

複数人談話における笑いの情報を考慮した盛り上がり判定

嶋田 和孝[†] 楠本 章裕[†] 横山 貴彦[†] 遠藤 勉[†]

[†]九州工業大学 情報工学部 知能情報工学科

〒 820-8502 福岡県飯塚市川津 680-4

E-mail: †{shimada,endo}@pluto.ai.kyutech.ac.jp

あらまし 複数人談話を対象とし、談話中の盛り上がり箇所を検出する手法について提案する。対話の盛り上がりを推定できれば、ユーザ心理の把握や、それに基づく詳細な対話状態の理解などが可能になる。提案手法では、言語的特徴の他に、笑いという状態に着目する。対話に現れる笑いを外部からの発話や行動に対して発生したものか、自発的な笑いかに分類する。また、笑いの大きさを3段階に分けて、特徴とする。これらの笑い特徴と、bag-of-words や時間的特徴、その他の言語的特徴などを組み合わせて、分類器に適用する。実験では、5分程度の複数人対話のデータを10セット用意し、提案手法で評価した。実験結果より、bag-of-words のみによる分類器と比較して、笑い特徴を利用した提案手法の有効性が確認された。

キーワード 複数人談話, 盛り上がり判定, 笑い特徴

Hot spot detection in multi-party conversation using laughing feature

Kazutaka SHIMADA[†], Akihiro KUSUMOTO[†], Takahiko YOKOYAMA[†], and Tsutomu ENDO[†]

[†] Department of Artificial Intelligence, Kyushu Institute of Technology

680-4 Kawazu Iizuka Fukuoka, 820-8502 Japan

E-mail: †{shimada,endo}@pluto.ai.kyutech.ac.jp

Abstract In this paper, we propose a method for hot spot detection in multi-party conversation. Detecting hot spots in conversation helps to recognize user's minds, and leads to the improvement of dialogue understanding. We focus on laughing of each user. We divide laughing situations into "internal" and "external". The "internal" and the "external" denote unprompted laughing and laughing that is caused by actions or utterances from other persons, respectively. Besides, we classify the power of laughing into "smile", "laugh" and "burst". We generate a classifier with these laughing features and basic features such as bag-of-words. We evaluate the proposed method with 10 dialogue sets. The experimental result shows the effectiveness of the proposed method with laughing features.

Key words Multi-party conversation, Hot spot detection, Laughing feature

1. はじめに

音声対話において、対話の盛り上がり状態を分析することは、重要な課題の一つである。例えば、盛り上がり様子から、ユーザの心理状態を把握し、それに合わせて円滑な対話を行う対話エージェントや、対話の盛り上がり箇所からユーザの興味を獲得することで、様々な情報の推薦システムに利用することができる。盛り上がりの分析のためには、まず盛り上がり箇所を検出する必要があり、これらの研究は盛んに行われている[1][2]。

音声対話には、声の大きさ、速さ、高さ、笑い、オーバーラップの頻度、フィルター、特定の言語表現の有無といった多くの特徴が存在している。これらの特徴と盛り上がりとの関連性は、

いくつかの研究で示されている。徳久ら[3]は、発話行為や修辭構造の分類を行い、これらを分析した結果、主観的な発話やユーモラスな発話、相手の発話を補完するような発話は、盛り上がり時に頻出し、また、客観的な発話や関心を示す発話、肯定的な発話の出現により、対話が収束し、非盛り上がりを引き起こす要因となることを述べている。このことから、発話の言語的な情報が、対話の盛り上がり判定に有効であることが分かる。また、西村ら[4]は、オーバーラップの頻度、声の高さ、フィルターと盛り上がりとの関連性を調査し、特に、オーバーラップの頻度、声の高さは、盛り上がり強い相関があることを述べており、言語的な情報だけではなく、音声のみに表れる特徴も盛り上がりにも影響すると考えられる。

