

Real conversations with artificial intelligence

Jennifer Hill

W. Randolph Ford

Ingrid G. Farreras

Computers in Human Behavior
Volume 49 Issue C, August 2015
Pages 245-250

目次

- 背景
- 目的
- **Cleverbot**
- 実験内容
- 実験結果
- まとめ
- 所感

背景

- 近年では,様々なシステムにAI技術が用いられている.
- 人工知能研究において,コンピュータが人間の言葉の意味を理解することや,言葉を伝えることの難しさが問題視されている.

目的

- 人間が行うコミュニケーションにおいて,人間相手とチャットボット相手とで,どのような違いが生まれるのか.

Cleverbot

- 1997年に作られたbot.
- 過去の会話を学習して,返事を行う.
- かつての実験では,参加者の59%がCleverbotとの会話を,人間の会話であると思い込んだ.

Cleverbot

User: Hi.

Cleverbot: Do you like cheese?

User: No.

Cleverbot: Why not?

User: Not a big fan.

Cleverbot: You like animals?

User: Yes. I like cats.

Cleverbot: Do you like me the best though?

User: I have to get to know you more to see.

Cleverbot: Do you like herman cain?

User: No.

Cleverbot: Why?

User: Cause's I dont know nothing about him.

Cleverbot: My name is Cleverbot.

Fig. 1. A typical conversation with Cleverbot.

実験内容

- インスタントメッセージング(IM)と,ボットとの対話を,それぞれランダムに100個用意する.
- その中から,書かれたテキストの分量,使われた単語の種類,顔文字などの特殊な表現などを比較する.
- 比較には多変量分散分析を用いる.

実験結果

Table 1

The means, standard deviations, and univariate ANOVA results of the seven dependent variables measured in human–human and human–chatbot conversations.

	Human (<i>N</i> = 100) <i>M</i> (<i>SD</i>)	Chatbot (<i>N</i> = 100) <i>M</i> (<i>SD</i>)	<i>F</i>	<i>p</i>	η^2
Words/conversation	190.42 (212.61)	219.95 (201.95)	1.01	0.315	–
Messages/conversation	23.03 (22.33)	49.58 (37.11)	37.58	0.001*	0.16
Words/message	7.95 (4.02)	4.29 (1.64)	70.86	0.001*	0.26
Type-token	0.67 (0.14)	0.47 (0.09)	144.40	0.001*	0.10
% Profanity	0.16 (0.39)	4.29 (4.55)	81.65	0.001*	0.29
Shorthand	2.69 (4.17)	2.28 (4.03)	0.50	0.480	–
Emoticons	0.70 (1.37)	1.01 (2.90)	0.93	0.335	–

* Statistically significant.

実験結果

Table 2

The means, standard deviations, *F* values, degrees of freedom, *p* values, and effect sizes for the 14 variables analyzed by the LIWC program between human-human and human-chatbot conversations.

	Human (<i>N</i> = 100) <i>M</i> (<i>SD</i>)	Chatbot (<i>N</i> = 100) <i>M</i> (<i>SD</i>)	<i>F</i>	<i>P</i>	η^2
WC	191.12 (212.00)	219.43 (202.42)	0.93	0.335	
WPS	42.77 (114.85)	4.65 (2.03)	11.00	0.001*	0.05
Six-letter	10.50 (4.96)	8.25 (3.38)	13.97	0.001*	0.06
Dic	87.49 (5.51)	86.72 (6.77)	0.77	0.380	
Funct	53.77 (7.58)	54.75 (8.29)	0.76	0.384	
Pronoun	17.25 (4.29)	22.90 (4.99)	73.69	0.001*	0.27
Article	4.50 (2.37)	4.21 (2.31)	0.76	0.382	
Swear	0.23 (0.67)	1.60 (2.14)	37.08	0.001*	0.15
Social	10.61 (5.75)	16.40 (4.73)	60.45	0.001*	0.23
Affect	9.68 (6.01)	8.95 (4.55)	0.93	0.335	
Pos Emot	7.78 (5.64)	6.00 (3.92)	6.67	0.011*	0.03
Neg Emot	1.89 (1.99)	2.94 (2.56)	10.41	0.001*	0.05
Sexual	0.28 (1.03)	3.45 (4.25)	52.51	0.001*	0.21
Assent	3.84 (3.32)	3.65 (3.01)	0.17	0.676	

* Statistically significant.

実験結果

- 人間相手とロボット相手とで,出現するメッセージ数,1文あたりの単語数,特定の言葉が出てくる頻度,口の悪い表現の数に違いが見られた.
- また出現する言葉の内容や品詞にも違いが見られた.

まとめ

- 相手が人間かロボットかで,送信するメッセージの数が変化していた.
- 出現する言葉の内容にも違いが見られた.
- 今回の実験では品詞の明確な分類はしていなかったため,今後品詞に注目した研究がなされるかもしれない.

所感

- ボット相手の場合に否定的な言葉や罵倒が多かったのは,相手が人間でないとわかっていたからではないか.
- 分析手法に関して勉強が足りていないと感じた.