

実世界インタラクション理解のための マルチモーダルデータ分析環境の構築

來嶋宏幸[†] 坊農真弓^{†‡} 角康之[†] 西田豊明[†]

京都大学情報学研究科知能情報学専攻[†] 日本学術振興会[‡]

1. はじめに

近年、日常的なミーティングにおける会話の音声や映像を収録し、後からサマリなどに再利用するための研究が盛んに行われている[1]. 我々はこのような会話の中に含まれる知識の伝達や創造の強化・支援に向けて、より効果的かつ汎用性の高い会話再利用のための技術の構築を目指している. そのためには人と人, 人とモノ, 人と環境の間にあるインタラクションのプロトコルを理解し, それらを機械可読な形にする必要がある. そのためのアプローチとして, 我々はインタラクションの場を複数センサを用いて記録する環境として **IMADE** ルーム[2]を構築し, そこから得られるデータにインデクスを与え検索可能な形で蓄積したコーパスの構築を試みた[3]. インタラクション・コーパスを用いた研究フレームワークの概観を図1に示す. コーパス構築に向けて, 収録したセンサの生データを蓄積したレイヤー, 発話・注視・指差しといった基本的な非言語情報からなるプリミティブなレイヤー, プリミティブなデータの分析から見出したメタ的なインデクスの付与されたレイヤーといった複数レイヤーの階層構造で実世界インタラクションの理解を目指す.

本稿では, インタラクション・コーパス構築に向けて実世界会話のデータから社会的インタラクションのパターンを発見・評価するために試作した **iCorpusStudio** について述べる.

2. システムの構成と実装

従来インタラクションデータの分析に用いられて来たシステム[4]はアノテーションの付与に特化しており, 付与されたアノテーションの統計処理や複数データ間の演算といった処理は別々に行ってきた. 前節で述べたようなインタラクションを階層的に分析したり, その自動化

Development of an Environment for Multimodal Interaction Analysis, Hiroyuki Kijima[†], Mayumi Bono[‡], Yasuyuki Sumi[†], Toyooki Nishida[†], [†]Graduate School of Informatics, Kyoto University, [‡]JSPS Research Fellow

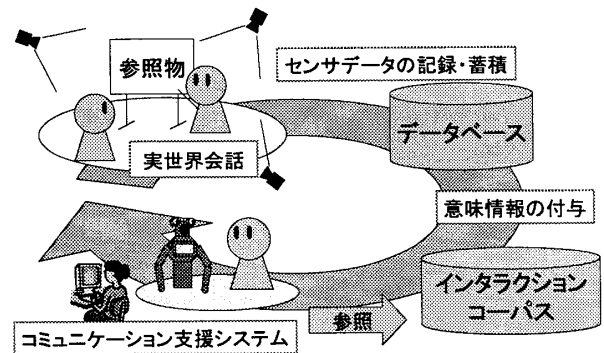


図1 インタラクション・コーパスを用いた研究フレームワーク

処理の試作を行うには, より統合的にデータ処理や分析を支援する環境が必要となった.

今回試作した **iCorpusStudio** はインタラクションのデータ分析基盤として, (a)アノテーション等データの可視化(2.1 節), (b)データ間演算や検索等の分析支援(2.2 節), (c)ユーザによるカスタマイズを前提としたプラグイン機構(2.3 節)を提供する. これにより, ユーザのニーズに応じたデータ分析や様々なセンサデータ, アノテーションに対応することが可能となる.

2.1. データの閲覧とアノテーションの付与

図2にシステムの概観を示す. インタラクションの場を捉えた複数映像や各音声の波形, モーションキャプチャの3次元座標を3D空間にマッピングしたもの, 及び映像に対応したアノテーションを時系列表示することで各データの対応関係を容易に把握することが可能である. また, 対応する映像を閲覧しながらユーザがアノテーションを付与することも可能である.

