

複数地点間常時遠隔会話のための プライバシーを考慮した対話状況伝達手法

河島 健司^{1,a)} 吉野 孝^{1,b)}

概要: 近年、容易にビデオチャットを行える環境が整ってきた。また、通信速度の向上により、今までは技術的に難しかった通信が誰でも行えるようになった。例として、複数の遠隔地間で、対話環境を半永久的に維持する遠隔会話が可能になってきている。また、上記の会話形態には、(1) 会話の映像と音声をシステム利用者全員に対して公開する状態と、(2) 会話の映像と音声をシステム利用者全員に対して非公開にする状態の2つがある。しかし、遠隔地と半永続的にビデオチャットをする場合、プライバシーの問題がある。(1) 誰が誰と何を話しているかが分かり、気軽にインフォーマルな会話をするができない。(2) 対話先の会話の様子や状況などが不明確であるため、遠隔地の人物に用事がある場合、会話を遮っても問題ないと判断することが難しい。そこで、プライバシーを考慮に入れた状況下でも、対話状況を提示することを実現する第3の情報提示環境の提案をする。本稿では、システムの概要と、実験の結果と考察について述べる。

Conversation Situation Conveying Method Considering Privacy for Multipoint Always-on Remote Connection

1. はじめに

近年、Skypeなどの無料でビデオチャットができるソフトウェアやスマートフォンの普及により、容易にビデオチャットを行える環境が整ってきた。これにより、様々な形態の遠隔会話手法が研究されてきた。

その一例として、対話環境を半永続的に維持して遠隔会話を行う研究がなされている。家族間で対話環境を維持し遠隔会話を行う場合、交換される間接的情報の種類によらず、偶発的に会話が発生することが分かっている [1]。また、子供や独居高齢者が、疎外感や孤独感などの感情を払拭する効果がある [2] ということが分かっている。同様に、複数の遠隔地間で対話環境を半永続的に維持する遠隔会話の研究がなされている。Portholes[3] では、複数地点間のリモートワークを可能にすることにより、コミュニティ意

識の構築に役立つことが分かっている。VideoWindow[4] では、分散拠点で24時間ビデオ通話をすることにより、インフォーマルコミュニケーションの実現を試みた。

以上の常時複数地点遠隔会話の先行研究では、大きく分けて2つの会話形態が用いられていた。1つ目は、会話の映像と音声をシステム利用者全員に対して公開する状態である。2つ目は、着手している仕事や会話内容を知られたくない時に、会話の映像と音声をシステム利用者全員に対して非公開にする状態である。しかし、遠隔地と半永続的にビデオチャットをする場合、プライバシーの問題を考慮する必要がある。システム利用者全員に対して情報を完全に公開する状態だと、対話環境をそのまま遠隔接続しているため、誰が誰と何を話しているかが分かり、個人間の会話をシステム利用者全員が聞くことができるようになっていく。そのため、気軽にインフォーマルな会話をするができないという問題がある。逆に、プライバシー保護のため、システム利用者全員に対して情報を完全に非公開にしてしまうと、対話先の会話の様子や状況などのアウェアネスが不明確であるため、遠隔地の人物に用事がある場合でも、会話を遮っても問題ないと判断することが難しいという問題がある。

¹ 和歌山大学

Wakayama University, Wakayama 640-8510, Japan

a) s206075@wakayama-u.ac.jp

b) yoshino@wakayama-u.ac.jp

*1 総務省:平成29年度版 情報通信白書
<http://www.soumu.go.jp/johotsusintokei/whitepaper/ja/h29/html/nc111110.html> (参照 2019.1.5)

また、複数地点間で常時遠隔会話をする際、情報を完全に公開するか、非公開にするかという手法のみであったため、システム利用者全員に対して同じ情報提示しか行えなかった。この場合、特定の人物に伝えたい情報でも、他の人物に聞かれたくない、もしくは聞かれてはならない場合、個別で話すしかないという常時接続のメリットを損なっていた。ゆえに、すべてを公開する状態とすべてを非公開する間の状態の必要性和各システム利用者に合わせて情報量を操作できる必要があると考えられる。

