社団法人　電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
TECHNICAL REPORT OF IEICE.
OFS98-21, AI98-30 (1998-07)

インターネットにおけるコミュニティ形成支援

吉田 仙　　亀井 剛次　　服部 文夫

NTT コミュニケーション科学研究所
phone: 0742–35–1511
e-mail: yoshida,kamei,hattori@cslab.kecl.ntt.co.jp

あらまし　本稿は，ユーザ間の興味の近さにもとづいてユーザ集団を可視化することにより，インターネット上の新たなコミュニティの形成を支援するシステムについて説明する．本システムは，ユーザアイコンを，ユーザ間の関連性を反映するように平面上に並べることにより，ユーザ集団のビューを提供する．ユーザ間の関連性は，ユーザの個人的なウェブページやメールのアーカイブから抽出されたキーワードの比較によって算出される．抽出された各キーワードには重みを付け，また，望ましい重みをユーザフィードバックから学習する機構を導入する．

キーワード　コミュニティ，可視化，ソーシャルウェア，エージェント，ネットワーク，学習

Supporting the Formation of Communities on the Internet

Sen Yoshida　　Koji Kamei　　Fumio Hattori

NTT Communication Science Laboratories
phone: 0742–35–1511
e-mail: yoshida,kamei,hattori@cslab.kecl.ntt.co.jp

Abstract　We explain the development of a system that supports the formation of new communities on the Internet, by the visualization of a user group in accordance with the degree of the users' common interests. The system provides a view of the user group by placing user icons in a plane to reflect the relevance among the users. The relevance is estimated based on the comparison of the keywords extracted from the users' personal web pages or mail archive. We use a weight system for the extracted keywords as well as a learning mechanism that determines the desired weights from user feedback.

key words　community, visualization, socialware, agent, network, learning

1 はじめに

私達の日常生活において，人々が共通の興味を持って集団を形成しローカルに活動する場であるコミュニティの存在は欠かせない．ところで近年では，インターネットやモバイルコンピューティングシステムのような広域情報ネットワークが急速に発達したのにともない，私達の日頃の活動は現実社会だけでなくそれらネットワーク上のサイバー社会でも行われるようになりつつある．このサイバー社会をより魅力的で便利なものにするためには，ネットワーク上でも現実社会と同じくサイバーコミュニティが必要である．

このような理由から，サイバーコミュニティを様々な場面や角度から支援する環境が望まれている．このような環境はソーシャルウェアやコミュニティウェアなどと呼ばれ，最近こうした分野の研究に対する気運が高まってきている [2, 3]．

ソーシャルウェアと似た概念に，これまで盛んに研究開発されているグループウェアがある．グループウェアとソーシャルウェアを比較すると，グループウェアはすでに組織化された人々（企業におけるプロジェクトのメンバーなど）の協調作業を支援するものであり，そのユーザはLANを通して固定的な相手と通信する．これに対しソーシャルウェアは，インターネットなどのオープンな環境下で，より多様であいまいな集団を対象にする．ソーシャルウェアの目的は，

- 人との出会いと新たなサイバーコミュニティの形成
- サイバーコミュニティ内でのインフォーマルな対話や集団活動
- サイバーコミュニティ内の情報統合やコミュニティの組織化

といったことの支援である．

本稿では，ソーシャルウェアの例として，上の一番目の目的，すなわち新たなコミュニティの形成を支援するシステム [5] について説明する．このシステムは，ユーザ間の興味の近さの度合いに基づき潜在的なコミュニティを可視化することにより，共通の興味を持つ人との出会いと新たなサイバーコミュニティの形成を支援する．

具体的には，ある人々の集団において，それぞれの人の間の興味の近さの度合い（関連度）を求める．そして，ユーザエージェントが，人を表すアイコンを，アイコン間の距離がその人々の間の関連度を反映するように，平面上に配置する．ユーザ自身のアイコンはそのビューの中心に置かれる．各アイコンをクリックするとその人の個人情報が表示される．また，ビュー上のある範囲を指定して，その中にあるアイコンが表す人々にメッセージを同報することができる．本システムはこのようにして，興味を同じくする人々が出会い話をする場を提供し，サイバーコミュニティの形成を支援する．

