3F4-3

意思決定型議論における非言語情報に基づく重要発言の推定

Predicting Important Statements based on Non-Verbal Behaviors in Decision Making Discussions

二瓶 芙巳雄[*1] Nihei Fumio
高瀬 裕[*2] Takase Yutaka
中野 有紀子[*2] Yukiko Nakano

[*1] 成蹊大学大学院理工学研究科 Graduate School of Science and Technology, Seikei University
[*2] 成蹊大学理工学部 Faculty of Science and Technology, Seikei University

Discussion summary can be very useful resource obtained from group work. This study aims to propose classification models that predict important statements to be included in a discussion summary. As prediction features, we focused on nonverbal behaviors, such as attention, head motions, prosodic information, and co-occurrence patterns among them. We created different prediction models according to the degree of importance of the statements. A balanced model was able to predict important statements and non-important statements quite well: F-Measures were 0.677, 0.813 respectively.

1. はじめに

近年グループ議論は意思決定や新規アイデア創出など様々な目的のため広く利用されている. 多様な立場により議論されたその結果や過程は, 将来的にその組織において有益な資源になりうる. 後に効率的に議論を振り返り, 論点や意見を再確認するには, 議論の要点をまとめた議事録のような形式で保存されていることが望ましい. しかし, その作成には経験や手間, また議論に対する理解を要するため, 認知的な負荷が大きい.

そこで本研究では議論の自動要約作成技術の達成を目指し, 議論中の重要発言を議論の要約作成に貢献する発言と定義し, 議論参加者の非言語行動に基づいて推定するモデルを機械学習により作成する.

2. 関連研究

要約作成の研究では, 文書から重要文を抽出し, それらを接合する方法が取られてきた. 重要文の抽出にはその文書のタイトルを利用する方法や tf*idf を用いて文の重要さの重み付けを行う方法など, 様々なアプローチが提案されている. そして作成された要約文書と人手により作成された要約文書を比較することにより, 手法を評価している[Spärck 2007]. 一方, 動画像に対する要約の研究も行われており, Qian らは映像の色や動きからサッカー映像に含まれるゴール場面やシュート場面などハイライト場面を検出している[Qian 2010]. しかし, 議論が記録された映像は動きが少なく, ハイライト場面の検出手法を議論映像に応用するのは難しい. 一方, 動きが少ないニュース映像を要約する研究も行われている[Eickeler 1999]が, これはニュースキャスターによるニュースの概要紹介からリポート, カットアウトの後にはじめに戻るといったニュース映像の定型に着目したものとなっており, 定型が定義しづらい議論に対しては適切でない. また Murray らは文書要約技術をミーティングド要約に適用し[Murray 2006], キーフレーズなどの言語情報, 韻律情報, 発話長などの音声情報を用いて要約作成における対話行為に重みづけを行っている. この研究では使用された情報は全て人手により作成され, 自動化を目指したものではない. そこで本研究では, 議論の自動要約を目指し, 要約に含めるべき重要発言を議論参加者のコミュニケーション行動に基づき推定することを目的とする.

連絡先: 二瓶芙巳雄, 成蹊大学理工学研究科, 東京都武蔵野市吉祥寺北町 3-3-1, dm146211@cc.seikei.ac.jp

3. グループ議論対話コーパス

グループ議論における実験参加者の言語, 非言語行動を収集するための実験を実施した[林 2015].

40 人の実験参加者を 10 グループに分割し, 各グループはテーマが与えられた約 20 分の議論を 3 セッション行った. そして議論の様子を多様なセンサを用いて記録・計測した(図 1). 本研究ではそのうち「学園祭出店課題」(学園祭への来場者数などのデータが記された資料を基に, 学園祭での出店場所・内容を計画するもの) の議論で得られた非言語行動を分析対象とする. また参加者は, 自身の行動や言動が後に専門家により評価されることを告げられた.

実験では次の機材を用いて身体・頭部の動作や発話音声, 視線等の行動データを収集した: (a)モーションキャプチャ (b)加速度センサ (c)Microsoft Kinect (d)アイトラッカ (e)ビデオ (f)ヘッドセットマイク (g)性格特性アンケート.

