

# 言葉を理解するコンピュータ

## 情報工学系 知能情報コース 岡崎 直観 研究室

岡崎 直観 教授 19XX年〇〇〇〇生まれ。東京大学大学院 情報理工学系 電子情報学専攻 博士課程修了。  
2017年より、東京工業大学 情報理工学系 知能情報コース 教授。



近年、計算機を用いた自然言語処理技術の発展は目覚ましい。機械翻訳など実社会での応用も広がりを  
見せている。岡崎先生の研究室では、特に機械学習を用いた自然言語処理を研究している。本稿では自然  
言語を計算機で解釈する難しさ、機械学習を用いた自然言語処理の技術に触れ、さらに先生が行なってい  
る研究を紹介する。

### 自然言語処理とは

私たちは言葉をどのように理解しているのだろうか。私たちは自然に話せるにもかかわらず、その仕組みを説明することはなかなか難しい。もちろん私たちは仕組みをしらなくても自然と会話ができるから問題はないのだが、コンピュータ相手ではそうはいかない。人間同様の方法で言葉を習得することはできないからだ。コンピュータに直接話しかけても、学校に通わせてみても何も変わらないだろう。自然言語処理とは、人間の言葉を理解することができるコンピュータを作る研究のことをいう。これはいわゆる人工知能の中でも言語に関する知能を研究するものである。特に岡崎研ではテキストや単語を受け取り、それを理解して生成する研究を行っている。

自然言語処理は様々な形で実社会に応用されている。ネット上にある多くの情報はテキストでできており、テキストが大量に集まることで、人々

が何を思っているのか、どのようなことに関心を持っているのか、そして何に注目しているのかを分析することができるのだ。Google翻訳に始まりtwitterのトレンド分析まで私たちの周りの多くのところで自然言語処理の技術は使われるようになっていく。

### 言語処理のその先

さらに区切り方だけでなく、単語同士の修飾関係が分かることも言葉を理解する際に重要な要素となってくる。同じ品詞の組で文ができていたとしても、係り受けが替わるということはよくあることだ。「友達と仙台に行きました」と「盛岡と仙台に行きました」の二つの文はどちらも「名詞＋と＋名詞＋に＋行きました」という形をとっているが、係り受けの形態は異なるものになっている。これらを認識する能力を備え一文を理解できるようになっても、さらに文脈を理解することができ



写真 顔の皮膚血流のイメージ図

なければその意味を完全にくみ取ることは難しい。このように言葉を理解するという作業には様々な事柄が絡んでおり、これらをどのようにコンピュータに教えるかが自然言語処理の課題となっている。

## 自然言語処理と深層学習

自然言語処理の一番の応用例として機械翻訳がある。機械翻訳はコンピュータを利用して、ある言語で書かれた文を別の言語に翻訳する変換を自動で行うものである。ここでは古典的な機械翻訳において、「私は東京に行きました」という文を“I went to Tokyo”に翻訳する過程でどのような処理がなされるかを見ていこう。

翻訳は与えられた文がどのような構造を持っているのか分析するところから始まる。まずは「形態素」という言葉の意味をなす最小の単位に分解し、文の品詞を推測する。次に、単語と単語の関係性を調べることで、どのような構文でできているのかを分析する。この構文解析は、主に文節間の係り受けの構造を発見することを目的として行われる。

こうして文の意味を解釈したところで、次に翻訳するステップに移る。ここでは数百万文にもなる日本語の文と英語の対訳が収録されている「訓練データ対訳コーパス」を用いて、翻訳モデルと言語モデルという二つの統計的なモデルを作ることによって翻訳を実行する。翻訳モデルは単語の訳され方の傾向に関するモデルであり、例えば「私は」は“I”に、「東京」は“Tokyo”に、「行った」は“went to”に訳される傾向があるなどの情報を持っている。それに対し、言語モデルは単語の並び方

の傾向に関するモデルであり、“I”の後ろには“Tokyo”よりも“went to”が出てきやすいといった情報を持っている。

このとき入力が $x$ 、出力が $y$ となるような確率 $P(y|x)$ を構築する。つまり、 $y$ には“I went Osaka”や“I go to Tokyo”, “He went to Tokyo”などの様々な候補を考え、その中で一番高い確率となるものを出力するというわけだ。 $p(y|x)$ はこの二つのモデルの組み合わせであり、 $x$ と $y$ の単語それぞれがどのくらい対応するのか、 $y$ の単語の並びがどれほど英語として自然なのか、という2つの特徴をもとに一番よい翻訳を選ぶのだ。

