ではスケジューリング支援システムと他のシステムとの連携について説明を行う。スケジューリング支援システムは他の 4 つのシステムとの連携を行った。人間関係ネットワークシステムと Comic FOAF は SOAP を用いて、位置情報システムと発表 Weblog とは POST リクエスト付きのハイパーリンクを用いて連携を行った。

## 位置情報システム

位置情報システムは、CoBIT の位置情報履歴を元に、任意の利用者が現在会場内のどこにいるかを提示するシステムである。

スケジューリング支援システムでは、人のページに位置検索ボタンを用意している。ボタンをクリックすると位置情報システムへジャンプし、そのページの持ち主がどこにいるかが表示される。

## 人間関係ネットワークシステム

人間関係ネットワークシステムは、Web コンテンツ中の記述を元に、あらかじめ登録した人々の関係を自動的に抽出するシステムである。

スケジューリング支援システムでは、人間関係ネットワークシステムが抽出した人間関係データを元に、知り合い推薦サービスを行った。さらに、自分と相手との間の人のつながりを示すサービスも、人間関係ネットワークシステムのデータを用いて行った。

人間関係ネットワークシステムでは、抽出した人間関係データをグラフ表示するサービスを行っている。そのグラフ表示に用いるリンクデータのの一つとして Know リンクや Check リンクの類似関係をスケジューリング支援システムから提供した。

## Comic FOAF

Comic FOAF とは FOAF データを元に漫画を生成するシステムで、漫画という形式によって FOAF データを元に作られる人のネットワークをブラウジングできる所が特徴である。

スケジューリング支援システムからは、共著関係や利用者の聴講データを提供した。Comic FOAF ではこれらのデータを用いて、漫画中に共著者と会うシーンを作ったり、自分の発表を聞きに来た人を登場させたりした。

## 発表 Weblog ( Blog 情報支援 )

Blog 情報支援では、アナウンス用の Weblog の設置、大会参加者の Weblog の aggregate サービスに加え、各発表が一つのエントリとして登録された発表 Weblog の設置を行った。発表 Weblog によって、論文発表へのコメント書き込みに加え、利用者が持っている Weblog に書かれた論文発表に関するコメントを Trackback によって集約できるようになる。

スケジューリング支援システムでは、各論文のページに対応するエントリへのハイパーリンクを載せている。比較的煩雑なコメント書き込み機能を完全に発表 Weblog ( Weblog ツールとしては Movable Type を利用 ) に任せることは、システム開発・管理上メリットの大きいことであると考えられる。

## 運用結果

### 基本データ

表 6.1 は、JSAI2003 と JSAI2004 での基本データ数の違いを示している。登録者数はシステムに登録されている人の数で、著者と新規登録者が含まれる。利用者はシステムに一度でもアクセスしたことがある人を指す。なお、セッション数、発表件数および著者数は初期値のまま変化しない。

データより、全てに置いて JSAI2004 の方が多くなっていることがわかる。特に登録者数が多くなっているが、これはスタッフの増加や、他のシステムとユーザ管理を一部統合したために他システムの利用者も登録だけなされたことが理由であると思われる。

表 6.1 基本データの比較

	JSAI2003	JSAI2004
セッション数	49 件	64 件
発表件数	259 人	288 件
著者数	510 人	544 人
登録者数	558 人	639 人
利用者数	276 人	257 人

### 利用状況

システムは 4 月 30 日からサービスを開始した。システムは現在も稼働中だが、基盤システムは学会終了と共に停止しているため新規登録はできなくなっている。以後のデータは 4 月 30 日から 6 月 7 日までの利用ログを元にしたものである。

表 6.2 は、利用者数の変化と告知活動等のイベントとの関係を示している。大会開始をのぞく四つのイベントでは、メールアドレスを把握している著者および共著者に対して個別にメールによる通知を行った。

利用者数の変化が大きいのは、人間関係ネットワークシステムと PDF ファイル公開・論文推薦サービス開始の二つのイベントであった。これは両イベントがスケジューリング

表 6.2 ユーザ数とイベント

日付	イベント	ユーザ数	三日後のユーザ数
4/30	サービス開始	25	42 (+17)
5/11	人間関係ネットワークシステムサービス開始	60	104 (+44)
5/18	CoBIT 利用登録通知, Blog 情報支援サービス開始	121	127 (+8)
5/25	PDF ファイル公開, 論文推薦サービス開始	128	200 (+72)
6/1	大会開始	229	251 (+22)

