

時系列イベント発見のための グラフクラスタリング手法の提案

高橋昌史^{†1} 角康之^{†2} 伊藤禎宣^{†3}
間瀬健二^{†4} 小暮潔^{†5} 西田豊明^{†2}

本稿では、複数のセンサ群によって取得された人のインタラクションに関連するデータから意味のあるシーンを発見して人の行動にインデキシングを行うことを目的として、インタラクションの主体となる人や物をノードとして表現する重み付きグラフを複数の部分グラフに分割するための非階層クラスタリング手法を提案する。提案手法は、ノードとノードを結合するリンクに与えられた重みが時間にもなって変化するグラフを対象とし、時間の関数として定義される重み付き隣接行列に基づいて部分グラフを効率的かつ柔軟に抽出する。重みが動的に変化するグラフから抽出される部分グラフは、重みの変化に応じて生成や消滅、統合、分裂、拡大、縮小を繰り返す。提案手法では、重みの変化を時系列に追跡し、部分グラフに変化が生じる可能性がある場面においてのみ評価計算を行う。これにより、時間的かつ空間的に幅広い観点から評価できる。今回、ポスター展示会場という開放的な空間において人のインタラクションを観測し、人が集団で討論を行っているシーンや、複数の人が同じ展示物を見ているシーンを自動的に抽出することにより、提案手法の有用性を確認した。また、提案手法は強力なデータマイニングツールとして幅広く利用可能であり、たとえば複数人の会話の流れを構造化したり、ハイパーテキスト構造を有する Web ページの関連性を発見したりするなど、様々な事例に適用できる。

Detecting Time Series Events by Graph Clustering

MASASHI TAKAHASHI,^{†1} YASUYUKI SUMI,^{†2}
SADANORI ITO,^{†3} KENJI MASE,^{†4} KIYOSHI KOGURE^{†5}
and TOYOAKI NISHIDA^{†2}

This paper proposes a new non-hierarchical clustering method, which divides weighted graphs into some significant subgraphs. Our goal is to find significant scenes using interaction data captured by multiple sensors and to add meaningful indices to human activities. Therefore, we express people and

objects which consist of interactions as nodes. Our target is the graphs where weights associated with links are changeable while, on the contrary, the targets of conventional clustering methods are those where weights are fixed. In short, proposed method can divide those graphs whose weights can be defined as time functions into some subgraphs according to their weighted adjacency matrices. Subgraphs are often created, extinguished, divided, expanded, and diminished according to the change of weights. Proposed method watches the change of subgraphs and judge whether they change only when they can, which results in flexible evaluations from spacio-temporal viewpoint. We have confirmed the usefulness of proposed method by the experiments to find such significant scenes automatically as, for example, those where people discussed together and those where people watched something together, in poster exhibitions. Moreover we believe that proposed method is widely available as a useful tool for data mining and can be applied to many problems.

1. はじめに

筆者らのグループでは、新しいインタフェースのパラダイム創出と人のインタラクションの分析を目的として、インタラクションのコーパスを構築する試みを進めてきた^{1),2)}。すなわち、我々は人が装着するウェアラブルなビデオカメラや各種センサに加えて環境に偏在するセンサ群を利用することで人の行動を多角的に観測し、映像、音声、視線情報、生体情報など、人のインタラクションを構成している様々なモダリティを蓄積している。これらのデータは社会心理学者が人間の行動分析に活用できるほか、体験者自身がビデオデータを鑑賞して自分の体験を振り返り新たな発見を得ることができる。しかし、これらのデータは非常に大量になる傾向があるため、ユーザが目的の場面を探し当てるのが難しい。そこで、我々は各種センサから取得されたデータを利用して人と人、人と物、人と環境の間で発生するインタラクションを自動的に解釈し、大量のビデオデータに対してインデキシングを行う

^{†1} 株式会社日立製作所中央研究所
Central Research Laboratory, Hitachi Ltd.

^{†2} 京都大学大学院情報学研究所
Graduate School of Informatics, Kyoto University

^{†3} 東京農工大学
Tokyo University of Agriculture and Technology

^{†4} 名古屋大学情報連携基盤センター
Information Technology Center, Nagoya University

^{†5} ATR 知識科学研究所
ATR Knowledge Science Laboratories

システムを開発した³⁾。本稿では、人のインタラクションを時間的および空間的に意味のある単位で分割した「シーン」を発見してインデキシングに利用するためのクラスタリング手法を提案する。

これまでも、様々な状況下で人の行動に対してインデキシングを行う研究が行われてきた。たとえば、会議場内で撮影したビデオをシーンごとに自動分割する文献⁴⁾、⁵⁾などでは、会議の場面で有意とされるイベントを抽出することによって蓄積された映像の可用性を高めている。しかしこれらの研究では、固定カメラでとらえられた画像の変化量や発話者の音源位置などからシーンを切り替えるといった、短絡的な規則によりシーン分割を行っているため、会議場における限られた行動にしか適用できなかった。我々は、「見る」、「話す」といった人間の基本的なインタラクションに対するクラスタリング手段を提供することで、適用場面に依存することなく適正かつ柔軟にシーンの切り出しを行うことができるようになった。

また、観測データの時系列推移から人の行動を判定する研究も数多く行われてきた。たとえば、人の滞在場所の推移から確率的な手法によって人の行動パターンを解釈する文献⁶⁾では、観測された行動パターンがその人物にとって日常的か否かを判断することを可能とし、独居高齢者の見守りシステムなどへの応用が検討されている。しかしこの場合、対象人物の日常的な行動パターンをあらかじめシステムに学習させておく必要がある。それに対し我々は、時系列データに対するデータマイニングツールを提供することにより、特定の生活パターンに特化したパターン検出ではなく、複数人の共在、共同注視、会話といった、より一般的な行動検出を可能にすることに注力した。そのため、我々のツールはビデオのインデキシングに用いるだけでなく、人のインタラクションのモデル化を試みる社会心理学者が人の行動を分析するために利用することもできる。

通常、インタラクションは主体となる人（あるいは物、環境）から相手（人、物、環境）に向けて行われるため、方向性を有する。さらに、インタラクションの種類や継続時間などによって重要度を付けることもできる。よって、その構造はインタラクションの主体となる人や物、環境などをノードとする重み付き有向グラフによって表すことができるため、前述した「シーン」の探索問題は、有向グラフの中から特徴的な部分グラフを抽出するクラスタリング問題に帰着できる。

クラスタリングは、蓄積情報から意味的なつながりを持つデータの集合を発見するための重要なデータマイニングツールとして幅広く利用されている⁷⁾、⁸⁾。しかし、グラフを対象とした従来のクラスタリングアルゴリズム⁹⁾–¹¹⁾はどれも時間の概念を扱うことができない。

