

Detecting Concealed Information in Text and Speech

Shengli Hu

論文 <https://www.aclweb.org/anthology/P19-1039>

ポスター https://shenglih.github.io/research/acl_poster.pdf

紹介者: 笹野遼平 (名大)

論文の構成

1. Introduction

- 扱う問題(=Concealed Information detection)の説明

2. Related Work

3. Data

- データの作成方法の説明


4. Feature Extraction

- 検出に利用する素性の説明と分析

5. Classification Experiments

6. Conclusion, Limitations, and Future Directions

© 10/18/2018



あるマスター・ソムリエが、
テイスティングされるワインに関する情報を事前に
漏らしていた

情報を教えられた受験
者は、情報を知っていることを隠そうとする

9月のマスター・ソムリエ試験で不正が行われ、ソムリエのクリス・ラメルブ氏を含む23人がマスター・ソムリエ資格を剥奪された。マスター・ソムリエ協会の理事会は、10月9日の声明で、2018年マスター・ソムリエ試験のワイン・テイスティングの部の結果を無効にすると発表した。あるマスター・ソムリエが、テイスティングされるワインに関する情報を事前に漏らしていた「明らかな証拠が見つかった」ためである。理事会はそのマスター・ソムリエの名前や、54人の受験者との関係を明らかにしていない。マスター・ソムリエは、ワイン専門家の最高レベルで、その取得には何年にもわたる勉学と995ドルの受験料を含む高額投資が必要である。1969年に試験が開始されて以来、合格者は世界中でわずか273人である。ラメルブ氏はハワイで4人目の認定取得者だった。「試験プロセスの完全性を維持することが最優先で、マスター・ソムリエの価値と威厳が損なわれないようにしなければならない。マスター・ソムリエの資格は、世界中で認められた、ワイン業界の専門家の中で最も高い資格である。手を抜いた試験ではその保証は得られない」と、デボン・プロルリー会長は述べた。ラメルブ氏は、マスター・ソムリエ協会主催のソムリエ予備試験を管理する5人のマスター・ソムリエの1人である。本試験には、理論、ワイン・サービス、6つの銘柄のワインのブラインド・テイスティング（名前、ヴィンテージ、原産地を当てる）の3つのパートがある。9月にセント・ルイスで行われたこのブラインド・テイスティングの部分が無効になった。協会は、無効になったこの試験部分の受験料を払い戻し、再試験の受験料を免除し、旅行費用を負担する、と発表した。再試験は、本年末及び2019年春か初夏に行われる。

Concealed information Detection

- 重要な情報を知っているのに知らないかのように振る舞っていること(隠蔽情報)を検出するタスク

⇔ (狭義の) Deception(虚偽):

– 知らない情報を知っているかのように振る舞うこと

The Information Grid		Appearance	
		Information	No Information
Truth	Information	Honesty	Concealed Information
	No Information	Deception	Honesty

Table 1: The Information Grid: Concealed Information vs. Deception

Research Questions

1. 人はどのくらい隠蔽情報を検出できるか？
2. 音声とテキスト情報を用いることで人の精度を超えた分類器を作成できるか？
3. **隠蔽**(concealed)情報検出と**虚偽**(deception)検出の手掛かりとどのくらい関連している？
4. 機械学習に基づく分類器はどのような場合に専門家より良い精度を実現できるか？

データ: Blind Tasting Game

- ソムリエ試験の不正に似せたデータを作成
 - 5-10人のワインのプロが7-10ラウンド参加
 - 各ラウンドごとに1つのワインが注がれるが、その際、無作為に参加者の1人(=情報隠蔽者)に正解が伝えられる
 - (隠蔽者も含む)参加者は、(正解率が高くなるよう)taste, describe, reason, concludeを行い、その後、ワインの情報が公開
 - 続けて、各参加者は、誰が情報隠蔽者であったかを予測(何人書いてもOK)
- もっともうまく隠蔽した参加者は fine wine を貰える

