マルチモーダルインタラクション記録からのパターン発見手法

森田友幸[†1,†4] 平野靖[†2] 角康之[†3,†4]
梶田将司[†2] 間瀬健二[†2,†4] 萩田紀博[†5]

本論文では，複数人のインタラクション行動の記録の中から，インタラクションの重要なパターンを発見・抽出するための手法を提案する．抽出したパターンを用いると，センサ群により記録された人間のインタラクションのデータ集合であるインタラクション・コーパスを構築する際に，設計時には気づかなかったインデックスを付与することができるようになる．イベントのインデックスはインタラクションに関する様々な文脈の情報を保持しており，出来事を要約したり，特定の出来事を検索するといった用途に用いることができる．提案手法はインタラクションにおいて基本的なイベントである注視や発話といった要素イベントの同時発生パターンを抽出する．提案手法により，トップダウンによる直感的な定義では得られなかったインタラクションイベントを実データをもとに統計的にもっともらしく定義できる．この手法を評価するために，擬似ポスター展示会を行って得られたデータに対し適用した結果，いくつかの興味深いインタラクション・パターンが抽出され，それらに対して解釈を試みた．

A Method for Mining Patterns from Multimodal Interaction Log

Tomoyuki Morita,[†1,†4] Yasushi Hirano,[†2] Yasuyuki Sumi,[†3,†4]
Shoji Kajita,[†2] Kenji Mase[†2,†4] and Norihiro Hagita[†5]

This paper proposes a novel mining method for multimodal interactions to extract important patterns of group activities. These extracted patterns can be used as machine-readable event indices in developing an interaction corpus based on a huge collection of human interaction data captured by various sensors. The event indices can be used, for example, to summarize a set of events and to search for particular events because they contain various pieces of context information. The proposed method extracts simultaneously occurring patterns of primitive events in interaction, such as gaze and speech, that in combination occur more consistently than randomly. The proposed method provides a statistically plausible definition of interaction events that is not possible through intuitive top-down definitions. We demonstrate the effectiveness of our method for the data captured in a experimental setup of a poster-exhibition scene. Several interesting patterns are extracted by the method, and we examined their interpretations.

1. はじめに

我々は，人や物の間でのインタラクションを記録し分析することを目的として，機械可読なインデックスデータの集合であるインタラクション・コーパスの開発を行っている．その中で我々は，単に映像や音声，注視や発話の区間といった情報を記録するだけでなく，「共同注視」や「グループ討論」といった，より上位のインタラクションや状況に関するインデックスを付加する試みを行ってきた[11]．文献 10) の中で我々は，インデックス付与対象となるインタラクションをいくつか定義した．しかしこれらはコーパス設計者の直感や経験により定められていた．このように，インデックスの種類はコーパス設計者の直感によりトップダウンで設計することも可能であり，そのような方法が採られることが多い．しかし直感による設計では，間違いや見落としを含む可能性が高い．

そこで本論文では，注視状況や発話状況といったイ

†1 名古屋大学大学院情報科学研究科
Graduate School of Information Science, Nagoya University
†2 名古屋大学情報連携基盤センター
Information Technology Center, Nagoya University
†3 京都大学大学院情報学研究科
Graduate School of Informatics, Kyoto University
†4 ATR メディア情報科学研究所
ATR Media Information Science Laboratories
†5 ATR 知能ロボティクス研究所
ATR Intelligent Robotics and Communication Laboratories

ンタラクションの基本要素を実際の場面で記録したデータの集合から，重要なインタラクションのパターンを抽出する方法を提案する．提案手法は，インタラクションの基本要素の集合で表される同時発生的なパターンを抽出することを特徴とする．適切な同時発生的なパターンがいくつか抽出されると，それらを用いてインタラクションの状況を効率良く，また分かりやすく記述することができるようになる．

本研究は，人対人や人対物のインタラクションが生じる場面で暗黙的に用いられている社会的プロトコルや頻出するインタラクションを抽出して記述することによって，場や人間行動の分析と知識化を目指している．その対象はオフィスや工場における作業フローの分析をはじめ，会議の要約，教室でのやりとり，住宅内や購買行動の動線分析など広範囲に及ぶ．それぞれの分野において特徴的なインタラクションのパターンは，一部の共通的なパターンを除いては異なることが予想されるため，ある基準で重要度を規定し自動的にパターンを抽出する手法の開発が望まれる．そこで，本論文では手始めに注視と発話を含む人と物とのインタラクションが生じる比較的単純な場面として，ポスターの展示説明を取り上げて，データマイニング手法を用いたパターン抽出手法を開発した．また，提案手法は本論文で用いるデバイスに特化したものではなく，他のデバイスを用いた場合にも適用可能なものである．

データマイニングは，大きなデータ集合の中から知識発見を行うための強力な手法であり，医療や遺伝子の研究，またはマーケティングといった様々な分野で用いられ，多くの成果を残している．インタラクションの分析においても，インタラクションの情報を記録したデータに対してデータマイニングの考え方を適用することにより，インタラクションに関する様々な新しい知識が発見できると期待される．

インタラクション・コーパスは様々な目的で利用できる．イベントのインデックスはインタラクションに関する様々な文脈の情報を保持しており，出来事を要約したり，特定の出来事を検索するといった用途に用いることができる．また，このようなコーパスは，認知科学の研究者にとって人間のインタラクションを分析するための便利なツールになると考えられる．

