NAIST-IS-MT1651094

修士論文

欺瞞検知における統計的学習手法の有効性に関する研究

細見 直希

2018 年 3 月 15 日

奈良先端科学技術大学院大学
情報科学研究科 情報科学専攻

本論文は奈良先端科学技術大学院大学情報科学研究科に
修士（工学）授与の要件として提出した修士論文である.

細見 直希

審査委員：

中村 哲 教授	（主指導教員）
松本 裕治 教授	（副指導教官）
Sakriani Sakti 特任准教授	（副指導教官）
吉野 幸一郎 助教	（副指導教官）

欺瞞検知における統計的学習手法の有効性に関する研究*

細見 直希

内容梗概

欺瞞はコミュニケーションにおいて利益獲得や不利益回避のためにしばしば用いられるが，受け手にとっては不当に不利益を被る恐れがあるため，これを検知することが望まれる．しかし，欺瞞/非欺瞞の発話における特徴の差は僅かであり，加えて，人間は発話の真偽に関わらず発話を真実とみなす傾向 (真実バイアス) などのバイアスを有しているため，欺瞞を正確に検知することは容易ではない．実際に，人間による欺瞞検知の正解率はチャンスレベル程度であることが複数の研究で示唆されている．このような問題に対して，人間のようなバイアスを持たず，また人間にとって把握することが難しい特徴を用いることができる計算機を利用することで，高精度な欺瞞検知を実現できる可能性がある．特に，統計的学習手法 (機械学習) の一種である教師あり学習によって，大量の欺瞞/非欺瞞の発話の事例を計算機に学習させることで，人間より高精度な欺瞞検知を行える識別器を構築可能なことが先行研究で示唆されている．本研究では，人間同士の欺瞞を含む音声対話を対象に，人間と統計的学習手法による識別器のそれぞれで欺瞞検知実験を実施した．その結果，本課題においても人間が真実バイアスを示すこと，母語やモーダルに関わらず検知能力がチャンスレベル程度であることを確認した．一方で，このような人間にとって困難な課題に対して，音声に含まれる感情関連の特徴や文の分散表現を組み合わせて学習させた識別器を用いることで，人間及びチャンスレベルと比較して有意に高精度な欺瞞検知が実現可能であることを確認した．

キーワード

欺瞞検知，真実バイアス，統計的学習

*奈良先端科学技術大学院大学 情報科学研究科 情報科学専攻 修士論文, NAIST-IS-MT1651094, 2018 年 3 月 15 日.

A Study on the Effectiveness of Statistical Learning Method for Deception Detection*

Naoki Hosomi

Abstract

Deception is often used for making profit or avoiding disadvantages in communication. Since there is a possibility that the person who received deceptive message suffer disadvantages, it is expected to detect. However, it is difficult for humans to detect any deception because differences between deceptive and non-deceptive speech are small. Humans also have some bias which tends to regard a message as a truth regardless of whether it is true or not (i.e. Truth-Bias). In fact, it is suggested in some existing works that the ability of deception detection by humans is about chance level. In order to solve such problems, we propose to use a computer for deception detection by building a classifier based on supervised learning which is a method of statistical learning (machine learning). It is expected that we can construct an accurate classifier, which can capture the characteristics of deception by letting the computer learn a large amount of deceptive/non-deceptive speech.

In this study, we conducted a task of deception detection with humans and computers by using dialogue corpus which includes deception. As a result, we confirmed that human subjects also have Truth-Bias in this task, and their detection accuracy is about the chance level regardless of their native language or given modals. On the other hand, the classifier that combines emotion-related features of speech and word embeddings of sentences shows significantly high performance than humans and chance level even though it is difficult tasks for humans.

*Master's Thesis, Department of Information Science, Graduate School of Information Science, Nara Institute of Science and Technology, NAIST-IS-MT1651094, March 15, 2018.

Keywords:

Deception detection, Truth-Bias, Statistical Learning

目次

図目次

第 1 章　緒言

1.1　背景

欺瞞 (嘘) はコミュニケーションにおいて利益獲得や不利益回避のためにしばしば用いられるものである．欺瞞が生じる状況は様々であるが，話し手が不当な利益を獲得し，それによって受け手が不利益を被る例として就職などの面接があり，実際に多くの人は職を得るためならば嘘をつくことが報告されている [25, 13, 33]．欺瞞の検知が可能になれば，このように受け手が不当な不利益を被ることを回避できるが，人間にとってそれが容易ではないことは過去の研究から明らかになっている．Bond らは過去の 200 以上の様々な欺瞞検知に関する研究のメタ分析を行い，特別な訓練を受けていない人間による欺瞞検知の平均正解率が 54% 程度だったことを報告している [3]．また，Levine らは真実と欺瞞のメッセージを収録したビデオを用いて，実験協力者に対して呈示するメッセージに含まれる真実の割合を変化させて欺瞞を検知させる実験を行うことで，人間による検知の正解率が呈示される真実の割合に依存することを示している [14]．つまり，人間が欺瞞検知を試みても，一般にその正解率はチャンスレベルを大きく上回らないことが示されている．人間による欺瞞検知がうまくいかない原因については様々な主張がなされているが，原因の 1 つに人間がバイアスを有していることが挙げられる．例えば，人間は真実バイアスと呼ばれる相手の発言を実際の真偽に関わらず真実だと判断する傾向 [17] や，相手を嘘つきだと思って見ているとそのように見えてしまう確証バイアス [22] などを有していることが知られている [34]．また，欺瞞時に特徴的な現象を人間が把握できない，欺瞞とは関連しない手がかりに着目してしまうことも欺瞞検知が困難である原因だと考えられる．そこで本研究では，人間が有しているようなバイアスの影響がない計算機を利用し，統計的学習手法（機械学習）を用いて欺瞞検知を行うことを提案する．欺瞞検知には生体信号や顔表情を用いることで高い検知性能が得られるという報告があるが [31]，検知の対象者に専用の計測機器を繋いだり，鮮明に顔表情を取得するために被験者の行動を制約する方法では，自由なコミュニケーション中の欺瞞を検知することが難しい．そこで本研究では，そのような特別な手続きを必要としない音声言語を用いた欺瞞検知を対象として，人間よりも高性能な検知性能を実現する識別器の構築を目指す．統計的学習手法の一種である教師

あり学習では，人間の有するタイプのバイアスの影響を受けずに，純粋に事例が持つラベルと特徴量からの学習を行うことができる．つまり，欺瞞発話の声に含まれる音響的な特徴量や書き起こし文の言語的な特徴を計算機が学習し，検知性能を最大化するような特徴量の選択を行うことができる．このように，計算機であれば人間よりも正確な欺瞞検知を実現できることが期待される．

1.2 目的

本研究では人間同士のコミュニケーションを対象に，音声言語を用いて欺瞞検知の為の識別器構築を目指す．また，その性能の検証の為に人間及び提案する識別器に対して同様の欺瞞検知実験を実施し，その結果を分析することで，それぞれの正解や誤りの傾向の解明を目指す．近年では，さまざまな状況で人間と計算機が対話するための対話システムが普及しており，提案手法が人間と同等以上の欺瞞検知を行うことができれば，対話システムに組み込むことで，より人間らしいコミュニケーション実現の一助になると考えられる．

1.3 本論文の構成

本論文の構成は以下の通りである．第 2 章で関連研究を挙げ，第 3 章では実験に用いるデータセットであるコーパスについて説明する．第 4 章で本研究の提案手法について説明し，第 5 章では人間と提案手法に対する評価実験と結果及び考察を述べる．第 6 章では，本論文のまとめと今後の課題について述べる．

第 2 章　関連研究

本章では，まず 2.1 節で一般的な欺瞞の定義について述べ，2.2 節で人間及び機械学習による欺瞞検知に関する関連研究について述べる．また，2.3 節でこれらの関連研究を踏まえて，本研究で取り組む課題について明らかにする．

2.1　欺瞞とは

欺瞞については様々な研究者が定義をしている．例えば Krauss は「欺瞞者が虚偽であるとみなす信念や理解を他者に抱かせようとする行為」[11]，Vrij は「メッセージの伝達者が真実でないとみなす信念を他者に形成しようとする，事前の通告の無い成功する可能性も失敗する可能性もある意図的試み」[31] と定義した．ここで重要な点は，欺瞞が真実ではなくかつ意図的な行為であるということであり，虚偽記憶や無知，誤りなどについてはここでは欺瞞として扱わない．また，「欺瞞」の関連語である「嘘」は厳密には意味が異なる語であるが，Vrij[31] に倣い，同様の意味を成す語として文脈に応じてそれらを適宜用いる．