作成データの詳細 & 前処理

- 41人のプロが参加し、49セッションを記録
 - 各参加者は、発話に加え、ワインの描写、結論、情報隠蔽者の予測結果をシート(text)に記載し提出
 - セッション終了後、参加者の性別、母語、資格、ドメイン知識情報も収集(Individual information)
- 音声情報を書き起こし(自動+人手修正) + 分割
 - ポーズを含まない連続する発話を“turn”と定義
 - Single turn分割: 1つずつの turn で構成
 - Multiple turn分割: 各ラウンドの各参加者のturnの集合
- 最終的に、164時間、3288のmultiple turn分割、9104のsingle turn分割(7:1:2に分けて5分割交差検定)

論文の構成

1. Introduction

- 扱う問題(=Concealed Information detection)の説明

2. Related Work

3. Data

- データの作成方法の説明

4. Feature Extraction

- 検出に利用する素性の説明と分析

5. Classification Experiments

6. Conclusion, Limitations, and Future Directions

音声・韻律的手掛かり

- 既存の研究で有効そうなものを使用
 - low-level音声素性(大きさ、ピッチ、声質など) (8)
 - メル周波数ケプストラム係数(MFCC) (13)
 - IS 2013: the 2013 Computational Paralinguistics Challenge baseline feature set (6373)
 - IS 2009: the 2009 Emotion challenge baseline feature set (384) (実際の実験ではこれだけ使用?)

low-level素性と情報隠蔽の関係

- 多重検定の設定で $\alpha=0.05$ 相当になるように補正を行った上で有意差検定 (S: 有意差あり)

Feature	Male	Female	Low Skill	High Skill	All
Pitch (max)	S				S
Pitch (mean)					
Intensity (max)	S	S	(S)		S
Intensity (mean)		(S)			
Speaking Rate			S		S
Duration		(-)(S)	(-)(S)		(-)(S)
Voice Quality					

情報隠蔽者は、ピッチの最高値、最大音量、話速が大きく、話す時間は短くなる傾向

Durationのみ虚偽(deception)では確認できなかった特徴

言語的手掛かり

- LIWC (Linguistic Inquiry and Word Count):
 - 語彙がカテゴリごとに分割された辞書
- Linguistic:
 - hedging (はぐらかし, e.g.: I think ..., sort of ...), subjectivity, contraction (短縮形) など10種類
- Length:
 - trunの平均語数、文の平均語数、語の平均長, etc.
- Ngrams:
 - unigram, bigram, trigram
- Emgeddings:
 - ワイン記事とtweetで学習したGloVeベクトル

言語的掛かりと情報隠蔽の関係

Feature	Concealed Information	Truthful
N-grams	<i>yeah</i> , but it, citrus, <u>correct</u> , ruby, did not, lift, botrytis, would not	<i>uh um</i> , there is, there are, was like, so, slight, not sure, blossom, clear
LIWC	<i>clout</i> , <u>certain</u> , <u>function</u> , <u>cogproc</u> , <u>negate</u> , <i>discrep</i> , <u>differ</u> , <u>assent</u> , <u>posemo</u>	<u>compare</u> , <u>pronoun</u> , <u>verb</u> , <i>ingest</i> , <u>feel</u>
Syntax	<i>adj</i> , <i>adverb</i> , syn_distinct	
Else	<i>specificity</i> , $\Delta(Trans, Text)$	<u>hedging</u> , <u>#word</u> , <u>length</u>

Detecting Concealed Information in Text and Speech

Shengli Hu, Cornell University

Motivation

Why a cheating scandal is shaking the sommelier world

With question-leaking, CNN has a scandal on its hands

Deception vs. Information Concealment

The Information Grid	Appearance	
	Information	No Information
Truth	Information	Honesty
	No Information	Deception

Consistent or Inconsistent with most recent deception literature²

Feature	Consistent	Inconsistent	Truthful
		but it, citrus, ruby, did not, botrytis, would not	uh um, there is, there are, was like, so, slight, not sure, blossom, clear
			compare, pronoun, verb, ingest, feel
Else	specificity, $\Delta(\text{Trans, Text})$	hedging ³ , #word, length	

Consistent or Inconsistent with most recent deception literature

Research Questions

- How good are humans at detecting concealed information in technical settings?
- Can we improve on human performance?
- How are indicators of concealed information related to those of deception?
- When are Machine Learning classifiers better (or worse) than human domain experts?