すなわち、従来手法は部分グラフが時間にもなって変化することのない、いわば普遍的に存在するものとして、ノードとリンクの空間的な接続関係にのみ基づいてクラスタリングを行っていた。しかし、当然人と人、人と物、人と環境の関係性は時間に依存して変化し、その関係性の変化を発見することは、人の行動に対してインデキシングを行うための大きな手がかりになる。そこで、我々は各リンクに時間関数として重みが与えられた有向グラフから部分グラフを効率的に抽出するためのクラスタリングアルゴリズムを開発した。本稿では、時間の概念を扱わない従来手法と区別するために、提案手法を「時空間クラスタリング」と呼ぶ。提案手法は、時間にもなって変化する重みに基づいてクラスタリングを行うことで、部分グラフの生成や消滅、統合、分裂、拡大、縮小を検出できる。

以下、本稿では人のインタラクションの記録から重要シーンの抽出までを順に説明する。2章では、まずインタラクションのコーパスを構築するための我々の試みについて述べ、続いて関連性の高いインタラクションを統合して意味のあるシーンを抽出するために解決すべき問題について説明する。3章では、これらの問題を解決するために我々が開発したグラフの時空間クラスタリングアルゴリズムについて述べる。続いて4章において提案手法の有効性を確認するために実施した実験結果について示し、5章において提案手法の特徴と発展性について考察を行う。そして最後に6章で本稿のまとめを行う。

2. インタラクションの記録と解釈

我々は不特定多数の人や物との自由なインタラクションが発生するポスター展示会場という開放的な空間を選び、人のインタラクションに関わるマルチモーダルなデータを収集してインタラクションのコーパスを構築している¹⁾、²⁾。

インタラクションを記録するためのセンサ部は運用形態に応じて柔軟な構成の変更が可能であり、図1のように、ウェアラブルなヘッドセットタイプのもので、設置型の据え置きタイプのものから構成される。ここでは、多視点から人のインタラクションを撮影するためのカメラとマイクに加え、映像内にとらえられている対象物の認識・位置測定を行うために、赤外線 ID タグシステム¹²⁾を利用している。赤外線 LED の点滅パターンによって固有の ID を発信できる赤外線 ID タグを認識対象に取り付け、それを認識する赤外線 ID センサをカメラの光軸と一致させて固定する。こういったデバイスをユーザの顔の向きに一致させて装着することにより、視野内のどこに何が存在するかを実時間で記録することができる。また、同じデバイスを環境側に設置することによって、そのエリアに滞在する人や物を検出することができる。さらに、喉に取り付けて声帯の振動から発話のボリュームを測定するこ

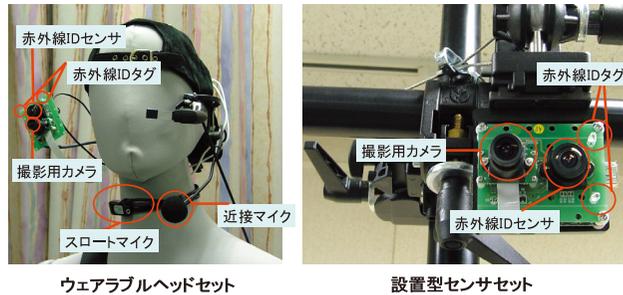


図1 インタクションを記録するためのセンサ群
Fig. 1 Sensor sets for recording human interactions.

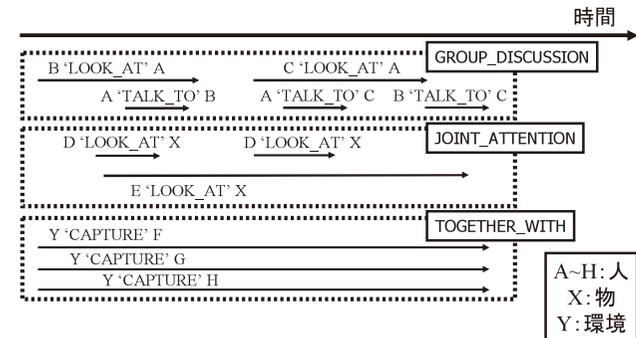
表1 自動的に解釈されるインタクション
Table 1 Interactions interpreted automatically by our system.

インデックス	意味	主体	対象
LOOK_AT	～を見る	人	人/物
TALK_TO	～に話しかける	人	人
CAPTURE	～を捕らえる	環境	人/物

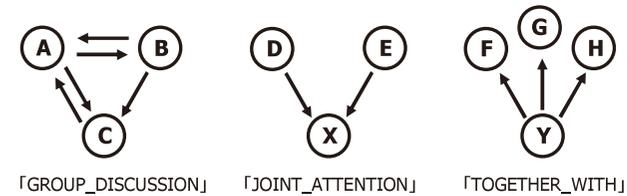
とができるスロートマイクを利用した。これにより、閾値処理を施すことで装着者が発話しているか否かの判定を行うことが可能である。

さらに、我々は、取得されたデータをもとに人のインタクションを解釈し、表1に示すような機械可読なインデックスを自動的に付与するシステムを試作した³⁾。これにより、動作主体(だれ/何が)、動作対象(だれ/何に対して)、動作開始時間(いつから)、動作終了時間(いつまで)、動作の種類(どのようなインタクションを行ったのか)などをキーワードとする検索が可能になり、映像データなどの利便性が向上した。

しかし、これらのインタクションは非常に断片的であり発生頻度も高いため、ユーザが目的の場面を探し出すためのインデックス情報として必ずしも適切であるとはいえない。そのため、我々はより抽象度の高いインデックスを生成するために、関連性の高いインタクションを統合して意味のあるまとまりを表す「シーン」を抽出する手法を開発した。すなわち、図2(a)のように、時間的かつ空間的近傍で発生するインタクションを統合することにより、人にとってより意味深い情報の抽出を試みた。これは、インタクションの主体と



(a) 時間的かつ空間的近傍で発生するインタクションの統合



(b) 部分グラフとして抽出される「シーン」

図2 「シーン」の発見とインデキシング

Fig. 2 Process of finding “scenes” to make indices.

なる人や物をノードとする有向グラフとしてインタクションの構造を表現した場合、その中からたとえば図2(b)に示す特徴的な部分グラフを抽出することに相当する。なお、図中のリンクは、インタクションの動作主から動作対象に向かってのびているものとする。

このように、人にとって意味のあるインタクションのまとまりを「シーン」として抽出することによって、たとえばポスター展示会場では、複数の人が特定の場所で討論を行う「GROUP_DISCUSSION」や、複数の人が一緒に同じ展示物を閲覧する「JOINT_ATTENTION」といったインデックスをビデオに対して付与することが可能になる。

しかし、これまでに何度もこのようなシーンを抽出するシステムの試作と評価を繰り返してきたが、大別して2種類の問題が原因となりこういったシーンを発見することが困難であることが分かった。まず1つ目は、蓄積されたデータの中に雑音が多く含まれるという問題である。これは、たまたま視界に入ってきた対象物を人が装着するセンサが認識してしまっ

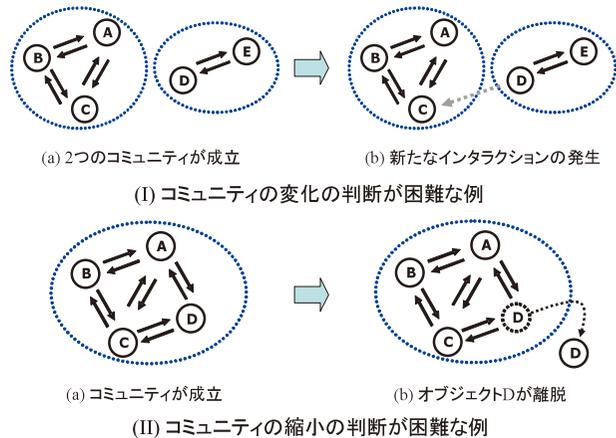


図3 「シーン」変化の把握が困難な例
 Fig. 3 Example case where it is difficult to find "scene".

