

擬人化エージェントへの信頼をもたらす設計論 Anthropomorphic-Agent Design deriving trustworthy

松井哲也^{1*} 山田誠二^{1,2,3}
Author1¹ Author2^{1,2,3}

¹ 国立情報学研究所

¹ National Informatics of Informatics

² 総合研究大学院大学

² the Graduate University for Advanced Studies (SOKENDAI)

² 東京工業大学

² Tokyo Institute of Technology

Abstract: Our goal is to suggest the agent design model to improve user's trust. We experimented about PRVA (product recommending virtual agent) because they are needed to be trusted by consumers. We suggest that improving agent's trust by increasing user's positive emotion and agent's intelligence. We experimented with 15 participants about PRVA's trustworthiness. We showed our model is practical

1 はじめに

本研究の目的は、人（ユーザー）と擬人化エージェントの間に信頼関係を生じさせ、それを高める擬人化エージェントの設計論を提出することである。本研究では擬人化エージェントの中でも、特に PRVA（商品推薦エージェント）とユーザーの間の信頼について考える。その理由は、PRVA はオンラインショッピングなどで利用者に商品を推薦し、購入を促すことが目的のエージェントであり、その役割を果たすためには利用者に信頼感を与えることが重要であると考えられるためである。日常生活などで「信頼する」という語を使う場合、信頼される対象は他の人間である場合もあれば、機械やコンピューターである場合も考えられる。他人に対する信頼については、Yamagishi Yamagishi[13]が信頼の理由に基づいて分類を行っている。ここで著者らは、信頼を一般的信頼（"General trust"）と知識ベース信頼（"Knowledge-based trust"）の二つに分けている。一般的信頼は不特定多数に対する信頼であり、信頼者の持つ固有の属性である。知識ベース信頼とは、信頼者と特定の関係にある相手に対して、「自分の利益を上げてくれる」という信頼である。知識ベース信頼は被信頼者の人格・能力に加えて、信頼者と被信頼者の関係性にも大きく左右されると言える。一方で人（ユーザー）とコンピューターの間の信頼につ

いては、Artz[2]によるサーベイ論文を元に概観しておきたい。Artsらは過去の研究をリストアップし、そこで使われていた信頼概念を大きく三つに分類する。その三つは「過去の評判に基づく信頼」[9]、「能力に対する信頼」[7]、「行為に対する信頼」[10]である。このように、人と人との間の信頼では信頼者と被信頼者の関係性が重視され、それと比較すると人とコンピューターとの間の信頼では被信頼者（コンピューター）の能力が重視されている。関係性に基づく信頼とは、「これだけ仲がいいのだから、自分を裏切りはしないだろう」という信頼であり、被信頼者の人格に起因するものだと言える。擬人化エージェントは人とコンピューターの双方の側面を持つものである。そのため、信頼するにあたっては人格と能力の双方が重視されると思われる。オンラインショッピングサイトにおいて、消費者がどのように購入の意思決定をするかについては、Kimら[8]がモデルを作成している。ここでは、消費者のサイトへの信頼に影響を与えるのは、主に「サイトの安全性」と「与えられる情報の質」であるとされている。ショッピングサイトを擬人的に考えるなら、サイトの安全性は人格に、情報の質は能力に相当するものであると考えられる。

*連絡先：国立情報学研究所
〒101-8430 東京都千代田区一ツ橋 2-1-2.
E-mail: tmatsui@nii.ac.jp

2 擬人化エージェントへの信頼

2.1 擬人化エージェントへの信頼の定義

PRVAにも、同様の二つの側面からユーザーの信頼を得ることが求められることになる。PRVAの能力は、提示される情報量などから論理的に判断ができるが、人格は短期間では論理的に判断することが難しく、最終的には直感的に判断することになると考えられる。意思決定において、論理的に判断できるほどの材料がない場合などに、感情などをもとにして直感的な判断を行うというモデルは、PettyCacioppo[11]の精緻化見込みモデルでも採用されている。よって本研究では、ユーザーからPRVAを含めた擬人化エージェントに対する信頼の定義として、「相手が自分の利益を上げることのできる能力を持っており、かつ自分を裏切らないという期待」と定義する。これは能力への信頼と人格への信頼の双方の側面を持ち、かつ相手の行為そのものではなく本質的に持っている属性に対する信頼である。