本論文の構成は以下のようになっている．まず 2 章で，本論文に関連する研究について触れる．3 章では，インタラクションの記録方法について簡単に説明する．4 章では，まずインタラクション・パターンについて考察し，インタラクション・パターンのモデルを提案し，最後にパターンの抽出方法を示す．5 章では，擬似ポスター展示会を行って取得されたデータに対して提案手法を適用した実験結果および考察を示す．最後に 6 章では，本論文のまとめと今後の課題について示す．

2. 関 連 研 究

部屋の中にセンサ群を設置することにより，部屋の中にいる人間に対してサービスを行うことを目的とする様々な研究がなされている．たとえば，Smart rooms[9]，Intelligent room[1]，AwareHome[5]，EasyLiving[2] などがそれである．これらの研究の目的は，個々の人間の振舞いの認識と，意思の理解と利用といえよう．一方，我々の興味は単に 1 人の人間の振舞いだけでなく，複数の人間の間でのインタラクションを記録することにある．

複数の人物の行動を記録し，「誰が」「いつ」「どこで」「何を」していたのかを推定する研究として PEPYS[8] がある．PEPYS では，ActiveBadge を利用して人物の滞在，移動情報を取得し，ユーザの行動を記録する．また，複数の人物が同じ場所に滞在していると「ミーティング」と認識するといったように，複数の人物の情報を統合して集団としての行動をスタティックに推定し記録する．我々の目的も複数の人物間でのインタラクションの記録であるが，ポスター展示会のように，よりインタラクションがダイナミックに変化する場の記述を想定している点で異なる．そのような場で PEPYS のようなシステムを用いた場合，各場所に人物が集まっておりつねに「ミーティング」と認識されてしまい場の記述の粒度が粗いという問題がある．

映像と音声のデータベースに対してインデックスの付与を行う研究には様々なものがある．たとえば，Miyamori ら[6] はスポーツの場面において物やプレーヤの様々な種類の動きに対してインデックスを付与することにより，特定のシーンを検索する方法を提案している．この場合，対象とする領域がスポーツであり，試合やプレーの内容といったものがすでに体系化されているためインデックスを付与する対象の選別は比較的容易である．同様に，映像と音声のデータベースに対してインデックスを付与する場合，対象とする領域に関してすでに体系化されている知識を用いることがほとんどである．それに対して，本論文で扱うミーティングやポスター発表といったインタラクションに関してはいまだ体系化がなされておらず知識が乏しい．そこで我々は，知識発見，特にデータマイニングのアプローチを採用して，この領域の知識発見を視野に入れて研究を進めることにした．

データマイニングの分野では，文献 3) や文献 4) で

時間軸上に配置されたイベントの集合の中から頻出するパターンを抽出する方法が提案されている．これらの手法はパターンの重要度の指標として頻度を用いて，イベントの発生順序のパターンを抽出する．我々の手法は，グラフ構造により表現されるデータを扱うグラフマイニングと関係がある．グラフマイニングに関する研究としては，たとえば，文献 7) では大きなグラフ構造の中で頻出する部分グラフを抽出する方法が提案されている．我々の手法は，グラフの表現を時間的な構造の変化が扱えるよう拡張し，その中から興味深い部分構造パターンを抽出するものと位置づけることができる．データマイニングの分野では，パターンの興味深さをはかる尺度として，頻度やサポート（全パターンに対する対象パターンの割合）といった比較的単純なものが用いられる場合が多い．しかし本論文では，パターンを抽出するために発生量の期待値とその実測値との比で表される新しい尺度を導入することを提案する．

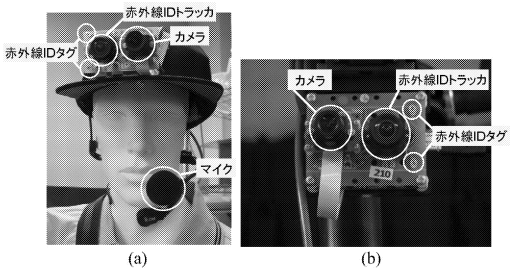


図 1 (a) ウェアラブルタイプ/(b) ユビキタスタイプセンサセット
Fig. 1 (a) Wearable/(b) Ubiquitous sensor client.

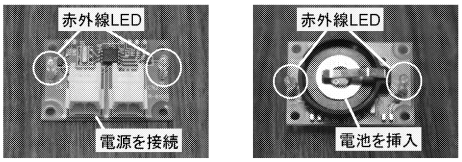


図 2 赤外線 ID タグ
Fig. 2 Infrared ID tag unit.

3. センサ群によるインタラクションの記録

この章では，インタラクション記録システム[11] を紹介し，センサ群を用いてどのようにインタラクションを記録するかを示す．人や物の間でのインタラクションを記録するために，ウェアラブルタイプとユビキタスタイプの 2 種類のセンサセットを用いる．ウェアラブルタイプは赤外線 ID システム，マイク，およびカメラからなる．各センサは **図 1** (a) に示すように帽子に取り付けられている．ユビキタスタイプは赤外線 ID システム，およびカメラからなり，環境に設置されるかまたはポスターパネルといった物に取り付けられる．図 1 (b) はユビキタスタイプのセンサセットを補助ポールに設置した例である．

