

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 東京大学大学院 情報理工学系研究科 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、自然言語処理技術の発展は目覚ましく、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。本稿ではまず自然言語処理の技術に触れ、次にコンピュータはどのように翻訳を行っているかを説明し、最後に深層学習を用いた翻訳や見出し生成の研究を紹介する。

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。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。もちろん私たち人間は仕組みを知らなくても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だ。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。岡崎研では特に文章から見出しや要約を生成する研究を行っている。

自然言語処理はさまざまな形で実社会に応用されている。例えば、Google翻訳などの機械翻訳によって私たちは知らない言語で書かれた文章を読

むことができる。また、Twitterのトレンド分析によって人々が今何を思っているのか、どのようなことに興味を持っているのか、そして何に注目しているのかを知ることができる。このように私たちの周りの多くの場面で自然言語処理の技術は使われているのだ。

言葉を理解することの難しさ

そもそも言葉を理解するとはどういうことだろうか。**次の図**を見てもらいたい。これはフィンランド語のWikipediaの「日本」の記事の導入部だ。これを見て、何が書かれているかわかるだろうか。どこかの国の言語であることは推測できても、内容を理解することは難しいだろう。

これを理解することができない理由の一つは単語の意味がわからないことだ。**私たちはフィンランド語の単語を知らない**ので、意味を取ることは難しい。これはコンピュータにとっても同じこと

Japani (jap. 日本, *Nihon* tai *Nippon*, virallisesti 日本国, *Nihon koku* tai *Nippon koku*) on saarivaltio Tyynessämeressä Itä-Aasiassa. Se sijaitsee Kiinan, Korean niemimaan ja Venäjän itäpuolella, ulottuen pohjoisesta Ohtanmereltä etelään Itä-Kiinan merelle. Maan nimen muodostavat kirjoitusmerkit tarkoittavat "auringon juurta" tai "auringon lähde", minkä vuoksi Japania kutsutaan joskus "nousevan auringon maaksi". Japanin pääkaupunki ja suurin asutuskeskus on Tokio.

図1 フィンランド語の例

フィンランド語版Wikipediaの「日本」の記事の導入部

だ。ただ、現代のコンピュータはハードディスクもメインメモリもたくさんの容量を持つので、すべての単語を教えることは可能かもしれない。

しかし、単語の意味がわかるだけでは言葉を理解したとは言えない。というのも、文を正しく分割できるということも言葉を理解するためには重要な要素だからだ。辞書に存在する単語の区切り方に限ったとしても区切り方のパターンの何通りにもなり、単に単語を暗記しているだけではわからない。例えば「～になった」という言葉は「元気になった」という意味で使われることもあれば、「責任をになった」という「担う」という意味で使われることもある。

さらに、単語同士の修飾関係も言葉を理解する際に重要な要素だ。同じ品詞の組み合わせでできた文でも係り受けが変わることはよくある。例えば、「友達と仙台に行きました」と「盛岡と仙台に行きました」の2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。また、文脈を理解することができなければ文章中で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さまざまな事柄が絡みあっていて、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研究の要なのである。

機械翻訳における自然言語処理

自然言語処理の応用の最たる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。ここでは伝統的な、統計的

機械翻訳において「私は東京に行きました」という文が"I went to Tokyo."に翻訳される過程でどのような処理がなされるかを見ていこう。

まずは与えられた文を「形態素」という言葉の意味をなす最小の単位に分解し、その品詞を推測する。次にそれらの関係性を調べ、文がどのような構文からできているのかを分析する。この構文解析により係り受けの構造を発見する。(図2)こうしてコンピュータは文の意味を解釈する。

次に翻訳するステップに移る。ここでは、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用いる。それぞれのモデルは言語に関する統計的な情報を持っている。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」"went to"に訳される傾向があるという情報を持っている。一方、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、例えば"I"の後ろには"Tokyo"よりも"went to"の出現頻度が高いという情報を持っている。

これらのモデルは数百万もの日本語の文と英語の文の対訳が収録された「訓練データ対訳コーパス」を用いて作成される。

この二つのモデルを組み合わせることで、無数の候補の中から最もよい翻訳を選び出すことができる。つまり、多くの候補 y について、入力された文 x に対して候補が y である確率 $p(y|x)$ を求め、2つのモデルにどの程度適合しているかを判断することができるのだ。これにより x と y の単語が互いに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つの観点から最もありうる候補を選ぶことができる。