本論文では、笑いという情報に着目する。笑いに関しては、

須見ら [5] が、盛り上がり強い影響を与えると考え、笑いを検出する手法について述べているが、盛り上がりどのような影響を与えるか言及していない。このため、笑いの分析を行い、盛り上がりへの影響について検証する必要がある。

本研究では、まず、盛り上がり強い影響を及ぼしている笑いを見つけるため、笑いの分析を行う。次に、笑い特徴および盛り上がり影響すると考えられる時間、極性、接続表現に基づく特徴を取得する。そして、タグ付けを人手で行った対話コーパスに対して、それらの特徴量を機械学習に適用することで、盛り上がり箇所を検出する。

2. 笑いの分析

本節では、本研究の着目点となる笑いの分析について述べる。須見ら [5] は、笑いが盛り上がり強く影響することを示唆しているが、笑いの検出手法について述べており、笑いの有効性について言及していない。笑いが談話の活性化に影響することは容易に想像できるが、笑いにも多くの種類があり、それらが盛り上がり及ぼす影響は、異なると考えられる。高精度な盛り上がり判定を行うためには、笑いの種類や特性を考慮した適用が必要になると考えられる。

笑いに関する研究として志水 [6] の分類がある。志水は、笑いを「快の笑い」、「社交場の笑い」、「緊張緩和の笑い」の3つに分類している。「快の笑い」には、本能充足の笑いや優越の笑いなど、「社交場の笑い」には、協調の笑いや価値無化の笑いなど、「緊張緩和の笑い」には、強い緊張や弱い緊張が緩んだ時の笑いが含まれる。志水は、これらのうちのいくつかが複合し、笑いが発生すると述べている。尾崎 [7] は、授業中に教室内で発生した発生した笑いについて、「外部の誘引によって起こる笑い」と「内部の誘因によって起こる笑い」の2種類があることを述べている。「外部の誘引によって起こる笑い」とは、間違い・勘違い、驚き・困惑・失敗、場の空気からの逸脱などを意味し、「内部の誘因によって起こる笑い」とは、同調や深刻さ・気まずさを包み隠すために生じるものだとしている。

このような多岐にわたる笑いの分類を行うことは、主観的な要素が強いため、人手でさえも正確な判定が難しく、コストも大きい。そのため、人手で容易に分類を行うことができ、かつ盛り上がり時に頻出する笑いとうでないものに分ける必要がある。そこで、本研究では、外部からの発話(行動)に対して発生する笑い(external)と自発的な笑い(internal)に注目する。externalな笑いの周辺には、笑いを引き起こした発話があり、その発話は、笑いを引き起こすような意味を持つと考えられる。一方で、internalな笑いとは、恥ずかしさを紛らわすためや、場の雰囲気が重たい時に雰囲気を変えるための笑いであり、このような笑いは、談話の盛り上がり時に必ず発生するとは限らない。

さらに、笑いの分類としては、単純で、影響力の高いと考えられる笑いの大きさにも着目する。笑いの大きさ順に(smile, laugh, burst)として3値に分類する。その他の先行研究で紹介したような多岐にわたる笑いの分類については、本論文では扱わず、言語的な特徴や時間的な特徴など複数の特徴を組み合

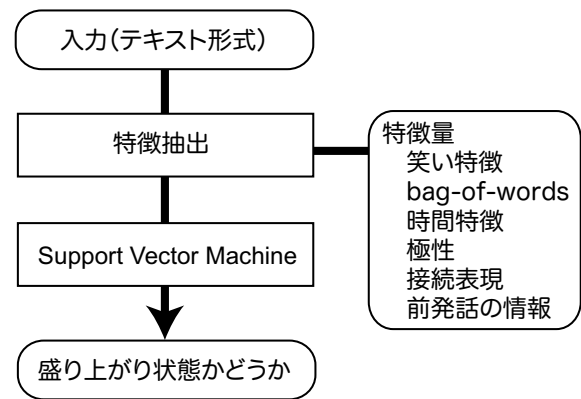


図1 提案手法の概要

わせることによって盛り上がりを判定する。

3. 提案手法

本研究では、複数人による談話を対象に各発話が盛り上がりしているかどうかを判定する手法を提案する。本節では、盛り上がりを検出するために用いる特徴量の詳細や盛り上がり判定のための分類器について述べる。まず、本研究の主となる笑いに関する特徴を説明し、その後、基本となる bag-of-words 特徴(単語の有無)について述べる。さらに、時間、極性、接続表現に基づく特徴量に関して説明する。図1に提案手法における処理の流れを示す。