たり、一時的に通りがかった場所に設置してあるセンサが人を認識してしまうことにより、シーンを適切に推定するための判断が妨げられてしまうことである。この問題を改善するためには、何らかの形で人のインタラクションに重み付けを行い、重要度の低いものを排除する必要がある。2つ目の問題は、シーンの変化を瞬時に判断することが困難であることである。この問題に関しては、これまでの評価を通じて、次の2種類の場面においてシーンの変化を把握することが困難になることが分かっている。ここでは、インタラクションの主体となる人や物、および環境を「オブジェクト」、シーンを構成するオブジェクトの集合を「コミュニティ」と呼ぶことにする。

(1) 異なるコミュニティに属するオブジェクト間で新たにインタラクションが発生した場面 (図3(I))

たとえば、図3(I)(a)のように、オブジェクトA, B, CとD, Eがそれぞれ別々のコミュニティを構成している場面について考える。このとき、図3(I)(b)のように、それぞれ異なるコミュニティに属するオブジェクトCとDの間に新たなインタラクションが発生した場合、コミュニティの構成に変化が生じる可能性がある。しかしこの場合、CがA, Bとともに構成していたコミュニティから抜けてD, Eが構成するコミュニティに加わったのか、DがEとともに構成していたコミュニティから抜けてA, B, Cが構成するコミュニティに加わったのか、CとDの両方がそれぞれA, BおよびEと構成していたコミュニ

ティから抜けてCとDで構成される新たなコミュニティを構成したのか、2つのコミュニティが統合されて1つのコミュニティを構成したのか、それともただの雑音データにすぎないのか、といったように、実際には様々な解釈が可能であり、これらの解釈の中でどれが最も妥当であるのかを瞬時に判断することが難しい。

(2) コミュニティを構成するオブジェクトが離脱した場面 (図3(II))

たとえば、図3(II)(a)のように、オブジェクトA, B, C, Dが1つのコミュニティを構成している場面について考える。このとき、図3(II)(b)のように、このコミュニティに属するDが、たとえばどこか別の場所に行ってしまうことによりこのコミュニティから離脱した場合、この変化を瞬時に判断することが困難となる。たとえばこの後でDがどのオブジェクトともインタラクションを行わなければ、Dがこのコミュニティを離脱したことは容易に判断できるが、いつ離脱したのかを判断することが難しい。また、この後でDが同じコミュニティに属するA, B, C以外のオブジェクトとインタラクションを行った場合は、コミュニティの構成に変化が生じたことは容易に推定できるが、(1)の問題と同じくDがこのコミュニティを抜けたのかそれともインタラクションの相手がこのコミュニティに加わったのかを瞬時に判断することが難しい。

上記2種類の場面においても柔軟にシーンの変化を判断するためには、インタラクションが発生した時点で即時にその意味を判断するのではなく、時間的かつ空間的に幅広い観点からシーンを評価する必要がある。

3. グラフの時空間クラスタリング

本章では、本稿で提案する重み付きグラフに対する時空間クラスタリング手法について説明する。本手法を利用することにより、2章で述べたような場面においても柔軟にクラスタリングを行い、シーンの変化を追跡することができる。

本手法は、オブジェクトどうしの関係を表す有向グラフから意味のある部分グラフを柔軟に抽出することにより、シーンを構成するコミュニティを推定する。ここで、ノード間のリンクにはオブジェクト間で発生するインタラクションの強さを表す重みが付与されており、その重みは時間に依存して変化する。すなわち、リンクの重みは時間関数となる。重み付きグラフから部分グラフを抽出するためのクラスタリング手法は従来から様々なものが提案されてきたが、このように重みが動的に変化する場合を対象とする手法はほかにない。

提案手法では、ノード間にリンクを「確立」することによってコミュニティの規模を拡大

し、そのリンクを「切断」することによってコミュニティの規模を縮小する。すなわち、コミュニティは、2つのノード間に強いリンクが確立されることによって新たに生成され、その後そのコミュニティを構成するノードとそうでないノードの間にリンクが確立されるたびにコミュニティは拡大する。一方、同じコミュニティに属するノードどうしの間に張られるリンクが弱くなって切れるとそのコミュニティは縮小し、最後の1本が切断されるとコミュニティは消滅する。このように、リンクの重みの変化にともなってコミュニティは生成や消滅、統合、分裂、拡大、縮小を繰り返す。提案手法では、リンクの確立や切断を行うタイミングを制御し、時系列にコミュニティの変化を監視する。コミュニティの構成が変化する可能性がある時点でのみリンクの確立や切断の評価を行うことにより、計算量を削減して効率的にコミュニティの変化を推定することができる。

3.1 評価指標の導入

コミュニティの変化を評価するために、「リンクの強度」、「コミュニティの強度」、「コミュニティへの貢献度」という3種類の指標を導入する。提案手法では、これらの評価指標に基づいてリンクの確立や切断を行うための「コスト」を計算する。コストとはある処理を行うための代償を表し、この値が低い処理が優先的に実行される。以下、各評価指標について説明する。

3.1.1 リンクの強度

オブジェクト間で発生するインタラクションの強さを表す指標として、各リンクに対して重みを与える。この値が大きい場合、両者が同じコミュニティに属する可能性が高い。そのため、このリンクを切断して両者を別々のコミュニティに分割する処理を行うためのコストが高くなる。そこで、時刻 t におけるノード x からノード y へ向かうリンクの強度 $L_{xy}(t)$ を、それらのオブジェクト間において時刻 t から一定時間の間に発生するインタラクションの総時間として定義し、ウィンドウ幅 w_1 を用いて次式のように計算する。

$$L_{xy}(t) = \frac{1}{w_1} I_{xy}(t, t + w_1)$$

ここでは、時刻 t_1 から時刻 t_2 の区間において x が y に対して行うインタラクションの総時間を $I_{xy}(t_1, t_2)$ と表している。