2.2　欺瞞検知

2.2.1　社会心理学における欺瞞検知

人間による欺瞞検知の正解率については，個々の研究によりばらつきはあるものの，一般的には高くないことが Bond らのメタ分析によって示されている [3]．このような検知能力の低さは，欺瞞検知の為の特別な訓練を受けていない一般人に限らず，尋問を行う警察官などの専門家であっても同様だと報告されている [18, 10]．また，検知能力の低さは初対面の相手を対象とした場合に限らず，友人などの親しい人から欺瞞を検知する場合でも同様であり，親密さと欺瞞検知の正確性について関連はみられていない [2]．このように欺瞞検知が人間にとって困難である原因として，人の判断が真実バイアスや確証バイアスなどのバイアスによって左右されることが挙げられ，加えて，そもそも人間にとって欺瞞の特徴を捉えることが難しい

ことも，人間が正確に欺瞞検知を行えない原因だと考えられる．そして，これまでの研究から，あらゆる事例で顕著である欺瞞の手がかりは存在しないことが示唆されており，欺瞞と真実の発話に差があったとしても，その差は概して非常に小さいと考えられている．また，たとえ同じ状況で嘘をつく場合であっても人によって欺瞞時の特徴は異なり，同様に，同じ人が嘘をつく場合でも状況によって特徴が異なる．さらに，一般的に人は欺瞞の手がかりについて誤った信念を持っており，欺瞞と関連しない手がかりに基づいて欺瞞検知を試みることや，実際には欺瞞と関連のある手がかりについて関連することに気づいていないことも欺瞞検知が困難な要因である [31].

ところで，ある特定の状況における欺瞞，またはある集団の欺瞞の特徴が，別の状況や別の集団で有効であるとは限らないが，個別の研究では欺瞞時の音声言語について様々な特徴が報告されている [5, 32, 4]．欺瞞検知の手がかりとなる特徴については，単独の特徴のみを用いるのではなく，いくつかの特徴を組み合わせることで，より正確な欺瞞検知が実現できる可能性がある．Vrij はどのような手がかりの組み合わせが良いか明確な答えは存在しないが，非言語的な手がかりをそれぞれ個別に分析するよりも，その組み合わせを用いた方が真実と欺瞞をより正確に識別できると述べている [31].

2.2.2 機械学習による欺瞞検知

2.2.1 で述べたように，人間が欺瞞検知を試みる場合，一般人，専門家に関わらず真実バイアスなどのバイアスによって判断に偏りが生じる為，正確な欺瞞検知を行うことは難しい．また，基本的に欺瞞と真実の発話の特徴の差は僅かであり，その差を人間が検知することは難しく，検知精度向上の為に特徴を組み合わせて判断することも困難である．そこで，こうした人間が持つバイアスに左右されず，特徴の組み合わせを考慮することが可能である計算機を用いて欺瞞検知を行うことで，高性能な識別器の実現が期待できる．特に，統計的学習手法である教師あり学習では，大量の欺瞞，真実 (非欺瞞) 発話の事例を計算機に与えることで，識別に有効な差異を獲得し，より高精度な欺瞞検知を行うことができる可能性がある．

人間同士の対話からの音声言語を用いた統計的学習手法による欺瞞検知につい

ては，いくつか先行研究が行われており，統計的学習に用いるデータセットとして Hirschberg らが標準アメリカ英語を母国語とするインタビュイーの欺瞞を含む発話を約 7 時間分収録したコーパス (CSC コーパス) を作成している [8]．このコーパスに対して様々な特徴量や識別器を適用して欺瞞発話を検知する取り組みがなされているが，チャンスレベルに対して最も性能が改善したものとしては Ripper rule-induction classifier を用いて語彙，音響，韻律及び話者依存の特徴量による識別器を構築したもので，正解率がチャンスレベルに対して約 6% 改善したことが報告されている [8]．また，人間と対話エージェントのインタビューを収録したコーパス [28] を対象とした研究では，人間の音声特徴から推定された感情スコアと Support Vector Machine(SVM) による欺瞞検知実験により，音声から得られる感情情報が欺瞞検知に有効であることを示唆している [1]．さらに，近年では Levitan らが標準英語及び北京語を母国語とする英語話者を対象とした収録時間 120 時間を超える大規模な対話形式のコーパス (CXD コーパス) を作成しており [15]，このコーパスに対して，Mendels らは音声に含まれる感情関連の特徴量を学習させた全結合ニューラルネットワークと，GloVe[23] を用いて作成した分散表現を特徴量として学習させた Bidirectional LSTM モデルを組み合わせることで，F 値が 0.64(Precision=0.67, Recall=0.61) の識別器を構築したことを報告している [19]．ただし，このコーパスは現在入手できないため再現が困難である．

このように，音声言語からの統計的学習手法による欺瞞検知はいくつかの成果が報告されており，ルールベースによる識別器の他，音声に含まれる感情関連の特徴量や発話文の分散表現などを特徴量として大量に学習させることで，性能の良い識別器が構築可能であることが示唆されている．

2.3 本研究で取り組む課題

欺瞞検知について，これまでの研究から人間は一般に真実バイアスなどのバイアスを示し，その検知能力は高くないと考えられる．また，欺瞞時の人間の音声言語の特徴については，あらゆる状況で有効な特徴は無いと考えられる一方で，個別の研究では欺瞞検知に有効な特徴が様々報告されており，特徴を組み合わせることで欺瞞検知の性能を向上させることができる可能性が示唆されている．統計的学習手

法による音声言語を用いた欺瞞検知の研究では，欺瞞と非欺瞞の発話を多数含むコーパスを用いた事例の学習，モデル設計の工夫によって，性能の高い識別器が構築可能であることが報告されている．

本研究では音声対話を対象に，人間と統計的学習手法による識別器を用いて欺瞞検知実験を行い，検知性能や人間が有するバイアスの影響について比較検証を行う．

第 3 章　欺瞞発話を含むコーパス

本章では，3.1 節で本研究で用いるコーパスについての説明，3.2 節でその内容についての分析を行う．

3.1　CSC コーパス

現在，入手可能な欺瞞を含むインタビューの場面が収録されたコーパスとして CSC Deceptive Speech(CSC コーパス) がある [8]．これは標準アメリカ英語を話す男性 16 人，女性 16 人の合計 32 人の実験協力者であるインタビュイーを対象に各 25 分から 50 分間のインタビューを実施し，合計約 7 時間分の実験協力者の発話を収録したコーパスである．このコーパスには録音された音声とそれを書き起こしたテキスト，回答の真偽についてのラベル情報が含まれている．コーパスに含まれる発話の例は表 3.1 の通りである．また，このコーパスの収録の条件については，以下の通りである [6].

1. 実験協力者のパフォーマンスがある目標プロフィールに適合するかどうか調査することを告げられる．
2. 実験協力者は 6 項目（music, interactive, survival skills, food and wine knowledge, NYC geography, civics）について事前の調査を受ける．
3. この 6 項目の回答について，2 項目が目標プロフィールに適合，4 項目が不適合となるよう実験者が調査結果を操作し，その結果を実験協力者に伝える．
4. 実験協力者には本人のインタビュー結果（4/6 項目が不適合）が告げられ，次のインタビューでインタビュアーに対して彼ら自身が目標プロフィールに適合していることを納得させられれば，賞金がもらえることを告げられる．
5. 次のインタビューで，実験協力者はインタビュアーに対して全項目の調査の結果において目標プロフィールに適合していることを主張する．実験協力者はインタビュアーからの各質問に対する自身の回答ついて，真実と欺瞞を示す 2 つのペダルのうちの 1 つを押すことによって，回答が真実であるか欺瞞であるかを示す．

表 3.1 ：CSC Deceptive Speech に含まれる発話例

発話	ラベル
well, yeah, there's a chance.	真実
uh actually, I did well. excellent.	欺瞞

コーパス中の発話に対する真偽のラベルは，実験協力者が発話中に自らアノテーションしたラベル情報に基づいて付与されている．本研究では，句読点や休止によって機械的に分割されたものを発話単位として以降の実験に利用する．