本研究では, 各議論参加者の加速度センサデータから得られる頭部動作, 顔を記録した映像から得られる頭部注視方向, そして発話音声を分析対象とした. 加速度センサは各議論参加者の後頭部に装着され, 3 軸における加速度・角速度を約 30 フレーム毎秒で出力する. 頭部注視方向は各議論参加者の顔を拡大し 30 フレーム毎秒で記録した映像に対して画像処理による顔検出器を用いて得られる顔回転角度・顔位置から, 4 方向の顔注視方向(正面, 右, 左, 下)を推定するモデルを機械学習により作成し(分類精度 0.896), フレームごとの顔注視方向を得た. そして分類により得られた顔方向を, 手作業にて修正した. そして音声解析ツールにより発話音声の 10 ミリ秒ごとの音声インテンシティを取得した.

4. 発言の重要性の決定

要約文書作成の研究領域において, 対象とする文書に含まれる文から高い重要度を持つ文を複数のアノテータが選択する, 抽出要約化法がある[Filatova 2004]. 本研究でもこの方法を採用し, 議論の要約作成に貢献する重要発言を決定する.

そこで 10 名のアノテータが議論コーパスに含まれる計 7778 発話から, 議論の要約作成に貢献する重要な発話を選択する

表 1 重要発言として扱われる, 各閾値での発言数

閾値	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
重要発言数	4063	3237	2566	1850	1256	750	379	177	41

図 1 実験の様子

作業を行った. 音声認識器により音声区間を検出し, 音声信号の前後の無音区間の合計が 300ms 以上である場合に 1 発話と認定し, これを作業単位とした. 10 名のアノテーション結果を集計し, 各発話区間について, 何割のアノテータが重要であると判断したかを算出した. これを概略化合意割合と呼ぶ. 例えばある発話に対して, 10 人のアノテータの内 6 人が議論の要約作成に貢献する重要な発話と判断した場合, その発話に割り当てられる概略化合意割合は 0.6 となる. すなわちこれは, 発言の重要さを示す指標といえる.

5. 重要発言推定モデルの作成

議論の要約作成に貢献する重要発言を推定するモデルを提案する. 提案モデルは対象とする発言を重要な発言か, 重要でない発言かの二値として分類する.

5.1 予測対象

発話ごとに割り当てられる概略化合意割合が, ある閾値を超えた発話を概略化に寄与する重要発言として定義し, それを予測対象とする. 閾値は 0.1 から 0.9 まで, 0.1 刻みに複数設定した. ここで各閾値において重要発言となる発話数を表 1 に示す. 閾値を複数設定することにより, 様々な重要度の発言を推定することが可能になると考える.

5.2 特徴量

重要発言を推定するために使用する特徴量は, 話者・非話者の特徴量, 発話数順位の各順位における議論参加者の特徴量, 非言語行動の共起関係に基づく特徴量の 3 種類に大別される. また推定に使用する特徴量は各発話区間に対して抽出された.

(1) 話者・非話者の特徴量

重要発言を行う話者と, 傾聴する非話者の非言語行動を対象としたこの特徴量は, 以下の 3 種に分類される.

(A) 注視方向:得られた頭部注視方向から, 以下の 2 つを抽出した. これらは発話長で正規化される.

・話者が発話中に注視方向を変化させた回数

・話者が発話中に 2 人以上の非話者から注視を受けた頻度

(B) 頭部移動量:加速度センサが持つ 3 軸の角速度を以下の式(1)に適用することで, 各議論参加者の頭部合成角速度を算出した. ここでx_i^2, y_i^2, z_i^2はそれぞれ, i フレーム目におけるx, y, z軸の角速度である.

$$\text{頭部合成角速度} = \sqrt{x_i^2 + y_i^2 + z_i^2} \quad \cdots (1)$$

そして各発話区間について, 話者の頭部合成角速度の平均, 分散, 最大値, 最小値, 更に非話者の平均, 分散, 最大値, 最小値を算出することで, 計 8 つの特徴量を設定した.