本システムは，ユーザ個人のウェブページやメールのアーカイブなどからキーワードを抽出し，各ユーザの特徴ベクトルを作成してそれらを比較することにより関連度を求める．抽出されるキーワードにはユーザの性格付けに重要でないとか関係ないものも含まれるので，システムはキーワードに重みを付け，望ましい重み付けをユーザからのフィードバックにより決定する学習機構を持つ．

本稿では，学習過程を高速化するために，エージェント間の協調の概念を導入する．つま

り，各エージェントが学習結果を交換することによって，ユーザからのフィードバック入力回数を減らすことができる．評価実験では，エージェント間の協調により，ある精度に達するのに必要なフィードバックの回数が 45% 減少することが示される．

本稿では，まず第 2章でコミュニティの可視化の詳細について述べる．次に第 3章で各エージェントがキーワードの適切な重み付けを学習する手法を提示し，第 4章では学習過程を高速化するためにエージェント間の協調機構を導入する．これらの手法の評価結果は第 5章で示す．

2　潜在的コミュニティの可視化

この章では，提案するシステムがどのようにユーザの興味を判断し潜在的コミュニティを可視化するかについて説明する．

本システムは，ユーザ間の興味の近さに基づき，潜在的なコミュニティを可視化する．システムは，ユーザの興味を判断するために，ユーザのプロファイルを獲得する．これには，例えばユーザ自身に書式にプロファイルを記入してもらう方法があるが，ユーザにとっては面倒で時間がかかる．この問題を避けるため，システムはユーザが過去に書いた文書，例えば個人的なウェブページやメーリングリストのアーカイブを用いる．これらから情報検索の技術を用いてキーワードを抽出する．

システムはまずあるグループ内の各ユーザの個人的な文書を集める．具体的には，ユーザの URL からリンクをたどってそのユーザの個人的なウェブページを集めたり，メーリングリストのアーカイブを送信者別に分けるなどの作業を行う．次に全ユーザの個人的文書から，重要と思われるキーワードをいくつか選びだす．そして，それらキーワードの文書中での出現頻度を調べることにより，各ユーザの特徴ベクトルが作成される．すべての特徴ベクトル間の関連度が内積を用いて計算され，最終的にグループ内の関連度行列が出来上がる．

関連度の計算手順の詳細は以下の通りである．n をグループの人数，m をキーワードの数とする．人 i の特徴ベクトルは，

$$\vec{f_i} = \langle f_i^1, \ldots, f_i^m \rangle$$

と表され，i と j の間の関連度は i の特徴ベクトルと j の特徴ベクトルの内積

$$r_{i,j} = \vec{f_i} \cdot \vec{f_j}$$

で定義される．関連度行列 $\boldsymbol{R}$ は全てのペアの間の関連度の行列である．

$$\begin{aligned}\boldsymbol{R} &= \left| \begin{array}{ccc} r_{1,1} & \ldots & r_{1,n} \\ & \vdots & \\ r_{n,1} & \ldots & r_{n,n} \end{array} \right| \\ &= \left| \begin{array}{ccc} \vec{f_1} \cdot \vec{f_1} & \ldots & \vec{f_1} \cdot \vec{f_n} \\ & \vdots & \\ \vec{f_n} \cdot \vec{f_1} & \ldots & \vec{f_n} \cdot \vec{f_n} \end{array} \right|\end{aligned}$$

得られた関連度行列 $\boldsymbol{R}$ を可視化するために，システムはバネモデル [1] を用いてユーザアイコンを二次元平面上に配置する．すなわち，システムはユーザアイコン間にメッシュ状に仮想的なバネを張り，バネの自然長が，繋がれた二つのアイコンが表す人の間の関連度に比例するようにする．これを平面上に置くと，各アイコンはバネが均衡するように移動する．その結果，ユーザアイコンは関連するユーザどうしが近くに，関連しないユーザは遠くに置かれる．ユーザはこのビューを見ることによって，自分の周