3.2 コーパスの内容

本節ではコーパスの内容について分析したものを示す．以下の表 3.2 から表 3.4 は，コーパス中のインタビュイーの全発話を対象として，ストップワード*を除く真実，欺瞞発話における n-gram 上位 10 個の出現頻度を表している．ここで，真実と欺瞞は発話数が均衡でないため，ここでの出現頻度は，真実，欺瞞それぞれの n-gram の総数で割って正規化したものを示している．

*is や and などの，どのようなテキストでも出現するごくありふれており，真実，欺瞞発話の区別に有益な情報を含んでいないと考えられる単語

表 3.2　1-gram 頻度

欺瞞		真実	
1-gram	頻度	1-gram	頻度
know	0.047978	know	0.046539
like	0.044012	um	0.041455
um	0.036177	uh	0.041064
uh	0.030857	like	0.034513
don	0.022538	don	0.021901
really	0.020700	just	0.017697
just	0.019540	really	0.015252
think	0.016734	ve	0.013786
mean	0.015864	think	0.013395
did	0.013252	mean	0.012026

表 3.3　2-gram 頻度

欺瞞		真実	
2-gram	頻度	2-gram	頻度
don know	0.009645	don know	0.009720
new york	0.004626	new york	0.004055
know like	0.003543	high school	0.003862
don think	0.002952	know like	0.002961
don really	0.002756	don think	0.002768
um know	0.002559	little bit	0.002639
like know	0.002460	don really	0.002446
high school	0.002460	um know	0.001931
just like	0.002067	like know	0.001738
uh know	0.001968	just like	0.001674

表 3.4　3-gram 頻度

欺瞞		真実	
3-gram	頻度	3-gram	頻度
empire state building	0.001214	new york city	0.001037
know don know	0.000971	uh don know	0.000638
new york city	0.000850	um don know	0.000638
um don know	0.000850	hand eye coordination	0.000558
don know like	0.000729	uh don think	0.000479
went high school	0.000607	don know don	0.000479
don really know	0.000607	don know ve	0.000479
don know know	0.000607	empire state building	0.000479
mean don know	0.000486	high school um	0.000399
don know really	0.000486	know don know	0.000399

表の通り，出現頻度の高い n-gram は真実，欺瞞ともに同様のものが出現している．加えて，欺瞞発話に特有な n-gram についてもその出現の割合は非常に低く，それらに基づいて欺瞞検知をするのは困難であると考えられる．

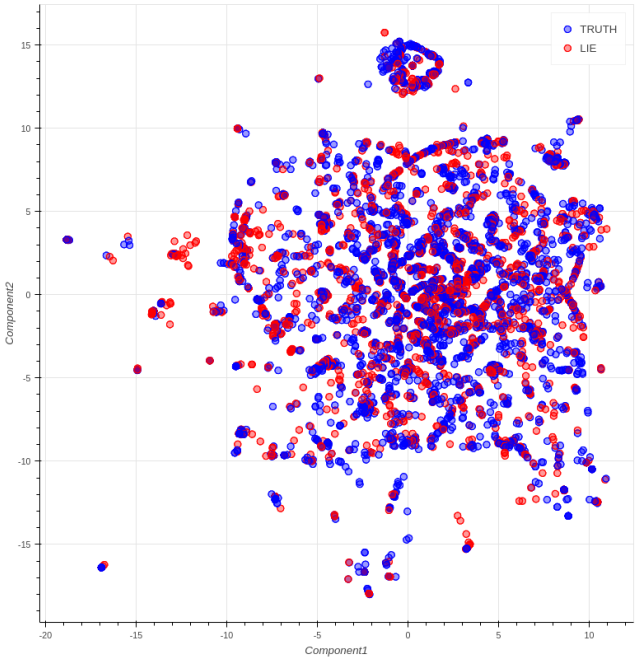

図 3.1 ： 真実と欺瞞発話の分布

また図 3.1 は，発話内容の書き起こし文とラベルに基づいた発話の分布である．ここでは，発話文を One-hot ベクトル†に変換し，このベクトルを t-SNE(4.1.3 参照) を用いて次元数を 2 次元まで削減して可視化したものであり，図の通り，真実と欺瞞発話は概ね同様の分布をしていることが確認された．以上の通り，CSC コーパスにおいて欺瞞と真実の発話には同様の語彙が含まれており，またそれぞれに特有な語彙の出現頻度は低く，その分布も概ね同様であった．

†語彙と同数の次元を事前に用意し，発話に含まれている単語に対応する次元を 1，それ以外を 0 で表現したベクトル

第 4 章　機械学習による欺瞞検知

本章では提案する統計的学習手法において用いる特徴量を 4.1 節，識別器のモデルを 4.2 節，ハイパーパラメータの最適化について 4.3 で説明する．

4.1　特徴量

ここでは本研究で利用する特徴量と，特徴量の次元数を削減する手法について述べる．

4.1.1　音響特徴量

2.2.1 の通り，先行研究では音声に含まれる感情関連の特徴が欺瞞検知に有効であることが示唆されている．そこで本研究では過去の感情認識コンペティションの為に用意された 384 次元の特徴量 [27](IS09) を用いる．これらの特徴量はコーパス中の音声ファイルから OpenSMILE[7] を利用して抽出した．この特徴量には以下のような特徴が含まれている．

表 4.1　：音響特徴量 (IS09)

LLD (16 · 2)	Functionals(12)
(Δ) ZCR	mean
(Δ) RMS Energy	standard deviation
(Δ) F0	kurtosis, skewness
(Δ) HNR	extremes: value, rel. position, range
(Δ) MFCC 1-12	linear regression: offset, slope, MSE

4.1.2 言語特徴量

本研究では識別器を構築，学習させる際に分散表現を用いることで低次元で密な文の意味表現を学習し，これを言語特徴量として利用する．欺瞞を含む対話は一般的な状況ではない為，ここでは学習済モデルは用いずコーパスからの学習を行う．分散表現を構築するには，まずコーパス中の全ての単語 (または n-gram) に対して固定次元のベクトルをランダムに割り当てる．そして，このベクトルを用いて文章のラベルを予測できるように，それぞれの単語に割り当てられたベクトル値を更新することで意味表現ベクトルの学習を行う．

本研究では Joulin らによって提案された fastText[9] のアーキテクチャを利用して分散表現を構築する．fastText のアーキテクチャは図 4.1 のように Word2Vec の Continuous Bag of Words モデル [20] とよく似た構造をもち，文の bag of n-grams を入力として埋め込み層を通して分散表現 $\boldsymbol{e_i}$ を構築し，それらのベクトルを平均化した $\boldsymbol{m}$ を識別器の入力として用いる．識別器はこれを入力としてラベルを予測する全結合ニューラルネットワークである．本研究ではこれを真実と欺瞞の 2 クラス分類に用い，識別器の出力である $\hat{y}$ と正解ラベル y を用いて N 個の文章に対して交差エントロピー

$$-\frac{1}{N}\sum_{n=1}^{N} y_n \log\left(\hat{y_n}\right) \tag{4.1.1}$$

を最小化するように学習を行う．本手法では文のラベル情報を利用して分散表現を構築し，同時に識別器の重みを学習するため，真実と欺瞞の識別に最適化された分散表現を学習することが期待できる．実際に，文章の評判分析や文に対するタグ予測の課題において高い性能を発揮することが報告されている [9].

