

参加者の非言語インタラクションと協調的アノテーションに基づく ミーティングキャプチャシステム

勝木 弘[†] 角 康之[†]

本稿では対面議論型ミーティングから聞き手反応に着目することで、参加者の興味が深いシーンや、盛り上がっているシーンなど、再利用性の高いシーンを取り出し利用するミーティングキャプチャシステムを提案する。そのために発話や視線と言った聞き手の非言語行動に加えて、聞き手による写真撮影や書き込みと言った協調的アノテーション、またそれらを基にした参加者間のインタラクションを利用する。実際に収録した会話データの中で、これらの行動が活発に行われていた箇所を確認したところ、活発な議論が行われ聞き手の疑問が解消されたシーンが見られた

Meeting Capture System based on Participants' Nonverbal Interactions and Collaborative Annotation

HIROSHI KATSUKI[†] YASUYUKI SUMI[†]

This paper proposes a meeting capture system based on participants' behavior so that we can extract the useful conversation scene. We use participant's nonverbal interactions and collaborative annotation. Collaborative annotation is shared data in participants such as taking a photo, writing in a slide or a picture. We observed the acquired conversation data where participants' nonverbal interactions and collaborative annotation were seen concentratively, and found the scene where the participants were discussing actively and one participant's question were resolved.

1. はじめに

ミーティングはグループ内の知識交流を図ったり、議論によって新たな知の創造が進んだりといった有益な活動が行われる場である。そのためミーティングの様子を記録・構造化することで知的活動支援を行う試みは今まで数多く行われている。

長尾ら[4]は対面議論型ミーティングにおいて、詳細な議事録と参加者のつけるメタ情報を基にミーティングの構造化を行っている。AMI/AMIDA プロジェクト[5]では発話者検出や音声認識、ジェスチャ認識などによってミーティングのコーパスやブラウジングシステムを作っている。

これらの研究が話し手の発話内容や振る舞いに着目しているのに対して、本研究では参加者の非言語インタラクションと聞き手による写真撮影、書き込みと言った協調的アノテーションを利用して、聞き手の反応に着目した構造化を行う。それによって会話を検索性・再利用性の高い知識コンテンツ化し、それらを提示してやることでグループ内の知識活動の活性化を行うミーティング支援システムを提案する。

2. 知識流通を促進する支援システム

本研究で提案するミーティング支援システムの流れを図1に示す。カメラやセンサを用いたミーティング記録、会話シーン認識とインデクス付与による知識コンテンツ化、実世界への提示・支援を行い、グループ内の知識活動の活性化を目指す。

2.1 対象とするミーティング

本研究では、5,6人前後の参加者が、スライド、ビデオ、配布資料などを参照しながら議論を行う対面議論型ミーティングを対象としている。例えばグループ内の進捗報告会や、文献発表会などである。参加者は手元に Photochat 端末[2]を持ち、写真やスライド、それに対するメモを共有する。また、非言語行動を取得するためのセンサを装着する。

2.2 利用シナリオ

本研究によって実現する知識活動の活性化の例を示す。2.1節で述べたようなミーティングを繰り返し記録することで多くの知識コンテンツが蓄えられる。これらの知識コンテンツには発表者や参加者の情報、誰が議論をしているか、聞き手の反応はどうかといった会話の様子、スライドやメモなどから得たキーワードなどのインデクスがある。これらを利用して状況に応じたシーンを提示してやることで、話題提供や議論の

