コンピュータと数学とAIと
～“形式化数学”敢えて人間の直観を取り除くことで見えてくるもの～

国立大学法人茨城大学工学部　宮島　啓一

1. “形式化数学”？？

“形式化数学”、聞き慣れない言葉だと思います。それもそのはずで、本格的に一学問分野として立ち上がってからまだ二十数年しか経っていない、まだまだ新しい発展途上の学問分野です。

2. “形式化”って？

分野のタイトルに“数学”の文字が入っているのですから、何か数学に関係する分野なのであろうことは想像がつくと思います。

では、その前にある“形式化”とはなんでしょう？

結論を、それもかなり簡潔に述べますと、形式化数学の“形式化”とは**数学の証明をコンピュータ言語化**することです。

知っての通り、数学には様々な“定義”や“定理”、そして“証明”があります。

そういったものを全てコンピュータ言語化し、その**正しさ**をコンピュータで検証（チェック）させるのが、“形式化数学”なのです（“形式証明”と呼ぶこともあります）。

3. なぜ数学を“形式化”するのか？（その1）

結論から先に述べれば、

「数学者でも間違いを起こすことはあるから」

です。

1995年に「フェルマーの大定理（最終定理）」が完全に証明されたとして全世界で話題になったニュースがありました。この「フェルマーの大定理」とは一体どんなものかというと、

「3以上の自然数 n について、$x^n + y^n = z^n$ となる 0 でない自然数（x, y, z）の組み合わせは存在しない」

というもので、一見すると非常に単純で簡単そうな問題に見えますが、きちんと（厳密に）証明しようとすると、実は極めて難しいことが知られていた問題です。

この問題を造った（発見した？）フェルマーとは多くの業績を残した17世紀のフラ

ンス人数学者（1601～1665）であり、彼はこの問題について

「この定理に関して、私は真に驚くべき証明を見つけたが、この余白はそれを書くには狭すぎる」

と書き残して亡くなってしまいます。

彼の死後多くの数学者たちがこの問題に挑み続けることになりますが、実に300年以上にわたって未解決のまま残されるというまさに難問中の難問となってしまいました。

結局、20世紀も終わりになってようやくワイルズというイギリス人数学者によって最終的な解決を見るわけですが、その証明法はフェルマーが当初想定していたと推測される証明の方法では絶対に証明出来ないものでした。

なぜなら17世紀当時の数学では存在しない様々な概念や理論を使わなくてはならなかったためです。

ということは、フェルマーが間違っていたか間違っていなかったか？という2分法で見るならば、このような2分法が適当であるかどうかはこの際おいておくとして、**「フェルマーは間違っていた」**という結論にならざるを得ないということになります。

実はこのようなことは数学の世界ではよくある話で、理論を書いた**数学者の死後30年も経ってから**その理論の証明部分に誤りが発見された、ということがあったりします。

このように、**後世に名を残すような偉大な数学者であっても**人間である以上**誤りは起こりえる**のです。

無論、後から誤りが発見されたからといって、その数学者の業績が全て否定されるものではありません。

4. なぜ数学を“形式化”するのか？（その2）

もう一つお話をしましょう。

1990年に森重文（もり しげふみ）京都大学名誉教授が（数学のノーベル賞と言われる）フィールズ賞を受賞されたとき、森教授の理論を理解出来るのは世界に10人いない、と報道されました。

それならば、理解出来る人が世界に**10人といない理論の正しさを**

一体何処の誰が保証してくれるのでしょう？

先に述べたように、**“偉い人”が証明したからといって、必ずしもそれが絶対に正しいとは限らない**のです。

（蛇足ですが、例に挙げた森教授の業績は既に複数の人間の手によって検証が

済んでおり、その正しさは確認されています。）

5. なぜ数学を“形式化”するのか？（結論）

このように数学の世界では、意図的なねつ造は論外として、しばしば誤ったままの証明が発表されてしまうということがあります。

一方で、数学は同時に最先端の科学・技術を支える強力な道具でもあります。

特に我々が日常的に使用しているインターネットを含むコンピュータシステムなどは言ってみれば**巨大な数学の塊**と言っても過言ではありません。

その巨大なコンピュータシステムを支えている数学に、**実は誤りがあった**、などということがあったらどうでしょう？

これは大変なことになる、と想像出来ませんか？

ところで、プログラミングを学んだことがある方であれば経験があると思うのですが、コンピュータってとても融通が利きませんよね。

ほんの一文字間違っていても「エラーです」と言って動いてくれない。

「人間だったらコレくらい融通を利かせてくれるのに～～～！！」

って思ったことありませんか？

私はもう数えきれないほどあります（笑）。

数学を形式化し、コンピュータという**敢えて全く融通が利かないものに数学の証明をチェックさせる**意義とはまさにこの部分にあります。

人間の直観を排し融通が利かないコンピュータだからこそ、数学の証明という誤りが許されないところを徹底的に検証させるという点で、極めて優れた力を発揮します。

6. AIについて

ここまで“形式化数学”という研究分野について説明してきました。

ここからは今私が行っている具体的な研究についてお話ししたいと思います。

近年、人工知能AIが爆発的に普及していることは皆さんご存じだと思います。

文章や画像の生成AIをはじめ、自動運転、音声認識、自動翻訳、画像認識、データ分析、予測、さらにカスタマーサポートなど様々な分野に活用されつつあります。

さてこのAIですが無論「最初から非常に優秀」というわけではありません。

皆さんも聞いたことがあると思いますが“学習”をすることによって、“優秀”になって

いく、というシステムになっています。

7. AIの学習と学習の“収束”

ではこのAIの“学習”とは一体どういうものなのでしょうか？

AIの学習は“繰り返し学習”と言って、同じことを何度も何度も、それこそ数百どころか数万回、場合によっては数億回というレベルで繰り返し学習を行って、その精度を高める（優秀になっていく）という仕組みです。

人間と違って「飽きる」ということが無いので使える技ですね（笑）。

この仕組みを簡単な例を使って説明します。

例えば「1＋1＝2」を学習させたいとしましょう。

そうするとAIは

1回目の学習では　1＋1＝1　だったものが

2回目の学習で　1＋1＝1.5

3回目の学習で　1＋1＝1.7

4回目の学習で　1＋1＝1.8

5回目の学習で　1＋1＝1.88

.