3.1 盛り上がり判定のための特徴量

3.1.1 笑いに関する特徴

本研究では、盛り上がり判定のために以下の笑いの特徴を用いる。

- 笑いの発生要因 (internal, external)
- 笑いの大きさ (smile, laugh, burst)
- 笑いの発生時間
- 同時に起きた笑いの人数
- 笑いの密度 (集中度)
- 発話 - 笑いの時間
- 笑いの人物情報

前節での笑いの分類に基づいて、外部からの発話(行動)に対して発生する笑い(external)であるか、自発的な笑い(internal)であるかを笑いの発生要因として用い、笑いの大きさについては、3段階に人手で分けた値(smile, laugh, burst)を用いる。

また、時間や笑った人に関する情報も、その笑いの意味的な強さに関連していると考えられる。笑っている時間が長いほど、その発話は盛り上がりしていると考えられるため、笑いの時間に関して、笑いの発生開始から終了までの時間を特徴量とする。どのくらいの密度で笑いが起こっているかも重要な特徴であると考えられる。そこで、笑いの密度として、盛り上がり判定の対象となる発話(対象発話)から数秒間前までに発生したすべての笑いの回数と時間の合計を特徴量として用いる。予備実験を行い、今回は最適であった対象発話から15秒間前までに発生した笑いを対象としている。発話の終了から笑いの発生までの時間が短ければ、その笑いが対象発話に与える影響が強いと考

えられる。そこで、この時間を、発話 - 笑いの時間特徴量として導入する。また、笑いの特徴は個人差が大きい。例えば、発話のたびに笑う人物と滅多に笑わない人が笑った場合では、その笑いの持つ意味は大きく異なる。このような違いを取り扱うために、誰が笑ったのかや、どのくらいの人が同時に笑ったのかなどの情報を特徴量として用いる。

提案手法で笑い情報を扱う場合、どの笑いがどの発話に関連しているかを推定しなければならない。発話中に生じた笑いが、その発話以外にも関連している場合を考える必要がある。例えば、ある発話のあとに起きた笑いが、その発話によって生じた笑いの可能性もある。そこで、発話の開始時間から終了時間+2秒間の間に発生した笑いの特徴をその発話に付加する。この時間間隔の長さは、あらかじめ予備実験を行い、最適な値になるように設定した。

3.1.2 bag of words

言語処理において用いられる最も基本的な特徴量は、単語の有無 (bag of words) である。本研究でもこれをベースラインとする。単語への分割には、形態素解析器 MeCab [8] を利用する。単語の有無をそのまま扱くと、全データで 1~2 回程度しか出現しない低頻度語によって分類器が過学習し、盛り上がり判定が失敗する可能性がある。このような問題を防ぐため、本研究では、全対話データ中に 5 回以上出現した単語の有無を bag-of-words 特徴量として利用する。

3.1.3 時間に関する特徴

本論文では、時間に関する特徴量として、以下のものを用いる。

- 発話時間
- 沈黙時間
- 形態素数
- 発話速度
- オーバーラップ・被オーバーラップ

発話の時間の長さは、場の雰囲気を変化させると考えられる。一人の発話時間が長い場合、盛り上がりにも影響することがあると考えられるため、発話の開始時間から終了時間までを発話時間の時間の特徴量として用いる。発話自体の長さだけでなく、発話間の時間の長さも盛り上がりにも関連していると考えられる。具体的には発話間の時間 (沈黙時間) が短いほど、談話は盛り上がっていると考えられる。そのため直近の発話の終了から対象発話の開始までの時間差を特徴量として用いる。

