

Google翻訳の性能が大幅に向上した理由とは？

-- 自然言語処理に応用されるディープラーニングの原理と限界

執筆者 KDDI 総合研究所 特別研究員 小林雅一

▼記事のポイント

<サマリー>

グーグルが提供する翻訳サービス「Google 翻訳」の性能が、2016 年 11 月のリニューアルに伴い、大幅にアップした。特にスマートフォンから入力する会話文の場合、たとえば外国に出張したビジネス・パーソンが現地のスタッフと打ち合わせをしたり、運悪く急病にかかった場合に現地病院の医師らと意思疎通が可能なレベルにまで、その翻訳精度は向上した。

一方で外国メディアの記事や外国語で書かれた小説など長文を日本語に訳そうとする場合などには、Google 翻訳の性能はまだ不十分と言わざるを得ない。両者の落差はどこから来るのか？本レポートでは、Google 翻訳の性能が大きく向上したにも関わらず、翻訳対象となる文章の長さなどに応じて、その性能に大きな違いが出る理由を、グーグル翻訳の仕組みにまで踏み込んで解説する。

<主な登場人物>

グーグル Google Andrew Ng New York Times

<キーワード>

Google 翻訳 ディープラーニング Deep Neural Network 機械翻訳 自然言語処理 パターン認識 回帰分析 トレーニング・セット 過剰適応（過学習） overfitting スケーラビリティ

<地域>

米国 日本 世界

What is the reason behind the significant increase in the performance of Google Translate? – The Principles and Limits of Deep Learning Applied to Natural Language Processing

Masakazu Kobayashi

Research Fellow, KDDI Research Inc.

Abstract

With its November 2016 renewal the performance of Google's translation service "Google Translate" has been significantly improved, especially so in the case of conversation text input from a smartphone. For example, a business person who unfortunately suffered a sudden illness while travelling abroad was able to consult with local staff and communicate with the doctors of the local hospital. A significant improvement in translation accuracy.

On the other hand, for example, when trying to translate a long sentence into Japanese, such as those found in a media article or novel written in a foreign language, the performance of Google Translate is still inadequate. Why does the technology perform poorly under these conditions? Despite the great improvement in the performance of Google Translate, accuracy is highly dependent on such factors as the length of sentences, and this report seeks to explain this shortcoming by delving into the structure of Google's translation service.

Main Players

Google Andrew Ng New York Times

Keywords

Google Translate deep learning Deep Neural Network machine translation natural language processing pattern recognition regression analysis training /set over adaptation (over learning) overfitting scalability

Region

US Japan world

以下は、参考のため Google 翻訳によるサマリー。

The following is a summary by Google translation for reference.

What is the reason why the performance of Google Translate has improved significantly? - Principle and Limit of Deep Learning Applied to Natural Language Processing

Masakazu Kobayashi

Research Fellow, KDDI Research Inc.

Abstract

The performance of the Google translation service "Google Translate" provided by Google has been significantly improved with the renewal of November 2016. Especially in the case of a conversation text input from a smartphone, for example, a business person who traveled abroad consulted with local staff, to a level where communication with the doctors of the local hospital is possible when it is unfortunately suffered from a sudden illness , Its translation accuracy improved.

On the other hand, for example, when trying to translate a long sentence into Japanese, such as a foreign media article or a foreign language written novel, the performance of Google translation is still inadequate. Where do the falls of both come from? In this report, despite the great improvement in the performance of Google Translate, why there is a big difference in performance depending on the length of sentences to be translated, etc. will be stepped into the structure of Google translation and explained .

1 はじめに

2016年11月にリニューアルされた「Google翻訳」が、それまでに比べて著しく性能がアップしたと評判になっている。過去には外国語での短い挨拶や極めて単純な日常会話などには対応できたが、たとえばビジネス・パーソンが海外業務などに使うための、少し長めで複雑な文章になると歯が立たなかった。

リニューアル後はこの問題がかなり解消され、ある程度、複雑な文章も相応に訳せるようになってきた。特にスマートフォンから音声入力モードで使うと、その便利さが実感できる（翻訳結果も音声出力される）。以下は試しに筆者が入力してみた日本語文章と、それらをGoogle翻訳が英訳した結果である。

為替レートが1円、円高になると、私の会社は100億円の損失を出します。
If the exchange rate gets higher by 1 yen, my company will lose 10 billion yen.