つまり、統計的機械翻訳ではまず与えられた文を要素ごとに区切り、次に文の要素それぞれを翻

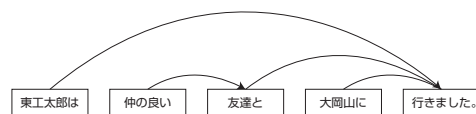


図2 係り受け解析の例

それぞれの要素の修飾関係を矢印を用いて現している。

訳先の言語の要素に置き換え、最後にそれを翻訳先の言語においてふさわしい順番に並べる。この作業が伝統的な機械翻訳の原理だ。

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に飛躍的な進展をもたらしている。その一例として機械翻訳の精度を示す指標であるBLEUを見ると、**伝統的な機械学習では20年間で20の上昇にとどまっていたが、深層学習の導入によりわずか4年間で35もの上昇を見せた**。最近では岡崎研究室でも深層学習を用いた研究が中心になっている。

深層学習とは、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関係をよく近似するパラメータを設定する。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。

深層学習を用いた自然言語処理をするにあたって、言葉を多次元のベクトルとして上手に表現することは重要だ。そこで用いるのが分布仮説だ。この仮説は「単語の意味は、**周囲の単語によって形成される**」というものだ。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ではないだろうか。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統計的に分析し、その分布をもとに単語に一つのベクトルを対応させるのだ。(図3)

単語をベクトルとして表現することには深層学習が活用できること以外にも、さまざまな情報を取り出しやすくなるというメリットがある。例えば、ベクトルのなす角と内積の関係を使い、ベクトル同士のなす角のコサインの絶対値を計算すれ

単語	文脈 (周辺に出現する単語)					
	enjoy	play	watch	ball	listen	read
soccer	249	412	128	50	3	2
tennis	252	563	137	57	0	5
piano	180	506	3	2	58	3
violin	113	483	3	0	72	1
book	130	3	2	0	2	398

…… “soccer” の意味を表すベクトル

図3 分散表現のイメージ

単語の周辺に出現する単語を調べることで、その単語の意味をベクトルとして表すことができる。

ば、類似度を0から1の範囲で求めることができる。

また、ベクトルの足し算・引き算によって単語の類推（アナロジー）を行うことができる。例えば、「香川県にとっての高松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、「高松市－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－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類似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。このような類推ができること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一つの要因である。

RNNによる文の意味の理解

深層学習の手法の一つであるリカレントニューラルネットワーク(RNN)を用いることで、文全体の意味を理解した翻訳や要約が可能となる。

例えば、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ることを考える。このとき、リカレント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 $W^{\{hv\}}$ をかけることで、"This"のベクトルからどのくらいの意味を取り出すのかを計算する。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 $W^{\{hh\}}$ をかけあわせることで計算する。次に"is"が来たときも、"is"の意味をどのくらい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。この行列 h_2 はその時点での文の意味を表すものである。行列 $W^{\{hh\}}$ と $W^{\{hv\}}$ は文全体で同じものを用い、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をエンコーダという。

(図4)

その文全体の意味を表す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作作用させ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。すなわち、ベクトルの各成分を確率とみなせる量に変換することができる。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1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リカレント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という。デコーダから「ペン」と出力されたら、"pen"が「ペン」に訳されたということになる。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翻訳されていく。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である。

なお学習を行う際は伝統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。まずはエンコーダを用いて文全体の意味を表すベクトルを作る。その後デコーダを用いて、そのベクトルから翻訳先の言語の文を生成する。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