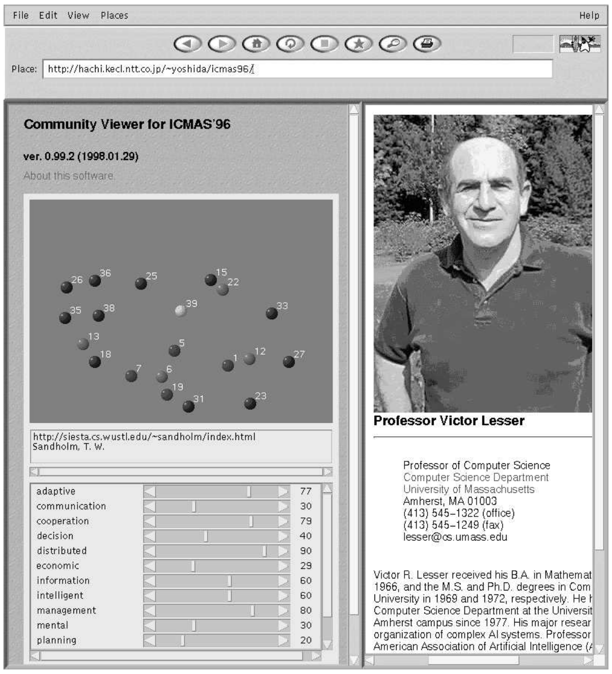

図 1: システムの画面例

りに同じような興味を持つ人々がどのように分布しているかを知ることができる．

ビューの作成は各ユーザに対応するエージェントによって行われ，そのユーザはビューの中央に配置される．また，バネの強さは，中央のユーザとそれ以外のユーザの間の自然長がより保たれるよう調節される．すなわち，中央のユーザとそれ以外のユーザの間のバネは強く，まわりのユーザ間のバネは弱くする．このため，関連度行列がすべてのエージェントで共通であっても，作成されるビューは各ユーザに特化したものとなる．

図1は，ある国際会議の参加者のためのビューの例である．この例では，会議の参加者のプロファイルをウェブから取得し，研究に関する興味の近さに応じてビューを構成している．アイコンをクリックすると，その研究者のホームページが図右側のように表示される．

本システムは興味の近さにもとづく同報の機能を持つ．すなわち，ユーザは，自分からある一定の範囲内にいる全ユーザに対してメッセージを同報することができる．この機能によって，メーリングリストのメンバーのような固定的なグループに対する通信でなく，境界が曖昧な潜在的コミュニティのメンバーに対する通信が実現される．

石田らは，ICMAS'96 という国際会議において，我々のシステムと同様にユーザ集合を可視化し出会いを支援するシステムを構築し運用した [2, 3]．このシステムは，会議の参加者が携帯端末を通して利用するもので，実際に広く使われ，ソーシャルウェアの有効性が確認された．

我々のシステムが石田らのシステムと異なる点は，

- 個人的文書からのキーワード抽出によるプロファイルの作成
- 各ユーザに特化したビューの提供
- ビューを用い興味の近さにもとづくメッセージの同報を実現

などである．

3 キーワードの重み付けのフィードバック

前章において，ユーザの個人的な文書からキーワードを抽出する手法について述べた．しかしながら，抽出されたキーワード集合はユーザを特徴づけるのに重要でなかったり関係ないようなキーワードを含みうるので，ユーザ間の関連度の計算において，すべてのキーワードを等しく用いることは適当でない．このため我々は，関連度の計算において，キーワード間に重みを付け，この重み付けの望ましい値をユーザ

からのフィードバックによって学習する機構を導入する．

ユーザからのフィードバックは次のように与えられる．ユーザははじめに，自分が中央に配置されているビューを見る．そこで自分と興味を同じくしないと思われる人が自分のすぐ近くに配置されていると，それをエージェントに伝える．するとエージェントは，示された人とユーザ自身との間の関連度が減少するよう，キーワードの重み付けを変更する．そして，新しい重み付けを用いて全ユーザ間の関連度を再計算し，変化したビューを表示する．