発話に含まれる形態素数が大きいほど、相手に伝達したい情報が多く、談話に積極的であると考えられる。一方で、これが小さいと、短い発話が飛び交い、話者たちの談話が盛り上がっている状態である可能性がある。そこで、形態素数も特徴量として用いる。ここで、形態素数特徴量では、句読点を除いた形態素数を利用する。

次に、発話の速度が速いほど、話者は談話に夢中で、盛り上がっているという仮定に基づき、形態素数を発話時間で除算した値を特徴量として用いる。また、発話のオーバーラップも盛り上がりにも関連した特徴である。ここで、オーバーラップとは、2 つ以上の発話が被っていることを指し、これは、複数の話者が

表 1 評価表現辞書 [9]

語	極性値
心配がない	15.76
コストが安い	15.00
丈夫だ	15.76
楽	14.42
...	...
不足	-13.99
駅から遠い	-14.32

表 2 日本語評価極性辞書 [10]

単語	読み	品詞	極性値
優れる	すぐれる	動詞	1.000000
良い	よい	形容詞	0.999995
喜ぶ	よるこぶ	動詞	0.999979
...
死ぬ	しぬ	動詞	-0.999999
悪い	わるい	形容詞	-1.000000

談話に参加したいという意志を表していると考えられる。そこで、発話が 1 発話前に被っている状態をオーバーラップ、発話が 1 発話後に被せられている状態を被オーバーラップとして、その有無を特徴量とする。

3.1.4 極性

次に、極性に基づく特徴量を導入する。ここで極性とは、発話が肯定的 (positive: P) または否定的 (negative: N) な意味を含んでいるか否かを指す。肯定的もしくは否定的な発話は、対話参加者の心理状態に影響を与え、場の雰囲気を変化させるため、談話の盛り上がりにも影響すると考えられる。徳久ら [3] は、発話が肯定的であるか否定的であるかを発話毎に判定し、その結果、肯定的である発話は、盛り上がり度が低い箇所では盛り上げの効果があり、高い箇所では話をまとめる効果があることを指摘している。今回は、発話の PN 判定に、鍛冶らの構築した評価表現辞書 [9] と高村らの構築した日本語評価極性辞書 [10] を利用する。これらの辞書では、表 1、表 2 のように、単語やフレーズに肯定、否定の極性値が割り当てられている。本研究では、この極性値を用いて、発話の極性を判定する。

- P らしさ (肯定の強さ) の合計値・平均値
- N らしさ (否定の強さ) の合計値・平均値

P らしさ、N らしさの合計値は、形態素解析器によって得られた単語に関する発話中での極性値の合計である。ただし、この手法だけでは、発話数が大きいほど、P らしさまたは N らしさの値が大きくなるため、P らしさ、N らしさの合計を形態素数で除算した値を P らしさ、N らしさの平均値として、特徴量に組み込む。それぞれの値は、予備実験に基づき、図 2 に示すような 3 段階に分けた。

3.1.5 接続表現

接続表現とは、文と文を繋ぐ表現であり、中道ら [11] は接続表現の出現は、発話の極性を変化させると述べている。発話中に極性の変化があるということは、密な発話が行われていると考えられる。また、中道らは接続表現は、文の前半と後半の極

評価表現辞書 [9]	
P の合計値	6 以下:P らしさなし 6 ~ 12:弱い P 12 以上:強い P
N の合計値	-6 以上:N らしさなし -12 ~ -6:弱い N -12 以下:強い N
P の平均値	0.5 以下:P らしさなし 0.5 ~ 1:弱い P 1 以上:強い P
N の平均値	-0.5 以上:N らしさなし -1 ~ -0.5:弱い N -1 以下:強い N