支援システムと密接な関係にあることが理由として考えられる。

新規利用者数が一番増えたのが PDF ファイル公開であったということは、スケジューリング支援システムがオンラインスケジュールとして認知されていることを示す共に、それ以上のサービスを提供できていない現状を示していると考えられる。

図 6.1 は利用者数の変化をグラフ表示したものである。先述のイベントがあった際に大きくユーザ数が増えていることがわかる。その他では微増するだけであり、宣伝活動の重要性がわかる。

ユーザ数の推移

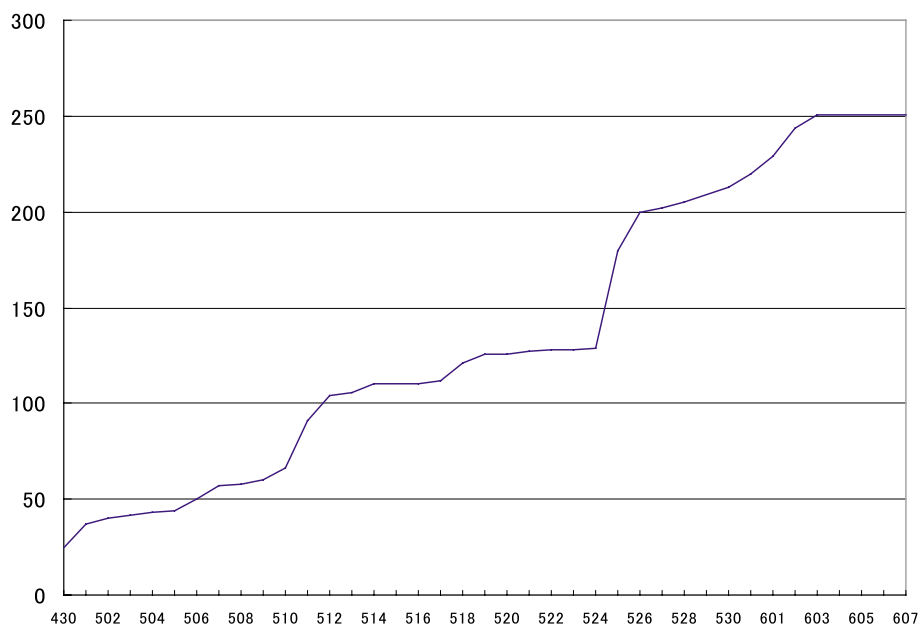


図 6.1 利用者数の推移

図 6.2 は一日あたりのページビュー数 (PV 数) とアクセス者数を示したものである。

ほぼ一人あたりのページビュー数（PV 数）は一定であることがわかる．ただ，会期中のみ一人あたりの PV 数が大きく増えていることがわかる．

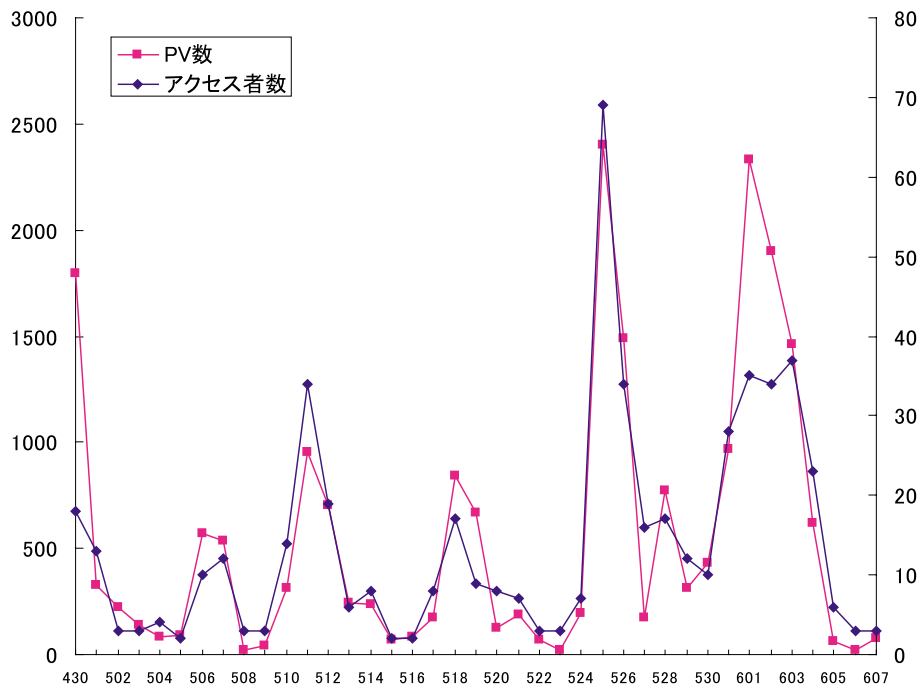


図 6.2 一日あたりの PV 数とアクセス者数

図 6.1 と図 6.2 の結果から，大会のオンラインプログラムは基本的に会期中にもっとも注目され，その他の期間ではあまり積極的に多くの（自分と直接には関係のない）コンテンツを見るものではないという性質がわかる．出会いや議論の場の提供を目標とするならば，このようなアクセス傾向を変えさせる効力を持つサービスの提供が必要であると考ええる．

## 利用者によるリンクの追加について

表 6.3 は Check リンクの追加状況の比較を示している．リンク追加者数，リンク数共に微減している．括弧内の数字は，それぞれ利用者のうち Check リンクを張った人の割合と，全論文のうち Check リンクを張られた論文の割合を示している．

表 6.4 と表 6.5 は，JSAI2003 と JSAI2004 での Know リンクの追加状況を比較したものである．表 6.5 の括弧内の数字は，利用者のうち Know リンクを張った人の割合を示している．

