

チューリングテスト「合格」のシステム

東中竜一郎（NTT メディアインテリジェンス研究所）

2014年6月 Eugene Goostman と呼ばれるチャットシステム（チャットボットとも呼ぶ）がチューリングテストに世界で初めて合格したとの報道がなされた。このニュースはネット上でも話題となり、ついに人工知能が実現したのではといった憶測も呼んだ。本稿では、チューリングテストについて簡単に説明するとともに、対話システム研究者の立場から、チューリングテスト合格のニュースの詳細、および、合格したというシステムについて解説する。また、今後の展望として、対話システム研究の現状と課題について述べる。

チューリングテスト

アラン・チューリングは1950年に“Computing Machinery and Intelligence”という論文³⁾の中で「コンピュータは思考することができるか」という問題を扱い、コンピュータが思考しているとみなせるかどうかを判定するためのテストとして Imitation game を提案した。Imitation game とは、パーティゲームの1つで、男性と女性、および、質問者がいるとき、質問者が別室から男性、女性のそれぞれとテキストメッセージのやりとりを行い、どちらの対話相手が男性で、どちらが女性かを当てるといったゲームである。ここで、男性・女性の一方はもう一方を模倣する。たとえば、男性が女性を模倣する場合、男性は女性のように発言し、女性は女性であることを信じてもらうように発言する。

チューリングは、このゲームでの男性、もしくは

は、女性をコンピュータに置き換えて同ゲームを行い、もし質問者が、人間同士がゲームをする場合と同程度に判定を誤るのであれば、コンピュータが思考していると判断してよいとした。なぜなら、人間か機械かどうかが分からないレベルの質疑応答ができるようなコンピュータは、客観的に見て、思考していると判定せざるを得ないと考えられるからである。この提案以降、対話相手が人間かコンピュータか分からないかどうかでコンピュータが思考しているかを判定するテストのことを「チューリングテスト」と呼び、コンピュータの知性を測るテストとして広く知られるようになった。

同論文の中でチューリングは1つの予想をしている。それは、「50年くらい経てば、5分程度の質疑応答では、質問者は対話相手が人間かコンピュータかを70%も当てられないだろう」というものである。このことから、質問者の30%以上が、対話相手が人間かコンピュータか判断がつかないということがチューリングテスト合格の1つの基準とされている。

チューリングテストに合格するような人工知能を実現することは大きなチャレンジであり、そのためのコンテストも開かれている。ローブナー賞^{☆1}はその1つであり、1991年以降、毎年開催されている。このコンテストには、さまざまなチャットシステムが出品されテストを受けている。ローブナー賞以外にも、チューリングにまつわる記念イベントではチ

☆1 <http://www.loebner.net/Prizef/loebner-prize.html>

ューリングテストが開催されることがある。チューリングは見た目の影響を取り除くためテキストメッセージでのテストを提案したが、音声や他のモダリティを扱うチューリングテストも考えられている。

チューリングテスト合格!?

2014年6月7日、Eugene Goostman^{☆2}と呼ばれるチャットシステムが、Turing Test 2014 というイベントにおいてチューリングテストに世界で初めて合格したと報道された^{☆3}。このイベントは、チューリングの没後60年を記念したイベントとして英国のレディング大で開かれたものである。30名の審査員がこのシステムと会話をし、そのうち10名がこのシステムを人間と判断した。33%の審査員が判断を誤ったのであるから、チューリングテスト合格の1つの基準とされる30%を達成したことになる。よって「チューリングテスト合格」と発表されたのである。

Eugene Goostman は、13歳のウクライナの少年を模したチャットシステムで、英語を話すが、英語は第二言語という設定である。Eugene Goostman は2001年に開発がスタートした。ローブナー賞にもたびたび参加しており、優勝したことはないものの2001年から2008年の間に5回ファイナリスト（ローブナー賞には予選がある）になっている。2012年のチューリング生誕100周年の記念に開催されたイベントでは、29%を人間だと思い込ませることに成功し、30%の達成は近いと考えられていた。

本システムの技術資料^{☆4}によれば、本システムはパターンマッチに基づく処理がメインで、基本的には、よくあるチャットシステムと同様、パターンに反応することで応答をする。ただ、ユーザのタイプのミス処理や、複雑なパターンマッチのルールを

簡単に書けたり、外部のデータベースを参照して応答を生成する機能などが充実しており、開発者は効率的に応答ルールを書けるという特長がある。開発者の1人であるVladimir Veselovは、システム開発にあたって、人間らしい応答の実現に力を入れたと語っているが、今回の「合格」は、人間らしい応答をするためのルールを労力をかけて書ききった結果であると言えるだろう。