日本語評価極性辞書 [10]	
P の合計値	2 以下:P らしさなし 2 ~ 4:弱い P 4 以上:強い P
N の合計値	-3 以上:N らしさなし -6 ~ -3:弱い N -6 以下:強い N
P の平均値	0.2 以下:P らしさなし 0.2 ~ 0.4:弱い P 0.4 以上:強い P
N の平均値	-0.3 以上:N らしさなし -0.6 ~ -0.3:弱い N -0.6 以下:強い N

図 2 評価表現の適用ルール

性を一致させる保持、文の前半の極性に関わらず、後半の極性を肯定にする共起、文の前半と後半の極性が反対になる反転に分類されると述べており、本研究では、以下を特徴量する。

- 接続表現の有無
- 接続表現：保持関係の有無
- 接続表現：共起関係の有無
- 接続表現：反転関係の有無

接続表現の判定には、中道らが用いた接続表現に基づいて行う。ただし、談話では、口語的な表現が多く、中道らが利用したような接続表現は出現しづらいため、表 3 のように接続表現の拡張を行った。表中の太字が拡張部分である。接続表現の取得数は、拡張前は全体の発話の 2 % 弱であったが、この拡張により、5 % 弱に上昇した。ここで、「**というか**」など一部の表現は、文頭や文末にも出現するが、このように接続表現としての意味を成していない場合は、対象としていない。

3.1.6 対話の流れの考慮

対話には必ず流れが存在し、対象発話の盛り上がりは、その発話の数発話前の特徴に大きく依存していると考えられる。特に、極性などの特徴は、対話の雰囲気に影響してくるので、極性を含む発話そのものよりも、それ以降の発話の盛り上がりへの影響が考えられる。そこで、対象発話の 1 および 2 発話前の特徴量を対象発話の特徴量に追加して、盛り上がり判定に利用する。ただし、笑いの密度では、発話の 15 秒前からの特徴を

表 3 接続表現の拡張

接続表現	種類
たところで とところで	反転
としても としても	反転
というのは というのは	保持
みたいに みたいな	保持
おかげで	共起
とはいえ ていっても	反転
...	...

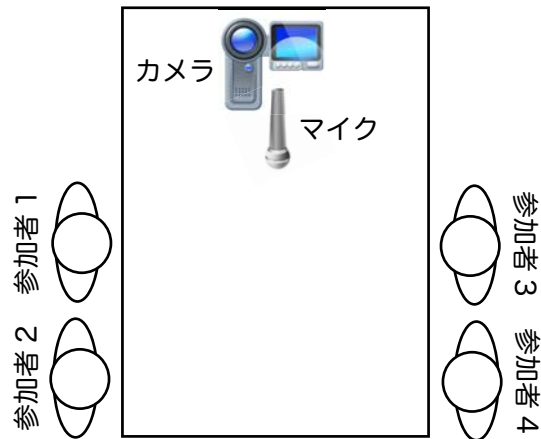


図 3 実験環境

取得しているため、数発話前の笑いの密度も、ほぼ同様の笑いの情報を含んでいるため、除外する。

3.2 分類器

前節で述べた特徴量を利用して、分類器を作成し、盛り上がり判定する。具体的には、各発話の盛り上がりか、盛り上がっていないかの 2 値問題を判別する。分類器としては Support Vector Machine (SVM) を利用する。SVM は Vapnik [12] が考案した Optimal Separating Hyperplane を起源とする超平面による特徴空間の分割法であり、現在、2 値分類問題を解決するための最も優秀な学習モデルの一つとして知られている。実装にはデータマイニングツール Weka [13] を利用する。