人物の大まかな注視情報を記録するために赤外線 ID システムを用いる．赤外線 ID システムは ID タグ（**図 2**）と ID トラッカからなる．ID タグは赤外線の発光パターンにより固有の ID を発行する．そして，ID トラッカは 8 ～ 10 fps 程度の速度で ID タグの位置と ID を認識する．ID タグをオブジェクト（人や物）に設置し，ID トラッカを人物の視線に方向を一致させて頭部に装着することにより，オブジェクトの ID とカメラの画像中でのオブジェクトの位置を認識できる．文献 11) の中で，展示会場のように立ち位置や姿勢を自由に変えられる状況では，人は注視対象物を視野全体を使って追いかけるのではなく，頭部を含む姿勢変更がともなう姿勢移動により注視対象方向に頭部を定位させていることが実験により明らかになっている．よって，本論文でもこの赤外線 ID システムから得られた情報を人物の注視の情報の近似として用いる．

人物の発話の音声と区間を記録するためにマイクを用いる．発話区間はマイク入力のボリュームを閾値処理することにより記録される．

これらの，センサ群によって記録された人物の注視および発話の情報はネットワークを介して転送され，すべてのイベントが同期できるようにタイムスタンプ付きでデータベースに蓄積される．記録された生データはそのままでは扱いにくいため，MaxInterval と MinInterval という 2 つの閾値を用いて時間でクラスタリング[10] することにより注視や発話の区間を推定し，開始時刻と終了時刻による時間区間の情報に変換される．クラスタリングの手順は以下のとおりである．まず MaxInterval 以上の間隔をあけずに起きている 2 つのイベント（タグの捕捉または発話区間）を 1 つのイベントとして結合する操作が繰り返される．その後，MinInterval 以下の長さのイベントはノイズとして除去される．本論文では，注視に対しては MaxInterval は 8 秒，MinInterval は 4 秒，発話に対しては MaxInterval は 4 秒，MinInterval は 2 秒という値を用いた．

本論文ではこの時間区間により表現されたデータを用いる．

4. パターン抽出法

4.1 インタラクション・パターン

本論文では，同時に発生するイベントの組合せで表現される同時発生的パターンを扱う．同時発生的パターンは,「Talking」や「Discussing」といったインタラクションの状態を表す．たとえば,「人物 A が人物 B に話しかけている」という状況を考えると,「人物 A が人物 B を見る」,「人物 B が人物 A を見る」,「人物 A が発話する」という 3 つのイベントが同時に起きている．

同時発生的パターンを用いることにより，場の状況やインタラクションの状態に関する情報を得ることができる．このような情報は，様々なことに用いることができる．たとえばガイドロボットを使って人に対してサービスを行いたい場合，活発に議論を行っているところに割って入ったりするようなことは明らかに不適切であり，自然なサービスを行うためには場の状況に関する情報が不可欠である．また，出来事の要約や検索を行いたい場合も，たとえばポスターセッションを考えると「説明を聞いている」のか「議論している」のかといった情報が重要になる．ショッピングにおいては，商品を見ながら友人や店員との会話がどのようになされるか行動分析が可能となりマーケティングの支援になる．

4.2 インタラクション・パターンのモデル

まず，以降の章で用いる用語を定義する．

定義 1 イベント（Event）：イベントは，発話や注視といったインタラクションの基本要素で，開始時刻と終了時刻を持つ．

イベントは式中では「e」と表記する．また，そのイベントの開始時刻および終了時刻は「$e.start$」および「$e.end$」と表記する．

定義 2 エピソード（Episode）：エピソードはイベントの集合である．

エピソードは式中では「Epi」と表記する．

定義 3 パターン（Pattern）：パターンはエピソードの部分集合でこれはエピソードの種類を表す．

パターンは式中では「Pat」と表記する．

本論文ではインタラクションを構成する重要な要素として 2 種類の行動，すなわち「注視」と「発話」に着目して検討した．また，これらは 3 章に紹介したインタラクション記録システムを用いて収録することができる.「注視」のイベントは「LOOK イベント」（または単に「LOOK」）と表記する．また,「発話」のイベントは「SPEAK イベント」（または単に「SPEAK」）と表記する．LOOK イベントは，その時間区間の情報とともに動作の主体と客体の情報を持つ．SPEAK イベントは，その時間区間と動作の主体の情報を持つ．

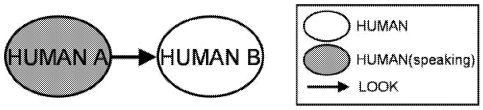


図 3 インタラクション・パターンの例
Fig. 3 Example of interaction pattern.

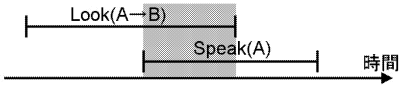


図 4 データ中でのパターンの発生
Fig. 4 Pattern occurred in data.

インタラクションのパターンを抽出するために，同時発生するイベントの組合せを表現するモデルを提案する．ここでは同時発生的パターンしか扱わないので，時間情報はモデルで表現される必要はない．そこで我々は，以下のような有向連結グラフをモデルとして用いる．

「人物」や「物」といったオブジェクトをグラフのノード（節点）として表す．各ノードは「HUMAN」や「DISPLAY」といったその実体の属性を表すラベルを持つ．有向エッジ（辺）はイベントの主体と客体関係を表し，またその種別（「LOOK」または「SPEAK」）をラベルとして持つ．LOOK イベントは主体と客体の情報を持つので，それを表現するエッジはその主体を表すノードを始点とし，客体を表すノードを終点とする．このようなイベントを「二項イベント」と呼ぶ．一方，SPEAK イベントは主体の情報しか持たないので，その主体を表すノードに対する自己ループとして表す☆．このようなイベントを「単項イベント」と呼ぶ．本論文の図中では，自己ループは見やすさのためにノードに着色することで表現する．ここで，グラフは連結なものに制限するが，それはパターンを構成するノードは互いにイベントを介して関連している必要があるからである．