.

.

と、このように「だんだん正解の2に近づいていく」という感じで学習していきます。

では

100回目の学習で　1＋1＝2　となりました。

1000回目の学習でも　1＋1＝2　のままです。

10000回学習させても　1＋1＝2　のままです。

.

.

.

10001回目の学習をさせたとき突然

1＋1＝100　です！

と言い出すことは無いのでしょうか？

私が大学で行っていることは

「そのような10001回目は**“絶対に起こらない”**」

ということを**数学で厳密に証明しよう**、という研究をしています。

それどころか1億回学習させても、1兆回学習させても……、∞回学習させても

1＋1＝2　　と答え続ける。

（これを学習の“収束”と言います）ということを数学で証明したいと考えています。

これはAIの安全性に関する研究の一つです。

8. 証明の過程で解ったAIの弱点

最後にこの証明の研究を行っている途中で解った、AIのある“弱点”についてお話しようと思います。

先ほどのAIの“学習”の説明で、実は私は意図的に書かなかったことがあります。

それは「1＋1＝？」という質問に対して、無条件に「“正しい”答えは2である」としている点です。

何をそんな当たり前のことを言っているのだ？　と思われるかもしれません。

でもちょっと考えてみてください。

AIが学習しているとき

「1＋1の答えは2が“正しい”んだよ」

と教えているのは**一体誰なのでしょうか？**

答えはそう、**人間**です。

人間が「1＋1の答えは2だよ」と教えているのです。（もちろん1回1回人間が手で入力する形で教えているわけではありません。まとまったデータという形で入力しています。）

このとき、「1＋1は100が“正しい”んだよ」としてデータを入力し学習させると、AIは「1＋1＝100」と答えるようになります。

実はAIは**学習するとき、そのデータが“正しい”か“正しくない”かをAI自身は判断していません。**

AIは「人間によって与えられる学習データは“無条件に正しい”」という前提で学習を行っています。

つまり初めから誤ったデータで学習し続けると、そのまま誤った結果を出力してしまう、という性質を持ちます。

これは非常に危うい性質だと思いませんか？

ここで上げた例では単純な足し算なのでAI以外の手段で簡単に正しい答え求めることができますが、もっと複雑な問題だったらどうでしょう？

実際、X（旧Twitter）社のAI「Grok」がヒトラーを礼賛する発言を度々する、ということで問題になりました。（https://www.techno-edge.net/article/2025/

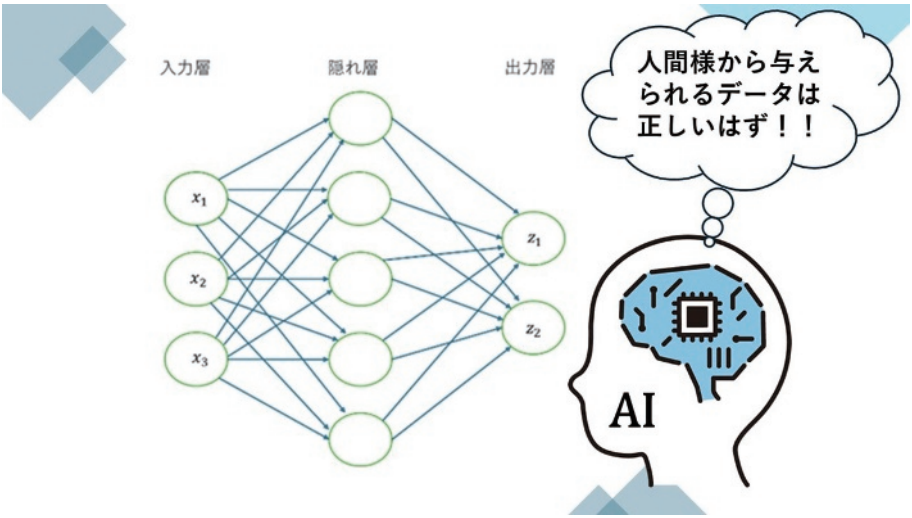

図1. AIは人間から与えられた学習データの“正しさ”をAI自身では判断していない！

07/09/4469.html）

おそらく何者かがヒトラーを礼賛する発言を出力するのが“正しい”として繰り返し学習させたのだろうと推測されます。

（もちろんX社の技術者達も先で説明したような単純な学習方法ではそんな出力はさせないようにいろいろ工夫していたハズですが……）

9. まとめ

AIというのは上手に使えば非常に便利で強力な“道具”です。

しかし所詮は人間が作った“道具”であり、決して完璧なものではありません。

そのことを踏まえたうえで、我々はAIを利用していかなくてはなりません。

重要なことは**我々人間がAIを利用するのであって、その逆ではない**のです。

著者紹介　宮島　啓一（みやじま　けいいち）

2006年4月より茨城大学工学部。数学の証明をコンピュータ言語化（形式化）する研究に従事。現在、AIの学習が収束することを数学的に証明する研究を行っています。