まとめると、まず文の要素をわけて、コーパスから学習することで作られる翻訳モデルと言語モデルを用いて、文の要素それぞれを翻訳先の言語におきかえ正しい順番に並べる。この作業が機械学習に基づく、古典的な機械翻訳の原理となっている。

## 自然言語処理と深層学習

近年では生物の脳機能の特徴を取り入れた深層学習と呼ばれる手法が自然言語処理でも使われ始め、飛躍的な進展をもたらしている。機械翻訳の精度を示す指標の1つであるBLEUを見ても、従来の手法では20年かけて20しか上がらなかったところを、深層学習を使うことで4年間で35まで上げることに成功した。岡崎研究室での最近の研究においても深層学習を手法として用いたものを中心となっている。

深層学習とは、一言でいうと「生物の神経回路網を模倣した大規模な神経回路網を用いて学習を行う手法」である。深層学習のモデルでは大量の入力と出力をベクトルとして受け取り、その対応付けを大規模な神経回路網で学習していく。学習に際しては「誤差逆伝搬法」と呼ばれる合成関数の微分をもとにしたパラメータの探索法を使って、与えられた入力と出力の間の関係をよく近似するパラメータを設定することになる。学習が完了したモデルは入力されたベクトルに対する出力として最も確率の高いベクトルを出力するようになる。したがって深層学習を活用するには入力も出

力もベクトルでなければならないため、自然言語処理の分野において深層学習を使うには言葉をベクトルとして上手に表現する必要がある。

ここで「単語の意味は、周囲の単語によって形成される」とする「分布仮説」と呼ばれるものがある。英語の穴埋め問題を想像してもらいたい。英語の穴埋め問題は、空欄になっている部分を前後の文章から予測して解く問題であるが、この問題が解けるのは単語の意味が周りの単語から推測できるからなのではないだろうか。一般に分からない単語が多少あったとしても、前後の単語さえ知っていれば問題なく読めるだろう。この「分布仮説」に基づき、単語を高次元の実数値ベクトルで表現することを「分散表現」という。「分散表現」ではその「分布仮説」に基づき、単語の意味をその周囲の文脈の分布で表現する。つまり意味を考えたい単語の前後に出現する単語の分布をコーパスからつくり、その分布をもとに単語に一つのベクトルを対応させるというわけだ。

単語をベクトルと表現することは、深層学習が活用できること以外にも、様々な情報を取り出しやすくなるというメリットがある。単語と単語の類似度を測るにはベクトル同士の類似度を測ればよい。例えばベクトルのなす角と内積の関係を使いベクトル通しのコサインを計算すれば、類似度を0から1の範囲で求めることができる。ほかにも単語と単語の足し算、引き算も考えることができる。

「香川県にとっての松山市は宮城県における何であるか?」といったことを考えるには、「松山市ー香川県+宮城県」のように各単語のベクトルの足し引きをすればよい。実際にこのベクトルと類似度の高いものを探すと、宮城県の県庁所在地である仙台市が出てくる。また「ManにとってのKingはWomanにとっての何であるか?」という問いに対しても、「KingーMan + Woman」のように各単語のベクトルの足し引きを行えばよく、この演算結果のベクトルと最も関連度が高いベクトルとして、「Queen」というベクトルが出力される。単語間の類似度だけでなく、このような「アナロジー（類推）」を行える点が分散表現の画期的な点であり、言語処理の分野における深層学習の活用が進んだ1つの要因ともされている。

## 単語文脈表（行列）

各単語の周辺に出現する単語（文脈）がどのように出現するのか、大量の文書を使って出現分布を計測する

単語	have	nom	tree	juice	ride	speed	read
apple	36	14	72	57	3	0	1
lemon	108	14	92	86	0	1	2
car	578	284	3	2	37	44	3
train	291	94	3	0	72	43	2
book	841	201	0	0	2	1	338

$m_{ij} = \#(i, j)$ : 単語*i*と文脈*j*の共起頻度  
(例: trainの周辺にtreeが3回出現)

“apple”の意味を表すベクトル (単語ベクトルに  $M_i$  で表す)