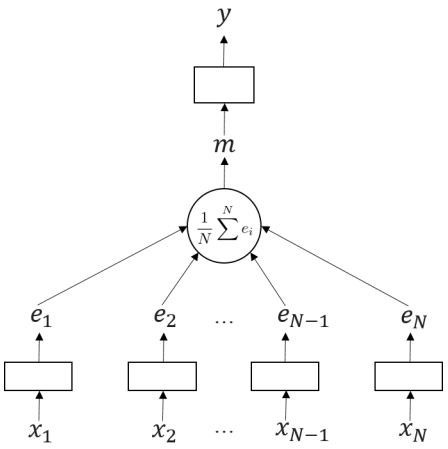

図 4.1 ： fastText のアーキテクチャ

4.1.3 特徴量の次元圧縮

本研究で扱う特徴量は 3 次元より大きい為，特に特徴量を可視化して分析する際には，次元数を 2 次元もしくは 3 次元に削減する必要がある．そこで本研究では特徴量の次元を削減する手法として t-distributed stochastic neighbor embedding(t-SNE)[16] を用いる．t-SNE では次元削減の際に非線形変換を行うため，一般的な主成分分析などの線形変換を用いる手法よりも，高次元での重要な構造をできる限り保ちながら次元を削減できると考えられる．この手法で次元を削減する方法は以下の通りである．

t-SNE において，ある高次元空間における点 $\boldsymbol{x}_i$ とその近傍の点 $\boldsymbol{x}_j$ との類似度は，点 $\boldsymbol{x}_i$ を中心とするガウス分布の $\boldsymbol{x}_i$ の近傍で $\boldsymbol{x}_j$ が選ばれる条件付き確率密度を用いて以下のように計算される．

$$p_{ij} = \frac{p_{j|i} + p_{i|j}}{2N} \tag{4.1.2}$$

$$p_{j|i} = \frac{\exp\left(-\left\|\boldsymbol{x}_i - \boldsymbol{x}_j\right\|^2 / 2\sigma_i^2\right)}{\sum_{k \neq i} \exp\left(-\left\|\boldsymbol{x}_i - \boldsymbol{x}_k\right\|^2 / 2\sigma_i^2\right)}, \quad p_{i|i} = 0 \tag{4.1.3}$$

このとき，$\boldsymbol{x}_i$,$\boldsymbol{x}_j$ よりも低次元空間のマップ点 $\boldsymbol{y}_i$,$\boldsymbol{y}_j$ の類似度は自由度 1 の Student の t 分布を用いて，以下のように表される．

$$q_{ij} = \frac{\left(1 + \|\boldsymbol{y}_i - \boldsymbol{y}_j\|^2\right)^{-1}}{\sum_{k \neq i} \left(1 + \|\boldsymbol{y}_i - \boldsymbol{y}_k\|^2\right)^{-1}}, \quad q_{ii} = 0 \tag{4.1.4}$$

t-SNE では p_{ji} と q_{ji} の Kullback–Leibler ダイバージェンス (KL ダイバージェンス)

$$C = KL\left(P \| Q\right) = \sum_{i \neq j} p_{ij} \log \frac{p_{ij}}{q_{ij}} \tag{4.1.5}$$

が最小となるような $\boldsymbol{y}_i$,$\boldsymbol{y}_j$ を求めることで，高次元での構造をできる限り保持したまま次元を削減している．

このように t-SNE は高次元，低次元空間における 2 点間の類似度をそれぞれ確率密度を用いて表し，それらの KL ダイバージェンスを最小化する低次元空間でのマップ点を求めることで，次元削減を実現している．ただし t-SNE の計算量は $O(N^2)$ で計算に時間が掛かる為，本研究では計算過程で近似を用いて計算量を $O(NlogN)$ に削減した Barnes-Hut t-SNE[30] を用いた．

4.2 識別器

本研究ではコーパス中の音響特徴量や言語特徴量を用いて識別器を学習することで，欺瞞と真実を識別する識別器を構築する．本節では，提案手法で用いる識別器のモデルについて述べる．

4.2.1 多層パーセプトロン

パーセプトロン (図 4.2) は Rosenblatt によって考案されたアルゴリズムである [26]．本節では 2 クラス分類に焦点をあてて．ある特徴量を入力 $\boldsymbol{x}$，その次元と対応する重みベクトルを $\boldsymbol{w}$ とし，入力全体を $u = w_1 x_1 + \cdots + w_n x_n$ と表現する．正例負例をそれぞれ 1,0 とすると，パーセプトロンでは入力の線形和 u が閾値 θ を以下の場合は負例，θ より大きい場合は正例というように線形識別を行う．ここで，

式 4.2.1 の $h(u)$ のような入力信号の線形和を出力値に変換する関数を一般に活性化関数と呼ぶ.

$$h(u) = \begin{cases} 0\,(u \leq \theta) \\ 1\,(u > \theta) \end{cases} \tag{4.2.1}$$

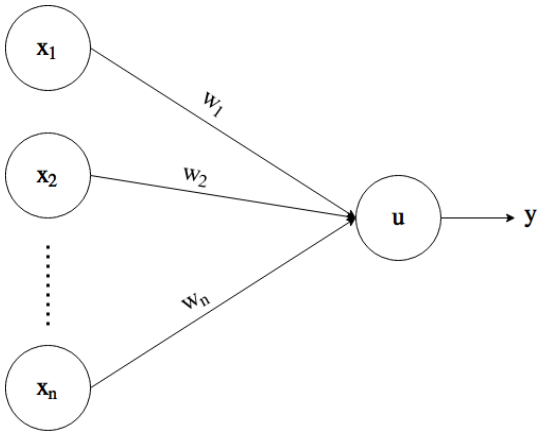

図 4.2 ：パーセプトロン

多層パーセプトロン (図 4.3) は，このようなパーセプトロンの層を重ねることで，非線形な関数を表すことが可能な順伝搬型ニューラルネットワークである．音響特徴量，言語特徴量，両方を結合した特徴量のそれぞれを学習する各識別器のハイパーパラメータは，バリデーションデータに対して Accuracy が最大化されるようにベイズ最適化 (4.3 節) を用いてチューニングした．本実験においてハイパーパラメータの探索範囲は多層パーセプトロンの中間層の層数 (2 から 5)，ユニット数 (128, 192, 256), ドロップアウトレート*(0.1 から 0.5) とした．活性化関数には標準シグモイド関数

$$h\,(x) = \frac{1}{1 + \exp\,(-1)} \tag{4.2.2}$$

を用いた．標準シグモイド関数は $\lim_{x\to-\infty} h(u) = 0,\ \lim_{x\to\infty} h(u) = 1$ であり，この関数は正規分布の累積分布関数とグラフの概形が概ね一致しているため，本研

*過学習（訓練データに対しては適合するが，未知のデータに対しては適合しない状態) を防ぐために，学習時に無効にするユニットの割合

究では標準シグモイド関数値を確率と捉え，発話が真実である確率として解釈する．従って，この値が 0.5 以上であれば発話を真実，それ未満であれば発話を欺瞞と分類する．学習時には K 個のサンプルに対して n をサンプル番号とすると，出力データ y_n と教師データ t_n を比較して誤差を算出する損失関数として 2 値交差エントロピー

$$E = -\sum_{n=1}^{k}\{t_n \log y_n + (1-t_n)\log(1-y_n)\} \tag{4.2.3}$$

を計算し，この損失関数値が小さくなるように学習を進める．提案手法では，学習中に誤差 (式 4.2.3) を監視し，この減少が止まってから 3epoch 連続して誤差の減少が見られなかった場合に学習を打ち止めた．

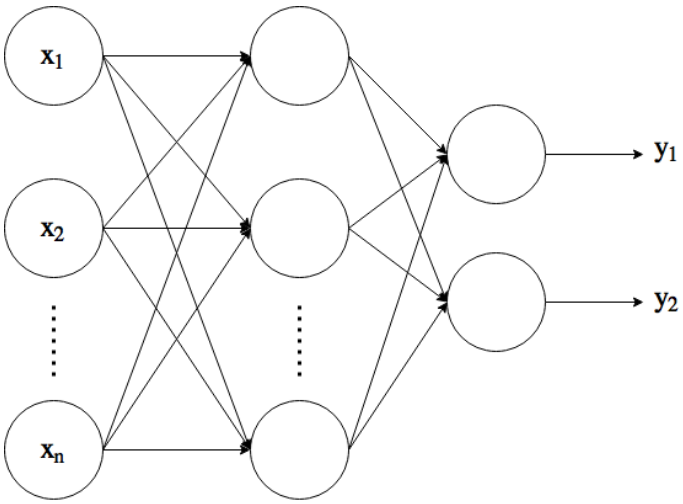

図 4.3 ：多層パーセプトロン

4.2.2 複数の識別器を結合させたモデル

言語特徴と音響特徴を用いた識別器は，それぞれ異なる種類の欺瞞の特徴を学習していることが期待される．そこで，個々のそれぞれの識別器を組み合わせて新たな 1 つの識別器を構築する．提案する識別器は図 4.4 の通りである．y_1，y_2 はそれぞれの識別器の出力であり，w_1，w_2 を重みとして，その重み付き線形和を出力