(C) 音声情報: 10ms ごとに計測したインテンシティを用いて, 各発話区間における話者・非話者それぞれのインテンシティの平均, 分散, 最大値, 最小値を算出し, 特徴量とした. これに加え, 話者の発話長, ポーズ長, 議論内発話位置を発話情報の特徴量として抽出し, 計 11 個の特徴量を設定した.

(2) 発話数順位を考慮した特徴量

コミュニケーション能力が高い議論参加者と低い参加者間の行動の差異を明らかにするため, 議論参加者をグループ内発話数で順位付けた [1]. そこで各順位の議論参加者について, 以下の 4 つの注視特徴量, 4 つの頭部動作特徴量, 4 つの韻律特徴量を算出した.

・注視特徴量:他者のいずれか(正面, 右, 左), あるいはメモ(下)への各注視割合

・頭部動作特徴量:頭部合成角速度の平均・分散・最大値・最小値

・韻律特徴量:発話インテンシティの平均・分散・最大値・最小値

また特徴量は各議論参加者が発話時・非発話時に細分化された. そのため計 96(=4 議論参加者×12 特徴量 ×発話時・非発話時)特徴量が抽出された.

(3) 非言語行動の共起関係による特徴量

話者の声の調子と視線方向から発言が終わるタイミングが予測できる[Harvey 1974]ことからわかるように, 複数の非言語行動の共起関係やタイミングを考慮することが望ましい. 本研究では多次元の非言語行動における特徴的な共起行動を Vahdatpour の共起関係探索アルゴリズム[Vahdatpour 2009]を用いることにより探索する. これは時系列かつ離散化されたデータを対象としたアルゴリズムであり, 高頻度で共起した多次元の行動を共起関係として抽出する. そのため得られる共起関係は議論の登場する顕著な行動といえ, さらに測定誤差により偶然生じた低頻度の共起関係は排除することができる.

共起関係を探索する対象とした行動データ, すなわち共起パターンの構成素は, 以下の注視行動(4), 頭部動作行動(2), 発話行動(1)の計 7 つとした.

・注視:他者のいずれか(正面, 右, 左), あるいはメモ(下)への注視

・頭部動作:5.2 節で算出した頭部合成角速度をコンポーネント数 2 で EM アルゴリズムによりクラスタリングし, その結果得られた 2 つのクラスタにより, 頭部動作データを頭部動作あり・なしの二値に離散化した.

・発話:当該音声区間において発話している

結果として 125 の特徴的な共起パターンが得られた. 得られた共起パターンには, 発話数順位 3 位の議論参加者が発話する様子を順位 1 位と 4 位が注視する様子を表す(Rank3Uttr + Rank1LaRank3 + Rank4LaRank3)ものや, (Rank1LaRank4+ Rank2LaRank4+Rank3LaRank4) のような, 特定の議論参加者に注視が集まる共起関係が得られた. そして得られた 125 の共起パターンが各特徴量抽出区間において生じた割合を特徴量として設定した. 加えて共起パターンの構成素となる各行動が特徴量抽出区間において生じた割合も特徴量として設定した. その結果, 計 173 の特徴量が設定された.

5.3 統計検定による特徴量の選定

設定した特徴量から概略化に寄与する重要発言の推定に有用な特徴量を選定するため, t 検定を行った. また, データ数に

[1] 評価されたコミュニケーション能力[林 2015]グループ内順位と発話数順位には高い相関があったため($\rho = 0.80, p < 0.1$).

表 2　t 検定で有意であった特徴量数
表内は，条件①:5%水準の有意があった数 / 条件②: Choen's d > 0.2 / 条件①&②:両方を満たした数，を示す．括弧内の総数は選定前の特徴量数を示す．

閾値	(非)話者特徴量（21 件）	コミュニケーション能力順位特徴量（96 件）	共起特徴量（153 件）
0.1	19/12/12	70/37/37	117/59/59
0.2	19/14/14	72/39/39	122/62/62
0.3	21/14/14	73/41/41	123/59/59
0.4	20/14/14	68/39/39	125/56/56
0.5	19/14/14	64/40/40	118/53/53
0.6	19/12/12	64/47/47	125/54/54
0.7	19/13/13	58/48/47	112/52/52
0.8	15/13/13	57/50/47	105/50/50
0.9	14/14/14	51/60/50	97/59/59
一貫して有意	14/9/9	39/29/27	80/35/35