2.2. インタラクションデータ分析支援

インタラクションデータを用いた分析を支援するための機能として,

- インタラクションの共起の抽出
- 各アノテーションの頻度情報算出と提示

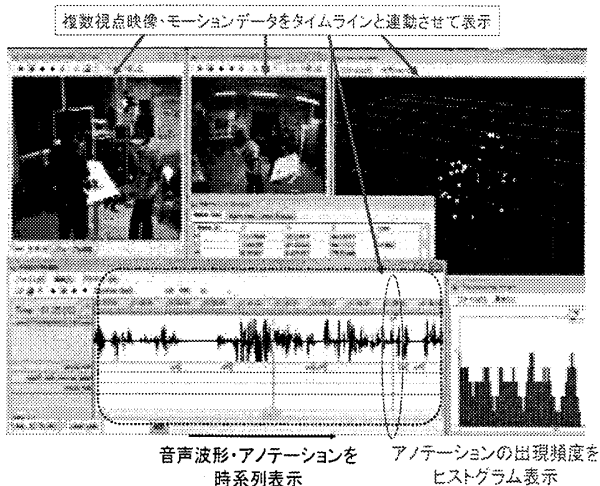


図2 iCorpusStudio

● インタラクションパターンによる類似シーン検索

などを分析支援の基本プラグインとして提供する。

大量にある会話データ、主に映像データの中から今自分が着目しているシーンと似たシーンを人手で見つけ出すには会話データの量に比例して莫大なコストがかかる。そのような状況を支援するための機能として、アノテーションパターンによる類似シーンの検索を提供している(図3)。分析者が指定した検索対象のシーンの開始時間と終了時間の間にあるアノテーションについて、一定の割合で時間方向に分割して系列データとして扱う。対象の系列と同じ時間長の系列を抽出してDPマッチングを行う処理をデータ全体の最初から最後までスライドさせることで各シーンとの類似度(複数アノテーションに対応した多次元のユークリッド距離)を算出し、類似度の高いシーンから順に提示する。

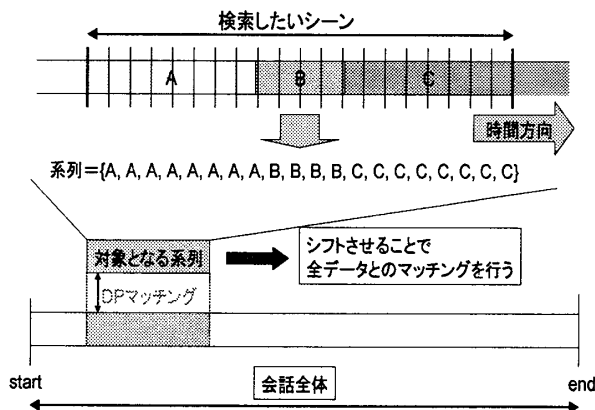


図3 類似シーン検索の処理

2.3. プラグインによる拡張

実世界インタラクションの計測と理解は、センサの発展に伴い常に変化し続ける。従って、今後会話の場を新たなモダリティを用いて記録したいというニーズがあれば、新たなセンサを導入することになるかもしれない。一方で、センサ以外のデータ収録システムを用いたインデックスの付与の可能性も模索している。このように分析対象となるデータが変更されようと、プラグインを開発することで柔軟に対応し、従来のデータと並列に扱ったり、新たな処理機能を実装出来るようにしている。

3. まとめと今後の課題

インタラクション・コーパス構築に向けて、蓄積されたデータの処理及び分析支援を行う環境である iCorpusStudio について紹介した。iCorpusStudio 上では、会話中のマルチモーダルなデータや他システムのデータを構造的に捉え、処理することが出来る。これにより、我々がインタラクション・コーパスに盛り込みたいインタラクション上有意義なインデックスを自動検出する方法の研究が加速されると考えている。今回実装したインタラクションパターンによるマッチングを応用し、将来的にはパターン認識技術を用いた自動アノテーションを実現したい。まずはその足がかりとして、作業者にアノテーションの推薦を行えるような枠組みを試作していければと考えている。

謝辞

本研究は、文部科学省科学研究費補助金「情報爆発時代に向けた新しい IT 基盤技術の研究」の一貫で実施された。IMADE ワーキンググループの関係者の皆様による日頃の支援と議論に感謝します。

参考文献

[1] S.Renals. AMI:Augmented multiparty interaction. In Proc. NIST Meeting Transcription Workshop, Montreal, 2004.
 [2] 松山隆司, 西田豊明, 國吉康夫. 情報爆発時代におけるヒューマンコミュニケーション基盤. 人工知能学会誌, Vol.22, No.2, pp.229-234, 2006
 [3] 来嶋宏幸, 坊農真弓, 角康之, 西田豊明. マルチモーダルインタラクション分析のためのコーパス環境構築. 情報処理学会研究報告(ヒューマンコンピュータインタラクション), Vol.HCI125, 2007
 [4] Michael Kipp. ANVIL : A generic annotation tool for multimodal dialogue. In Proc. Eurospeech 2001, pp.1367-1370, 2007