さて、今回の「合格」はどのような意味を持つのだろうか。そもそも「合格」なのだろうか。従来に比べ、より多くの審査員を騙せたのだから、進歩であることは間違いない。対話研究者の視点から見れば、どれほど頑張ってルールを書いても、ユーザの多様な発言を処理しきれずにすぐにコンピュータとばれてしまうことがほとんどである。そんな中、短い時間であっても人間と間違えるような入出力が実現できたこと自体、素晴らしい成果であると言ってよい。しかし、思考ができるコンピュータが実現できたかという点、そうとは言いがたい。批判のポイントとして以下の3つが考えられる。

1つ目は「13歳のウクライナの少年」という設定である。つまり、子供であるため知識がそれほど準備されていなくてもよく、また、英語は第二言語であるため、おかしい発言をしても、対話相手の発言を理解できなくても許容されやすい。システムに特殊な設定を持たせることで、対話能力の至らない点をカモフラージュするテクニックは、チャットシステムでよく用いられる。たとえば、1991年の初回のローブナー賞で優勝したシステム“Whimsical Conversation”は、jester（道化師）を真似るシステムであった²⁾。道化師はおかしなことを言っても問題と思われにくい。対話システムの始祖として有名なELIZAは、セラピストという設定であり、基本的にユーザの発言を繰り返すだけであるが、多くのユーザが自分のことを理解してくれていると解釈し、システムを人間のように感じてしまった。PARRYは、統合失調症の患者を真似たシステムであり、論理に整合性のない発言を行うが、医者も対話相手が人間の患者であるのかPARRYであるのか区別が

☆2 <http://www.princetonai.com/>

☆3 <http://www.reading.ac.uk/news-and-events/releases/PR583836.aspx>

☆4 <http://www.alicebot.org/chatbots3/Eugene.pdf>

つかなかった¹⁾。今回のような設定が許容されるなら、すでに合格したシステムがいくつもあることになる。

2つ目は、テストの時間についてである。Turing Test 2014 では、オリジナルのチューリングテストに即して、5分という時間制限でテストが行われた。また、オリジナルに合わせて審査員は人間とコンピュータの両方と同時に会話をを行った。すなわち、各システムとの対話時間はおよそ2分半である。タイピングによる対話であることを考えれば、それほど多くのやりとりができる時間はなかったと思われる。つまり、システムは、審査員と出会ってすぐの短い時間を乗り切るだけのプログラムが実現できれば、チューリングテストに合格できてしまう。基本的に、初対面の対話は文脈が乏しく、深くなりにくい。もし5分経ったあと、次の5分をまともに話せないとすれば、それはかなり限定された知性と言えるだろう。ローブナー賞創設者の Hugh Loebner も、より長い時間のテストが必要と指摘している。

3つ目は、審査員の質である。2008年のローブナー賞では、Elbot というシステムが12名中3名の審査員を騙すことに成功した。これは、25%であるため、2009年にもチューリングテストに合格するシステムが現れるのではないかという憶測があった。しかし、2009年には、前年を受けて審査員が「まじめ」に対話をし、審査員側の圧勝に終わっている。審査員の質を完全に統制することは難しいかもしれないが、異なる審査員のセットで何度も追試することは必要であろう。余談であるが、2009年のローブナー賞における審査員が執筆した「機械より人間らしくなれるか？」⁴⁾では審査の内実が面白く描かれておりお薦めである。ローブナー賞では、最も人間らしいコンピュータだけではなく、最も人間らしい審査員にも賞(The most human human award)が与えられることになっている。審査員による、コンピュータと思われないための奮闘ぶりが愉快であり、また、人間らしい会話とは何かについて考えさせてくれる。

チューリングはチューリングテストの提案に際し

て、システムの設定や具体的な評価手順について言及していない。よって、特定の実験設定において、Eugene Goostman はチューリングテストに合格したと言って良いだろう。しかし、特定の設定に依存していることから、世界初の合格は言い過ぎだろう。チューリングは「人間かコンピュータかを70%も当てられない頃には機械が考えているかどうかを議論することが無意味なレベルになっているだろう」とも述べている³⁾。今回のニュースが多くの議論を呼んでいることから、本システムがまだそのレベルに到達していないことは明らかである。今後、さまざまな設定や評価手順でチューリングテストは実施されていくだろう。システムによる合格が当たり前になったとき、チューリングテストは本当に「合格」となるのだと思う。