依存しない標準化された指標として，各特徴量における平均値の差の効果量を示す Cohen's d を算出した．

表 2 は，概略化合意割合に基づく 9 種類の閾値それぞれにおいて，条件①:5%水準で統計的有意差がみられた特徴量，条件②:Cohen's d が 0.2 を超えた特徴量，条件①&②:以上 2 つをともに満たした特徴量の数を示す．表最下段には，全ての閾値設定において上記の条件を満たす特徴量の数が示されている．これより，話者・非話者の特徴量 9 個，発話数順位特徴量 27 個，共起特徴量 35 個（表中，下線で示す），計 71 の特徴量は概略化合意割合に関わらず，両条件を満たすことがわかった．よって，この 71 個の特徴量は，概略化に寄与する発言の特徴をよく表す特徴量であるといえる．

これらのうち，Cohen's d の平均値(閾値 0.1～0.9 における Cohen's d から算出)が高いもの上位 30 個を表 3 に示す．表中$\bar{d}$の欄に，Cohen's d の平均値を示す．発話数順位を考慮しなければ，これら 30 個の特徴量のうち，より上位の特徴量には非話者時の行動に関するものが多いことがわかる．上位 15 個のうち 10 個が非話者時の特徴量であるのに対し，話者時の特徴量は 4 つだけである．一方，16 位～30 位を見ると，話者時の特徴量は 8 つであるのに対し非話者時の特徴量は 3 つにとどまる．このことから，重要発言の検出には，他者からどのように傾聴されているかがより重要な情報であることがわかる．また発話数順位特徴量の中の，話者時の特徴量に着目すると，上位 30 個中，発話数順位 1 位の話者に関する特徴量が 3 つ，順位 4 位の話者の特徴量は 4 つであったが，順位 2 位と 3 位の特徴量はそれぞれ 1 つだけである．これより，発話数順位 1 位と 4 位の参加者が話者である際の行動が重要発言推定に貢献すると推察できる．

共起関係による特徴量においては，173 件の特徴量のうち 80 件の特徴量が概略化合意割合に関わらず一貫して有意であった．中でも，発話数順位 2 位の議論参加者の発話行動($p < .01, \bar{d} = 0.317$)と，その頭部動作 ($p < .01, \bar{d} = 0.323$)の組み合わせから得られた共起パターン(Rank2Uttr + Rank2HM)による特徴量は，1%水準で統計的に有意であり，かつ$\bar{d} = 0.371$であった．さらにその共起パターンに，発話数順位 3 位の議論参加者の発話行動も追加した共起パターン(Rank2Uttr + Rank2HM + Rank3Uttr)でも，同様の統計的有意差がみられ，さらに$\bar{d} = 0.451$となった．単一の行動よりも共起パターンとして扱うことにより，より大きな効果量を得ることができる．すなわち得られた共起パターンは単一行動よりも概略化に寄与する発話の推定により貢献している可能性を示唆する．