たとえば,「人物 A が人物 B を見て発話する」というパターンは**図 3** のように表現される．このパターンは**図 4** のように 2 つのイベントの重なりとして発生する．

パターン中には，図 4 中の右側のノードのように SPEAK のエッジのない（着色されていない）ノードが含まれることがある．これは，一見するとそのノー

☆「SPEAK」の対象は「LOOK」の情報に頼る．

ドの人物が「発話していない」状態を示すように思われる．しかし，実際には「発話していない」のではなく「発話しているかしていないかは考慮されない」状態を示す．LOOK のエッジに関しても同様である．なぜなら，パターンはそのモデルに含まれるエッジの集合（=イベントの集合）のみが発生条件となるからである．パターンのモデルに含まれないエッジに関しては考慮されない．

本論文では，最も基本的かつ独立な事象である LOOK および SPEAK の 2 種類のイベントのみを取りあげた．しかし，それ以外のモダリティが利用可能な場合には，それらを単項もしくは二項イベントとして表現することで提案手法を適用可能である．たとえば,「触る」というイベントが検出可能で利用できるなら，二項イベントとして表現すればよい．また,「うなずく」というイベントの場合には，単項イベントとして表現する．さらに,「握手する」のように主体と客体が区別できないようなイベントの場合にも，両向きの 2 つの二項イベントとして表現すればよい．

4.3 インタラクション・パターンの抽出

4.3.1 発生時間尺度による抽出

パターンの抽出は，前節で示したモデルで表されるパターンについて，基本的にはより多く発生しているパターンほど重要と見なして行われる．より多く発生するパターンほど重要とするのは，発生量が多いものほどそのパターンを発生させやすくしている要因が存在していると考えられるからである．互いにまったく無関係に起こっているイベントであっても同時に起こることはある．しかし，そのような事柄と比較して，重要な関連性を持つ複数の事柄はより多く同時に起こると考えられるのである．データマイニングの分野でも，パターンの重要度を測る尺度としてパターンの発生頻度やそれに順ずるもの（サポート等）を用いることが多い[3),4),7)]．

パターンを抽出する対象となるデータは LOOK イベントと SPEAK イベントの集合である．まず最初に，全イベント集合から同時発生している部分集合をすべて抽出する．得られた部分集合の集合は，前節に示したモデルによって同型のグループに分割され，各グループ内で発生時間が集計される．この操作は，観測の開始時刻から終了時刻まで移動しながら，各時刻で起こっているイベントの考えられる組合せをすべて抽出し，各パターンの発生時間を集計するという操作と等価である．この操作は，以下のような手順で行われる．

まず，全イベント集合のうち，同時発生している部分集合（エピソード）をすべて抽出する．ここで,「同時発生している」とはそのエピソードに含まれるすべてのイベントが同時に起こっている時刻が一瞬でも存在するという意味である．たとえば**図 5** では，イベント A とイベント B は同時発生しているが，イベント B とイベント C はそうでない．換言すれば，エピソード（Epi）に含まれる全イベント（e）の開始時刻の最大値が終了時刻の最小値より小であるということである．

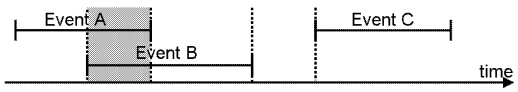


図 5 同時発生の例
Fig. 5 Example events.

$$\max_{e \in Epi}(e.start) \leq \min_{e' \in Epi}(e'.end)$$

次に，得られたエピソードの集合をそれぞれモデルで表現し，互いに同型であるか否かを判定していくことで，エピソードの集合を同型の部分集合に分割する．各部分集合がそれぞれ固有のパターンを表現する．

次に，各パターンに対して合計発生時間 $T(Pat)$ を計算する．これは，各パターンに含まれるエピソードの持続時間 $T(Epi)$ を合計することで行われる．

$$T(Epi) = \min_{e \in Epi}(e.end) - \max_{e' \in Epi}(e'.start)$$

$$\begin{aligned} T(Pat) &= \sum_{Epi \in Pat} T(Epi) \\ &= \sum_{Epi \in Pat} \left(\min_{e \in Epi}(e.end) - \max_{e' \in Epi}(e'.start) \right) \end{aligned}$$

この $T(Pat)$ の値を発生時間尺度と呼ぶ．これは，データマイニングの分野でよく用いられる，発生頻度に相当する尺度である．

4.3.2 正規化発生時間尺度による抽出

ここで，新たに正規化発生時間尺度 I を導入する．前項の発生時間尺度では，多く発生するイベント種別を含むパターンが多く抽出され，あまり起こらないイベント種別を含むパターンを見落としてしまう．そこで，尺度 I ではパターンの期待値で発生時間を正規化することにより，イベント種別間の発生量の偏りの影響を排除する．あるパターン Pat に対する尺度 $I(Pat)$ は，そのパターンの「実際の総発生時間 $T(Pat)$」と各イベント種別の総発生時間から求められる「総発生時間の期待値 $E(Pat)$」の比で表される．パターンの総発生時間の期待値は，各イベントが無作為に発生すると仮定したときの期待値である．これは，各人物が