する．重みについては，各々の出力を線型結合する際の重みの割合を 0 から 1 まで 0.1 ずつ変化させ，バリデーションデータに対して Accuracy が最大化されるようチューニングし，結果として $w_1 = w_2 = 0.5$ とした．

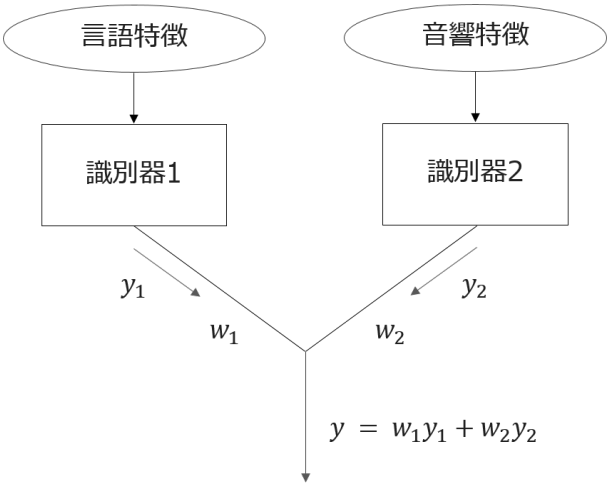

図 4.4 ：識別器を組み合わせたモデル

4.3 ハイパーパラメータの最適化

機械学習においてハイパーパラメータ (チューニングパラメータ) はモデルの性能を大きく左右するため，そのチューニングは重要である．しかし，多層パーセプトロンなどのハイパーパラメータの組み合わせが多数考えられるモデルではチューニングが容易ではない．例えば，グリッドサーチと呼ばれるハイパーパラメータのあらゆる組み合わせの探索を試みる方法では，膨大な組み合わせを検討する必要があるため現実的ではなく，また探索漏れが発生する可能性がある．そこで，本研究ではハイパーパラメータのチューニングにベイズ最適化 [12] を用いる．ベイズ最適化では提案手法のような入力が与えられるまで出力値が分からないようなモデルに対して，その事前分布がガウス過程に従うと仮定し，得られる事後分布の情報に基いて最適化を行う．なお，本研究では最適化の際に Mockus らの Expected Improvement(EI)[21] を用いた．

第 5 章　評価実験

本章では，5.1 節で実験設定，5.2 節で評価指標について述べる．5.3 節及び 5.4 節では，それぞれの実験設定における実験結果を示し，5.5 節では人間と識別器の予測について比較を行う．

5.1　実験設定

本研究では，提案する統計的学習手法による識別器と実験協力者，それぞれに対して複数の真実と欺瞞発話を呈示し，その発話が真実なのか欺瞞なのか，ラベルを予測する実験を行った．実験では，コーパス全体からランダムに選ばれた発話を纏めたデータセット (ランダムテスト) と，一連の対話を対象とした対話の履歴 (文脈) が把握できるデータセット (シーケンシャルテスト) の 2 つの実験設定で行った．

本実験では CSC コーパスを用いてデータセットを作成するため，テストデータの音声と書き起こし文は英語である．そこで，ラベルを予測する人間の母語が検知能力に影響するかを調べるために，英語を母語とする者とそうでない者に実験協力を依頼した．また，テスト時に利用するモーダルが音声のみ，書き起こし文のみ，その両方を用いるそれぞれの場合で検知能力に影響するかを調べるために，実験協力者の一部は異なるモーダルを用いて同様のテストに取り組んだ．以下では，ランダムテスト，シーケンシャルテストにおける実験設定について述べる．

5.1.1　ランダムテスト

本実験は，実験協力者がデータセット中のある発話が真実か欺瞞か予測することが可能であるか，使用するモーダルによって検知能力に差があるか，さらに，検知能力が母語によって異なるかの検証を目的とする．実験では，コーパス中の 32 名のインタビュイーの発話全体からランダムに抽出された 100 発話（真実 50 発話，欺瞞 50 発話）をテストデータとし，テストデータとは別に 4000 発話（真実 2000 発話，欺瞞 2000 発話）の学習データを用意した．ただし，コーパスには極めて短く，文脈無しでラベルを予測することが困難な発話が含まれているため，本実験で

は句読点を含めて 5 単語より長い発話を実験データとして利用した．実験ではテストデータに含まれる発話の音声の聴解，書き起こし文の読解，及び聴解と読解両方の 3 つの形式でテストデータのラベル予測を行った．また，実験協力者はコーパス中のどのような発話が欺瞞または真実であるかの傾向を把握するために，テスト実施前に学習データを自由に見聞きして発話とラベルを確認して良いものとした．

実験協力者について，音声の聴解，文の読解，聴解と読解のテストでは，20-49 歳の英語ネイティブスピーカーの 10 名 (男性 6 名, 女性 4 名)，非ネイティブスピーカーの大学院生 (男性 6 名) が参加した．また，非ネイティブスピーカーについては，一般財団法人国際ビジネスコミュニケーション協会が公開する PROFICIENCY SCALE*を英語能力の参考として，TOEIC®スコア 730 以上（Mean=861.7，SD=117.7）の者が参加した．また，音声と書き起こし文の両方を用いるテストのみ，5.1.2 でのテストと比較するために，上記に加えて 20-39 歳のネイティブスピーカー 5 名 (男性 2 名, 女性 3 名) が追加で参加している．なお，ネイティブスピーカーによる実験では，協力会社に委託して実験を実施した．

一方で識別器は，音響特徴量のみ，言語特徴量のみをそれぞれ学習する識別器とそれぞれの特徴量を結合して学習する識別器，また，音響，言語特徴を別々に学習したそれぞれのモデルを結合させた識別器を用いて人間と同様のテストを行った．

5.1.2 シーケンシャルテスト

5.1.1 の実験では，コーパスからランダムに抽出された発話を対象に欺瞞検知を行うが，これは普段，発話者の発話履歴や文脈を考慮しながら欺瞞を予測している我々人間にとって困難な課題である可能性がある．そこで，実験協力者が文脈を与えられた場合に欺瞞を検知できるかを検証することを目的として本実験を行う．実験では，5.1.1 の参加者の一部である 20-30 代のネイティブスピーカー 5 名 (男性 2 名, 女性 3 名) が参加し，コーパス中の 32 インタビューのうち，1 インタビューに含まれる 138 発話 (真実 70 発話，欺瞞 68 発話，チャンスレベル:0.51) をテストデータとして実験を行った．テストでは音声の聴解と書き起こし文の読解の両方を

*http://www.iibc-global.org/library/default/toeic/official_data/lr/pdf/proficiency.pdf

用いてラベルの予測を行った．また，実験被験者は，テスト前にテストデータとは異なる他の 1 つのインタビューの音声と書き起こし文を学習データとして自由に確認できるようにした．

5.2 評価指標

検知能力の評価指標として欺瞞発話を正例 (P)，真実発話を負例 (N) として，以下の評価指標を用いて評価した．

表 5.2 ： 予測結果の分類

		予測ラベル	
		正例	負例
真のラベル	正例	TP	FN
	負例	FP	TN

$$Accuracy = \frac{TP + TN}{FP + FN + TP + TN} \tag{5.2.1}$$

$$Precision = \frac{TP}{TP + FP} \tag{5.2.2}$$

$$Recall = \frac{TP}{FN + TP} \tag{5.2.3}$$

$$F - measure = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall} \tag{5.2.4}$$

上式の通り，Accuracy は全体の中でどれだけ正解があったかの比率，Precision は正例 (欺瞞) だと予測した中で真に欺瞞な発話の比率，Recall は真に正例 (欺瞞) な発話に対してどれだけ欺瞞だと予測できた比率を表す．また，F-measure は Precision と Recall の調和平均である．

5.3 人間による検知結果と分析

本節では実験結果について述べる．5.3.1 ではテスト時に実験協力者が利用するモーダルについて，5.3.2 では実験協力者の母語について，5.3.3 ではランダムテストとシーケンシャルテストの結果を比較する．なお，以降の表の値は実験協力者の平均値を表す．

5.3.1 モーダルの影響

表 5.3 ：ネイティブスピーカーのモーダルによる影響比較

	Accuracy	Precison	Recall	F-measure
音声	0.421	0.258	0.383	0.308
テキスト	0.447	0.470	0.449	0.459
音声 + テキスト	0.417	0.300	0.386	0.334