表 3　$\bar{d}$の値による上位 30 個の特徴量

順位	対象	特徴量名	$\bar{d}$
1	話者	発話長	0.972
2	非話者	発話インテンシティ・最小値	0.841
3	非話者	頭部角速度・最小値	0.767
4	Rank1(非話者時)	発話インテンシティ・最小値	0.767
5	Rank1(非話者時)	発話インテンシティ・平均	0.747
6	Rank2(非話者時)	発話インテンシティ・最小値	0.713
7	非話者	発話インテンシティ・平均	0.695
8	Rank2(非話者時)	発話インテンシティ・平均	0.624
9	Rank1(非話者時)	頭部角速度・最小値	0.623
10	Rank3(非話者時)	発話インテンシティ・最小値	0.588
11	Rank2(非話者時)	頭部角速度・最小値	0.566
12	Rank3(非話者時)	頭部角速度・最小値	0.553
13	Rank4(話者時)	発話インテンシティ・最小値	0.549
14	Rank1(話者時)	頭部角速度・最小値	0.520
15	Rank1(話者時)	発話インテンシティ・最小値	0.518
16	共起行動	Rank2Uttr,Rank3Uttr	0.518
17	Rank4(話者時)	Rank2 を注視した割合	0.518
18	Rank2(話者時)	発話インテンシティ・最小値	0.516
19	話者	発話インテンシティ・最小値	0.510
20	Rank4(非話者時)	発話インテンシティ・最小値	0.500
21	非話者	頭部角速度・平均	0.494
22	Rank4(話者時)	メモを注視した割合	0.484
23	Rank3(話者時)	頭部角速度・最小値	0.476
24	話者	頭部角速度・最小値	0.469
25	Rank1(話者時)	発話インテンシティ・平均	0.467
26	共起行動	Rank3Uttr,Rank4Uttr	0.451
27	共起行動	Rank2Uttr,Rank3Uttr,Rank2HM	0.451
28	Rank4(話者時)	頭部角速度・最小値	0.445
29	Rank3(非話者時)	発話インテンシティ・平均	0.440
30	共起行動構成素	Rank4Uttr	0.438

以上のように概略化合意割合に関わらず一貫して統計的に有意である特徴量がある一方，概略化合意割合の閾値が上昇するにつれ，統計的有意性が消失する特徴量もある．例えば，発話数順位 4 位の議論参加者の話者としてメモを注視した割合は，概略化合意割合が 0.8 を超えると，5%水準の統計的有意差がみられなくなる．

5.4 予測モデルの作成

対象とする発話が重要発言であるか否かを推定するモデルを機械学習により作成する．使用する特徴量は 5.3 節の検定により，5%水準の有意差が t 検定により得られ，かつ Cohen's d が 0.2 を超えた特徴量とした．0.1～0.9 の閾値ごとに推定モデルを作成した．モデルは Leave One Group Out 交差検証法により評価した．これは LOOCV をグループに対して適用したもので，対象とする 8 グループのデータの内，1 グループをテストデータ，残りを訓練データする交差検証法である．訓練されたモデルは未知のグループから得られたテストデータにより評価されるため，グループによる違いを考慮した評価方法となっている．各モデルを作成する際は，訓練データに含まれる重要発言数と非重要発言数が同数になるようリサンプリングを行った．学習アルゴリズムは Random Forest を用いた．

表 4 各閾値におけるモデル性能評価

閾値	Class	Precision	Recall	F-Measure
0.1	重要	0.809	0.689	0.744
	非重要	0.698	0.815	0.752
0.2	重要	0.741	0.673	0.706
	非重要	0.757	0.813	0.784
0.3	重要	0.677	0.677	0.677
	非重要	0.813	0.813	0.813
0.4	重要	0.552	0.633	0.589
	非重要	0.857	0.810	0.833
0.5	重要	0.411	0.645	0.502
	非重要	0.907	0.789	0.844
0.6	重要	0.293	0.650	0.404
	非重要	0.944	0.791	0.861
0.7	重要	0.211	0.713	0.326
	非重要	0.974	0.803	0.880
0.8	重要	0.103	0.658	0.177
	非重要	0.983	0.775	0.867
0.9	重要	0.062	0.781	0.115
	非重要	0.995	0.788	0.880

概略化合意割合の閾値ごとに作成された重要・非重要発言推定モデルの評価結果を表 4 に示す．作成されたモデルにおいて，重要発言を推定する再現率の平均は 0.680 であった．これは二値分類問題においては比較的良好であり，提案モデルは対象とする発言の重要さに関わらず，重要発言を取りこぼしなく推定できることを示す．一方で，適合率の平均は 0.429 となった．学習に使用した概略化に寄与する発言数の減少が適合率の低下をもたらしていると考えられる．高い重要さを持つ発言を推定する精度の改善には，学習時のインスタンス数を増加させることが必要である．加えて概略化に寄与する発言・しない発言をあえて同数にしないなどのリサンプルの工夫などにより改善が見込まれる可能性もある．