3.1.2 コミュニティの強度

コミュニティの結合力を表す指標として、その強度を定義する。これは、コミュニティが生成されてからオブジェクト間で発生したインタラクションの総時間として表される。この値が大きければ、コミュニティの推定結果が妥当である可能性が高いため、そのコミュニ

ティを変化させる処理を行うためのコストが高くなる。そこで、時刻 t におけるコミュニティ C の強度 $F_C(t)$ を次式のように計算する

$$F_C(t) = \frac{1}{|C|^2} \sum_i \sum_j T_C(t)[i, j]$$

ただし、 $|C|$ をコミュニティ C 内に存在するオブジェクトの数、 $T_C(t)$ を部分グラフ C の重み付き隣接行列とし、その (i, j) 成分の値を $T_C(t)[i, j]$ として表す。 $T_C(t)$ はウィンドウ幅 w_2 を用いて次のように定義される。

$$T_C(t)[i, j] = \begin{cases} \frac{1}{w_2} I_{ij}(t - w_2, t) & (t - t_s > w_2 \text{ の場合}) \\ \frac{1}{t - t_s} I_{ij}(t_s, t) & (t - t_s \leq w_2 \text{ の場合}) \end{cases}$$

なお、 t_s はコミュニティ C が生成された時刻であるとする。

3.1.3 コミュニティへの貢献度

コミュニティを構成するオブジェクトに対して、それぞれそのコミュニティに寄与している度合いを定式化するために、その貢献度を定義する。これは、対象となるオブジェクトが同じコミュニティに属する別のオブジェクトと行うインタラクションの総時間として表される。この値がコミュニティの強度に対して十分大きい場合、対象オブジェクトはそのコミュニティに属するという解釈が妥当である可能性が高いため、そのオブジェクトをコミュニティから排除する処理を行うためのコストが高くなる。そこで、コミュニティ C に属するノード x に対して、時刻 t におけるコミュニティへの貢献度 $P_C(x, t)$ を次式のように定義する。

$$P_C(x, t) = \frac{1}{|C|^2} \sum_i \sum_j A_C(x, t)[i, j]$$

ただし、 $A_C(x, t)$ は、部分グラフ C の重み付き隣接行列に対して、ノード x に関わる要素以外を0に設定したものであるとする。すなわち、 $A_C(x, t)$ はウィンドウ幅 w_3 を用いて次のように定義される。

$$A_C(x, t)[i, j] = \begin{cases} 0 & (i \text{ および } j \text{ が } x \text{ と一致しない場合}) \\ \frac{1}{w_3} I_{ij}(t - w_3, t) & (t - t_s > w_3 \text{ の場合}) \\ \frac{1}{t - t_s} I_{ij}(t_s, t) & (t - t_s \leq w_3 \text{ の場合}) \end{cases} \quad \left. \begin{array}{l} \\ \\ \end{array} \right\} \begin{array}{l} \\ (i \text{ または } j \text{ が } x \text{ と} \\ \text{一致する場合}) \end{array}$$

3.2 コミュニティの変化の検出

前節では、一定時間内に発生するインタラクションの総時間に基づいて各種評価指標を計算することを説明した。本節では、これらの指標に基づいてコミュニティの変化を抽出する方法について説明する。

インタラクションのまとまりについて適切に把握するには、インタラクションが始まる直前にコミュニティを生成・拡大し、インタラクションが終了した直後にコミュニティを縮小・消滅させたい。提案手法では、コミュニティの生成・拡大はリンクの強度によって判断されるため、そのタイミングを決定するにはある時点以降の数値を調べる必要がある。一方、コミュニティの縮小・消滅については、コミュニティの強度とコミュニティへの貢献度によって判断されるため、そのタイミングを決定するにはある時点までの数値変化を調べる必要がある。そのため、リンクの強度については未来の時間区間 $[t, t + w_1]$ に対して、コミュニティの強度とコミュニティへの貢献度については過去の時間区間 $[t - w_2, t]$, $[t - w_3, t]$ に対して計算を行っている。

提案手法では、コミュニティの構成に変化が生じる可能性がある5種類の場面(図4)においてのみ評価指標の計算を行う。こうすることにより、余分な計算を行うことなく効率的にコミュニティを推定することができる。以下、各場面におけるコミュニティの推定手順について示す。

3.2.1 新たなコミュニティが生成する場面

図4(a)のように、どのコミュニティにも属さない単独のオブジェクト間で新たなインタラクションが起ると、新しいコミュニティが生成する可能性がある。このとき、リンクの強度が十分大きい場合はノード間にリンクを確立してそれら2つのオブジェクトによって構成される新たなコミュニティを生成する。一方、このリンクの強度が十分大きくない場合は、そのインタラクションが雑音データにすぎないと判断し、新たなコミュニティは生成しない。

この場合、たとえば定数 $Th1$ を用いて次式が成立すると、リンクの強度が十分大きいと判断できる。

$$L_{xy}(t) + L_{yx}(t) > Th1$$

3.2.2 コミュニティが拡大する場面

図4(b)のように、あるコミュニティに属するオブジェクトが、どのコミュニティにも属さない単独のオブジェクトとの間でインタラクションを行うと、そのコミュニティは拡大する可能性がある。このとき、このリンクの強度がコミュニティの強度に対して十分大きければ、そのリンクはコミュニティを変化させるのに十分な力を備えていると判断してオブジェクト間にリンクを確立し、相手のオブジェクトをコミュニティに追加する。一方、リンクの強度がコミュニティの強度に対して十分大きくない場合は、そのインタラクションは重要で

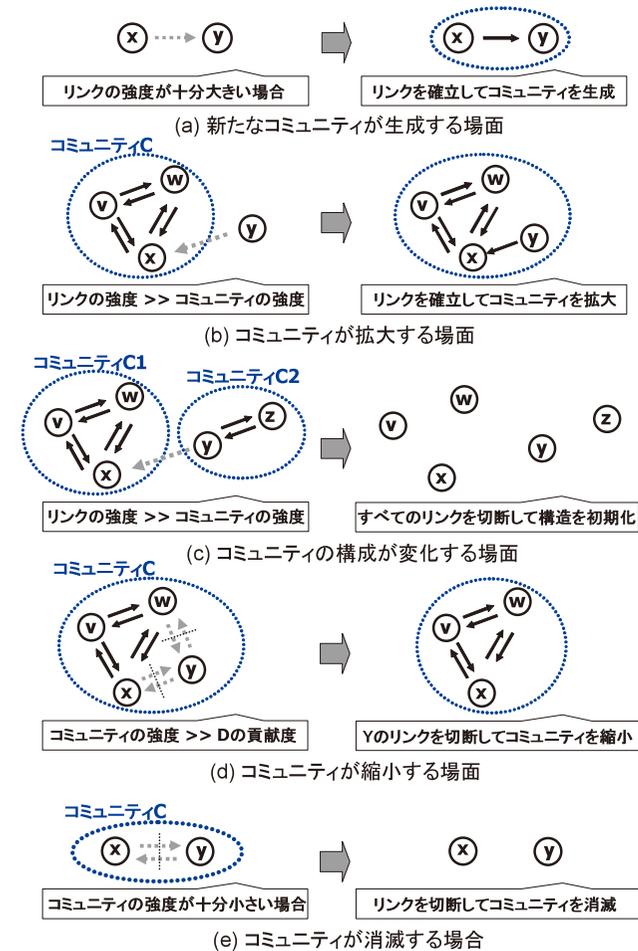


図4 コミュニティの変化の評価

Fig. 4 Evaluations for changing communities.