あの会社は取引先として信用できますか？
Can that company be trusted as a business partner?

空港から御社に行くには地下鉄がいいですか、それともタクシーの方がいいですか？
Is it better for the subway or taxi to go from the airport to your company?

以上の英訳文は厳密には、あるいは文法的には間違っているかもしれないが、意味は何とか相手に通じそうだ。次に英語から日本語への翻訳もやってみた。

If you want to accomplish this work as soon as possible, you need to read these two manuals.

できるだけ早くこの作業を行うには、これら2つのマニュアルを読む必要があります。

What I was trying to say was that he would have to complete this report by tomorrow.

私が言っていることは、彼が明日までにこの報告書を完成させなければならないということでした。

If you can not borrow money from me, how do you live from tomorrow?

あなたが私からお金を借りることができないなら、あなたは明日からどのように住んでいますか？

3番目のケースは明らかに誤訳である。正しくは「・・・明日から、どのように生きていくのですか？」。他の英・和文でも試してみたが、いずれの翻訳も概ね、以上と同じレベルだ。ただGoogle翻訳は世界中のユーザーによる使用データを機械学習して日々進化しているので、後で改めて同じ原文で試してみれば、上記とはかなり違う翻訳結果が表示されるかもしれない。

全体の感触としては、「海外出張したビジネス・パーソンが、かろうじて使えるかもしれない」といったレベルではないだろうか。それでも過去のGoogle翻訳に比べると、確かに大きく改善したという印象を受けた。

ただし、気になるのは翻訳の質よりも、インターフェース、つまり使い勝手の方だ。たとえばスマホに向かって文章を入力している途中で言い淀んで、「あー」とか「えー」とか言っているうちに、Google翻訳の方では勝手に見限って処理を始めてしまう。

それと見ていて面白いのは、スマホの画面に推敲中の翻訳文が次々と表示されることだ。語順を入れ替えたかと思えば、違う単語を出してみたり、さっきまでとは全く別の訳文が表示されたりする。「こうしょうか、ああしょうか」と人工知能が試行錯誤している様子が窺えるのだ。

2 高度な英語長文の和訳には難がある

一方、パソコンの方から、より長く複雑な英文を入力してみた。こうした文章となると、《リニューアル後の現在でも》Google翻訳の精度はかなり落ちてくる。以下は、つい先日のNew York Times記事の一節をGoogle翻訳で和訳した一例だ。

Meeting Between Japan and Russia Ends With Stalemate on Disputed Islands

By MOTOKO RICH DEC. 16, 2016

TOKYO — The Japanese news media gave breathless coverage to President Vladimir V. Putin's two-day visit to Japan this week, training television cameras on the airport in Yamaguchi Prefecture where the Russian leader arrived more than two hours late. Schoolchildren in Nagato, Prime Minister Shinzo Abe's hometown, ate borscht and piroshki in honor of Mr. Putin's trip to hot spring baths there. In Tokyo on Friday, Kodokan, the premier judo school, received the Russian leader, a black belt, for a demonstration and a meeting with Olympic medalists.

日本とロシアの間の紛争は、紛争を起こした諸島の膠着状態で終わる

MOTOKO RICH DEC. 16, 2016

プーチン大統領は2日、日本を訪問し、山口県の空港でテレビカメラを訓練し、ロシアの指導者が2時間以上遅れて到着したと報じた。プーチン大統領の温泉への旅に敬意を表して、長門の学童、安倍晋三首相の故郷、ボルシェットとピロシキを食べました。東京では金曜午後、柔道校の鼓道館がロシアの指導者黒帯を受け、デモやオリンピックメダリストとの会談を行った。

残念ながら、上の訳文では、まだ実用には不十分だろう。確かに部分的には正しく訳されている箇所もあるが、要所々々で誤訳や訳し抜けが起きているため、全体としては意味不明の奇怪な文章になっている。逆に実用に耐える和訳がどのレベルかを示すため、以下、筆者が同じ英文を訳してみた。