10 名のネイティブスピーカーがランダムテストを行った結果，テキストのみを用いた場合に検知能力が最も高くなる結果となった．しかし，いずれのモーダルにおいても Accuracy は有意水準 5% の二項検定において，チャンスレベルと有意差はなかった．また，音声を用いた場合に Precision が低くなる傾向が確認された．

5.3.2 母語の影響

ランダムテストについて，3 つのモーダルを使用して，ネイティブスピーカー 10 名とノンネイティブスピーカー 6 名で検知能力を比較した．また，それぞれの評価指標に対して Welch の t 検定を用いて 2 群の平均値について有意水準 5% で検定を行った．

音声のみを用いた場合

表 5.3 ：母語による影響比較 (音声)

	Accuracy	Precison	Recall	F-measure
ネイティブ	0.421	0.258	0.383	0.308
ノンネイティブ	0.515	0.370	0.524	0.415

いずれの評価指標においても，ノンネイティブスピーカーがネイティブスピーカーを上回る結果となり，Accuracy と Precision については 2 群の平均値に有意差があるという結果になった ($p < 0.05$)．ただし，有意水準 5% の二項検定の結果，2 群いずれも Accuracy はチャンスレベル程度であった．

書き起こし文のみを用いた場合

表 5.3 ：母語による影響比較 (テキスト)

	Accuracy	Precison	Recall	F-measure
ネイティブ	0.447	0.470	0.449	0.459
ノンネイティブ	0.510	0.387	0.515	0.425

2 群いずれも Accuracy はチャンスレベル程度であり，その他のいずれの指標においても，2 群の平均値に有意差があるという結果は得られなかった．

音声と書き起こし文の両方を用いた場合

表 5.3 ：母語による影響比較 (音声 + テキスト)

	Accuracy	Precison	Recall	F-measure
ネイティブ	0.417	0.300	0.386	0.334
ノンネイティブ	0.512	0.360	0.498	0.405

いずれの評価指標においてもノンネイティブスピーカーがネイティブスピーカーを上回っており，Accuracy については 2 群の平均値はいずれもチャンスレベル程度であるが，その差は有意であるという結果になった ($p < 0.05$).

以上の結果より，いずれのモーダルを用いた場合でも，多くの評価指標において，ノンネイティブスピーカーがネイティブスピーカーの検知能力を上回ることが確認された．ただし，その結果はチャンスレベル程度であった．このことから，英語の欺瞞検知を行うにあたって，ネイティブスピーカー，ノンネイティブスピーカーどちらにおいても，正確に欺瞞検知を行うことは困難であり，チャンスレベルに近い予測になっていることが確認された.

5.3.3 発話履歴 (文脈情報) の影響

5.3.1，5.3.2 の実験では，発話者の発話履歴 (文脈情報) が無いために人間が正確に欺瞞を検知できなかった可能性が考えられる．そこで以下では，これまでの実験協力者とは異なる 5 名のネイティブスピーカーを対象に，コーパスからランダムにテストデータを抽出したランダムテストと，1 つのインタービュー全体をテストデータとして扱ったシーケンシャルテストでの検知能力の差を比較し，発話者の発話履歴が人間の欺瞞検知能力に影響を及ぼすかを検証した．以下に 2 群それぞれの検知能力の平均値を示す.

表 5.3 ： 文脈有無による影響比較

	Accuracy	Precison	Recall	F-measure
ランダム	0.504	0.432	0.499	0.457
シーケンシャル	0.478	0.291	0.440	0.339

いずれの評価指標でもランダムテストの結果が，シーケンシャルテストを上回る結果となった．ただし，有意水準 5% で二項検定を行った結果，どちらのテスト形式でも Accuracy はチャンスレベルと比べて有意差が無く，文脈情報の有無に関わらず人間の正解率はチャンスレベル程度であることが確認された．また，表 5.3 の通り，シーケンシャルテストにおいて Precision は低い値を示し，文脈情報があることで，むしろ真に欺瞞な発話に対して真実であると予測してしまう傾向が強まる結果となった．

5.3.4 真実バイアス

ここまでの音声と書き起こし文両方を用いるランダムテストに参加した全 15 名のネイティブスピーカーの欺瞞検知能力の結果について分析を行った．ここで，任意の発話に対する予測ラベルの一致率は平均約 79% となり，ある 1 発話に対して実験協力者の平均約 8 割が同一ラベルを予測していた．また，今回の実験協力者 15 人は，平均して 100 発話のうち約 60% を真実，約 40% を欺瞞と予測しており，この結果は Bond らによる人間の予測バイアスは真実約 56%，欺瞞約 44% とする報告 [3] と概ね一致している．

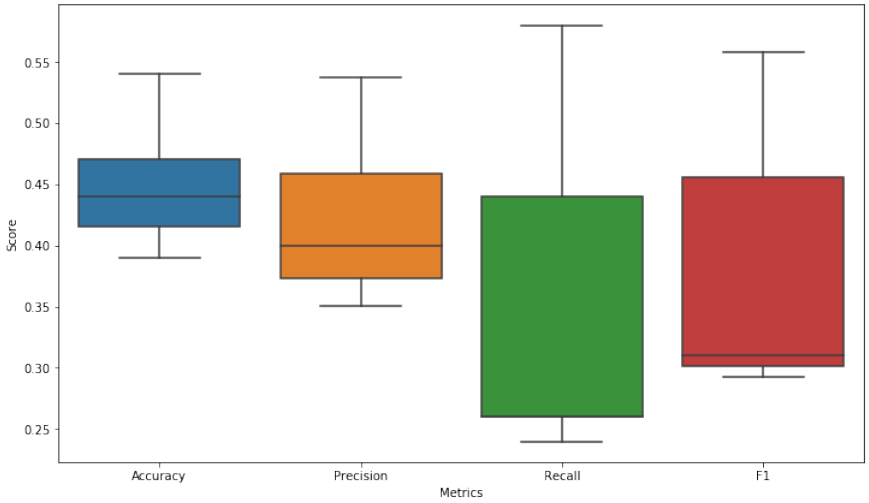

図 5.1 ：実験協力者 15 名の検知能力分布

図 5.1 は実験協力者 15 名の検知能力の分布を示している．本課題においても実験協力者は発話を真実だと予測する傾向が強く，実際に欺瞞発話から欺瞞を予測できなかったため，図 5.1 のように Recall の最小値から中央値までが約 0.26 と著しく低い結果となった．本検証により，本課題においても人間は真実バイアスを有していることが確認された．

5.4 識別器による検知結果と分析

5.4.1 識別器によるランダムテスト

本節では，人間がランダムテストで使用したものと同じテストデータを用いて，識別器が同様のテストを行った結果について述べる．

実験結果

表 5.4 ： 識別器の検知性能

	Accuracy	Precison	Recall	F-measure
音響特徴	0.580	0.583	0.560	0.571
言語特徴†	0.620	0.658	0.500	0.568
モデル結合	0.590	0.605	0.520	0.559
特徴量結合	**0.680**	**0.661**	**0.740**	**0.698**

結果は表 5.4 の通りである．特に音響，言語特徴量を結合して学習した識別器の性能が最も高く‡，この識別器の Accuracy について二項検定を行ったところ，人間 (ネイティブ，ノンネイティブ) 及びチャンスレベルと比較して有意に性能が高い結果となった ($p < 0.001$)．また，以下にこの特徴量結合した識別器についてクローズドテスト§を行った結果を示す．