4. 実験

4.1 実験環境

まず、複数人による自由対話を収集した。対話環境を図 3 の示す。対話収集では、被験者は 10 名の大学生から対話ごとにランダムに 4 名選出し、対話ごとに 1 つのテーマを提示し、異なるテーマの 10 対話を収集した。対話時間を 5 分程度とし、指向性マイクによって録音したものを人手で書き起こした。同一発話者が連続して発話する場合、発話中に 0.6 秒以上間隔がある場合は、異なる発話として扱い、発話の書き起こしとともにそれぞれ発話者の ID、発話の開始時間、終了時間なども記録した。表 4 が作成された対話コーパスの一部である。10 対話は、1697 発話で構成される。

盛り上がり判定の実験のため、対話コーパスの発話ごとに、2 名のアノテータによって盛り上がり度を評価した。ここでの盛り上がりとは「話者がある発話を行った時点で、その話者が談

表 4 対話コーパス

発話者	開始時間	終了時間	発話内容
M00	02:06.99	02:07.92	食ったことねえな。
M09	02:07.14	02:07.14	嫌いなんですか？
M03	02:08.56	02:09.61	トマト駄目なんですか？
M00	02:09.42	02:09.97	トマト駄目。
M06	02:10.60	02:14.93	でも、トマト駄目っていう人でも何かあのトマトは良いみたいいな結構多いですよ。
M09	02:13.64	02:14.69	へえ。
M00	02:14.37	02:14.81	ええっ？
M06	02:14.93	02:15.56	駄目ですか？
M00	02:15.63	02:17.07	食ったこと無いから、分かんないよ。
M06	02:17.07	02:18.03	あー、本当ですか？
M00	02:18.03	02:19.15	食ってみたいかもしれない。
M06	02:18.84	02:20.96	いや、なかなかシャキシャキしてて、
M00	02:21.07	02:21.62	トマトが？
M06	02:21.15	02:22.54	美味しいです。あそこのは、
M00	02:23.11	02:24.32	超汁溢れそうじゃない？
M06	02:24.99	02:25.77	まあ、確かに。

話にどれだけ積極的か」と定義し、話者の積極性は、アノテータが収録された音声より知覚可能な発話内容の意味や声のトーンなども踏まえて、主観的に1~5の5段階で判断した。ここで、1が盛り上がっていない状態、5が盛り上がった状態とし、3をどちらでもないと設定した。2名のアノテータが評価した盛り上がり度の平均が3.5以上ならば、その発話を盛り上がり発話とした。表4の対話コーパスでは、太字の部分が盛り上がり発話を意味する。盛り上がり発話は、全体の33%程度であった。次にアノテータ間の評価の一致度を式(1)で計算した。

$$\text{一致度} = \frac{\text{二人が盛り上がり度4以上の評価した発話数}}{\text{盛り上がり発話数}} \quad (1)$$

その結果、一致度は37%程度であった。この一致度の低さは、盛り上がり判定の潜在的な難しさを表している。

精度評価には、盛り上がり発話から見た適合率、再現率、F値を用いた。それぞれの計算式を式(2)、(3)、(4)に表す。

$$\text{適合率} = \frac{R}{N} \quad (2)$$

$$\text{再現率} = \frac{R}{C} \quad (3)$$

$$F \text{ 値} = \frac{2 \times \text{適合率} \times \text{再現率}}{\text{適合率} + \text{再現率}} \quad (4)$$

ここで、Rは盛り上がり判定された発話のうち、盛り上がりの正解発話数、Nは盛り上がり判定された発話数、Cは実験対象のすべての発話で、盛り上がりの正解発話数である。

4.2 実験結果

実験では2つの点について評価した。1つめは笑い特徴の中で最も効果的なものが何であるかを検証することであり、2つめはそれぞれの特徴の有効性検証である。実験ではともに、10対話1697発話について10分割交差検定によって評価した。

4.2.1 笑いに関する特徴の検証実験

まず、笑い特徴のうち、どれが有効であったかを検証する。

表 5 実験結果:笑いに関する特徴

利用した特徴	適合率	再現率	F 値
ベースライン (bag of words)	0.704	0.633	0.667
+笑いの有無	0.697	0.633	0.664
+笑いの発生要因	0.697	0.640	0.667
+笑いの大きさ	0.701	0.627	0.662
+笑いの時間	0.713	0.642	0.675
+笑いの人物	0.688	0.624	0.654
+笑いに関するすべての特徴量	0.702	0.651	0.675