領有権が争われている島々に関する、日露首脳会談は膠着状態に終わる

モトコ・リッチ 2016年12月16日

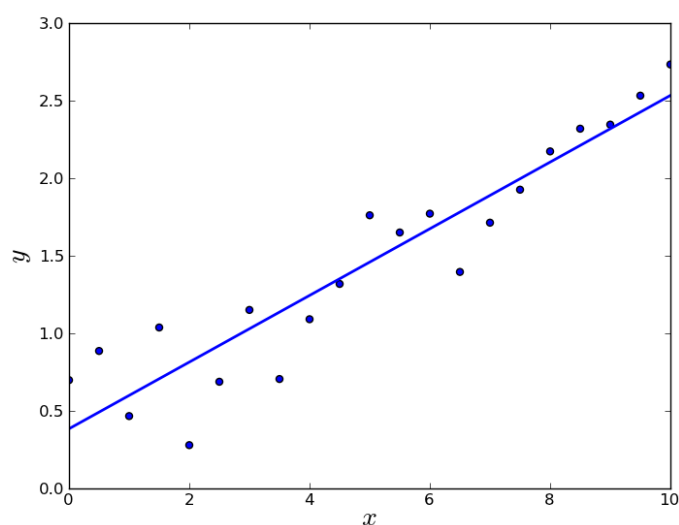
東京 —— 日本のメディアは、今週二日間の日程で日本を訪問したウラジミール・プーチン大統領を熱狂的に報じた。予定より2時間以上遅れて山口県の空港に降り立ったロシアの指導者に、無数のテレビカメラが向けられた。安倍晋三首相の故郷である長門市の学童たちは、この日、学校でボルシチとピロシキを食べたが、これは、この地の温泉を訪れるプーチン大統領への敬意を示すためだ。黒帯の資格を有するロシア指導者は金曜日に東京で、世界最高の柔道学校である講道館を訪れ、オリンピックのメダリスト達と面会し、彼らと柔道の稽古をしてみせた。

3 Google翻訳の正体は「統計的な言語変換」

上記2つの訳文の落差は、何に起因するのか？ それは筆者（つまり人間）が英文を意識しているのに対し、Google翻訳では原文の意味などお構いなしに、ある種の統計的な言語変換を行っていることにある。こうした作業は、統計の専門用語で「回帰分析」と呼ばれる。

回帰分析とは、データにモデルを当てはめる分析手法であり、大量のデータ（いわゆるビッグ・データ）から、何らかのパターンや規則性を見出すためによく用いられる。中でも一番、単純で分かりやすいのは「線形回帰分析」と呼ばれるもので、それは以下の図表1で示される。

【図表1】線形回帰分析では、グラフ上にプロットした点々にフィットする近似直線を求める



(出典) <http://netprophetblog.blogspot.jp/2013/10/local-regression.html>

このグラフでは、たとえば横軸（x 軸）に不動産物件の床面積、縦軸（y 軸）に物件の家賃をとって、そこに不動産会社から貰ってきたデータをプロットする。それらがグラフに記された無数の点々だ。これらの点々（データ）に最もフィットする直線を引く（算出する）作業が「線形回帰分析」である（線形とは直線を意味する）。

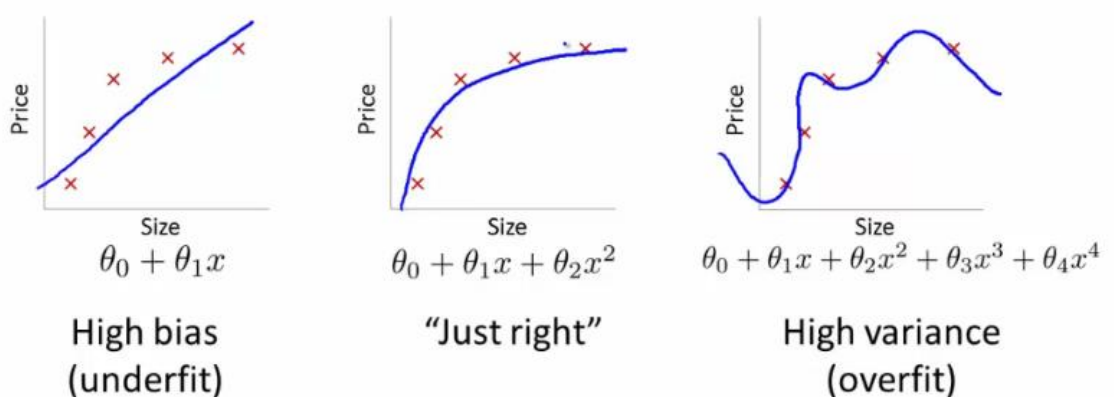
あらかじめグラフにプロットした多数の点々（データ）は、同じく統計、ないしは機械学習の専門用語で「トレーニング・セット」と呼ばれる。線形回帰分析によって、これらの点々にフィットする直線が一旦、引かれてしまうと、トレーニング・セットはお役御免になる。