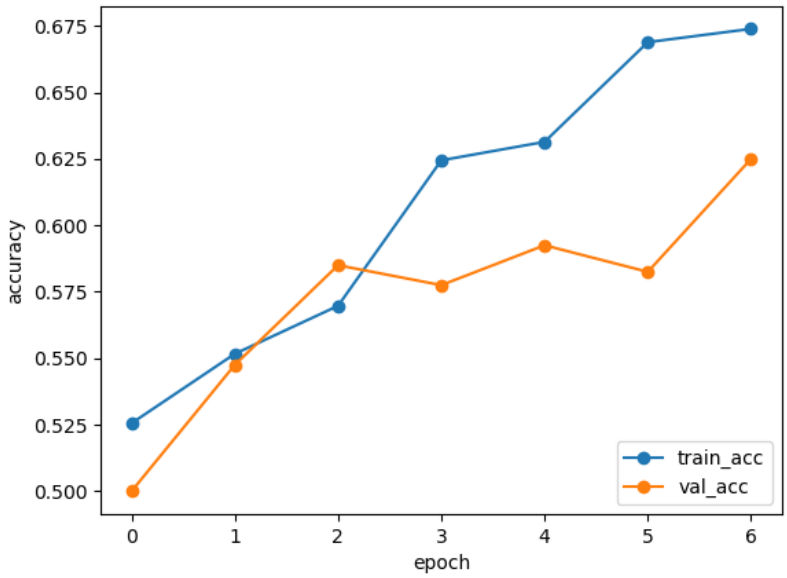

図 5.2 ： クローズドテストの結果

図 5.2 の通り，epoch が進むにつれて学習データ，バリデーションデータのどち

†One-hot ベクトルを用いた場合の Accuracy は 0.56 で分散表現を用いた場合を下回った．
‡関連研究で用いられている 2 値分類器 SVM による Accuracy は 0.54 で提案手法を下回った．
§学習，バリデーションに使用したものと同じデータを用いて行うテスト

らに対しても Accuracy が上昇しており，このモデルについてある程度の学習ができていることを確認した．

識別器と閾値

上記の結果については，識別器が出力する発話が真実である確率が 0.5 以上であれば真実，それ未満であれば欺瞞だと分類している．ここでは，テストデータを使ってこの閾値を変化させることで，各評価指標がどのように変動するかを検証した．結果は図 5.3 の通りである．ただし，グラフ描画の都合上,Positive と予測したものが無い場合は Precision, F1 は 0 とした．

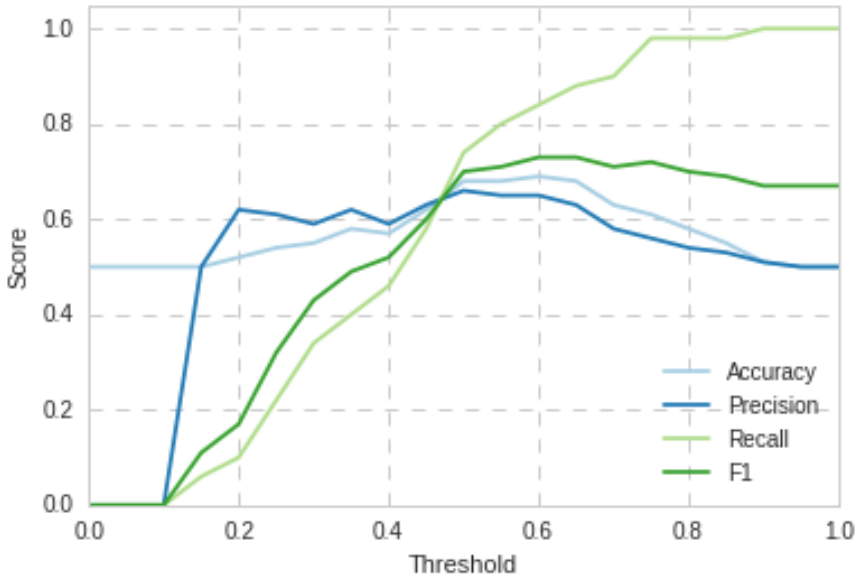

図 5.3 ：閾値と性能の比較

図 5.3 の通り，この識別器では Accuracy は閾値 0.62 付近で最大となり，事前に真実だとは予測しにくい条件を付与する，つまり欺瞞バイアスを付与した場合に性能が最も高なっている．この結果は，識別器では真実または欺瞞バイアスを事前に付与することで性能を調整することが可能であることを示唆している．

5.5　人間と識別器による予測の比較

本節では，ランダムテストにおいて音声とテキスト両方を利用した場合の結果について，人間と識別器の比較を行う．ここではネイティブスピーカーの実験参加者 15 名のうち 8 割 (12 名) 以上が同一のラベルを予測したものを人間の予測とする．テストデータの 100 発話のうち，実験参加者の 8 割以上が同じラベルを予測した発話は全体の 55 発話だった．この 55 発話について，人間の予測と 5.4.1 で最も性能が高かった識別器の予測における，それぞれの正解，不正解の比率は以下の表の通りであった (ただし四捨五入したため，合計が 100% になっていない)．表 5.2 の通り，人間の予測については正解率が 43.7% とチャンスレベル程度であったが，識別器は 74.6% であり，二項検定の結果，有意水準 5% でチャンスレベルより正解率が高い結果となった．

表 5.2　人間と識別器の予測結果比率 (単位:%)

		人間の予測		計
		正解	不正解	
識別器の予測	正解	36.4	38.2	74.6
	不正解	7.3	18.2	25.5
計		43.7	56.4	100.1

それぞれの事例には表 5.3 から表 5.6 のような発話が含まれる．統計的学習手法を用いることで，人間が予測を誤る問題に正解できた例として表 5.5 の発話が挙げられる．表中の *"uh, visiting museums, um, uh, um, you know, theaters, things like that. yeah."* は *"uh"* や *"um"* などのフィラーが多数含まれているためか，実験協力者の *14/15* 名が欺瞞だと誤った予測をした．一方で識別器は，学習データから真実と欺瞞発話の特徴を学習した結果，約 *73%* の確率で真実だと予測し，正解することができた．ただし，人間が正解した問題で識別器が誤った事例もいくつか存在する．例えば，表 *5.6* 中の *2* つの発話は実験協力者の *14/15* 名の予測が一致して正解した発話である．しかし，識別器は実際は真実である発話 *"you have to learn these things."* を約 *79%* の確率で欺瞞，実際は欺瞞である発話 *"I came I arrived, uh,"* を

約 *68%* で真実だと予測した．ただし，このように人間の予測が正解で識別器が誤った事例は全体の *7.3%* に過ぎず，人間が不正解で識別器が正解した割合 *38.2%* より少数であった．

表 5.3　人間も識別器も正解だった発話例

発話	実際のラベル	人間の予測	識別器の予測
she'd have been surprised. yeah.	真実	真実	真実
I think so, yeah. I mean, um,	欺瞞	欺瞞	欺瞞

表 5.4　人間も識別器も不正解だった発話例

発話	実際のラベル	人間の予測	識別器の予測
you know, end up some place	真実	欺瞞	欺瞞
stems a lot from my parents. they	欺瞞	真実	真実

表 5.5　人間が不正解で識別器が正解だった発話例

発話	実際のラベル	人間の予測	識別器の予測
uh, visiting museums, um, uh, um, you know, theaters, things like that. yeah.	真実	欺瞞	真実
yeah, I'm a little bit embarrassed.	欺瞞	真実	欺瞞

表 5.6　人間が正解で識別器が不正解だった発話例

発話	実際のラベル	人間の予測	識別器の予測
you have to learn these things.	真実	真実	欺瞞
I came I arrived, uh,	欺瞞	欺瞞	真実

第 6 章　結言

6.1　本論文のまとめ

本研究では，欺瞞を含むインタビュー形式の対話を対象に，人間と統計的学習手法による識別器の欺瞞検知能力の比較検証を行った．人間を対象とした実験結果から，今回の実験では人間が欺瞞検知能力は文脈 *(*対話履歴*)* の有無，実験協力者の母語や利用するモーダルに関わらず，チャンスレベル程度であることが確認された．また，実験協力者の予測結果から，本課題においても人間の予測には真実バイアスが生じていることが確認された．一方で，人間が持つようなバイアスを有しない統計的学習手法に基づく識別器による欺瞞検知では，人間及びチャンスレベルよりも有意に性能が高いことが確認された．また，単一の特等量のみを用いるのではなく，音響特徴量と言語特徴量を組み合わせた場合の方が，識別性能が向上することが確認された．

6.2　今後の課題

今後は，系列情報の学習が可能である *Long short-term memory(LSTM)[29]* やその拡張である *Bidirectional LSTM[29]* などを用いて，対話の発話履歴を考慮するモデルを構築する予定である．また，識別器が正解，不正解する場合についても，より詳細な分析を行いたいと考えている．例えば，*Ribeiro* らによって提案された *Local Interpretable Model-Agnostic Explanations(LIME)[24]* と呼ばれる識別結果の解釈手法や，発話文に含まれる特定の語のマスキングを行って予測の変化を調査することで，識別器による予測に強い影響を与えている単語を特定することができると考えられる．このような分析を通して，欺瞞検知に利用する特徴量及び識別器の改良を行い，より高性能な識別器の構築を目指したい．