Bag-of-wordsのみを利用したものをベースラインとし、それに笑いの各特徴を追加し、精度差を比較する。ただし、この実験では3.1.6節で説明した前発話の情報は利用していない。ここでは、(1) 笑いの有無、(2) 発生要因、(3) 大きさ、(4) 時間、(5) 人物に関する特徴を比較する。(1)は笑いが存在すれば1、存在しなければ0という特徴量である。(2)および(3)の特徴は、3.1.1節で述べたものである。(4)については、時間に関するいくつかの特徴をまとめて扱っている。具体的には、笑いの発生時間や笑いの密度などがこれにあたる。(5)についても同様に、笑いの人数や笑った人物のIDなどをまとめたものを意味する。

表5に実験結果を示す。笑いに関する時間特徴やすべてを利用した場合に、F値が最高になるが、ベースラインと比較して、その向上はわずかであり(1ポイント弱)、単純に笑いに関する特徴のみを追加しても、大幅な精度の向上は得られないという結果となった。これは、すべての発話で笑いが生じているわけではなく、また、盛り上がり箇所必ずしも笑いが生じるとは限らないため^(注1)、笑いのみで盛り上がり判定することは、情報量の面で十分ではないためだと考えられる。

(注1): 実際に笑いが生じたのは、1697発話中で498発話であり、そのうち盛り上がり箇所であったのはその半分程度であった。

表 6 実験結果:対象発話における特徴

利用した特徴	適合率	再現率	F 値
ベースライン (bag of words)	0.704	0.633	0.667
+笑い	0.702	0.651	0.675
+時間	0.711	0.683	0.696
+極性	0.696	0.668	0.682
+接続表現	0.698	0.626	0.660
+時間+極性	0.695	0.661	0.678
+笑い+時間	0.714	0.706	0.710
+笑い+極性	0.716	0.686	0.701
+笑い+接続表現	0.697	0.649	0.672
+笑い+時間+極性	0.701	0.706	0.703
+すべて	0.710	0.711	0.711
+すべて (文脈あり)	0.711	0.688	0.699

4.2.2 各特徴量の有効性検証

続いて、笑い以外の特徴 (3.1.3 節 ~ 3.1.6 節) と組み合わせた場合の有効性について検証する。前節同様、bag-of-words のみによる手法をベースラインとし、各特徴を加えていき、精度差を比較する。表 6 にその実験結果を示す。ベースラインと笑い特徴の精度は、表 5 と同じものである。表中の最後の行「すべて (文脈あり)」以外の特徴は、判別対象となっている発話の情報のみを利用している^(注2)。

最も F 値が良かったのは、すべての特徴を利用した場合で、ベースラインと比較して、5 ポイント程度の向上がみられている。時間や極性などの特徴単体と比べても、それらと笑い特徴を同時に利用する方が F 値が向上しており、さらに、笑いを含んでいない「時間+極性」と比較しても、笑いを含む場合の方が高い F 値を示している。これらの結果から、前節ではみられなかった笑い特徴の有効性が確認できた。一方で、接続表現は、単体としても、笑いの特徴量のみと複合しても、十分な有効性が確認できなかった。これは、今回対象とした接続表現がそもそも対話中にあまり多く含まれていないことに起因している。今後は接続表現辞書の拡充が必要である。

実験結果から、対話の流れを考慮した「すべて (文脈あり)」は有効に機能していない。徳久ら [3] の先行研究などから、文脈の利用は、特に極性との組み合わせで有効に働くのではないかと考えられたが、結果として表れなかった。これは、対話中に「はい」や「うん」といったほぼ極性を含まないあいづちが頻出することが影響していると考えられる。発話と発話の間に極性を持たないあいづちが頻出することによって、談話構造が意図していたものと異なっていたことが、精度の向上に繋がらなかった理由だと考えられる。