そこで、最終的には各遠隔地点のプライバシーレベルに合った会話環境を提示することを視野にいれ、プライバシーを考慮に入れた状況下でも、対話状況を提示することを実現する第3の情報提示環境の提案をする。これにより、プライバシーを部分的に保護した状況下でも、各遠隔地点に合わせて状況を知らせることができ、システム利用者が他のシステム利用者に話しかけたり、会話に参加したりすることができると思われる。

本研究では、プライバシーを考慮した状態での会話状況提示手法の1例として、対話内容から抽出した対話のトピックを表す単語を表示することで、プライバシーを保護した状態で情報提示をすることができるかどうかを検証した。

2. 関連研究

2.1 プライバシー保護

2地点間の常時遠隔会話におけるプライバシーの研究では、先行研究として、半田らのなめらカーテン [5] がある。カーテンに見立てた布の開閉を利用し、開けた時は直接的な実際に会話している音声を、閉めた時は間接的でアンビエントな雰囲気が伝わる音声を提示することができ、接続地点のプライバシー状況に合わせて柔軟に遠隔会話を続けることができるようになっている。

また、ビデオチャットを複数地点と接続しながら行う研究がなされている。Web会議やテレビ会議のような、参加者が疑似的に1か所に遠隔接続し会議を行い、移動にかかる費用を抑えつつ、メールや電話よりも情報量の多いやり取りを可能にするとして注目されてる。Portholes[3]やVideoWindow[4]、CRUISER[6]では、複数地点間の遠隔接続により、複数人を1つのブラウザに表示して、誰がどのような状況であるかをソーシャルブラウジングできるようにしている。他のシステム利用者から見られるのは、自分が他の参加者を見ている時だけで、常時接続するうえで問題となっているプライバシーを守ることのできるシステム形態をとっている。しかし、システムに接続しているすべてのユーザが、同様に接続している他のユーザ同士の会話や様子を閲覧することができる形式であるため、個別で話したい会話であっても、すべてのユーザに共有してしまうという問題がある。

2.2 会話への割り込み

電話中に第三者が電話をかけてきたことを知らせるシステムとしてコールウェイティングという機能が存在する。しかし、重要な話をしている時にシステムが鳴ると焦ってしまったり、誰からの呼び出しなのか気になってしまったりといった問題がある。

会話時の割り込み拒否が存在し、[7][8] 会話に割り込みをしたい時には、会話が終わりつつあることや、会話が盛り上がり段階であることや、話者同士が笑顔で話しかけても問題なさそうというような、会話の雰囲気をある程度把握し、問題のないタイミングを見定める必要がある。本研究では、リアルタイムの映像に加え、会話のトピックを提示し、会話の変化を明確にすることで、会話の雰囲気を把握する助けとすることを想定する。

2.3 トピック推定

リアルタイムの会話内容から、会話のトピックを推定する研究がなされている。丸田 [9] の研究では、対話ログから動的に推移する話題を検出する手法を提案し、継続した会話である雑談などの非タスク型対話システムへの導入がなされた。

3. 提案システム

本章では、開発したシステム概要について述べる。

3.1 システム機能

本システムは、常時複数地点間遠隔会話において、話かけたい遠隔地の相手が、別の人物と会話をしている時に、実際の対話音声を利用する代わりに、音声から会話内容に即したトピックを提示することで、プライバシーを守りつつ、抽象的な会話内容や会話に参加するきっかけを提示し、トピックに関連する人物なら参加できるというような、完全公開よりプライバシーレベルを上げたシステムである。システムの機能として、対話先のリアルタイムの映像や、トピックとなる文字列を提示する。