[†] 京都大学情報学研究科

Graduation School of Informatics, Kyoto University

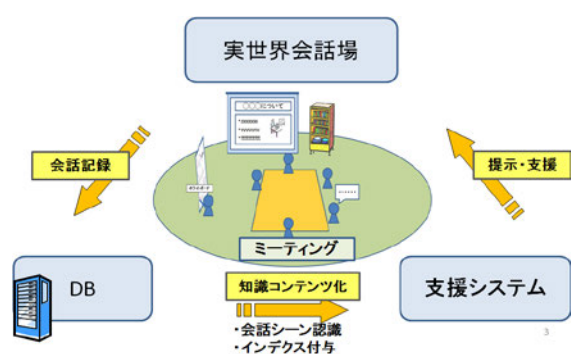


図1 ミーティング支援システムの流れ

活性化、知識の共有などが可能になると考えられる。

例えばミーティングが行われている際に、発表に対して聞き手の一人が「～の研究と近いね」などの書き込みを行うとする。既にその研究、もしくは関連研究に関する知識コンテンツがある場合、システムはそのシーンのスナップショットと閲覧へのリンク情報を Photochat に投稿する。過去の関連シーンを提示することで、ミーティング中の議論の活性化を見込むことができる。また部屋の中に情報端末を設置し、その付近に人が近づいた際に、彼らが所属するグループが行ったミーティングの場面を提示することで、話題提供を行い、議論の誘発を行う。

2.3 再利用性の高い知識コンテンツ化

久保田ら[1]によると再利用性の高いコンテンツを得るには参加者の意図が重要だとされる。本研究では聞き手の興味・関心の強いシーンを獲得することで、盛り上がっているシーンや重要な議論が行われていたシーンを取り出す。そのために発言や視線といった非言語行動、聞き手の行う写真撮影、書き込みと言った協調的アノテーションを利用する。

2.3.1 非言語行動

非言語行動は会話の状況や参加者の意図を掴む手がかりになると言われ、多くの会話分析研究で着目。実際知らない言語で行われている会話でも、非言語行動パターンから、会話がどういう状況にあるのか、参加者の立場の違い、盛り上がりの様子、参加者の心理状況がある程度察することは可能である。そのため、非言語行動から機械的に会話状況や参加者の意図を推測するという試みもいくつか行われている。

本研究では非言語行動として視線、うなずき、発話の3種類を用いる。これらは人数会話において重要な行動とされており、会話分析や自動認識でもよく使われているものである。

2.3.2 協調的アノテーション

近年会議において議論を活性化するためにテキスト・チャット・システムを併用する手法が広まりつつある。チャットを利用することで、聞き手は発表者などの発言を阻害することなく自分の意見を伝えることができる。また参加者同士による議論が行われたり、場合によっては聞き手と発表者のやり取りなどが出てきたりもする。小林ら[6]によると、こうしたチャットの内容は実世界における発表やスライドに関連するものが多く、発言の重要性を判断する基準になりえるとされる。

本研究では普通のテキストチャットではなく、Photochat[2]を利用する。Photochat はカメラ付き小型タブレット PC 上で動作する写真撮影とそれに対して書き込みを 실시간 共有出来るシステムである。これを利用することで、参加者は自分の興味のあるシーンを撮影したり、写真やスライドに対してメモや質問などの書き込みを行ったりできる。これらの情報を共有していくことでチャットと同じように他の参加者間とのやり取りも可能になる。また、スライドに対する書き込みを見ることで、参加者がどの内容に対してどういう風に感じているのかを一目で分かることができる。こうしたミーティングに対して協調的につけられていく行動を協調的アノテーションと呼ぶ。

2.4 収録環境

本研究におけるミーティング収録環境（図 2）について説明する。

2.4.1 カメラレイシステム

ミーティングの映像を収録する際、重要なシーンが影に隠れてしまわないよう、ネットワークカメラアレイシステムを利用して複数視点映像を撮影する。

2.4.2 加速度センサ

参加者のうなずきを取得するために、参加者は小型無線化速度センサ WAA-006 を前頭部に装着する。これは X, Y, Z の三軸の加速度情報を Bluetooth で PC に送信するものである。

2.4.3 モーションキャプチャ

参加者の頭部の動きや方向を取得するために Motion Analysis 社製の再帰反射性モーションキャプチャ MAC3D を使用する。参加者は4点のマーカがついた帽子を頭に被る。

2.4.4 指向性マイク

参加者の発話を記録するために、頭に指向性マイクを装着する。マイクはワイヤレスが4台、有線が4台であるが、有線の場合でも十分な長さがあるため参加者の動きを制限することはない。