このようにして得られた直線は「モデル」と呼ばれる（モデルは必ずしも直線とは限らず、何らかの曲線になることもある）。このようにモデル（この事例では直線）が決まると、あとは新しい物件の床面積（x 座標）を入力すれば、この直線に従って、その物件の家賃（y 座標）を計算（推定、予測）することができる。

4 回帰分析の要点とは

ここで注意すべきは、現実世界に存在する多くのデータは、線形回帰分析、つまり単なる直線ではフィットできないことだ。その場合、（前述の通り）直線よりも複雑な、ある種の曲線を使って、点々（トレーニング・セット）にフィットさせる必要がある。その様子を示したのが、以下の図表2である。

【図表2】回帰分析で最適なモデル（近似曲線）を求める作業



（出典）”Machine Learning” by Andrew Ng, Coursera

上の3つのグラフには、いずれも同じ点々（トレーニング・セット）がプロットされている。このうち一番左のグラフでは、これらの点々に対し直線でフィットさせようとしているが、明らかに無理がある。そこで中央のグラフでは、直線ではなく「 x の二乗項」を含む2次曲線でフィットさせようとしている。このように直線（線形）ではなく、ある種の曲線で点々にフィットさせる場合は「非線形回帰分析」と呼ばれる。

明らかに中央のグラフは、これらの点々に上手くフィットしている。が、敢えて難を言えば、全部の点々にぴったり2次曲線が乗っているわけではない。そこで「 x の3乗、4乗項」まで導入して、言わば無理矢理にでも点々の全てに乗る4次曲線を描いたのが、一番右のグラフである。これも当然、非線形回帰分析の一種である。

さて、これら3つのうち（今回のようなケースに関する）正解はどれかというところ、それは中央のグラフ（2次曲線）である。確かに、一番右のグラフに描かれた4次曲線は全ての点々（トレーニング・セット）にピッタリ乗っているが、それは実は回帰分析の目的ではない。つまりトレーニング・セットはあくまで最適なモデル（点々に無理なくフィットする直線や曲線）を求めるためのデータに過ぎないので、実際にはモデルが点々から多少ずれても構わないのだ。

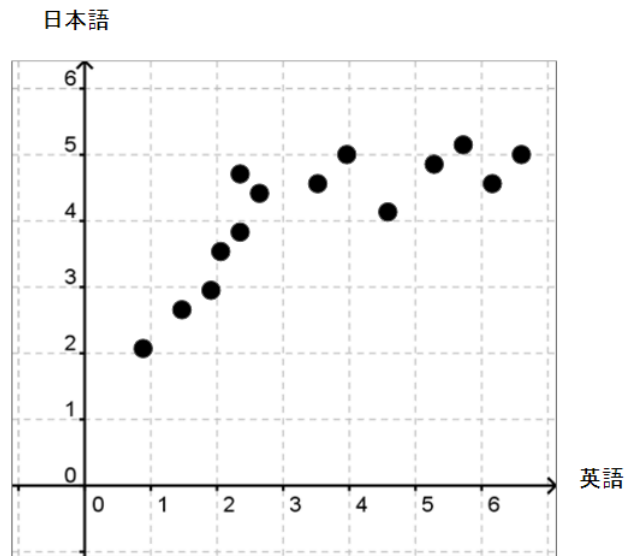
逆に、それらの点々にピッタリ合わせようとしたばかりに、あまりにも不自然な形状の曲線（モデル）を引いてしまうと、そのモデルは本来の目的である「推定・予測」に使えなくなってしまう。このような状況は、機械学習の専門用語で「overfit（過剰適応）」と呼ばれる。上の図表2の場合、一番右のグラフがそれに当たる。逆に一番左のグラフでは、モデル（この場合は直線）が単純過ぎて、点々（トレーニング・セット）とは余りにもかけ離れている。この状態は「underfit（適応不足）」と呼ばれる。

underfitもoverfitも望ましい状態ではない。あらゆる回帰分析の目的は、多数の点々（トレーニング・セット）に過不足なくフィットする「just right（最適）」のモデル（直線や曲線）を導き出すことにある。