ば、そのリンクはコミュニティを変化させるのに十分な力を備えていると判断してオブジェクト間にリンクを確立し、相手のオブジェクトをコミュニティに追加する。一方、リンクの強度がコミュニティの強度に対して十分大きくない場合は、そのインタラクションは重要で

ないと判断してコミュニティは拡大しない。

この場合、たとえば定数 $Th2$ を用いて次式が成立すると、リンクの強度が十分大きいと判断できる。

$$L_{xy}(t) + L_{yx}(t) > Th2 \times F_C(t)$$

3.2.3 コミュニティの構成が変化する場面

図 4(c) のように、別々のコミュニティに属する 2 つのオブジェクト間で新たにインタラクションが行われると、2 章で説明したように様々な解釈が可能となる。このとき、新たなリンクの強度がこれら 2 つのコミュニティの強度に対して十分大きい場合、コミュニティの構成に変化が生じたと判断して両コミュニティ内に存在するすべてのリンクを切断し、これらのコミュニティを消滅させる。そしてその時点からコミュニティの解釈を再開する。一方、新たなリンクの強度が 2 つのコミュニティの強度に対して十分大きくない場合は、コミュニティの構成に変化はないものとしてコミュニティは消滅させない。

この場合、たとえば定数 $Th3$ を用いて次式が成立すると、リンクの強度が十分大きいと判断できる。

$$L_{xy}(t) + L_{yx}(t) > Th3 \times (F_{C1}(t) + F_{C2}(t))$$

3.2.4 コミュニティが縮小する場面

図 4(d) のように、3 つ以上のオブジェクトから構成されるコミュニティにおいて、そのコミュニティを構成するオブジェクトの 1 つがしばらくどのオブジェクトともインタラクションを行わない場合、そのコミュニティが縮小する可能性がある。このとき、コミュニティ全体の強度に比べてあるオブジェクトのコミュニティ貢献度が十分小さくなれば、そのオブジェクトをコミュニティから切り離してコミュニティを縮小する。

この場合、たとえば定数 $Th4$ を用いて次式が成立すると、コミュニティ貢献度が十分小さいと判断できる。

$$P_C(y, t) < Th4 \times F_C(t)$$

3.2.5 コミュニティが消滅する場面

図 4(e) のように、2 つのオブジェクトから構成されるコミュニティに対して、それらのオブジェクト間でインタラクションがしばらく検出されなくなると、そのコミュニティは消滅する可能性がある。この場合、コミュニティの強度が十分小さくなればノード間のリンクを切断してコミュニティを消滅させる。

この場合、たとえば定数 $Th5$ を用いて次式が成立すると、コミュニティ貢献度が十分小さいと判断できる。

$$F_C(t) < Th5$$

3.3 コミュニティ推定アルゴリズム

以上で述べた手順によりコミュニティの変化を追跡するアルゴリズムについて説明する。本アルゴリズムは、表 1 のように解釈されたインタラクションの情報（動作主体、動作対象、動作開始時間、動作終了時間、動作の種類）を入力して人のインタラクションに対して時空間クラスタリングを行った結果、検出されたコミュニティに関する情報（生成時刻、消滅時刻、構成オブジェクトのリスト）を出力する。以下、時刻 t_{start} から t_{end} までの間に検出されたインタラクションに基づいてコミュニティを時系列に検出するアルゴリズムについて示す。ここでは、表 1 に基づいて解釈されたインタラクションの集合を I と表している。

【コミュニティを検出するアルゴリズム】

- 1.【前処理】抽出されたコミュニティの情報を一時的に格納するためのリスト $CList$ を用意し、これを空の状態に設定して初期化を行う。
- 2.【前処理】時刻 t_{start} から t_{end} までの間に発生したインタラクションを I から抽出し時刻 t_{start} で整列させる。
- 3.【ループ 1】 I からインタラクションを 1 つ抽出し、その動作主と動作対象を特定する。 $CList$ を調べ、両者がどのコミュニティにも属していなければステップ 4 へ。片方が何らかのコミュニティに属していればステップ 5 へ。両者がそれぞれ別々のコミュニティに属していればステップ 6 へ。両者がともに同じコミュニティに属していればステップ 7 へ。 I が空であれば終了。
- 4.【3.2.1 項の評価】リンクの強度が十分大きければ、そのインタラクションの動作主と動作対象の 2 者からなるコミュニティを新たに生成し、そのコミュニティの情報を $CList$ に追加する。続いてステップ 7 へ。
- 5.【3.2.2 項の評価】リンクの強度がそのコミュニティの強度に比べて十分大きければ、動作主と動作対象を同じコミュニティに含めて拡大し、 $CList$ を更新する。続いてステップ 7 へ。
- 6.【3.2.3 項の評価】リンクの強度が、動作主と動作対象が属するコミュニティの強度に対して十分大きければ、 $CList$ からそれら 2 つのコミュニティに関する情報を削除してそれらのコミュニティを消滅させる。続いてステップ 7 へ。
- 7.【ループ 2】 $CList$ に含まれるコミュニティを 1 つ抽出し、そのコミュニティを構成するオブジェクトの数を調べる。3 以上であればステップ 8 へ。2 であればステップ 9 へ。 $CList$ に含まれるすべてのコミュニティに対して処理が終了すればステップ 3 へ。
- 8.【3.2.4 項の評価】コミュニティを構成するすべてのオブジェクトに対してコミュニティ

貢献度を調べる．これがコミュニティの強度に比べて十分小さいオブジェクトをそのコミュニティから排除して *CList* を更新する．続いてステップ 7 へ．

9. 【3.2.5 項の評価】コミュニティの強度が十分小さければそのコミュニティを消滅させて *CList* から排除する．続いてステップ 7 へ．

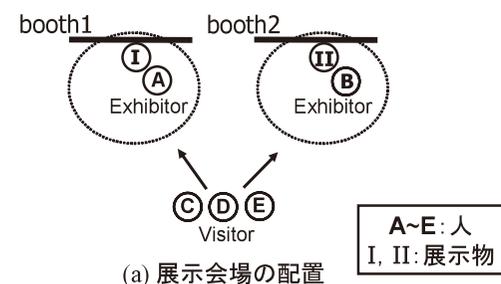
4. 実験

本稿で提案したアルゴリズムの動作を確認してその有効性を検証するために，ポスター展示会における展示者と見学者のインタラクションを記録し，抽出されるコミュニティの変化を観測した．