図2 ミーティング収録環境

2.4.5 Photochat

協調的アノテーションを取得するために、参加者間で写真、書き込みの共有が可能なシステム Photochat を利用する。

3. 非言語行動及び協調的アノテーションの獲得方法

収録したデータから参加者の非言語行動、協調的アノテーションを作成する方法について述べる。非言語行動に関しては、多人数インタラクション分析環境 iCorpusStudio[3]を利用して作成する。

3.1 発話

指向性マイクで収録した音声データから以下のようにして発話区間を取り出す。

1. 無音区間から閾値を定め、それ以上のパワーの区間を取り出す
2. 細かい外れ音声を除去する
3. 300msec 以下の無音区間を消去する
4. 50msec 以下の有音区間を消去する

3.2 視線

頭部につけたモーションキャプチャから頭部方向ベクトルを求め、対象物（人、スクリーン）を見ている区間を取り出す。

1. 後頭部と前頭部のマーカーを結んだ直線を下向きに 40 度回転させ、頭部方向ベクトルとする
2. 頭部方向ベクトルが鉛直方向となす角度から、手元を見ている区間、顔を上げている区間を取得
3. 顔を上げている区間のうち、頭部方向ベクトルと対象物のなす角度からそれぞれの対象物を見ている区間を取得
4. それぞれについて細かい区間を除去

なお、各角度の閾値に関しては、帽子の被る角度の違い、対象物との相対位置の違いなどで変わってくるため、今回は手動で調整を行った。

3.3 写真閲覧

photochat のログ及び、手元を見ている区間から、参加者がある写真を見ている区間を取り出す。

1. photochat ログより画像を閲覧していた区間を取得し、手元を見ている区間との AND をとる
2. 細かい区間を除去する。

3.4 協調的アノテーション

photochat のログより、協調的アノテーションとしての書き込み、写真撮影が行われた区間を取り出す。

3.4.1 撮影

1. photochat ログより、カメラモードを立ち上げてから撮影が終わるまでの区間を取り出す。また撮影を行わずカメラモードを終了した場合も取り出す。
2. 撮影者以外の参加者による書き込みのない写真、もしくは撮影者の書き込みに対する閲覧のない写真を取り除く。

3.4.2 書き込み

1. photochat から書き込みに関するログをとる。
2. ログは 1 ストロークごとに区切れているため、4 秒以内に同じ写真へのストロークがある場合は同一の書き込みと見なす。
3. 既に書き込みされている写真への書き込み、または書き込み後他の参加者から 2 秒以上閲覧された書き込み以外を取り除く。

4. 収録

非言語行動及び協調的アノテーションを用いて会話シーン認識を行うために、参加者 5 人による会話を収録した。発表は筆者が参加してきた学会の報告について行った。使用したスライドは 28 枚で、参加者でスライドを共有するために、新しいスライドに移る度に発表者が手動で photochat に投稿した、収録時間は 1 時間 14 分であった。協調的アノテーションとしての撮影は 7 枚、書き込みは 86 回行われた。

5. 考察

収録したデータに対して、発話・視線・写真閲覧・協調的アノテーションを獲得した。また iCorpusStudio を用いて以下のような参加者間のインタラクション行動を求めた。

- 写真の同時閲覧
- 発表者と聞き手の相互注視
- 発表者に対する聞き手の視線集中
- スクリーンに対する聞き手の視線集中

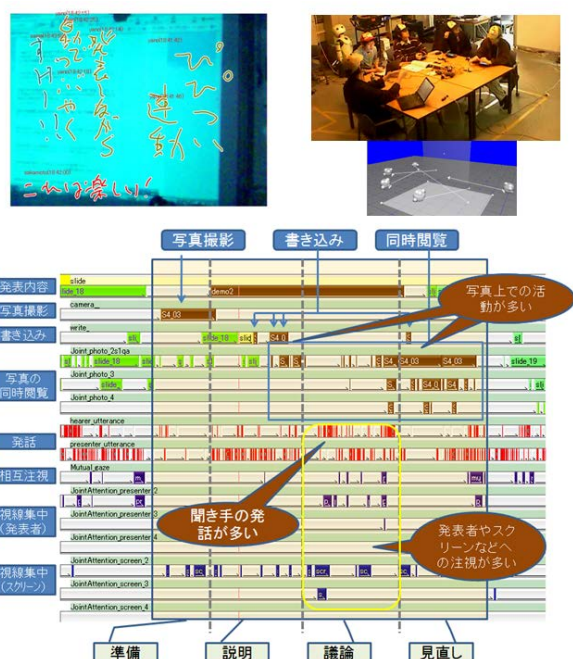


図3 PhotoChat (左上) 会話の様子 (右上)