6. 議論要約動画作成への応用

提案モデルにより重要発言として分類された発言区間を，議論を記録した動画から抽出することで，議論動画を短縮することができる．提案モデルによる分類結果にもとづき議論動画を圧縮した場合の動画時間の圧縮率と，アノテータにより与えられた真値の概略化合意割合を用いて重要発言を決定し，これらを用いて作成された動画の圧縮率を図 2 に示す．高い概略化合意割合では，推定結果に基づき作成された要約では，充分な圧縮率は得られていない．一方，概略化合意割合の閾値を 0.3 とした推定モデルにおいては，真値による圧縮率と比較して，あまり劣らない圧縮率を示している．その際の圧縮率は約 0.5 であり，これにより，約 20 分の議論を 10 分程度に短縮することが可能となることがわかる．

7. おわりに

本研究はグループ議論対話コーパスを分析することにより，議論の要約作成に貢献する重要発言を推定する複数のモデルを提案した．提案モデルは，議論参加者が表す非言語行動とその共起関係に着目した特徴量を用いている．モデルは発言の重要さに応じて複数作成され，中でも 30%を超える人数が重要と判断した発言はよくバランスする推定精度(重要:0.677，非重要:0.813)で推定が可能である．

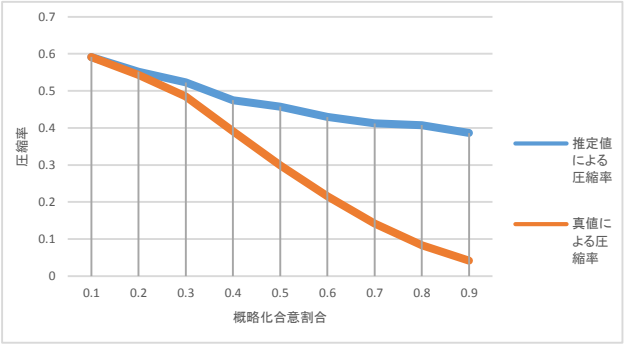

図 2 期待される動画時間圧縮率

将来的にはより高い精度で概略化に寄与する発言を推定できるよう，さらに特徴量を探索すると共に，学習に使用するインスタンスを増加させ高い概略化合意割合の発言を推定する精度を改善する．加えて推定された概略化に寄与する発言に基づく議論動画要約ツールを完成させ，提案モデルが議論の概略化に貢献するかを評価する．

参考文献

[Eickeler 1999] Eickeler, S. & Muller, S.: Content-based video indexing of TV broadcast news using hidden Markov models, In *Acoustics, Speech, and Signal Processing, Proceedings., 1999 IEEE International Conference on*. pp. 2997–3000 vol.6, *1999*.

[Filatova 2004] Filatova, E. & Hatzivassiloglou, V.: Event-based extractive summarization. *Proceedings of ACL Workshop on Summarization*, pp.104–111, 2004.

[Harvey 1974] Harvey Sacks Emanuel A. Schegloff, G.J.: A Simplest Systematics for the Organization of Turn-Taking for Conversation. *Language*, 50(4), pp.696–735, 1974.

[Murray 2006] Murray, G. et al.: Incorporating Speaker and Discourse Features into Speech Summarization. In *Proceedings of the Main Conference on Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics*. HLT-NAACL '06. Stroudsburg, PA, USA: Association for Computational Linguistics, pp. 367–374, 2006.

[Qian 2010] Qian, X. et al.: Highlight Events Detection in Soccer Video Using HCRF. In *Proceedings of the Second International Conference on Internet Multimedia Computing and Service*. ICIMCS '10. New York, NY, USA: ACM, pp. 171–174, 2010.

[Spärck 2007] Spärck Jones, K.: Automatic Summarising: The State of the Art. *Inf. Process. Manage.*, 43(6), pp.1449–1481, 2007.

[Vahdatpour 2009] Vahdatpour, A., Amini, N. & Sarrafzadeh, M.: Toward Unsupervised Activity Discovery Using Multi-dimensional Motif Detection in Time Series. In *Proceedings of the 21st International Jont Conference on Artifical Intelligence*. IJCAI'09. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., pp. 1261–1266, 2009.

[林 2015] 林佑樹 et al.: グループディスカッションコーパスの構築および性格特性との関連性の分析. 情報処理学会論文誌, 56(4), pp.1217–1227, 2015.