

リアクションを考慮した HMM による コミュニケーション場の雰囲気推定アルゴリズム

A proposal of atmospheric estimation algorithm
for communication field applying HMM with reaction

西田 悠[†]
Haruka Nishida

中平 勝子[†]
Katsuko T. Nakahira

北島 宗雄[†]
Muneo Kitajima

1 はじめに

近年、急速な情報処理技術の発展とともに、その応用先の一つとして教育への寄与が検討されている。その一つの可能性として、Learning Analytics(LA)が挙げられる。LAが提唱された時点では、解析に用いられるデータとして念頭に置かれていた主たる情報源は、eラーニングで収集される学習ログや学習者プロフィール、eテストの結果などが候補として上がっていた。また、高等教育機関では、アクティブ・ラーニング(AL)の必要性も説かれ、各機関が創意工夫をこらす形で実施している。中でも、事前に学生に予習をさせ、授業時間時は課題を解かせる反転学習型のALでは、事前学習にeラーニングが活用されることも多く、その結果、LAに活用できるデータの取得も可能となっている。他方、ALでは、学修者の能動的な活動を促す精神のもと、ディベートやグループ討論、グループワークを積極的に取り入れることが好ましいとされている。これらはいずれも学修者間の能動的なコミュニケーションによって問題を解決したり新たな知識を構築するといった成果が期待される。

コミュニケーションを伴う学習活動を評価する場合、単に解決された問題をレポートなどで報告させるといった論述能力に対する評価のほか、問題解決や知識構築のプロセスとしてのコミュニケーション進行自体を評価できると、より良い教育評価に繋がることが予想される。その実現のための一つの方策として、コミュニケーション場の記述・符号化作業が必要である。さきがけ的な研究は後藤ら[1]によって部分的に行われているが、場の記述・符号化の自動化が進むことで、最終的にはLAと親和性の高い電子データ保存が可能となると考えられる。

本稿では、コミュニケーション場の記述・符号化の発展を念頭に、その巧拙の評価を行うための基礎研究として、複数人が会話を行うコミュニケーション場における

雰囲気推定手法の検討を、隠れマルコフモデル(HMM)を用いて行う。

2 先行研究

近年、empathやcorevoなどに代表される音声から個人の感情を推定・認識する技術の開発が進んでいる。

門谷ら[2]は音声に含まれる感情の音声認識の精度低下への影響度を調べており、どの感情がどのくらい影響するかを示している。その過程で怒り・悲しみ・喜びの感情の判定に寄与する音響特徴量を示しており、怒りについて文中の最大振幅 A_{max} が最も大きく寄与し、 $F0$ の平均値が次に寄与するとしている。

後藤ら[3]は音声中の有声休止を検出し、そこを起点として音声認識を行うことで発話区間検出を容易にすることを提案している。この中で、「えー」や「あー」などの言い淀みの検出手法の提案を行っている。

波多野ら[4]は質問発話の句末において $F0_{move}$ と $F0_{reset}$ の2つの特徴量を定義してこれらの増減を調べ、上昇調や下降調などを検出している。

栗田ら[5]や田中ら[6]は、音声から笑い声の検出を試みており、高い精度での検出を行っている。

これらのように、音声から文章でなく音響情報を抽出・検出したり個人の感情を推定する手法が数多く研究されている。

しかし複数人会話において感情推定や雰囲気推定を行った例は少なく、技術や手法の確立がなされていないと考える。

3 提案モデル

雰囲気を推定において、単体のマイクのみから得られる音声情報のみを用い、会話内容を考えないこととする。プライバシー保護の観点から、画像処理や音声認識技術を利用しないことは有用であると考えられる。さらに、複数チャンネルの音声を利用しないことで複数人の声が入り混じった場の状況に近い情報を得ることができると考える。

[†]長岡技術科学大学

場の雰囲気、観測できず推定が必要なものとし、音響情報を観測可能な変数として考えることでHMMによる推定可能な問題として扱う。HMMは状態の推定や状態が移り変わる遷移確率を求めることができるモデルであり、時系列変化する雰囲気の推定に有用であると考える。

まず、音声から検出する6つのリアクションを(1)笑う、(2)怒る、(3)黙る、(4)質問、(5)相槌、(6)平常、で定義し、これを可観測な変数とする。

これらのリアクションは、視覚情報が不要でかつ会話内容を伴わずに利用できることを前提として選定した。さらに、先行研究などから検出や抽出が可能であり、実際に抽出可能である点も考慮した。