Contributions

- The first corpus and study on concealed information in technical settings – please let me know if I am totally wrong here!
- Novel insights from identified key features (cf. deception)
- Multi-task learning

- Acoustic-prosodic deception
- Linguistic deception
- Algorithmic deception
- Multi-task learning

Selected Papers

- Wolfgang Ambrazaitis, Guozhen An, et al. 2018. Deception detection in technical settings.
- Stefan Bensus, et al. 2018. Deception detection in technical settings.
- Sarah Ita Levitan et al. 2018a. Individual differences in deception and deception detection.
- Sarah Ita Levitan, et al. 2018b. Acoustic-prosodic indicators of deception and trust in interview dialogues.
- Sarah Ita Levitan, et al. 2018b. Linguistic cues to deception and perceived deception in interview dialogues.
- Gideon Mandels, et al. 2017. Hybrid acoustic-lexical deep learning approach for deception detection.
- Rada Mihalcea and Carlo Strapparava. 2009. The liedetector: Explorations in the automatic recognition of deceptive language.
- Esther Mobley. 2018. Why a cheating scandal is shaking the sommelier world.
- Myle Ott, et al. 2011. Finding deceptive opinion span by any stretch of the imagination.

Demographics: gender, native language, wine credential, self confidence.

Discrepancy: the author herself is a sommelier, a wine student, certified sommelier, and certified specialist of wine.



Feature	Male	Female	Low Skill	High Skill	All
Pitch (max)	S				S
Pitch (mean)					
Intensity (max)	S	S	(S)		S
			S		S
			(-)S		(-)S

we denote significant features consistent with deception literature as red and underlined, and those opposite with deception literature as *blue and italicized*.

Model	Features	Accuracy
BiLSTM	GloVe	61.41 / 67.35
MLP + BiLSTM	IS 2009, GloVe	64.12 / 68.57
MLP + BiLSTM	IS 2009, Individual Features, GloVe.	64.14 / 70.02
MLP + BiLSTM + Multi-task	IS 2009, Individual Features, GloVe.	65.16 / 71.51



言語的手掛かりと情報隠蔽の関係

Feature	Concealed Information	Truthful
	<p>確信レベルが増加するのと同様に、情報を隠さなければというプレッシャーから認知負荷 (“cause”, “think”, “know”) も増加</p>	<p><i>uh um</i>, there is, there are, was like, so, slight, not sure, blossom, clear</p>
LIWC	<p><i>clout</i>, <u>certain</u>, <u>function</u>, <u>cogproc</u>, <u>negate</u>, <i>discrep</i>, <u>differ</u>, <u>assent</u>, <u>posemo</u></p>	<p><u>compare</u>, <u>pronoun</u>, <u>verb</u>, <i>inaest</i> <u>feel</u></p>
Syntax	<p><i>adj</i>, <i>adverb</i>, <i>syn_distinct</i></p>	<p>はぐらかすような表現が多い</p>
Else	<p><i>specificity</i>, $\Delta(Trans, Text)$</p>	<p><u>hedging</u>, <u>#word</u>, <u>length</u></p>

