

対話中の音声言語特徴量に着目した嘘の検出法と日英間比較*

☆角森唯子, Graham Neubig, Sakriani Sakti, 戸田智基, 中村哲 (奈良先端大)

1 はじめに

対話中の嘘¹を見破ることは決して容易ではなく、刑事などは嘘を見抜く高度な技術を身につける必要がある [3]。この問題に対し、機械学習を用いた嘘の自動検出について研究が行われており、英語において一定の成果が得られている [4]。[4]では、英語話者によって収録された嘘を含むコーパス (以下、CSC コーパス) に対して、音響特徴量と言語特徴量を用いて嘘の自動検出を行った結果、チャンスレート (60.2%) を有意に上回る分類精度 (66.4%) が得られることが確認されている。一方で、言語間、文化間によって嘘の言語的な特徴に差異があることも指摘されており [6]、他言語においても嘘の自動検出が有効に機能する保証はない。しかしながら、音響・言語特徴量を両方用いた嘘の自動検出に関する言語間比較は、著者の知る限り行われていない。

本研究では、日本語の嘘を含むコーパスを新たに収集し、これを用いて嘘の検出実験を行う。また、検出に有効な特徴量について日英間比較を行う。

2 日本語偽言コーパスの収集

本研究では CSC コーパスと同様の収録設定の下、日本語で対話収録を行い、書き起こしと嘘ラベルの付与を行う。

2.1 収録設定

嘘が誘発されやすい場面の一例として、面接 [2] における質問者と対象者の 2 名の対話を想定する。収録の手順を以下に示す。

1. 6 項目 (政治や音楽など) に対する、ある“目標プロフィール”に適合するかどうかを調べるための実験と対象者に伝える。
2. 対象者に 6 項目のテストを収録前に受けてもらう。
3. 2 項目が適合、4 項目が不適合となるように操作した結果を対象者に伝える。
4. この実験の本当の目的は“目標プロフィールに適合している”と主張し、質問者を納得させることができる人物を探すことであり、上手く納得させれば賞金があると対象者に伝える。
5. 対象者は、全項目のテスト結果において“目標プロフィールに適合している”と、面接で質問者に主張する。質問者は、目標プロフィールやテスト内容についての知識はなく、いかなる質問をしても構わないとする。

収録した日本語偽言コーパス²の一例を Table 1 に示す。20 代の男女計 2 名を質問者とし、20 代の男女計 10 名を対象者として対話収録を行った。収録した総対話数は 10 対話であり、総時間は約 150 分である。対象者の総発話数は 1069 で、SU³の総数は 1671 である。

Table 1 対話例 (I: 質問者, P: 対象者)

話者	書き起こし文	ラベル
I	音楽に関して、あなたはマッチしていませんか?	
P	はい、マッチしていました。	嘘
I	それはなぜだと思いますか?	
P	えーと、そこそこ答えたからです。	真実
P	小さい頃からずっとピアノをやっていたので。	嘘

Table 2 音響および言語特徴量と個人性の情報

カテゴリ	説明
一般記述	テスト項目, 笑い, 雑音, 言い間違い
文構造	代名詞, 否定, YesNo, 終助詞, 動詞の原形, 遊び言葉, 合図句, 質問, ポジティブ・ネガティブ単語
バラ言語情報	同意, 言いよどみ
F_0	中央値, SU 長に対する中央値の割合
音素継続長	母音, 平均, 最大
パワー	平均, SU の第 1・最終フレーム
個人性	性別, 言いよどみ・合図句の頻度

2.2 ラベル付与

対象者の SU に対してラベル付与を行う方法として“真実”と“嘘”のボタンを用意し、対象者には対話収録中の SU ごとにどちらかのボタンを押してもらう。なお、一部分でも嘘が含まれる SU は、嘘として定義する。それぞれのボタンが押された時間と書き起こしから、SU 毎にラベル付けを行う。1671 SU のうち真実ラベルが付与された SU は 1401 で、嘘ラベルが付与された SU は 270 である。