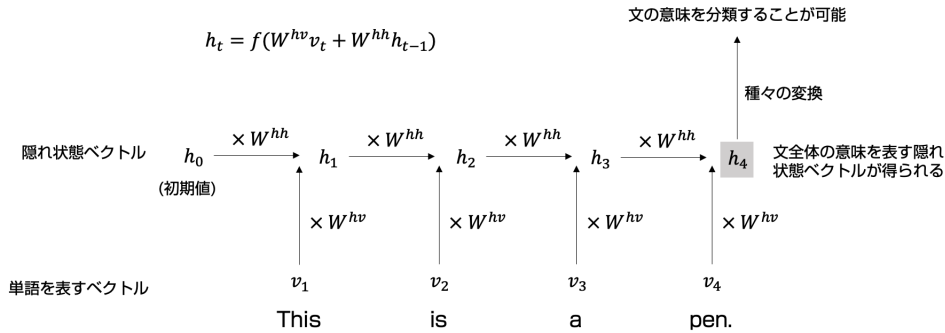


図4 リカレントニューラルネットワークのイメージ図

RNNを用いて文全体の意味を表す一つのベクトルを得ることができる。

て、入力される文を、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。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、すべての情報を保持できないことがある。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。この注意機構では翻訳元の文全体の意味を表すベクトルだけでなく、各単語の意味を表す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文全体の意味を取ってきてそれをもとに翻訳しよ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.」という文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意味を表す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、次はどのような日本語の単語が来るべきかを考える。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の中から、次は原文のうちのどの要素が必要なのかをニューラルネットワークを用いて判断する。そのニューラルネットワークは次の単語には「pen」という情報が必要だということを判断し、「pen」の情報を文のベクトルから抽出する。そしてその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。すなわち、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精度の指標であるBLEUが悪くなるが、注意機構を使うと文の長さが長くなったとしても精度が落ちることはなかった。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

深層学習研究の進展がもたらした成果の一つに分野間の垣根が取り払われたことがあげられる。言葉だけでなく画像や音声もベクトルで表現するようになったので、それらを同じ土台で取り扱うことができるようになったのだ。このような分野を横断した応用の一つが写真を説明する文章(キャプション)の生成だ。これを行うにはまず与えられた画像を、その特徴を表すベクトルに変換する。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。(図5)

また岡崎研では時事通信社と共同で新聞記事か

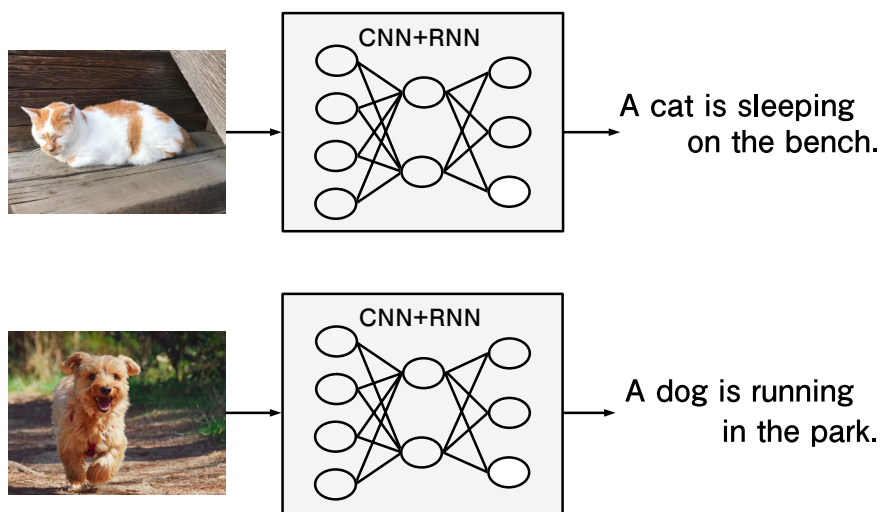


図5 キャプション生成のイメージ

深層学習を用いて、写真を説明する文章（キャプション）を生成することができる。

から見出しを自動で作る研究を行っている。岡崎研で作った見出し生成ソフトの斬新な点は見出しの長さを指定できることである。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できるようになったという。

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作る際も、そのネットワークの形を記述するだけでパラメータの推定などを自動で行うことができ、容易に学習エンジンを作ることができるようになった。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。

その一方で、近年の深層学習の課題としてモデルの解釈性に乏しいことが挙げられる。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、再現性の低下も問題となっている。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のだ。

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。これは自動翻訳や音声認識など

の自然言語処理を用いたサービスの発展からも明らかである。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、自分が興味のあることを見つけてどんどん勉強してほしい。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。岡崎先生は自然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、岡崎先生にはさまざま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岡崎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)