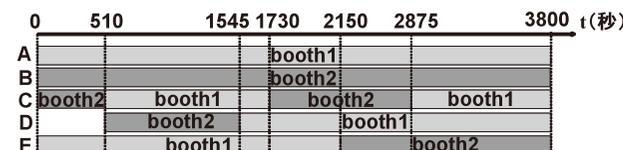
ポスター展示会場には図 5 (a) のように 2 つのブースを設けてそれぞれ 1 セットずつの環境型センサを取り付け，それぞれ 1 人の展示者を配置した．各ブース (booth1 と booth2) にはそれぞれ展示者 A と展示者 B が配置され，見学者に対して展示物 I, II について説明を行った．さらに，3 人の見学者 C, D, E は自由に好きな方のブースを訪れて展示者と討論を行った．

5 人がどちらのブースに滞在しているのかを目視で判断した結果を図 5 (b) に示す．ここでは水平方向に時間軸 (秒) を設定し，各人がどの時間帯にどちらのブースを訪れているかについて示している．

本実験に先駆けて，3.1 節で説明した評価指針を計算する際に利用するパラメータ w_1, w_2, w_3 をすべて等しい値 w に設定し，同様の環境下で w を変化させてコミュニティの検出を行った．このとき，検出されたコミュニティの数とその平均長の変化を図 6 に示す． w が小さい場合，一時的な状況の変化に対して過剰に反応してしまい，コミュニティが細分化されて安定した結果が得られていない．そのため，コミュニティの数は多く，その平均長は短くなっている．一方， w が大きい場合は，コミュニティの評価対象となる時間区間が長く，余分なインタラクションを評価対象としてしまうため安定した結果が得られていない．以上の考察により，我々は，検出されるコミュニティ数が少なく，その平均長が長い場合に安定した結果が得られると判断した．本実験ではまず，以上の判断に基づいて $w = 480$ ， $w = 600$ ， $w = 720$ という 3 種類の値を w の候補としてあげた．続いて，これらの値の中で，目視で判断した場合と比べて最も妥当な結果が得られると判断した $w = 480$ を最終的に採用した．また，3.2 節で言及した定数については， $Th1 = 0.1$ ， $Th2 = 0.5$ ， $Th3 = 0.35$ ， $Th4 = 0.15$ ， $Th5 = 0.015$ に設定した．これらはアプリケーションの種類によって最適値が異なると考えられるが，本実験ではビデオのインデキシングに適した値を経験的に定めた．



(a) 展示会場の配置



(b) 各人のブース滞在状況

図 5 実験環境

Fig. 5 Environments for our experiment.

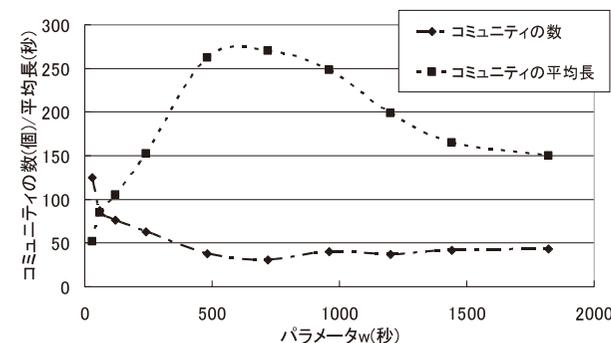


図 6 検出されたコミュニティの数と平均長

Fig. 6 Number and mean duration of communities.

提案手法を利用してコミュニティの抽出を行った結果を図 7 に示す．図では，垂直方向に時間軸 (秒) を設定し，抽出されたコミュニティを，それを構成するオブジェクト名とともに記載している．図 5 (b) と比較しても，コミュニティの変化を柔軟に追跡できていることが分かる．ここでは，短期間の間にコミュニティの推定結果が頻繁に変化して安定した結

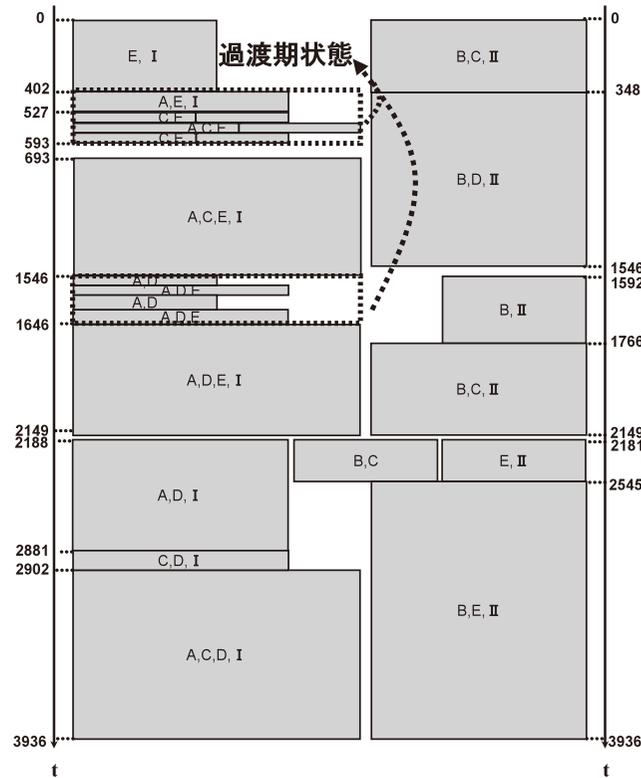


図 7 コミュニティの抽出結果
Fig. 7 Communities found by our system.

果が得られない過渡期状態が図 7 の点線部に示す 2 カ所で観測されているが、隣接する短時間のコミュニティを連結するという簡単なヒューリスティクスを用いることにより、これらのコミュニティを統合することができた。

続いて、特徴的な時間区間を抜粋し、目視でビデオを観測した結果とシステムがコミュニティを検出した結果を比較する。t = 510 (秒) から t = 1,766 (秒) におけるコミュニティの検出過程を図 8 に示す。t = 510 (秒) から 1,545 (秒) までは展示者 A が booth1 において見学者 C と見学者 E に対して、展示者 B は booth2 において見学者 D に対して説明を行っている。ここではまず、いくつかの過渡期状態を経た後、t = 693 (秒) から