以下にシステムの各機能について説明していく。

(1) 会話途中参加支援

抽象的な会話のトピックを提示する(図1(1))ことで、遠隔地で、自分に関わりのある話をしているときや、自分の興味のある話をしているときは、会話への途中参加が可能になると考えられる。

(2) リアルタイム雰囲気提示

体の動きから話しかけられる最適なタイミングを考える研究 [10] があるため、実際に、対話している映像を提示し続ける(図1(2))ことで、対話先の雰囲気を提示することができる。トピックを見て、話しかけようとするとき、対話先の人物が笑っていれば、話しかけても大丈夫だと思えるようになることが考え



図 1 システム映像

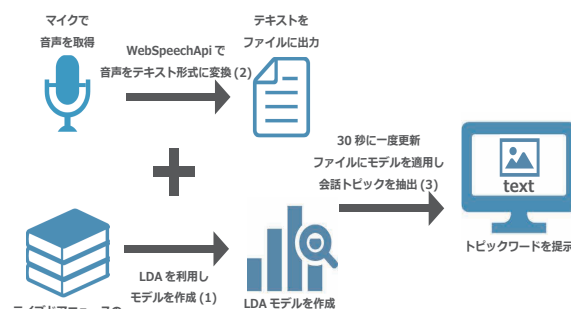


図 2 システム構成

られる。また、常時接続の際、他の作業をしながらの閲覧などが考えられる。その際、表示されている複数の文字を見て、内容を理解する以上に直観的な映像が役に立つと考えられる。

(3) プライバシー保護

遠隔地での会話内容がわかってしまうと、感情や思考などの個人がどうしているかがわかってしまう。そのため音声を非表示にする。

3.2 システム構成

図 2 にシステム構成図を示す。あらかじめ、公開されているライブドアニユースのコーパスを利用し、LDA のモデルを作成する (図 2(1))。各遠隔地の対話音声を遠隔地に設置されているマイクで取得する。WebSpeechAPI を利用して、マイクで取得した音声をリアルタイムに文字に変換する (図 2(2))。変換された文字を一定間隔 (30 秒) でテキストファイルに保存する。保存された文字を一定時間ごとに MeCab を利用して形態素解析を行い、抽出された名詞から会話内容のトピックを LDA を利用して一定間隔 (30 秒) ごとに推定する (図 2(3))。以上の情報を、プライバシーレベルに合わせて提示する。

今回、会話内容のトピックを推定するにあたり、トピック推定の精度の関係から LDA を利用することを考えた。また、コーパスを利用するにあたり、ライブドアニユースのコーパスの他に、Wikipedia のコーパスの使用も考えたが、Wikipedia のデータには、日常的に会話で使われないような単語や文章が多く含まれ、推定の精度が悪くなる恐れがあったため、今回の実験にはライブドアニユースを利用した。

4. 実験

4.1 検証項目と目標

本実験では、以下の項目について検証を行う。

- (1) 実験協力者が本システムを利用し、対話先の話題を推測することが可能か。
- (2) 実験協力者が本システムを利用し、話題を推測する際



システム適用映像を閲覧中の様子

図 3 実験風景

に、どんな情報を利用し推定を行うか。

今回の実験の目的は、システムに用いたモデルの汎用性や正確性の検証が目的ではなく、プライバシーの保護下で情報提示ができているかが検証の目的である。そのため、実験では、話題の推測ができるかや、プライバシー保護下で話題の変化を推測することができるかを検証する。また、話題の推測や、話題の変化の推測をする際、どの単語に着目すると、どう情報量が増減するか、あるいはどうして着目したかを分析し、どんな単語に着目することが推測の助けになったかを検証する。

4.2 実験手順・実験協力者タスク

あらかじめ、2 本の映像を録画しておく。録画した各映像をシステムに適用する。適用した映像を見てもらいながら、同時並行で、映像に出てきた単語から、会話の話題を推測してもらう。また、推測に使えそうな単語をメモしてもらう。映像を見終わった後に、参考にした単語を重要度の順番に並べ替えてもらう (図 3)。

