3D-03

ポスター対話における会話参与者の振舞いの定量化

吉本 廣雅　　中村 裕一　　京都大学 学術情報メディアセンター

1. はじめに

学会などで一般的に行われているポスター発表では，ポスターを介して発表者と複数の聴衆が自由に討論を行う多人数会話の場が形成されている．本稿はこのような**ポスター対話**の場を対象とし，参与者の振舞いを定量的に比較・分析する手法を提案する．

本提案法は画像処理により注視点を３次元計測し，その時空間分布をクラスタリングすることで振舞いの時空間パターンを自動抽出するものである．その特徴としては，体格や髪型のような身体的な個人差，会話中の立ち位置や姿勢の動的変動を３次元で正しく計測しつつそれらの影響を除外する正規化を行うことで，振舞いのパターンを出力する点にある．そのためパターンに対する単純な演算で参与者間の振舞いが定量的に比較できる利点がある．

本稿では，実験結果と共に，1) パターンの時間分布から話題の転換点が推定できること，2) パターンの空間分布からポスターの意味的な配置構造が推定できること，3) 聴衆間のパターンの類似度から議論の活発さが計算できること，の３点を報告する．

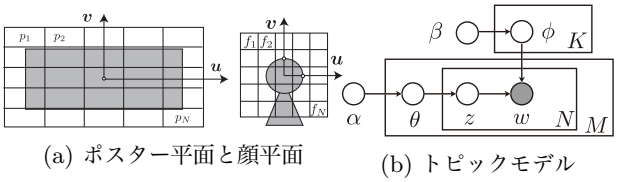
2. 振舞いの定量化

頷きや視線配分のような会話中の振舞いは非言語行動とも呼ばれ，会話状況の認識や理解への大きな手がかりとなる．例えば，河原らは次発話者の予測や興味度が推定できること [1]，中田らは会話の構造が抽出できること [2] を報告している．

これらの手法は，非言語行動を手や頭部の方向や位置などの物理量として観測し，観測値から参与者の興味や会話の文脈のような直接観測できない心的状態を推定する処理を基本的枠組みとしている．ここで問題が二つある．一つは非言語行動と心的状態の関係は個人差や話題の影響を受ける点である．二つ目は非言語行動と観測値の関係も個人差や状況に応じて変動する点である．例えば画像から顔の正面方向の角度を推定する場合には，顔の形・髪型・メガネの有無などの個人差と立ち位置や姿勢のような状況を考慮しなければならない．

個人差や状況の多様性を考えると，実際の会話を観測した生の対話データから興味や関心のような心的状態を正確に計算することはまだ困難であると言える．そこで本研究は以下のアプローチを提案する．まず興味や関心に対応する変数は陽に定義せず，興味や関心や話題には N_m 個の**モード**があることだけを仮定する．そしてモードに応じて，振舞いが表出する**振舞いの生成モデル**を考える．これにより，興味や話題の変遷をモードの時系列として，個人差を各モードに対する振舞いの表出パターンとして算出することを考える．

Quantitative measurement of participant's behaviors in poster conversations: *Hiromasa YOSHIMOTO (Kyoto Univ.) and Yuichi NAKAMURA (Kyoto Univ.).*



(a) ポスター平面と顔平面　　(b) トピックモデル

図 1: 提案法の概要

工学的応用を考えると，モードや表出パターンを定量的に値として算出できるだけで，それは場面の転換点の候補や，個人差を比較する指標としての利用が十分に期待できる．また対話分析の研究という面でも，これはアノテーション作業の支援や，本来の目的である対話の意味や参与者の心的状態に関する新しい知見の発見に寄与すると期待できる．