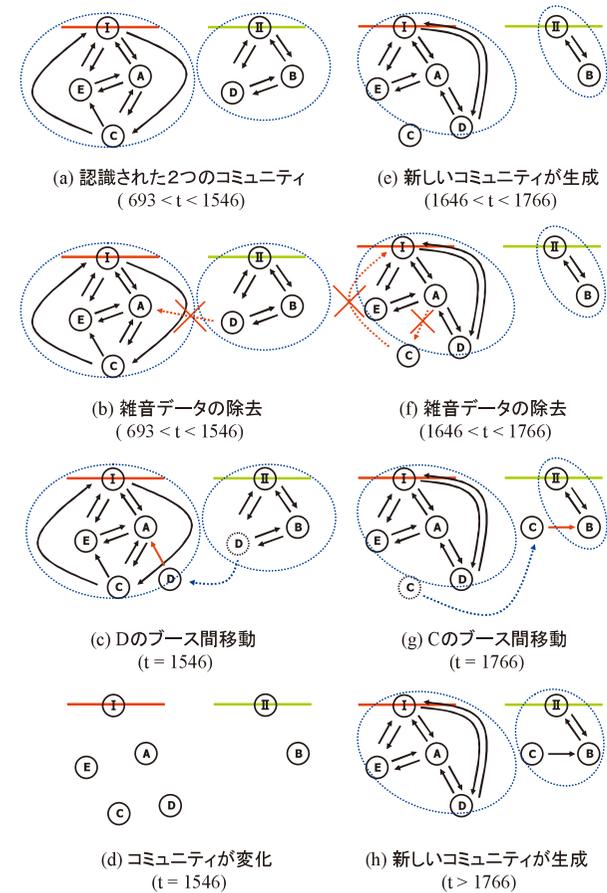


図 8 アルゴリズムの詳細な動作例
Fig. 8 Detailed performance of our algorithm.

1,546 (秒) までの間に図 8 (a) のような 2 つのコミュニティが検出された (リンクの構成は t = 1,546 (秒) 時点のもの)。この間、図 8 (b) のように D の A に対するインタラクション (D LOOK_AT A) がたびたび検出されたが、これらはたまたま D の視界に A が入っただけの雑音データであったため、リンクの強度を考慮することにより除去することができた。さらに、t = 1,546 (秒) において D がブース間を移動し、A とのインタラクショ

ン (D LOOK_AT A, A LOOK AT D) を行うと (図 8(c)), このリンクの強度が十分大きかったためにコミュニティが変化したことをシステムが検知し, コミュニティの構成が初期化された (図 8(d)). その後, コミュニティが再構成され, いくらかの過渡期状態を経た後, $t = 1,646$ (秒) には図 8(e) のような 2 つのコミュニティが検出された (リンクの構成は $t = 1,766$ (秒) 時点のもの). この間, C は booth1 に滞在していたにもかかわらず, 同じく booth1 に滞在していた A や I とのインタラクションが希薄であったためコミュニティから除外された (図 8(f)). 続いて C がブース間を移動し, $t = 1,766$ (秒) において B とのインタラクション (C LOOK_AT B) を行うと (図 8(g)), そのリンクの強度が十分大きかったためコミュニティが拡大されて図 8(h) のような新たなコミュニティが生成された. このように, 本実験ではオブジェクト間で非常に複雑なインタラクションが行われていたが, 提案手法を利用することにより柔軟にコミュニティの変化を追跡することができた. また同時に, 2 章で説明を行ったシーンを抽出することが困難な場面に対しても, 提案手法が効果的であることが確認された.

5. 考 察

従来研究で開発されてきたグラフのクラスタリング手法は, 時間の概念を排除してグラフの接続関係のみに基づいて部分グラフを検出するものであった. たとえば, 十分少ない数のクラスタから推論を始め, リンクの重みに応じてそのクラスタを分割していく手法⁹⁾ や, authority と hub と呼ばれるノードの集合を同定することで精度の高いクラスタリングが実現できる HITS アルゴリズム¹⁰⁾ などは Web コミュニティの発見などに応用されているが, 検出できるコミュニティは時間に依存しないものに限られる. 一方, 本稿で提案した時空間クラスタリング手法では, 時間にもなって変化する情報を扱って部分グラフの変化を検出できる.

リンクに付与された重みが時間の関数で記述される場合, 従来のクラスタリング手法を利用して部分グラフの変化を発見しようとする, 特定の時刻においてクラスタリングを行う処理を繰り返し実行し, 異なる時刻で検出されたコミュニティどうしの対応関係を調べることによってその変化を検出しなければならない. しかし, 特定の時刻における空間方向の情報のみを利用して部分グラフを抽出した場合, 雑音データに対して敏感に反応してしまい, クラスタリングの結果にばらつきが出てしまうことを確認している. 一方, 提案手法では, 空間方向に加えて時間方向の情報をも考慮に入れることにより, 一時的な雑音データに影響されずにロバストな推定を行うことができる. さらに, 従来手法を利用する場合, 細かい時間間隔でクラスタリングを繰り返し実行する必要があるため, 膨大な計算量を必要とする.

一方, 提案手法では, 部分グラフの構成に変化が生じる可能性のある場面でのみ評価を行うため, 計算量はきわめて小さい. たとえば手法 9) では, ノードの数に対して 2 乗のオーダーで計算量が増加するのに対し, 提案手法の場合はノード数に対してたかだか 1 乗のオーダーにすぎない.

最後に, 提案手法の可能性と発展性について議論を行う. 本稿では, 人のインタラクションの発生が時間的および空間的な制約を受けにくいポスター展示会場を選んでシーンの推定を行った. しかし, 我々は人が固定された席に着いて議論を行う会議や講義の場においても同様にインタラクションの記録と解釈を行うシステムの開発を進めており¹³⁾, このような場所においても提案手法が有効であることを確認している.

また, 我々は, 提案手法が本稿の目的である人と人, 人と物, 人と環境の関係性をシーンとして発見する問題だけでなく, 幅広い事例に適用することができる強力なデータマイニングツールであると考えている. たとえば, ハイパーテキスト構造を有する Web ページ間の関連性を構造化したい場合, 各ページをノードとして設定し, それらの間の参照関係を参照回数が多いほど太いリンクとして有向グラフを設定することにより, 関連のある Web ページの集合を時系列に抽出することができる. また, 電子掲示板に投函される一意見を示すデータをノードとして設定し, 賛成や反対といった, 意見間の応答回数を反映させた重みを付けた有向グラフを設定することで, 掲示板における会話の流れを構造化することができる. さらに, 人と電子機器をノードとして設定し, それらの対面関係を反映させた重みを付けた有向グラフを設定することで, 人が操作する対象となる機器を特定することが可能となり, ネットワークで連結されたデジタル家電製品の制御に応用できると考えられる. 今後, これらの事例に対する提案手法の適用を試み, その可能性について検証したい.

6. おわりに

本稿では, 時系列に発生する意味のあるシーンを発見的に抽出することを目的として, 時間にもなって変化する重み付きグラフに対してクラスタリングを行い, 部分グラフを抽出する手法について提案した. さらに, ポスター展示会場といった開放的な空間において見学者と説明者のインタラクションを記録し, 重要なシーンの抽出を試みた. その結果, 実際の状況と比較しても妥当であるといえる解釈を行うことができた.