2.2 エージェントの信頼状態モデルと遷移オペレータ

前節の議論を踏まえて、具体的なエージェントの設計を考える。信頼は論理的な判断による部分と直感的な判断による部分とに分けられると考え、それぞれの判断を強化するような手法を考える。論理的に判断できる能力は、「エージェントの知性」と言い換えられる。知的なエージェントほど信頼できるということは、Gevenら[6]によって示されている。知性を表出するには大量の正確な情報を提示するという方法も考えられるが、それでは非常に時間がかかる。本研究ではそれに代わって、Carnegie[3]が示している知的な話し方によって知性を高く見せることを目指す。具体的には、聞き手への質問、例え話の使用などの盛り込まれた話し方である。

一方、エージェントの人格に対する直感的な判断に大きく寄与しているのは、「ユーザーの感情」と考えられる。ポジティブな感情の時ほど相手を信頼するということは、DunnSchweitzer[5]によって示されている。また、擬人化エージェントの表情によってユーザーの感情が変化する、すなわち擬人化エージェントの感情がユーザーに伝染することはTsaiら[12]やDimasら[4]によって示されている。これらの先行研究から、エージェントにポジティブな感情を表出させて、ユーザーの感情をポジティブに遷移させ、それによってユーザーからエージェントへの信頼を高めることが可能であると考えられる。「エージェントの知性」と「ユーザーの感情」は、それぞれ論理的な判断と直感的な判断に関わるため、互いに独立であると考えられる。どちら

かを高めれば、通常状態よりも信頼が一段階上昇し、さらにもう片方も高めることによって信頼をさらに高くすることができるというのが本研究のモデルである。このモデルを示す図を図1に示す。

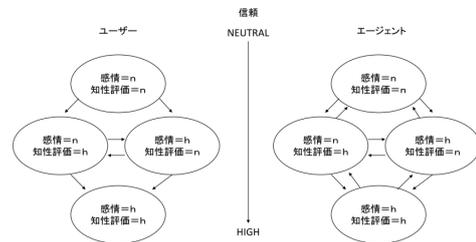


図1: ユーザーとエージェントの状態遷移モデル。番号が同じ状態同士はそれぞれ対応している。

このモデルでは、ユーザーの状態は自身の感情とエージェントに対する知性評価という二要素、エージェントの状態は自身の感情と知性という二要素で表現されている。それぞれの要素は、n(neutral)とh(high)という二状態で表現される。エージェントの各状態から他の状態への矢印方向への変化は、対応するユーザーの状態の変化を誘発する。例えばエージェントの感情がnからhへと変化することで、エージェントの状態がAからBへ変化すると、ユーザーの感情も同様にnからhへと変化し、ユーザーの状態もAからBへと遷移する。Aが最も信頼が低い状態であり、BとCは信頼が一段階のみ上がった状態、Dは信頼がさらに上がった最大になった状態である。

本研究では、このようなエージェント側の状態の変化を表す出力を遷移オペレーターと呼ぶ。遷移オペレーターはエージェントの状態遷移を表現すると同時に、ユーザーの状態変化を誘発するものである。

エージェントの感情がnからhに遷移したことを示す遷移オペレーターとしては、エージェントの表情(ほほえみ)とジェスチャーを実装する。エージェントの知性がnからhに遷移したことを示す遷移オペレーターとしては、Carnegie[3]における知的な話し方を実装する。我々は、このモデルを確かめるために実験を行った。