図4 参加者の振る舞い (下)

収録した会話データについて、これらの非言語行動、協調的アノテーション、参加者間のインタラクションが活発に行われていた場所を見たと、図 3, 4 のようなシーンが観測できた。

このシーンはある発表についての説明の中で、参考資料として外部資料を参照していた時のシーンである。聞き手の一人が撮影した写真はあまり鮮明でなく字は読み取れるものではなかったが、その上に発表者の説明や感想などが書かれていった。書き込み以外にも、聞き手の発話や視線集中、写真の同時閲覧なども多く観測され、活発な議論が行われていたことが分かる。また、発表者と相互注視しながら発話していた聞き手が最後に「すげー!!」と言う書き込みをしていたことから聞き手の疑問が解消された重要なシーンであったことが分かる。実際に映像や音声で聞くと聞き手の質問と発表者の応答が繰り返し、その中で「へー」や「なるほど」、「そういう事だったのか」などと言った発話が行われていた。

協調的アノテーションの例としては以下のような書き込みが見られた。

- 発表者の説明を聞いてスライドの内容を補足
- スライドにない説明の付け足し
- 罫線や囲いなどによるスライドや書き込みの強調
- 疑問や感想など
- 写真についての説明・メモ
- 書き込みに対する応答

今後、これらの書き込みの前後で何が行われているかを分析していくことで、聞き手が興味を持っているシーンを取り出すモデルを構築していく。

6. おわりに

本研究では対面議論型ミーティングから聞き手反応に着目することで、参加者の興味が深いシーンや、盛り上がっているシーンなど、再利用性の高いシーンを取り出し利用するミーティングキャプチャシステムを提案した。そのために発話や視線と言った聞き手の非言語行動に加えて、聞き手による写真撮影や書き込みと言った協調的アノテーション、またそれらを基にした参加者間のインタラクションを利用する。実際に収録した会話データに対して、これらの行動が活発に行われていた個所を見たと、活発な議論が行われ聞き手の疑問が解消されたシーンが見られた。

今後、これらの書き込みの前後で何が行われているかを分析していくことで、聞き手が興味を持っているシーンを取り出すモデルを構築していく。また今回使わなかったうなずきや、手作業で閾値を定めていた視線に関して自動で獲得できるようにする。取り出したシーンにインデックスを付与し、その後のミーティングや立ち話に対する支援に利用できるようにする。

謝辞 本研究は科研費・基盤研究(B)No.22300042 および特定領域研究 No.18049045 の助成を受けた。

参 考 文 献

- 1) 久保田 秀和, 齊藤 憲, 角 康之, 西田 豊明: 会話量子化器を用いた会話場面の記録, 情報処理学会論文誌, Vol. 48, No. 12, pp. 3703-3714 (2007).
- 2) 角 康之, 伊藤 惇, 西田 豊明: PhotoChat: 写真と書き込みの共有によるコミュニケーション支援システム, 情報処理学会論文誌, Vol.49, No.6, pp.1993-2003, (2008).
- 3) Yasuyuki Sumi, Masaharu Yano, Toyooki Nishida: Analysis Environment of Conversational Structure with Nonverbal Multimodal Data, in Proc. ICMI2010, (2010).
- 4) Nagao, K., Kaji, K., Yamamoto, D. and Tomobe, H.. Discussion Mining: Annotation-Based Knowledge Discovery from Real World Activities. in Proc. PCM 2004, pp.522-531, (2004).
- 5) S.Renals, T.Hain, and H.Bourlard, Recognition and Understanding of Meetings The AMI and AMIDA Projects, in Proc. ASRU, pp.238-247, (2007)
- 6) 小林 智也, 西本 一志: 超メディア参照機能を有するチャット併用会議のためのチャットシステム, WISS2009 論文集, pp.101-106, (2009)