今後の課題としては, 非階層的クラスタリングを行う本手法を階層的なものに拡張することで, より柔軟に部分グラフを抽出可能にすることがあげられる. また, 本手法はデータマイニングツールとして提案しているため, 抽出されたシーンの意味が特定できていない. そ

のため、抽出されたシーンを分析してインタラクションのパターンを構造化したり、コミュニティ内で発生するインタラクションの時系列推移から確率的な手法を用いて人の行動を理解したりすることにより、発見したシーンに対してその意味を自動的に解釈する研究にも着手したい。さらに、提案手法はデータとデータのつながりを時系列に発見する別の事例にも適用できると考えられるため、今後 Web コミュニティを抽出する問題などに適用して評価を行いたい。

謝辞 本研究開発の一部は情報通信研究機構の研究委託「超高速知能ネットワーク社会に向けた新しいインタラクション・メディアの研究開発」により実施したものである。

参 考 文 献

- 1) 角 康之, 伊藤禎宣, 松口哲也, Fels, S., 間瀬健二: 協調的なインタラクションの記録と解釈, 情報処理学会論文誌, Vol.44, No.11, pp.2628-2637 (2003).
- 2) Mase, K., Sumi, Y., Toriyama, T., Tsuchikawa, M., Ito, S., Iwasawa, S., Kogure, K. and Hagita, N.: Ubiquitous Experience Media, *IEEE Multimedia*, Vol.13, No.4, pp.20-29 (2006).
- 3) Takahashi, M., Ito, S., Sumi, Y., Tsuchikawa, M., Kogure, K., Mase, K. and Nishida, T.: A Layered Interpretation of Human Interaction Captured by Ubiquitous Sensors, *The 1st ACM Workshop on Continuous Archival and Retrieval of Personal Experiences (CARPE 2004)*, pp.32-38 (2004).
- 4) Cutler, R., Rui, Y., Gupta, A., Cadiz, J.J., Tashev, I., He, L.-W., Colburn, A., Zhang, Z., Liu, Z. and Silverberg, S.: Distributed Meetings: A Meeting Capture and Broadcasting System, *Proc. ACM Multimedia 2002* (2002).
- 5) Chiu, P., Kapuskar, A. and Wilcox, L.: Meeting Capture in a Media Enriched Conference Room, *Proc. CoBuild'99*, pp.79-88 (1999).
- 6) 青木茂樹, 岩井嘉男, 大西正輝, 小島篤博, 福永邦雄: 人物の位置・姿勢に注目した行動パターンの学習・認識と非日常状態検出への応用, 電子情報通信学会論文誌, Vol.J87-D-II, No.5, pp.1083-1093 (2004).
- 7) 神鷹敏弘: データマイニング分野のクラスタリング手法 (1) — クラスタリングを使ってみよう!, 人工知能学会誌, Vol.18, No.1, pp.59-65 (2003).
- 8) 神鷹敏弘: データマイニング分野のクラスタリング手法 (2) — 大規模データへの挑戦と次元の呪いの克服, 人工知能学会誌, Vol.18, No.2, pp.170-176 (2003).
- 9) Ding, C., He, X., Zha, H., Gu, M. and Simon, H.: Spectral Min-Max Cut for Graph Partitioning and Data Clustering, *Proc. 1st IEEE Int'l Conf. Data Mining*, pp.107-114 (2001).
- 10) Kleinberg, J.: Authoritative sources in a hyperlinked environment, *Proc. 9th ACM-SIAM Symposium on Discrete Algorithms* (1998).

- 11) 堀田政二, 井上光平, 浦浜喜一: 重み付きグラフからのファジークラスタ抽出, 電子情報通信学会論文誌, Vol.J84-A, No.3, pp.351-359 (2001).
- 12) 伊藤禎宣, 角 康之, 間瀬健二: 赤外線 ID センサを用いたインタラクション記録装置, 情報処理学会研究報告 (ヒューマンインタフェース), HI104-4 (2003).
- 13) 高橋昌史, 伊藤禎宣, 土川 仁, 角 康之, 間瀬健二, 小暮 潔: インタラクション解釈における階層構造の検討, 第 18 回人工知能学会全国大会 (2004).

(平成 19 年 9 月 9 日受付)

(平成 20 年 2 月 5 日採録)



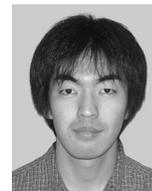
高橋 昌史

2003 年大阪府立大学工学部情報工学科卒業。2005 年京都大学大学院情報学研究所修士課程修了。同年 (株) 日立製作所中央研究所入社。在学中は、人のインタラクションの理解とモデル化に関する研究に従事。現在は、動画像符号化技術の研究開発に従事。電子情報通信学会、映像情報メディア学会各会員。2006 年度映像情報メディア学会鈴木記念奨励賞受賞。



角 康之 (正会員)

1990 年早稲田大学理工学部電子通信学科卒業。1995 年東京大学大学院工学系研究科情報工学専攻修了。同年 (株) 国際電気通信基礎技術研究所 (ATR) 入所。2003 年より京都大学大学院情報学研究所助教授 (現在は准教授)。博士 (工学)。研究の興味は、知識や体験の共有を促すシステムの開発や、人のインタラクションの理解と支援に関わるメディア技術。



伊藤 禎宣 (正会員)

2003 年北陸先端科学技術大学院大学知識科学研究科博士後期課程修了。同年 ATR メディア情報科学研究所研究員。2006 年より東京農工大学大学院工学府特任講師 (2007 年より特任准教授), NICT ユニバーサルメディア研究センター短時間研究員, ATR 知識科学研究所客員研究員。博士 (知識科学)。HCI, CSCW に興味を持つ。



間瀬 健二 (正会員)

1981年名古屋大学大学院工学研究科情報工学専攻修士課程修了。同年NTT入社。1995～2002年(株)国際電気通信基礎技術研究所研究室長。2002年より名古屋大学情報連携基盤センター教授。コンピュータによるコミュニケーション支援の研究を推進している。IEEE, ACM, 電子情報通信学会, 人工知能学会, 日本VR学会, 画像電子学会各会員。1999年人工知能学会論文賞。博士(工学)。



小暮 潔 (正会員)

1981年慶應義塾大学大学院工学研究科電気工学専攻修士課程修了。同年日本電信電話公社に入社。現在, ATR 知識科学研究所所長。博士(工学)。自然言語処理, エージェント, ロボット, 知的環境等の研究に従事。電子情報通信学会, 人工知能学会, 言語処理学会, 日本認知科学会, 日本音響学会各会員。



西田 豊明 (フェロー)

1977年京都大学工学部卒業。1979年同大学大学院修士課程修了。1993年奈良先端科学技術大学院大学教授, 1999年東京大学大学院工学系研究科教授, 2001年東京大学大学院情報理工学系研究科教授を経て, 2004年4月京都大学大学院情報学研究科教授。会話情報学, 原書知識モデル, 社会知のデザインの研究に従事。日本学術会議連携会員。本学会理事, フェロー。最近の編著: Toyooki Nishida (Ed.). Conversational Informatics: An Engineering Approach, John Wiley & Sons Ltd., 2007。