5. おわりに

本研究では、談話の盛り上がり判定を行うための手法を提案し、その有効性の検証を行った。本論文では、まず、盛り上がり

行った。談話の盛り上がりを検出するために、笑いに関する特徴を定義し、その他の特徴として、単語の有無 (bag of words) を基本として、時間的な特徴、人間の情緒への影響が考えられる極性、接続表現など特徴を利用した。これらの特徴量を機械学習 (Support Vector Machine) に適用することで、各発話の盛り上がり

を判定する手法を提案した。実験の結果、笑いの特徴については、単体では十分な有効性を確認できなかったが、その他の特徴と統合的に利用することで、その有効性を確認できた。実験では、すべての特徴量を組み合わせることで F 値が 5 ポイント程度向上した。一方で、接続表現については十分な有効性を確認できなかった。

今後の課題としては、今回利用した特徴の洗練化と、談話構造や声の大きさ、高さなどの盛り上がり判定においてによく利用される特徴の導入などが挙げられる。また、今回は笑いの種類の判定 (external, internal, smile, laugh, burst) は人手で行った。これらの自動判定も重要な課題の一つである。

文 献

- [1] 稲葉通将, 鳥海不二夫, 石井健太郎, “語の共起情報を用いた対話における盛り上がりの自動判定”, 電子情報通信学会論文誌, J94-D(1), pp. 59-67, 2011.
- [2] B. Wrede and E. Shriberg, “Spotting “Hot Spots” in meetings: Human Judgments and Prosodic Cues”, In Proceedings of EUROSPEECH, pp. 2805-2808, 2003.
- [3] 徳久良子, 寺島立太, “雑談における発話のやりとりと盛り上がりの関連”, 人工知能学会論文誌, Vol. 21, No. 2, pp. 132-142, 2006.
- [4] 西村良太, 北岡教英, 中川聖一, “音声対話における韻律変化をもたらし要因分析”, 日本音声学会 音声研究, Vol. 13, No. 3, pp. 66-84, 2009.
- [5] 須見康平, 河原達也, 緒方淳, 後藤真孝, “ボッドキャストを対象とした音リアクションイベント検出”, 情報処理学会研究報告, Vol. 2009-SLP-77, No. 24, pp. 132-142, 2009.
- [6] 志水彰, “笑い その異常と正常”, 講談社, 2000.
- [7] 尾崎由美子, “教室の中のコミュニケーションに関する質的研究: 「笑い」の場面を通して”, 一橋大学留学生センター紀要, No. 11, pp. 97-108, 2008.
- [8] Taku Kudo, Kaoru Yamamoto and Yuji Matsumoto, “Applying Conditional Random Fields to Japanese Morphological Analysis”, In Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing, EMNLP-2004, pp. 230-237, 2004.
- [9] Nobuhiro Kaji and Masaru Kitsuregawa, “Automatic Construction of Polarity-tagged Corpus from HTML Documents”, In Proceedings of COLING/ACL (Poster Sessions), pp. 452-459, 2006.
- [10] Hiroya Takamura, Takashi Inui, Manabu Okumura, “Extracting Semantic Orientations of Words using Spin Model”, In Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL2005), pp. 133-140, 2005.
- [11] 中道龍三, 徳久雅人, 村上仁一, 池原悟, “情緒推定の手がかりとなる接続表現の収集”, 電子情報通信学会技術報告, TL, 思考と言語 108(353), pp. 1-6, 2008.
- [12] V. N. Vapnik, “Statistical Learning Theory”, John Wiley & Sons, 1999.
- [13] Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann and Ian H. Witten, “The WEKA Data Mining Software: An Update; SIGKDD Explorations”, Volume 11, Issue 1, 2009.

(注2): すなわち「すべて (文脈あり)」のみ, 3.1.6 節で説明された対話の流れを考慮している。