3 実験方法

3.1 参加者

実験は2015年9月11日に実施した。実験参加者は15人の工学部の学生および大学院生であり、全て男性であった。実験の所要時間は一人当たり約30-45分であった。

3.2 タスク

参加者は PRVA による日本国内の城への日帰り旅行の推薦動画を見て、その後アンケートに回答した。推薦とアンケートは全部で 8 回繰り返された。各動画の長さは 40 秒ほどである。PRVA は名古屋工業大学が公開している MMDAgent¹ を使用して作成し、エージェントも同サイトで公開されている”Mei” を使用した。推薦する城は実験を実施した大学から日帰り旅行が可能であり、財団法人日本城郭協会が選定した「日本 100 名城」² に選定されており世界遺産には登録されていない、かつ天守閣が現存もしくは再建されているという条件の中から選定した。これは利便性や知名度・おおよその外観などの条件を可能な限り揃えるためである。参加者は一つの城の推薦が終わる度に、以下の 8 つの質問をされた。

- 質問 1・あなたはこの城にどれくらい興味を持ちましたか？
- 質問 2・あなたはこのお城に行ってみたくありませんか？
- 質問 3・もしこのお城に行く、日帰りでおいしい昼食付の 19800 円の旅行パックがあったら、購入したいですか？
- 質問 4・あなたは説明中どれだけ楽しい気分になりましたか？
- 質問 5・あなたは説明中どれだけくつろいだ気分になりましたか？
- 質問 6・あなたは説明中にキャラクターは正しい知識を持っていると思いましたか？
- 質問 7・あなたは説明中にキャラクターがどれくらい賢そうだと思いますか？
- 質問 8・あなたは説明中にキャラクターは嘘をつかないと感じましたか？
- 質問 9・あなたは説明中にキャラクターがどれだけ信頼できると感じましたか？

質問 1 から 3 は購買意欲に関わるものであり、4 と 5 はユーザーの感情に、6 と 7 はエージェントの知性に、8 と 9 は信頼にそれぞれ関わるものである。また、被験者は推薦動画を見る前に、エージェントが推薦をせずに立っている 10 秒ほどの動画を見て、同様の質問に回答した。回答は 7 段階のリカード尺度で行われた。最も低く感じた場合は 1、最も高く感じた場合は 7 である。

前半の推薦では、遷移オペレーターのうち、PRVA の知性を表す知的な話し方は表出されるが、感情に関

する遷移オペレーターは表出されない。後半の 4 推薦では、知的な話し方に加えて、表情とジェスチャーという遷移オペレーターが表出される。

我々のモデルに照らし合わせれば、実験の後半でユーザーの感情がポジティブに変化し、同時にエージェントへの信頼も高まると予想される。前半の推薦と後半の推薦におけるエージェントの表情とジェスチャーの変化を図 2 に示す。

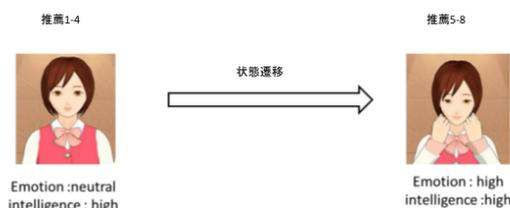


図 2: 前半 4 推薦におけるエージェントの外見と後半の 4 推薦におけるエージェントの外見。後半の推薦では、エージェントの表情とジェスチャーがポジティブな感情を表すものに変化する。

4 結果

4.1 解析方法

本研究では、ユーザーの感情・知性評価・信頼のそれぞれの状態を”neutral”か”high”かの二値で評価する。この二値の分け方は、推薦前のアンケート結果からの差によって行った。各質問項目について、各推薦後のアンケート結果の数値が推薦を始める前と比べてどれだけ増加しているかを計測した。そして、その増加量とその質問項目全体の平均値よりも高い場合を high、それ以外を neutral と定義した。このような手法を取ったのは、初期の状態からの変化を見ることに重点を置くためである。