JSAI2003 と JSAI2004 とでは基本的な傾向は変わらないが，JSAI2004 の方が Know リンクを追加した人は減ったが追加された Know リンク数そのものは増えている．

図 6.3 と図 6.4 は，それぞれ Know リンク数および被 Know リンク数の累積分布を示し



表 6.3 追加された Check リンクの比較

	JSAI2003	JSAI2004
Check リンクの数	1840 本	1713 本
Check リンクを張った人	149 人 (0.54)	140 人 (0.54)
Check リンクを張られた論文	245 件 (0.95)	281 件 (0.96)

表 6.4 追加された Know リンクの比較

	JSAI2003	JSAI2004
全 Know リンク	840	883
双方向 Know リンク	171	167
片方向 Know リンク	498	549
無方向全 Know リンク	669	716

表 6.5 Know リンクを追加した人の比較

	JSAI2003	JSAI2004
Know リンクを張った人	99 人 (0.36)	94 人 (0.37)
Know リンクを張られた人	260 人	289 人
Know リンクを張った or 張られた人	273 人	300 人

ている。傾向としては JSAI2003 と同じで、スケールフリーに似た傾向が見られる。特に、被 Know リンク数の方にその傾向（グラフが直線的）が強く表れていることがわかる。

## アクセスパターン

次に示す 3 つの表は 4 月 30 日の 16:42:50 以降（内部的にはアクセスログの id=1922 以降）のアクセス状況を示したものである。それぞれ、各ページのページビュー数（表 6.6）、人のページへの移動経路（表 6.7）、論文のページへの移動経路（表 6.8）を示している。

## 推薦結果

推薦サービスは 64 人が利用し、システムは 430 アイテム（論文）を推薦した。去年と違って、推薦サービスに直接アクセスするリンクを通知しなかったことや、他のサービスの一つとして提供したため、利用者数をあまり増やせなかった。表 6.9 は推薦に対して利