5 Google翻訳は大量の対訳文書から学ぶ

さて前準備が大分長くなったが、Google翻訳も基本的には、以上のような回帰分析の手法に基づいている。ここで大前提として、Google翻訳が回帰分析のために使うトレーニング・セットは、グーグル（公式にはアルファベット）がこれまでに収集した大量の対訳文書である。日英翻訳を例にとって考えると、たとえば国連で作成される英語と日本語間の対訳文書などがそれに当たる。これを2次元座標上で直感的に表現したのが、以下の図表3である。

【図表3】 Google翻訳に使われるトレーニング・セット（大量の対訳文書＝グラフ上の点々）



(出典) 筆者作成

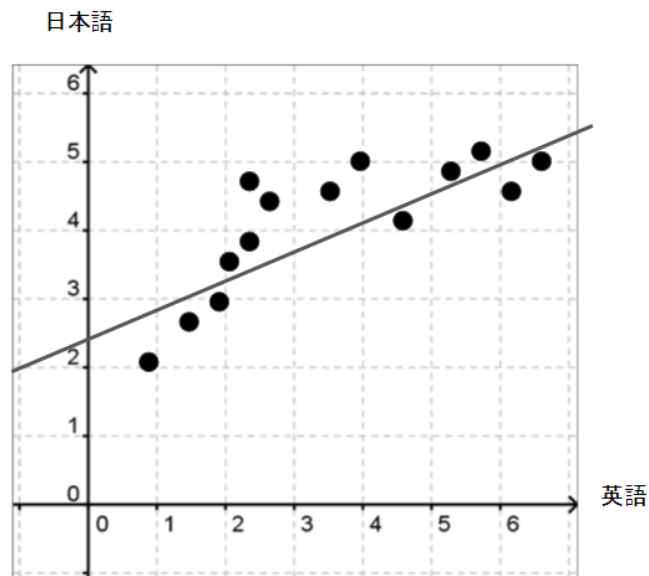
上のグラフにプロットされた多数の点々が、Google翻訳の回帰分析でトレーニング・セットとして使われる日英の対訳文書だ。そしてGoogle翻訳が実際に行っているのは、実はこれらの点々にフィットするモデル（直線や曲線）を求める作業に過ぎないのである。

これはGoogle翻訳がリニューアルされた現在でも同じだ。ただし以前との違いは、最適なモデルを求めるために「ディープラーニング」あるいは「ディープ・ニューラルネット」などと呼ばれる最先端の機械学習技術を導入したことにある。

ディープラーニングは人工知能の一種だが、統計学的な観点からは「現時点において、最も高度で洗練された非線形回帰分析の手法」という位置づけにある。つまりGoogle翻訳は今回、ディープラーニングを導入したとは言え、統計的な回帰分析に頼っている点では以前と同じなのだ。

ただし、その性能はリニューアルの前後では大きく異なる。この違いを見るために、まず従来のGoogle翻訳による回帰分析の様子を示したのが以下の図表4だ（必ずしもモデルが直線になるとは限らないが、ここではそれが比較的単純なモデルあることを示すため、敢えて直線を採用した。あくまで象徴的な意味しかない）。

【図表4】従来のGoogle翻訳はunderfitの状態にあったため、点々（実データ＝対訳文書）とモデルの間に大きな乖離があった



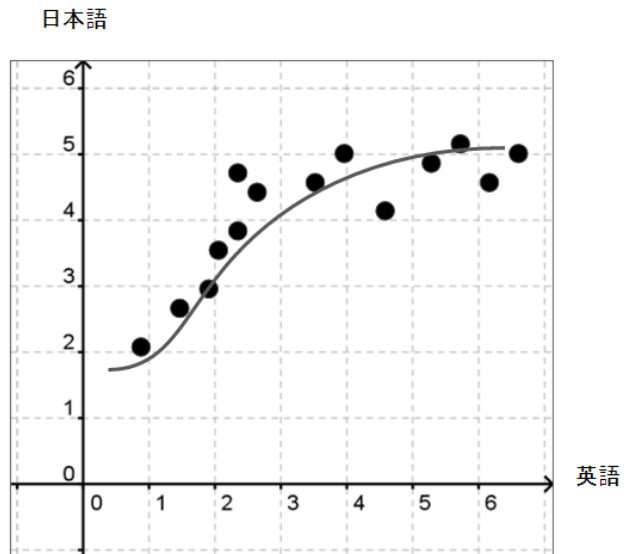
(出典) 筆者作成

図表4に示したように、従来のGoogle翻訳は明らかにunderfitの状態にあった。これまでのGoogle翻訳が（英語、ドイツ語、フランス語など互いに親和性の高い欧州言語族などを除く）日本語と英語などの間で正直使い物にならなかったのは、そのためだ。つまりモデルを求めるための回帰分析として、（恐らくベイズ統計に基づく）比較的単純な方式を採用していたため、モデルがunderfitの状態にあったからなのだ。