3. 提案法の概要

本稿では，振舞いの観測値として注視点の３次元位置を考え，以下，注視点の時空間分布からモードと注視点分布を算出する方法について詳述する．

3.1 頭部運動の３次元計測

自然な会話の収録と分析を実現するために本研究は RGB-D カメラのみを用いた非侵襲な計測方法を考える．また先行研究の知見に基づいて視線方向はおおよそ頭部正面方向と一致すると仮定する [1]．そして個人差の影響を受けない頭部運動計測を実現するために，動画像から参与者の頭部３次元形状と３次元運動の同時推定を行う．これは，我々が文献 [3] で提案した手法を応用したものであり，顔つきや髪型のような個人差を頭部形状として３次元で正確に計測することで，正面顔だけでなく横顔からも頭部正面方向を頑健に３次元で推定できる方法となっている．

3.2 注視点の推定

ポスターに対応する平面と，発表者の頭部位置に仮想的配置した平面の二平面を考え，注視点の３次元位置はこれら平面と顔向きに対応する直線の交点として算出する．さらに図 1a に示すように各平面を格子状に N 分割することで注視点を離散化し，$\boldsymbol{p}_i$ と $\boldsymbol{f}_i$ の記号を割り当てる．これにより注視点を記号列として表現する．

3.3 モードと注視点分布の計算

この処理は Latent Dirichlet Allocation(LDA) [4] を利用する．図 1b に LDA のトピックモデルを示す．本研究におけるモードは LDA におけるトピック z に，注視点の記号列はワード w に対応し，全体ではモード z に応じて注視点 w が生成される生成モデルとなる．また $\boldsymbol{\phi}$ はモード z に対する注視点 w の生起確率となる行列であり，これは個人差 $\boldsymbol{\beta}$ をハイパーパラメータとした注視点分布のパターンとなる．ここで Gibbs sampling を行うと，記号列から z と $\boldsymbol{\phi}$ の組が逆算できる [4]．

Copyright ©2015 Information Processing Society of Japan.
All Rights Reserved.

4. 実験と考察

3 名の被験者 (A，B，C) を一組とし，A に発表者役，B,C に聴衆役を割り当て，30 分を目安に自由に議論を行うよう指示した．3 名の振舞いは 2 台の Kinect センサを用いて RGB-D 動画像として記録し，検証用に磁気式姿勢センサを用いて姿勢の真値を，マイクを用いて音声データを記録した．収録は計 8 セッション行い，のべ 24 名，720 分のデータを用意した．

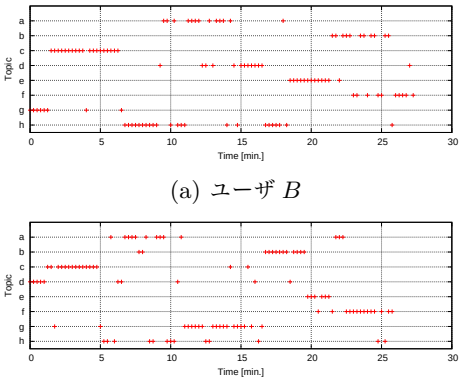
実験ではまず画像から頭部運動を計算した．運動は 1/30 秒間隔で計測でき，その誤差は並進成分で 12mm 程度，回転成分で 5 度程度であった．次に注視点の計算を行い，記号列を求めた．ポスター平面と顔平面の分割数はそれぞれ 100 とした．

次にモードと注視点分布を計算した．本実験ではモードは計 8 個あるとし，記号列は 15 秒間隔で分割しそれを一単位とした．また Gibbs sampling アルゴリズムの反復回数は予備実験の結果から 10,000 回とした．

計算結果の一部を図 2，3 に示す．図 2b は聴衆 B のモードの時間推移を表し，モードが g，c，h と順に変遷していることが判る．図 3b はモード毎の注視点分布をポスター平面・顔平面上に可視化したものである．モード g ではポスター右上，c では右下，h では中央上を注視していることが読み取れる．これは，実験で用いたポスターは 6 枚のスライドを右上，右下，中央上，中央下と配置しており，その順番で発表者が説明を進めたことと対応している．このように提案法は，ポスターの意味的なレイアウトや，その説明順序を数値として算出できることが確認された．