次に、システムを適用した 2 本の映像を見てもらった後、元の音声があり、文字表示のないシステム適用前の映像を 2 本見てもらい、最後にアンケートに答えてもらう。



図 4 映像 1 : LDA のみのシステム映像

4.3 元の映像

元の映像は、それぞれ2～3分ほどの長さの映像を撮った。映像の会話内容はライブドアニースのカテゴリーに含まれるものを会話トピックとして利用した(スポーツ, 芸能, 経済など)。1つ目の映像はサッカーから野球に会話内容が変更するもの、そして、2つ目の映像は芸能から就活に会話内容が変更するものを録画した。

4.4 システムを適用した映像

システムを適用した映像は、両映像ともに、音声を利用しない動画と、会話のトピックとなる単語を表示したものであるという部分は同じだが、以下の点において違いがある。

- (1) 映像 1(LDA のみ):学習モデルによって得られた単語のみ(図 4(1))を利用。
- (2) 映像 2(LDA+固有名詞):映像 1 で利用された学習モデルによって得られた単語(図 5(1))に加え、実際の会話中で使われていた固有名詞(図 5(2))を追加した映像。以上の2つの映像を利用した。

4.5 実験詳細

今回の実験協力者は、和歌山大学の男子学生9名、女性学生1名である。実験結果の考察にはアンケートと、前述したシステム適用映像閲覧中に実験協力者に推測した話題を記述してもらったり、推定に利用できそうな単語を記録してもらったりしたメモをもとに行った。

5. 考察

5.1 会話内容の推測の可能性の実験結果と考察

表1に会話内容推測に関するアンケート結果を示す。「私は、表示された単語を見た時、具体的な会話内容を推測できた。」という質問をしたところ、LDA のみの場合は、5段階評価で中央値2.5、最頻値4という結果となった。また、LDA+固有名詞の場合は、5段階評価で中央値4、最頻値4という結果となった。

LDA のみで、同意すると評価した人の意見としては「大

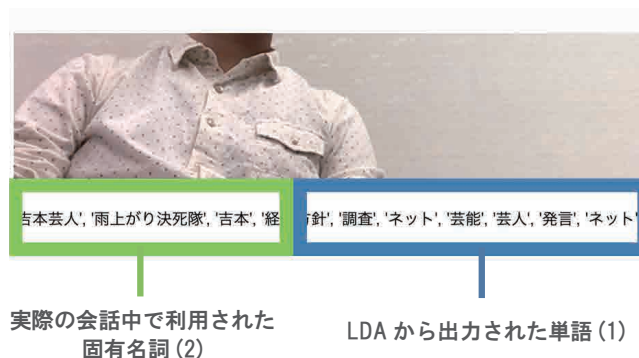


図 5 映像 2 : LDA+固有名詞のシステム映像

雑把なトピックは推測できた。」という意見があった。同意しないと評価した人の意見としては「具体的な内容までは分からなかった。」という意見があった。

LDA+固有名詞で、同意すると評価した人の意見としては「話の核となるキーワードが表示され具体的な推測ができた。」という意見があった。

実験の結果から、LDA のみの場合は抽象的な会話内容の提示に有用であること、提示する単語の中に、実際の会話で用いられた固有名詞を利用すると情報量が増加することが考えられる。

また、システム適用映像を閲覧中のメモから、両映像共に、参考にされた単語が、固有名詞が必ず上位に入っていたことが分かった。加えて、LDA+固有名詞を適用した映像では、LDA の出力が会話に沿ったものであるとき、話題の推測はおおむね正しいものであったことから、会話中で利用されている、固有名詞を利用することによって情報量提示量が増加することが考えられる。

5.2 会話内容と表示単語の正当性の実験結果と考察

表2に会話内容推測に関するアンケート結果を示す。「私は、実際の対話映像と表示された単語があっていると思ったか。」という質問をしたところ、LDA のみの場合は、5段階評価で中央値2、最頻値1,2という結果となった。また、LDA+固有名詞の場合は、5段階評価で中央値4、最頻値4という結果となった。