6 モデルを最適化した

これに対し今回のリニューアル後は、最も高度な回帰分析の手法であるディープラーニングを搭載したので、just rightの状態へと改善された。その様子を示したのが、以下の図表5だ。

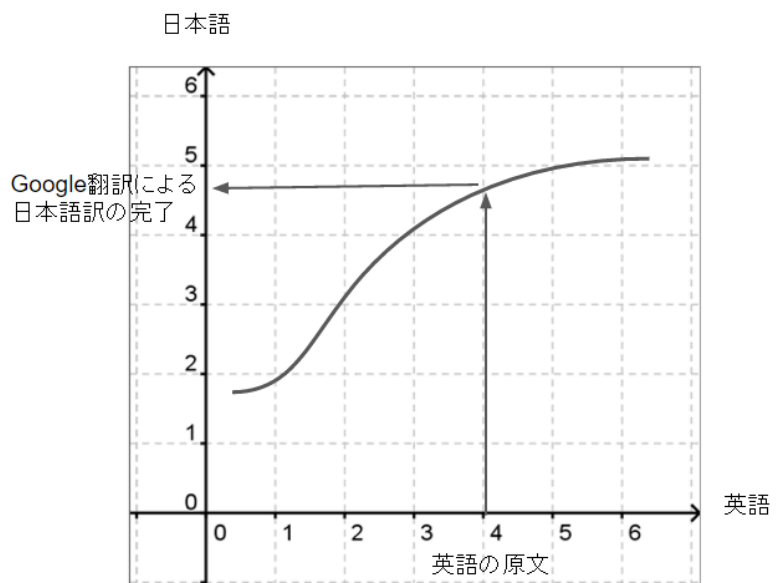
【図表5】リニューアル後は（ディープラーニングによって）実データにフィットするモデルになった



(出典) 筆者作成

一旦、このように最適モデルが作られると、あとは新しい英文（x座標）を入力すると、このモデルに従って日本語訳（y座標）が出力される（図表6）。これこそGoogle翻訳が英・日の翻訳時に、実際行っている作業なのだ。

【図表6】英語の原文を入力すると、事前に算出されたモデル（回帰曲線）に従って日本語訳が出力される



(出典) 筆者作成

ただし正確を期すために断っておくと、以上の説明は実態よりも極端に単純化されている。つまり実際のGoogle翻訳では、約1000種類の言語的な特徴量を変数にして回帰分析のモデルを作成している。

従って、ここまで紹介してきたようなX軸、Y軸による2次元座標で、その様子を表現するのは実際には無理で、厳密には1000次元の座標を使う必要がある。しかし私たち人間が1000次元の座標、つまり1000次元の空間を頭の中に思い描くことは不可能。そこで筆者は、分かりやすく説明するために、2次元の座標で直感的、視覚的に表現した。

こうした回帰分析によって算出されるモデルとは、ある物事と別の物事の相関関係に過ぎない。逆に物事の背後に潜む「因果関係」や「自然法則」、あるいは「真理」や「論理」といったものは一切無視されている。Google翻訳のような自然言語サービスで無視されたのは「言葉の意味」だ。

このやり方に従う限り、いくらディープラーニングのような洗練された手法を導入しようと、実在データ（私たち人間が話したり、書いたりする自然な文章）とモデルの間には、(図表5から見て取れるように) ある程度のズレ（落差）が生じるのは止むを得ない。

そのズレは、Google翻訳の処理対象が短く単純な文章である場合はあまり目立たない。これに対し、たとえば2.で紹介したNew York Times記事のように、比較的高度な長文を入力した場合、両者の落差は激しくなり、結果的にその日本語訳は意味不明の文章になってしまうのだ。

7 天井知らずのスケラビリティとは

しかし問題は、現在ではなく将来のGoogle翻訳の性能である。そこに搭載されたディープラーニングの最大の特徴は、その並外れたスケラビリティにある。スケラビリティとは、システムに入力されるトレーニング・セットのデータ量に比例して、システムの性能がどんどんアップすることを意味する。