次に振舞いの類似度として注視点分布の相関を計算した．図 4a はセッション 1 での類似度であり，例えば B のモード g(以下 B-g と表記) は C-d と一体一対応している．これは会話冒頭で B，C の両者が聴き手のモードになっていた事を意味している．一方 B-h には C-a と C-h の二つが対応している．これは C-a と C-h の注視点分布を比較すると判るように，C-a はポスター側を注視した聴き手のモード，C-h はポスターと A の両者を注視する話し手のモードに対応している．一方 B-h はポスターを注視した聴き手のモードであるため，B-h と C-a の類似度が高く，B-h と C-h の類似度が低くなっている．データを解析した結果，B，C の両者が聴き手である場合は類似度が一体一対応し，B か C の片方が話し手に変わった場合は類似度は一体多対応となる．また B と C が頻繁に話し手となる場合は多対多の対応となることが明らかになった．まとめると，図 4a に示すような一対一対応のセッションでは，B と C は基本的に聴き手となっており，図 4b のようなセッションでは C よりも B の方が話し手となる場合が多く，図 4c に示すようなセッションでは C も B も頻繁に話し手となっている．このように分布の類似度から，会話の状況や議論の活発さのような物を定量的に比較できることが示された．

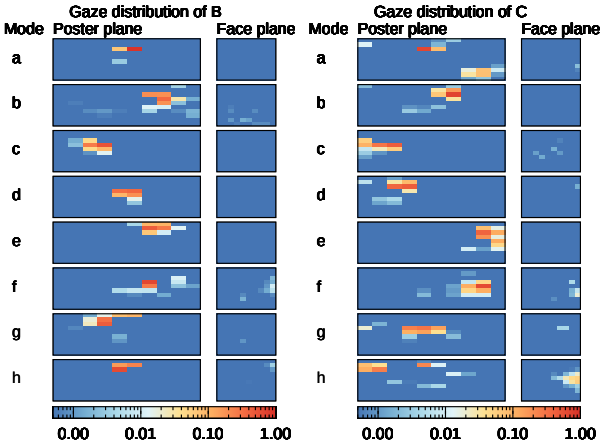
謝辞: 本研究の一部は，独立行政法人科学技術振興機構 (JST) 戦略的創造研究推進事業 (CREST)「マルチモーダルな場の認識に基づくセミナー・会議の多層的支援環境」の助成を受けて行った．



(a) ユーザ B

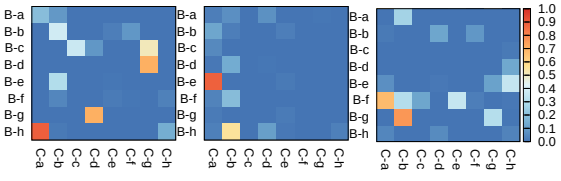
(b) ユーザ C

図 2: セッション 1 におけるモードの推移



(a) ユーザ B (b) ユーザ C

図 3: セッション 1 におけるモード毎の注視点分布



(a) セッション 1 (b) セッション 2 (c) セッション 5

図 4: 聴衆 B-C 間の注視点分布の類似度

参考文献

[1] 河原達也. スマートポスターボード：ポスター会話のマルチモーダルなセンシングと認識. 信学技報. SP, 音声, Vol. 112, No. 141, pp. 7–12, jul 2012.

[2] 中田篤志, 角康之, 西田豊明. 非言語行動の出現パターンによる会話構造抽出. 信学論 D, Vol. 94, No. 1, pp. 113–1232, 2011.

[3] 吉本廣雅, 中村祐一. 未知剛体の形状と姿勢の実時間同時推定のための cubistic 表現. 信学論 D, Vol. 97, No. 8, pp. 1218–1227, Aug. 2014.

[4] Thomas L. Griffiths and Mark Steyvers. Finding scientific topics. *Proceedings of the National Academy of Sciences*, Vol. 101, No. suppl 1, pp. 5228–5235, 2004.

Copyright ©2015 Information Processing Society of Japan.
All Rights Reserved.