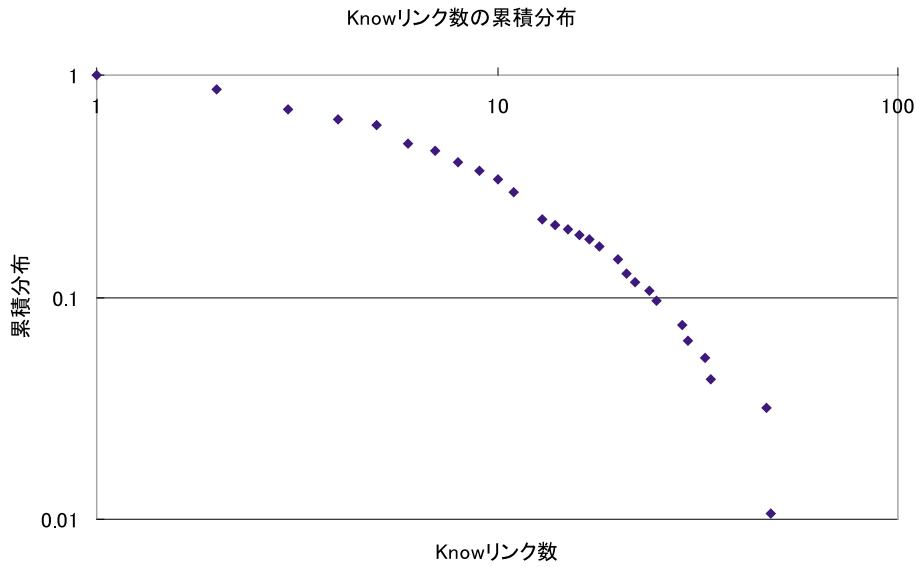


図 6.3 Know リンクの累積分布

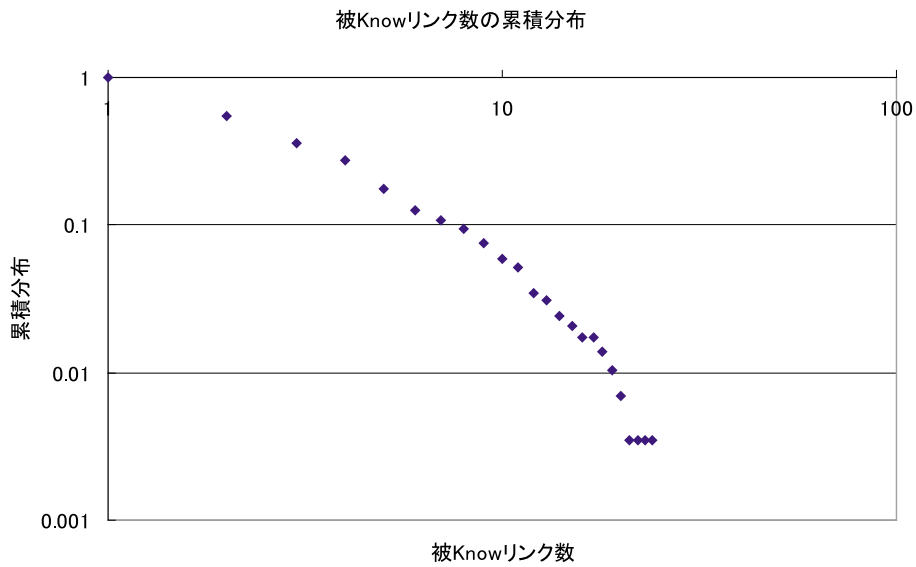


図 6.4 被 Know リンクの累積分布

表 6.6 各ページのページビュー数

ページの種類	ページビュー
セッション	3575 pv
人 (マイページ除く)	4264 pv
論文	3638 pv

表 6.7 人のページへの移動経路

移動経路	アクセス回数		
Know リンクを使って人へ移動	811 件	人	人
Check リンクを使って人へ移動	530 件	論文	人
著者リンクを使って人へ移動	769 件	人・論文	人
松尾システムから人へ移動	660 件	外部	人

表 6.8 論文のページへの移動経路

移動経路	アクセス回数		
セッションから論文へ移動	1242 件	セッション	論文
チェックリンクを使って論文へ移動	147 件	人	論文
著作リンクを使って論文へ移動	1413 件	人	論文

表 6.9 推薦結果

推薦論文に対する振る舞い	件数 (割合)
何もしない	328 件 (0.76)
アクセスのみ	31 件 (0.07)
Check リンク追加のみ	39 件 (0.09)
アクセスして Check リンク追加	32 件 (0.07)

用者がとった振る舞いを示している。

推薦結果に対して利用者が何らかのアクション（推薦した論文のページにアクセスする，Check リンクを追加する）をした場合に，推薦が成功したとみなすならば，今回の推薦サービスの成功率は 24% ということになる。