図2. 分散表現

だが単語に意味が分かるだけでは翻訳や要約はできない。1文全体の意味を理解する必要がある。そこで次に各単語のベクトルから”This is a pen”という文全体の意味を表すベクトルを求めたい。このとき用いるのが深層学習の手法の1つである「再帰型ニューラルネットワーク (RNN)」である。再帰型ニューラルネットワークでは、“This”の意味を表すベクトルに行列をかけて”This”のベクトルからどのくらいの意味を取り出すのかというのを行列  $W^{\{hv\}}$  をかけることで計算する。次に行列  $W^{\{hv\}}$  をかけることで抽出されたベクトルのうち、どのくらいの部分を次の単語の状態に引き継ぐかを行列  $W^{\{hh\}}$  を掛け合わせることで計算する。次に“is”が来たときも、“is”の意味をどの位読み込むのかということを計算するための行列  $W^{\{hv\}}$  と、前の状態の意味をどれくらい引き継ぐのかということを計算するための行列  $W^{\{hh\}}$  の両方の行列をかけて足すことで行列  $h_2$  を作る。行列  $W^{\{hh\}}$  と  $W^{\{hv\}}$  は文全体で同じものを継承し、同じサイクルを繰り返すことで、“This is a pen”の意味を表すベクトルが作られる。このように文章自体を一つのベクトルに変換する深層学習のアーキテクチャのことを機械学習の世界ではエンコーダをいう。

さらにそのベクトルに新たな行列をかけあわせてソフトマックス関数を用いることで各タスクに対応したベクトルを取り出すことができる。ソフトマックス関数はベクトルからベクトルへの写像を行うことで、出力されるベクトルの成分をすべて足すと1になるように変換する関数である。

## 分散表現 (ベクトル) の合成： 再帰型ニューラルネットワーク

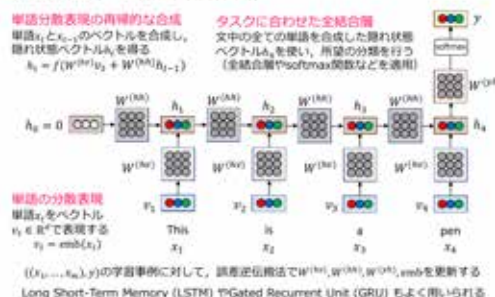


図3 RNN

すなわちベクトルの各成分があたかも確率であるかのように解釈できるようになるのだ。例えば自然言語処理により評判分析を行い、その文が positive か negative か neutral かを判断したい場合には、positive だった場合は一次元目の要素が反応し negative だった場合には3次元目のベクトルの要素が反応するようにすればよい。

翻訳を行う場合もまずは文章をベクトルに変換するエンコーダを用いてその文を表すベクトルを抽出する。”this is a pen”を「これはペンです」と翻訳する場合、まず”this is a pen”という文の意味を表すベクトルを出力する。そしてこのベクトルをもとに、翻訳結果の単語を予測する再帰型ニューラルネットワークを作る。このアーキテクチャをデコーダーという。デコーダーから「ペン」と出力されたら、pen が「ペン」に訳し終えられたことになる。その次の単語を予測しようとする”と「を」が予測される。これを繰り返していくことで順序よく訳されていく。これがニューラルネットワークを使った翻訳アーキテクチャになる。

なお学習を行う際は古典的な機械学習による翻訳と同様に、“I have a pen”と「ペンを持つ」、「It’s fine today」と「今日はいい天気ですね」といった英語と日本語の翻訳の関係にある文のペアを大量にニューラルネットワークに読み込ませる。そして入力された英語に対して正しい日本語を出力するように各行列の要素を学習する。

ここまでの翻訳の過程を整理すると、まずは文の単語の意味をベクトルとして表現し、次に文全体の意味を表すベクトルを再帰型ニューラルネット

ワークを用いて作る。(エンコーダ)そして文全体のベクトルを翻訳先の言語にある単語に分解する別の再帰型ニューラルネットワークを用いてつくる。(デコーダ)エンコーダとデコーダの二つにより翻訳を実現しているのだ。このような翻訳モデルを「エンコーダ・デコーダモデル」という。

このエンコーダ・デコーダモデルの問題点として、入力される文の長さに関係なく、常に一定の大きさの一つのベクトルに変換することがある。しかしそれでは文が長くなればなるほど、一つのベクトルに含まれる情報量が多くなり、全部の情報を保持できなくなってしまった。エンコーダ・デコーダモデルを用いた機械翻訳において長い文になると翻訳精度が落ちてしまうことが難点であった。