重み付けの詳細な手順は以下の通りである．キーワードの数を m，キーワードへの重み付けを

$$\vec{w} = \langle w^1, \ldots, w^m \rangle$$

とする．特徴ベクトル $\vec{f_i}$ は $\vec{w}$ によって

$$\begin{aligned} \vec{g}_i &= \langle g_i^1, \ldots, g_i^m \rangle \\ &= \langle w^1 f_i^1, \ldots, w^m f_i^m \rangle \end{aligned}$$

のように重み付けされる．i と j の間の重み付けされた関連度は，

$$r_{i,j} = \vec{g}_i \cdot \vec{g}_j$$

関連度行列 $\boldsymbol{R}$ は，

$$\begin{aligned} \boldsymbol{R} &= \begin{vmatrix} r_{1,1} & \ldots & r_{1,n} \\ & \vdots & \\ r_{n,1} & \ldots & r_{n,n} \end{vmatrix} \\ &= \begin{vmatrix} \vec{g}_1 \cdot \vec{g}_1 & \ldots & \vec{g}_1 \cdot \vec{g}_n \\ & \vdots & \\ \vec{g}_n \cdot \vec{g}_1 & \ldots & \vec{g}_n \cdot \vec{g}_n \end{vmatrix} \end{aligned}$$

となる．

ユーザはビューを見て，フィードバックの標的（関連のないユーザ）を示す．ここでユーザ i が自分と近すぎるユーザ j を選んだとする．このとき，各キーワード k について，k に対するそれぞれの重み付き特徴量 g_i^k および g_j^k の差が大きい場合，エージェントは重み w^k を大きくし，差が小さい場合は w^k を小さくする．すなわち，フィードバックを表す関数を F，特徴量の差の閾値を s とすると，

$$w^k \leftarrow F(w^k) = \begin{cases} w^k + \delta & \text{if } |g_i^k - g_j^k| > s \\ w^k - \delta & \text{otherwise} \end{cases}$$

これにより，i と j で特徴量に差があるようなキーワードがより重視されるようになり，結果として j が i から遠ざかる．

4　協調フィードバック

前章で述べた学習過程は，エージェント間の協調を導入することにより改善することができる．具体的には，エージェント間で重み付けの値を交換することにより，ユーザからのフィードバックの回数を減らすことが可能である．

協調の手順は以下のとおりである．エージェント h による重み付けを

$$\vec{w}_h = \langle w_h^1, \ldots, w_h^m \rangle$$

とする．h によって重み付けされた i の特徴ベクトルは

$$\begin{aligned} \vec{g}_{h,i} &= \langle g_{h,i}^1, \ldots, g_{h,i}^m \rangle \\ &= \langle w_h^1 f_i^1, \ldots, w_h^m f_i^m \rangle \end{aligned}$$

h によって重み付けされた i と j の間の関連度は

$$r_{h,i,j} = \vec{g}_{h,i} \cdot \vec{g}_{h,j}$$

であり，h によって計算された関連度行列は

$$\boldsymbol{R}_h = \begin{vmatrix} r_{h,1,1} & \ldots & r_{h,1,n} \\ & \vdots & \\ r_{h,n,1} & \ldots & r_{h,n,n} \end{vmatrix}$$

となる．

エージェント h は得られた $\boldsymbol{R}_h$ を用いて単独でフィードバック F_h を行い重み付け $F_h(\vec{w}_h)$ を計算する．フィードバックが行われるたびに，全エージェントはそれぞれの重み付け $F_1(\vec{w}_1), \ldots, F_n(\vec{w}_n)$ を交換する．そして，エージェント h は，受け取ったすべての重み付けから新たな重み付けを合成する．重み付けの合成を表す関数を X とすると，

$$w_h \leftarrow X(F_1(\vec{w}_1), \ldots, F_n(\vec{w}_n))$$

重み付け合成関数 X の実装の方法はいくつか考えられる．我々は二通りの方法でこれを実装した．一つは受け取った重み付けの平均を計算する．この場合，すべてのエージェントの重み付けは等しくなる．

$$X(w_1^k) = \ldots = X(w_n^k) = \frac{\sum_{i=1}^{n} F_i(w_i^k)}{n}$$

別の方法は，受け取った重み付けを加味するときに，関連度を考慮する．自分と近いエージェントの重み付けほど，自分にとっても有効であるという仮定のもとで，エージェント h は，i から受け取った重みと h と i の間の関連度の値を掛け合わせる．関数 X は，

$$X(w_h^k) = \sum_{i=1}^{n} r_{h,h,i} F_i(w_i^k)$$

となる．

5 評価

本章では，これまでの章で述べた手法の評価結果を示す．我々は，国際会議の参加者の支援について評価を行った．第 2 回マルチエージェントシステム国際会議（ICMAS'96）の発表者のうち 40 名の URL を用い，第 2章で述べた手法により 277 個のキーワードを選択した．