周囲との関わりとはまったく関係なくランダムに発話や注視を行っているような状態で期待される発生量といえる．期待値に対して実際の値が大きいほど，つまり尺度 I の値が大きいほどそのパターンが実際の場面で発生しやすくなっており重要なパターンであると考えられる．

$$I(Pat) = \frac{\text{パターンの実際の総発生時間}}{\text{パターンの総発生時間の期待値}} = \frac{T(Pat)}{E(Pat)}$$

パターンの総発生時間の期待値は以下のように求められる．まず，各イベント種別ごとの総発生量を求める．次に，各イベント種別は種別間で互いに独立であり観測区間全体で均一に発生しているという仮定のもとに，各時刻におけるそのイベント種別の発生確率を求める．

ここで，導出のために時刻 t において各イベントが発生しているか否かを表す 2 値関数 L と S を導入する．

$$L_{O_a \to O_b}(t) = \begin{cases} 1, & \text{時刻 } t \text{ で } LOOK(O_a \to O_b) \\ & \quad \text{が起きている場合} \\ 0, & \text{それ以外の場合} \end{cases}$$

$$S_{O_a}(t) = \begin{cases} 1, & \text{時刻 } t \text{ で } SPEAK(O_a) \\ & \quad \text{が起きている場合} \\ 0, & \text{それ以外の場合} \end{cases}$$

すると，$LOOK(O_a \to O_b)$ および $SPEAK(O_a)$ の総発生時間 $T(L_{O_a \to O_b})$ および $T(S_{O_a})$ は，観測の開始時刻 t_0 および終了時刻 t_1 を用いて以下のように表される．

$$T(L_{O_a \to O_b}) = \int_{t_0}^{t_1} L_{O_a \to O_b}(t)dt$$

$$T(S_{O_a}) = \int_{t_0}^{t_1} S_{O_a}(t)dt$$

すべてのイベント種別は独立で時間的に均一に発生すると仮定しているので，各イベント種別の発生確率は時間に依存せず以下のように表される．

$$P(L_{O_a \to O_b}) = \frac{T(L_{O_a \to O_b})}{t_1 - t_0} = \frac{\int_{t_0}^{t_1} L_{O_a \to O_b}(t)dt}{t_1 - t_0} \quad (1)$$

$$P(S_{O_a}) = \frac{T(S_{O_a})}{t_1 - t_0} = \frac{\int_{t_0}^{t_1} S_{O_a}(t)dt}{t_1 - t_0} \quad (2)$$

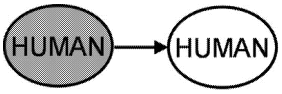


図 6 期待値計算例のパターン
Fig. 6 Example pattern.

ある特定のイベントの組合せが同時に起こる確率は，各イベントの発生確率の積で表される．

$$P(e_1, e_2 ... e_n) = \prod_{i=1}^{n} P(e_i)$$

次に，パターンを構成しうるイベント種別の組合せすべてについてその発生確率を求める．それらの値を合計することにより，パターンの発生確率が求められる．

$$P(Pat) = \sum_{E \in Pat} P(E)$$

ここで，E はイベント種別の組合せを表し，$E \in Pat$ は E がパターン Pat を構成することを意味する．パターンの総発生量の期待値は，この値に対して観測時間を掛けることにより求められる．

$$E[Pat] = \int_{t_0}^{t_1} P(Pat)dt = (t_1 - t_0)P(Pat)$$

ここで例として，**図 6** に示すパターンに対して発生量の期待値を求める．ただし，人物 A，B，C の 3 名が存在するものとする．このパターンを構成するイベントの組合せは，以下の 6 種類である．

$$\{LOOK(A \to B), SPEAK(A)\},$$
$$\{LOOK(A \to C), SPEAK(A)\},$$
$$\{LOOK(B \to A), SPEAK(B)\},$$
$$\{LOOK(B \to C), SPEAK(B)\},$$
$$\{LOOK(C \to A), SPEAK(C)\},$$
$$\{LOOK(C \to B), SPEAK(C)\}$$

期待値 $E[Pat]$ は以下の式で表される．

$$\begin{aligned} E[Pat] &= \int_{t_0}^{t_1} P(Pat)dt \\ &= (t_1 - t_0)P(Pat) \\ &= (t_1 - t_0) \sum_{E \in Pat} P(E) \\ &= (t_1 - t_0)(P(L_{A \to B} \wedge S_A) + P(L_{A \to C} \wedge S_A) \\ &\qquad + P(L_{B \to A} \wedge S_B) + \ldots) \\ &= (t_1 - t_0)(P(L_{A \to B}) \times P(S_A) \\ &\qquad + P(L_{A \to C}) \times P(S_A) + \ldots) \end{aligned}$$

ここで，$P(L_{A \to B}), P(L_{A \to C}) \ldots, P(S_A), P(S_B), P(S_C)$ は式 (1) および式 (2) から求められる．

5. 実　　験

5.1 実験の概要

提案手法を評価するために，**図 7** に示すように擬似ポスター展示会を行った．人物 5 人が参加し，ディスプレイ 2 枚を用いた．人物 2 人は説明員として参加し，それぞれ 1 枚ずつのディスプレイの横に付いてディスプレイに表示した展示について説明を行う．残りの人物 3 人は見学者として参加し，自由に移動して展示を見学するよう指示された．人物 5 人はそれぞれ図 1 (a) のウェアラブルタイプセンサセットを装着した．また**図 8** に示すように，図 1 (b) のユビキタスタイプセンサセットをディスプレイの上部に設置し，さらにそれぞれ左右の 2 カ所に赤外線 ID タグを設置した．擬似ポスター展示会は約 1 時間半に及んだ．実験中に取得されたカメラ画像を**図 9** に示す．