対話システム研究の現状と課題

ELIZA以降、自由な会話をするシステムは望まれながらも、技術的な難しさから研究が進んでいるとは言いがたい。ローブナー賞に参加しているシステムも、最先端の人工知能技術の結晶というよりは、一部のシステムを除き、チャットシステム愛好家によるエンジニアリング（主にルール記述の労力）の結晶といった状況である。

これは、自由に会話をするための要素技術が十分でないからである。発話の意味をどう表現すればよいか明らかでないし、複数の発話からなる談話をどう扱えばよいかはさらに不明である。会話における状況やユーザの状態を理解したり、状況に合わせてどのような行動をとるべきかという理論もほとんど確立されていない。問題が難しすぎて、一般的な解法が見つからないために、対話システム研究者は、自由な会話をするシステムは諦めて、タスクを限定することで実用的な対話システムを作ってきたと言える。タスクを限定することでユーザ発話の意味は決定できる。たとえば、乗り換え案内タスクにおいて「東京まで」と言えば、「東京までの経路を知りたい」のことであると解釈できる。Siri やしゃべっ

てコンシェルは、実用化されたタスク指向型対話システムの一例である。これらのシステムは、多少の雑談めいたこともできるが、それはエンジニアリングの結果である。

近年、スマートフォン上の音声エージェントシステムが普及するにつれ、自由な会話をするシステムに対する要望が増えている。しゃべってコンシェルでは、雑談がメインの機能でないにもかかわらず、雑談をしようとするユーザも多い⁵⁾。また、高齢世帯や独居世帯の増加に伴い、雑談ができるシステムのニーズも増加している。タスク指向型対話システムであっても雑談を挟むことでタスクが達成されやすくなったり、システムへの愛着が増すという結果も報告されている。

ニーズが顕在化してきていることから、今後、自由な会話をするシステムの研究が増えていくだろう。筆者も数年前から雑談対話システムのプロジェクトを始めている⁶⁾。また、今年から、Project Next NLP^{☆5}と呼ばれる言語処理分野におけるエラー分析のワークショップが立ち上がっており、その中に対話タスクというものがある。筆者を含む関係者で立ち上げたタスクであるが、人間とシステムによる雑談を分析し、対話の破綻箇所を自動的に検出することが目標の1つとなっている。破綻箇所が分かるということは、破綻しないように会話ができることにつながる。この営みが、日本語の対話システムのチューリングテスト合格につながるのではないかと考えている。

高まる期待

本稿では、チューリングテストについて紹介し、チューリングテストに世界で初めて合格したとされる Eugene Goostman というチャットシステムについて説明した。今回の報道された内容は素晴らしい成果ではあるが、限定された上でのチューリングテスト合格と考えられることについて述べた。加えて、これまでの対話システム研究があまり自由な会話を行うシステムに注力していなかった状況について述べ、また、近年雑談の研究が立ち上がろうとしていることについても触れた。次回のローブナー賞は2014年11月15日に英国のブレッチリー・パークで行われる。ブレッチリー・パークはチューリングが暗号解読を行った場所としても有名である。そのような場所で、今年どのような結果が見られるか楽しみである。

参考文献

- 1) Colby, K. M. : Human-computer Conversation in a Cognitive Therapy Program, In *Machine Conversations*, pp.9-19. Springer (1999).
- 2) Epstein, R. : The Quest for the Thinking Computer, *AI Magazine*, Vol.13, No.2, pp.81-95 (1992).
- 3) Turing, A. M. : Computing Machinery and Intelligence, *Mind*, pp. 433-460 (1950).
- 4) ブライアン・クリスチャン (著), 吉田晋治 (翻訳): 機械より人間らしくなるか? : AI との対話が、人間でいることの意味を教えてくれる, 草思社 (2012) .
- 5) 大西可奈子, 東中竜一郎, 吉村 健: コンピュータとの自然な会話を実現する雑談対話システム, 人工知能学会研究会資料 SIG-SLUD-B303, pp.39-40 (2014).
- 6) 東中竜一郎: 雑談対話システムに向けた取り組み, 人工知能学会研究会資料 SIG-SLUD-B303, pp.65-70 (2014). (2014年7月15日受付)

謝辞 本稿のドラフトについて有益なコメントをいただいた南泰浩氏、堂坂浩二氏、駒谷和範氏、中野幹生氏、および、音声・言語基盤技術グループの諸氏に感謝します。

● 東中竜一郎 (正会員) higashinaka.ryuichiro@lab.ntt.co.jp

日本電信電話 (株), NTT メディアインテリジェンス研究所にて勤務。博士 (学術)。質問応答システム・音声対話システムの研究に従事。しゃべってコンシェルの Q&A 機能の研究開発に携わる。

☆5 <https://sites.google.com/site/projectnextnlp/>