本実験では，フィードバックの性能を測る目安として，適合率 (precision) と再現率 (recall) を用いた．適合率とは得たデータのうち正しいデータ（実際に関連のあるユーザ）の割合，再現率とはすべての正しいデータのうち得ることができたデータの割合である．一般に，適合率と再現率はトレードオフの関係にある．

ここでは，ユーザは自分と興味を同じくしない人のうちもっとも近くに置かれているものを選んでフィードバックを与えると仮定する．あるユーザが別のユーザと興味を同じくするかどうかは，著者らの知識と直観に基づき真偽が決定される．本実験では 40 名からなるユーザ集合について 40 × 40 の真理値（関連する / しない）を持つ正解行列が著者らによって作成された．

図 2は，ある二つのエージェントにおける，第 3章で述べたフィードバック手法による適合率と再現率の変化である．左右の軸は再現率，上下は適合率，前後はユーザインプットの回数を表す．グラフから，フィードバックを重ねる毎に同じ再現率での適合率が向上しているのがわかる．

図 3は，図 2と同じ条件下での，単独フィードバックと協調フィードバックの比較の結果を示している．左右の軸は全ユーザからの入力の回数の合計を表し，上下の軸は再現率が 60% のときの適合率の平均値を表す．グラフ中で，`independent`の線が協調しないフィードバックの結果を，`relevance`が関連度を考慮した協調フィードバックの結果を，`average`が平均を用いる協調フィードバックの結果を表す．

55% の適合率と 60% の再現率を得るために必要なユーザインプットの回数が，エージェント間の協調を導入することによって，約 130 回から 70 回に（一ユーザあたり 3.2 回から 1.8 回