4.2 解析結果

各推薦後の参加者の内部状態が、推薦前と比べてどれだけ変動したかを表に示す。ここでは感情・知性評価・信頼に関する結果を示す。感情は質問 4 と 5 の回答の平均値を、知性評価は質問 6 と 7 の回答の平均値を、信頼は質問 8 と 9 の回答の平均値をそれぞれ示す。括弧内の数値は分散を示す。この表に現れた数値を、上記の手法で high と neutral に分けた。上記の手法で high と定義された数値を灰色で示した。この表の数値を元に、参加者の感情・知性評価・信頼がどのように変化したかを図 3 に示す。

¹<http://www.mmdagent.jp/>

²<http://www1a.biglobe.ne.jp/nihonjokaku/castle/100meijo.html>

	感情	知性評価	信頼
推薦1	0.533(1.57)	0.233(1.52)	-0.8(1.44)
推薦2	1.17(1.44)	0.567(1.22)	-0.2(1.07)
推薦3	1.23(1.47)	0.733(1.28)*	-0.4(1.21)
推薦4	1.43(1.05)*	0.6(1.29)	-0.0667(1.24)*
推薦5	2.2(1.19)*	1.13(0.663)*	0.3(0.952)*
推薦6	1.7(1.11)*	0.867(0.977)*	0(0.907)*
推薦7	1.53(1.27)*	0.8(1.12)*	-0.0667(1.34)*
推薦8	1.4(1.56)*	0.7(1.49)	-0.267(1.40)
平均	1.4	0.704	-0.19

表 1: 感情・知性評価・信頼それぞれに関する質問への回答が、各推薦後に推薦前と比べてどれだけ増加したかを示す表、括弧内は分散である。平均以上に増加した数値を*で示す。

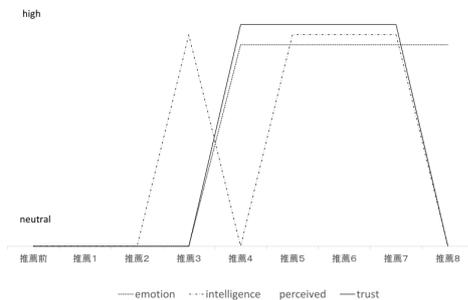


図 3: 感情・知性評価・信頼がどのように変化したかを示すグラフ。

5 考察

図 3 では、感情は推薦 4 の終了後に high に遷移し、そのまま推薦 8 の終了時まで変化しなかった。一方知性評価については、推薦 3 の後でいったん high に遷移しているが、その次の推薦 4 の後には neutral に戻っている。その後、推薦 5 の後に再び high に遷移するが、最後の推薦 8 の次には neutral に再び戻った。最後の信頼については、感情と同じく推薦 4 の後に high に遷移し、その後知性評価と同じく推薦 8 の後に neutral に戻っている。この結果は、我々のモデルと概ね一致していると言える。この図 3 に現れた遷移の経路を図 4 に示す。

図 3 の中で、知性評価は推薦 3 の後で high になった後推薦 4 の後で再び neutral に戻っている。この時感情は high に遷移していることから、感情と知性評価は互いに独立ではなく、干渉しあっている可能性が考えられる。しかしその後は双方とも high の状態を維持していることから、この干渉は一時的なものであると考えられる。推薦 8 の後では、知性評価と信頼の状態が neutral に戻っている。これは実験時間が長かったための飽きの効果であるとも考えられるが、この説明は感情は high のままであるという結果と合致しない。もう一つの可能な説明は、感情が信頼に与える効果は限定

的であり、信頼の低下には知性評価がより大きく関わるといえる。

この実験では参加者の信頼状態は、最も低い状態から一段階高い状態を経て、最も高い状態に遷移している。「一段階だけ高い状態」の必要性は、Allen[1] の user-model と mental-model のインタラクションモデルと適合している。Allen のモデルでは、user-model はコンピューター的设计者が持つユーザーの内部状態に関するモデルであり、mental-model はユーザーが持つコンピューターの内部状態に関するモデルである。ユーザーはコンピューターを使用する中で、入力に対してコンピューターからもたらされる出力を受け取り、それに応じて一度の出力ごとに mental-model を更新していくとされる。すなわち、mental-model は一度に大きく更新することはできず、コンピューターとの相互作用の中で徐々に更新されていく。これに照らし合わせれば、エージェントに関する信頼モデルも、エージェントとの相互作用の中で徐々に更新されるものであると言える。信頼を一度に最大の状態に上げるよりも、一段階ずつ上げるほうが Allen のモデルには適合している。