3 嘘検出のための特徴量

本研究は先行研究 [4] に基づき、SU 毎に言語特徴量と音響特徴量をそれぞれ抽出する。先行研究で有効と示された特徴量のうち、英語に特化しておらず日本語でも使用可能な特徴量に“終助詞の数”を新たな言語特徴量として加えている。抽出した特徴量を Table 2 に示す。

- 音響特徴量
音響特徴量として、基本周波数 (F_0) とパワー、音素継続長を用いる。 F_0 の抽出には Snack Sound Toolkit [8]、音素継続長の抽出には Kaldi [7] を用いる。
- 言語特徴量
人手による書き起こし文に対して MeCab [9] で形態素解析を行い、得られた形態素から言語特徴量を抽出する。雑音は、咳やマイクとの接触音など対象者によって生成されたものに限定する。ポジティブ・ネガティブ単語の頻度の抽出には、単語感情極性対応表 [10] を用いる。
- 個人性
対象者の性別と、1 対話中の全 SU 数に対する合図句・言いよどみの頻度を個人特徴量とする。

*Detecting deception with speech and language features during interaction and comparison between Japanese and English. by TSUNOMORI, Yuiko, NEUBIG, Graham, SAKTI, Sakriani, TODA, Tomoki, NAKAMURA, Satoshi (NAIST)

¹ “騙そうとする意図をもった事実とは異なる言語的陳述”と定義する [5]。

² <http://ahclab.naist.jp/resource/ja-deception/>

³ 発話を句読点や休止で区切ったスラッシュユニットの略称

4 実験的評価

前述した音響特徴量と言語特徴量を合わせた特徴量を用いて、SU に対して嘘の検出を行う。2 値分類するための機械学習手法として、Bagging[1] を用いる。嘘の検出率の評価には、1670 SU を学習用のデータにし、残り 1 SU をテスト用のデータにする一個抜き交差検定を用いる。

4.1 結果と考察

Table 3 に、各特徴量を用いた際の嘘の検出率と F 値を示す。“日本語”は本研究における日本語の嘘検出率，“英語”は先行研究 [4] に基づき CSC コーパスの一部を用いて英語の再現実験を行った際の嘘検出率である。“人間による検出”とは、被験者と別の人間に同じ対話データを聞いてもらい、嘘の検出を行ってもらった結果である。

日本語では音響特徴量+個人性を用いた場合が最も精度が高く、言語特徴量のみ比べて嘘検出率が約 7% 高かった。全特徴量を用いた場合でも、音響特徴量+個人性、音響特徴量のみ検出率より低い精度であった。英語でも音響特徴量+個人性を用いたものが最も精度が高く、チャンスレートに比べて約 17% 高かった。また、日英ともに言語特徴量を用いた場合の検出率はチャンスレートとほぼ同等であった。先行研究 [4] においても、言語特徴量のみを用いた場合はチャンスレートに対して約 1% 程度の向上しか確認されていないことから、今回用いた言語特徴量は単体では嘘検出にあまり有効ではないと言える。さらに、言いよどみと合図句の頻度、性別などを追加して検出を行った場合に精度の向上が見られたことから、日本語においても嘘検出にはそのような個人性が影響を及ぼすことがわかった。また、人間による検出はチャンスレートとほぼ同等の精度であり、どの特徴量を用いた検出もこれを上回る精度であった。なお、フィッシャーの正確確率検定の 1% 水準を用いて特徴量間の違いを比較したところ、日英ともにチャンスレートと音響特徴量、音響特徴量+個人性、音響特徴量+言語特徴量、音響特徴量+言語特徴量+個人性で p 値が 0.01 以下となり、有意差が見られた。

Table 4 に、話者ごとに嘘検出を行った場合に、検出率が高かった話者と低かった話者の対話例を示す。SN は雑音、MP は言い間違いのことを指す。質問者がテストの出来について質問した際の対話であり、A は検出率が高かった話者、B は低かった話者である。A は雑音や言い間違いが多く、声も上ずっていた。その一方で、B は真実の発話に比べて際立った変化は見られなかった。