平常は他のどのリアクションも検出されなかった場合に与えられる。基本的に大きなリアクションが無い音声の場合、雰囲気の変化は見られないと考え、平常を得た場合には状態遷移が起こらない確率が高いと考える。

次に、推定したい隠れ変数として4つの雰囲気を想定し、表1の様な用語で表現する。

さらに、各雰囲気とリアクションの関係性を表2に示す。数値は大きいほどその雰囲気との関係が深いことを示す。

表1 雰囲気大別

TA-	TA+	EA-	EA+
落ち着いている	平素	頭の動きが鈍い	活動的
穏やか	びくびく	ぼんやり	エネルギー
ゆったり	ピリピリ	無気力	元気
リラックス	不安	活気がない	やる気
冷静		気が進まない	生き生き
		い	

さらに、各雰囲気とリアクションの関係性を表2に示す。数値は大きいほどその雰囲気との関係が深いことを示す。

表2 リアクションと雰囲気の関係

雰囲気	笑う	怒る	黙る	相槌	質問	平常
TA-	1	1	3	2	4	2
TA+	1	4	3	1	1	2
EA-	2	1	4	2	1	2
EA+	4	1	1	2	3	2

これを元に出力確率の初期値を与えてHMMによる状態遷移確率推定を行う。

4 モデル検証

Baum-Welch アルゴリズムを用いて各雰囲気間の状態遷移確率を推定することを考える。

今回は10個のEA+と感じられる4~5分の会話音声を利用する。これらからリアクションを手動で抽出して学習させ、モデルを構築する。その後、新たに与えた入力リアクション列から状態の推定を行う。この入力データは学習に使ったもの以外で、5個のEA+と1個のTA-と感じられる会話音声を利用した。

またリアクション抽出範囲をセクションを定義し、ここでは10秒を区切りとする。

この結果、状態推定の正解率は50.02[%]となった。学習データ・検証データともに非常に少ないことから、十分に学習が行われていない。

しかし、定義したリアクションからある程度雰囲気の推定が可能であることが示唆され、データ量や種類を増加させることで精度の向上が見込まれる。

5 まとめ

本稿では、音声データから検出可能なリアクションを定義した。それらを可観測な変数・コミュニケーション場の雰囲気を隠れ変数としたHMMを提案した。

実際にモデルを検証し、雰囲気の推定が可能であることを示唆した。

状態遷移確率は学習を重ねて精度が向上されることから、より多くのシチュエーションでの会話音声を利用することでさらに精度を向上していけると考える。

また、今回は手作業にて行ったリアクションのラベリングだが、多くの先行研究からその多くが自動的に検出可能な技術が開発・研究されている。実際にこれらを取り入れることで、リアクション検出から雰囲気推定までを自動的に行えるようにしていきたい。

さらに、検出可能なリアクションの種類増加や精度の向上などから、より実用的な推定が可能になると考えられる。

今後の予定として、各リアクションを自動的に検出可能にすることと、N階マルコフ過程を考えることで悪い雰囲気や良い雰囲気を引きずった推定を考える。

【謝辞】本研究は、科研費(基盤研究(C), 16K01061)の支援を受けて行われたものである。

参考文献

- [1] 後藤義貴, 中平勝子, 北島宗雄. 円滑なコミュニケーション形成支援を目的とした符号化・記述手法. Technical Report 6, 長岡技術科学大学, jan 2016.
- [2] 門谷信愛希, 阿曾弘具, 鈴木基之, 牧野正三. 音声に含まれる感情の判別に関する検討. 電子情報通信学会技術研究報告. SP, 音声, Vol. 100, No. 522, pp. 43-48, dec 2000.
- [3] 後藤真孝, 北山広治, 伊藤克亘, 小林哲則. 音声スタート:有声休止による発話開始の指定が可能な音声入力インタフェース. 情報処理学会論文誌, Vol. 48, No. 5, pp. 2001-2011, may 2007.
- [4] 波多野博頭, 石井カルロス寿憲. 日本語自然対話に現れる質問発話の句末音調. 音声研究, Vol. 21, No. 1, pp. 1-11, 2017.
- [5] 栗田将史, 鈴木隆広, 杉山雅英. Vcml player 字幕生成のための笑い声の検出. 電子情報通信学会技術研究報告. SP, 音声, Vol. 102, No. 417, pp. 13-18, oct 2002.
- [6] 田中爽太, 鈴木健嗣. 情動表現理解のための笑い声判別システムの開発. 電子情報通信学会 HCG シンポジウム 2015 論文集, pp. 65-68, 2015.