Consistent or Inconsistent with most recent deception literature

分類実験

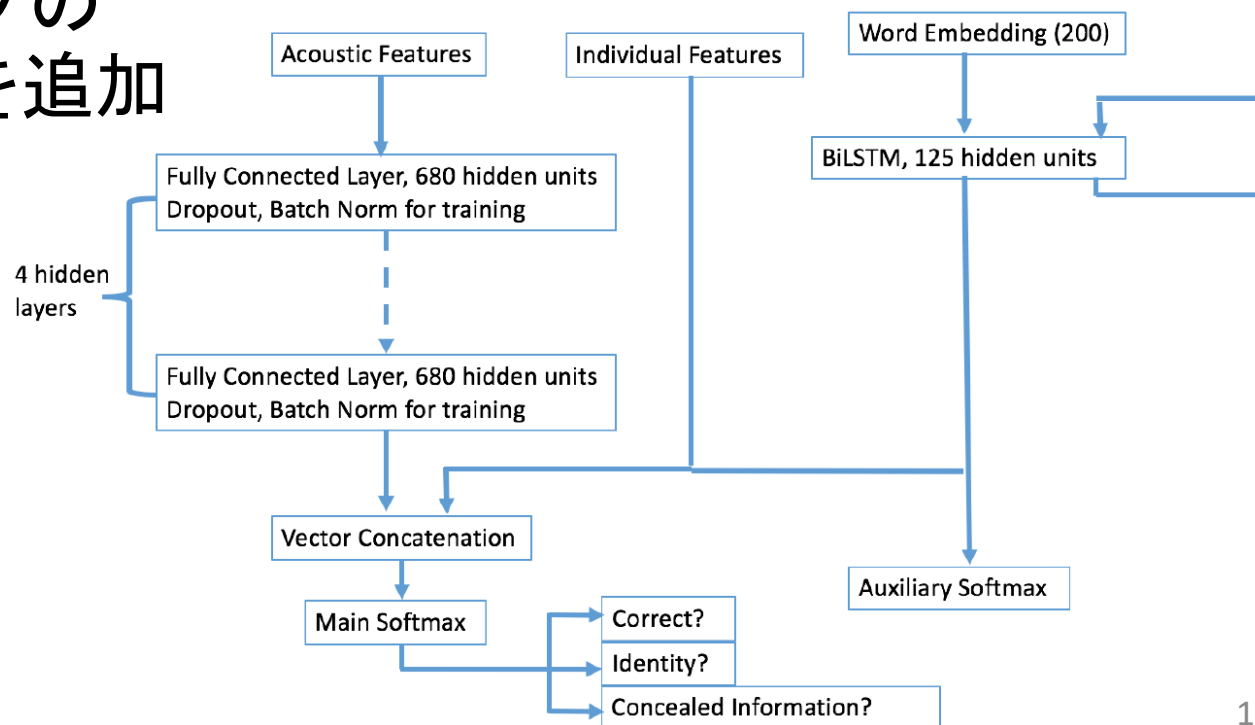
- Baseline Models
 - 言語的手掛かり with Logistic Regression (LR)
 - 音声・音韻的手掛かり with Random Forest (RF)
- Deep Learning Models
 - 音声・音韻的手掛かり with MLP
 - GloVeベクトル with BiLSTM
 - 上2つの統合モデル (MLP+BiLSTM)
 - さらに Individual素性 (性別、母語、etc.) を追加

Multi-task Learning Models

- 他に2つのタスクを追加
 1. 参加者の予想が正解かどうかを予測
 2. ワインの銘柄等の正解が何であることを予測

- テイスティングのデータセットを追加

- 5.5時間
- 21ラウンド



実験結果

Model	Features	F1 (single / multiple turn)
Logistic Regression	Bigrams	Human: NA / 56.28 61.18 / 65.45
Random Forest	IS 2009	59.23 / 60.03
MLP	IS 2009	63.96 / 67.27
BiLSTM	GloVe	61.41 / 67.35
MLP + BiLSTM	IS 2009, GloVe	64.12 / 68.57
MLP + BiLSTM	IS 2009, Individual Features, GloVe.	64.14 / 70.02
MLP + BiLSTM + Multi-task	IS 2009, Individual Features, GloVe.	65.16 / 71.51

まとめ

- Concealed information検出に取り組んだ論文
- プロのソムリエを集めデータセットを作成
 - 知人がいるなどの要因でややバイアスされている可能性
 - モデルの頑健性を示すには他分野のデータも必要
- 検出の手掛かりを、Deception(虚偽)検出に有効な手掛かりと比較し、特徴を分析
 - 多くの手掛かりは一致するものの一部異なっている
 - e.g., 説明時間は短くなる傾向, 前向きな感情や確信度は増加
- マルチタスク学習を取り入れたニューラルネットに基づく手法で、人を大きく超える予測精度を実現