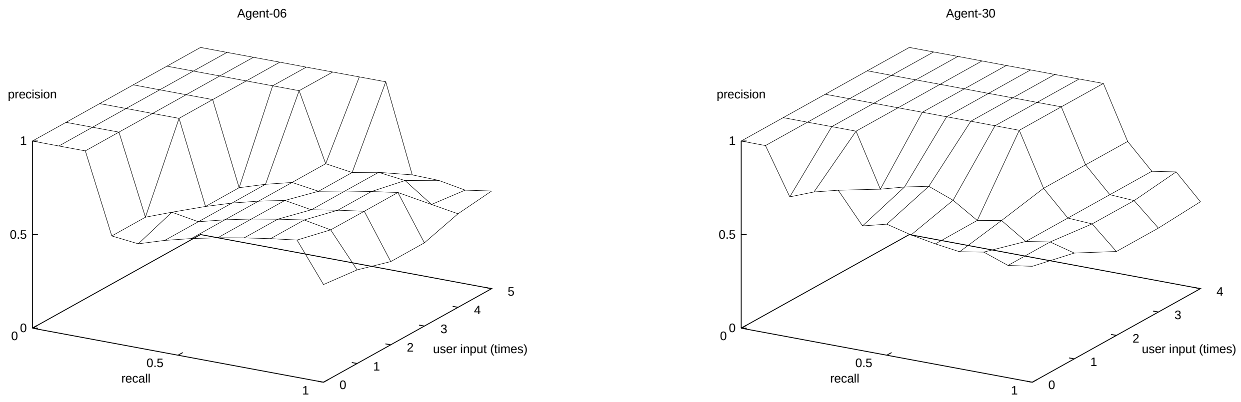

図 2: フィードバックによる適合率と再現率の変化

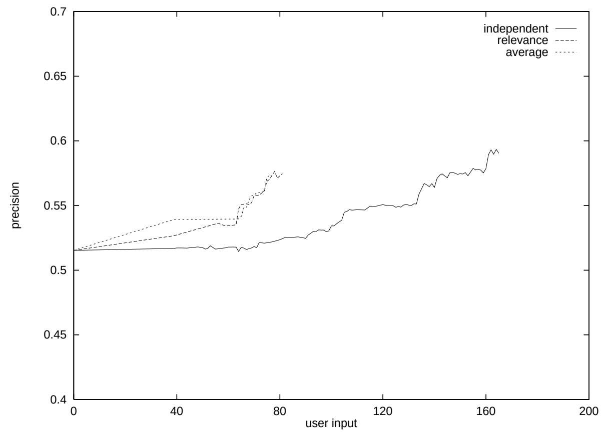

図 3: 単独 / 協調フィードバックの比較

に）約 45% 減少している．協調の方法（平均または関連度考慮）による違いはあまり影響しない．この理由の一つとして，ユーザの数が少ないためユーザの観点に偏りがなかったことが考えられる．望ましい重み付けがユーザ毎に異なるときには関連度の考慮が重要になると思われる．

6　おわりに

本稿では，ユーザどうしの興味の近さにもとづいて潜在的コミュニティを可視化することにより，新たなサイバーコミュニティの形成を支援するシステムについて述べた．本システムは，ユーザの個人的なウェブページやメールのアーカイブからキーワードを抽出して各ユーザの特徴ベクトルを作り，それらを比較することにより関連度を計算する．そしてユーザ間の距離がそのユーザ間の関連度を反映するように，各ユーザを平面上に配置する．

さらに，ユーザからのフィードバックを用いて，各エージェントがユーザのプロファイルを表すキーワード間の適切な重み付けを学習する手法を開発した．各エージェントが学習結果を交換することによって，学習過程を高速化できることが示された．ICMAS'96 の発表者のデータを用いた評価実験では，エージェント間の協調を導入することにより，55% の適合率と 60% の再現率を得るために必要なユーザからのフィードバックの回数を，一ユーザあたり約 3.2 回から 1.8 回に減らすことができた．

第 5章で示した実験では，協調を導入することにより学習のスピードは速くなるが，フィードバックを行っても適合率は比較的低いままになっている．この原因の一つとして，ウェブページが，頻繁に更新されていないなどの理由で，現在のユーザの特徴を十分に表していないことが考えられる．より正しい特徴ベクトルを求めるためには，なんらかのさらなる工夫が

必要である．これには，特徴ベクトルそのものを動的に変化させる関連フィードバック（relevance feedback）[4] の手法を用いることなどが考えられる．

その他の今後の課題としては，望ましい重み付けがユーザ毎に異なる場合の対処法について調べること，本システムを実際の場面で試験することなどが挙げられる．

最後に, 議論していただいた NTT コミュニケーション科学研究所の諸氏に感謝します．

参考文献

[1] Matthew Chalmers and Paul Chitson. Bead: Explorations in information visualization. In Nicholas Belkin, Peter Ingwersen, and Annelise Mark Pejtersen, editors, *Proceedings of the Fifteenth Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR Forum, pp. 330–337. ACM Press, 1992.

[2] Toru Ishida. Towards CommunityWare. In *Proceedings of the Second International Conference and Exhibition on The Practical Application of Intelligent Agents and Multi-Agent Technology (PAAM 97)*, pp. 7–21. The Practical Application Co. Ltd., 1997.

[3] 石田亨, 西村俊和. 広域情報ネットワークによるコミュニティ支援. 情報処理, Vol. 38, No. 1, pp. 48–53, 1997.

[4] Gerard Salton and Chris Buckley. Improving retrieval performance by relevance feedback. *Journal of the American Society for Information Science*, Vol. 41, pp. 288–297, 1990.

[5] Sen Yoshida, Koji Kamei, Makoto Yokoo, Takeshi Ohguro, Kaname Funakoshi, and Fumio Hattori. Community visualizing agent. In *Proceedings of the Third International Conference and Exhibition on The Practical Application of Intelligent Agents and Multi-Agent Technology (PAAM 98)*, pp. 643–644. The Practical Application Co. Ltd., 1998.