謝辞

本学情報科学研究科の中村哲教授には主指導教官として，常に熱心なご指導をして頂き，大変有意義な研究生活を送ることができました．心より感謝致します．本学情報科学研究科の松本裕治教授には副指導教官として，的確で貴重なご助言を賜りました．心より感謝致します．本学情報科学研究科の *Sakriani Sakti* 特任准教授には副指導教官として，いつも優しく，多数のご助言を賜りました．心より感謝致します．本学情報科学研究科の吉野幸一郎助教には副指導教官として，とても丁寧なご指導，多数のご助言を賜りました．心より感謝致します．本学情報科学研究科の須藤克仁准教授には親身になって頂き，的確なご助言を多数賜りました．心より感謝致します．本学情報科学研究科の田中宏季助教には優しいご助言を賜りました．心より感謝致します．知能コミュニケーション研究室の秘書である松田真奈美様，林美保様，軽部瑞穂様には，諸手続きの際に大変お世話になりました．心より感謝致します．知能コミュニケーション研究室の皆様には，公私にわたり大変お世話になりました．心より感謝致します．

参考文献

[1] *S. Amiriparian, J. Pohjalainen, E. Marchi, S. Pugachevskiy, and B. W. Schuller, "Is deception emotional? an emotion-driven predictive approach." in* INTERSPEECH*, 2016, pp. 2011–2015.*

[2] *D. E. Anderson, B. M. DePaulo, and M. E. Ansfield, "The development of deception detection skill: A longitudinal study of same-sex friends,"* Personality and Social Psychology Bulletin*, vol. 28, no. 4, pp. 536–545, 2002.*

[3] *C. F. Bond Jr and B. M. DePaulo, "Accuracy of deception judgments,"* Personality and social psychology Review*, vol. 10, no. 3, pp. 214–234, 2006.*

[4] *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.*

[5] *P. Ekman, M. O'Sullivan, W. V. Friesen, and K. R. Scherer, "Invited article: Face, voice, and body in detecting deceit,"* Journal of nonverbal behavior*, vol. 15, no. 2, pp. 125–135, 1991.*

[6] *F. Enos,* Detecting deception in speech*. Columbia University, 2009.*

[7] *F. Eyben, F. Weninger, F. Gross, and B. Schuller, "Recent developments in opensmile, the munich open-source multimedia feature extractor," in* 21st ACM international conference on Multimedia*. ACM, 2013, pp. 835–838.*

[8] *J. 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." in* Interspeech*, 2005, pp. 1833–1836.*

[9] *A. Joulin, E. Grave, P. Bojanowski, and T. Mikolov, "Bag of tricks for efficient text classification,"* arXiv preprint arXiv:1607.01759*, 2016.*

[10] *S. M. Kassin, C. A. Meissner, and R. J. Norwick, "" i'd know a false confession if i saw one": a comparative study of college students and police investigators."* Law and Human Behavior*, vol. 29, no. 2, p. 211, 2005.*

[11] *R. M. Krauss, "Impression formation, impression management, and nonverbal behaviors," in* Social cognition: The Ontario Symposium*, vol. 1. Erlbaum Hillsdale, NJ, 1981, pp. 323–341.*

[12] *H. J. Kushner, "A new method of locating the maximum point of an arbitrary multipeak curve in the presence of noise,"* Journal of Basic Engineering*, vol. 86, no. 1, pp. 97–106, 1964.*

[13] *J. Levashina and M. A. Campion, "Measuring faking in the employment interview: development and validation of an interview faking behavior scale."* Journal of applied psychology*, vol. 92, no. 6, p. 1638, 2007.*

[14] *T. R. Levine, R. K. Kim, H. Sun Park, and M. Hughes, "Deception detection accuracy is a predictable linear function of message veracity baserate: A formal test of park and levine's probability model,"* Communication Monographs*, vol. 73, no. 3, pp. 243–260, 2006.*

[15] *S. I. Levitan, G. An, M. Wang, G. Mendels, J. Hirschberg, M. Levine, and A. Rosenberg, "Cross-cultural production and detection of deception from speech," in* Proceedings of the 2015 ACM on Workshop on Multimodal Deception Detection*. ACM, 2015, pp. 1–8.*

[16] *L. v. d. Maaten and G. Hinton, "Visualizing data using t-sne,"* Journal of Machine Learning Research*, vol. 9, no. Nov, pp. 2579–2605, 2008.*

[17] *S. A. McCornack and M. R. Parks, "Deception detection and relationship development: The other side of trust,"* Annals of the International Communication Association*, vol. 9, no. 1, pp. 377–389, 1986.*

[18] *C. A. Meissner and S. M. Kassin, "" he's guilty!": investigator bias in judgments of truth and deception."* Law and human behavior*, vol. 26, no. 5, p. 469, 2002.*

[19] *G. Mendels, S. I. Levitan, K.-Z. Lee, and J. Hirschberg, "Hybrid acoustic-lexical deep learning approach for deception detection,"* Proc. Interspeech 2017*, pp. 1472–1476, 2017.*

[20] *T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space,"* arXiv preprint arXiv:1301.3781*,*

2013.

[21] J. Mockus, V. Tiesis, and A. Zilinskas, "The application of bayesian methods for seeking the extremum," vol. 2, pp. 117–129, 1978.

[22] R. S. Nickerson, "Confirmation bias: A ubiquitous phenomenon in many guises." Review of general psychology*, vol. 2, no. 2, p. 175, 1998.*

[23] J. Pennington, R. Socher, and C. Manning, "Glove: Global vectors for word representation," in Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, 2014, pp. 1532–1543.*

[24] M. T. Ribeiro, S. Singh, and C. Guestrin, "Why should i trust you?: Explaining the predictions of any classifier," in Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 2016, pp. 1135–1144.*

[25] W. Robinson, A. Shepherd, and J. Heywood, "Truth, equivocation concealment, and lies in job applications and doctor-patient communication," Journal of Language and Social Psychology*, vol. 17, no. 2, pp. 149–164, 1998.*

[26] F. Rosenblatt, "The perceptron: A probabilistic model for information storage and organization in the brain." Psychological review*, vol. 65, no. 6, p. 386, 1958.*

[27] B. Schuller, S. Steidl, and A. Batliner, "The interspeech 2009 emotion challenge," in Tenth Annual Conference of the International Speech Communication Association*, 2009.*

[28] B. W. Schuller, S. Steidl, A. Batliner, J. Hirschberg, J. K. Burgoon, A. Baird, A. C. Elkins, Y. Zhang, E. Coutinho, and K. Evanini, "The interspeech 2016 computational paralinguistics challenge: Deception, sincerity & native language." in INTERSPEECH*, 2016, pp. 2001–2005.*

[29] M. Schuster and K. K. Paliwal, "Bidirectional recurrent neural networks," IEEE Transactions on Signal Processing*, vol. 45, no. 11, pp. 2673–2681, 1997.*

[30] L. Van Der Maaten, "Accelerating t-sne using tree-based algorithms." Jour-

nal of machine learning research, *vol. 15, no. 1, pp. 3221–3245, 2014.*

[31] A. Vrij, Detecting lies and deceit: Pitfalls and opportunities. *John Wiley & Sons, 2008.*

[32] A. Vru and S. Heaven, "Vocal and verbal indicators of deception as a function of lie complexity," Psychology, crime and law, *vol. 5, no. 3, pp. 203–215, 1999.*

[33] B. Weiss and R. S. Feldman, "Looking good and lying to do it: Deception as an impression management strategy in job interviews," Journal of Applied Social Psychology, *vol. 36, no. 4, pp. 1070–1086, 2006.*

[34] 村井潤一郎, "社会心理学における嘘研究: 現状と展望, 嘘に対する雑感 (特集うそ・ウソ・嘘)," 心理学ワールド, *no. 71, pp. 5–8, 2015.*

発表リスト

研究会

[1] 細見 直希*, Sakriani Sakti,* 吉野 幸一郎*,* 中村 哲，*fastText* を用いた対話からの欺瞞検知と分析，研究報告音声言語情報処理（*SLP*）*, 2017-SLP-117, 6, 1-5, Jul. 2017*