LDA のみで、同意すると評価した人の意見としては「大雑把にはあっていた。」という意見があった。同意しないと評価した人の意見としては「話題の核心を突く単語はなかった。」という意見があった。

実験の結果から、具体的な単語がないことが、会話内容の抽象化を可能にしていると考えられる。

次にLDA+固有名詞で、同意すると評価した人の意見としては「話の内容を推測できる単語が表示されている。」という意見があった。同意しないと評価した人の意見としては「具体的な単語が表示されたが、会話中で使われていなかった。」という意見があった。

実験の結果から、実際の会話で話されていない固有名詞

表 1 表示された単語からの具体的な会話内容の推測可能性に関する質問。

質問内容	システム種別	評価の分布					中央値	最頻値
		1	2	3	4	5		
私は、表示された単語を見た時、具体的な会話内容を推測できた。	LDA のみ	2	3	1	4	0	2.5	4
	LDA+固有名詞	0	1	1	7	1	4	4

評価項目：1: 強く同意しない, 2: 同意しない, 3: どちらとも言えない, 4: 同意する, 5: 強く同意する

表 2 実際の対話映像と表示された単語の正当性に関する質問。

質問内容	システム種別	評価の分布					中央値	最頻値
		1	2	3	4	5		
私は、実際の対話映像と表示された単語があっていると思った。	LDA のみ	3	3	2	2	0	2	1,2
	LDA+固有名詞	0	1	0	8	1	4	4

評価項目：1: 強く同意しない, 2: 同意しない, 3: どちらとも言えない, 4: 同意する, 5: 強く同意する

が提示単語の中に含まれている場合、推測を誤らせる可能性があることが分かった。

また、システム適用映像を閲覧中のメモから、LDA+固有名詞の時、LDA の方から出力された単語が、会話に沿っていない単語 (固有名詞を含む) を出力していた場合、会話内容の推測の誤りを引き起こしていた。このことから、LDA から出力される単語には固有名詞を利用しないなどの LDA から出力される単語の出力を調整する必要性が考えられる。

5.3 会話内容推定と会話内容変化の関係性の実験結果と考察

表 3 に会話内容推測に関するアンケート結果を示す。「私は、表示された単語の変化から対話内容の変化が分かった。」という質問をしたところ、LDA のみの場合は、5 段階評価で中央値 3、最頻値 4 という結果となった。また、LDA+固有名詞の場合は、5 段階評価で中央値 4、最頻値 5 という結果となった。

LDA のみで、同意すると評価した人の意見としては「違う系統の単語が登場したので、話題が変わったと推測した。」という意見があった。同意しないと評価した人の意見としては「表示される単語に規則性がなく、どう変化しているか不明。」という意見があった。

LDA+固有名詞で、同意すると評価した人の意見としては「違う系統の単語が登場したので、話題が変わったと推測した。」という意見があった。

実験の結果から、固有名詞による正確な会話内容の理解が、対話内容の理解を実現したと考えられる。また、具体的な会話内容が不明瞭だと会話内容の変化の推測が難しい

と考えられる。

また、LDA のみを適用した映像閲覧中のメモから、会話内容が変化した前後でトピックが似ているときは、話題の推測が難化することが分かった。例えば、高校野球の話をしている時は、「現役、共同、学園、スター、スタジアム」などの単語が出力される。しかし、高校野球の話の前にサッカーの話をしていると、サッカーの話が続いているとも受けとれる単語が出力されているため、同じ内容の話をしていると考える傾向にあることが分かった。この問題を解決するために、固有名詞を利用することで、会話内容が近い場合などに、変化を明確にする使い方が考えられる。

6. おわりに

プライバシーを考慮した常時遠隔会話のための会話状況判別手法の提案として、プライバシーを考慮した情報提示による、第 3 の情報提示方法を提案した。今回は、プライバシーの保護下で情報提示が可能かを検証の目的とし、具体的に以下の点に着目し実験を行った。