この問題を解決する方法として、「注意機構」というものが作られた。この手法は少し難しいものであるが、岡崎先生がかみ砕いて説明してくださった。その話をもとに骨格を見ていくことにする。この注意機構ではエンコーダ・デコーダモデルのように、入力されたベクトルをまず1つのベクトルに圧縮した後に翻訳を行うということはない。文全体の意味を表すベクトルから各単語に意味を抽出する際に、もとの文の単語のベクトルも同時に参照するのである。確かに人間が翻訳するときも一文全体の意味を取ってきてそれをもとに日本語を作ろうとするのではなく、次の単語は元の文のどの単語に当たるのかを考え、次の単語を訳すのに必要な部分のみに注目して訳しているだろう。

例えば”This is a pen”を注意機構を用いて日本語に訳すことを考えよう。まずはエンコーダを用いて文全体のベクトルを作りだす。そしてそのベクトルから日本語の「これは」までを訳し終え、次はどのような日本語が来るべきかをコンピュータは考えているとする。このとき入力側の“This”, “is”, “a”, “pen”といった各単語を表すベクトルを見て、次は原文のうちどの要素が必要かを判断するニューラルネットワークをはさむ。そのニューラルネットワークは次の単語には”pen”という情報がたくさん必要ということを判断し、そのベクトルの成分を中心に文のベクトルから抽出してくる。そしてとってきたベクトルを

もとに出力されるべき単語を判断し、「これはペン」と訳すことになる。まとめると、元の各単語のベクトルをどのくらい混ぜてとるのかということを決める注意機構を、エンコーダ・デコーダモデルのニューラルネットワークに組み込むことで長文でも精度を失うことなく、翻訳できるようになったのだ。

注意機構により翻訳の精度はかなり良くなったという。注意機構を使わなかった場合、文の長さが長くなればなるほど翻訳の指標であるBLEUがどんどん悪くなるが、注意機構を使うと文の長さが長くなっても精度が落ちなくなった。これにより自然言語処理の研究に深層学習を取り入れる流れが進み、技術革新のスピードが一気に上がったという。

## 言語処理のその先

深層学習がもたらした変革として言葉や文の意味をベクトルで表現することで自然言語処理や画像処理などのベクトルで表現される全てのものと繋がっている状態になったことがある。これにより従来の分野やタスクの垣根を越えて、新しいことができるようになった。例えばどのような写真であるかの説明文を生成することができるようになった。これを行うにはまず与えられた画像の特徴を表すベクトルに深層学習を用いて変換する。そしてその出力されたベクトルを今度は新たな入力として、キャプションの単語列を予測する問題を深層学習に解かせる。〈例〉

また岡崎研では新聞記事から見出しを自動で作る研究を時事通信社と共同で研究を行っている。岡崎研で作った見出し生成ソフトの新しい点は見出しの長さを指定できることである。これにより最終的に出てくる見出しの長さを30文字でやりなさい、12文字でやりなさいというように指定することができる。

最近では汎用的な深層学習ツールが普及したことにより、ネットワーク構造が複雑なニューラルネットワークをつくる際も、そのネットワークの形を記述するだけで全部自動でパラメータの推定

などを行うことができ、容易に学習エンジンを作れるようになった。これにより、研究開発のスピードは以前より急速に上がっているという。

その一方で、モデルの解釈性に乏しいというニューラルネットワークの問題点は自然言語処理においても現れているという。すなわち深層学習による推論では、どうしてそのように結論付けたのかの根拠を示すのが難しいのである。また公開される論文はパラメータをチューニングした結果うまくいったもののみがシェアされるため、その結果にたどり着くためにどのような過程を踏んだのかがわかりにくくなっている。そのために再現性が低くなり、研究がやりにくくなっていると岡崎先生は語った。

このような問題を抱えているものの、自然言語処理の研究は私たちの生活の豊かさに直結する重要な研究である。これはgoogleの翻訳サービスやalexaやsiriといった音声処理などから明らかなことである。これからの自然言語処理の発展に期待したい。

## 学部生の人たちへ

岡崎先生は最後に『大学では授業だけをこなすのではなく、自分が興味があることを見つけてどんどん勉強して行ってほしい。今は本やサイトも充実しているので自分で勉強する敷居は低くなっているはずです。』と語ってくれた。岡崎先生も言語処理に興味がある人向けに『言語処理100本ノック』を公開している。興味のある人はこのようなツールを積極的に使っていくといいだろう。

## 執筆者より

本稿の執筆にあたり、林先生には二度にわたる取材や資料提供、校正など、様々なお力添えを頂きました。記事に入らなかったことまで、非常に興味深いお話をたくさん伺うことができました。お忙しい中、取材を快く引き受けてくださった林先生に心より御礼申し上げます。ありがとうございました。

(石川 智貴)