4.2 特に有効な特徴量

最良優先探索を用いて、分類率が学習データに対して最大となるように特徴量選択を行った。その結果、特に有効とされた特徴量を Table 5 に示す。音響特徴量に関しては F_0 の中央値、母音の平均継続長、パワーの最終フレームが日英ともに有効であることが確認された。パワーの最終フレームについては、一般に言われている感情の変化は語尾に出やすいという特徴を捉えていると考えられる。言語特徴量に関しては日英で大きく異なっており、英語では雑音、代名詞、Yes と No の有無が有効とされた。日本語では動詞の原形の有無のみであったことから、今回用いた言語特徴量はあまり貢献しないと考えられる。

Table 3 嘘の検出率

特徴量	検出率 / F 値 (%)	
	日本語	英語
チャンスレート	83.8/0.0	71.4/0.0
音響	90.5/60.2	86.8/74.5
言語	84.2/7.6	71.4/14.7
音響+個人性	90.7/61.4	88.1/77.7
言語+個人性	85.2/31.5	76.8/52.9
音響+言語	89.9/56.9	86.8/74.6
音響+言語+個人性	90.2/58.1	87.8/77.2
人間による検出	83.0/28.4	

Table 4 対話例 (A : 高検出率, B : 低検出率)

対象者	書き起こし文
A	SN は一、そうですね、まあ MP 七八割は答えれたかなっていう位ですね。
B	たぶん大丈夫だと思います。

Table 5 特に有効な特徴量

カテゴリ	英語	日本語
言語	雑音, 代名詞, YesNo	動詞の原形
個人性	合図句の頻度	
F_0	中央値	中央値
音素継続長	平均, 母音	母音
パワー	第 1・最終フレーム, 平均	最終フレーム

5 おわりに

本稿では、日本語の嘘を含むコーパスを収集し、嘘の自動検出を行った。英語に対する先行研究で有効性が確認されている特徴量を用いて分類を行うことで、ほぼ同等の精度が確認された。言語特徴量に関しては、英語で有効とされた特徴量が日本語ではあまり有効でないことがわかった。今後は、言いよどみの種類や対話行為を表すタグを新たな特徴量として加えるなどして、日本語において有効な特徴量を検討することで、さらなる検出精度の向上を目指す。

参考文献

- [1] L. Breiman. Bagging predictors. *Machine learning*, Vol. 24, No. 2, pp. 123–140, 1996.
- [2] B. M. DePaulo, J. J. Lindsay, B. E. Malone, L. Muhlenbruck, K. Charlton, and H. Cooper. Cues to deception. *Psychological bulletin*, Vol. 129, No. 1, p. 74, 2003.
- [3] P. Ekman. 暴かれる嘘. 誠信書房, 1992.
- [4] J. B. Hirschberg, S. Benus, J. M. Brenier, F. Enos, S. Friedman, S. Gilman, C. Girand, M. Graciarena, A. Kathol, L. Michaelis, et al. Distinguishing deceptive from non-deceptive speech. *Proc. Eurospeech*, 2005.
- [5] R. Hopper and R. A. Bell. Broadening the deception construct. *Quarterly Journal of Speech*, Vol. 70, No. 3, pp. 288–302, 1984.
- [6] V. Pérez-Rosas and R. Mihalcea. Cross-cultural deception detection. *Proc. ACL*, pp. 440–445, 2014.
- [7] D. Povey, A. Ghoshal, G. Boulianne, L. Burget, O. Glembek, N. Goel, M. Hannemann, P. Motlicek, Y. Qian, P. Schwarz, J. Silovsky, G. Stemmer, and K. Vesely. The kaldi speech recognition toolkit. In *IEEE 2011 Workshop on Automatic Speech Recognition and Understanding*. IEEE Signal Processing Society, 2011.
- [8] K. Sjolander. Tcl/tk snack toolkit. 2004.
- [9] 工藤拓, 山本薫, 松本裕治. Conditional random fields を用いた日本語形態素解析. 情報学 NL 研報, pp. 161–173, 2004.
- [10] 高村大也, 乾孝司, 奥村学. スピンモデルによる単語の感情極性抽出. 情報処理学会論文誌, Vol. 47, No. 2, pp. 627–637, 2006.