(1) 実験協力者が本システムを利用し、対話先の話題を推測することが可能か。

(2) 協力者が本システムを利用し、話題を推測する際に、どんな情報を利用し推定を行うか。

実験の結果、以下の点が明らかになった。

(1) LDA のみの場合は抽象的な会話内容の提示に有用

(2) 固有名詞を利用すると提示する情報量が増加し、LDA から出力される内容が会話に沿っている場合、LDA+固有名詞を利用すると会話内容の推測は実際の会話内容とほぼ一致

(3) LDA により出力された、会話に使われていない固有

表 3 トピック提示の内容から対話先の対話内容の変化の把握に関する質問

質問内容	システム種別	評価の分布					中央値	最頻値
		1	2	3	4	5		
私は、実際の対話映像と表示された単語があっていると思った。	LDA のみ	2	3	0	4	1	3	4
	LDA+固有名詞	0	1	2	3	4	4	5

評価項目：1: 強く同意しない, 2: 同意しない, 3: どちらとも言えない, 4: 同意する, 5: 強く同意する

名詞がある場合、会話の推測を誤る場合がある。そのため、LDA から固有名詞を出力しないなどの工夫が必要であると考えられる。

- (4) 具体的な会話内容が不明瞭だと会話内容の変化の推測が難しい。特に、会話内容が似ており、変化が分かりにくいものには、固有名詞を利用して、変化を明確にするなどの方法が考えられる。

今後の予定は、音声同様、映像においてもプライバシーを考慮した情報提示手法の提案を行う。

参考文献

- [1] 望月崇由, 久保宏一郎, 藤村香央里, 佐藤仁美, 下倉健一郎: 家族間における常時接続型コミュニケーション環境の評価, 電子情報通信学会技術研究報告, WIT, 福祉情報工学, Vol.107, pp.51-57(2007).
- [2] 望月 崇由, 井上 雅之, 佐藤 仁美, 大塚 晃一郎, 久保 宏一郎, 藤野 雄一, 波多江 健, 福重 淳一郎: 常時接続型コミュニケーションシステムにおけるアバタデザインの評価 (福祉工学および一般), 映像情報メディア学会技術報告, Vol.102, No.280, pp.33-38(2002).
- [3] P. Dourish and S. Bly.: Portholes: Supporting Awareness in a Distributed Work Group Proc. of Conf. on Human factors in computing systems (CHI'92), ACM, pp.541-547(1992).
- [4] R.S.Fish, R.E.Kraut, B.L.Chalfonte.: The VideoWindow System in Informal Communications, Proceedings of CSCW'90, pp.1-11(1990).
- [5] 半田 智子, 神原 啓介, 塚田 浩二, 椎尾 一郎: なめらカーテン, ヒューマンインタフェースシンポジウム 2009 論文集, pp.117-120(2009).
- [6] R.W.Root.: Design of a Multi-Media Vehicle for Social Browsing, Proceedings of CSCW'88, pp.25-38(1988).
- [7] 田中 貴紘, 藤田 欣也: ユーザの割り込み拒否度を考慮した円滑な会話開始支援エージェント, 電子情報通信学会論文誌 Vol.J92-A, No.11, pp.852-863(2009).
- [8] 田中貴紘, 松村京平, 藤田欣也: アプリケーションスイッチに着目した情報提示タイミング制御のための作業履歴の分析, 情報処理学会論文誌, vol.50(1), No.1, pp.314-322(2009).
- [9] 丸田 要: 自然な対話継続のための推移する話題推定, 言語処理学会第 24 回年次大会講演論文集, pp.4-11(2018).
- [10] 鷺見 和彦, 田中 宏一, 松山 隆司: 三次元姿勢計測を用いた人の動作特徴の記述, 画像の認識理解シンポジウム, MIRU2004, vol.1, pp.660-665(2004).