各人物が装着したセンサセットの ID には「HUMAN」というラベルを関連付けた．よって，各人物を表すノードのラベルは「HUMAN」となる．ディスプレイの上部に設置したセンサセットの ID には「DISPLAY」というラベルを関連付けた．各ディスプレイの左右のタグの ID は，そのディスプレイのセンサセットに付いたタグの ID と関連付けた．つまり，センサセットの ID と左右のタグの ID を捕えることは等価である．

実験を行い約 87 分間のインタラクションを記録した．得られたデータに対して，正規化発生時間尺度を用いてインタラクション・パターンの抽出を行った．

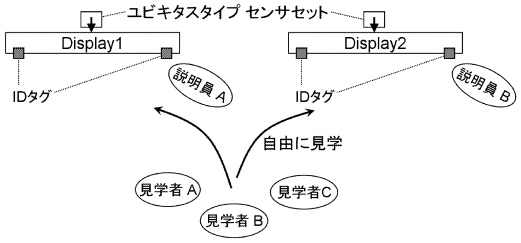


図 7 実験の概要
Fig. 7 Setup of the experiment.

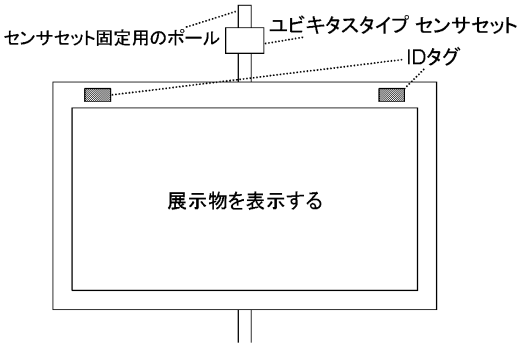


図 8 ディスプレイ設置概要
Fig. 8 Setup of the display.

5.2 結　　果

抽出を行った結果，ノード数 2，3 および 4 のパターンがそれぞれ 10，113 および 402 種類抽出された．ノード数が 5 以上のパターンは得られなかった．各ノード数のパターンに対して，I の値が上位 5 位までのパターンを**図 10** に示す．

5.3 考　　察

各ノード数に対して上位 5 位までのパターンに対してレビューを行った．

ノード数が 2 のパターンは，インタラクションの最も基本的なパターンを表現している．図 10 の 2-1, 2-4 および 2-5 のパターンは，「2 者のうち少なくとも一方が発話している」，「2 者が互いに向き合っている」および「2 者がともに発話している」というようにそれぞれが会話の中での状態を表している．2-2 および 2-3 のパターンは「人物がディスプレイを見ている」および「人物がディスプレイを見ながら発話している」

図 9 実験中に取得されたカメラ画像
Fig. 9 Captured images from the experiment.

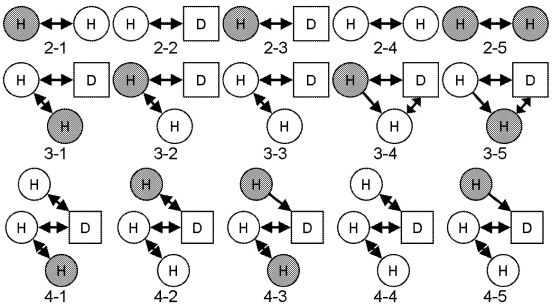


図 10 ノード数 2，3，4 に対する上位 5 位までのパターン．「H」は「HUMAN」，「D」は「DISPLAY」を意味する．パターンの下に示した数字はノード数と順位である．たとえば，パターン 2-3 はノード数 2 の 3 位のパターンである
Fig. 10 Top five Patterns for two, three, and four nodes. “H” indicates “HUMAN” and “D” indicates “DISPLAY”. The numbers under the pattern mean the number of nodes and the rank. For example, pattern 2-3 is the third ranked of two-node patterns.

という状態をそれぞれ表している．このパターンは，見学者はディスプレイの前でディスプレイを見ながら会話をする時間が多かったため抽出された．

ノード数が 3 のパターンは，展示会という場面の中での基本的なパターンを表現している．3-1，3-2，および 3-3 のパターンは，説明員と見学者との会話の場面を表している．より詳しくいうと，展示員は見学者を見て，見学者はディスプレイを見て，展示物について会話しているという場面である（これは，実験で収録された映像のうち，パターンの発生している付近の時間のものを見ることにより確認した）．3-4 および 3-5 のパターンは，2 人の人物が展示物を見ながら会話している場面を表している．

ノード数が 4 のパターンは，基本的にはノード数が 2 および 3 のパターンの単純な組合せである．パターン 4-1 は 3-1 と 2-2 の組合せであり，パターン 4-2 は 3-3 と 2-1 の組合せであり，パターン 4-4 は 3-3 と 2-2 の組合せである．