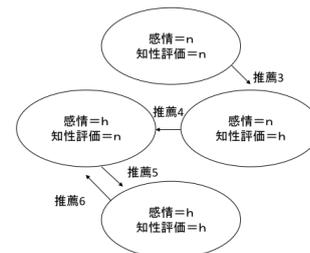


図 4: 実験で再現された遷移の経路。信頼は推薦 3 の後 1 段階上昇したが、次の推薦では信頼のレベルは変化しないままに遷移が起こっている。その次の遷移で信頼が最も高い状態に到達した。

6 まとめ

本研究では、ユーザーの擬人化エージェントへの信頼を、ユーザーの感情とエージェントの知性によって操作することができることが示された。今回はエージェントの感情は表情と動きで、知性は話し方で表出した。今後は、特に知性についてはよりよい表出の仕方を考える必要があると考える。また感情と知性評価を干渉させずに共に上昇させる方法についても、さらに検討を進めていきたい。

参考文献

- [1] Robert B Allen. Mental models and user models. *Handbook of human-computer interaction*, 1:49–63, 1997.
- [2] Donovan Artz and Yolanda Gil. A survey of trust in computer science and the semantic web. *Web Semantics: Science, Services and Agents on the World Wide Web*, 5(2):58–71, 2007.
- [3] Dale Carnegie. *Public speaking for success*. Penguin, 2006.
- [4] Joana Dimas, Gonçalo Pereira, Pedro A Santos, Rui Prada, and Ana Paiva. I’m happy if you are happy.: a model for emotional contagion in game characters. In *Proceedings of the 8th International Conference on Advances in Computer Entertainment Technology*, pages 2:1–2:7, 2011.
- [5] Jennifer R Dunn and Maurice E Schweitzer. Feeling and believing: the influence of emotion on trust. *Journal of personality and social psychology*, 88(5):736, 2005.
- [6] Arjan Geven, Johann Schrammel, and Manfred Tscheligi. Interacting with embodied agents that can see: how vision-enabled agents can assist in spatial tasks. In *Proceedings of the 4th Nordic conference on Human-computer interaction (NordiCHI’06)*, pages 135–144, 2006.
- [7] Tyrone Grandison and Morris Sloman. A survey of trust in internet applications. *Communications Surveys & Tutorials, IEEE*, 3(4):2–16, 2000.
- [8] Dan J Kim, Donald L Ferrin, and H Raghav Rao. A trust-based consumer decision-making model in electronic commerce: The role of trust, perceived risk, and their antecedents. *Decision support systems*, 44(2):544–564, 2008.
- [9] Lik Mui, Mojdeh Mohtashemi, and Ari Halberstadt. A computational model of trust and reputation. In *System Sciences, 2002. HICSS. Proceedings of the 35th Annual Hawaii International Conference on*, pages 2431–2439. IEEE, 2002.
- [10] Daniel Olmedilla, Omer F Rana, Brian Matthews, and Wolfgang Nejdl. Security and trust issues in semantic grids. *Semantic grid*, 5271, 2005.
- [11] Richard E Petty and John T Cacioppo. *The elaboration likelihood model of persuasion*. Springer, 1986.
- [12] Jason Tsai, Emma Bowring, Stacy Marsella, Wendy Wood, and Milind Tambe. A study of emotional contagion with virtual characters. In Yukiko Nakano, Michael Neff, Ana Paiva, and Marilyn Walker, editors, *Intelligent Virtual Agents*, pages 81–88. Springer, 2012.
- [13] Toshio Yamagishi and Midori Yamagishi. Trust and commitment in the united states and japan. *Motivation and emotion*, 18(2):129–166, 1994.