従来の機械学習技術では、ある時点でスケラビリティが限界に達し、それ以上はいくらデータを入力しても、システムの性能は頭打ちになってしまった。これに対しディープラーニングでは、入力データ量を増やせば増やすほど、システムの性能が天井知らずに上昇し続けるのだ。

この効果が如実に現れたのが、画像認識や音声認識など、いわゆる「パターン認識」と呼ばれる分野だ。世界中のAI研究者が飽くことを知らず、毎日々々、大量の画像や音声データを入力した結果、今ではディープラーニングのパターン認識力は私たち人間を追い抜き、今もアップし続けている。たとえば以下の2枚の写真、どちらが「チーター」でどちらが「レパード (豹)」か私たち人間には判別し難いが、現在のディープラーニングはそれを正確に判定することができる。

【図表7】どちらがチーターで、どちらがレパードか？（答えは脚注1¹）



出典：ウィキペディア

興味深いのは、以上のようなパターン認識で起きたのと同じ事が、果たしてGoogle翻訳のような自然言語処理の分野でも起きるのか、ということだ。それは確かに私たち人間のように言葉の意味を理解することはできないが、回帰分析のような一種のパターン認識によって翻訳の精度を日々向上させている。特に英仏独など欧州言語族の間では、現時点においても、新聞記事など比較的複雑な長文でも、ほぼ支障なく使えるレベルに達していると見てよい。

8 言葉を理解しなくても、言葉を操ることはできるのか

これに対し、日本語と英語のように言語構造が全く異なる場合は、冒頭で示したようにGoogle翻訳の精度はまだ不十分である。しかし今後、入力するトレーニング・セットのデータ量をどんどん増やしていけば、例によって天井知らずのスケールビリティに従って、いつかは私たち人間に匹敵する翻訳精度に達するかもしれない。仮に、そうなれば、それは「言葉の意味を理解することなく、言葉を自在に操る一種のイディオサバン²」の誕生である。ユーザーは喜ぶが、言語学者や心理学者、哲学者らは渋い顔をするだろう。

もちろん、それとは相反する見方もある。確かに、これまでチェスや将棋、囲碁などは、人工知能のパターン認識技術によって次々と制覇されてしまった。しかし厳密なルールと論理に縛られた、これらボードゲームとは対照的に、私たちが普段使っている言葉は本質的に融通無碍で捉えどころがない（確かに文法というルールも存在するが、これは私たちの日常的な言語活動には、あまり関与していない印象

¹ 左がチーター、右がレパード

² イディオサバン (Idiot Savant)：知的障害や発達障害などのある者のうち、ごく特定の分野に限って優れた能力を発揮する者の症状 出典：ウィキペディア

を受ける)。このため、たとえディープラーニングによって、どれほど回帰分析のモデルを精緻化したところで、実際に私たちが普段行っている言語活動を、このモデルでカバーすることは無理という見方である。

逃げを打つようで恐縮だが、現時点の筆者には、ディープラーニングによる自然言語処理が今後、どこまで発達するのか予想できない。見極めるには、もう少し時間が必要だ。たとえば今から1年後に、より大量のデータを消化したGoogle翻訳の精度がどこまで向上しているか？これを見れば、その将来性は自ずと明らかになるだろう。

【執筆者プロフィール】

氏名：小林 雅一（こばやし まさかず）

所属：KDDI総研

専門：メディア・IT・コンテンツ産業の調査研究

経歴：東京大学大学院理学系研究科を終了後、雑誌記者などを経てアメリカに留学。ボストン大学でマスコミ論を専攻し、ニューヨークで新聞社勤務。慶應義塾大学メディア・コミュニケーション研究所などで教鞭をとった後、現職。

主な著書：

『クラウドからAIへ アップル、グーグル、フェイスブックの次なる主戦場』（朝日新書）

『日本企業復活へのHTML5戦略』（光文社）

『スマートフォンのすすめ一手のひらのクラウドで未来を生きる』（ぱる出版）

『ウェブ進化 最終形 「HTML5」が世界を変える』（朝日新書）

『モバイル・コンピューティング』（PHP研究所）

『社員監視時代』（光文社ペーパーバックス）

『欧米メディア・知日派の日本論』（光文社ペーパーバックス）

ほか多数。