ノード数が 5 以上のパターンは抽出されなかった．これは，見学者が 3 人であり，4 人以上の人物が 1 つの展示の前に集まることがなかったためである．

我々は，得られたこれらのパターンのうち，2-1，3-1，3-2，3-3 が特に重要であると考える．まず，2-1 は 2 者が向き合って会話している場面を示している．これは，文献 10) の中でイベントの解釈としてあげられた「会話」のパターンと類似する．ただし，文献 10) の中では単に 2 者が向き合っている 2-4 のような場面を想定していたが，2-1 では一方が発話している場面が切り取られている．3-1，3-2，3-3 のパターンは，説明員が説明を行う場合に見学者を見て話すという重要な一面を切り取っている．これまで我々は，複数の人物がある物を見て会話するという「TALK ABOUT」というインタラクションを想定していた．しかし，今回発見されたこれらのパターンにより，説明員が説明を行う場面や見学者が発言する場面を認識しインデックスを付加することが可能となる．このように，提案手法を用いることで，これまでの設計者の直感による方法ではまったく想定されなかったインタラクションのパターンが発見された．

5.4 他の尺度との比較

新たに導入した正規化発生時間尺度の有効性を示すために，他の単純な尺度との比較を行う．比較対象として，4.3.1 項の発生時間尺度と，パターンの発生した頻度による発生頻度尺度の 2 種類をあげる（**図 11**）．ここでは，抽出結果が特に有益であったと考えられるノード数が 3 のパターンについて詳しく比較する．

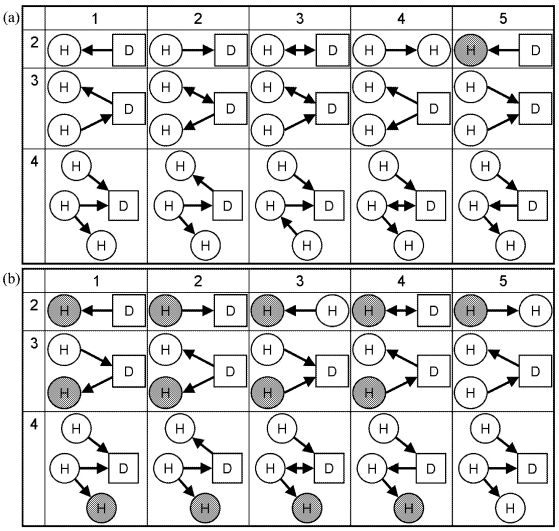


図 11 (a) 発生時間尺度/(b) 発生頻度尺度を用いたときの結果．列は順位，行はノード数を表す
Fig. 11 The results measured by (a) total occurrence time/(b) total occurring frequency.

まず，発生時間尺度を用いた場合のパターンを見る（図 11 (a)）．人間とディスプレイの間の LOOK イベントを含み，SPEAK イベントを含まないパターンが上位に抽出されている．これらのパターンは以下の 2 点の理由から抽出されたと考えられる．第 1 に，人間とディスプレイの間での LOOK イベントの総発生時間が長かった．第 2 に，SPEAK イベントの発生時間（すなわち発声時間）は総じて短いため，パターンに SPEAK イベントを付加すると発生条件が厳しくなりパターンの総発生時間が短くなる．

次に，発生頻度尺度を用いた場合のパターンを見る（図 11 (b)）．人間とディスプレイの間の LOOK イベントを含み，SPEAK イベントも含むパターンが上位に抽出されている．こららのパターンは以下の 2 点の理由から抽出されたと考えられる．第 1 に，人間とディスプレイの間での LOOK イベントの発生回数が多い．第 2 に，SPEAK イベントの発生回数が多い．第一の理由から人間とディスプレイの間の LOOK イベントを含むパターンが抽出されやすくなっている．また，第二の理由から，SPEAK イベントを含むことはマイナスには働かないため SPEAK イベントを含むパターンも抽出されている．

このように，発生時間尺度および発生頻度尺度では，イベント種別ごとの発生量の差の影響が大きく，好ましい結果が得られていない．ノード数が 2 および 4 の結果に対しても同様のことがいえる．環境に設置したディスプレイとのインタラクションが多いことは，

ディスプレイ中心の行動分析では必ず生じる事態であるが，そのような状況にあっても人間中心の分析をする必要がある．それに対して，我々の提案する正規化発生時間尺度ではイベント種別間の偏りに影響されない結果が得られていることが分かる．

6. まとめと今後の課題

マルチモーダルなインタラクション記録の中から，重要なパターンを抽出する手法を提案した．また，擬似ポスター展示会の実験を行い得られたデータに提案手法を適用した．得られた結果から発見された重要なインタラクション・パターンを示し，提案手法により重要なパターンが抽出可能であることを示した．また，提案した正規化発生時間尺度が発生時間尺度や発生頻度尺度に比べて，イベント種別間の発生量の差に影響されにくいパターン抽出に効果があることを示した．

正規化発生時間尺度を用いて得られた順位から，上位何位までが重要なパターンとして抽出されるべきかを決定する手法が未開発である．ここで我々は，各ノード数のグループに対して 5 位までというマジックナンバを用いた．これは，何位までレビューする必要があるかを設計者が決定する必要があり，設計者の感覚に影響される可能性がある．よって，それを決定する手法を開発する必要があると考える．この問題に対しては，抽出したパターン群が全パターン空間をどの程度効率的に記述できるかというような評価基準を導入する必要がある．たとえば全空間を重複を少なくかつ広範囲にカバーするパターン集合を良いとするような基準が考えられる．これについては今後の課題としたい．

本論文では，同時発生的パターンを扱った．インタラクションのパターンには，同時発生的パターンのほかに時系列的パターンが存在すると考えられる．よって，インタラクション・コーパスの語彙を豊かにするためにも，時系列的なパターンを抽出する手法を開発する必要があると考える．それにより，社会的プロトコルやそのほかのインタラクションが形式的にもしくは体系的に表現される．

謝辞 本実験を実施するにあたって ATR メディア情報科学研究所の研究員諸氏にご協力，ご議論いただいた．これらの諸氏に感謝する．本研究の一部は，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」ならびに文部科学省平成 16 年度「知的資産の電子的な保存・活用を支援するソフトウェア技術基盤の構築」研究開発課題「ユビキタス環境下での高等教育機関向けコース管理システム」により実施したものである．

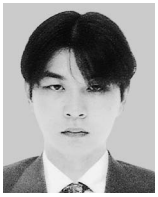
参 考 文 献

1) Brooks, R.A., Coen, M., Dang, D., De Bonet, J., Kramer, J., Lozano-Perez, T., Mellor, J., Pook, P., Stauffer, C., Stein, L., Torrance, M. and Wessler, M.: The intelligent room project, *Proc. 2nd International Cognitive Technology Conference* (*CT'97*), pp.271–278, IEEE (1997).
2) Brumitt, B., Meyers, B., Krumm, J., Kern, A. and Shafer, S.: EasyLiving: Technologies for intelligent environments, *Proc. HUC2000*, pp.12–29, Springer LNCS1927 (2000).
3) Casas-Garriga, G.: Discovering unbounded episodes in sequential data., *PKDD*, pp.83–94 (2003).
4) Mannila, H., Toivonen, H. and Verkamo, A.I.: Discovery of frequent episodes in event sequences, *Data Mining and Knowledge Discovery*, Vol.1, No.3, pp.259–289 (1997).
5) Kidd, C.D., Orr, R., Abowd, G.D., Atkeson, C.G., Essa, I.A., MacIntyre, B., Mynatt, E., Startner, T.E. and Newstetter, W.: The aware home: A living laboratory for ubiquitous computing research, *Cobuild'99*, pp.190–197, Springer LNCS1670 (2000).
6) Miyamori, H. and Iisaku, S.: Video annotaion for content-based retrieval using human behavior analysis and domain knowledge, *FG2000*, p.320, IEEE (Mar. 2000).
7) Kuramochi, M. and Karypis, G.: Frequent Subgraph Discovery, *Proc. ICDM'01*, pp.313–320, IEEE (2001).
8) Newman, W.M., Eldridge, M.A. and Lamming, M.G.: PEPYS: Generating autobiographies by automatic tracking, *Proc. ECSCW'91*, pp.175–188 (1991).
9) Pentland, A.: Smart rooms, *Scientific American*, Vol.274, No.4, pp.68–76 (1996).
10) 角 康之，伊藤禎宣，松口哲也，シドニーフェルス，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).
11) 伊藤禎宣，岩澤昭一郎，土川 仁，角 康之，間瀬健二，片桐恭弘，小暮 潔，萩田紀博：装着型体験記録装置による対話インタラクションの判別機能実装と評価，ヒューマンインタフェース学会論文誌，Vol.7, No.1, pp.167–178 (2005).

(平成 17 年 5 月 31 日受付)
(平成 17 年 11 月 1 日採録)

森田　友幸（学生会員）

1981 年生．2005 年名古屋大学工学部電気電子・情報工学科卒業．現在，同大学院情報科学研究科社会システム情報学専攻博士課程前期課程に在籍．研究の興味はコンピュータによるコミュニケーション支援とヒューマンインタフェース．



平野　　靖

1995 年名古屋大学工学部電子情報工学科卒業．1997 年同大学院博士課程前期課程（電子情報学専攻）修了．1999 年同大学院博士課程後期課程（情報工学専攻）修了．2000 年同大学院工学研究科助手．2002 年同大学情報連携基盤センター助手．2004 年同大学情報連携基盤センター助教授．博士（工学）．3 次元画像処理とその肺腫瘤の良悪性鑑別への応用に関する研究，および大学内・大学間の認証システムに関する業務に従事．電子情報通信学会，IEEE 等の会員．

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科助教授．博士（工学）．研究の興味は知識処理システムとヒューマンインタフェース．

梶田　将司（正会員）

1990 年名古屋大学工学部情報工学科卒業．1995 年同大学院工学研究科情報工学専攻博士課程満了，名古屋大学工学部助手．1998 年同情報メディア教育センター助手．2002 年情報連携基盤センター助教授，文部科学省メディア教育開発センター客員助教授併任，2003 年株式会社エミットジャパン取締役兼任．博士（工学）．大学における教育・研究活動での IT 活用に関する研究に従事．1996 年電気関係学会東海支部連合大会奨励賞，1998 年日本音響学会第 15 回粟屋潔学術奨励賞，2001 年電子情報通信学会第 56 回論文賞．電子情報通信学会，日本音響学会，日本教育工学会，IEEE 各会員．

間瀬　健二（正会員）

1979 年名古屋大学工学部電気学科卒業．1981 年同大学院工学研究科情報工学専攻修士課程修了．同年日本電信電話公社（現 NTT）入社．1988〜1989 年米国 MIT メディア研究所客員研究員．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．人工知能学会 1999 年度論文賞．IEEE，ACM，電子情報通信学会，VR 学会，画像電子学会各会員．博士（工学）．

萩田　紀博（正会員）

1978 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．同年電電公社（現 NTT）武蔵野電気通信研究所入所．文字認識，画像認識等の研究に従事．NTT 基礎研究所などを経て，現在，ATR 知能ロボティクス研究所長，博士（工学），IEEE，電子情報通信学会，人工知能学会，日本ロボット学会各会員．